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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 September 2020 (24.09.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/191102 A1

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

(21) International Application Number:
PCT/US2020/023441

(22) International Filing Date:
18 March 2020 (18.03.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/820,109 18 March 2019 (18.03.2019) 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).

(72) Inventors: **ZHANG, Feng**; c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **ALTAE-TRAN, Han**; c/o 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US).

(74) Agent: **MILLER, Carin R.** et al.; Johnson, Marcou, Isaacs & Nix, LLC, 27 City Square, Suite 1, Hoschton, Georgia 30548 (US).

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

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

(54) **Title:** TYPE VII CRISPR PROTEINS AND SYSTEMS

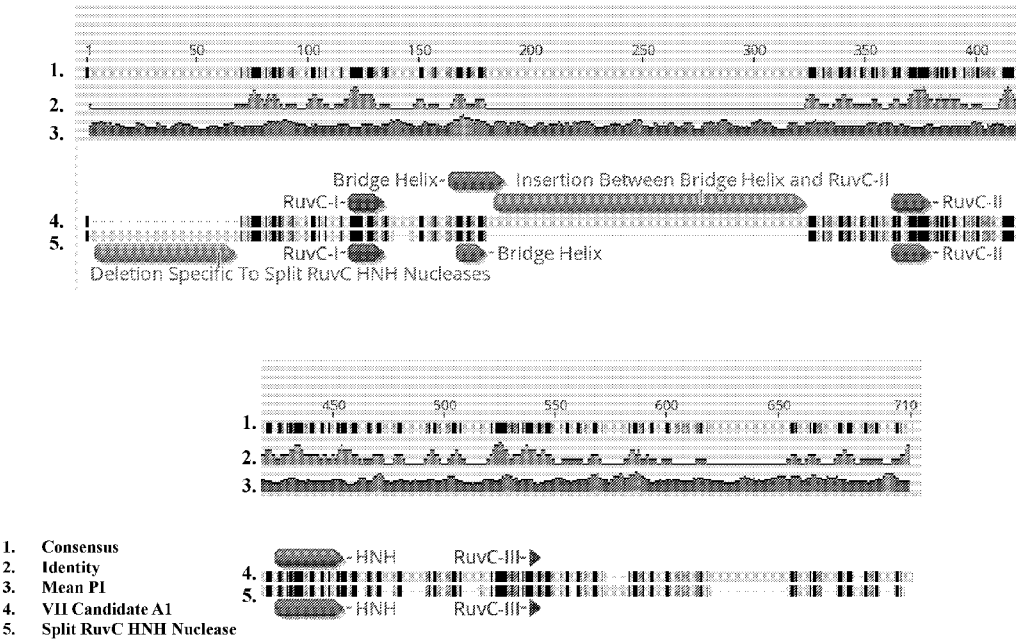


FIG. 4

(57) **Abstract:** The present application provides systems, methods and compositions used for targeted gene modification, targeted insertion, perturbation of gene transcripts, and nucleic acid editing. Novel nucleic acid targeting systems comprise components of Type VII Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems and transposable elements.

[Continued on next page]

WO 2020/191102 A1

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))*

TYPE VII CRISPR PROTEINS AND SYSTEMS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/820,109, filed March 18, 2019. The entire contents of the above-identified application(s) is/are hereby fully incorporated herein by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under Grant No.(s) HG009761, MH110049 and HL141201 awarded by the National Institutes of Health. The government has certain rights in the invention.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0003] This application contains a sequence listing filed in electronic form as an ASCII.txt file entitled BROD_4550WP_ST25.txt, created on March 18, 2020 and having a size of 56,961 bytes. The content of the sequence listing is incorporated herein in its entirety.

TECHNICAL FIELD

[0004] The subject matter disclosed herein is generally directed to systems, methods and compositions used for targeted gene modification, targeted insertion, perturbation of gene transcripts, nucleic acid editing. The nucleic acid targeting systems described herein can include components of Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) systems and/or transposable elements.

BACKGROUND

[0005] Recent advances in genome sequencing techniques and analysis methods have significantly accelerated the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. Precise genome targeting technologies are needed to enable systematic reverse engineering of causal genetic variations by allowing selective perturbation of individual genetic elements, as well as to advance synthetic biology, biotechnological, and medical applications. Although genome-editing techniques such as designer zinc fingers, transcription activator-like effectors (TALEs), or homing meganucleases are available for producing targeted genome perturbations, there remains a need for new

genome engineering technologies that employ novel strategies and molecular mechanisms and are affordable, easy to set up, scalable, and amenable to targeting multiple positions within the eukaryotic genome. This would provide a major resource for new applications in genome engineering and biotechnology.

[0006] The CRISPR-Cas systems of bacterial and archaeal adaptive immunity show extreme diversity of protein composition, genomic loci architecture, and system function, and systems comprising CRISPR-like components are widespread and continue to be discovered. Novel Class 1 multisubunit effector complexes and Class 2 single-subunit effector modules may be developed as powerful genome engineering tools.

[0007] Class II systems discovered to date generally fall into three types: II, V, and VI. Effectors containing a RuvC-like nuclease domain and an HNH nuclease are categorized as Type II. Effectors in which RuvC is the only recognizable enzymatic domain (e.g., lacking an HNH nuclease) are categorized as Type V. Effectors with two HEPN RNase domains are categorized as Type VI. Signature genes include cas9 for Type II, cas12 for Type V, and cas13 for Type VI.

[0008] There exists a pressing need for alternative and robust systems and techniques for targeting nucleic acids or polynucleotides (e.g. DNA or RNA or any hybrid or derivative thereof) with a wide array of applications.

[0009] 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

[0010] In certain example embodiments, the invention provides a non-naturally occurring or engineered composition comprising a Type VII CRISPR effector protein and a crRNA comprising a direct repeat sequence and a guide sequence that is designed to hybridize to a target nucleic acid sequence, whereby there is formed a CRISPR complex comprising the Type VII CRISPR effector protein complexed with the crRNA. In certain embodiments, the crRNA comprises a DR sequence located 5' to the guide sequence. In certain embodiments, the crRNA comprises a DR sequence located 3' to the guide sequence. In certain embodiments, the target nucleic acid comprises DNA. In certain embodiments, the target nucleic acid comprises RNA. In certain embodiments, the composition further comprises a tracrRNA. In certain embodiments, the tracrRNA is fused to the DR of the crRNA.

[0011] In certain embodiments, the guide is designed to hybridize to a target nucleic acid in a prokaryotic cell. In certain embodiments, the guide is designed to hybridize to a target nucleic acid in a eukaryotic cell or organelle. In certain embodiments, the Type VII effector protein comprises one or more trafficking domains, such as but not limited to one or more nuclear localization signal (NLS) or one or more nuclear export signal (NES).

[0012] In certain embodiments, the Type VII effector protein is a nickase. In certain embodiments, the Type VII effector is catalytically inactive.

[0013] In certain embodiments, the Type VII CRISPR effector protein is associated with one or more functional domains. In one such embodiment, the functional domain comprises cleavage activity whereby a target nucleic acid can be cleaved. In another such embodiment, the functional domain comprises an activator or inhibitor of transcription or translation. In yet another embodiment, the functional domain comprises an adenosine deaminase, or catalytic domain thereof, or cytidine deaminase, or catalytic domain thereof, and may further comprise one or more mutations that increase activity or specificity of the adenosine deaminase relative to wild type or modify the functionality of the adenosine deaminase relative to wild type.

[0014] In certain example embodiments, the invention provides a vector system for expressing or delivering a Type VII CRISPR system. In an embodiment, the vector system comprises a regulatory element operably linked to a nucleotide sequence encoding a Type VII CRISPR effector protein and a regulatory element operably linked to a nucleotide sequence encoding a Type VII CRISPR system crRNA. In an embodiment, the crRNA is fused to a tracrRNA. In an embodiment, the vector system further comprises a regulatory element operably linked to a tracrRNA. The nucleotide sequence encoding the Type VII effector can be codon optimized for expression in a target cell, including but not limited to a eukaryotic cell, a human cell, a plant cell, or a yeast cell. In certain embodiments, the vector system is comprised in one or more vectors. In certain embodiments, the vector system is comprised in a single vector. In certain embodiments, the vector system comprises one or more viral vectors, including but not limited to retroviral, lentiviral, adenoviral, adeno-associated or herpes simplex viral vectors.

[0015] The invention provides a method of modifying nucleic acid sequences associated with or at a target locus of interest, the method comprising delivering to said nucleic acid or locus a non-naturally occurring or engineered composition comprising a Type VII CRISPR-Cas effector protein and one or more CRISPR system nucleic acid components, wherein the

CRISPR-Cas effector protein and one or more CRISPR nucleic acid components is designed to form a complex, and wherein the CRISPR-Cas effector protein - nucleic acid complex is capable of promoting the modification of the sequences associated with or at the target locus of interest. In one embodiments, the modification comprises the introduction of a strand break. In another embodiment, the modification comprises a base substitution. In another embodiment, the modification comprises modulating gene expression, including but not limited to, increasing or decreasing expression. In another embodiment, the modification comprises a change in methylation. In certain embodiments, the target nucleic acid comprises DNA. In certain embodiments, the target nucleic acid comprises RNA. In certain embodiments, a non-target nucleic acid is collaterally modified. In certain embodiments, the target nucleic acid is in a prokaryotic cell. In other embodiments, the target nucleic acid is in a eukaryotic cell.

[0016] In certain example embodiments, the invention provides cells, cell lines, tissues, model plants, and model animals comprising a target nucleic acid of interest, wherein the target has been modified by an engineered Type VII CRISPR system. In certain embodiments, the cell a prokaryotic cell. In other embodiments, the cell is in a eukaryotic cell. In non-limiting examples, the cellular modification comprises altered expression of at least one gene product; increased expression of at least one gene product, decreased expression of at least one gene product, or alteration of the expressed gene product such as by mutation.

[0017] In certain example embodiments, the invention provides a gene product from the cells, cell lines, tissues, model plants, or model animals.

[0018] In certain example embodiments, the invention provides novel CRISPR proteins and systems described and exemplified herein. The novel CRISPR systems and proteins may be referred to as Type VII CRISPR effector proteins and systems. Type VII CRISPR effector proteins comprise a RuvC-like nuclease domain. In certain embodiments, nuclease activity is present while in other embodiments, the nuclease is inactive. In certain embodiments, a Type VII CRISPR effector protein comprises an HNH-like domain. In certain embodiments, the HNH-like domain comprises nuclease activity. In other embodiments, the HNH-like domain lacks nuclease activity. In certain embodiments, the CRISPR effector proteins comprise a split RuvC domain. In certain embodiments, the split RuvC domain comprises RuvC-I, RuvC-II, and/or RuvC-III. In certain embodiments, the Type VII effector comprises one, two, or all three of RuvC-I, RuvC-II, and/or RuvC-III. In certain embodiments, the split RuvC domain

comprises an HNH-like domain. In certain embodiments, the split RuvC domain comprises structural domains linked in the following order: RuvC-I, Bridge Helix, RuvC-II, HNH, RuvC-III. In certain embodiments, the HNH domain is separate from the RuvC domain. In certain embodiments, the Type VII CRISPR proteins lack regions or domains homologous to a Cas9 P1 domain. In certain embodiments, the Type VII CRISPR effector proteins lack regions or domains homologous to a Cas9 REC1/REC2 domains of Cas9.

[0019] In certain embodiments, the Type VII effector protein comprises a protospacer adjacent motif (PAM) interacting domain. In certain embodiments, target nucleic acid binding specificity and/or cleavage is determined by the crRNA sequence and the PAM interacting domain. In certain embodiments, the PAM interacting domain allows or promotes binding to and/or cleavage of a target nucleic acid and reduces or inhibits binding and/or cleavage of an off-target nucleic acid that lacks a flanking PAM. In certain embodiments, the PAM is 5' to the target sequence. In certain embodiments, the PAM is 3' to the target sequence. In embodiments of the invention, the PAM comprises one, two, three, or four nucleotides and is directly adjacent to the target nucleic acid sequence or separated from the target nucleic acid sequence by one, two, or three nucleotides. In certain embodiments, the PAM is 5' to the target sequence. In certain embodiments, the PAM is 3' to the target sequence.

[0020] In certain embodiments, Type VII CRISPR effector proteins comprise an insertion between the bridge helix and RuvCII as compared to RuvC HNH nucleases (Fig. 4). The insertion is believed to contain residues that bind to crRNA. In certain embodiments, the Type VII CRISPR effector proteins comprise a PAM interacting domain.

[0021] The Type VII CRISPR effector proteins are compact relative to heretofore characterized Class II CRISPR proteins such as Cas9 and Cas12s. In certain embodiments, a Type VII CRISPR effector protein comprises from about 500 to about 525 aa, or from about 520 to about 540 aa, or from about 530 to about 550 aa, or from about 540 to about 560 aa, or from about 550 to about 570 aa, or from about 560 to about 580 aa, or from about 570 to about 590 aa, or from about 580 to about 600 aa, or from about 600 to about 650 aa, or from about 625 to about 675 aa, or from about 650 to about 700 aa.

[0022] Amino acid sequences of exemplary Type VII effector proteins are shown in **Table 1** below.

[0023] 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

[0024] 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:

[0025] **FIG. 1** - Phylogenetic tree. The phylogenetic tree depicts HNH Nucleases, IscB Nucleases (which are HNH Nucleases with a split RuvC and a Bridge Helix domain), VII Candidate A1, VII Candidate A2, as well as a representative panel of Cas9s. VII Candidate A is a distinct lineage of proteins.

[0026] **FIG. 2** – Phylogenetic tree of the HNH domains. The HNH domains of VII Candidate A1-A2 are distinct from the IscB transposases and HNH Endonuclease HNH domains and the more distant Cas9 HNH domains.

[0027] **FIG. 3** – Phylogenetic tree. The phylogenetic tree depicts novel Type VII effectors with Cas9 proteins, IscB proteins, and HNH Endonucleases.

[0028] **FIG. 4** – The top panel depicts sequence similarity of Type VII proteins A1 and A2. The bottom panel compares domain structure of A1 protein with RuvC HNH nucleases.

[0029] **FIG. 5** – Sequence alignment of Type VII proteins cA1 and cA2.

[0030] **FIG. 6** – Sequence alignment of Type VII proteins cB1 and cB2.

[0031] **FIG. 7** – Map of synthetic construct comprising Type VII cA1 locus with human codon optimized coding sequence.

[0032] **FIG. 8** – Map of synthetic construct comprising Type VII cA2 locus with human codon optimized coding sequence.

[0033] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

General Definitions

[0034] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.): Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.): Antibodies A Laboratory Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011)

.

[0035] As used herein, the singular forms “a”, “an”, and “the” include both singular and plural referents unless the context clearly dictates otherwise.

[0036] 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.

[0037] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0038] The terms “about” or “approximately” as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, and +/-0.1% or less of and from the specified value, insofar such

variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier “about” or “approximately” refers is itself also specifically, and preferably, disclosed.

[0039] Whereas the terms "one or more" or "at least one" or "X or more", where X is a number and understand to mean X or increases one by one of X, such as one or more or at least one member(s) or "X or more" of a group of members, is clear per se, by means of further exemplification, the term encompasses inter alia a reference to any one of said members, or to any two or more of said members, such as, e.g., any >3, >4, >5, >6 or >7 etc. of said members, and up to all said members.

[0040] As used herein, a “biological sample” may contain whole cells and/or live cells and/or cell debris. The biological sample may contain (or be derived from) a “bodily fluid”. The present invention encompasses embodiments wherein the bodily fluid is selected from amniotic fluid, aqueous humour, vitreous humour, bile, blood serum, breast milk, cerebrospinal fluid, cerumen (earwax), chyle, chyme, endolymph, perilymph, exudates, feces, female ejaculate, gastric acid, gastric juice, lymph, mucus (including nasal drainage and phlegm), pericardial fluid, peritoneal fluid, pleural fluid, pus, rheum, saliva, sebum (skin oil), semen, sputum, synovial fluid, sweat, tears, urine, vaginal secretion, vomit and mixtures of one or more thereof. Biological samples include cell cultures, bodily fluids, cell cultures from bodily fluids. Bodily fluids may be obtained from a mammal organism, for example by puncture, or other collecting or sampling procedures.

[0041] 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.

[0042] Various embodiments are described hereinafter. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s). Reference throughout this specification to “one embodiment”, “an embodiment,” “an example embodiment,” means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present

invention. Thus, appearances of the phrases "in one embodiment," "in an embodiment," or "an example embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

[0043] 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

[0044] Embodiments disclosed herein provide CRISPR-Cas systems that can be classified as Type VII systems.

[0045] In general, a CRISPR-Cas or CRISPR system as used in herein and in documents, such as International Patent Publication No. WO 2014/093622 (PCT/US2013/074667), refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas, such as Cas9, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). See, e.g, Shmakov et al. (2015) "Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems", Molecular Cell, DOI: dx.doi.org/10.1016/j.molcel.2015.10.008.

[0046] Currently known CRISPR-Cas systems can generally fall into two classes based on their architectures of their effector molecules, which are each further subdivided by type and subtype. The two classes are Class 1 and Class 2. Class 1 CRISPR-Cas systems have effector modules composed of multiple Cas proteins, some of which form crRNA-binding complexes, while Class 2 CRISPR-Cas systems include a single, multi-domain crRNA-binding protein. Each class can be divided into Types, which are further subdivided into subtypes. *See e.g.,* Makarova et al. 2020. Nat. Rev. 18: 67-83. Class 1 CRISPR-Cas systems are divided into types I, II, and IV. Makarova et al. 2020. Nat. Rev. 18: 67-83, particularly as described in Figure 1. Type I CRISPR-Cas systems are divided into 9 subtypes (I-A, I-B, I-C, I-D, I-E, I-F1, I-F2, I-F3, and IG). Makarova *et al.*, 2020. Each type of Class 2 system is further divided into subtypes. See Markova et al. 2020, particularly at Figure. 2. Class 2, Type II systems can be divided into 4 subtypes: II-A, II-B, II-C1, and II-C2. Class 2, Type V systems can be divided into 17 subtypes: V-A, V-B1, V-B2, V-C, V-D, V-E, V-F1, V-F1(V-U3), V-F2, V-F3, V-G, V-H, V-I, V-K (V-U5), V-U1, V-U2, and V-U4. Class 2, Type IV systems can be divided into 5 subtypes: VI-A, VI-B1, VI-B2, VI-C, and VI-D.

[0047] In certain embodiments, a protospacer adjacent motif (PAM) or PAM-like motif directs binding of the effector protein complex as disclosed herein to the target locus of interest. In some embodiments, the PAM may be a 5' PAM (i.e., located upstream of the 5' end of the protospacer). In other embodiments, the PAM may be a 3' PAM (i.e., located downstream of the 5' end of the protospacer). The term "PAM" may be used interchangeably with the term "PFS" or "protospacer flanking site" or "protospacer flanking sequence".

[0048] In the context of formation of a CRISPR complex, "target sequence" or "target nucleic acid" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise RNA polynucleotides. The term "target RNA" refers to an RNA polynucleotide being or comprising the target sequence. In other words, the target RNA may be an RNA polynucleotide or a part of an RNA polynucleotide to which a part of the gRNA, i.e. the guide sequence, is designed to have complementarity and to which the effector function mediated by the complex comprising CRISPR effector protein and a gRNA is to be directed. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell.

[0049] The Type VII CRISPR-Cas systems of the present invention are a new Type of CRISPR-Cas system not currently identified by current classification systems. In some embodiments, the Type VII CRISPR-Cas systems described herein can have an effector molecule (e.g. a Cas) that is distinct from Cas9 as well as known split RuvC HNH nucleases.

[0050] As described in greater detail elsewhere herein, the Type VII systems and/or components thereof (e.g., guide molecules, effector molecules, and/or polynucleotide(s) encoding the same) can be delivered to and/or produced by a cell. The Type VII systems and/or components can be used to modify a polynucleotide, in screening methods, generate modified organisms. Other features, advantages, uses, and applications for the Type VII systems and/or components thereof are described in greater detail herein.

TYPE VII SYSTEMS

Type VII Effector and Additional Effector Molecules

Type VII Effectors

[0051] The Type VII CRISPR-Cas systems described herein can include one or more Type VII effector molecules. In some embodiments, the type VII CRISPR effector proteins can include a RuvC-like nuclease domain. Type VII CRISPR effector proteins comprise a RuvC-like nuclease domain. In certain embodiments, nuclease activity is present while in other embodiments, the nuclease is inactive. In certain embodiments, a Type VII CRISPR effector protein comprises an HNH-like domain. In certain embodiments, the HNH-like domain comprises nuclease activity. In other embodiments, the HNH-like domain lacks nuclease activity. In certain embodiments, the CRISPR effector proteins comprise a split RuvC domain. In certain embodiments, the split RuvC domain comprises RuvC-I, RuvC-II, and/or RuvC-III. In certain embodiments, the Type VII effector comprises one, two, or all three of RuvC-I, RuvC-II, and/or RuvC-III. In certain embodiments, the split RuvC domain comprises an HNH-like domain. In certain embodiments, the split RuvC domain comprises structural domains linked in the following order: RuvC-I, Bridge Helix, RuvC-II, HNH, RuvC-III. In certain embodiments, the HNH domain is separate from the RuvC domain. In certain embodiments, the Type VII CRISPR proteins lack regions or domains homologous to a Cas9 P1 domain. In certain embodiments, the Type VII CRISPR effector proteins lack regions or domains homologous to a Cas9 REC1/REC2 domains of Cas9.

[0052] Type VII CRISPR effector proteins comprise a RuvC-like nuclease domain. In certain embodiments, nuclease activity is present while in other embodiments, the nuclease is

inactive. In certain embodiments, a Type VII CRISPR effector protein comprises an HNH-like domain. In certain embodiments, the HNH-like domain comprises nuclease activity. In other embodiments, the HNH-like domain lacks nuclease activity. In certain embodiments, the CRISPR effector proteins comprise a split RuvC domain. In certain embodiments, the split RuvC domain comprises RuvC-I, RuvC-II, and/or RuvC-III. In certain embodiments, the Type VII effector comprises one, two, or all three of RuvC-I, RuvC-II, and/or RuvC-III. In certain embodiments, the split RuvC domain comprises an HNH-like domain. In certain embodiments, the split RuvC domain comprises structural domains linked in the following order: RuvC-I, Bridge Helix, RuvC-II, HNH, RuvC-III. In certain embodiments, the HNH domain is separate from the RuvC domain. In certain embodiments, the Type VII CRISPR proteins lack regions or domains homologous to a Cas9 P1 domain. In certain embodiments, the Type VII CRISPR effector proteins lack regions or domains homologous to a Cas9 REC1/REC2 domains of Cas9.

[0053] In certain embodiments, the Type VII effector protein comprises a protospacer adjacent motif (PAM) interacting domain. In certain embodiments, target nucleic acid binding specificity and/or cleavage is determined by the crRNA sequence and the PAM interacting domain. In certain embodiments, the PAM interacting domain allows or promotes binding to and/or cleavage of a target nucleic acid and reduces or inhibits binding and/or cleavage of an off-target nucleic acid that lacks a flanking PAM. In certain embodiments, the PAM is 5' to the target sequence. In certain embodiments, the PAM is 3' to the target sequence. In embodiments of the invention, the PAM comprises one, two, three, or four nucleotides and is directly adjacent to the target nucleic acid sequence or separated from the target nucleic acid sequence by one, two, or three nucleotides. In certain embodiments, the PAM is 5' to the target sequence. In certain embodiments, the PAM is 3' to the target sequence.

[0054] Amino acid sequences of exemplary Type VII effector proteins are shown in **Table 1** below.

Table 1 - Type VII effector proteins	
VII_cA1 aa 1-611	MPCHPARARELVRKGRAYWVKRDTVKLKTFLQNAVCQHTVLGIDAGAIHIGLAVVAKREK RKPKVLFRAQVDARPMSEIKSVMDKRRSLRRNRRSRKGYRQPWYSPTRRVGLTAKLFKKS GEFLCKIPLAEGKRLLKPDKGGNRQAKRLGGNKYQLSTMKEIKAPKREKEPGKNDIVVYR FDGRPLTWFSPKSRKFVKQLCKCLRGYGKVRRLPVADVKKEGDKVKSIWLKSRRAERLIP

SEQ ID NO: 1	KNILPEVVPPFRKMGQKRKPSGWLAPSANALVNAYFNGIKRVGSYVPIASVRFEDVAFDT RKLEDCLDGKGDCSTVNYARQNRKDFVLARDGWKCVYCGRKSNFSKKGLRLTQDHVVPKS KNGSNAIGNLVACCEKCNSDKGSLDLDDYLKKREQQGDKPSDTKRIKRYVVGLSQKSESF RSKAHVGQVKTKVLRLLRERFEDVSITKGYVTKKDRDWMGLTKSHDCDAVAIAAYGERVG FMPPLRTYHVRRYTWEGSRRRQYSANPMKKYSSKCVLIAKGTPRERWVQPNSYNPYADTK VGRVYSRDVVLAVGGNKGTEPVLGYIRGICKDGALFISRLPTGGKKKERSKRQAKYVKKI ISRCRGLMEVL
VII_cA2 aa 1-625 **SEQ ID NO: 2**	MNQDNKPLMP CHPARARELV RRGRAYWVRR DTIKLSEYKE NAVLQRTFLG IDAGAEHIGLAVVAEREKHK PRVIFRAQVE ARPMREIKEC LDQRRSYRRG RRSRKAYRRPWFSPTRRVSL SARLYNKQGE FLCRIPLAYG KYLTKPDKSG RPQAKRLYGNNYQLLTSKNH ESPIREKKPG KQDILVYRND GRLLTWYDPK RGDLEGIEKRLLGTSRRHNI PIASLEKEGS KKIIRLLSVS AEDAIPKAAL PNPMPRFRKM GQKRKPPGWL SPSANSLLNA YYRGIKEIGS YVPISLVRLE VAAFDTQKIE DPTIHGVGYQ KPALYAKQNR KDFVIARDGW RCVYCGVNNV PLTQDHIQPQ SKGGPDRIGN LVACCFKCQN EKSNLTLDEF LKTKSPQQAK RIKEYVSSIT QKNESLQKNE SLRSAAHVGQ IKSKLLQGLI DLFGDSNVKR TWGYITKRDR LALHLPKSHD YDAIAIASLG RRIDTENIPP LRLYYMRRYT KDSPRRQKYK ANPLSNAKKD NKNLAPLFIG QGSRSRWVQQ VQVNSYAETP VGIVRLRDLI KTIDGRIGYV RKINSGRKPK PGQKPVRILT LSHKPQGKEG QFSTSANKVE KIIRRRCGLL EMLYV
VII-cB1 aa 1-567 **SEQ ID NO: 3**	MAQLARQLRE EAARWLGVFG DDANVRQKIG TPTGYQTSVC RESWGENLPK KDRSNLRHHL WDAAVISHIP PGKGMNLTNC AGIFLHTGTN QHGNITMRAI PGLGPDLNSF ENGTKDQCLV DERRATKSKK QRFDQTIYSP PDTEGVMWAR DPLDKLATKK NQSAEGLLQL LRDAGMDENQ LPSSRFEEWW NGVQAQWFTR NDAQALFQTL EIPSPKIISE DALNKWWGNY EKKKINYKSL RLLLKESGIT PVEVSDLRLQ NALSGHSEPG SLTRKDGTPI RGVRGTAKQM TPMAVIPHRN HKGETIGFKF AKESFVRAEI WVTEKTTKDG KVELDYHCRL IPHPRGLANL QLRVMKCTGQ PLTWERRLTD DELVELGLAK KLATLKNERE RLKNSHAKAV EKWQKSQQKK KDSDTLGLAA DAATSSPKEI PELWLSLRKI YTGLPPHAKR LTNANGADIS RVTKGDLLFV PLRKVPGAGK GHGKFCKRGE VATGGSCWFR VSAIKAAGEI ELHLAEFKLP KIKDDKKATE QQEWLTTIWE QRPSSEDDIA WLLEQSRAHD QSPHPVK
VII-cB2 aa 1-591 **SEQ ID NO: 4**	MGRGFRASIK SKNHHNLSYG ETNGLARQLR EEAAHWLGIF GDDAKVRQRI GTPTGYQTSV CRGSWSINLP ADFYPKKNRN NLRHHLWDAA ILANIPPGKG LNLTNCGGIF FHAGRNKHGN IAMQALPGLG PDLNAFEKAT SNQCLVEKKQ TTNSKKSRFQ QWIYGFPDKN GEMWARDAIE TLADKPNLLS RLRDAGIDEA QLPTRHFEVW QERRQAQYFT REEAWVAVQA LALPAEHAIP VSTFDEWWND RLKSDKKKVT DKSLRALLAK ARIPKAAVTD VKLTGVLVNR GDPGPLTRKD GTPIYGVSGT AKQMTPMAVV PHRNHKGETI GFKLATESFI RAEIWTTEKR DKKGDIVKDD TGKPVLEYHR RLIPHPRGLK NLRLRTMRCT GKRLSWERRL TDEEIVELGL KEHAEVKRLR KDYDKAVSQH VKHTRAQAPE CELALCQSKV TMPGQPKPPV ISLRKIYTGL PPNAKRLTNA QGADISCFAK GDLLLVPVSQ SGAICESDKA PYRELWYRIT ALKANGQVEL KLAEFKEIKR PSEKELKEGK SLTPEQEWFA DAFLQQPGND ELIAGLLQRT RVNDQSSHSV K

Table 2 - Type VII Direct Repeats	
VII cA1 **SEQ ID NO: 5**	gttacaacgcacggtgcagcaaccctgtgactgaaac
VII cA2 **SEQ ID NO: 6**	gttgcaatgtatagtgcagcgaccctgggaatgaaac
VII cB1 **SEQ ID NO: 7**	gttttgatctcgctccccccggagcaatttggtatt
VII cB2 **SEQ ID NO: 8**	gttttgatctcgcccccacggggcaatttggtaag

[0055] The exemplified Type VII effectors are relatively small, as small as or smaller than 578 amino acids in length and are associated with CRISPR arrays. In certain embodiments, the Type VII effectors comprise RuvCI, RuvCII, RuvCIII, Bridge Helix, and HNH domains, and comprise catalytic amino acid residues corresponding to D10, 893H and 896D of SpCas9, as well as conserved Asp and His catalytic residues in their HNH domains. In certain embodiments of the invention, a Type VII effector comprises the RuvC and HNH catalytic residues of SpCas9. In certain embodiments, one or more catalytic residues is mutated and the Type VII effector lacks RuvC and or HNH nuclease activity. Accordingly, in certain embodiments, the Type VII effector is a nickase. In certain embodiments, the Type VII effector lack nuclease activities and is a dead Type VII effector.

[0056] The exemplified Type VII cA1 and cA2 effectors have 66% sequence similarity (52% identity) to each other, and their respective direct repeats (DRs) have 78% identity. In this respect, the Type VII proteins are distinct from Cas9s, demonstrating 5-6% identity (13% similarity) to SpCas9 and 9% identity (16% similarity) to CjCas9. The Type VII cB1 and cB2 effectors are also similar. In certain embodiments of the invention, the Type VII effector protein has greater than about 40%, or greater than about 50%, or greater than about 60%, or greater than about 70%, or greater than about 80%, or greater than about 90%, or greater than about 95% identity to a Type VII effector set forth herein. In certain embodiments of the invention, the Type VII effector protein has greater than about 40%, or greater than about 50%, or greater than about 60%, or greater than about 70%, or greater than about 80%, or greater than about 90%, or greater than about 95% identity to a Type VII effector set forth in **Table 1**.

[0057] In some embodiments, the Type VII effector polypeptide can have about 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 % identity to any one of **SEQ ID NOs**: **1-4**. In some embodiments, the Type VII effector polypeptide can have about 50 to 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 % identity to any one of **SEQ ID NOs**: **1-4**.

[0058] In some embodiments a Type VII effector direct repeat can have about 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 % identity to any one of **SEQ ID NOs**: 5-8 as shown in **Table 2**. In some embodiments a Type

VII effector direct repeat can have about 50 to 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 % identity to any one of **SEQ ID NOs**: **5-8** as shown in **Table 2**.

[0059] In some embodiments, a polynucleotide encoding a Type VII effector polypeptide can have about 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 % identity to any one of **SEQ ID NOs**: **34-37** as shown in **Table 14** (see Examples). In some embodiments, a polynucleotide encoding a Type VII effector polypeptide can have about 50 to about 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 % identity to any one of **SEQ ID NOs.: 34-37** as shown in **Table 14** (see Examples).

[0060] In one example embodiment, the CRISPR system effector (e.g. a Type VII effector or additional effector protein) is or is engineered to be an RNA-targeting effector protein. In certain embodiments, the CRISPR system effector protein is engineered to be used like a Type VI CRISPR system targeting RNA (e.g., Cas13a, Cas13b, Cas13c or Cas13d). Example RNA-targeting effector proteins include Cas13b and C2c2 (now known as Cas13a). It will be understood that the term "C2c2" herein is used interchangeably with "Cas13a". "C2c2" is now referred to as "Cas13a", and the terms are used interchangeably herein unless indicated otherwise. As used herein, the term "Cas13" refers to any Type VI CRISPR system targeting RNA (e.g., Cas13a, Cas13b, Cas13c or Cas13d). When the CRISPR protein is a C2c2 protein, a tracrRNA is not required. C2c2 has been described in Abudayyeh et al. (2016) "C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector"; Science; DOI: 10.1126/science.aaf5573; and Shmakov et al. (2015) "Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems", Molecular Cell, DOI: dx.doi.org/10.1016/j.molcel.2015.10.008; which are incorporated herein in their entirety by reference. Cas13b has been described in Smargon et al. (2017) "Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNases Differentially Regulated by Accessory Proteins Csx27 and Csx28," Molecular Cell. 65, 1-13; dx.doi.org/10.1016/j.molcel.2016.12.023., which is incorporated herein in its entirety by reference.

Additional Effectors

[0061] Additional effectors for use according to the invention can be identified by their proximity to cas1 genes, for example, though not limited to, within the region 20 kb from the start of the cas1 gene and 20 kb from the end of the cas1 gene. In certain embodiments, the effector protein comprises at least one HEPN domain and at least 500 amino acids, and wherein the C2c2 effector protein is naturally present in a prokaryotic genome within 20 kb upstream or downstream of a Cas gene or a CRISPR array. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologues thereof, or modified versions thereof. In certain example embodiments, the C2c2 effector protein is naturally present in a prokaryotic genome within 20kb upstream or downstream of a Cas 1 gene. The terms "orthologue" (also referred to as "ortholog" herein) and "homologue" (also referred to as "homolog" herein) are well known in the art. By means of further guidance, a "homologue" of a protein as used herein is a protein of the same species which performs the same or a similar function as the protein it is a homologue of. Homologous proteins may but need not be structurally related, or are only partially structurally related. An "orthologue" of a protein as used herein is a protein of a different species which performs the same or a similar function as the protein it is an orthologue of. Orthologous proteins may but need not be structurally related, or are only partially structurally related.

Modified Effectors

[0062] The Type VII and/or additional effector molecules can be modified such that they have one or more additional or other functionalities or lack one or more native functionalities. In some embodiments one or more domains are added. In some embodiments, the effector molecules can be modified such that they are split between two molecules, such that they are a "split" effector molecule.

Additional Functional Domains

[0063] In some embodiments, the system is a Cas-based system that is capable of performing a specialized function or activity. For example, the Cas protein may be fused, operably coupled to, or otherwise associated with one or more functionals domains. In certain example embodiments, the Cas protein may be a catalytically dead Cas protein ("dCas") and/or

have nickase activity. A nickase is a Cas protein that cuts only one strand of a double stranded target. In such embodiments, the dCas or nickase provide a sequence specific targeting functionality that delivers the functional domain to or proximate a target sequence. Example functional domains that may be fused to, operably coupled to, or otherwise associated with a Cas protein can be or include, but are not limited to, an HPEN domain or a catalytically active domain that is homologous to an HPEN domain, a nuclear localization signal (NLS) domain, a nuclear export signal (NES) domain, a translational activation domain, a transcriptional activation domain (e.g. VP64, p65, MyoD1, HSF1, RTA, and SET7/9), a translation initiation domain, a transcriptional repression domain (e.g., a KRAB domain, NuE domain, NcoR domain, and a SID domain such as a SID4X domain), a nuclease domain (e.g., FokI), a histone modification domain (e.g., a histone acetyltransferase), a light inducible/controllable domain, a chemically inducible/controllable domain, a transposase domain, a homologous recombination machinery domain, a recombinase domain, an integrase domain, and combinations thereof.

[0064] In some embodiments, the functional domains can have one or more of the following activities: methylase activity, demethylase activity, translation activation activity, translation initiation activity, translation repression activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity, single-strand DNA cleavage activity, double-strand DNA cleavage activity, molecular switch activity, chemical inducibility, light inducibility, and nucleic acid binding activity. In some embodiments, the one or more functional domains may comprise epitope tags or reporters. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporters include, but are not limited to, glutathione-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP).

[0065] The one or more functional domain(s) may be positioned at, near, and/or in proximity to a terminus of the effector protein (e.g., a Cas protein). In embodiments having two or more functional domains, each of the two can be positioned at or near or in proximity

to a terminus of the effector protein (e.g., a Cas protein). In some embodiments, such as those where the functional domain is operably coupled to the effector protein, the one or more functional domains can be tethered or linked via a suitable linker (including, but not limited to, GlySer linkers) to the effector protein (e.g., a Cas protein). When there is more than one functional domain, the functional domains can be same or different. In some embodiments, all the functional domains are the same. In some embodiments, all of the functional domains are different from each other. In some embodiments, at least two of the functional domains are different from each other. In some embodiments, at least two of the functional domains are the same as each other.

[0066] Other suitable functional domains can be found, for example, in International Application Publication No. WO 2019/018423.

Split CRISPR-Cas systems

[0067] In some embodiments, the CRISPR-Cas system is a split CRISPR-Cas system. *See* e.g., Zetche et al., 2015. Nat. Biotechnol. 33(2): 139-142, the compositions and techniques of which can be used in and/or adapted for use with the present invention. Split CRISPR-Cas proteins are set forth herein and in documents incorporated herein by reference in further detail herein. In certain embodiments, each part of a split CRISPR protein 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 some embodiments, CRISPR proteins may preferably split between domains, leaving domains intact. In particular embodiments, said Cas split domains (e.g., RuvC and HNH domains in the case of Cas9) can be simultaneously or sequentially introduced into the cell such that said split Cas domain(s) process the target nucleic acid sequence in the algae cell. The reduced size of the split Cas compared to the wild type Cas allows other methods of delivery of the systems to the cells, such as the use of cell penetrating peptides as described herein.

Base Editing

[0068] In some embodiments, a Cas protein (e.g. a Type VII effector or additional effector protein described herein) is connected or fused to a nucleotide deaminase. Thus, in some embodiments the Cas-based system can be a base editing system. As used herein "base editing"

refers generally to the process of polynucleotide modification via a CRISPR-Cas-based or Cas-based system that does not include excising nucleotides to make the modification. Base editing can convert base pairs at precise locations without generating excess undesired editing byproducts that can be made using traditional CRISPR-Cas systems.

[0069] In certain example embodiments, the nucleotide deaminase may be a DNA base editor used in combination with a DNA binding Cas protein such as, but not limited to, Class 2 Type II and Type V systems. Two classes of DNA base editors are known: cytosine base editors (CBEs) and adenine base editors (ABEs). CBEs convert a C•G base pair into a T•A base pair (Komor et al. 2016. Nature. 533:420-424; Nishida et al. 2016. Science. 353; and Li et al. Nat. Biotech. 36:324-327) and ABEs convert an A•T base pair to a G•C base pair. Collectively, CBEs and ABEs can mediate all four possible transition mutations (C to T, A to G, T to C, and G to A). Rees and Liu. 2018.Nat. Rev. Genet. 19(12): 770-788, particularly at Figures 1b, 2a-2c, 3a-3f, and Table 1. In some embodiments, the base editing system includes a CBE and/or an ABE. In some embodiments, a polynucleotide of the present invention described elsewhere herein (e.g., a MARC polynucleotide, such as a MARC1 or MARC2 gene or transcript) can be modified using a base editing system. Rees and Liu. 2018. Nat. Rev. Gent. 19(12):770-788. Base editors also generally do not need a DNA donor template and/or rely on homology-directed repair. Komor et al. 2016. Nature. 533:420-424; Nishida et al. 2016. Science. 353; and Gaudeli et al. 2017. Nature. 551:464-471. Upon binding to a target locus in the DNA, base pairing between the guide RNA of the system and the target DNA strand leads to displacement of a small segment of ssDNA in an "R-loop". Nishimasu et al. Cell. 156:935-949. DNA bases within the ssDNA bubble are modified by the enzyme component, such as a deaminase. In some systems, the catalytically disabled Cas protein can be a variant or modified Cas can have nickase functionality and can generate a nick in the non-edited DNA strand to induce cells to repair the non-edited strand using the edited strand as a template. Komor et al. 2016. Nature. 533:420-424; Nishida et al. 2016. Science. 353; and Gaudeli et al. 2017. Nature. 551:464-471.

[0070] Other Example Type V base editing systems are described in International Patent Publication Nos. WO 2018/213708 and WO 2018/213726, and International Patent Application Nos. PCT/US2018/067207, PCT/US2018/067225, and PCT/US2018/067307 which are incorporated herein by reference.

[0071] In certain example embodiments, the base editing system may be a RNA base editing system. As with DNA base editor a nucleotide deaminase capable of converting nucleotide bases may be fused to a Cas protein. However, in these embodiments, the Cas protein will need to be capable of binding RNA. Example RNA binding Cas proteins include, but are not limited to, RNA-binding Cas9s such as *Francisella novicida* Cas9 ("FnCas9"), and Class 2 Type VI Cas systems. The nucleotide deaminase may be a cytidine deaminase or an adenosine deaminase, or an adenosine deaminase engineered to have cytodine deaminase activity. In certain example embodiments, the RNA based editor may be used to delete or introduce a post-translation modification site in the expressed mRNA. In contrast to DNA base editors, whose edits are permanent in the modified cell, RNA base editors can provide edits where finer temporal control may be needed, for example in modulating a particular immune response. Example Type VI RNA-base editing systems are described in Cox et al. 2017. Science 358: 1019-1027, International Patent Publication Nos. WO 2019/005884, WO 2019/005886, and WO 2019/071048, and International Patent Application Nos. PCT/US20018/05179, PCT/US2018/067207, which are incorporated herein by reference. An example FnCas9 system that may be adapted for RNA base editing purposes is described in International Patent Publication No. WO 2016/106236, which is incorporated herein by reference.

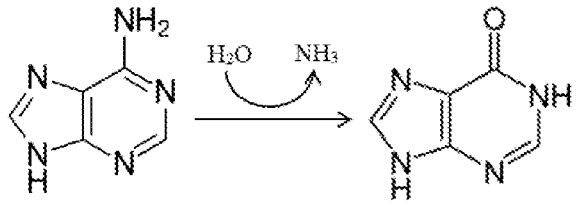
[0072] An example method for delivery of base-editing systems, including use of a split-intein approach to divide CBE and ABE into reconstitutable halves, is described in Levy et al. Nature Biomedical Engineering doi.org/10.1038/s41441-019-0505-5 (2019), which is incorporated herein by reference.

[0073] In certain example embodiments, a dead Type VII CRISPR protein can be fused with a adenosine deaminase or cytidine deaminase for base editing purposes. Reference is made to WO 2019/005884, WO 2019/005886, PCT/US2018/052247, PCT/US2018/054469, PCT/US2018/0571179, PCT/US2018/057177, WO2018/035388, WO/2018/035387, WO/2018/170333, WO/2018/191388, WO/2018/213708, PCT/US2018/067207, PCT/US2018/067225, PCT/US2018/067307, WO/2016/205749, PCT/US2018/067328, WO/2016/205764, WO/2017/070605, WO/2017/219027, US-2017-0211142-A1, WO/2019/005866, WO/2019/018423; Cox DBT, et al., RNA editing with CRISPR-Cas13, Science. 2017 Nov 24;358(6366):1019-1027; Gootenberg JS, et al., Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6., Science. 2018 Apr

27;360(6387):439-444; Gootenberg JS, et al., Nucleic acid detection with CRISPR-Cas13a/C2c2., Science. 2017 Apr 28;356(6336):438-442; Abudayyeh OO, et al., RNA targeting with CRISPR-Cas13, Nature. 2017 Oct 12;550(7675):280-284; Smargon AA, et al., Cas13b Is a Type VI-B CRISPR-Associated RNA-Guided RNase Differentially Regulated by Accessory Proteins Csx27 and Csx28. Mol Cell. 2017 Feb 16;65(4):618-630.e7; Abudayyeh OO, et al., C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector, Science. 2016 Aug 5;353(6299):aaf5573; Yang L, et al., Engineering and optimizing deaminase fusions for genome editing. Nat Commun. 2016 Nov 2;7:13330, each of which is incorporated herein by reference in its entirety.

Adenosine Deaminase

[0074] The term "adenosine deaminase" or "adenosine deaminase protein" as used herein refers to a protein, a polypeptide, or one or more functional domain(s) of a protein or a polypeptide that is capable of catalyzing a hydrolytic deamination reaction that converts an adenine (or an adenine moiety of a molecule) to a hypoxanthine (or a hypoxanthine moiety of a molecule), as shown below. In some embodiments, the adenine-containing molecule is an adenosine (A), and the hypoxanthine-containing molecule is an inosine (I). The adenine-containing molecule can be deoxyribonucleic acid (DNA) or ribonucleic acid (RNA).



Adenine Hypoxanthine

[0075]

[0076] According to the present disclosure, adenosine deaminases that can be used in connection with the present disclosure include, but are not limited to, members of the enzyme family known as adenosine deaminases that act on RNA (ADARs), members of the enzyme family known as adenosine deaminases that act on tRNA (ADATs), and other adenosine deaminase domain-containing (ADAD) family members. According to the present disclosure, the adenosine deaminase is capable of targeting adenine in a RNA/DNA and RNA duplexes. Indeed, Zheng et al. (Nucleic Acids Res. 2017, 45(6): 3369-3377) demonstrate that ADARs can carry out adenosine to inosine editing reactions on RNA/DNA and RNA/RNA duplexes.

In particular embodiments, the adenosine deaminase has been modified to increase its ability to edit DNA in a RNA/DNA heteroduplex of in an RNA duplex as detailed herein below.

[0077] In some embodiments, the adenosine deaminase is derived from one or more metazoa species, including but not limited to, mammals, birds, frogs, squids, fish, flies and worms. In some embodiments, the adenosine deaminase is a human, squid or Drosophila adenosine deaminase.

[0078] In some embodiments, the adenosine deaminase is a human ADAR, including hADAR1, hADAR2, hADAR3. In some embodiments, the adenosine deaminase is a Caenorhabditis elegans ADAR protein, including ADR-1 and ADR-2. In some embodiments, the adenosine deaminase is a Drosophila ADAR protein, including dAdar. In some embodiments, the adenosine deaminase is a squid Loligo pealeii ADAR protein, including sqADAR2a and sqADAR2b. In some embodiments, the adenosine deaminase is a human ADAT protein. In some embodiments, the adenosine deaminase is a Drosophila ADAT protein. In some embodiments, the adenosine deaminase is a human ADAD protein, including TENR (hADAD1) and TENRL (hADAD2).

[0079] In some embodiments, the adenosine deaminase is a TadA protein such as E. coli TadA. See Kim et al., Biochemistry 45:6407-6416 (2006); Wolf et al., EMBO J. 21:3841-3851 (2002). In some embodiments, the adenosine deaminase is mouse ADA. See Grunebaum et al., Curr. Opin. Allergy Clin. Immunol. 13:630-638 (2013). In some embodiments, the adenosine deaminase is human ADAT2. See Fukui et al., J. Nucleic Acids 2010:260512 (2010).

[0080] In some embodiments, the adenosine deaminase protein recognizes and converts one or more target adenosine residue(s) in a double-stranded nucleic acid substrate into inosine residues (s). In some embodiments, the double-stranded nucleic acid substrate is a RNA-DNA hybrid duplex. In some embodiments, the adenosine deaminase protein recognizes a binding window on the double-stranded substrate. In some embodiments, the binding window contains at least one target adenosine residue(s). In some embodiments, the binding window is in the range of about 3 bp to about 100 bp. In some embodiments, the binding window is in the range of about 5 bp to about 50 bp. In some embodiments, the binding window is in the range of about 10 bp to about 30 bp. In some embodiments, the binding window is about 1 bp, 2 bp, 3 bp, 5 bp, 7 bp, 10 bp, 15 bp, 20 bp, 25 bp, 30 bp, 40 bp, 45 bp, 50 bp, 55 bp, 60 bp, 65 bp, 70 bp, 75 bp, 80 bp, 85 bp, 90 bp, 95 bp, or 100 bp.

[0081] In some embodiments, the adenosine deaminase protein comprises one or more deaminase domains. Not intended to be bound by a particular theory, it is contemplated that the deaminase domain functions to recognize and convert one or more target adenosine (A) residue(s) contained in a double-stranded nucleic acid substrate into inosine (I) residue(s). In some embodiments, the deaminase domain comprises an active center. In some embodiments, the active center comprises a zinc ion. In some embodiments, during the A-to-I editing process, base pairing at the target adenosine residue is disrupted, and the target adenosine residue is "flipped" out of the double helix to become accessible by the adenosine deaminase. In some embodiments, amino acid residues in or near the active center interact with one or more nucleotide(s) 5' to a target adenosine residue. In some embodiments, amino acid residues in or near the active center interact with one or more nucleotide(s) 3' to a target adenosine residue. In some embodiments, amino acid residues in or near the active center further interact with the nucleotide complementary to the target adenosine residue on the opposite strand. In some embodiments, the amino acid residues form hydrogen bonds with the 2' hydroxyl group of the nucleotides.

[0082] In some embodiments, the adenosine deaminase comprises human ADAR2 full protein (hADAR2) or the deaminase domain thereof (hADAR2-D). In some embodiments, the adenosine deaminase is an ADAR family member that is homologous to hADAR2 or hADAR2-D.

[0083] Particularly, in some embodiments, the homologous ADAR protein is human ADAR1 (hADAR1) or the deaminase domain thereof (hADAR1-D). In some embodiments, glycine 1007 of hADAR1-D corresponds to glycine 487 hADAR2-D, and glutamic Acid 1008 of hADAR1-D corresponds to glutamic acid 488 of hADAR2-D.

[0084] In some embodiments, the adenosine deaminase comprises the wild-type amino acid sequence of hADAR2-D. In some embodiments, the adenosine deaminase comprises one or more mutations in the hADAR2-D sequence, such that the editing efficiency, and/or substrate editing preference of hADAR2-D is changed according to specific needs.

[0085] Certain mutations of hADAR1 and hADAR2 proteins have been described in Kuttan et al., Proc Natl Acad Sci U S A. (2012) 109(48):E3295-304; Want et al. ACS Chem Biol. (2015) 10(11):2512-9; and Zheng et al. Nucleic Acids Res. (2017) 45(6):3369-337, each of which is incorporated herein by reference in its entirety.

[0086] In some embodiments, the adenosine deaminase comprises a mutation at glycine336 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 336 is replaced by an aspartic acid residue (G336D).

[0087] In some embodiments, the adenosine deaminase comprises a mutation at Glycine487 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 487 is replaced by a non-polar amino acid residue with relatively small side chains. For example, in some embodiments, the glycine residue at position 487 is replaced by an alanine residue (G487A). In some embodiments, the glycine residue at position 487 is replaced by a valine residue (G487V). In some embodiments, the glycine residue at position 487 is replaced by an amino acid residue with relatively large side chains. In some embodiments, the glycine residue at position 487 is replaced by a arginine residue (G487R). In some embodiments, the glycine residue at position 487 is replaced by a lysine residue (G487K). In some embodiments, the glycine residue at position 487 is replaced by a tryptophan residue (G487W). In some embodiments, the glycine residue at position 487 is replaced by a tyrosine residue (G487Y).

[0088] In some embodiments, the adenosine deaminase comprises a mutation at glutamic acid488 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glutamic acid residue at position 488 is replaced by a glutamine residue (E488Q). In some embodiments, the glutamic acid residue at position 488 is replaced by a histidine residue (E488H). In some embodiments, the glutamic acid residue at position 488 is replace by an arginine residue (E488R). In some embodiments, the glutamic acid residue at position 488 is replace by a lysine residue (E488K). In some embodiments, the glutamic acid residue at position 488 is replace by an asparagine residue (E488N). In some embodiments, the glutamic acid residue at position 488 is replace by an alanine residue (E488A). In some embodiments, the glutamic acid residue at position 488 is replace by a Methionine residue (E488M). In some embodiments, the glutamic acid residue at position 488 is replace by a serine residue (E488S). In some embodiments, the glutamic acid residue at position 488 is replace by a phenylalanine residue (E488F). In some embodiments, the glutamic acid residue at position 488 is replace by a lysine residue (E488L). In some embodiments, the glutamic acid residue at position 488 is replace by a tryptophan residue (E488W).

[0089] In some embodiments, the adenosine deaminase comprises a mutation at threonine490 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the threonine residue at position 490 is replaced by a cysteine residue (T490C). In some embodiments, the threonine residue at position 490 is replaced by a serine residue (T490S). In some embodiments, the threonine residue at position 490 is replaced by an alanine residue (T490A). In some embodiments, the threonine residue at position 490 is replaced by a phenylalanine residue (T490F). In some embodiments, the threonine residue at position 490 is replaced by a tyrosine residue (T490Y). In some embodiments, the threonine residue at position 490 is replaced by a serine residue (T490R). In some embodiments, the threonine residue at position 490 is replaced by an alanine residue (T490K). In some embodiments, the threonine residue at position 490 is replaced by a phenylalanine residue (T490P). In some embodiments, the threonine residue at position 490 is replaced by a tyrosine residue (T490E).

[0090] In some embodiments, the adenosine deaminase comprises a mutation at valine493 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the valine residue at position 493 is replaced by an alanine residue (V493A). In some embodiments, the valine residue at position 493 is replaced by a serine residue (V493S). In some embodiments, the valine residue at position 493 is replaced by a threonine residue (V493T). In some embodiments, the valine residue at position 493 is replaced by an arginine residue (V493R). In some embodiments, the valine residue at position 493 is replaced by an aspartic acid residue (V493D). In some embodiments, the valine residue at position 493 is replaced by a proline residue (V493P). In some embodiments, the valine residue at position 493 is replaced by a glycine residue (V493G).

[0091] In some embodiments, the adenosine deaminase comprises a mutation at alanine589 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the alanine residue at position 589 is replaced by a valine residue (A589V).

[0092] In some embodiments, the adenosine deaminase comprises a mutation at asparagine597 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the asparagine residue at position 597 is replaced by a lysine residue (N597K). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine

residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by an arginine residue (N597R). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by an alanine residue (N597A). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by a glutamic acid residue (N597E). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by a histidine residue (N597H). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by a glycine residue (N597G). In some embodiments, the adenosine deaminase comprises a mutation at position 597 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 597 is replaced by a tyrosine residue (N597Y). In some embodiments, the asparagine residue at position 597 is replaced by a phenylalanine residue (N597F). In some embodiments, the adenosine deaminase comprises mutation N597I. In some embodiments, the adenosine deaminase comprises mutation N597L. In some embodiments, the adenosine deaminase comprises mutation N597V. In some embodiments, the adenosine deaminase comprises mutation N597M. In some embodiments, the adenosine deaminase comprises mutation N597C. In some embodiments, the adenosine deaminase comprises mutation N597P. In some embodiments, the adenosine deaminase comprises mutation N597T. In some embodiments, the adenosine deaminase comprises mutation N597S. In some embodiments, the adenosine deaminase comprises mutation N597W. In some embodiments, the adenosine deaminase comprises mutation N597Q. In some embodiments, the adenosine deaminase comprises mutation N597D. In certain example embodiments, the mutations at N597 described above are further made in the context of an E488Q background

[0093] In some embodiments, the adenosine deaminase comprises a mutation at serine599 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR

protein. In some embodiments, the serine residue at position 599 is replaced by a threonine residue (S599T).

[0094] In some embodiments, the adenosine deaminase comprises a mutation at asparagine613 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the asparagine residue at position 613 is replaced by a lysine residue (N613K). In some embodiments, the adenosine deaminase comprises a mutation at position 613 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 613 is replaced by an arginine residue (N613R). In some embodiments, the adenosine deaminase comprises a mutation at position 613 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 613 is replaced by an alanine residue (N613A) In some embodiments, the adenosine deaminase comprises a mutation at position 613 of the amino acid sequence, which has an asparagine residue in the wild type sequence. In some embodiments, the asparagine residue at position 613 is replaced by a glutamic acid residue (N613E). In some embodiments, the adenosine deaminase comprises mutation N613I. In some embodiments, the adenosine deaminase comprises mutation N613L. In some embodiments, the adenosine deaminase comprises mutation N613V. In some embodiments, the adenosine deaminase comprises mutation N613F. In some embodiments, the adenosine deaminase comprises mutation N613M. In some embodiments, the adenosine deaminase comprises mutation N613C. In some embodiments, the adenosine deaminase comprises mutation N613G. In some embodiments, the adenosine deaminase comprises mutation N613P. In some embodiments, the adenosine deaminase comprises mutation N613T. In some embodiments, the adenosine deaminase comprises mutation N613S. In some embodiments, the adenosine deaminase comprises mutation N613Y. In some embodiments, the adenosine deaminase comprises mutation N613W. In some embodiments, the adenosine deaminase comprises mutation N613Q. In some embodiments, the adenosine deaminase comprises mutation N613H. In some embodiments, the adenosine deaminase comprises mutation N613D. In some embodiments, the mutations at N613 described above are further made in combination with a E488Q mutation.

[0095] In some embodiments, to improve editing efficiency, the adenosine deaminase may comprise one or more of the mutations: G336D, G487A, G487V, E488Q, E488H, E488R, E488N, E488A, E488S, E488M, T490C, T490S, V493T, V493S, V493A, V493R, V493D,

V493P, V493G, N597K, N597R, N597A, N597E, N597H, N597G, N597Y, A589V, S599T, N613K, N613R, N613A, N613E, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above.

[0096] In some embodiments, to reduce editing efficiency, the adenosine deaminase may comprise one or more of the mutations: E488F, E488L, E488W, T490A, T490F, T490Y, T490R, T490K, T490P, T490E, N597F, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In particular embodiments, it can be of interest to use an adenosine deaminase enzyme with reduced efficacy to reduce off-target effects.

[0097] In some embodiments, to reduce off-target effects, the adenosine deaminase comprises one or more of mutations at R348, V351, T375, K376, E396, C451, R455, N473, R474, K475, R477, R481, S486, E488, T490, S495, R510, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase comprises mutation at E488 and one or more additional positions selected from R348, V351, T375, K376, E396, C451, R455, N473, R474, K475, R477, R481, S486, T490, S495, R510. In some embodiments, the adenosine deaminase comprises mutation at T375, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at N473, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at V351, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at E488 and T375, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at E488 and N473, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation E488 and V351, and optionally at one or more additional positions. In some embodiments, the adenosine deaminase comprises mutation at E488 and one or more of T375, N473, and V351.

[0098] In some embodiments, to reduce off-target effects, the adenosine deaminase comprises one or more of mutations selected from R348E, V351L, T375G, T375S, R455G, R455S, R455E, N473D, R474E, K475Q, R477E, R481E, S486T, E488Q, T490A, T490S, S495T, and R510E, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase comprises mutation E488Q and one or more additional mutations selected from

R348E, V351L, T375G, T375S, R455G, R455S, R455E, N473D, R474E, K475Q, R477E, R481E, S486T, T490A, T490S, S495T, and R510E. In some embodiments, the adenosine deaminase comprises mutation T375G or T375S, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation N473D, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation V351L, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation E488Q, and T375G or T375G, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation E488Q and N473D, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation E488Q and V351L, and optionally one or more additional mutations. In some embodiments, the adenosine deaminase comprises mutation E488Q and one or more of T375G/S, N473D and V351L.

[0099] In certain examples, the adenosine deaminase protein or catalytic domain thereof has been modified to comprise a mutation at E488, preferably E488Q, of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein and/or wherein the adenosine deaminase protein or catalytic domain thereof has been modified to comprise a mutation at T375, preferably T375G of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In certain examples, the adenosine deaminase protein or catalytic domain thereof has been modified to comprise a mutation at E1008, preferably E1008Q, of the hADAR1d amino acid sequence, or a corresponding position in a homologous ADAR protein.

[0100] Crystal structures of the human ADAR2 deaminase domain bound to duplex RNA reveal a protein loop that binds the RNA on the 5' side of the modification site. This 5' binding loop is one contributor to substrate specificity differences between ADAR family members. See Wang et al., Nucleic Acids Res., 44(20):9872-9880 (2016), the content of which is incorporated herein by reference in its entirety. In addition, an ADAR2-specific RNA-binding loop was identified near the enzyme active site. See Mathews et al., Nat. Struct. Mol. Biol., 23(5):426-33 (2016), the content of which is incorporated herein by reference in its entirety. In some embodiments, the adenosine deaminase comprises one or more mutations in the RNA binding loop to improve editing specificity and/or efficiency.

[0101] In some embodiments, the adenosine deaminase comprises a mutation at alanine454 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR

protein. In some embodiments, the alanine residue at position 454 is replaced by a serine residue (A454S). In some embodiments, the alanine residue at position 454 is replaced by a cysteine residue (A454C). In some embodiments, the alanine residue at position 454 is replaced by an aspartic acid residue (A454D).

[0102] In some embodiments, the adenosine deaminase comprises a mutation at arginine455 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 455 is replaced by an alanine residue (R455A). In some embodiments, the arginine residue at position 455 is replaced by a valine residue (R455V). In some embodiments, the arginine residue at position 455 is replaced by a histidine residue (R455H). In some embodiments, the arginine residue at position 455 is replaced by a glycine residue (R455G). In some embodiments, the arginine residue at position 455 is replaced by a serine residue (R455S). In some embodiments, the arginine residue at position 455 is replaced by a glutamic acid residue (R455E). In some embodiments, the adenosine deaminase comprises mutation R455C. In some embodiments, the adenosine deaminase comprises mutation R455I. In some embodiments, the adenosine deaminase comprises mutation R455K. In some embodiments, the adenosine deaminase comprises mutation R455L. In some embodiments, the adenosine deaminase comprises mutation R455M. In some embodiments, the adenosine deaminase comprises mutation R455N. In some embodiments, the adenosine deaminase comprises mutation R455Q. In some embodiments, the adenosine deaminase comprises mutation R455F. In some embodiments, the adenosine deaminase comprises mutation R455W. In some embodiments, the adenosine deaminase comprises mutation R455P. In some embodiments, the adenosine deaminase comprises mutation R455Y. In some embodiments, the adenosine deaminase comprises mutation R455E. In some embodiments, the adenosine deaminase comprises mutation R455D. In some embodiments, the mutations at R455 described above are further made in combination with a E488Q mutation.

[0103] In some embodiments, the adenosine deaminase comprises a mutation at isoleucine456 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the isoleucine residue at position 456 is replaced by a valine residue (I456V). In some embodiments, the isoleucine residue at position 456 is replaced by a leucine residue (I456L). In some embodiments, the isoleucine residue at position 456 is replaced by an aspartic acid residue (I456D).

[0104] In some embodiments, the adenosine deaminase comprises a mutation at phenylalanine457 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the phenylalanine residue at position 457 is replaced by a tyrosine residue (F457Y). In some embodiments, the phenylalanine residue at position 457 is replaced by an arginine residue (F457R). In some embodiments, the phenylalanine residue at position 457 is replaced by a glutamic acid residue (F457E).

[0105] In some embodiments, the adenosine deaminase comprises a mutation at serine458 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the serine residue at position 458 is replaced by a valine residue (S458V). In some embodiments, the serine residue at position 458 is replaced by a phenylalanine residue (S458F). In some embodiments, the serine residue at position 458 is replaced by a proline residue (S458P). In some embodiments, the adenosine deaminase comprises mutation S458I. In some embodiments, the adenosine deaminase comprises mutation S458L. In some embodiments, the adenosine deaminase comprises mutation S458M. In some embodiments, the adenosine deaminase comprises mutation S458C. In some embodiments, the adenosine deaminase comprises mutation S458A. In some embodiments, the adenosine deaminase comprises mutation S458G. In some embodiments, the adenosine deaminase comprises mutation S458T. In some embodiments, the adenosine deaminase comprises mutation S458Y. In some embodiments, the adenosine deaminase comprises mutation S458W. In some embodiments, the adenosine deaminase comprises mutation S458Q. In some embodiments, the adenosine deaminase comprises mutation S458N. In some embodiments, the adenosine deaminase comprises mutation S458H. In some embodiments, the adenosine deaminase comprises mutation S458E. In some embodiments, the adenosine deaminase comprises mutation S458D. In some embodiments, the adenosine deaminase comprises mutation S458K. In some embodiments, the adenosine deaminase comprises mutation S458R. In some embodiments, the mutations at S458 described above are further made in combination with a E488Q mutation.

[0106] In some embodiments, the adenosine deaminase comprises a mutation at proline459 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the proline residue at position 459 is replaced by a cysteine residue (P459C). In some embodiments, the proline residue at position 459 is replaced by a

histidine residue (P459H). In some embodiments, the proline residue at position 459 is replaced by a tryptophan residue (P459W).

[0107] In some embodiments, the adenosine deaminase comprises a mutation at histidine460 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the histidine residue at position 460 is replaced by an arginine residue (H460R). In some embodiments, the histidine residue at position 460 is replaced by an isoleucine residue (H460I). In some embodiments, the histidine residue at position 460 is replaced by a proline residue (H460P). In some embodiments, the adenosine deaminase comprises mutation H460L. In some embodiments, the adenosine deaminase comprises mutation H460V. In some embodiments, the adenosine deaminase comprises mutation H460F. In some embodiments, the adenosine deaminase comprises mutation H460M. In some embodiments, the adenosine deaminase comprises mutation H460C. In some embodiments, the adenosine deaminase comprises mutation H460A. In some embodiments, the adenosine deaminase comprises mutation H460G. In some embodiments, the adenosine deaminase comprises mutation H460T. In some embodiments, the adenosine deaminase comprises mutation H460S. In some embodiments, the adenosine deaminase comprises mutation H460Y. In some embodiments, the adenosine deaminase comprises mutation H460W. In some embodiments, the adenosine deaminase comprises mutation H460Q. In some embodiments, the adenosine deaminase comprises mutation H460N. In some embodiments, the adenosine deaminase comprises mutation H460E. In some embodiments, the adenosine deaminase comprises mutation H460D. In some embodiments, the adenosine deaminase comprises mutation H460K. In some embodiments, the mutations at H460 described above are further made in combination with a E488Q mutation.

[0108] In some embodiments, the adenosine deaminase comprises a mutation at proline462 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the proline residue at position 462 is replaced by a serine residue (P462S). In some embodiments, the proline residue at position 462 is replaced by a tryptophan residue (P462W). In some embodiments, the proline residue at position 462 is replaced by a glutamic acid residue (P462E).

[0109] In some embodiments, the adenosine deaminase comprises a mutation at aspartic acid469 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the aspartic acid residue at position 469 is replaced by

a glutamine residue (D469Q). In some embodiments, the aspartic acid residue at position 469 is replaced by a serine residue (D469S). In some embodiments, the aspartic acid residue at position 469 is replaced by a tyrosine residue (D469Y).

[0110] In some embodiments, the adenosine deaminase comprises a mutation at arginine470 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 470 is replaced by an alanine residue (R470A). In some embodiments, the arginine residue at position 470 is replaced by an isoleucine residue (R470I). In some embodiments, the arginine residue at position 470 is replaced by an aspartic acid residue (R470D).

[0111] In some embodiments, the adenosine deaminase comprises a mutation at histidine471 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the histidine residue at position 471 is replaced by a lysine residue (H471K). In some embodiments, the histidine residue at position 471 is replaced by a threonine residue (H471T). In some embodiments, the histidine residue at position 471 is replaced by a valine residue (H471V).

[0112] In some embodiments, the adenosine deaminase comprises a mutation at proline472 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the proline residue at position 472 is replaced by a lysine residue (P472K). In some embodiments, the proline residue at position 472 is replaced by a threonine residue (P472T). In some embodiments, the proline residue at position 472 is replaced by an aspartic acid residue (P472D).

[0113] In some embodiments, the adenosine deaminase comprises a mutation at asparagine473 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the asparagine residue at position 473 is replaced by an arginine residue (N473R). In some embodiments, the asparagine residue at position 473 is replaced by a tryptophan residue (N473W). In some embodiments, the asparagine residue at position 473 is replaced by a proline residue (N473P). In some embodiments, the asparagine residue at position 473 is replaced by an aspartic acid residue (N473D).

[0114] In some embodiments, the adenosine deaminase comprises a mutation at arginine 474 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 474 is replaced by a

lysine residue (R474K). In some embodiments, the arginine residue at position 474 is replaced by a glycine residue (R474G). In some embodiments, the arginine residue at position 474 is replaced by an aspartic acid residue (R474D). In some embodiments, the arginine residue at position 474 is replaced by a glutamic acid residue (R474E).

[0115] In some embodiments, the adenosine deaminase comprises a mutation at lysine475 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the lysine residue at position 475 is replaced by a glutamine residue (K475Q). In some embodiments, the lysine residue at position 475 is replaced by an asparagine residue (K475N). In some embodiments, the lysine residue at position 475 is replaced by an aspartic acid residue (K475D).

[0116] In some embodiments, the adenosine deaminase comprises a mutation at alanine476 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the alanine residue at position 476 is replaced by a serine residue (A476S). In some embodiments, the alanine residue at position 476 is replaced by an arginine residue (A476R). In some embodiments, the alanine residue at position 476 is replaced by a glutamic acid residue (A476E).

[0117] In some embodiments, the adenosine deaminase comprises a mutation at arginine477 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 477 is replaced by a lysine residue (R477K). In some embodiments, the arginine residue at position 477 is replaced by a threonine residue (R477T). In some embodiments, the arginine residue at position 477 is replaced by a phenylalanine residue (R477F). In some embodiments, the arginine residue at position 474 is replaced by a glutamic acid residue (R477E).

[0118] In some embodiments, the adenosine deaminase comprises a mutation at glycine478 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 478 is replaced by an alanine residue (G478A). In some embodiments, the glycine residue at position 478 is replaced by an arginine residue (G478R). In some embodiments, the glycine residue at position 478 is replaced by a tyrosine residue (G478Y). In some embodiments, the adenosine deaminase comprises mutation G478I. In some embodiments, the adenosine deaminase comprises mutation G478L. In some embodiments, the adenosine deaminase comprises mutation G478V. In some embodiments, the adenosine deaminase comprises mutation G478F.

In some embodiments, the adenosine deaminase comprises mutation G478M. In some embodiments, the adenosine deaminase comprises mutation G478C. In some embodiments, the adenosine deaminase comprises mutation G478P. In some embodiments, the adenosine deaminase comprises mutation G478T. In some embodiments, the adenosine deaminase comprises mutation G478S. In some embodiments, the adenosine deaminase comprises mutation G478W. In some embodiments, the adenosine deaminase comprises mutation G478Q. In some embodiments, the adenosine deaminase comprises mutation G478N. In some embodiments, the adenosine deaminase comprises mutation G478H. In some embodiments, the adenosine deaminase comprises mutation G478E. In some embodiments, the adenosine deaminase comprises mutation G478D. In some embodiments, the adenosine deaminase comprises mutation G478K. In some embodiments, the mutations at G478 described above are further made in combination with a E488Q mutation.

[0119] In some embodiments, the adenosine deaminase comprises a mutation at glutamine479 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glutamine residue at position 479 is replaced by an asparagine residue (Q479N). In some embodiments, the glutamine residue at position 479 is replaced by a serine residue (Q479S). In some embodiments, the glutamine residue at position 479 is replaced by a proline residue (Q479P).

[0120] In some embodiments, the adenosine deaminase comprises a mutation at arginine348 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 348 is replaced by an alanine residue (R348A). In some embodiments, the arginine residue at position 348 is replaced by a glutamic acid residue (R348E).

[0121] In some embodiments, the adenosine deaminase comprises a mutation at valine351 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the valine residue at position 351 is replaced by a leucine residue (V351L). In some embodiments, the adenosine deaminase comprises mutation V351Y. In some embodiments, the adenosine deaminase comprises mutation V351M. In some embodiments, the adenosine deaminase comprises mutation V351T. In some embodiments, the adenosine deaminase comprises mutation V351G. In some embodiments, the adenosine deaminase comprises mutation V351A. In some embodiments, the adenosine deaminase comprises mutation V351F. In some embodiments, the adenosine deaminase comprises

mutation V351E. In some embodiments, the adenosine deaminase comprises mutation V351I. In some embodiments, the adenosine deaminase comprises mutation V351C. In some embodiments, the adenosine deaminase comprises mutation V351H. In some embodiments, the adenosine deaminase comprises mutation V351P. In some embodiments, the adenosine deaminase comprises mutation V351S. In some embodiments, the adenosine deaminase comprises mutation V351K. In some embodiments, the adenosine deaminase comprises mutation V351N. In some embodiments, the adenosine deaminase comprises mutation V351W. In some embodiments, the adenosine deaminase comprises mutation V351Q. In some embodiments, the adenosine deaminase comprises mutation V351D. In some embodiments, the adenosine deaminase comprises mutation V351R. In some embodiments, the mutations at V351 described above are further made in combination with a E488Q mutation.

[0122] In some embodiments, the adenosine deaminase comprises a mutation at threonine375 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the threonine residue at position 375 is replaced by a glycine residue (T375G). In some embodiments, the threonine residue at position 375 is replaced by a serine residue (T375S). In some embodiments, the adenosine deaminase comprises mutation T375H. In some embodiments, the adenosine deaminase comprises mutation T375Q. In some embodiments, the adenosine deaminase comprises mutation T375C. In some embodiments, the adenosine deaminase comprises mutation T375N. In some embodiments, the adenosine deaminase comprises mutation T375M. In some embodiments, the adenosine deaminase comprises mutation T375A. In some embodiments, the adenosine deaminase comprises mutation T375W. In some embodiments, the adenosine deaminase comprises mutation T375V. In some embodiments, the adenosine deaminase comprises mutation T375R. In some embodiments, the adenosine deaminase comprises mutation T375E. In some embodiments, the adenosine deaminase comprises mutation T375K. In some embodiments, the adenosine deaminase comprises mutation T375F. In some embodiments, the adenosine deaminase comprises mutation T375I. In some embodiments, the adenosine deaminase comprises mutation T375D. In some embodiments, the adenosine deaminase comprises mutation T375P. In some embodiments, the adenosine deaminase comprises mutation T375L. In some embodiments, the adenosine deaminase comprises mutation T375Y. In some embodiments, the mutations at T375Y described above are further made in combination with an E488Q mutation.

[0123] In some embodiments, the adenosine deaminase comprises a mutation at Arg481 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 481 is replaced by a glutamic acid residue (R481E).

[0124] In some embodiments, the adenosine deaminase comprises a mutation at Ser486 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the serine residue at position 486 is replaced by a threonine residue (S486T).

[0125] In some embodiments, the adenosine deaminase comprises a mutation at Thr490 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the threonine residue at position 490 is replaced by an alanine residue (T490A). In some embodiments, the threonine residue at position 490 is replaced by a serine residue (T490S).

[0126] In some embodiments, the adenosine deaminase comprises a mutation at Ser495 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the serine residue at position 495 is replaced by a threonine residue (S495T).

[0127] In some embodiments, the adenosine deaminase comprises a mutation at Arg510 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the arginine residue at position 510 is replaced by a glutamine residue (R510Q). In some embodiments, the arginine residue at position 510 is replaced by an alanine residue (R510A). In some embodiments, the arginine residue at position 510 is replaced by a glutamic acid residue (R510E).

[0128] In some embodiments, the adenosine deaminase comprises a mutation at Gly593 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 593 is replaced by an alanine residue (G593A). In some embodiments, the glycine residue at position 593 is replaced by a glutamic acid residue (G593E).

[0129] In some embodiments, the adenosine deaminase comprises a mutation at Lys594 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the lysine residue at position 594 is replaced by an alanine residue (K594A).

[0130] In some embodiments, the adenosine deaminase comprises a mutation at any one or more of positions A454, R455, I456, F457, S458, P459, H460, P462, D469, R470, H471, P472, N473, R474, K475, A476, R477, G478, Q479, R348, R510, G593, K594 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein.

[0131] In some embodiments, the adenosine deaminase comprises any one or more of mutations A454S, A454C, A454D, R455A, R455V, R455H, I456V, I456L, I456D, F457Y, F457R, F457E, S458V, S458F, S458P, P459C, P459H, P459W, H460R, H460I, H460P, P462S, P462W, P462E, D469Q, D469S, D469Y, R470A, R470I, R470D, H471K, H471T, H471V, P472K, P472T, P472D, N473R, N473W, N473P, R474K, R474G, R474D, K475Q, K475N, K475D, A476S, A476R, A476E, R477K, R477T, R477F, G478A, G478R, G478Y, Q479N, Q479S, Q479P, R348A, R510Q, R510A, G593A, G593E, K594A of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein.

[0132] In certain embodiments the adenosine deaminase is mutated to convert the activity to cytidine deaminase. Accordingly, in some embodiments, the adenosine deaminase comprises one or more mutations in positions selected from E396, C451, V351, R455, T375, K376, S486, Q488, R510, K594, R348, G593, S397, H443, L444, Y445, F442, E438, T448, A353, V355, T339, P539, T339, P539, V525 I520, P462 and N579. In particular embodiments, the adenosine deaminase comprises one or more mutations in a position selected from V351, L444, V355, V525 and I520. In some embodiments, the adenosine deaminase may comprise one or more of mutations at E488, V351, S486, T375, S370, P462, N597, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above.

[0133] In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, based on amino acid sequence positions of hADAR2-D, and

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

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

[0134] In some embodiments, the adenosine deaminase comprises a mutation at any one or more of positions T375, V351, G478, S458, H460 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein, optionally in combination a mutation at E488. In some embodiments, the adenosine deaminase comprises one or more of mutations selected from T375G, T375C, T375H, T375Q, V351M, V351T, V351Y, G478R, S458F, H460I, optionally in combination with E488Q.

[0135] In some embodiments, the adenosine deaminase comprises one or more of mutations selected from T375H, T375Q, V351M, V351Y, H460P, optionally in combination with E488Q.

[0136] In some embodiments, the adenosine deaminase comprises mutations T375S and S458F, optionally in combination with E488Q.

[0137] In some embodiments, the adenosine deaminase comprises a mutation at two or more of positions T375, N473, R474, G478, S458, P459, V351, R455, R455, T490, R348, Q479 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein, optionally in combination a mutation at E488. In some embodiments, the adenosine deaminase comprises two or more of mutations selected from T375G, T375S, N473D, R474E, G478R, S458F, P459W, V351L, R455G, R455S, T490A, R348E, Q479P, optionally in combination with E488Q.

[0138] In some embodiments, the adenosine deaminase comprises mutations T375G and V351L. In some embodiments, the adenosine deaminase comprises mutations T375G and R455G. In some embodiments, the adenosine deaminase comprises mutations T375G and R455S. In some embodiments, the adenosine deaminase comprises mutations T375G and T490A. In some embodiments, the adenosine deaminase comprises mutations T375G and R348E. In some embodiments, the adenosine deaminase comprises mutations T375S and V351L. In some embodiments, the adenosine deaminase comprises mutations T375S and R455G. In some embodiments, the adenosine deaminase comprises mutations T375S and R455S. In some embodiments, the adenosine deaminase comprises mutations T375S and T490A. In some embodiments, the adenosine deaminase comprises mutations T375S and R348E. In some embodiments, the adenosine deaminase comprises mutations N473D and V351L. In some embodiments, the adenosine deaminase comprises mutations N473D and R455G. In some embodiments, the adenosine deaminase comprises mutations N473D and R455S. In some embodiments, the adenosine deaminase comprises mutations N473D and T490A. In some embodiments, the adenosine deaminase comprises mutations N473D and R348E. In some embodiments, the adenosine deaminase comprises mutations R474E and V351L. In some embodiments, the adenosine deaminase comprises mutations R474E and R455G. In some embodiments, the adenosine deaminase comprises mutations R474E and R455S. In some embodiments, the adenosine deaminase comprises mutations R474E and T490A. In some embodiments, the adenosine deaminase comprises mutations R474E and R348E. In some embodiments, the adenosine deaminase comprises mutations S458F and T375G. In some embodiments, the adenosine deaminase comprises mutations S458F and T375S. In some embodiments, the adenosine deaminase comprises mutations S458F and N473D. In some embodiments, the adenosine deaminase comprises mutations S458F and R474E. In some embodiments, the adenosine deaminase comprises mutations S458F and

G478R. In some embodiments, the adenosine deaminase comprises mutations G478R and T375G. In some embodiments, the adenosine deaminase comprises mutations G478R and T375S. In some embodiments, the adenosine deaminase comprises mutations G478R and N473D. In some embodiments, the adenosine deaminase comprises mutations G478R and R474E. In some embodiments, the adenosine deaminase comprises mutations P459W and T375G. In some embodiments, the adenosine deaminase comprises mutations P459W and T375S. In some embodiments, the adenosine deaminase comprises mutations P459W and N473D. In some embodiments, the adenosine deaminase comprises mutations P459W and R474E. In some embodiments, the adenosine deaminase comprises mutations P459W and G478R. In some embodiments, the adenosine deaminase comprises mutations P459W and S458F. In some embodiments, the adenosine deaminase comprises mutations Q479P and T375G. In some embodiments, the adenosine deaminase comprises mutations Q479P and T375S. In some embodiments, the adenosine deaminase comprises mutations Q479P and N473D. In some embodiments, the adenosine deaminase comprises mutations Q479P and R474E. In some embodiments, the adenosine deaminase comprises mutations Q479P and G478R. In some embodiments, the adenosine deaminase comprises mutations Q479P and S458F. In some embodiments, the adenosine deaminase comprises mutations Q479P and P459W. All mutations described in this paragraph may also further be made in combination with a E488Q mutations.

[0139] In some embodiments, the adenosine deaminase comprises a mutation at any one or more of positions K475, Q479, P459, G478, S458of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein, optionally in combination a mutation at E488. In some embodiments, the adenosine deaminase comprises one or more of mutations selected from K475N, Q479N, P459W, G478R, S458P, S458F, optionally in combination with E488Q.

[0140] In some embodiments, the adenosine deaminase comprises a mutation at any one or more of positions T375, V351, R455, H460, A476 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein, optionally in combination a mutation at E488. In some embodiments, the adenosine deaminase comprises one or more of mutations selected from T375G, T375C, T375H, T375Q, V351M, V351T, V351Y, R455H, H460P, H460I, A476E, optionally in combination with E488Q.

[0141] In certain embodiments, improvement of editing and reduction of off-target modification is achieved by chemical modification of gRNAs. gRNAs which are chemically modified as exemplified in Vogel et al. (2014), Angew Chem Int Ed, 53:6267-6271, doi:10.1002/anie.201402634 (incorporated herein by reference in its entirety) reduce off-target activity and improve on-target efficiency. 2′-O-methyl and phosphothioate modified guide RNAs in general improve editing efficiency in cells.

[0142] ADAR has been known to demonstrate a preference for neighboring nucleotides on either side of the edited A (www.nature.com/nsmb/journal/v23/n5/full/nsmb.3203.html, Matthews et al. (2017), Nature Structural Mol Biol, 23(5): 426-433, incorporated herein by reference in its entirety). Accordingly, in certain embodiments, the gRNA, target, and/or ADAR is selected optimized for motif preference.

[0143] Intentional mismatches have been demonstrated *in vitro* to allow for editing of non-preferred motifs (https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gku272; Schneider et al (2014), Nucleic Acid Res, 42(10):e87); Fukuda et al. (2017), Scientific Reports, 7, doi:10.1038/srep41478, incorporated herein by reference in its entirety). Accordingly, in certain embodiments, to enhance RNA editing efficiency on non-preferred 5' or 3' neighboring bases, intentional mismatches in neighboring bases are introduced.

[0144] Results suggest that A's opposite C's in the targeting window of the ADAR deaminase domain are preferentially edited over other bases. Additionally, A's base-paired with U's within a few bases of the targeted base show low levels of editing by Type VII-ADAR fusions, suggesting that there is flexibility for the enzyme to edit multiple A's. These two observations suggest that multiple A's in the activity window of Type VII effector-ADAR fusions could be specified for editing by mismatching all A's to be edited with C's. Accordingly, in certain embodiments, multiple A:C mismatches in the activity window are designed to create multiple A:I edits. In certain embodiments, to suppress potential off-target editing in the activity window, non-target A's are paired with A's or G's.

[0145] The terms "editing specificity" and "editing preference" are used interchangeably herein to refer to the extent of A-to-I editing at a particular adenosine site in a double-stranded substrate. In some embodiment, the substrate editing preference is determined by the 5' nearest neighbor and/or the 3' nearest neighbor of the target adenosine residue. In some embodiments, the adenosine deaminase has preference for the 5' nearest neighbor of the substrate ranked as U>A>C>G (">" indicates greater preference). In some embodiments, the adenosine deaminase

has preference for the 3' nearest neighbor of the substrate ranked as G>C~A>U (">" indicates greater preference; "~" indicates similar preference). In some embodiments, the adenosine deaminase has preference for the 3' nearest neighbor of the substrate ranked as G>C>U~A (">" indicates greater preference; "~" indicates similar preference). In some embodiments, the adenosine deaminase has preference for the 3' nearest neighbor of the substrate ranked as G>C>A>U (">" indicates greater preference). In some embodiments, the adenosine deaminase has preference for the 3' nearest neighbor of the substrate ranked as C~G~A>U (">" indicates greater preference; "~" indicates similar preference). In some embodiments, the adenosine deaminase has preference for a triplet sequence containing the target adenosine residue ranked as TAG>AAG>CAC>AAT>GAA>GAC (">" indicates greater preference), the center A being the target adenosine residue.

[0146] In some embodiments, the substrate editing preference of an adenosine deaminase is affected by the presence or absence of a nucleic acid binding domain in the adenosine deaminase protein. In some embodiments, to modify substrate editing preference, the deaminase domain is connected with a double-strand RNA binding domain (dsRBD) or a double-strand RNA binding motif (dsRBM). In some embodiments, the dsRBD or dsRBM may be derived from an ADAR protein, such as hADAR1 or hADAR2. In some embodiments, a full length ADAR protein that comprises at least one dsRBD and a deaminase domain is used. In some embodiments, the one or more dsRBM or dsRBD is at the N-terminus of the deaminase domain. In other embodiments, the one or more dsRBM or dsRBD is at the C-terminus of the deaminase domain.

[0147] In some embodiments, the substrate editing preference of an adenosine deaminase is affected by amino acid residues near or in the active center of the enzyme. In some embodiments, to modify substrate editing preference, the adenosine deaminase may comprise one or more of the mutations: G336D, G487R, G487K, G487W, G487Y, E488Q, E488N, T490A, V493A, V493T, V493S, N597K, N597R, A589V, S599T, N613K, N613R, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above.

[0148] Particularly, in some embodiments, to reduce editing specificity, the adenosine deaminase can comprise one or more of mutations E488Q, V493A, N597K, N613K, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein

corresponding to the above. In some embodiments, to increase editing specificity, the adenosine deaminase can comprise mutation T490A.

[0149] In some embodiments, to increase editing preference for target adenosine (A) with an immediate 5' G, such as substrates comprising the triplet sequence GAC, the center A being the target adenosine residue, the adenosine deaminase can comprise one or more of mutations G336D, E488Q, E488N, V493T, V493S, V493A, A589V, N597K, N597R, S599T, N613K, N613R, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above.

[0150] Particularly, in some embodiments, the adenosine deaminase comprises mutation E488Q or a corresponding mutation in a homologous ADAR protein for editing substrates comprising the following triplet sequences: GAC, GAA, GAU, GAG, CAU, AAU, UAC, the center A being the target adenosine residue.

[0151] In some embodiments, the adenosine deaminase comprises the wild-type amino acid sequence of hADAR1-D:

MGSGGGGSEGAPKKKRKVGSSLGTGNRCVKGDSLSLKGETVNDCHAEIISRRGFIRF LYSELMKYNSQTAKDSIFEPAKGGEKLQIKKTVSFHLYISTAPCGDGALFDKSCSDRA MESTESRHYPVFENPKQGKLRTKVENGQGTIPVESSDIVPTWDGIRLGERLRTMSCSD KILRWNVLGLQGALLTHFLQPIYLKSVTLGYLFSQGHLTRAICCRVTRDGSAFEDGL RHPFIVNHPKVGRVSIYDSKRQSGKTKETSVNWCLADGYDLEILDGTRGTVDGPRNE LSRVSKKNIFLLFKKLCSFRYR DLLRLSYGEA KKAARDYETAKNYFKKGLKD MGYGNWISKPQEEKNF (**SEQ ID NO: 14**)

[0152] In some embodiments, the adenosine deaminase comprises one or more mutations in the hADAR1-D sequence, such that the editing efficiency, and/or substrate editing preference of hADAR1-D is changed according to specific needs.

[0153] In some embodiments, the adenosine deaminase comprises a mutation at Glycine1007 of the hADAR1-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glycine residue at position 1007 is replaced by a non-polar amino acid residue with relatively small side chains. For example, in some embodiments, the glycine residue at position 1007 is replaced by an alanine residue (G1007A). In some embodiments, the glycine residue at position 1007 is replaced by a valine residue (G1007V). In some embodiments, the glycine residue at position 1007 is replaced by an amino acid residue with relatively large side chains. In some embodiments, the glycine

residue at position 1007 is replaced by an arginine residue (G1007R). In some embodiments, the glycine residue at position 1007 is replaced by a lysine residue (G1007K). In some embodiments, the glycine residue at position 1007 is replaced by a tryptophan residue (G1007W). In some embodiments, the glycine residue at position 1007 is replaced by a tyrosine residue (G1007Y). Additionally, in other embodiments, the glycine residue at position 1007 is replaced by a leucine residue (G1007L). In other embodiments, the glycine residue at position 1007 is replaced by a threonine residue (G1007T). In other embodiments, the glycine residue at position 1007 is replaced by a serine residue (G1007S).

[0154] In some embodiments, the adenosine deaminase comprises a mutation at glutamic acid1008 of the hADAR1-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the glutamic acid residue at position 1008 is replaced by a polar amino acid residue having a relatively large side chain. In some embodiments, the glutamic acid residue at position 1008 is replaced by a glutamine residue (E1008Q). In some embodiments, the glutamic acid residue at position 1008 is replaced by a histidine residue (E1008H). In some embodiments, the glutamic acid residue at position 1008 is replaced by an arginine residue (E1008R). In some embodiments, the glutamic acid residue at position 1008 is replaced by a lysine residue (E1008K). In some embodiments, the glutamic acid residue at position 1008 is replaced by a nonpolar or small polar amino acid residue. In some embodiments, the glutamic acid residue at position 1008 is replaced by a phenylalanine residue (E1008F). In some embodiments, the glutamic acid residue at position 1008 is replaced by a tryptophan residue (E1008W). In some embodiments, the glutamic acid residue at position 1008 is replaced by a glycine residue (E1008G). In some embodiments, the glutamic acid residue at position 1008 is replaced by an isoleucine residue (E1008I). In some embodiments, the glutamic acid residue at position 1008 is replaced by a valine residue (E1008V). In some embodiments, the glutamic acid residue at position 1008 is replaced by a proline residue (E1008P). In some embodiments, the glutamic acid residue at position 1008 is replaced by a serine residue (E1008S). In other embodiments, the glutamic acid residue at position 1008 is replaced by an asparagine residue (E1008N). In other embodiments, the glutamic acid residue at position 1008 is replaced by an alanine residue (E1008A). In other embodiments, the glutamic acid residue at position 1008 is replaced by a Methionine residue (E1008M). In some embodiments, the glutamic acid residue at position 1008 is replaced by a leucine residue (E1008L).

[0155] In some embodiments, to improve editing efficiency, the adenosine deaminase may comprise one or more of the mutations: E1007S, E1007A, E1007V, E1008Q, E1008R, E1008H, E1008M, E1008N, E1008K, based on amino acid sequence positions of hADAR1-D, and mutations in a homologous ADAR protein corresponding to the above.

[0156] In some embodiments, to reduce editing efficiency, the adenosine deaminase may comprise one or more of the mutations: E1007R, E1007K, E1007Y, E1007L, E1007T, E1008G, E1008I, E1008P, E1008V, E1008F, E1008W, E1008S, E1008N, E1008K, based on amino acid sequence positions of hADAR1-D, and mutations in a homologous ADAR protein corresponding to the above.

[0157] In some embodiments, the substrate editing preference, efficiency and/or selectivity of an adenosine deaminase is affected by amino acid residues near or in the active center of the enzyme. In some embodiments, the adenosine deaminase comprises a mutation at the glutamic acid 1008 position in hADAR1-D sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the mutation is E1008R, or a corresponding mutation in a homologous ADAR protein. In some embodiments, the E1008R mutant has an increased editing efficiency for target adenosine residue that has a mismatched G residue on the opposite strand.

[0158] In some embodiments, the adenosine deaminase protein further comprises or is connected to one or more double-stranded RNA (dsRNA) binding motifs (dsRBMs) or domains (dsRBDs) for recognizing and binding to double-stranded nucleic acid substrates. In some embodiments, the interaction between the adenosine deaminase and the double–stranded substrate is mediated by one or more additional protein factor(s), including a CRISPR/CAS protein factor. In some embodiments, the interaction between the adenosine deaminase and the double-stranded substrate is further mediated by one or more nucleic acid component(s), including a guide RNA.

[0159] In certain example embodiments, directed evolution may be used to design modified ADAR proteins capable of catalyzing additional reactions besides deamination of a adenine to a hypoxanthine.

Modified Adenosine Deaminase Having C to U Deamination Activity

[0160] In certain example embodiments, directed evolution may be used to design modified ADAR proteins capable of catalyzing additional reactions besides deamination of an adenine to a hypoxanthine. For example, the modified ADAR protein may be capable of

catalyzing deamination of a cytidine to a uracil. While not bound by a particular theory, mutations that improve C to U activity may alter the shape of the binding pocket to be more amenable to the smaller cytidine base.

[0161] In some embodiments, the modified adenosine deaminase having C-to-U deamination activity comprises a mutation at any one or more of positions V351, T375, R455, and E488 of the hADAR2-D amino acid sequence, or a corresponding position in a homologous ADAR protein. In some embodiments, the adenosine deaminase comprises mutation E488Q. In some embodiments, the adenosine deaminase comprises one or more of mutations selected from V351I, V351L, V351F, V351M, V351C, V351A, V351G, V351P, V351T, V351S, V351Y, V351W, V351Q, V351N, V351H, V351E, V351D, V351K, V351R, T375I, T375L, T375V, T375F, T375M, T375C, T375A, T375G, T375P, T375S, T375Y, T375W, T375Q, T375N, T375H, T375E, T375D, T375K, T375R, R455I, R455L, R455V, R455F, R455M, R455C, R455A, R455G, R455P, R455T, R455S, R455Y, R455W, R455Q, R455N, R455H, R455E, R455D, R455K. In some embodiments, the adenosine deaminase comprises mutation E488Q, and further comprises one or more of mutations selected from V351I, V351L, V351F, V351M, V351C, V351A, V351G, V351P, V351T, V351S, V351Y, V351W, V351Q, V351N, V351H, V351E, V351D, V351K, V351R, T375I, T375L, T375V, T375F, T375M, T375C, T375A, T375G, T375P, T375S, T375Y, T375W, T375Q, T375N, T375H, T375E, T375D, T375K, T375R, R455I, R455L, R455V, R455F, R455M, R455C, R455A, R455G, R455P, R455T, R455S, R455Y, R455W, R455Q, R455N, R455H, R455E, R455D, R455K.

[0162] In connection with the aforementioned modified ADAR protein having C-to-U deamination activity, the invention described herein also relates to a method for deaminating a C in a target RNA sequence of interest, comprising delivering to a target RNA or DNA an AD-functionalized composition disclosed herein.

[0163] In certain example embodiments, the method for deaminating a C in a target RNA sequence comprising delivering to said target RNA: (a) a catalytically inactive (dead) Cas; (b) a guide molecule which comprises a guide sequence linked to a direct repeat sequence; and (c) a modified ADAR protein having C-to-U deamination activity or catalytic domain thereof; wherein said modified ADAR protein or catalytic domain thereof is covalently or non-covalently linked to said dead Cas protein or said guide molecule or is adapted to link thereto after delivery; wherein guide molecule forms a complex with said dead Cas protein and directs

said complex to bind said target RNA sequence of interest; wherein said guide sequence is capable of hybridizing with a target sequence comprising said C to form an RNA duplex; wherein, optionally, said guide sequence comprises a non-pairing A or U at a position corresponding to said C resulting in a mismatch in the RNA duplex formed; and wherein said modified ADAR protein or catalytic domain thereof *deaminates* said C in said RNA duplex.

[0164] In connection with the aforementioned modified ADAR protein having C-to-U deamination activity, the invention described herein further relates to an engineered, non-naturally occurring system suitable for deaminating a C in a target locus of interest, comprising: (a) a guide molecule which comprises a guide sequence linked to a direct repeat sequence, or a nucleotide sequence encoding said guide molecule; (b) a catalytically inactive Type VII protein, or a nucleotide sequence encoding said catalytically inactive Type VII protein; (c) a modified ADAR protein having C-to-U deamination activity or catalytic domain thereof, or a nucleotide sequence encoding said modified ADAR protein or catalytic domain thereof; wherein said modified ADAR protein or catalytic domain thereof is covalently or non-covalently linked to said Type VII protein or said guide molecule or is adapted to link thereto after delivery; wherein said guide sequence is capable of hybridizing with a target RNA sequence comprising a C to form an RNA duplex; wherein, optionally, said guide sequence comprises a non-pairing A or U at a position corresponding to said C resulting in a mismatch in the RNA duplex formed; wherein, optionally, the system is a vector system comprising one or more vectors comprising: (a) a first regulatory element operably linked to a nucleotide sequence encoding said guide molecule which comprises said guide sequence, (b) a second regulatory element operably linked to a nucleotide sequence encoding said catalytically inactive Type VII protein; and (c) a nucleotide sequence encoding a modified ADAR protein having C-to-U deamination activity or catalytic domain thereof which is under control of said first or second regulatory element or operably linked to a third regulatory element; wherein, if said nucleotide sequence encoding a modified ADAR protein or catalytic domain thereof is operably linked to a third regulatory element, said modified ADAR protein or catalytic domain thereof is adapted to link to said guide molecule or said Type VII effector protein after expression; wherein components (a), (b) and (c) are located on the same or different vectors of the system, optionally wherein said first, second, and/or third regulatory element is an inducible promoter.

[0165] In an embodiment of the invention, the substrate of the adenosine deaminase is an RNA/DNA heteroduplex formed upon binding of the guide molecule to its DNA target which then forms the CRISPR-Cas complex with the CRISPR-Cas enzyme. The RNA/DNA or DNA/RNA heteroduplex is also referred to herein as the "RNA/DNA hybrid", "DNA/RNA hybrid" or "double-stranded substrate".

[0166] According to the present invention, the substrate of the adenosine deaminase is an RNA/DNAn RNA duplex formed upon binding of the guide molecule to its DNA target which then forms the CRISPR-Cas complex with the CRISPR-Cas enzyme. The substrate of the adenosine deaminase can also be an RNA/RNA duplex formed upon binding of the guide molecule to its RNA target which then forms the CRISPR-Cas complex with the CRISPR-Cas enzyme. The RNA/DNA or DNA/RNAn RNA duplex is also referred to herein as the "RNA/DNA hybrid", "DNA/RNA hybrid" or "double-stranded substrate". The particular features of the guide molecule and CRISPR-Cas enzyme are detailed below.

[0167] The term "editing selectivity" as used herein refers to the fraction of all sites on a double-stranded substrate that is edited by an adenosine deaminase. Without being bound by theory, it is contemplated that editing selectivity of an adenosine deaminase is affected by the double-stranded substrate's length and secondary structures, such as the presence of mismatched bases, bulges and/or internal loops.

[0168] In some embodiments, when the substrate is a perfectly base-paired duplex longer than 50 bp, the adenosine deaminase may be able to deaminate multiple adenosine residues within the duplex (e.g., 50% of all adenosine residues). In some embodiments, when the substrate is shorter than 50 bp, the editing selectivity of an adenosine deaminase is affected by the presence of a mismatch at the target adenosine site. Particularly, in some embodiments, adenosine (A) residue having a mismatched cytidine (C) residue on the opposite strand is deaminated with high efficiency. In some embodiments, adenosine (A) residue having a mismatched guanosine (G) residue on the opposite strand is skipped without editing.

[0169] In particular embodiments, the adenosine deaminase protein or catalytic domain thereof is delivered to the cell or expressed within the cell as a separate protein, but is modified so as to be able to link to either the Type VII effector protein or the guide molecule. In particular embodiments, this is ensured by the use of orthogonal RNA-binding protein or adaptor protein / aptamer combinations that exist within the diversity of bacteriophage coat proteins. Examples of such coat proteins include but are not limited to: MS2, Qβ, F2, GA, fr, JP501, M12, R17,

BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1. Aptamers can be naturally occurring or synthetic oligonucleotides that have been engineered through repeated rounds of *in vitro* selection or SELEX (systematic evolution of ligands by exponential enrichment) to bind to a specific target.

[0170] In particular embodiments, the guide molecule is provided with one or more distinct RNA loop(s) or distinct sequence(s) that can recruit an adaptor protein. A guide molecule may be extended, without colliding with the Type VII effector protein by the insertion of distinct RNA loop(s) or distinct sequence(s) that may recruit adaptor proteins that can bind to the distinct RNA loop(s) or distinct sequence(s). Examples of modified guides and their use in recruiting effector domains to the C2c1 complex are provided in Konermann (Nature 2015, 517(7536): 583–588). In particular embodiments, the aptamer is a minimal hairpin aptamer which selectively binds dimerized MS2 bacteriophage coat proteins in mammalian cells and is introduced into the guide molecule, such as in the stem loop and/or in a tetraloop. In these embodiments, the adenosine deaminase protein is fused to MS2. The adenosine deaminase protein is then co-delivered together with the Type VII effector protein and corresponding guide RNA.

[0171] In some embodiments, the Type VII effector-ADAR base editing system described herein comprises (a) a Type VII effector protein, which is catalytically inactive or a nickase; (b) a guide molecule which comprises a guide sequence; and (c) an adenosine deaminase protein or catalytic domain thereof; wherein the adenosine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to the Type VII effector protein or the guide molecule or is adapted to link thereto after delivery; wherein the guide sequence is substantially complementary to the target sequence but comprises a non-pairing C corresponding to the A being targeted for deamination, resulting in a A-C mismatch in a DNA-RNA or RNA-RNA duplex formed by the guide sequence and the target sequence. For application in eukaryotic cells, the Type VII effector protein and/or the adenosine deaminase are preferably NLS-tagged.

[0172] In some embodiments, the components (a), (b) and (c) are delivered to the cell as a ribonucleoprotein complex. The ribonucleoprotein complex can be delivered via one or more lipid nanoparticles.

[0173] In some embodiments, the components (a), (b) and (c) are delivered to the cell as one or more RNA molecules, such as one or more guide RNAs and one or more mRNA

molecules encoding the Type VII effector protein, the adenosine deaminase protein, and optionally the adaptor protein. The RNA molecules can be delivered via one or more lipid nanoparticles.

[0174] In some embodiments, the components (a), (b) and (c) are delivered to the cell as one or more DNA molecules. In some embodiments, the one or more DNA molecules are comprised within one or more vectors such as viral vectors (e.g., AAV). In some embodiments, the one or more DNA molecules comprise one or more regulatory elements operably configured to express the Type VII effector protein, the guide molecule, and the adenosine deaminase protein or catalytic domain thereof, optionally wherein the one or more regulatory elements comprise inducible promoters.

[0175] In some embodiments of the guide molecule is capable of hybridizing with a target sequence comprising the Adenine to be deaminated within a first DNA strand or a RNA strand at the target locus to form a DNA-RNA or RNA-RNA duplex which comprises a non-pairing Cytosine opposite to said Adenine. Upon duplex formation, the guide molecule forms a complex with the Type VII effector protein and directs the complex to bind said first DNA strand or said RNA strand at the target locus of interest. Details on the aspect of the guide of the Type VII effector-ADAR base editing system are provided herein below.

[0176] In some embodiments, a Type VII effector guide RNA is used to form a DNA-RNA or RNA-RNA duplex with the target DNA or RNA. In some embodiments, a Type VII guide molecule longer than the canonical length is used to form a DNA-RNA or RNA-RNA duplex with the target DNA or RNA including outside of the Type VII-guide RNA-target DNA complex. In certain example embodiments, the guide sequence has a length of about 29-53 nt capable of forming a DNA-RNA or RNA-RNA duplex with said target sequence. In certain other example embodiments, the guide sequence has a length of about 40-50 nt capable of forming a DNA-RNA or RNA-RNA duplex with said target sequence. In certain example embodiments, the distance between said non-pairing C and the 5' end of said guide sequence is 20-30 nucleotides. In certain example embodiments, the distance between said non-pairing C and the 3' end of said guide sequence is 20-30 nucleotides.

[0177] In at least a first design, the Type VII CRISPR-ADAR system comprises (a) an adenosine deaminase fused or linked to a Type VII CRISPR protein, wherein the Type VII CRISPR protein is catalytically inactive or a nickase, and (b) a guide molecule comprising a guide sequence designed to introduce a A-C mismatch in a DNA-RNA or RNA-RNA duplex

formed between the guide sequence and the target sequence. In some embodiments, the Type VII CRISPR protein and/or the adenosine deaminase are NLS-tagged, on either the N- or C-terminus or both.

[0178] In at least a second design, the Type VII CRISPR-ADAR system comprises (a) a Type VII CRISPR protein that is catalytically inactive or a nickase, (b) a guide molecule comprising a guide sequence designed to introduce a A-C mismatch in a DNA-RNA or RNA-RNA duplex formed between the guide sequence and the target sequence, and an aptamer sequence (e.g., MS2 RNA motif or PP7 RNA motif) capable of binding to an adaptor protein (e.g., MS2 coating protein or PP7 coat protein), and (c) an adenosine deaminase fused or linked to an adaptor protein, wherein the binding of the aptamer and the adaptor protein recruits the adenosine deaminase to the DNA-RNA or RNA-RNA duplex formed between the guide sequence and the target sequence for targeted deamination at the A of the A-C mismatch. In some embodiments, the adaptor protein and/or the adenosine deaminase are NLS-tagged, on either the N- or C-terminus or both. The Type VII CRISPR protein can also be NLS-tagged.

[0179] The use of different aptamers and corresponding adaptor proteins also allows orthogonal gene editing to be implemented. In one example in which adenosine deaminase are used in combination with cytidine deaminase for orthogonal gene editing/deamination, sgRNA targeting different loci are modified with distinct RNA loops in order to recruit MS2-adenosine deaminase and PP7-cytidine deaminase (or PP7-adenosine deaminase and MS2-cytidine deaminase), respectively, resulting in orthogonal deamination of A or C at the target loci of interested, respectively. PP7 is the RNA-binding coat protein of the bacteriophage Pseudomonas. Like MS2, it binds a specific RNA sequence and secondary structure. The PP7 RNA-recognition motif is distinct from that of MS2. Consequently, PP7 and MS2 can be multiplexed to mediate distinct effects at different genomic loci simultaneously. For example, an sgRNA targeting locus A can be modified with MS2 loops, recruiting MS2-adenosine deaminase, while another sgRNA targeting locus B can be modified with PP7 loops, recruiting PP7-cytidine deaminase. In the same cell, orthogonal, locus-specific modifications are thus realized. This principle can be extended to incorporate other orthogonal RNA-binding proteins.

[0180] In at least a third design, the Type VII CRISPR-ADAR CRISPR system comprises (a) an adenosine deaminase inserted into an internal loop or unstructured region of a Type VII CRISPR protein, wherein the Type VII CRISPR protein is catalytically inactive or a nickase, and (b) a guide molecule comprising a guide sequence designed to introduce a A-C mismatch

in a DNA-RNA or RNA-RNA duplex formed between the guide sequence and the target sequence.

[0181] Type VII CRISPR protein split sites that are suitable for insertion of adenosine deaminase can be identified with the help of a crystal structure. For example, with respect to Type VII CRISPR mutants, it should be readily apparent what is the corresponding position for, for example, a sequence alignment. For other Type VII effector protein one can use the crystal structure of an ortholog if a relatively high degree of homology exists between the ortholog and the intended Type VII protein.

[0182] The split position may be located within a region or loop. Preferably, the split position occurs where an interruption of the amino acid sequence does not result in the partial or full destruction of a structural feature (e.g. alpha-helixes or β-sheets). Unstructured regions (regions that did not show up in the crystal structure because these regions are not structured enough to be "frozen" in a crystal) are often preferred options. Splits in all unstructured regions that are exposed on the surface of Type VII CRISPR protein are envisioned in the practice of the invention. The positions within the unstructured regions or outside loops may not need to be exactly the numbers provided above, but may vary by, for example 1, 2, 3, 4, 5, 6, 7, 8, 9, or even 10 amino acids either side of the position given above, depending on the size of the loop, so long as the split position still falls within an unstructured region of outside loop.

[0183] The Type VII CRISPR-ADAR system described herein can be used to target a specific Adenine within a DNA sequence for deamination. For example, the guide molecule can form a complex with the Type VII CRISPR protein and directs the complex to bind a target sequence at the target locus of interest. Because the guide sequence is designed to have a non-pairing C, the heteroduplex formed between the guide sequence and the target sequence comprises a A-C mismatch, which directs the adenosine deaminase to contact and deaminate the A opposite to the non-pairing C, converting it to a Inosine (I). Since Inosine (I) base pairs with C and functions like G in cellular process, the targeted deamination of A described herein are useful for correction of undesirable G-A and C-T mutations, as well as for obtaining desirable A-G and T-C mutations.

Base Excision Repair Inhibitor

[0184] In some embodiments, the AD-functionalized CRISPR system further comprises a base excision repair (BER) inhibitor. Without wishing to be bound by any particular theory, cellular DNA-repair response to the presence of I:T pairing may be responsible for a decrease

in nucleobase editing efficiency in cells. Alkyladenine DNA glycosylase (also known as DNA-3-methyladenine glycosylase, 3-alkyladenine DNA glycosylase, or N-methylpurine DNA glycosylase) catalyzes removal of hypoxanthine from DNA in cells, which may initiate base excision repair, with reversion of the I:T pair to a A:T pair as outcome.

[0185] In some embodiments, the BER inhibitor is an inhibitor of alkyladenine DNA glycosylase. In some embodiments, the BER inhibitor is an inhibitor of human alkyladenine DNA glycosylase. In some embodiments, the BER inhibitor is a polypeptide inhibitor. In some embodiments, the BER inhibitor is a protein that binds hypoxanthine. In some embodiments, the BER inhibitor is a protein that binds hypoxanthine in DNA. In some embodiments, the BER inhibitor is a catalytically inactive alkyladenine DNA glycosylase protein or binding domain thereof. In some embodiments, the BER inhibitor is a catalytically inactive alkyladenine DNA glycosylase protein or binding domain thereof that does not excise hypoxanthine from the DNA. Other proteins that are capable of inhibiting (e.g., sterically blocking) an alkyladenine DNA glycosylase base-excision repair enzyme are within the scope of this disclosure. Additionally, any proteins that block or inhibit base-excision repair as also within the scope of this disclosure.

[0186] Without wishing to be bound by any particular theory, base excision repair may be inhibited by molecules that bind the edited strand, block the edited base, inhibit alkyladenine DNA glycosylase, inhibit base excision repair, protect the edited base, and/or promote fixing of the non-edited strand. It is believed that the use of the BER inhibitor described herein can increase the editing efficiency of an adenosine deaminase that is capable of catalyzing a A to I change.

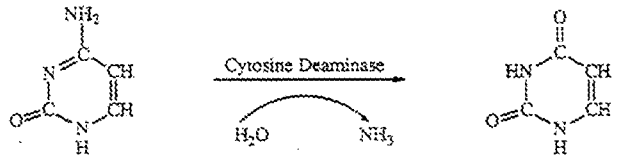
[0187] Accordingly, in the first design of the AD-functionalized CRISPR system discussed above, the CRISPR-Cas protein or the adenosine deaminase can be fused to or linked to a BER inhibitor (e.g., an inhibitor of alkyladenine DNA glycosylase). In some embodiments, the BER inhibitor can be comprised in one of the following structures (nVII = Type VII CRISPR nickase; dVII =dead Type VII CRISPR): [AD]-[optional linker]-[nVII/dVII]-[optional linker]-[BER inhibitor]; [AD]-[optional linker]-[BER inhibitor]-[optional linker]-[nVII/dVII]; [BER inhibitor]-[optional linker]-[AD]-[optional linker]-[nVII/dVII]; [BER inhibitor]-[optional linker]-[nVII/dVII]-[optional linker]-[AD]; [nVII/dVII]-[optional linker]-[AD]-[optional linker]-[BER inhibitor]; [nVII/dVII]-[optional linker]-[BER inhibitor]-[optional linker]-[AD].

[0188] Similarly, in the second design of the AD-functionalized CRISPR system discussed above, the CRISPR-Cas protein, the adenosine deaminase, or the adaptor protein can be fused to or linked to a BER inhibitor (e.g., an inhibitor of alkyladenine DNA glycosylase). In some embodiments, the BER inhibitor can be comprised in one of the following structures: [nVII/dVII]-[optional linker]-[BER inhibitor]; [BER inhibitor]-[optional linker]-[nVII/dVII]; [AD]-[optional linker]-[Adaptor]-[optional linker]-[BER inhibitor]; [AD]-[optional linker]-[BER inhibitor]-[optional linker]-[Adaptor]; [BER inhibitor]-[optional linker]-[AD]-[optional linker]-[Adaptor]; [BER inhibitor]-[optional linker]-[Adaptor]-[optional linker]-[AD]; [Adaptor]-[optional linker]-[AD]-[optional linker]-[BER inhibitor]; [Adaptor]-[optional linker]-[BER inhibitor]-[optional linker]-[AD].

[0189] In the third design of the AD-functionalized CRISPR system discussed above, the BER inhibitor can be inserted into an internal loop or unstructured region of a CRISPR-Cas protein.

Cytidine deaminase

[0190] In some embodiments, the deaminase is a cytidine deaminase. The term "cytidine deaminase" or "cytidine deaminase protein" as used herein refers to a protein, a polypeptide, or one or more functional domain(s) of a protein or a polypeptide that is capable of catalyzing a hydrolytic deamination reaction that converts a cytosine (or a cytosine moiety of a molecule) to an uracil (or a uracil moiety of a molecule), as shown below. In some embodiments, the cytosine-containing molecule is an cytidine (C), and the uracil-containing molecule is an uridine (U). The cytosine-containing molecule can be deoxyribonucleic acid (DNA) or ribonucleic acid (RNA).



Cytosine
(4-amino-2-oxopyrimidine)

Uracil
(2,4-dioxopyrimidine)

[0191]

[0192] According to the present disclosure, cytidine deaminases that can be used in connection with the present disclosure include, but are not limited to, members of the enzyme family known as apolipoprotein B mRNA-editing complex (APOBEC) family deaminase, an activation-induced deaminase (AID), or a cytidine deaminase 1 (CDA1). In particular

embodiments, the deaminase in an APOBEC1 deaminase, an APOBEC2 deaminase, an APOBEC3A deaminase, an APOBEC3B deaminase, an APOBEC3C deaminase, and APOBEC3D deaminase, an APOBEC3E deaminase, an APOBEC3F deaminase an APOBEC3G deaminase, an APOBEC3H deaminase, or an APOBEC4 deaminase.

[0193] In the methods and systems of the present invention, the cytidine deaminase is capable of targeting Cytosine in a DNA single strand. In certain example embodiments the cytidine deaminase may edit on a single strand present outside of the binding component e.g. bound Type VII effector. In other example embodiments, the cytodine deaminasemay edit at a localized bubble, such as a localized bubble formed by a mismatch at the target edit site but the guide sequence. In certain example embodiments the cytodine deaminase may contain mutations that help focus the area of activity such as those disclosed in Kim et al., Nature Biotechnology (2017) 35(4):371-377 (doi:10.1038/nbt.3803.

[0194] In some embodiments, the cytidine deaminase is derived from one or more metazoa species, including but not limited to, mammals, birds, frogs, squids, fish, flies and worms. In some embodiments, the cytidine deaminase is a human, primate, cow, dog rat or mouse cytidine deaminase.

[0195] In some embodiments, the cytidine deaminase is a human APOBEC, including hAPOBEC1 or hAPOBEC3. In some embodiments, the cytidine deaminase is a human AID.

[0196] In some embodiments, the cytidine deaminase protein recognizes and converts one or more target cytosine residue(s) in a single-stranded bubble of an RNA duplex into uracil residues (s). In some embodiments, the cytidine deaminase protein recognizes a binding window on the single-stranded bubble of an RNA duplex. In some embodiments, the binding window contains at least one target cytosine residue(s). In some embodiments, the binding window is in the range of about 3 bp to about 100 bp. In some embodiments, the binding window is in the range of about 5 bp to about 50 bp. In some embodiments, the binding window is in the range of about 10 bp to about 30 bp. In some embodiments, the binding window is about 1 bp, 2 bp, 3 bp, 5 bp, 7 bp, 10 bp, 15 bp, 20 bp, 25 bp, 30 bp, 40 bp, 45 bp, 50 bp, 55 bp, 60 bp, 65 bp, 70 bp, 75 bp, 80 bp, 85 bp, 90 bp, 95 bp, or 100 bp.

[0197] In some embodiments, the cytidine deaminase protein comprises one or more deaminase domains. Not intended to be bound by theory, it is contemplated that the deaminase domain functions to recognize and convert one or more target cytosine (C) residue(s) contained in a single-stranded bubble of an RNA duplex into (an) uracil (U) residue (s). In some

embodiments, the deaminase domain comprises an active center. In some embodiments, the active center comprises a zinc ion. In some embodiments, amino acid residues in or near the active center interact with one or more nucleotide(s) 5' to a target cytosine residue. In some embodiments, amino acid residues in or near the active center interact with one or more nucleotide(s) 3' to a target cytosine residue.

[0198] In some embodiments, the cytidine deaminase comprises human APOBEC1 full protein (hAPOBEC1) or the deaminase domain thereof (hAPOBEC1-D) or a C-terminally truncated version thereof (hAPOBEC-T). In some embodiments, the cytidine deaminase is an APOBEC family member that is homologous to hAPOBEC1, hAPOBEC-D or hAPOBEC-T. In some embodiments, the cytidine deaminase comprises human AID1 full protein (hAID) or the deaminase domain thereof (hAID-D) or a C-terminally truncated version thereof (hAID-T). In some embodiments, the cytidine deaminase is an AID family member that is homologous to hAID, hAID-D or hAID-T. In some embodiments, the hAID-T is a hAID which is C-terminally truncated by about 20 amino acids.

[0199] In some embodiments, the cytidine deaminase comprises the wild-type amino acid sequence of a cytosine deaminase. In some embodiments, the cytidine deaminase comprises one or more mutations in the cytosine deaminase sequence, such that the editing efficiency, and/or substrate editing preference of the cytosine deaminase is changed according to specific needs.

[0200] Certain mutations of APOBEC1 and APOBEC3 proteins have been described in Kim et al., Nature Biotechnology (2017) 35(4):371-377 (doi:10.1038/nbt.3803); and Harris et al. Mol. Cell (2002) 10:1247-1253, each of which is incorporated herein by reference in its entirety.

[0201] In some embodiments, the cytidine deaminase is an APOBEC1 deaminase comprising one or more mutations at amino acid positions corresponding to W90, R118, H121, H122, R126, or R132 in rat APOBEC1, or an APOBEC3G deaminase comprising one or more mutations at amino acid positions corresponding to W285, R313, D316, D317X, R320, or R326 in human APOBEC3G.

[0202] In some embodiments, the cytidine deaminase comprises a mutation at tryptophane90 of the rat APOBEC1 amino acid sequence, or a corresponding position in a homologous APOBEC protein, such as tryptophane285 of APOBEC3G. In some

embodiments, the tryptophane residue at position 90 is replaced by an tyrosine or phenylalanine residue (W90Y or W90F).

[0203] In some embodiments, the cytidine deaminase comprises a mutation at Arginine118 of the rat APOBEC1 amino acid sequence, or a corresponding position in a homologous APOBEC protein. In some embodiments, the arginine residue at position 118 is replaced by an alanine residue (R118A).

[0204] In some embodiments, the cytidine deaminase comprises a mutation at Histidine121 of the rat APOBEC1 amino acid sequence, or a corresponding position in a homologous APOBEC protein. In some embodiments, the histidine residue at position 121 is replaced by an arginine residue (H121R).

[0205] In some embodiments, the cytidine deaminase comprises a mutation at Histidine122 of the rat APOBEC1 amino acid sequence, or a corresponding position in a homologous APOBEC protein. In some embodiments, the histidine residue at position 122 is replaced by an arginine residue (H122R).

[0206] In some embodiments, the cytidine deaminase comprises a mutation at Arginine126 of the rat APOBEC1 amino acid sequence, or a corresponding position in a homologous APOBEC protein, such as Arginine320 of APOBEC3G. In some embodiments, the arginine residue at position 126 is replaced by an alanine residue (R126A) or by a glutamic acid (R126E).

[0207] In some embodiments, the cytidine deaminase comprises a mutation at arginine132 of the APOBEC1 amino acid sequence, or a corresponding position in a homologous APOBEC protein. In some embodiments, the arginine residue at position 132 is replaced by a glutamic acid residue (R132E).

[0208] In some embodiments, to narrow the width of the editing window, the cytidine deaminase may comprise one or more of the mutations: W90Y, W90F, R126E and R132E, based on amino acid sequence positions of rat APOBEC1, and mutations in a homologous APOBEC protein corresponding to the above.

[0209] In some embodiments, to reduce editing efficiency, the cytidine deaminase may comprise one or more of the mutations: W90A, R118A, R132E, based on amino acid sequence positions of rat APOBEC1, and mutations in a homologous APOBEC protein corresponding to the above. In particular embodiments, it can be of interest to use a cytidine deaminase enzyme with reduced efficacy to reduce off-target effects.

[0210] In some embodiments, the cytidine deaminase is wild-type rat APOBEC1 (rAPOBEC1, or a catalytic domain thereof. In some embodiments, the cytidine deaminase comprises one or more mutations in the rAPOBEC1 sequence, such that the editing efficiency, and/or substrate editing preference of rAPOBEC1 is changed according to specific needs.

[0211] rAPOBEC1:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQNT NKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYIAR LYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEAHWPRYPHLW VRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK **(SEQ ID NO: 15)**

[0212] In some embodiments, the cytidine deaminase is wild-type human APOBEC1 (hAPOBEC1) or a catalytic domain thereof. In some embodiments, the cytidine deaminase comprises one or more mutations in the hAPOBEC1 sequence, such that the editing efficiency, and/or substrate editing preference of hAPOBEC1 is changed according to specific needs.

[0213] APOBEC1:

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGKN TTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVIYV ARLFWHMDQQNRQGLRDLVNSGVTIQIMRASEYYHCWRNFVNYPPGDEAHWPQY PPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLATGLI HPSVAWR **(SEQ ID NO: 16)**

[0214] In some embodiments, the cytidine deaminase is wild-type human APOBEC3G (hAPOBEC3G) or a catalytic domain thereof. In some embodiments, the cytidine deaminase comprises one or more mutations in the hAPOBEC3G sequence, such that the editing efficiency, and/or substrate editing preference of hAPOBEC3G is changed according to specific needs.

[0215] hAPOBEC3G:

MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPKV TLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKIMNYDEFQHCWSKFVYSQRE LFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVERM HNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRVTCFT SWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISIMTYS EFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN **(SEQ ID NO: 17)**

[0216] In some embodiments, the cytidine deaminase is wild-type Petromyzon marinus CDA1 (pmCDAl) or a catalytic domain thereof. In some embodiments, the cytidine deaminase comprises one or more mutations in the pmCDAl sequence, such that the editing efficiency, and/or substrate editing preference of pmCDAl is changed according to specific needs.

[0217] pmCDAl:

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVNK PQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQELRG NGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSSHN QLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV **(SEQ ID NO: 18)**

[0218] In some embodiments, the cytidine deaminase is wild-type human AID (hAID) or a catalytic domain thereof. In some embodiments, the cytidine deaminase comprises one or more mutations in the pmCDAl sequence, such that the editing efficiency, and/or substrate editing preference of pmCDAl is changed according to specific needs.

[0219] hAID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPYLSLRIFTAR LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHEN SVRLSRQLRRILLPLYEVDDLRDAFRTLGLLD **(SEQ ID NO: 19)**

[0220] In some embodiments, the cytidine deaminase is truncated version of hAID (hAID-DC) or a catalytic domain thereof. In some embodiments, the cytidine deaminase comprises one or more mutations in the hAID-DC sequence, such that the editing efficiency, and/or substrate editing preference of hAID-DC is changed according to specific needs.

[0221] hAID-DC:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNGC HVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFTAR LYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGLHEN SVRLSRQLRRILL **(SEQ ID NO: 20)**

[0222] Additional embodiments of the cytidine deaminase are disclosed in International Patent Publication No. WO WO2017/070632, titled "Nucleobase Editor and Uses Thereof," which is incorporated herein by reference in its entirety.

[0223] In some embodiments, the cytidine deaminase has an efficient deamination window that encloses the nucleotides susceptible to deamination editing. Accordingly, in some

embodiments, the "editing window width" refers to the number of nucleotide positions at a given target site for which editing efficiency of the cytidine deaminase exceeds the half-maximal value for that target site. In some embodiments, the cytidine deaminase has an editing window width in the range of about 1 to about 6 nucleotides. In some embodiments, the editing window width of the cytidine deaminase is 1, 2, 3, 4, 5, or 6 nucleotides.

[0224] Not intended to be bound by theory, it is contemplated that in some embodiments, the length of the linker sequence affects the editing window width. In some embodiments, the editing window width increases (e.g., from about 3 to about 6 nucleotides) as the linker length extends (e.g., from about 3 to about 21 amino acids). In a non-limiting example, a 16-residue linker offers an efficient deamination window of about 5 nucleotides. In some embodiments, the length of the guide RNA affects the editing window width. In some embodiments, shortening the guide RNA leads to a narrowed efficient deamination window of the cytidine deaminase.

[0225] In some embodiments, mutations to the cytidine deaminase affect the editing window width. In some embodiments, the cytidine deaminase component of the CD-functionalized CRISPR system comprises one or more mutations that reduce the catalytic efficiency of the cytidine deaminase, such that the deaminase is prevented from deamination of multiple cytidines per DNA binding event. In some embodiments, tryptophan at residue 90 (W90) of APOBEC1 or a corresponding tryptophan residue in a homologous sequence is mutated. In some embodiments, the catalytically inactive Type VII effector is fused to or linked to an APOBEC1 mutant that comprises a W90Y or W90F mutation. In some embodiments, tryptophan at residue 285 (W285) of APOBEC3G, or a corresponding tryptophan residue in a homologous sequence is mutated. In some embodiments, the catalytically inactive Type VII protein is fused to or linked to an APOBEC3G mutant that comprises a W285Y or W285F mutation.

[0226] In some embodiments, the cytidine deaminase component of CD-functionalized CRISPR system comprises one or more mutations that reduce tolerance for non-optimal presentation of a cytidine to the deaminase active site. In some embodiments, the cytidine deaminase comprises one or more mutations that alter substrate binding activity of the deaminase active site. In some embodiments, the cytidine deaminase comprises one or more mutations that alter the conformation of DNA to be recognized and bound by the deaminase active site. In some embodiments, the cytidine deaminase comprises one or more mutations

that alter the substrate accessibility to the deaminase active site. In some embodiments, arginine at residue 126 (R126) of APOBEC1 or a corresponding arginine residue in a homologous sequence is mutated. In some embodiments, the catalytically inactive Type VII CRISPR is fused to or linked to an APOBEC1 that comprises a R126A or R126E mutation. In some embodiments, tryptophan at residue 320 (R320) of APOBEC3G, or a corresponding arginine residue in a homologous sequence is mutated. In some embodiments, the catalytically inactive Type VII CRISPR is fused to or linked to an APOBEC3G mutant that comprises a R320A or R320E mutation. In some embodiments, arginine at residue 132 (R132) of APOBEC1 or a corresponding arginine residue in a homologous sequence is mutated. In some embodiments, the catalytically inactive Type VII CRISPR is fused to or linked to an APOBEC1 mutant that comprises a R132E mutation.

[0227] In some embodiments, the APOBEC1 domain of the CD-functionalized CRISPR system comprises one, two, or three mutations selected from W90Y, W90F, R126A, R126E, and R132E. In some embodiments, the APOBEC1 domain comprises double mutations of W90Y and R126E. In some embodiments, the APOBEC1 domain comprises double mutations of W90Y and R132E. In some embodiments, the APOBEC1 domain comprises double mutations of R126E and R132E. In some embodiments, the APOBEC1 domain comprises three mutations of W90Y, R126E and R132E.

[0228] In some embodiments, one or more mutations in the cytidine deaminase as disclosed herein reduce the editing window width to about 2 nucleotides. In some embodiments, one or more mutations in the cytidine deaminase as disclosed herein reduce the editing window width to about 1 nucleotide. In some embodiments, one or more mutations in the cytidine deaminase as disclosed herein reduce the editing window width while only minimally or modestly affecting the editing efficiency of the enzyme. In some embodiments, one or more mutations in the cytidine deaminase as disclosed herein reduce the editing window width without reducing the editing efficiency of the enzyme. In some embodiments, one or more mutations in the cytidine deaminase as disclosed herein enable discrimination of neighboring cytidine nucleotides, which would be otherwise edited with similar efficiency by the cytidine deaminase.

[0229] In some embodiments, the cytidine deaminase protein further comprises or is connected to one or more double-stranded RNA (dsRNA) binding motifs (dsRBMs) or domains (dsRBDs) for recognizing and binding to double-stranded nucleic acid substrates. In

some embodiments, the interaction between the cytidine deaminase and the substrate is mediated by one or more additional protein factor(s), including a CRISPR/CAS protein factor. In some embodiments, the interaction between the cytidine deaminase and the substrate is further mediated by one or more nucleic acid component(s), including a guide RNA.

[0230] According to the present invention, the substrate of the cytidine deaminase is an DNA single strand bubble of a RNA duplex comprising a Cytosine of interest, made accessible to the cytidine deaminase upon binding of the guide molecule to its DNA target which then forms the CRISPR-Cas complex with the CRISPR-Cas enzyme, whereby the cytosine deaminase is fused to or is capable of binding to one or more components of the CRISPR-Cas complex, i.e. the CRISPR-Cas enzyme and/or the guide molecule. The particular features of the guide molecule and CRISPR-Cas enzyme are detailed below.

[0231] The cytidine deaminase or catalytic domain thereof may be a human, a rat, or a lamprey cytidine deaminase protein or catalytic domain thereof.

[0232] The cytidine deaminase protein or catalytic domain thereof may be an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase. The cytidine deaminase protein or catalytic domain thereof may be an activation-induced deaminase (AID). The cytidine deaminase protein or catalytic domain thereof may be a cytidine deaminase 1 (CDA1).

[0233] The cytidine deaminase protein or catalytic domain thereof may be an APOBEC1 deaminase. The APOBEC1 deaminase may comprise one or more mutations corresponding to W90A, W90Y, R118A, H121R, H122R, R126A, R126E, or R132E in rat APOBEC1, or an APOBEC3G deaminase comprising one or more mutations corresponding to W285A, W285Y, R313A, D316R, D317R, R320A, R320E, or R326E in human APOBEC3G.

[0234] The system may further comprise an uracil glycosylase inhibitor (UGI). Inn some embodiments, the cytidine deaminase protein or catalytic domain thereof is delivered together with a uracil glycosylase inhibitor (UGI). The GI may be linked (e.g., covalently linked) to the cytidine deaminase protein or catalytic domain thereof and/or a catalytically inactive type VII effector protein.

Base Editing Guide Molecule Design Considerations

[0235] In some embodiments, the guide sequence is an RNA sequence of between 10 to 50 nt in length, but more particularly of about 20-30 nt advantageously about 20 nt, 23-25 nt or 24 nt. In base editing embodiments, the guide sequence is selected so as to ensure that it

hybridizes to the target sequence comprising the adenosine to be deaminated. This is described more in detail below. Selection can encompass further steps which increase efficacy and specificity of deamination.

[0236] In some embodiments, the guide sequence is about 20 nt to about 30 nt long and hybridizes to the target DNA strand to form an almost perfectly matched duplex, except for having a dA-C mismatch at the target adenosine site. Particularly, in some embodiments, the dA-C mismatch is located close to the center of the target sequence (and thus the center of the duplex upon hybridization of the guide sequence to the target sequence), thereby restricting the adenosine deaminase to a narrow editing window (e.g., about 4 bp wide). In some embodiments, the target sequence may comprise more than one target adenosine to be deaminated. In further embodiments the target sequence may further comprise one or more dA-C mismatch 3' to the target adenosine site. In some embodiments, to avoid off-target editing at an unintended Adenine site in the target sequence, the guide sequence can be designed to comprise a non-pairing Guanine at a position corresponding to said unintended Adenine to introduce a dA-G mismatch, which is catalytically unfavorable for certain adenosine deaminases such as ADAR1 and ADAR2. See Wong et al., RNA 7:846-858 (2001), which is incorporated herein by reference in its entirety.

[0237] In some embodiments, a Type VII CRISPR guide sequence having a canonical length (e.g., about 20 nt) is used to form a heteroduplex with the target DNA. In some embodiments, a Type VII CRISPR guide molecule longer than the canonical length (e.g., >20 nt) is used to form a heteroduplex with the target DNA including outside of the Type VII CRISPR-guide RNA-target DNA complex. This can be of interest where deamination of more than one adenine within a given stretch of nucleotides is of interest. In alternative embodiments, it is of interest to maintain the limitation of the canonical guide sequence length. In some embodiments, the guide sequence is designed to introduce a dA-C mismatch outside of the canonical length of Type VII CRISPR guide, which may decrease steric hindrance by the Type VII CRISPR and increase the frequency of contact between the adenosine deaminase and the dA-C mismatch.

[0238] In some base editing embodiments, the position of the mismatched nucleobase (e.g., cytidine) is calculated from where the PAM would be on a DNA target. In some embodiments, the mismatched nucleobase is positioned 12-21 nt from the PAM, or 13-21 nt from the PAM, or 14-21 nt from the PAM, or 14-20 nt from the PAM, or 15-20 nt from the PAM, or 16-20 nt

from the PAM, or 14-19 nt from the PAM, or 15-19 nt from the PAM, or 16-19 nt from the PAM, or 17-19 nt from the PAM, or about 20 nt from the PAM, or about 19 nt from the PAM, or about 18 nt from the PAM, or about 17 nt from the PAM, or about 16 nt from the PAM, or about 15 nt from the PAM, or about 14 nt from the PAM. In a preferred embodiment, the mismatched nucleobase is positioned 17-19 nt or 18 nt from the PAM.

[0239] Mismatch distance is the number of bases between the 3' end of the Type VII CRISPR spacer and the mismatched nucleobase (e.g., cytidine), wherein the mismatched base is included as part of the mismatch distance calculation. In some embodiment, the mismatch distance is 1-10 nt, or 1-9 nt, or 1-8 nt, or 2-8 nt, or 2-7 nt, or 2-6 nt, or 3-8 nt, or 3-7 nt, or 3-6 nt, or 3-5 nt, or about 2 nt, or about 3 nt, or about 4 nt, or about 5 nt, or about 6 nt, or about 7 nt, or about 8 nt. In a preferred embodiment, the mismatch distance is 3-5 nt or 4 nt.

[0240] In some embodiments, the editing window of a Type VII CRISPR-ADAR system described herein is 12-21 nt from the PAM, or 13-21 nt from the PAM, or 14-21 nt from the PAM, or 14-20 nt from the PAM, or 15-20 nt from the PAM, or 16-20 nt from the PAM, or 14-19 nt from the PAM, or 15-19 nt from the PAM, or 16-19 nt from the PAM, or 17-19 nt from the PAM, or about 20 nt from the PAM, or about 19 nt from the PAM, or about 18 nt from the PAM, or about 17 nt from the PAM, or about 16 nt from the PAM, or about 15 nt from the PAM, or about 14 nt from the PAM. In some embodiment, the editing window of the Type VII -ADAR system described herein is 1-10 nt from the 3' end of the Type VII spacer, or 1-9 nt from the 3' end of the Type VII spacer, or 1-8 nt from the 3' end of the Type VII spacer, or 2-8 nt from the 3' end of the Type VII spacer, or 2-7 nt from the 3' end of the Type VII spacer, or 2-6 nt from the 3' end of the Type VII spacer, or 3-8 nt from the 3' end of the Type VII spacer, or 3-7 nt from the 3' end of the Type VII spacer, or 3-6 nt from the 3' end of the Type VII spacer, or 3-5 nt from the 3' end of the Type VII spacer, or about 2 nt from the 3' end of the Type VII spacer, or about 3 nt from the 3' end of the Type VII spacer, or about 4 nt from the 3' end of the Type VII spacer, or about 5 nt from the 3' end of the Type VII spacer, or about 6 nt from the 3' end of the Type VII spacer, or about 7 nt from the 3' end of the Type VII spacer, or about 8 nt from the 3' end of the Type VII spacer.

Prime Editors

[0241] In some embodiments, the Type VII effector or system thereof can be modified such that it can operate as a prime editing system, similar to that described in e.g. Anzalone et al. 2019. Nature. 576: 149-157. Like base editing systems, prime editing systems can be

capable of targeted modification of a polynucleotide without generating double stranded breaks and does not require donor templates. Further prime editing systems can be capable of all 12 possible combination swaps. Prime editing can operate via a "search-and-replace" methodology and can mediate targeted insertions, deletions, all 12 possible base-to-base conversion, and combinations thereof. Generally, a prime editing system, as exemplified by PE1, PE2, and PE3 (*Id.*), can include a reverse transcriptase fused or otherwise coupled or associated with an RNA-programmable nickase, and a prime-editing extended guide RNA (pegRNA) to facility direct copying of genetic information from the extension on the pegRNA into the target polynucleotide. Embodiments that can be used with the present invention include these and variants thereof. Prime editing can have the advantage of lower off-target activity than traditional CRIPSR-Cas systems along with few byproducts and greater or similar efficiency as compared to traditional CRISPR-Cas systems.

[0242] In some embodiments, the prime editing guide molecule can specify both the target polynucleotide information (e.g. sequence) and contain new polynucleotide information that replaces target polynucleotides. Information transfer from the guide molecule to the target polynucleotide, the PE system can nick the target polynucleotide at a target side to expose a 3'hydroxyl group, which can prime reverse transcription of an edit-encoding extension region of the guide molecule (e.g. a prime editing guide molecule or peg guide molecule) directly into the target site in the target polynucleotide. See e.g. Anzalone et al. 2019. Nature. 576: 149-157, particularly at Figures 1b, 1c, related discussion, and Supplementary discussion.

[0243] In some embodiments, a prime editing system can be composed of a Cas polypeptide having nickase activity, a reverse transcriptase, and a guide molecule. The Cas polypeptide can lack nuclease activity. The guide molecule can include a target binding sequence as well as a primer binding sequence and a template containing the edited polynucleotide sequence. The guide molecule, Cas polypeptide, and/or reverse transcriptase can be coupled together or otherwise associate with each other to form an effector complex and edit a target sequence. In some embodiments, the Cas polypeptide is a Class 2, Type V Cas polypeptide. In some embodiments, the Cas polypeptide is a Cas9 polypeptide (e.g. is a Cas9 nickase). In some embodiments, the Cas polypeptide is fused to the reverse transcriptase. In some embodiments, the Cas polypeptide is linked to the reverse transcriptase.

[0244] In some embodiments, the prime editing system can be a PE1 system or variant thereof, a PE2 system or variant thereof, or a PE3 (e.g. PE3, PE3b) system. See e.g., Anzalone

et al. 2019. Nature. 576: 149-157, particularly at pgs. 2-3, Figs. 2a, 3a-3f, 4a-4b, Extended data Figs. 3a-3b, 4,

[0245] The peg guide molecule can be about 10 to about 200 or more nucleotides in length, such as 10 to/or 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, or 200 or more nucleotides in length. Optimization of the peg guide molecule can be accomplished as described in Anzalone et al. 2019. Nature. 576: 149-157, particularly at pg. 3, Fig. 2a-2b, and Extended Data Figs. 5a-c.

CRISPR Associated Transposase (CAST) Systems

[0246] In some embodiments, the Type VII effector or system thereof can be modified such that it operates similarly to a CAST system and can include a Cas protein (e.g. a Type VII or additional effector protein) that is catalytically inactive, or engineered to be catalytically active, and further comprises a transposase (or subunits thereof) that catalyze RNA-guided DNA transposition. Such systems are able to insert DNA sequences at a target site in a DNA molecule without relying on host cell repair machinery. Exemplary CAST systems can be Class1 or Class 2 CAST systems and can be used as a reference for modification of the Type VII Systems described herein. An example Class 1 system is described in Klompe et al. Nature, doi:10.1038/s41586-019-1323, which is in incorporated herein by reference. Example Class 2 systems are described in Strecker et al. Science. 10/1126/science. aax9181 (2019), and PCT/US2019/066835 which are incorporated herein by reference.

Cas truncations

[0247] In certain embodiments, the effector protein (Type VII CRISPR enzyme; effector protein) according to the invention as described herein is a catalytically inactive or dead Type VII effector protein (dVII). In some embodiments, the dVII effector comprises mutations in the nuclease domain. In some embodiments, the dType VII effector protein has been truncated.

[0248] To reduce the size of a fusion protein of the Type VII effector and the one or more functional domains, the C-terminus of the Type VII effector can be truncated while still maintaining its nucleic acid binding function. For example, at least 20 amino acids, at least 40 amino acids, at least 50 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 150 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 250 amino acids, at least 260 amino acids, or at least 300 amino acids, or at least 350 amino acids, or up to 120 amino acids, or up to 140 amino acids, or up to 160 amino acids, or up to 180 amino acids, or up to 200 amino acids, or up to 250 amino acids, or up to 300 amino acids, or up to 350 amino acids, or up to 400 amino acids, may be truncated at the C-terminus of the Type VII effector. The skilled person will understand that similar truncations can be designed for various Type VII orthologues or subtypes.

[0249] In some embodiments, the N-terminus of the Type VII effector protein may be truncated. For example, at least 20 amino acids, at least 40 amino acids, at least 50 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 150 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 250 amino acids, at least 260 amino acids, or at least 300 amino acids, or at least 350 amino acids, or up to 120 amino acids, or up to 140 amino acids, or up to 160 amino acids, or up to 180 amino acids, or up to 200 amino acids, or up to 250 amino acids, or up to 300 amino acids, or up to 350 amino acids, or up to 400 amino acids, may be truncated at the N-terminus of the Type VII effector.

[0250] In some embodiments, both the N- and the C- termini of the Type VII effector protein may be truncated. For example, at least 20 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 40 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino

acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 60 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 80 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 100 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 120 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 140 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector.

For example, at least 160 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 180 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 200 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 220 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 240 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 260 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at

least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 280 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 300 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 350 amino acids may be truncated at the C-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the N-terminus of the Type VII effector. For example, at least 20 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 40 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at

least 60 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 80 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 100 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 120 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 140 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 160 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at

least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 180 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 200 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 220 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 240 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 260 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 280 amino acids

may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 300 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector. For example, at least 350 amino acids may be truncated at the N-terminus of the Type VII effector, and at least 20 amino acids, at least 40 amino acids, at least 60 amino acids, at least 80 amino acids, at least 100 amino acids, at least 120 amino acids, at least 140 amino acids, at least 160 amino acids, at least 180 amino acids, at least 200 amino acids, at least 220 amino acids, at least 240 amino acids, at least 260 amino acids, at least 300 amino acids, or at least 350 amino acids may be truncated at the C-terminus of the Type VII effector.

Guide Molecules

[0251] The methods described herein may be used to screen inhibition of CRISPR systems employing different types of guide molecules. As used herein, the term "guide sequence" and "guide molecule" in the context of a CRISPR-Cas system, comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. The guide sequences made using the methods disclosed herein may be a full-length guide sequence, a truncated guide sequence, a full-length sgRNA sequence, a truncated sgRNA sequence, or an E+F sgRNA sequence. In some embodiments, the degree of complementarity of the guide sequence to a given target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. In certain example embodiments, the guide molecule comprises a guide sequence that may be designed to have at least one mismatch with the target sequence, such that a RNA duplex formed between the guide sequence and the target sequence. Accordingly, the degree of complementarity is preferably

less than 99%. For instance, where the guide sequence consists of 24 nucleotides, the degree of complementarity is more particularly about 96% or less. In particular embodiments, the guide sequence is designed to have a stretch of two or more adjacent mismatching nucleotides, such that the degree of complementarity over the entire guide sequence is further reduced. For instance, where the guide sequence consists of 24 nucleotides, the degree of complementarity is more particularly about 96% or less, more particularly, about 92% or less, more particularly about 88% or less, more particularly about 84% or less, more particularly about 80% or less, more particularly about 76% or less, more particularly about 72% or less, depending on whether the stretch of two or more mismatching nucleotides encompasses 2, 3, 4, 5, 6 or 7 nucleotides, etc. In some embodiments, aside from the stretch of one or more mismatching nucleotides, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid -targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence (or a sequence in the vicinity thereof) may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at or in the vicinity of the target sequence between the test and control guide sequence reactions. Other assays are possible, and

will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting guide RNA may be selected to target any target nucleic acid sequence.

[0252] In certain embodiments, the guide sequence or spacer length of the guide molecules is from 15 to 50 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 27-30 nt, e.g., 27, 28, 29, or 30 nt, from 30-35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer. In certain example embodiment, the guide sequence is 15, 16, 17,18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39 40, 41, 42, 43, 44, 45, 46, 47 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 nt.

[0253] In some embodiments, the guide sequence is an RNA sequence of between 10 to 50 nt in length, but more particularly of about 20-30 nt advantageously about 20 nt, 23-25 nt or 24 nt. The guide sequence is selected so as to ensure that it hybridizes to the target sequence. This is described more in detail below. Selection can encompass further steps which increase efficacy and specificity.

[0254] In some embodiments, the guide sequence has a canonical length (e.g., about 15-30 nt) is used to hybridize with the target RNA or DNA. In some embodiments, a guide molecule is longer than the canonical length (e.g., >30 nt) is used to hybridize with the target RNA or DNA, such that a region of the guide sequence hybridizes with a region of the RNA or DNA strand outside of the Cas-guide target complex. This can be of interest where additional modifications, such deamination of nucleotides is of interest. In alternative embodiments, it is of interest to maintain the limitation of the canonical guide sequence length.

[0255] In some embodiments, the sequence of the guide molecule (direct repeat and/or spacer) is selected to reduce the degree secondary structure within the guide molecule. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the nucleic acid-targeting guide RNA participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker

and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g., A.R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[0256] In some embodiments, a nucleic acid-targeting guide is designed or selected to modulate intermolecular interactions among guide molecules, such as among stem-loop regions of different guide molecules. It will be appreciated that nucleotides within a guide that base-pair to form a stem-loop are also capable of base-pairing to form an intermolecular duplex with a second guide and that such an intermolecular duplex would not have a secondary structure compatible with CRISPR complex formation. Accordingly, it is useful to select or design DR sequences in order to modulate stem-loop formation and CRISPR complex formation. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of nucleic acid-targeting guides are in intermolecular duplexes. It will be appreciated that stem-loop variation will often be within limits imposed by DR-CRISPR effector interactions. One way to modulate stem-loop formation or change the equilibrium between stem-loop and intermolecular duplex is to vary nucleotide pairs in the stem of the stem-loop of a DR. For example, in one embodiment, a G-C pair is replaced by an A-U or U-A pair. In another embodiment, an A-U pair is substituted for a G-C or a C-G pair. In another embodiment, a naturally occurring nucleotide is replaced by a nucleotide analog. Another way to modulate stem-loop formation or change the equilibrium between stem-loop and intermolecular duplex is to modify the loop of the stem-loop of a DR. Without be bound by theory, the loop can be viewed as an intervening sequence flanked by two sequences that are complementary to each other. When that intervening sequence is not self-complementary, its effect will be to destabilize intermolecular duplex formation. The same principle applies when guides are multiplexed: while the targeting sequences may differ, it may be advantageous to modify the stem-loop region in the DRs of the different guides. Moreover, when guides are multiplexed, the relative activities of the different guides can be modulated by balancing the activity of each individual guide. In certain embodiments, the equilibrium between intermolecular stem-loops vs. intermolecular duplexes is determined. The determination may be made by physical or biochemical means and can be in the presence or absence of a CRISPR effector.

[0257] In some embodiments, it is of interest to reduce the susceptibility of the guide molecule to RNA cleavage, such as cleavage by a CRISPR system that cleaves RNA. Accordingly, in particular embodiments, the guide molecule is adjusted to avoid cleavage by a CRISPR system or other RNA-cleaving enzymes.

[0258] In certain embodiments, the guide molecule comprises non-naturally occurring nucleic acids and/or non-naturally occurring nucleotides and/or nucleotide analogs, and/or chemically modifications. Preferably, these non-naturally occurring nucleic acids and non-naturally occurring nucleotides are located outside the guide sequence. Non-naturally occurring nucleic acids can include, for example, mixtures of naturally and non-naturally occurring nucleotides. Non-naturally occurring nucleotides and/or nucleotide analogs may be modified at the ribose, phosphate, and/or base moiety. In an embodiment of the invention, a guide nucleic acid comprises ribonucleotides and non-ribonucleotides. In one such embodiment, a guide comprises one or more ribonucleotides and one or more deoxyribonucleotides. In an embodiment of the invention, the guide comprises one or more non-naturally occurring nucleotide or nucleotide analog such as a nucleotide with phosphorothioate linkage, a locked nucleic acid (LNA) nucleotides comprising a methylene bridge between the 2′ and 4′ carbons of the ribose ring, or bridged nucleic acids (BNA). Other examples of modified nucleotides include 2'-O-methyl analogs, 2'-deoxy analogs, or 2'-fluoro analogs. Further examples of modified bases include, but are not limited to, 2-aminopurine, 5-bromo-uridine, pseudouridine, inosine, 7-methylguanosine. Examples of guide RNA chemical modifications include, without limitation, incorporation of 2′-O-methyl (M), 2′-O-methyl 3′phosphorothioate (MS), *S*-constrained ethyl(cEt), or 2′-O-methyl 3′thioPACE (MSP) at one or more terminal nucleotides. Such chemically modified guides can comprise increased stability and increased activity as compared to unmodified guides, though on-target vs. off-target specificity is not predictable. (See, Hendel, 2015, Nat Biotechnol. 33(9):985-9, doi: 10.1038/nbt.3290, published online 29 June 2015 Ragdarm et al., 0215, *PNAS*, E7110-E7111; Allerson et al., *J. Med. Chem*. 2005, 48:901-904; Bramsen et al., *Front. Genet*., 2012, 3:154; Deng et al., *PNAS*, 2015, 112:11870-11875; Sharma et al., *MedChemComm*., 2014, 5:1454-1471; Hendel et al., *Nat. Biotechnol*. (2015) 33(9): 985-989; Li et al., *Nature Biomedical Engineering*, 2017, 1, 0066 DOI:10.1038/s41551-017-0066). In some embodiments, the 5' and/or 3' end of a guide RNA is modified by a variety of functional moieties including fluorescent dyes, polyethylene glycol, cholesterol, proteins, or detection tags. (See Kelly et al.,

2016, *J. Biotech.* 233:74-83). In certain embodiments, a guide comprises ribonucleotides in a region that binds to a target RNA and one or more deoxyribonucletides and/or nucleotide analogs in a region that binds to a Type VII effector. In an embodiment of the invention, deoxyribonucleotides and/or nucleotide analogs are incorporated in engineered guide structures, such as, without limitation, stem-loop regions, and the seed region. In certain embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides of a guide is chemically modified. In some embodiments, 3-5 nucleotides at either the 3' or the 5' end of a guide is chemically modified. In some embodiments, only minor modifications are introduced in the seed region, such as 2'-F modifications. In some embodiments, 2'-F modification is introduced at the 3' end of a guide. In certain embodiments, three to five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-O-methyl (M), 2'-O-methyl 3' phosphorothioate (MS), *S*-constrained ethyl(cEt), or 2'-O-methyl 3' thioPACE (MSP). Such modification can enhance genome editing efficiency (see Hendel et al., *Nat. Biotechnol.* (2015) 33(9): 985-989). In certain embodiments, all of the phosphodiester bonds of a guide are substituted with phosphorothioates (PS) for enhancing levels of gene disruption. In certain embodiments, more than five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-O-Me, 2'-F or *S*-constrained ethyl(cEt). Such chemically modified guide can mediate enhanced levels of gene disruption (see Ragdarm et al., 0215, *PNAS*, E7110-E7111). In an embodiment of the invention, a guide is modified to comprise a chemical moiety at its 3' and/or 5' end. Such moieties include, but are not limited to amine, azide, alkyne, thio, dibenzocyclooctyne (DBCO), or Rhodamine, peptides, nuclear localization sequence (NLS), peptide nucleic acid (PNA), polyethylene glycol (PEG), triethylene glycol, or tetraethyleneglycol (TEG). In certain embodiment, the chemical moiety is conjugated to the guide by a linker, such as an alkyl chain. In certain embodiment, the chemical moiety is conjugated to the guide by a linker, such as an alkyl chain. In certain embodiments, the chemical moiety of the modified guide can be used to attach the guide to another molecule, such as DNA, RNA, protein, or nanoparticles. Such chemically modified guide can be used to identify or enrich cells generically edited by a CRISPR system (see Lee et al., *eLife*, 2017, 6:e25312, DOI:10.7554).

[0259] In some embodiments, 3 nucleotides at each of the 3' and 5' ends are chemically modified. In a specific embodiment, the modifications comprise 2'-O-methyl or phosphorothioate analogs. In a specific embodiment, 12 nucleotides in the tetraloop and 16

nucleotides in the stem-loop region are replaced with 2'-O-methyl analogs. Such chemical modifications improve *in vivo* editing and stability (see Finn et al., *Cell Reports* (2018), 22: 2227-2235). In some embodiments, more than 60 or 70 nucleotides of the guide are chemically modified. In some embodiments, this modification comprises replacement of nucleotides with 2'-O-methyl or 2'-fluoro nucleotide analogs or phosphorothioate (PS) modification of phosphodiester bonds. In some embodiments, the chemical modification comprises 2'-O-methyl or 2'-fluoro modification of guide nucleotides extending outside of the nuclease protein when the CRISPR complex is formed or PS modification of 20 to 30 or more nucleotides of the 3'-terminus of the guide. In a particular embodiment, the chemical modification further comprises 2'-O-methyl analogs at the 5' end of the guide or 2'-fluoro analogs in the seed and tail regions. Such chemical modifications improve stability to nuclease degradation and maintain or enhance genome-editing activity or efficiency, but modification of all nucleotides may abolish the function of the guide (see Yin et al., *Nat. Biotech.* (2018), 35(12): 1179-1187). Such chemical modifications may be guided by knowledge of the structure of the CRISPR complex, including knowledge of the limited number of nuclease and RNA 2'-OH interactions (see Yin et al., *Nat. Biotech.* (2018), 35(12): 1179-1187). In some embodiments, one or more guide RNA nucleotides may be replaced with DNA nucleotides. In some embodiments, up to 2, 4, 6, 8, 10, or 12 RNA nucleotides of the 5'-end tail/seed guide region are replaced with DNA nucleotides. In certain embodiments, the majority of guide RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, 16 guide RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, 8 guide RNA nucleotides of the 5'-end tail/seed region and 16 RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, guide RNA nucleotides that extend outside of the nuclease protein when the CRISPR complex is formed are replaced with DNA nucleotides. Such replacement of multiple RNA nucleotides with DNA nucleotides leads to decreased off-target activity but similar on-target activity compared to an unmodified guide; however, replacement of all RNA nucleotides at the 3' end may abolish the function of the guide (see Yin et al., *Nat. Chem. Biol.* (2018) 14, 311-316). Such modifications may be guided by knowledge of the structure of the CRISPR complex, including knowledge of the limited number of nuclease and RNA 2'-OH interactions (see Yin et al., *Nat. Chem. Biol.* (2018) 14, 311-316).

[0260] In some embodiments, the guide molecule forms a stem loop with a separate non-covalently linked sequence, which can be DNA or RNA. In particular embodiments, the sequences forming the guide are first synthesized using the standard phosphoramidite synthetic protocol (Herdewijn, P., ed., Methods in Molecular Biology Col 288, Oligonucleotide Synthesis: Methods and Applications, Humana Press, New Jersey (2012)). In some embodiments, these sequences can be functionalized to contain an appropriate functional group for ligation using the standard protocol known in the art (Hermanson, G. T., Bioconjugate Techniques, Academic Press (2013)). Examples of functional groups include, but are not limited to, hydroxyl, amine, carboxylic acid, carboxylic acid halide, carboxylic acid active ester, aldehyde, carbonyl, chlorocarbonyl, imidazolylcarbonyl, hydrozide, semicarbazide, thio semicarbazide, thiol, maleimide, haloalkyl, sufonyl, ally, propargyl, diene, alkyne, and azide. Once this sequence is functionalized, a covalent chemical bond or linkage can be formed between this sequence and the direct repeat sequence. Examples of chemical bonds include, but are not limited to, those based on carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, fulfones, sulfoxides, ureas, thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[0261] In some embodiments, these stem-loop forming sequences can be chemically synthesized. In some embodiments, the chemical synthesis uses automated, solid-phase oligonucleotide synthesis machines with 2'-acetoxyethyl orthoester (2'-ACE) (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989).

[0262] In certain embodiments, the guide molecule comprises (1) a guide sequence capable of hybridizing to a target locus and (2) a tracr mate or direct repeat sequence whereby the direct repeat sequence is located upstream (i.e., 5') or downstream (i.e. 3') from the guide sequence. In a particular embodiment the seed sequence (i.e. the sequence essential critical for recognition and/or hybridization to the sequence at the target locus) of the guide sequence is approximately within the first 10 nucleotides of the guide sequence.

[0263] In a particular embodiment the guide molecule comprises a guide sequence linked to a direct repeat sequence, wherein the direct repeat sequence comprises one or more stem

loops or optimized secondary structures. In particular embodiments, the direct repeat has a minimum length of 16 nts and a single stem loop. In further embodiments the direct repeat has a length longer than 16 nts, preferably more than 17 nts, and has more than one stem loops or optimized secondary structures. In particular embodiments the guide molecule comprises or consists of the guide sequence linked to all or part of the natural direct repeat sequence. A typical Type V or Type VI CRISPR-Cas guide molecule comprises (in 3' to 5' direction or in 5' to 3' direction): a guide sequence a first complimentary stretch (the "repeat"), a loop (which is typically 4 or 5 nucleotides long), a second complimentary stretch (the "anti-repeat" being complimentary to the repeat), and a poly A (often poly U in RNA) tail (terminator). In certain embodiments, the direct repeat sequence retains its natural architecture and forms a single stem loop. In particular embodiments, certain aspects of the guide architecture can be modified, for example by addition, subtraction, or substitution of features, whereas certain other aspects of guide architecture are maintained. Preferred locations for engineered guide molecule modifications, including but not limited to insertions, deletions, and substitutions include guide termini and regions of the guide molecule that are exposed when complexed with the CRISPR-Cas protein and/or target, for example the stem loop of the direct repeat sequence.

[0264] In particular embodiments, the stem comprises at least about 4 bp comprising complementary X and Y sequences, although stems of more, e.g., 5, 6, 7, 8, 9, 10, 11 or 12 or fewer, e.g., 3, 2, base pairs are also contemplated. Thus, for example X2-10 and Y2-10 (wherein X and Y represent any complementary set of nucleotides) may be contemplated. In one aspect, the stem made of the X and Y nucleotides, together with the loop will form a complete hairpin in the overall secondary structure; and, this may be advantageous and the amount of base pairs can be any amount that forms a complete hairpin. In one aspect, any complementary X:Y basepairing sequence (e.g., as to length) is tolerated, so long as the secondary structure of the entire guide molecule is preserved. In one aspect, the loop that connects the stem made of X:Y basepairs can be any sequence of the same length (e.g., 4 or 5 nucleotides) or longer that does not interrupt the overall secondary structure of the guide molecule. In one aspect, the stem loop can further comprise, e.g. an MS2 aptamer. In one aspect, the stem comprises about 5-7bp comprising complementary X and Y sequences, although stems of more or fewer basepairs are also contemplated. In one aspect, non-Watson Crick basepairing is contemplated, where such pairing otherwise generally preserves the architecture of the stem loop at that position.

[0265] In particular embodiments the natural hairpin or stem loop structure of the guide molecule is extended or replaced by an extended stem loop. It has been demonstrated that extension of the stem can enhance the assembly of the guide molecule with the CRISPR-Cas protein (Chen et al. Cell. (2013); 155(7): 1479–1491). In particular embodiments the stem of the stem loop is extended by at least 1, 2, 3, 4, 5 or more complementary basepairs (i.e. corresponding to the addition of 2,4, 6, 8, 10 or more nucleotides in the guide molecule). In particular embodiments these are located at the end of the stem, adjacent to the loop of the stem loop.

[0266] In particular embodiments, the susceptibility of the guide molecule to RNAses or to decreased expression can be reduced by slight modifications of the sequence of the guide molecule which do not affect its function. For instance, in particular embodiments, premature termination of transcription, such as premature transcription of U6 Pol-III, can be removed by modifying a putative Pol-III terminator (4 consecutive U's) in the guide molecules sequence. Where such sequence modification is required in the stem loop of the guide molecule, it is preferably ensured by a basepair flip.

[0267] In a particular embodiment, the direct repeat may be modified to comprise one or more protein-binding RNA aptamers. In a particular embodiment, one or more aptamers may be included such as part of optimized secondary structure. Such aptamers may be capable of binding a bacteriophage coat protein as detailed further herein.

[0268] In some embodiments, the guide molecule forms a duplex with a target RNA comprising at least one target cytosine residue to be edited. Upon hybridization of the guide RNA molecule to the target RNA, the cytidine deaminase binds to the single strand RNA in the duplex made accessible by the mismatch in the guide sequence and catalyzes deamination of one or more target cytosine residues comprised within the stretch of mismatching nucleotides.

[0269] A guide sequence, and hence a nucleic acid-targeting guide RNA may be selected to target any target nucleic acid sequence. The target sequence may be mRNA.

[0270] In certain embodiments, the target sequence should be associated with a PAM (protospacer adjacent motif) or PFS (protospacer flanking sequence or site); that is, a short sequence recognized by the CRISPR complex. Depending on the nature of the CRISPR-Cas protein, the target sequence should be selected such that its complementary sequence in the DNA duplex (also referred to herein as the non-target sequence) is upstream or downstream of

the PAM. In the embodiments of the present invention where the CRISPR-Cas protein is a Type VII protein, the complementary sequence of the target sequence is downstream or 3' of the PAM or upstream or 5' of the PAM. The precise sequence and length requirements for the PAM differ depending on the Type VII protein used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence). Skilled person will be able to identify PAM sequences for use with a given Type VII protein.

[0271] Further, engineering of the PAM Interacting (PI) domain may allow programing of PAM specificity, improve target site recognition fidelity, and increase the versatility of the CRISPR-Cas protein, for example as described for Cas9 in Kleinstiver BP et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592. As further detailed herein, the skilled person will understand that Type VII proteins may be modified analogously.

[0272] In particular embodiment, the guide is an escorted guide. By "escorted" is meant that the CRISPR-Cas system or complex or guide is delivered to a selected time or place within a cell, so that activity of the CRISPR-Cas system or complex or guide is spatially or temporally controlled. For example, the activity and destination of the 3 CRISPR-Cas system or complex or guide may be controlled by an escort RNA aptamer sequence that has binding affinity for an aptamer ligand, such as a cell surface protein or other localized cellular component. Alternatively, the escort aptamer may for example be responsive to an aptamer effector on or in the cell, such as a transient effector, such as an external energy source that is applied to the cell at a particular time.

[0273] The escorted CRISPR-Cas systems or complexes have a guide molecule with a functional structure designed to improve guide molecule structure, architecture, stability, genetic expression, or any combination thereof. Such a structure can include an aptamer.

[0274] Aptamers are biomolecules that can be designed or selected to bind tightly to other ligands, for example using a technique called systematic evolution of ligands by exponential enrichment (SELEX; Tuerk C, Gold L: "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." Science 1990, 249:505-510). Nucleic acid aptamers can for example be selected from pools of random-sequence oligonucleotides, with high binding affinities and specificities for a wide range of biomedically relevant targets, suggesting a wide range of therapeutic utilities for aptamers (Keefe, Anthony D., Supriya Pai, and Andrew Ellington. "Aptamers as therapeutics." Nature Reviews Drug

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

[0275] Accordingly, in particular embodiments, the guide molecule is modified, e.g., by one or more aptamer(s) designed to improve guide molecule delivery, including delivery across the cellular membrane, to intracellular compartments, or into the nucleus. Such a structure can include, either in addition to the one or more aptamer(s) or without such one or more aptamer(s), moiety(ies) so as to render the guide molecule deliverable, inducible or responsive to a selected effector. The invention accordingly comprehends a guide molecule that responds to normal or pathological physiological conditions, including without limitation pH, hypoxia, O_2 concentration, temperature, protein concentration, enzymatic concentration, lipid structure, light exposure, mechanical disruption (e.g. ultrasound waves), magnetic fields, electric fields, or electromagnetic radiation.

[0276] Light responsiveness of an inducible system may be achieved via the activation and binding of cryptochrome-2 and CIB1. Blue light stimulation induces an activating conformational change in cryptochrome-2, resulting in recruitment of its binding partner CIB1. This binding is fast and reversible, achieving saturation in <15 sec following pulsed stimulation and returning to baseline <15 min after the end of stimulation. These rapid binding kinetics result in a system temporally bound only by the speed of transcription/translation and transcript/protein degradation, rather than uptake and clearance of inducing agents. Crytochrome-2 activation is also highly sensitive, allowing for the use of low light intensity stimulation and mitigating the risks of phototoxicity. Further, in a context such as the intact

mammalian brain, variable light intensity may be used to control the size of a stimulated region, allowing for greater precision than vector delivery alone may offer.

[0277] The invention contemplates energy sources such as electromagnetic radiation, sound energy or thermal energy to induce the guide. Advantageously, the electromagnetic radiation is a component of visible light. In a preferred embodiment, the light is a blue light with a wavelength of about 450 to about 495 nm. In an especially preferred embodiment, the wavelength is about 488 nm. In another preferred embodiment, the light stimulation is via pulses. The light power may range from about 0-9 mW/cm^2. In a preferred embodiment, a stimulation paradigm of as low as 0.25 sec every 15 sec should result in maximal activation.

[0278] The chemical or energy sensitive guide may undergo a conformational change upon induction by the binding of a chemical source or by the energy allowing it act as a guide and have the Type VII CRISPR-Cas system or complex function. The invention can involve applying the chemical source or energy so as to have the guide function and the Type VII CRISPR-Cas system or complex function; and optionally further determining that the expression of the genomic locus is altered.

[0279] There are several different designs of this chemical inducible system: 1. ABI-PYL based system inducible by Abscisic Acid (ABA) (see, e.g., stke.sciencemag.org/cgi/content/abstract/sigtrans;4/164/rs2), 2. FKBP-FRB based system inducible by rapamycin (or related chemicals based on rapamycin) (see, e.g., www.nature.com/nmeth/journal/v2/n6/full/nmeth763.html), 3. GID1-GAI based system inducible by Gibberellin (GA) (see, e.g., www.nature.com/nchembio/journal/v8/n5/full/nchembio.922.html).

[0280] A chemical inducible system can be an estrogen receptor (ER) based system inducible by 4-hydroxytamoxifen (4OHT) (see, e.g., www.pnas.org/content/104/3/1027.abstract). A mutated ligand-binding domain of the estrogen receptor called ERT2 translocates into the nucleus of cells upon binding of 4-hydroxytamoxifen. In further embodiments of the invention any naturally occurring or engineered derivative of any nuclear receptor, thyroid hormone receptor, retinoic acid receptor, estrogen receptor, estrogen-related receptor, glucocorticoid receptor, progesterone receptor, androgen receptor may be used in inducible systems analogous to the ER based inducible system.

[0281] Another inducible system is based on the design using Transient receptor potential (TRP) ion channel-based system inducible by energy, heat or radio-wave (see, e.g., www.sciencemag.org/content/336/6081/604). These TRP family proteins respond to different stimuli, including light and heat. When this protein is activated by light or heat, the ion channel will open and allow the entering of ions such as calcium into the plasma membrane. This influx of ions will bind to intracellular ion interacting partners linked to a polypeptide including the guide and the other components of the Type VII CRISPR-Cas complex or system, and the binding will induce the change of sub-cellular localization of the polypeptide, leading to the entire polypeptide entering the nucleus of cells. Once inside the nucleus, the guide protein and the other components of the Type VII CRISPR-Cas complex will be active and modulating target gene expression in cells.

[0282] While light activation may be an advantageous embodiment, sometimes it may be disadvantageous especially for *in vivo* applications in which the light may not penetrate the skin or other organs. In this instance, other methods of energy activation are contemplated, in particular, electric field energy and/or ultrasound which have a similar effect.

[0283] Electric field energy is preferably administered substantially as described in the art, using one or more electric pulses of from about 1 Volt/cm to about 10 kVolts/cm under *in vivo* conditions. Instead of or in addition to the pulses, the electric field may be delivered in a continuous manner. The electric pulse may be applied for between 1 μs and 500 milliseconds, preferably between 1 μs and 100 milliseconds. The electric field may be applied continuously or in a pulsed manner for 5 about minutes.

[0284] As used herein, 'electric field energy' is the electrical energy to which a cell is exposed. Preferably the electric field has a strength of from about 1 Volt/cm to about 10 kVolts/cm or more under *in vivo* conditions (see WO97/49450).

[0285] As used herein, the term "electric field" includes one or more pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave and/or modulated square wave forms. References to electric fields and electricity should be taken to include reference the presence of an electric potential difference in the environment of a cell. Such an environment may be set up by way of static electricity, alternating current (AC), direct current (DC), etc, as known in the art. The electric field may be uniform, non-uniform or otherwise, and may vary in strength and/or direction in a time dependent manner.

[0286] Single or multiple applications of electric field, as well as single or multiple applications of ultrasound are also possible, in any order and in any combination. The ultrasound and/or the electric field may be delivered as single or multiple continuous applications, or as pulses (pulsatile delivery).

[0287] Electroporation has been used in both *in vitro* and *in vivo* procedures to introduce foreign material into living cells. With *in vitro* applications, a sample of live cells is first mixed with the agent of interest and placed between electrodes such as parallel plates. Then, the electrodes apply an electrical field to the cell/implant mixture. Examples of systems that perform *in vitro* electroporation include the Electro Cell Manipulator ECM600 product, and the Electro Square Porator T820, both made by the BTX Division of Genetronics, Inc (see U.S. Pat. No 5,869,326).

[0288] The known electroporation techniques (both *in vitro* and *in vivo*) function by applying a brief high voltage pulse to electrodes positioned around the treatment region. The electric field generated between the electrodes causes the cell membranes to temporarily become porous, whereupon molecules of the agent of interest enter the cells. In known electroporation applications, this electric field comprises a single square wave pulse on the order of 1000 V/cm, of about 100 microseconds duration. Such a pulse may be generated, for example, in known applications of the Electro Square Porator T820.

[0289] Preferably, the electric field has a strength of from about 1 V/cm to about 10 kV/cm under *in vitro* conditions. Thus, the electric field may have a strength of 1 V/cm, 2 V/cm, 3 V/cm, 4 V/cm, 5 V/cm, 6 V/cm, 7 V/cm, 8 V/cm, 9 V/cm, 10 V/cm, 20 V/cm, 50 V/cm, 100 V/cm, 200 V/cm, 300 V/cm, 400 V/cm, 500 V/cm, 600 V/cm, 700 V/cm, 800 V/cm, 900 V/cm, 1 kV/cm, 2 kV/cm, 5 kV/cm, 10 kV/cm, 20 kV/cm, 50 kV/cm or more. More preferably from about 0.5 kV/cm to about 4.0 kV/cm under *in vitro* conditions. Preferably the electric field has a strength of from about 1 V/cm to about 10 kV/cm under *in vivo* conditions. However, the electric field strengths may be lowered where the number of pulses delivered to the target site are increased. Thus, pulsatile delivery of electric fields at lower field strengths is envisaged.

[0290] Preferably the application of the electric field is in the form of multiple pulses such as double pulses of the same strength and capacitance or sequential pulses of varying strength and/or capacitance. As used herein, the term "pulse" includes one or more electric pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave/square wave forms.

[0291] Preferably the electric pulse is delivered as a waveform selected from an exponential wave form, a square wave form, a modulated wave form and a modulated square wave form.

[0292] A preferred embodiment employs direct current at low voltage. Thus, described herein is the use of an electric field which is applied to the cell, tissue or tissue mass at a field strength of between 1V/cm and 20V/cm, for a period of 100 milliseconds or more, preferably 15 minutes or more.

[0293] Ultrasound is advantageously administered at a power level of from about 0.05 W/cm2 to about 100 W/cm2. Diagnostic or therapeutic ultrasound may be used, or combinations thereof.

[0294] As used herein, the term “ultrasound” refers to a form of energy which consists of mechanical vibrations the frequencies of which are so high they are above the range of human hearing. Lower frequency limit of the ultrasonic spectrum may generally be taken as about 20 kHz. Most diagnostic applications of ultrasound employ frequencies in the range 1 and 15 MHz' (From Ultrasonics in Clinical Diagnosis, P. N. T. Wells, ed., 2nd. Edition, Publ. Churchill Livingstone [Edinburgh, London & NY, 1977]).

[0295] Ultrasound has been used in both diagnostic and therapeutic applications. When used as a diagnostic tool ("diagnostic ultrasound"), ultrasound is typically used in an energy density range of up to about 100 mW/cm2 (FDA recommendation), although energy densities of up to 750 mW/cm2 have been used. In physiotherapy, ultrasound is typically used as an energy source in a range up to about 3 to 4 W/cm2 (WHO recommendation). In other therapeutic applications, higher intensities of ultrasound may be employed, for example, HIFU at 100 W/cm up to 1 kW/cm2 (or even higher) for short periods of time. The term "ultrasound" as used in this specification is intended to encompass diagnostic, therapeutic and focused ultrasound.

[0296] Focused ultrasound (FUS) allows thermal energy to be delivered without an invasive probe (see Morocz et al 1998 Journal of Magnetic Resonance Imaging Vol.8, No. 1, pp.136-142. Another form of focused ultrasound is high intensity focused ultrasound (HIFU) which is reviewed by Moussatov et al in Ultrasonics (1998) Vol.36, No.8, pp.893-900 and TranHuuHue et al in Acustica (1997) Vol.83, No.6, pp.1103-1106.

[0297] Preferably, a combination of diagnostic ultrasound and a therapeutic ultrasound is employed. This combination is not intended to be limiting, however, and the skilled reader will

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

[0298] Preferably, the exposure to an ultrasound energy source is at a power density of from about 0.05 to about 100 Wcm-2. Even more preferably, the exposure to an ultrasound energy source is at a power density of from about 1 to about 15 Wcm-2.

[0299] Preferably, the exposure to an ultrasound energy source is at a frequency of from about 0.015 to about 10.0 MHz. More preferably the exposure to an ultrasound energy source is at a frequency of from about 0.02 to about 5.0 MHz or about 6.0 MHz. Most preferably, the ultrasound is applied at a frequency of 3 MHz.

[0300] Preferably, the exposure is for periods of from about 10 milliseconds to about 60 minutes. Preferably the exposure is for periods of from about 1 second to about 5 minutes. More preferably, the ultrasound is applied for about 2 minutes. Depending on the particular target cell to be disrupted, however, the exposure may be for a longer duration, for example, for 15 minutes.

[0301] Advantageously, the target tissue is exposed to an ultrasound energy source at an acoustic power density of from about 0.05 Wcm-2 to about 10 Wcm-2 with a frequency ranging from about 0.015 to about 10 MHz (see International Patent Publication No. WO 98/52609). However, alternatives are also possible, for example, exposure to an ultrasound energy source at an acoustic power density of above 100 Wcm-2, but for reduced periods of time, for example, 1000 Wcm-2 for periods in the millisecond range or less.

[0302] Preferably, the application of the ultrasound is in the form of multiple pulses; thus, both continuous wave and pulsed wave (pulsatile delivery of ultrasound) may be employed in any combination. For example, continuous wave ultrasound may be applied, followed by pulsed wave ultrasound, or vice versa. This may be repeated any number of times, in any order and combination. The pulsed wave ultrasound may be applied against a background of continuous wave ultrasound, and any number of pulses may be used in any number of groups.

[0303] Preferably, the ultrasound may comprise pulsed wave ultrasound. In a highly preferred embodiment, the ultrasound is applied at a power density of 0.7 Wcm-2 or 1.25 Wcm-2 as a continuous wave. Higher power densities may be employed if pulsed wave ultrasound is used.

[0304] Use of ultrasound is advantageous as, like light, it may be focused accurately on a target. Moreover, ultrasound is advantageous as it may be focused more deeply into tissues

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

[0305] In particular embodiments, the guide molecule is modified by a secondary structure to increase the specificity of the CRISPR-Cas system and the secondary structure can protect against exonuclease activity and allow for 5' additions to the guide sequence also referred to herein as a protected guide molecule.

[0306] In one aspect, the invention provides for hybridizing a "protector RNA" to a sequence of the guide molecule, wherein the "protector RNA" is an RNA strand complementary to the 3' end of the guide molecule to thereby generate a partially double-stranded guide RNA. In an embodiment of the invention, protecting mismatched bases (i.e. the bases of the guide molecule which do not form part of the guide sequence) with a perfectly complementary protector sequence decreases the likelihood of target RNA binding to the mismatched base pairs at the 3' end. In particular embodiments of the invention, additional sequences comprising an extended length may also be present within the guide molecule such that the guide comprises a protector sequence within the guide molecule. This "protector sequence" ensures that the guide molecule comprises a "protected sequence" in addition to an "exposed sequence" (comprising the part of the guide sequence hybridizing to the target sequence). In particular embodiments, the guide molecule is modified by the presence of the protector guide to comprise a secondary structure such as a hairpin. Advantageously there are three or four to thirty or more, e.g., about 10 or more, contiguous base pairs having complementarity to the protected sequence, the guide sequence or both. It is advantageous that the protected portion does not impede thermodynamics of the CRISPR-Cas system interacting with its target. By providing such an extension including a partially double stranded guide molecule, the guide molecule is considered protected and results in improved specific binding of the CRISPR-Cas complex, while maintaining specific activity.

[0307] In particular embodiments, use is made of a truncated guide (tru-guide), i.e. a guide molecule which comprises a guide sequence which is truncated in length with respect to the

canonical guide sequence length. As described by Nowak et al. (Nucleic Acids Res (2016) 44 (20): 9555-9564), such guides may allow catalytically active CRISPR-Cas enzyme to bind its target without cleaving the target RNA. In particular embodiments, a truncated guide is used which allows the binding of the target but retains only nickase activity of the CRISPR-Cas enzyme.

Templates

[0308] In some embodiments, a recombination template is also provided. A recombination template may be a component of another vector as described herein, contained in a separate vector, or provided as a separate polynucleotide. 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.

[0309] 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.

[0310] 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 Type VII effector 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 Type VII mediated event, and a second site on the target sequence that is cleaved in a second Type VII mediated event.

[0311] 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.

[0312] 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.

[0313] 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, 1 1 , 12 or more nucleotides of the target sequence.

[0314] 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, 1 10+/- 10, 120+/- 10, 130+/- 10, 140+/- 10, 150+/- 10, 160+/- 10, 170+/- 10, 1 80+/- 10, 190+/- 10, 200+/- 10, 210+/-10, of 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, 1 10+/-20, 120+/-20, 130+/-20, 140+/-20, I 50+/-20, 160+/-20, 170+/-20, 180+/-20, 190+/-20, 200+/-20, 210+/-20, of 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 to300, 50 to 200, or 50 to 100 nucleotides in length.

[0315] 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.

[0316] 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.

[0317] The upstream and downstream sequences in the exogenous polynucleotide template are selected to promote recombination between the chromosomal sequence of interest and the donor polynucleotide. The upstream sequence is a nucleic acid sequence that shares sequence similarity with the genome sequence upstream of the targeted site for integration. Similarly, the downstream sequence is a nucleic acid sequence that shares sequence similarity with the chromosomal sequence downstream of the targeted site of integration. The upstream and downstream sequences in the exogenous polynucleotide template can have 75%, 80%, 85%, 90%, 95%, or 100% sequence identity with the targeted genome sequence. Preferably, the upstream and downstream sequences in the exogenous polynucleotide template have about 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with the targeted genome sequence. In some methods, the upstream and downstream sequences in the exogenous polynucleotide template have about 99% or 100% sequence identity with the targeted genome sequence.

[0318] 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.

[0319] 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

[0320] 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.

[0321] 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).

[0322] In certain embodiments, a template nucleic acids for correcting a mutation may 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.

[0323] Suzuki et al. describe *in vivo* genome editing via CRISPR/Cas9 mediated homology-independent targeted integration (2016, Nature 540:144–149).

[0324] Accordingly, when referring to the Type VII system herein, in some aspects or embodiments, the CRISPR system comprises (i) a CRISPR protein or a polynucleotide encoding a CRISPR effector protein and (ii) one or more polynucleotides engineered to: complex with the CRISPR protein to form a CRISPR complex; and to complex with the target sequence, and an insertion or repair template. In some embodiments, the Type VII effector protein is a nuclease directing cleavage of one or both strands at the location of the target sequence, or wherein the CRISPR protein is a nickase directing cleavage at the location of the target sequence.

[0325] In some embodiments, the system is a therapeutic for delivery (or application or administration) to a eukaryotic cell, either *in vivo* or *ex vivo*.

Target Sequences and PAMs

Target Sequences

[0326] In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise RNA polynucleotides. The term "target RNA" refers to an RNA polynucleotide being or comprising the target sequence. In other words, the target polynucleotide can be a polynucleotide or a part of a polynucleotide to which a part of the guide sequence is designed to have complementarity with and to which the effector function

mediated by the complex comprising the CRISPR effector protein and a guide molecule is to be directed. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell.

[0327] The guide sequence can specifically bind a target sequence in a target polynucleotide. The target polynucleotide may be DNA. The target polynucleotide may be RNA. The target polynucleotide can have one or more (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, etc. or more) target sequences. The target polynucleotide can be on a vector. The target polynucleotide can be genomic DNA. The target polynucleotide can be episomal. Other forms of the target polynucleotide are described elsewhere herein.

[0328] 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 an 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 (also referred to herein as a target polynucleotide) may be a sequence within an 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 an 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.

PAM Elements

[0329] PAM elements are sequences that can be recognized and bound by Cas proteins. Cas proteins/effector complexes can then unwind the dsDNA at a position adjacent to the PAM element. It will be appreciated that Cas proteins and systems that include them that target RNA do not require PAM sequences (Marraffini et al. 2010. Nature. 463:568-571). Instead, many rely on PFSs, which are discussed elsewhere herein. In certain embodiments, the target sequence should be associated with a PAM (protospacer adjacent motif) or PFS (protospacer flanking sequence or site), that is, a short sequence recognized by the CRISPR complex. Depending on the nature of the CRISPR-Cas protein, the target sequence should be selected, such that its complementary sequence in the DNA duplex (also referred to herein as the non-target sequence) is upstream or downstream of the PAM. In the embodiments, the

complementary sequence of the target sequence is downstream or 3' of the PAM or upstream or 5' of the PAM. The precise sequence and length requirements for the PAM differ depending on the Cas protein used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence). Examples of the natural PAM sequences for different Cas proteins are provided herein below and the skilled person will be able to identify further PAM sequences for use with a given Cas protein.

[0330] The ability to recognize different PAM sequences depends on the Cas polypeptide(s) included in the system. *See* e.g., Gleditzsch et al. 2019. RNA Biology. 16(4):504-517. **Table 3** below shows several Cas polypeptides and the PAM sequence they recognize.

Table 3 – Example PAM Sequences

Cas Protein	PAM Sequence
SpCas9	NGG/NRG
SaCas9	NGRRT or NGRRN
NmeCas9	NNNNGATT
CjCas9	NNNNRYAC
StCas9	NNAGAAW
Cas12a (Cpf1) (including LbCpf1 and AsCpf1)	TTTV
Cas12b (C2c1)	TTT, TTA, and TTC
Cas12c (C2c3)	TA
Cas12d (CasY)	TA
Cas12e (CasX)	5′-TTCN-3′

[0331] In a preferred embodiment, the CRISPR effector protein may recognize a 3' PAM. In certain embodiments, the CRISPR effector protein may recognize a 3' PAM which is 5'H, wherein H is A, C or U.

[0332] Further, engineering of the PAM Interacting (PI) domain on the Cas protein may allow programing of PAM specificity, improve target site recognition fidelity, and increase the versatility of the CRISPR-Cas protein, for example as described for Cas9 in Kleinstiver BP et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. Nature. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592. As further detailed herein, the skilled person will understand that Cas13 proteins may be modified analogously. Gao *et al*, "Engineered Cpf1

Enzymes with Altered PAM Specificities," bioRxiv 091611; doi: http://dx.doi.org/10.1101/091611 (Dec. 4, 2016). Doench et al. created a pool of sgRNAs, tiling across all possible target sites of a panel of six endogenous mouse and three endogenous human genes and quantitatively assessed their ability to produce null alleles of their target gene by antibody staining and flow cytometry. The authors showed that optimization of the PAM improved activity and also provided an on-line tool for designing sgRNAs.

[0333] PAM sequences can be identified in a polynucleotide using an appropriate design tool, which are commercially available as well as online. Such freely available tools include, but are not limited to, CRISPRFinder and CRISPRTarget. Mojica et al. 2009. Microbiol. 155(Pt. 3):733-740; Atschul et al. 1990. J. Mol. Biol. 215:403-410; Biswass et al. 2013 RNA Biol. 10:817-827; and Grissa et al. 2007. Nucleic Acid Res. 35:W52-57. Experimental approaches to PAM identification can include, but are not limited to, plasmid depletion assays (Jiang et al. 2013. Nat. Biotechnol. 31:233-239; Esvelt et al. 2013. Nat. Methods. 10:1116-1121; Kleinstiver et al. 2015. Nature. 523:481-485), screened by a high-throughput *in vivo* model called PAM-SCNAR (Pattanayak et al. 2013. Nat. Biotechnol. 31:839-843 and Leenay et al. 2016.Mol. Cell. 16:253), and negative screening (Zetsche et al. 2015. Cell. 163:759-771).

VECTORS

[0334] Also provided herein are vectors that can contain one or more of the Type VII systems or components thereof described herein. In aspects, the vector can contain one or more polynucleotides encoding one or more elements of an Type VII systems or components thereof described herein. 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 Type VII systems or components thereof 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 Type VII systems or components thereof 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 particles (e.g. viral particles, exosomes, and the like), such as those 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. In general, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked.

[0335] 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. As used herein, the term "host cell" can refer to any cell that can be capable of containing and/or expression or contains and/or expresses a polynucleotide, polypeptide, system, vector, and the like or a combination thereof described herein. In some embodiments, the host cell can produce one or more particles, such as viral particles and/or other vesicles (e.g. an exosome) that can contain one or more polynucleotides, polypeptides, systems, vectors, and the like or a combination thereof described herein.

[0336] 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 aspects of the vectors and vector systems are described elsewhere herein. Any suitable cloning or recombination method can be used to generate a vector, such as a recombinant expression vector. In addition to those described elsewhere, exemplary methods include, but are not limited to, those set forth in U.S. Pat. App. Pub. 2004-0171156 A1.

[0337] In some aspects, the vector can be a bicistronic vector. In some aspects, a bicistronic vector can be used for one or more elements of the Type VII systems or components thereof described herein. In some aspects, expression of elements of the Type VII systems or components thereof described herein can be driven by the CBh promoter. Where the element of the Type VII systems or components thereof is an RNA, its expression can be driven by a Pol III promoter, such as a U6 promoter. In some aspects, the two are combined.

[0338] The vectors described herein can be used to generate modified cells that include the CRISPR-Cas system and/or component(s) thereof described herein. Such cells are described in greater detail elsewhere herein. In certain example embodiments, the transgenic cell may function as an individual discrete volume. In other words, samples comprising a masking construct may be delivered to a cell, for example in a suitable delivery vesicle and if the target is present in the delivery vesicle the CRISPR effector is activated and a detectable signal generated.

[0339] The vector(s) can include the regulatory element(s), e.g., promoter(s). The vector(s) can comprise Cas encoding sequences, and/or a single, but possibly also can comprise at least 3 or 8 or 16 or 32 or 48 or 50 guide RNA(s) (e.g., sgRNAs) encoding sequences, such as 1-2, 1-3, 1-4 1-5, 3-6, 3-7, 3-8, 3-9, 3-10, 3-8, 3-16, 3-30, 3-32, 3-48, 3-50 RNA(s) (e.g., sgRNAs). In a single vector there can be a promoter for each RNA (e.g., sgRNA), advantageously when there are up to about 16 RNA(s); and, when a single vector provides for more than 16 RNA(s),

one or more promoter(s) can drive expression of more than one of the RNA(s), e.g., when there are 32 RNA(s), each promoter can drive expression of two RNA(s), and when there are 48 RNA(s), each promoter can drive expression of three RNA(s). By simple arithmetic and well established cloning protocols and the teachings in this disclosure one skilled in the art can readily practice the invention as to the RNA(s) for a suitable exemplary vector such as AAV, and a suitable promoter such as the U6 promoter.

[0340] By way of example, the packaging limit of AAV is ~4.7 kb. The length of a single U6-gRNA (plus restriction sites for cloning) is 361 bp. Therefore, the skilled person can readily fit about 12-16, e.g., 13 U6-gRNA cassettes in a single vector. This can be assembled by any suitable means, such as a golden gate strategy used for TALE assembly (genome-engineering.org/taleffectors/). The skilled person can also use a tandem guide strategy to increase the number of U6-gRNAs by approximately 1.5 times, e.g., to increase from 12-16, e.g., 13 to approximately 18-24, e.g., about 19 U6-gRNAs. Therefore, one skilled in the art can readily reach approximately 18-24, e.g., about 19 promoter-RNAs, e.g., U6-gRNAs in a single vector, e.g., an AAV vector. A further means for increasing the number of promoters and RNAs in a vector is to use a single promoter (e.g., U6) to express an array of RNAs separated by cleavable sequences. And an even further means for increasing the number of promoter-RNAs in a vector, is to express an array of promoter-RNAs separated by cleavable sequences in the intron of a coding sequence or gene; and, in this instance it is advantageous to use a polymerase II promoter, which can have increased expression and enable the transcription of long RNA in a tissue specific manner. (see, e.g., nar.oxfordjournals.org/content/34/7/e53.short and nature.com/mt/journal/v16/n9/abs/mt2008144a.html). In an advantageous embodiment, AAV may package U6 tandem gRNA targeting up to about 50 genes. Accordingly, from the knowledge in the art and the teachings in this disclosure the skilled person can readily make and use vector(s), e.g., a single vector, expressing multiple RNAs or guides under the control or operatively or functionally linked to one or more promoters—especially as to the numbers of RNAs or guides discussed herein, without any undue experimentation.

[0341]

Cell-based Vector Amplification and Expression

[0342] Vectors can be designed for expression of one or more elements of the Type VII systems or components thereof described herein (e.g. nucleic acid transcripts, proteins, enzymes, and combinations thereof) in a suitable host cell. In some aspects, 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. The vectors can be viral-based or non-viral based. In some aspects, the suitable host cell is a eukaryotic cell. In some aspects, 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 aspects, 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 aspects, the host cell is a suitable yeast cell. In some aspects, the yeast cell can be from Saccharomyces cerevisiae. In some aspects, 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).

[0343] In some aspects, 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, 2μ plasmids, yeast integrative plasmids, yeast replicative plasmids, shuttle vectors, and episomal plasmids.

[0344] In some aspects, the vector is a baculovirus vector or expression vector and can be suitable for expression of polynucleotides and/or proteins in insect cells. 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).

[0345] In some embodiments, the vector is a mammalian expression vector. In some aspects, 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.

[0346] 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.

[0347] 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 aspects 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 Type VII systems or components thereof so as to drive expression of the one or more elements of the Type VII systems or components thereof described herein.

[0348] In some embodiments, the guide RNA(s) encoding sequences and/or Cas encoding sequences, can be functionally or operatively linked to regulatory element(s) and hence the regulatory element(s) drive expression. The promoter(s) can be constitutive promoter(s) and/or conditional promoter(s) and/or inducible promoter(s) and/or tissue specific promoter(s). The promoter can be selected from the group consisting of RNA polymerases, pol I, pol II, pol III, T7, U6, H1, retroviral Rous sarcoma virus (RSV) LTR promoter, the cytomegalovirus (CMV) promoter, the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter. An advantageous promoter is the promoter is U6.

[0349] Vectors may be introduced and propagated in a prokaryote or prokaryotic cell. In some aspects, a prokaryote is used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a eukaryotic cell (e.g. amplifying a plasmid as part of a viral vector packaging system). In some aspects, 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.

[0350] In some aspects, the vector can be a fusion vector or fusion expression vector. In some aspects, 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 aspects, 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 aspects, 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).

[0351] In some embodiments, one or more vectors driving expression of one or more elements of a Type VII systems and/or components thereof described herein are introduced into a host cell such that expression of the elements of the engineered delivery system described herein direct formation of a Type VII system and/or components thereof described herein (including but not limited to an particles, which is described in greater detail elsewhere herein). For example, different elements of the Type VII systems and/or components thereof described herein can each be operably linked to separate regulatory elements on separate vectors. RNA(s) of different elements of the engineered delivery system described herein can be delivered to an

animal or mammal or cell thereof to produce an animal or mammal or cell thereof that constitutively or inducibly or conditionally expresses different elements of the Type VII systems and/or components thereof described herein that incorporates one or more elements of the Type VII systems or components thereof described herein or contains one or more cells that incorporates and/or expresses one or more elements of the Type VII systems or components thereof described herein.

[0352] In some aspects, 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. Type VII systems or components thereof 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 Type VII system or component thereof 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 Type VII systems and/or components thereof polynucleotides can be operably linked to and expressed from the same promoter.

Vector Features

[0353] 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 there from, 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

[0354] In aspects, the polynucleotides and/or vectors thereof described herein (such as the Type VII system and/or component thereof 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), and other expression control elements (e.g., transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990).

[0355] 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 EF1α 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).

[0356] In some aspects, 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 aspects, the vector can contain a minimal promoter. In some aspects, the minimal promoter is the Mecp2 promoter, tRNA promoter, or U6. In a further embodiment, the minimal promoter is tissue specific. In some aspects, the length of the vector polynucleotide the minimal promoters and polynucleotide sequences is less than 4.4Kb. The present application provides a vector for delivering an effector protein and at least one CRISPR guide

RNA to a cell comprising 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 is an AAV vector. In another embodiment, the effector protein is a CRISPR enzyme. In a further embodiment, the CRISPR enzyme is a Type VII CRISPR enzyme. In a related aspect, the invention provides a lentiviral vector for delivering an effector protein and at least one CRISPR guide RNA to a cell comprising a promoter operably linked to a polynucleotide sequence encoding Type VII effector and a second promoter operably linked to a polynucleotide sequence encoding at least one guide RNA, wherein the polynucleotide sequences are in reverse orientation. In another aspect, the invention provides a method of expressing an effector protein and guide RNA in a cell comprising introducing the vector according any of the vector delivery systems disclosed herein. In an embodiment of the vector for delivering an effector protein, the minimal promoter is the Mecp2 promoter, tRNA promoter, or U6. In a further embodiment, the minimal promoter is tissue specific.

[0357] 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 aspects 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, CBh, EF-1α, β-actin, RSV, ferritin heavy chain, ferritin light chain, mouse mammary tumor virus (MMTV), human immunodeficiency virus (HIV) long terminal repeat (LTR) promoter, MoMuLV promoter, avian leukemia virus promoter, Epstein-Barr virus immediate early promoter, human actin promoter, myosin promoter, hemoglobin promoter, creatine kinase promoter, 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. In some embodiments an AAV ITR can serve as a promoter. An AAV ITR can have the advantage of eliminating the need for an additional promoter element, which can free up space for other polynucleotides in the vector. Also, ITR activity can be relatively weaker, so such a promoter can be used to reduce potential toxicity due to over expression of a Type VII effector.

[0358] In some aspects, 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 aspects, 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. albumin, 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 (e.g. for all neurons), CaMKIIalpha (e.g. for excitatory neurons), GAD67 (e.g. for GABAergic neurons), GAD65 (e.g. for GABAergic neurons), VGAT (e.g. for GABAergic neurons), 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, Fer1l4), endothelial cell specific promoters (e.g. ENG and ICAM), pluripotent and embryonic germ layer cell specific promoters (e.g. Oct4, NANOG, Synthetic Oct4, T brachyury, NES, SOX17, FOXA2, MIR122), lung specific promoters (e.g. SP-B), hematopoeitc cell specific promoters (e.g. IFNbeta and CD45), osteoblast specific promoters (e.g. OG-2), 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.

[0359] 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 promotor 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. The use of an inducible promoter provides a molecular switch capable of turning on expression of the polynucleotide sequence which it is operatively linked when such expression is desired, or turning off the expression when expression is not desired. Additional examples of inducible promoters include, but are not limited to a metallothionine promoter, a glucocorticoid promoter, a progesterone promoter, and a tetracycline promoter.

[0360] Where expression in a plant cell is desired, the components of the Type VII systems and/or components thereof 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.

[0361] 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 Type VII systems or components thereof 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 Type VII systems or components thereof 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.

[0362] 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 Type VII systems or components thereof described herein, a light-responsive cytochrome heterodimer (e.g. from Arabidopsis thaliana), and a transcriptional activation/repression domain. In some aspects, 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. US 2015/0291966, US 2017/0166903, and US 2019/0203212, which describe e.g. aspects of inducible DNA binding proteins and methods of use and can be adapted for use with the present invention.

[0363] In some aspects, transient or inducible expression can be achieved by including, for example, chemical-regulated promotors, 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.

[0364] In some aspects, the vector or system thereof can include one or more elements capable of translocating and/or expressing a Type VII system or components thereof 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.

[0365] Additional promoter elements, e.g., enhancers, regulate the frequency of transcriptional initiation. Typically, these are located in the region 30-110 bp upstream of the start site, although a number of promoters have recently been shown to contain functional elements downstream of the start site as well. The spacing between promoter elements frequently is flexible, so that promoter function is preserved when elements are inverted or moved relative to one another. In the thymidine kinase (tk) promoter, the spacing between

promoter elements can be increased to 50 bp apart before activity begins to decline. Depending on the promoter(s) and/or other regulatory elements included, individual elements can function either cooperatively or independently to activate transcription.

Selectable Markers and Tags

[0366] One or more of the Type VII systems or components thereof polynucleotides can be 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 aspects, the polypeptide encoding a polypeptide selectable marker can be incorporated in the Type VII system and/or components thereof polynucleotide such that the selectable marker polypeptide, when translated, is inserted between two amino acids between the N- and C-terminus of the Type VII systems or components thereof polypeptide or at the N- and/or C-terminus of the Type VII system or component thereof polypeptide. In some aspects, the selectable marker or tag is a polynucleotide barcode or unique molecular identifier (UMI).

[0367] 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 Type VII system and/or component thereof 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.

[0368] 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.

[0369] Selectable markers and tags can be operably linked to one or more components of the Type VII system and/or component thereof described herein via suitable linker, such as a glycine or glycine serine linkers as short as GS or GG up to $(GGGGG)_3$ or $(GGGGS)_3$. Other suitable linkers are described elsewhere herein.

[0370] The vector or vector system can include one or more polynucleotides encoding one or more targeting moieties. In some aspects, 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 aspects, the targeting moiety encoding polynucleotides can be included in the vector or vector system such that the Type VII systems and/or components thereof polynucleotide(s) and/or products expressed therefrom include the targeting moiety and can be targeted to specific cells, tissues, organs, etc. In some aspects, 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 Type VII systems and/or components thereof polynucleotide(s) to specific cells, tissues, organs, etc.

[0371] In some embodiments, the selectable marker is incorporated into the system via a gene (also referred to as a "reporter gene,") which encodes a selectable marker protein. In general, a reporter gene is a gene that is not present in or expressed by the recipient organism or tissue and that encodes a polypeptide whose expression is manifested by some easily detectable property, e.g., enzymatic activity. Expression of the reporter gene is assessed at a suitable time after the DNA has been introduced into the recipient cells. Suitable reporter genes

may include genes encoding luciferase, beta-galactosidase, chloramphenicol acetyl transferase, secreted alkaline phosphatase, or the green fluorescent protein gene (e.g., Ui-Tei et al., 2000 FEBS Letters 479: 79-82). Suitable expression systems are well known and may be prepared using known techniques or obtained commercially. In general, the construct with the minimal 5′ flanking region showing the highest level of expression of reporter gene is identified as the promoter. Such promoter regions may be linked to a reporter gene and used to evaluate agents for the ability to modulate promoter-driven transcription.

[0372] Selectable markers and/or reporter genes can be used to facilitate identification and selection of expressing cells from the population of cells sought to be transfected or infected through viral vectors. In other aspects, the selectable marker may be carried on a separate piece of DNA and used in a co-transfection procedure. Both selectable markers and reporter genes may be flanked with appropriate regulatory sequences to enable expression in the host cells. Useful selectable markers include, for example, antibiotic-resistance genes, such as neo and the like.

Cell-free Vector and Polynucleotide Expression

[0373] In some aspects, the polynucleotide encoding one or more features of the Type VII system and/ or components thereof 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.

[0374] *In vitro* translation can be stand-alone (e.g. translation of a purified polyribonucleotide) or linked/coupled to transcription. In some aspects, 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.

Codon Optimization of Vector Polynucleotides

[0375] As described elsewhere herein, the polynucleotide encoding one or more aspects of the Type VII systems or components thereof described herein can be codon optimized. In some aspects, one or more polynucleotides contained in a vector ("vector polynucleotides") described herein that are in addition to an optionally codon optimized polynucleotide encoding aspects of the Type VII systems and/or components thereof 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.orjp/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 http://www.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.

[0376] The vector polynucleotide can be codon optimized for expression in a specific cell-type, tissue type, organ type, and/or subject type. In some aspects, 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 aspects, 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 aspects, 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 aspects, 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.

[0377] 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.

Non-Viral Vectors and Carriers

[0378] In some aspects, the vector is a non-viral vector or carrier. In some aspects, non-viral vectors can have the advantage(s) of reduced toxicity and/or immunogenicity and/or increased bio-safety as compared to viral vectors The terms of art "Non-viral vectors and carriers" and as used herein in this context refers to molecules and/or compositions 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 attaching to, incorporating, coupling, and/or otherwise interacting with a Type VII system and/or components thereof polynucleotide of the present invention and can be capable of ferrying the polynucleotide to a cell and/or expressing the polynucleotide. It will be appreciated that this does not exclude the inclusion of 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 and carriers include naked polynucleotides, chemical-based carriers, polynucleotide (non-viral) based vectors, and particle-based carriers. It will be appreciated that the term "vector" as used in the context of non-viral vectors and carriers refers to polynucleotide vectors and "carriers" used in this context refers to a non-nucleic acid or polynucleotide molecule or composition that be attached to or otherwise interact with a polynucleotide to be delivered, such as an Type VII systems or components thereof and/or polynucleotide of the present invention.

Naked Polynucleotides

[0379] In some aspects one or more Type VII system and or component thereof 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 Type VII systems and/or components thereof 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 aspects, the naked polynucleotide contains only the Type VII systems and/or components thereof polynucleotide(s) of the present invention. In some aspects, the naked polynucleotide can contain other nucleic acids and/or polynucleotides in addition to the Type VII systems and/or components thereof 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

[0380] In some aspects, one or more of the Type VII systems and/or components thereof 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.

[0381] In some aspects, the non-viral polynucleotide vector can have a conditional origin of replication. In some aspects, the non-viral polynucleotide vector can be an ORT plasmid. In some aspects, the non-viral polynucleotide vector can have a minimalistic immunologically defined gene expression. In some aspects, the non-viral polynucleotide vector can have one or more post-segregationally killing system genes. In some aspects, the non-viral polynucleotide vector is AR-free. In some aspects, the non-viral polynucleotide vector is a minivector. In some

aspects, the non-viral polynucleotide vector includes a nuclear localization signal. In some aspects, the non-viral polynucleotide vector can include one or more CpG motifs. In some aspects, 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 aspects, the S/MAR sequence is located downstream of an actively transcribed polynucleotide (e.g. one or more Type VII systems and/or components thereof polynucleotides of the present invention) included in the non-viral polynucleotide vector. In some aspects, 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.

[0382] In some aspects, 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 form 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 aspects, the non-viral polynucleotide vector can be a retrotransposon vector. In some aspects, the retrotransposon vector includes long terminal repeats. In some aspects, the retrotransposon vector does not include long terminal repeats. In some aspects, the non-viral polynucleotide vector can be a DNA transposon vector. DNA transposon vectors can include a polynucleotide sequence encoding a transposase. In some aspects, 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 aspects, the transposon vector lacks one or

more polynucleotide sequences encoding proteins required for transposition. In some aspects, the non-autonomous transposon vectors lack one or more Ac elements.

[0383] In some aspects a non-viral polynucleotide transposon vector system can include a first polynucleotide vector that contains the Type VII systems and/or components thereof 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 Type VII systems and/or components thereof polynucleotide(s) of the present invention) and integrate it into one or more positions in the host cell's genome. In some aspects the transposon vector or system thereof can be configured as a gene trap. In some aspects, 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 Type VII systems and/or components thereof 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.

[0384] 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.

Chemical Carriers

[0385] In some aspects the Type VII systems and/or components thereof polynucleotide(s) can be coupled to a chemical carrier. Chemical carriers that can be suitable for delivery of polynucleotides can be broadly classified into the following classes: (i) inorganic particles, (ii) lipid-based, (iii) polymer-based, and (iv) peptide based. They can be categorized as (1) those that can form condensed complexes with a polynucleotide (such as the Type VII systems and/or components thereof polynucleotide(s) of the present invention), (2) those capable of targeting specific cells, (3) those capable of increasing delivery of the polynucleotide (such as the Type VII systems and/or components thereof polynucleotide(s) of the present invention) to the

nucleus or cytosol of a host cell, (4) those capable of disintegrating from DNA/RNA in the cytosol of a host cell, and (5) those capable of sustained or controlled release. It will be appreciated that any one given chemical carrier can include features from multiple categories. The term "particle" as used herein, refers to any suitable sized particles for delivery of the CRISPR-Cas systems and components thereof described herein. Suitable sizes include macro-, micro-, and nano-sized particles. Particles can be viral or non-viral particles.

[0386] In some aspects, the non-viral carrier can be an inorganic particle. In some aspects, the inorganic particle, can be a nanoparticle. The inorganic particles can be configured and optimized by varying size, shape, and/or porosity. In some aspects, the inorganic particles are optimized to escape from the reticulo endothelial system. In some aspects, the inorganic particles can be optimized to protect an entrapped molecule from degradation., the Suitable inorganic particles that can be used as non-viral carriers in this context can include, but are not limited to, calcium phosphate, silica, metals (e.g. gold, platinum, silver, palladium, rhodium, osmium, iridium, ruthenium, mercury, copper, rhenium, titanium, niobium, tantalum, and combinations thereof), magnetic compounds, particles, and materials, (e.g. supermagnetic iron oxide and magnetite), quantum dots, fullerenes (e.g. carbon nanoparticles, nanotubes, nanostrings, and the like), and combinations thereof. Other suitable inorganic non-viral carriers are discussed elsewhere herein.

[0387] In some aspects, the non-viral carrier can be lipid-based. Suitable lipid-based carriers are also described in greater detail herein. In some aspects, the lipid-based carrier includes a cationic lipid or an amphiphilic lipid that is capable of binding or otherwise interacting with a negative charge on the polynucleotide to be delivered (e.g. such as an Type VII systems and/or components thereof polynucleotide of the present invention). In some aspects, chemical non-viral carrier systems can include a polynucleotide such as the Type VII systems and/or components thereof polynucleotide(s) of the present invention) and a lipid (such as a cationic lipid). These are also referred to in the art as lipoplexes. Other aspects of lipoplexes are described elsewhere herein. In some aspects, the non-viral lipid-based carrier can be a lipid nano emulsion. Lipid nano emulsions can be formed by the dispersion of an immisicible liquid in another stabilized emulsifying agent and can have particles of about 200 nm that are composed of the lipid, water, and surfactant that can contain the polynucleotide to be delivered (e.g. the Type VII systems and/or components thereof polynucleotide(s) of the present

invention). In some aspects, the lipid-based non-viral carrier can be a solid lipid particle or nanoparticle.

[0388] 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).

[0389] In some aspects, the non-viral carrier can be peptide-based. In some aspects, the peptide-based non-viral carrier can include one or more cationic amino acids. In some aspects, 35 to 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 99 or 100 % of the amino acids are cationic. In some aspects, peptide carriers can be used in conjunction with other types of carriers (e.g. polymer-based carriers and lipid-based carriers to functionalize these carriers). In some aspects, the functionalization is targeting a host cell. Suitable polymers that can be included in the polymer-based non-viral carrier can include, but are not limited to, polyethylenimine (PEI), chitosan, poly (DL-lactide) (PLA), poly (DL-Lactide-co-glycoside) (PLGA), dendrimers (see e.g. US Pat. Pub. 2017/0079916 whose techniques and compositions can be adapted for use with the Type VII systems and/or components thereof polynucleotide(s) of the present invention), polymethacrylate, and combinations thereof.

[0390] In some aspects, the non-viral carrier can be configured to release an engineered delivery system polynucleotide that is associated with or attached to the non-viral carrier in response to an external stimulus, such as pH, temperature, osmolarity, concentration of a specific molecule or composition (e.g. calcium, NaCl, and the like), pressure and the like. In some aspects, the non-viral carrier can be a particle that is configured includes one or more of the Type VII systems and/or components thereof polynucleotide(s) describe herein and an environmental triggering agent response element, and optionally a triggering agent. In some aspects, the particle can include a polymer that can be selected from the group of polymethacrylates and polyacrylates. In some aspects, the non-viral particle can include one or more aspects of the compositions microparticles described in US Pat. Pubs. 20150232883 and 20050123596, whose techniques and compositions can be adapted for use in the present invention.

[0391] In some aspects, the non-viral carrier can be a polymer-based carrier. In some aspects, the polymer is cationic or is predominantly cationic such that it can interact in a charge-dependent manner with the negatively charged polynucleotide to be delivered (such as the Type VII systems and/or components thereof polynucleotide(s) of the present invention). Polymer-based systems are described in greater detail elsewhere herein.

Viral Vectors

[0392] In some aspects, 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 an Type VII systems and/or components thereof polynucleotide(s) 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 Type VII systems and/or components thereof described herein. The viral vector can be part of a viral vector system involving multiple vectors. In some aspects, 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 aspects of viral vectors and viral particles produce therefrom are described elsewhere herein. In some aspects, the viral vectors are configured to produce replication incompetent viral particles for improved safety of these systems.

Retroviral and Lentiviral Vectors

[0393] 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 Type VII systems and/or components thereof can include, but are not limited to, 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); Sommnerfelt 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.

[0394] 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. 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. Cell type specific promoters can be used to target expression in specific cell types. Such vector components and systems are described in greater detail elsewhere herein.

[0395] 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.

[0396] In some embodiments, an HIV-based lentiviral vector system can be used. In some embodiments, a FIV-based lentiviral vector system can be used.

[0397] In some aspects, 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 Type VII systems and/or components thereof described herein.

[0398] In some aspects, 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, vprm 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.

[0399] In some aspects, 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 aspects, the second-generation vector lacks one or more accessory virulence factors (e.g. vif, vprm, 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 aspects, 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.

[0400] In some aspects, 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 aspects, 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 aspects, 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.

[0401] In some aspects, 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 Type VII systems and/or components thereof of the present invention.

[0402] In some aspects, 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 aspects, 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.

[0403] In some aspects, 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 aspects, 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. 2013. 21: 849-859.

[0404] In some aspects, 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 aspects, 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.

[0405] In some aspects, 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 aspects, a lentiviral vector can include an N-termial PDZ domain of InaD protein (PDZ1) and its pentapeptide ligand (TEFCA) from NorpA, which can conjugate the cell targeting peptide to the virus particle via a covalent bond (e.g. a disulfide bond). In some aspects, 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 aspects, the TEFCA can be fused to a cell targeting peptide and the TEFCA-CPT fusion construct can be incorporated into the same or a different lentiviral vector as the PDZ1-envenlope protein construct. During virus production, specific interaction between the PDZ1 and TEFCA 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.

[0406] 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 Type VII systems and/or components thereof polynucleotide(s) described herein to a cell.

[0407] In some aspects, 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.

[0408] Lentiviruses with a packaged cargo polynucleotide can be prepared by any suitable method. An exemplary method is as follows and can be adapted for production of lentiviral particles containing a Type VII CRISPR-Cas system molecule of the present invention

described herein. 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 supplemented with 10% fetal bovine serum and without antibiotics. After about 20 hours, media can be changed to OptiMEM (serum-free) media and transfection can be completed about 4 hours later. Cells can be transfected with about 10 μg of lentiviral transfer plasmid (pCasES10) and the following packaging plasmids: 5 μg of pMD2.G (VSV-g pseudotype), and 7.5ug of psPAX2 (gag/pol/rev/tat). Transfection can be completed in 4mL OptiMEM with a cationic lipid delivery agent (50uL Lipofectamine 2000 and 100ul Plus reagent). After about 6 hours, the media can be changed to antibiotic-free DMEM with 10% fetal bovine serum. This exemplary method uses serum during cell culture, but serum-free methods are can be used and, in some embodiments, are preferred.

[0409] Lentivirus particles produced can be purified using any suitable method. An exemplary method is as follows and can be adapted for production of lentiviral particles containing a Type VII CRISPR-Cas system molecule of the present invention described herein. Viral particle-containing supernatants and/or culture media can be harvested after about 48 hours. Supernatants and/or cell culture media can be first cleared of debris and filtered through a 0.45um low protein binding (PVDF) filter. They can then be spun in a ultracentrifuge for about 2 hours at 24,000 rpm. Viral particle pellets can be resuspended in 50ul of DMEM overnight at 4 °C. They can be aliquoted as desired and used immediately or frozen at -80 °C for storage.

[0410] In some embodiments, the lentiviral vector is an EIAV vector. EIAV vectors can be used to deliver the Type VII system or component thereof to the eye or be advantageous for use in ocular therapies (see, e.g., Balagaan, J Gene Med 2006; 8: 275 – 285). In some embodiments, RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostatin and angiostatin that is delivered via a subretinal injection for the treatment of the web form of age-related macular degeneration can be used to deliver the Type VII system or component thereof described herein (see, e.g., Binley et al., HUMAN GENE THERAPY 23:980–991 (September 2012)) and this vector may be modified for the CRISPR-Cas system of the present invention.

[0411] In another embodiment, self-inactivating lentiviral vectors with an siRNA targeting a common exon shared by HIV tat/rev, a nucleolar-localizing TAR decoy, and an anti–CCR5-specific hammerhead ribozyme (see, e.g., DiGiusto et al. (2010) Sci Transl Med 2:36ra43) may

be used/and or adapted to the CRISPR-Cas system of the present invention. A minimum of 2.5 × 10^6 CD34+ cells per kilogram patient weight may be collected and prestimulated for 16 to 20 hours in X-VIVO 15 medium (Lonza) containing 2 μmol/L-glutamine, stem cell factor (100 ng/ml), Flt-3 ligand (Flt-3L) (100 ng/ml), and thrombopoietin (10 ng/ml) (CellGenix) at a density of 2 × 10^6 cells/ml. Prestimulated cells may be transduced with lentiviral at a multiplicity of infection of 5 for 16 to 24 hours in 75-cm^2 tissue culture flasks coated with fibronectin (25 mg/cm^2) (RetroNectin,Takara Bio Inc.).

[0412]

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

[0413] In some aspects, the vector can be an adenoviral vector. In some aspects, 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 aspects, the polynucleotide to be delivered via the adenoviral particle can be up to about 8 kb. Thus, in some aspects, 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. Teramato 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.

[0414] In some aspects the vector can be a helper-dependent adenoviral vector or system thereof. These are also referred to in the art as "gutless" or "gutted" vectors and are a modified generation of adenoviral vectors (see e.g. Thrasher et al. 2006. Nature. 443:E5-7). In aspects 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 Type VII systems and/or components thereof 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 Type VII systems and/or components thereof polynucleotide(s) described herein. In some aspects, 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 aspects, 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).

[0415] In some aspects, 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 aspects, 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 Type VII system of the present invention. In some aspects, a hybrid-adenoviral vector can include one or more features of a retrovirus and/or an adeno-associated virus. In some aspects 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 Type VII systems 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. 156:146-156 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 Type VII systems of the present invention.

Adeno Associated Viral (AAV) Vectors

[0416] 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; International Patent Publication No. 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 that adenoviral vectors. In some aspects the AAV can integrate into a specific site on chromosome 19 of a human cell with no observable side effects. In some aspects, the capacity of the AAV vector, system thereof, and/or AAV particles can be up to about 4.75 kb. This can mean that a Type VII effector described herein as well as a promoter and transcription terminator can be incorporated into the same viral vector. Constructs larger than about 4.75 kb can result in reduced virus particle production and low virus titers produced by a producer cell. AAVs can have low toxicity and/or a low probability of causing insertional mutagenesis as they do not integrate into the host genome.

[0417] The AAV vector or system thereof can include one or more regulatory molecules. In some aspects the regulatory molecules can be promoters, enhancers, repressors and the like, which are described in greater detail elsewhere herein. In some aspects, the AAV vector or system thereof can include one or more polynucleotides that can encode one or more regulatory proteins. In some aspects, the one or more regulatory proteins can be selected from Rep78, Rep68, Rep52, Rep40, variants thereof, and combinations thereof.

[0418] 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 aspects, the AAV capsid can contain 60 capsid proteins. In some aspects, the ratio of VP1:VP2:VP3 in a capsid can be about 1:1:10.

[0419] In some aspects, 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 aspects, a producing host cell line expresses one or more of the adenovirus helper factors.

[0420] The AAV vector or system thereof can be configured to produce AAV particles having a specific serotype. In some aspects, 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 aspects, 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 aspects, 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 aspects, 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 aspects, 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 aspects, 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 2nd 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.

[0421] 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 below in **Table 4**.

Table 4

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

[0422] In some aspects, the AAV vector or system thereof is configured as a "gutless" vector, similar to that described in connection with a retroviral vector. In some aspects, 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 Type VII systems and/or components thereof polynucleotide(s)).

[0423] In some embodiments, the AAV contains or consists essentially of an exogenous nucleic acid molecule encoding a CRISPR system described herein (e.g. a Type VII system), e.g., 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 Type VII effector molecule or other effector molecule described herein and a terminator, and a two, or more, advantageously up to the packaging size limit of the vector. In some embodiments, the total number of cassettes (including the first cassette) is five, with 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 Type VII effector, Cas9 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. The promoter is in some embodiments advantageously human Synapsin I promoter (hSyn). In another embodiment, multiple gRNA expression cassettes along with the Type VII and/or Cas9

expression cassette can be delivered in a high-capacity adenoviral vector (HCAdV), from which all AAV coding genes have been removed. *See e.g,* Schiwon et al., "One-Vector System for Multiplexed CRISPR/Cas9 against Hepatitis B Virus cccDNA Utilizing High-Capacity Adenoviral Vectors" Mol Ther Nucleic Acids. 2018 Sep 7; 12: 242–253; and Ehrke-Schulz et al., "CRISPR/Cas9 delivery with one single adenoviral vector devoid of all viral genes" Sci Rep. 2017; 7: 17113. Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

[0424] In some embodiments, an AAV vector can include additional sequence information encoding sequences that facilitate transduction or that assist in evasion of the host immune system. In one embodiment, CRISPR-Cas9 can be delivered to astrocytes using an AAV vector that includes a synthetic surface peptide for transduction of astrocytes. *See, e.g.* Kunze et al., "Synthetic AAV/CRISPR vectors for blocking HIV-1 expression in persistently infected astrocytes" Glia. 2018 Feb;66(2):413-427. In another embodiment, CRISPR-Cas9 can be delivered in a capsid engineered AAV, for example an AAV that has been engineered to include "chemical handles" on the AAV surface and be complexed with lipids to produce a "cloaked AAV" that is resistant to endogenous neutralizing antibodies in the host. *See, e.g.* Katrekar et al., "Oligonucleotide conjugated multi-functional adeno-associated viruses" Sci Rep. 2018; 8: 3589.

[0425] In some embodiments, rAAV 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). In some embodiments, the rAAV producer cell (i.e. the cell that produced AAV viral particles) can be an insect cell.

Herpes Simplex Viral Vectors

[0426] In some aspects, 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 Type VII systems and/or components thereof of the present invention. In some aspects

where an HSV vector or system thereof is utilized, the host cell can be a complementing cell. In some aspects, 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 aspect the Type VII system and/or components thereof 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. Neruosci. 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 Type VII systems of the present invention.

Poxvirus Vectors

[0427] In some aspects, the vector can be a poxvirus vector or system thereof. In some aspects, the poxvirus vector can result in cytoplasmic expression of one or more Type VII systems and/or components thereof of the present invention. In some aspects the capacity of a poxvirus vector or system thereof can be about 25 kb or more. In some aspects, a poxivirus vector or system thereof can include one or more Type VII system or component there of polynucleotide(s).

Other Vectors

Dual Vector Systems

[0428] Also contemplated is delivery by dual vector systems. In one embodiment, expression cassettes of a Type VII effector and/or Cas9 and/or gRNA can be delivered via a dual vector system. Such systems can include, for example, a first AAV vector encoding a gRNA and an N-terminal Cas9 and/or Type VII effector and a second AAV vector containing a C- terminal Cas9 and/or Type VII effector. *See, e.g.* Moreno et al., "In Situ Gene Therapy via AAV-CRISPR-Cas9-Mediated Targeted Gene Regulation" Mol Ther. 2018 Jul 5;26(7):1818-1827. In another embodiment, the Type VII effector and/or other Cas protein (e.g. Cas9) can be separated into two parts that are expressed individually and reunited in the cell by various means, including use of 1) the gRNA as a scaffold for effector (e.g. Type VII Cas9 assembly; 2) the rapamycin-controlled FKBP/FRB system; 3) the light-regulated Magnet system; or 4) inteins. *See, e.g.* Schmelas et al., "Split Cas9, Not Hairs - Advancing the

Therapeutic Index of CRISPR Technology" Biotechnol J. 2018 Sep;13(9):e1700432. doi: 10.1002/biot.201700432. Epub 2018 Feb 2.

Cocal vesiculovirus Pseudotyped Vector Systems

[0429] In another embodiment, Cocal vesiculovirus envelope pseudotyped retroviral 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 epsilonretroviral vector particles that may comprise retroviral Gag, Pol, and/or one or more accessory protein(s) and a Cocal vesiculovirus envelope protein. Within certain aspects of these embodiments, the Gag, Pol, and accessory proteins are lentiviral and/or gammaretroviral.

Vector Construction

[0430] The vectors described herein can be constructed using any suitable process or technique. In some aspects, 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. Application publication No. US 2004-0171156 A1. Other suitable methods and techniques are described elsewhere herein.

[0431] Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260

(1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989). 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.

[0432] In some embodiments, the vector can have 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.

[0433] Delivery vehicles, vectors, particles, nanoparticles, formulations and components thereof for expression of one or more elements of a Type VII system(s) and/or component(s) thereof 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.

Virus Particle Production from Viral Vectors

Retroviral Production

[0434] In some aspects, 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 aspects, 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.

[0435] In some aspects, 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. a Type VII system and/or component thereof 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.

[0436] Mature virus particles can be collected from the culture media by a suitable method. In some aspects, 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 aspects, the resulting composition containing virus particles can contain 1 $X10^{1}$ -1 X 10^{20} particles/mL.

AAV Particle Production

[0437] 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 aspects, 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 Type VII systems and/or components thereof polynucleotide(s)). In some aspects, 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 Type VII systems and/or components thereof 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.

Vector and Virus Particle Delivery

[0438] A vector (including non-viral carriers) described herein can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides encoded by nucleic acids as described herein (e.g., Type VII systems and/or components thereof transcripts, proteins, enzymes, mutant forms thereof, fusion proteins thereof, etc.), and virus particles (such as from viral vectors and systems thereof).

[0439] One or more Type VII systems and/or components thereof polynucleotides can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other plasmid or viral

vector types, in particular, using formulations and doses from, for example, US Patents Nos. 8,454,972 (formulations, doses for adenovirus), 8,404,658 (formulations, doses for AAV) and 5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For examples, for AAV, the route of administration, formulation and dose can be as in US Patent No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in US Patent No. 8,404,658 and as in clinical trials involving adenovirus.

[0440] For plasmid delivery, the route of administration, formulation and dose can be as in US Patent No 5,846,946 and as in clinical studies involving plasmids. In some aspects, doses can be based on or extrapolated to an average 70 kg individual (e.g. a male adult human), and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed. The viral vectors can be injected into or otherwise delivered to the tissue or cell of interest.

[0441] In terms of *in vivo* delivery, AAV is advantageous over other viral vectors for a couple of reasons such as low toxicity (this may be due to the purification method not requiring ultra-centrifugation of cell particles that can activate the immune response) and a low probability of causing insertional mutagenesis because it doesn't integrate into the host genome.

[0442] The vector(s) and virus particles described herein can be delivered in to a host cell *in vitro*, *in vivo*, and or *ex vivo*. Delivery can occur by any suitable method including, but not limited to, physical methods, chemical methods, and biological methods. Physical delivery methods are those methods that employ physical force to counteract the membrane barrier of the cells to facilitate intracellular delivery of the vector. Suitable physical methods include, but are not limited to, needles (e.g. injections), ballistic polynucleotides (e.g. particle bombardment, micro projectile gene transfer, and gene gun), electroporation, sonoporation, photoporation, magnetofection, hydroporation, and mechanical massage. Chemical methods are those methods that employ a chemical to elicit a change in the cells membrane permeability or other characteristic(s) to facilitate entry of the vector into the cell. For example, the environmental pH can be altered which can elicit a change in the permeability of the cell membrane. Biological methods are those that rely and capitalize on the host cell's biological

processes or biological characteristics to facilitate transport of the vector (with or without a carrier) into a cell. For example, the vector and/or its carrier can stimulate an endocytosis or similar process in the cell to facilitate uptake of the vector into the cell. These and other methods are further described elsewhere herein.

[0443] Delivery of Type VII system components (e.g. polynucleotides encoding Type VII systems and/or components thereof polypeptides) to cells via particles. In some aspects, any of the of the Type VII systems and/or components thereof system components (e.g. polypeptides, polynucleotides, vectors and combinations thereof described herein) can be attached to, coupled to, integrated with, otherwise associated with one or more particles or component thereof as described herein. The particles described herein can then be administered to a cell or organism by an appropriate route and/or technique. In some aspects, particle delivery can be selected and be advantageous for delivery of the polynucleotide or vector components. It will be appreciated that in aspects, particle delivery can also be advantageous for other Type VII systems and/or components thereof molecules and formulations described elsewhere herein.

Packaging of CRISPR-Cas System Molecules

[0444] Methods and techniques to package or incorporate Type VII CRISRP-Cas system coding molecules (including, but not limited to, Type VII CRISRP-Cas system effector coding molecules), e.g., DNA, into vectors, e.g., viral vectors, include, but are not limited to: (a) those configured to achieve NHEJ-mediate gene knockout, which can be single virus vectors or double virus vector systems in some embodiments. Single virus vectors can contain, in some embodiments, two or more expression cassettes, where the expression cassettes each can be or include one or more of the following: (i) Promoter-Type VII effector coding nucleic acid molecule -terminator; (ii) Promoter-gRNA1-terminator; (iii) Promoter-gRNA2-terminator; (iv) Promoter-gRNA(N)-terminator, where the number and/or content of the cassette(s) can be up to the size limit of the vector they are contained in. Double virus vectors can include a first vector (Vector 1) that can contain one expression cassette for driving the expression of Type VII effector or Promoter-Type VII effector-coding nucleic acid molecule-terminator and a second vector (Vector 2), which can include one more expression cassettes for driving the expression of one or more gRNAs, where each cassette be or include one or more of the following: (i) Promoter-gRNA1-terminator and/or (ii) Promoter-gRNA(N)-terminator, where the number and/or content of the cassette(s) can be up to the size limit of the vector they are contained in.

[0445] To mediate homology-directed repair, in addition to the single and double virus vector approaches described above, an additional vector can be used to deliver a homology-direct repair template.

FORMULATIONS AND DELIVERY

Pharmaceutical Formulations

[0446] Also described 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 CRISPR-Cas systems or components thereof, compounds, molecules, compositions, vectors, vector systems, cells, or a combination thereof (which are also referred to as the primary active agent or ingredient elsewhere herein) described in greater detail elsewhere herein a pharmaceutically acceptable carrier. 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 VII CRISPR-Cas system and/or component(s) thereof, a vector or vector system containing the Type VII CRISPR-Cas system and/or component(s) thereof, a cell modified by the Type VII CRISPR-Cas system and/or component(s) thereof, a cell containing the Type VII CRISPR-Cas system and/or component(s) thereof, a cell capable of producing particles containing the Type VII CRISPR-Cas system and/or component(s) thereof, particle and other delivery compositions containing or otherwise incorporating or associating with the Type VII CRISPR-Cas system and/or component(s) thereof, and combinations thereof.

[0447] The pharmaceutical formulations described herein can be administered 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, intracoronal (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, intratym panic, 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).

[0448] 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 described 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, glucaronate, saccharate, formate, benzoate, glutamate, methanesulfonate, ethanesulfonate, benzenesulfonate, p-toluenesulfonate, camphorsulfonate, napthalenesulfonate, propionate, malonate, mandelate, malate, phthalate, and pamoate.

[0449] In some embodiments, it is desirable to modify the genome, epigenome, transcriptome, or other polynucleotide in the subject in need thereof. In some embodiments subject is in need of an improvement in one or more traits or characteristics, modification, modulation, addition, and/or deletion of one or more nucleotides in a polynucleotide or of polynucleotides. In some embodiments, the subject in need thereof is in need of treatment, prevention, or both, of a disease, infection, and/or condition. In some embodiments, the subject in need thereof has or is suspected of having a genetic or epigenetic disease or a symptom thereof. As used herein, “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. 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 Auxiliary Ingredients and Agents

[0450] 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.

[0451] The pharmaceutical formulations can be sterilized, and if desired, mixed with auxiliary 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.

[0452] In some embodiments, the pharmaceutical formulation can also include an effective amount of auxiliary active agents, including but not limited to, biologic agents or molecules (including but not limited to (e.g. polypeptides, polynucleotides, antibodies and fragments thereof, aptamers, and the like), chemotherapeutics, antineoplasic agents, hormones, antibiotics, antivirals, immunomodulating agents, antinausea, pain modifying compounds (such as opiates), anti-inflammatory agents, antipyretics, antibiotics, and combinations thereof.

Effective Amounts

[0453] In some embodiments, the amount of the primary active agent and/or optional auxiliary active agent can be an effective amount, least effective amount, and/or therapeutically effective amount. The effective amount, least effective amount, and/or therapeutically effective amount of the primary and optional auxiliary active agent described elsewhere herein contained in the pharmaceutical formulation can range 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 with any of these ranges. 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 range 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, μM, mM, or M or be any numerical value with any of these ranges.

[0454] In other embodiments, the effective amount, least effective amount, and/or therapeutically effective amount of the auxiliary active agent can range 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 with any of these ranges.

[0455] In some embodiments, a primary active agent can be present in the pharmaceutical formulation can range 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.

[0456] In some embodiments, the auxiliary active agent, when optionally present, can range 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.

[0457] In some embodiments where a cell population is delivered, the effective amount of cells can range from about $1X10^1$/mL to $1X10^{20}$/mL or more, such as about $1X10^1$/mL, $1X10^2$/mL, $1X10^3$/mL, $1X10^4$/mL, $1X10^5$/mL, $1X10^6$/mL, $1X10^7$/mL, $1X10^8$/mL, $1X10^9$/mL, $1X10^{10}$/mL, $1X10^{11}$/mL, $1X10^{12}$/mL, $1X10^{13}$/mL, $1X10^{14}$/mL, $1X10^{15}$/mL, $1X10^{16}$/mL, $1X10^{17}$/mL, $1X10^{18}$/mL, $1X10^{19}$/mL, to/or about $1X10^{20}$/mL.

[0458] In embodiments where there is an auxiliary active agent contained in the pharmaceutical formulation, the effective amount of the auxiliary active agent will vary depending on the auxiliary active agent.

[0459] When optionally present in the pharmaceutical formulation, the auxiliary 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. In yet other embodiments, the effective amount of the auxiliary active agent can range 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 auxiliary active agent pharmaceutical formulation. In additional embodiments, the effective amount of the auxiliary active agent can range 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.

[0460] Dosage Forms

[0461] In some embodiments, the pharmaceutical formulations described herein can be 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. In some cases, the dosage form contains a greater amount of the active ingredient 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.

[0462] 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.

[0463] 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 described herein can be microencapsulated.

[0464] 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 described 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 auxiliary 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.

[0465] 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.

[0466] 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.

[0467] Where appropriate, the dosage forms described herein can be a liposome. In these embodiments, primary active ingredient(s), and/or optional auxiliary 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.

[0468] 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 auxiliary 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 auxiliary 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.

[0469] 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 auxiliary 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 auxiliary) 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.

[0470] 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, auxiliary 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.

[0471] 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 auxiliary 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, or 3 doses are delivered each time. The aerosol formulations can be administered to a subject in need thereof.

[0472] 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 auxiliary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate, such a dosage form can contain a powder base such as lactose, glucose, trehalose, manitol, and/or starch. In some of these embodiments, a primary active agent, auxiliary 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 described herein.

[0473] 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.

[0474] 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.

[0475] For some embodiments, the dosage form contains a predetermined amount of a primary active agent, auxiliary active ingredient, and/or pharmaceutically acceptable salt thereof where appropriate per unit dose. In an embodiment, the predetermined amount of primary active agent, auxiliary 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 some embodiments, the predetermined amount can be effective to modify one or more polynucleotides, such as in one or more cells in/or of a subject in need thereof. Modification can occur *in vitro*, *ex vivo*, and/or *in vivo*.

[0476] In other embodiments, the predetermined amount of a primary active agent, auxiliary active agent, and/or pharmaceutically acceptable salt thereof where appropriate, 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.

Delivery

[0477] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding components of a nucleic acid-targeting system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, Science 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada

et al., in Current Topics in Microbiology and Immunology, Doerfler and Böhm (eds) (1995); and Yu et al., Gene Therapy 1:13-26 (1994).

[0478] Methods of non-viral delivery of nucleic acids include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, and International Patent Publication Nos. WO 91/17424 and WO 91/16024. Delivery can be to cells (e.g. *in vitro* or *ex vivo* administration) or target tissues (e.g. *in vivo* administration). In some embodiments, cells are transfected. In some embodiments, cells are transduced.

[0479] Physical methods for introducing a polynucleotide into a host cell include calcium phosphate precipitation, lipofection, particle bombardment, microinjection, electroporation, and the like. Methods for producing cells comprising vectors and/or exogenous nucleic acids are well-known in the art. See, for example, Sambrook et al., 2012, MOLECULAR CLONING: A LABORATORY MANUAL, volumes 1-4, Cold Spring Harbor Press, NY). Nucleic acids can be introduced into target cells using commercially available methods which include electroporation (Amaxa Nucleofector-II (Amaxa Biosystems, Cologne, Germany)), (ECM 830 (BTX) (Harvard Instruments, Boston, Mass.) or the Gene Pulser II (BioRad, Denver, Colo.), Multiporator (Eppendort, Hamburg Germany). Nucleic acids can also be introduced into cells using cationic liposome mediated transfection using lipofection, using polymer encapsulation, using peptide mediated transfection, or using biolistic particle delivery systems such as "gene guns" (see, for example, Nishikawa, et al. Hum Gene Ther., 12(8):861-70 (2001).

[0480] Biological methods for introducing a polynucleotide of interest into a host cell include the use of DNA and RNA vectors. RNA vectors include vectors having an RNA promoter and/other relevant domains for production of a RNA transcript. Viral vectors, and especially retroviral vectors, have become the most widely used method for inserting genes into mammalian, e.g., human cells. Other viral vectors may be derived from lentivirus, poxviruses, herpes simplex virus, adenoviruses and adeno-associated viruses, and the like. See, for example, U.S. Pat. Nos. 5,350,674 and 5,585,362.

[0481] Chemical means for introducing a polynucleotide into a host cell include colloidal dispersion systems, such as macromolecule complexes, nanocapsules, microspheres, beads, and lipid-based systems including oil-in-water emulsions, micelles, mixed micelles, and liposomes. An exemplary colloidal system for use as a delivery vehicle *in vitro* and *in vivo* is a liposome (e.g., an artificial membrane vesicle).

[0482] Expression and/or delivery of polynucleotides and/or polypeptides can be confirmed by any suitable method including, but not limited to, molecular biological assays, biochemical assays, enzymatic-based assays, sequencing-based methods, CRISPR-based screening assays, PCR-based methods, immunodetection-based methods, chromatography-based methods, flow-cytometry based methods, separation-based techniques, and any other technology well known and available in the art.

[0483] In one embodiment, one or more of the nucleic acid sequences described elsewhere herein are introduced by a method selected from the group consisting of transducing the population of cells, transfecting the population of cells, and electroporating the population of cells. In one embodiment, a population of cells comprises one or more of the nucleic acid sequences described herein.

[0484] Further delivery methods and details are discussed below and elsewhere herein.

Vector Delivery

[0485] Plasmid delivery involves the cloning of a guide RNA into a CRISPR effector protein expressing plasmid and transfecting the DNA in cell culture. Plasmid backbones are available commercially and no specific equipment is required. They have the advantage of being modular, capable of carrying different sizes of CRISPR effector coding sequences (including those encoding larger sized proteins) as well as selection markers. Both an advantage of plasmids is that they can ensure transient, but sustained expression. However, delivery of plasmids is not straightforward such that *in vivo* efficiency is often low. The sustained expression can also be disadvantageous in that it can increase off-target editing. In addition, excess build-up of the CRISPR effector protein can be toxic to the cells. Finally, plasmids have the risk of random integration of the dsDNA in the host genome, more particularly in view of the double-stranded breaks being generated (on and off-target). Plasmids are discussed in greater detail elsewhere herein.

[0486] In some embodiments, the Type VII CRISPR-Cas system and/or component thereof described herein can be delivered via a viral-based system. Viral vectors are described in

greater detail elsewhere herein. The use of RNA or DNA viral based systems for the delivery of nucleic acids takes advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (*in vivo*) or they can be used to treat cells *in vitro*, and the modified cells may optionally be administered to patients (*ex vivo*). Conventional viral based systems can include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer, which are described in greater detail elsewhere herein. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, and can in some embodiments, result in long-term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0487] Selection of a retroviral gene transfer system would therefore depend on the target tissue. In some embodiments, the retroviral delivery system is composed of cis-acting long terminal repeats and can be capable of packaging and delivering up to about 6 kb to about 10 kb of cargo (e.g. foreign polynucleotide to be delivered).

[0488] In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. Suitable adenoviral based vectors are described in greater detail elsewhere herein. These vectors can be produced in large quantities. In some embodiments, the adenoviral based vectors can be an adeno-associated virus ("AAV") vector. AAV vectors can be used to transduce cells with target nucleic acids, e.g., in the *in vitro* production of nucleic acids and peptides, and for *in vivo* and *ex vivo* gene therapy procedures (see, e.g., West et al., Virology 160:38-47 (1987); U.S. Pat. No. 4,797,368; International Patent Publication No. WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989).

[0489] Type VII effector and one or more guide RNA can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other plasmid or viral vector types, in particular, using formulations and doses from, for example, US Patents Nos. 8,454,972

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

[0490] Vector delivery, e.g., plasmid, viral delivery: The CRISPR enzyme, for instance a Type VII effector, and/or any of the present RNAs, for instance a guide RNA, can be delivered using any suitable vector, e.g., plasmid or viral vectors, such as adeno associated virus (AAV), lentivirus, adenovirus or other viral vector types, or combinations thereof. Type VII effector and one or more guide RNAs can be packaged into one or more vectors, e.g., plasmid or viral vectors. In some embodiments, the vector, e.g., plasmid or viral vector is delivered to the tissue of interest by, for example, an intramuscular injection, while other times the delivery is via intravenous, transdermal, intranasal, oral, mucosal, or other delivery methods. Such delivery may be either via a single dose, or multiple doses. One skilled in the art understands that the actual dosage to be delivered herein may vary greatly depending upon a variety of factors, such as the vector choice, the target cell, organism, or tissue, the general condition of the subject to be treated, the degree of transformation/modification sought, the administration route, the administration mode, the type of transformation/modification sought, etc.

[0491] Delivery via a retrovirus can result in long-term expression of the delivered exogenous polynucleotide, which can be a transgene in some cases. In a preferred embodiment, delivery can be by a lentivirus. Zou et al. administered about 10 µl of a recombinant lentivirus having a titer of 1 x 10^9 transducing units (TU)/ml by an intrathecal catheter. These sort of dosages can be adapted or extrapolated to use of a retroviral or lentiviral vector in the present invention.

[0492] In some embodiments, vector delivery includes delivery by a plasmid vector, viral vector, retrotransposon (e.g. piggyback, sleeping beauty), site directed insertion vector (e.g. CRISPR, zn finger nucleases, TALEN), or suicide expression vector, or other known vector in the art.

[0493] In some embodiments, delivery is by a third-generation lentiviral vector system.

[0494] In some embodiments, the vector delivery system is approved for use in human cells.

[0495] In some embodiments, delivery can be by an RNA vector.

Nucleic Acid delivery

[0496] In certain example embodiments, the CRISPR effector protein described herein (e.g. a Type VII effector protein) may be delivered using a nucleic acid molecule encoding the CRISPR effector protein. The nucleic acid molecule encoding a CRISPR effector protein described herein (e.g. a Type VII effector protein), may advantageously be a codon optimized CRISPR effector protein. An example of a codon optimized sequence, is in this instance a sequence optimized for expression in eukaryote, e.g., humans (i.e. being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed; see, e.g., SaCas9 human codon optimized sequence in International Patent Publication No. WO 2014/093622 (PCT/US2013/074667). Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species other than human, or for codon optimization for specific organs is known. In some embodiments, an enzyme coding sequence encoding a CRISPR effector protein is a 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 some embodiments, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals

which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g. about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at kazusa.orjp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. In some embodiments, one or more codons (e.g. 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a Cas correspond to the most frequently used codon for a particular amino acid.

[0497] In some embodiments it is envisaged to introduce the RNA and/or protein directly to the host cell. For instance, the CRISPR effector can be delivered as CRISPR effector-encoding mRNA together with an *in vitro* transcribed guide RNA. Such methods can reduce the time to ensure effect of the CRISPR effector protein and further prevents long-term expression of the CRISPR system components.

[0498] In some embodiments the RNA molecules of the invention are delivered in liposome or lipofectin formulations and the like and can be prepared by methods well known to those skilled in the art. Such methods are described, for example, in U.S. Pat. Nos. 5,593,972, 5,589,466, and 5,580,859, which are herein incorporated by reference. Delivery systems aimed specifically at the enhanced and improved delivery of siRNA into mammalian cells have been

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

[0499] Indeed, RNA delivery is a useful method of *in vivo* delivery. It is possible to deliver Type VII effector and gRNA (and, for instance, HR repair template) into cells using liposomes or nanoparticles. Thus, delivery of the CRISPR enzyme, such as a Type VII effector and/or delivery of the RNAs of the invention may be in RNA form and via microvesicles, liposomes or particle or particles. For example, Type VII effector mRNA and gRNA can be packaged into liposomal particles for delivery in vivo. Liposomal transfection reagents such as lipofectamine from Life Technologies and other reagents on the market can effectively deliver RNA molecules into the liver.

[0500] Delivery of RNA (or DNA) can include delivery via particles. (Cho, S., Goldberg, M., Son, S., Xu, Q., Yang, F., Mei, Y., Bogatyrev, S., Langer, R. and Anderson, D., Lipid-like nanoparticles for small interfering RNA delivery to endothelial cells, Advanced Functional Materials, 19: 3112-3118, 2010) or exosomes (Schroeder, A., Levins, C., Cortez, C., Langer, R., and Anderson, D., Lipid-based nanotherapeutics for siRNA delivery, Journal of Internal Medicine, 267: 9-21, 2010, PMID: 20059641). Indeed, exosomes have been shown to be particularly useful in delivery siRNA, a system with some parallels to the CRISPR system. For instance, El-Andaloussi S, et al. ("Exosome-mediated delivery of siRNA *in vitro* and *in vivo*." Nat Protoc. 2012 Dec;7(12):2112-26. doi: 10.1038/nprot.2012.131. Epub 2012 Nov 15.) describe how exosomes are promising tools for drug delivery across different biological barriers and can be harnessed for delivery of siRNA *in vitro* and *in vivo*. Their approach is to generate targeted exosomes through transfection of an expression vector, comprising an exosomal protein fused with a peptide ligand. The exosomes are then purified and characterized from transfected cell supernatant, then RNA is loaded into the exosomes. Delivery or administration according to the invention can be performed with exosomes, in particular but not limited to the brain. Vitamin E (α-tocopherol) may be conjugated with CRISPR Cas and delivered to the brain along with high density lipoprotein (HDL), for example in a similar manner as was done by Uno et al. (HUMAN GENE THERAPY 22:711–719 (June 2011)) for

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

[0501] In particular embodiments, RNA-based delivery is used. In these embodiments, mRNA of the CRISPR effector protein is delivered together with *in vitro* transcribed guide RNA. Liang et al. describes efficient genome editing using RNA based delivery (Protein Cell. 2015 May; 6(5): 363–372).

[0502] RNA delivery: The CRISPR enzyme, for instance a Type VII effector, and/or any of the present RNAs, for instance a guide RNA, can also be delivered in the form of RNA. Type VII effector mRNA can be generated using *in vitro* transcription. For example, Type VII effector mRNA can be synthesized using a PCR cassette containing the following elements: T7_promoter-kozak sequence (GCCACC)-Type VII effector-3' UTR from beta globin-polyA tail (a string of 120 or more adenines). The cassette can be used for transcription by T7 polymerase. Guide RNAs can also be transcribed using *in vitro* transcription from a cassette containing T7_promoter-GG-guide RNA sequence.

[0503] To enhance expression and reduce possible toxicity, the CRISPR enzyme-coding sequence and/or the guide RNA can be modified to include one or more modified nucleoside e.g. using pseudo-U or 5-Methyl-C.

[0504] mRNA delivery methods are especially promising for liver delivery currently.

[0505] Much clinical work on RNA delivery has focused on RNAi or antisense, but these systems can be adapted for delivery of RNA for implementing the present invention. References below to RNAi etc. should be read accordingly.

[0506] CRISPR enzyme mRNA and guide RNA might also be delivered separately. CRISPR enzyme mRNA can be delivered prior to the guide RNA to give time for CRISPR enzyme to be expressed. CRISPR enzyme mRNA might be administered 1-12 hours (preferably around 2-6 hours) prior to the administration of guide RNA.

[0507] Alternatively, CRISPR enzyme mRNA and guide RNA can be administered together. Advantageously, a second booster dose of guide RNA can be administered 1-12 hours (preferably around 2-6 hours) after the initial administration of CRISPR enzyme mRNA + guide RNA.

[0508] In one embodiment, the nucleic acids introduced into the cell are RNA. In another embodiment, the RNA is mRNA that comprises *in vitro* transcribed RNA or synthetic RNA. The RNA is produced by *in vitro* transcription using a polymerase chain reaction (PCR)-generated template. DNA of interest from any source can be directly converted by PCR into a template for *in vitro* mRNA synthesis using appropriate primers and RNA polymerase. The source of the DNA can be, for example, genomic DNA, plasmid DNA, phage DNA, cDNA, synthetic DNA sequence or any other appropriate source of DNA. The desired template for *in vitro* transcription is a chimeric intracellular signaling molecule and/or a bispecific antibody.

[0509] PCR can be used to generate a template for *in vitro* transcription of mRNA which is then introduced into cells. Methods for performing PCR are well known in the art. Primers for use in PCR are designed to have regions that are substantially complementary to regions of the DNA to be used as a template for the PCR. "Substantially complementary", as used herein, refers to sequences of nucleotides where a majority or all of the bases in the primer sequence are complementary, or one or more bases are non-complementary, or mismatched. Substantially complementary sequences are able to anneal or hybridize with the intended DNA target under annealing conditions used for PCR. The primers can be designed to be substantially complementary to any portion of the DNA template. For example, the primers can be designed to amplify the portion of a gene that is normally transcribed in cells (the open reading frame), including 5′ and 3′ UTRs. The primers can also be designed to amplify a portion of a gene that encodes a particular domain of interest. In one embodiment, the primers are

designed to amplify the coding region of a human cDNA, including all or portions of the 5′ and 3′ UTRs. Primers useful for PCR are generated by synthetic methods that are well known in the art. "Forward primers" are primers that contain a region of nucleotides that are substantially complementary to nucleotides on the DNA template that are upstream of the DNA sequence that is to be amplified. "Upstream" is used herein to refer to a location 5, to the DNA sequence to be amplified relative to the coding strand. "Reverse primers" are primers that contain a region of nucleotides that are substantially complementary to a double-stranded DNA template that are downstream of the DNA sequence that is to be amplified. "Downstream" is used herein to refer to a location 3′ to the DNA sequence to be amplified relative to the coding strand.

[0510] Chemical structures that have the ability to promote stability and/or translation efficiency of the RNA may also be used. The RNA preferably has 5′ and 3′ UTRs. In one embodiment, the 5′ UTR is between zero and 3000 nucleotides in length. The length of 5′ and 3′ UTR sequences to be added to the coding region can be altered by different methods, including, but not limited to, designing primers for PCR that anneal to different regions of the UTRs. Using this approach, one of ordinary skill in the art can modify the 5′ and 3′ UTR lengths required to achieve optimal translation efficiency following transfection of the transcribed RNA.

[0511] The 5′ and 3′ UTRs can be the naturally occurring, endogenous 5′ and 3′ UTRs for the gene of interest. Alternatively, UTR sequences that are not endogenous to the gene of interest can be added by incorporating the UTR sequences into the forward and reverse primers or by any other modifications of the template. The use of UTR sequences that are not endogenous to the gene of interest can be useful for modifying the stability and/or translation efficiency of the RNA. For example, it is known that AU-rich elements in 3′ UTR sequences can decrease the stability of mRNA. Therefore, 3′ UTRs can be selected or designed to increase the stability of the transcribed RNA based on properties of UTRs that are well known in the art.

[0512] In one embodiment, the 5′ UTR can contain the Kozak sequence of the endogenous gene. Alternatively, when a 5′ UTR that is not endogenous to the gene of interest is being added by PCR as described above, a consensus Kozak sequence can be redesigned by adding the 5′ UTR sequence. Kozak sequences can increase the efficiency of translation of some RNA transcripts, but does not appear to be required for all RNAs to enable efficient translation. The requirement for Kozak sequences for many mRNAs is known in the art. In other embodiments

the 5′ UTR can be derived from an RNA virus whose RNA genome is stable in cells. In other embodiments various nucleotide analogues can be used in the 3′ or 5′ UTR to impede exonuclease degradation of the mRNA.

[0513] To enable synthesis of RNA from a DNA template without the need for gene cloning, a promoter of transcription should be attached to the DNA template upstream of the sequence to be transcribed. When a sequence that functions as a promoter for an RNA polymerase is added to the 5′ end of the forward primer, the RNA polymerase promoter becomes incorporated into the PCR product upstream of the open reading frame that is to be transcribed. In one embodiment, the promoter is a T7 polymerase promoter, as described elsewhere herein. Other useful promoters include, but are not limited to, T3 and SP6 RNA polymerase promoters. Consensus nucleotide sequences for T7, T3 and SP6 promoters are known in the art.

[0514] In one embodiment, the mRNA has both a cap on the 5′ end and a 3′ poly(A) tail which determine ribosome binding, initiation of translation and stability mRNA in the cell. On a circular DNA template, for instance, plasmid DNA, RNA polymerase produces a long concatameric product which is not suitable for expression in eukaryotic cells. The transcription of plasmid DNA linearized at the end of the 3′ UTR results in normal sized mRNA which is not effective in eukaryotic transfection even if it is polyadenylated after transcription.

[0515] On a linear DNA template, phage T7 RNA polymerase can extend the 3′ end of the transcript beyond the last base of the template (Schenborn and Mierendorf, Nuc Acids Res., 13:6223-36 (1985); Nacheva and Berzal-Herranz, Eur. J. Biochem., 270:1485-65 (2003).

[0516] The conventional method of integration of polyA/T stretches into a DNA template is molecular cloning. However polyA/T sequence integrated into plasmid DNA can cause plasmid instability, which is why plasmid DNA templates obtained from bacterial cells are often highly contaminated with deletions and other aberrations. This makes cloning procedures not only laborious and time consuming but often not reliable. That is why a method which allows construction of DNA templates with polyA/T 3′ stretch without cloning highly desirable.

[0517] The polyA/T segment of the transcriptional DNA template can be produced during PCR by using a reverse primer containing a polyT tail, such as 100T tail (size can be 50-5000 T), or after PCR by any other method, including, but not limited to, DNA ligation or *in vitro* recombination. Poly(A) tails also provide stability to RNAs and reduce their degradation.

Generally, the length of a poly(A) tail positively correlates with the stability of the transcribed RNA. In one embodiment, the poly(A) tail is between 100 and 5000 adenosines.

[0518] Poly(A) tails of RNAs can be further extended following *in vitro* transcription with the use of a poly(A) polymerase, such as *E. coli* polyA polymerase (E-PAP). In one embodiment, increasing the length of a poly(A) tail from 100 nucleotides to between 300 and 400 nucleotides results in about a two-fold increase in the translation efficiency of the RNA. Additionally, the attachment of different chemical groups to the 3′ end can increase mRNA stability. Such attachment can contain modified/artificial nucleotides, aptamers and other compounds. For example, ATP analogs can be incorporated into the poly(A) tail using poly(A) polymerase. ATP analogs can further increase the stability of the RNA.

[0519] 5′ caps also provide stability to RNA molecules. In a preferred embodiment, RNAs produced by the methods disclosed herein include a 5′ cap. The 5′ cap is provided using techniques known in the art and described herein (Cougot, et al., Trends in Biochem. Sci., 29:436-444 (2001); Stepinski, et al., RNA, 7:1468-95 (2001); Elango, et al., Biochim. Biophys. Res. Commun., 330:958-966 (2005)).

[0520] The RNAs produced by the methods disclosed herein can also contain an internal ribosome entry site (IRES) sequence. The IRES sequence may be any viral, chromosomal or artificially designed sequence which initiates cap-independent ribosome binding to mRNA and facilitates the initiation of translation. Any solutes suitable for cell electroporation, which can contain factors facilitating cellular permeability and viability such as sugars, peptides, lipids, proteins, antioxidants, and surfactants can be included.

[0521] Some *in vitro*-transcribed RNA (IVT-RNA) vectors are known in the literature which are utilized in a standardized manner as template for *in vitro* transcription and which have been genetically modified in such a way that stabilized RNA transcripts are produced. Currently protocols used in the art are based on a plasmid vector with the following structure: a 5′ RNA polymerase promoter enabling RNA transcription, followed by a gene of interest which is flanked either 3′ and/or 5′ by untranslated regions (UTR), and a 3′ polyadenyl cassette containing 50-70 A nucleotides. Prior to *in vitro* transcription, the circular plasmid is linearized downstream of the polyadenyl cassette by type II restriction enzymes (recognition sequence corresponds to cleavage site). The polyadenyl cassette thus corresponds to the later poly(A) sequence in the transcript. As a result of this procedure, some nucleotides remain as part of the enzyme cleavage site after linearization and extend or mask the poly(A) sequence at the 3′ end.

It is not clear, whether this nonphysiological overhang affects the amount of protein produced intracellularly from such a construct.

[0522] In one aspect, the RNA construct is delivered into the cells by electroporation. See, e.g., the formulations and methodology of electroporation of nucleic acid constructs into mammalian cells as taught in US 2004/0014645, US 2005/0052630A1, US 2005/0070841A1, US 2004/0059285A1, US 2004/0092907A1. The various parameters including electric field strength required for electroporation of any known cell type are generally known in the relevant research literature as well as numerous patents and applications in the field. See e.g., U.S. Pat. No. 6,678,556, U.S. Pat. No. 7,171,264, and U.S. Pat. No. 7,173,116. Apparatus for therapeutic application of electroporation are available commercially, e.g., the MedPulser™ DNA Electroporation Therapy System (Inovio/Genetronics, San Diego, Calif.), and are described in patents such as U.S. Pat. No. 6,567,694; U.S. Pat. No. 6,516,223, U.S. Pat. No. 5,993,434, U.S. Pat. No. 6,181,964, U.S. Pat. No. 6,241,701, and U.S. Pat. No. 6,233,482; electroporation may also be used for transfection of cells *in vitro* as described e.g. in US20070128708A1. Electroporation may also be utilized to deliver nucleic acids into cells *in vitro*. Accordingly, electroporation-mediated administration into cells of nucleic acids including expression constructs utilizing any of the many available devices and electroporation systems known to those of skill in the art presents an exciting new means for delivering an RNA of interest to a target cell.

RNP

[0523] In particular embodiments, pre-complexed guide RNA and CRISPR effector protein are delivered as a ribonucleoprotein (RNP). RNPs have the advantage that they lead to rapid editing effects even more so than the RNA method because this process avoids the need for transcription. RNPs have the advantage that they lead to rapid editing effects even more so than the RNA method because this process avoids the need for transcription. Other advantages include that RNP delivery is transient, reducing off-target effects and toxicity issues. Efficient genome editing in different cell types has been observed by Kim et al. (2014, Genome Res. 24(6):1012-9), Paix et al. (2015, Genetics 204(1):47-54), Chu et al. (2016, BMC Biotechnol. 16:4), and Wang et al. (2013, Cell. 9;153(4):910-8). In some embodiments, CRISPR protein or adaptor is optionally fused to, linked, or otherwise associated with an adenosine deaminase or other deaminase.

[0524] In particular embodiments, the ribonucleoprotein is delivered by way of a polypeptide-based shuttle agent as described in WO2016161516. WO2016161516 describes efficient transduction of polypeptide cargos 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. Similarly, these polypeptides can be used for the delivery of CRISPR-effector based RNPs in eukaryotic cells.

[0525] Indeed, RNA delivery is a useful method of *in vivo* delivery. It is possible to deliver Type VII effector and gRNA (and, for instance, HR repair template) into cells using liposomes or particles. Thus delivery of the CRISPR enzyme, such as a Type VII effector and/or delivery of the RNAs of the invention may be in RNA form and via microvesicles, liposomes or particles . For example, Type VII effector mRNA and gRNA can be packaged into liposomal particles for delivery *in vivo*. Liposomal transfection reagents such as lipofectamine from Life Technologies and other reagents on the market can effectively deliver RNA molecules into the liver.

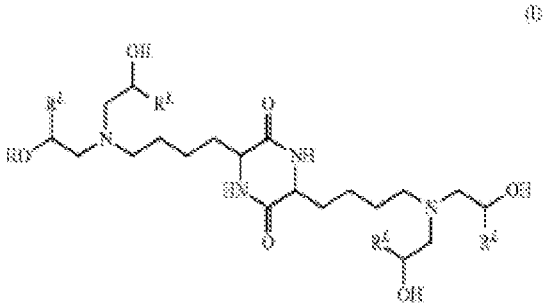
[0526] Means of delivery of RNA also preferred include delivery of RNA via nanoparticles (Cho, S., Goldberg, M., Son, S., Xu, Q., Yang, F., Mei, Y., Bogatyrev, S., Langer, R. and Anderson, D., Lipid-like nanoparticles for small interfering RNA delivery to endothelial cells, Advanced Functional Materials, 19: 3112-3118, 2010) or exosomes (Schroeder, A., Levins, C., Cortez, C., Langer, R., and Anderson, D., Lipid-based nanotherapeutics for siRNA delivery, Journal of Internal Medicine, 267: 9-21, 2010, PMID: 20059641). Indeed, exosomes have been shown to be particularly useful in delivery siRNA, a system with some parallels to the CRISPR system. For instance, El-Andaloussi S, et al. ("Exosome-mediated delivery of siRNA *in vitro* and *in vivo*." Nat Protoc. 2012 Dec;7(12):2112-26. doi: 10.1038/nprot.2012.131. Epub 2012 Nov 15.) describe how exosomes are promising tools for drug delivery across different biological barriers and can be harnessed for delivery of siRNA *in vitro* and *in vivo*. Their approach is to generate targeted exosomes through transfection of an expression vector, comprising an exosomal protein fused with a peptide ligand. The exosomes are then purified and characterized from transfected cell supernatant, then RNA is loaded into the exosomes. Delivery or administration according to the invention can be performed with exosomes, in particular but not limited to the brain. Vitamin E (α-tocopherol) may be conjugated with CRISPR Cas and delivered to the brain along with high density lipoprotein (HDL), for example in a similar manner as was done by Uno et al. (HUMAN GENE THERAPY 22:711–719 (June

2011)) for delivering short-interfering RNA (siRNA) to the brain. Mice were infused via Osmotic minipumps (model 1007D; Alzet, Cupertino, CA) filled with phosphate-buffered saline (PBS) or free TocsiBACE or Toc-siBACE/HDL and connected with Brain Infusion Kit 3 (Alzet). A brain-infusion cannula was placed about 0.5mm posterior to the bregma at midline for infusion into the dorsal third ventricle. Uno et al. found that as little as 3 nmol of Toc-siRNA with HDL could induce a target reduction in comparable degree by the same ICV infusion method. A similar dosage of CRISPR Cas conjugated to α-tocopherol and co-administered with HDL targeted to the brain may be contemplated for humans in the present invention, for example, about 3 nmol to about 3 μmol of CRISPR Cas targeted to the brain may be contemplated.

[0527] Zou et al. ((HUMAN GENE THERAPY 22:465-475 (April 2011)) describes a method of lentiviral-mediated delivery of short-hairpin RNAs targeting PKCγ for *in vivo* gene silencing in the spinal cord of rats. Zou et al. administered about 10 μl of a recombinant lentivirus having a titer of 1 x 109 transducing units (TU)/ml by an intrathecal catheter. A similar dosage of CRISPR Cas expressed in a lentiviral vector may be contemplated for humans in the present invention, for example, about 10-50 ml of CRISPR Cas in a lentivirus having a titer of 1 x 109 transducing units (TU)/ml may be contemplated. A similar dosage of CRISPR Cas expressed in a lentiviral vector targeted to the brain may be contemplated for humans in the present invention, for example, about 10-50 ml of CRISPR Cas targeted to the brain in a lentivirus having a titer of 1 x 109 transducing units (TU)/ml may be contemplated.

[0528] Anderson et al. (US 20170079916) provides a modified dendrimer nanoparticle for the delivery of therapeutic, prophylactic and/or diagnostic agents to a subject, comprising: one or more zero to seven generation alkylated dendrimers; one or more amphiphilic polymers; and one or more therapeutic, prophylactic and/or diagnostic agents encapsulated therein. One alkylated dendrimer may be selected from the group consisting of poly(ethyleneimine), poly(polyproylenimine), diaminobutane amine polypropylenimine tetramine and poly(amido amine). The therapeutic, prophylactic and diagnostic agent may be selected from the group consisting of proteins, peptides, carbohydrates, nucleic acids, lipids, small molecules and combinations thereof.

[0529] Anderson *et al.* (US 20160367686) provides a compound of Formula (I):



[0530] and salts thereof, wherein each instance of R L is independently optionally substituted C6-C40 alkenyl, and a composition for the delivery of an agent to a subject or cell comprising the compound , or a salt thereof; an agent; and optionally, an excipient. The agent may be an organic molecule, inorganic molecule, nucleic acid, protein, peptide, polynucleotide, targeting agent, an isotopically labeled chemical compound, vaccine, an immunological agent, or an agent useful in bioprocessing. The composition may further comprise cholesterol, a PEGylated lipid, a phospholipid, or an apolipoprotein.

[0531] Anderson et al. (US20150232883) provides a delivery particle formulations and/or systems, preferably nanoparticle delivery formulations and/or systems, comprising (a) a CRISPR-Cas system RNA polynucleotide sequence; or (b) Cas9; or (c) both a CRISPR-Cas system RNA polynucleotide sequence and Cas9; or (d) one or more vectors that contain nucleic acid molecule(s) encoding (a), (b) or (c), wherein the CRISPR-Cas system RNA polynucleotide sequence and the Cas9 do not naturally occur together. The delivery particle formulations may further comprise a surfactant, lipid or protein, wherein the surfactant may comprise a cationic lipid.

[0532] Anderson et al. (US20050123596) provides examples of microparticles that are designed to release their payload when exposed to acidic conditions, wherein the microparticles comprise at least one agent to be delivered, a pH triggering agent, and a polymer, wherein the polymer is selected from the group of polymethacrylates and polyacrylates.

[0533] Anderson et al (US 20020150626) provides lipid-protein-sugar particles for delivery of nucleic acids, wherein the polynucleotide is encapsulated in a lipid-protein-sugar matrix by contacting the polynucleotide with a lipid, a protein, and a sugar; and spray drying mixture of the polynucleotide, the lipid, the protein, and the sugar to make microparticles.

[0534] In terms of local delivery to the brain, this can be achieved in various ways. For instance, material can be delivered intrastriatally e.g. by injection. Injection can be performed stereotactically via a craniotomy.

[0535] Enhancing NHEJ or HR efficiency is also helpful for delivery. It is preferred that NHEJ efficiency is enhanced by co-expressing end-processing enzymes such as Trex2 (Dumitrache et al. Genetics. 2011 August; 188(4): 787–797). It is preferred that HR efficiency is increased by transiently inhibiting NHEJ machineries such as Ku70 and Ku86. HR efficiency can also be increased by co-expressing prokaryotic or eukaryotic homologous recombination enzymes such as RecBCD, RecA.

Particles

[0536] In some aspects or embodiments, a composition comprising a delivery particle formulation may be used. In some aspects or embodiments, the formulation comprises a CRISPR complex, the complex comprising a CRISPR protein and-a guide which directs sequence-specific binding of the CRISPR complex to a target sequence. In some embodiments, the delivery particle comprises a lipid-based particle, optionally a lipid nanoparticle, or cationic lipid and optionally biodegradable polymer. In some embodiments, the cationic lipid comprises 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP). In some embodiments, the hydrophilic polymer comprises ethylene glycol or polyethylene glycol. In some embodiments, the delivery particle further comprises a lipoprotein, preferably cholesterol. In some embodiments, the delivery particles are less than 500 nm in diameter, optionally less than 250 nm in diameter, optionally less than 100 nm in diameter, optionally about 35 nm to about 60 nm in diameter.

[0537] Several types of particle delivery systems and/or formulations are known to be useful in a diverse spectrum of biomedical applications. In general, a particle is defined as a small object that behaves as a whole unit with respect to its transport and properties. Particles are further classified according to diameter. Coarse particles cover a range between 2,500 and 10,000 nanometers. Fine particles are sized between 100 and 2,500 nanometers. Ultrafine particles, or nanoparticles, are generally between 1 and 100 nanometers in size. The basis of the 100-nm limit is the fact that novel properties that differentiate particles from the bulk material typically develop at a critical length scale of under 100 nm.

[0538] As used herein, a particle delivery system/formulation is defined as any biological delivery system/formulation which includes a particle in accordance with the present invention.

A particle in accordance with the present invention is any entity having a greatest dimension (e.g. diameter) of less than 100 microns (μm). In some embodiments, inventive particles have a greatest dimension of less than 10 μ m. In some embodiments, inventive particles have a greatest dimension of less than 2000 nanometers (nm). In some embodiments, inventive particles have a greatest dimension of less than 1000 nanometers (nm). In some embodiments, inventive particles have a greatest dimension of less than 900 nm, 800 nm, 700 nm, 600 nm, 500 nm, 400 nm, 300 nm, 200 nm, or 100 nm. Typically, inventive particles have a greatest dimension (e.g., diameter) of 500 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 250 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 200 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 150 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 100 nm or less. Smaller particles, e.g., having a greatest dimension of 50 nm or less are used in some embodiments of the invention. In some embodiments, inventive particles have a greatest dimension ranging between 25 nm and 200 nm.

[0539] In terms of this invention, it is preferred to have one or more components of CRISPR complex, e.g., CRISPR enzyme or mRNA or guide RNA delivered using nanoparticles or lipid envelopes. Other delivery systems or vectors are may be used in conjunction with the nanoparticle aspects of the invention.

[0540] In general, a "nanoparticle" refers to any particle having a diameter of less than 1000 nm. In certain preferred embodiments, nanoparticles of the invention have a greatest dimension (e.g., diameter) of 500 nm or less. In other preferred embodiments, nanoparticles of the invention have a greatest dimension ranging between 25 nm and 200 nm. In other preferred embodiments, nanoparticles of the invention have a greatest dimension of 100 nm or less. In other preferred embodiments, nanoparticles of the invention have a greatest dimension ranging between 35 nm and 60 nm. It will be appreciated that reference made herein to particles or nanoparticles can be interchangeable, where appropriate.

[0541] It will be understood that the size of the particle will differ depending as to whether it is measured before or after loading. Accordingly, in particular embodiments, the term “nanoparticles” may apply only to the particles pre-loading.

[0542] Nanoparticles encompassed in the present invention may be provided in different forms, e.g., as solid nanoparticles (e.g., metal such as silver, gold, iron, titanium), non-metal,

lipid-based solids, polymers), suspensions of nanoparticles, or combinations thereof. Metal, dielectric, and semiconductor nanoparticles may be prepared, as well as hybrid structures (e.g., core–shell nanoparticles). 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.

[0543] Semi-solid and soft nanoparticles have been manufactured, and are within the scope of the present invention. A prototype nanoparticle of semi-solid nature is the liposome. Various types of liposome nanoparticles are currently used clinically as delivery systems for anticancer drugs and vaccines. 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.

[0544] 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, concerning particles, methods of making and using them and measurements thereof.

[0545] Particles delivery systems within the scope of the present invention may be provided in any form, including but not limited to solid, semi-solid, emulsion, or colloidal

particles. As such any of the delivery systems described herein, including but not limited to, e.g., lipid-based systems, liposomes, micelles, microvesicles, exosomes, or gene gun may be provided as particle delivery systems within the scope of the present invention.

[0546] CRISPR enzyme mRNA and guide RNA may be delivered simultaneously using particles or lipid envelopes; for instance, CRISPR enzyme and RNA of the invention, e.g., as a complex, can be delivered via a particle as in Dahlman et al., WO2015089419 A2 and documents cited therein, such as 7C1 (see, e.g., James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84), e.g., delivery particle comprising lipid or lipidoid and hydrophilic polymer, e.g., cationic lipid and hydrophilic polymer, for instance wherein the cationic lipid comprises 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP) or 1,2-ditetradecanoyl-sn-glycero-3-phosphocholine (DMPC) and/or wherein the hydrophilic polymer comprises ethylene glycol or polyethylene glycol (PEG); and/or wherein the particle further comprises cholesterol (e.g., particle from formulation 1 = DOTAP 100, DMPC 0, PEG 0, Cholesterol 0; formulation number 2 = DOTAP 90, DMPC 0, PEG 10, Cholesterol 0; formulation number 3 = DOTAP 90, DMPC 0, PEG 5, Cholesterol 5), wherein particles are formed using an efficient, multistep process wherein first, effector protein and RNA are mixed together, e.g., at a 1:1 molar ratio, e.g., at room temperature, e.g., for 30 minutes, e.g., in sterile, nuclease free 1X PBS; and separately, DOTAP, DMPC, PEG, and cholesterol as applicable for the formulation are dissolved in alcohol, e.g., 100% ethanol; and, the two solutions are mixed together to form particles containing the complexes).

[0547] Nucleic acid-targeting effector protein mRNA and guide RNA may be delivered simultaneously using particles or lipid envelopes. Examples of suitable particles include but are not limited to those described in US 9,301,923.

[0548] For example, Su X, Fricke J, Kavanagh DG, Irvine DJ ("*In vitro* and *in vivo* mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles" Mol Pharm. 2011 Jun 6;8(3):774-87. doi: 10.1021/mp100390w. Epub 2011 Apr 1) describes biodegradable core-shell structured nanoparticles with a poly(β-amino ester) (PBAE) core enveloped by a phospholipid bilayer shell. These were developed for *in vivo* mRNA delivery. The pH-responsive PBAE component was chosen to promote endosome disruption, while the lipid surface layer was selected to minimize toxicity of the polycation core. Such are, therefore, preferred for delivering RNA of the present invention.

[0549] Liu et al. (US 20110212179) provides bimodal porous polymer microspheres comprising a base polymer, wherein the particle comprises macropores having a diameter ranging from about 20 to about 500 microns and micropores having a diameter ranging from about 1 to about 70 microns, and wherein the microspheres have a diameter ranging from about 50 to about 1100 microns.

[0550] Berg et al. (US20160174546) a nanolipid delivery system, in particular a nano-particle concentrate, comprising: a composition comprising a lipid, oil or solvent, the composition having a viscosity of less than 100 cP at 25 degrees C and a Kauri Butanol solvency of greater than 25 Kb; and at least one amphipathic compound selected from the group consisting of an alkoxylated lipid, an alkoxylated fatty acid, an alkoxylated alcohol, a heteroatomic hydrophilic lipid, a heteroatomic hydrophilic fatty acid, a heteroatomic hydrophilic alcohol, a diluent, and combinations thereof, wherein the compound is derived from a starting compound having a viscosity of less than 1000 cP at 50 degrees C, wherein the concentrate is configured to provide a stable nano emulsion having a D50 and a mean average particle size distribution of less than 100 nm when diluted.

[0551] Liu et al. (US 20140301951) provides a protocell nanostructure comprising: a porous particle core comprising a plurality of pores; and at least one lipid bilayer surrounding the porous particle core to form a protocell, wherein the protocell is capable of loading one or more cargo components to the plurality of pores of the porous particle core and releasing the one or more cargo components from the porous particle core across the surrounding lipid bilayer.

[0552] Chromy et al. (US 20150105538) provides methods and systems for assembling, solubilizing and/or purifying a membrane associated protein in a nanolipoprotein particle, which comprise a temperature transition cycle performed in presence of a detergent, wherein during the temperature transition cycle the nanolipoprotein components are brought to a temperature above and below the gel to liquid crystalling transition temperature of the membrane forming lipid of the nanolipoprotein particle.

[0553] Bader et al. (US 20150250725), provides a method for producing a lipid particle comprising the following: i) providing a first solution comprising denatured apolipoprotein, ii) adding the first solution to a second solution comprising at least two lipids and a detergent but no apolipoprotein, and iii) removing the detergent from the solution obtained in ii) and thereby producing a lipid particle.

[0554] Mirkin et al., (US20100129793) provides a method of preparing a composite particle comprising the steps of (a) admixing a dielectric component and a magnetic component to form a first intermediate, (b) admixing the first intermediate and gold seeds to form a second intermediate, and (c) forming a gold shell on the second intermediate by admixing the second intermediate with a gold source and a reducing agent to form said composite particle.

[0555] Other gold nanoparticle-based systems are also contemplated. In one embodiment, Cas9 RNP can be delivered in a vehicle composed of gold nanoparticles conjugated with DNA, which are complexed with donor DNA, Cas9 RNP, and the endosomal disruptive polymer PAsp(DET). *See, e.g.* Lee et al., "Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA *in vivo* induces homology-directed DNA repair" Nat Biomed Eng. 2017; 1: 889–901. In another embodiment, Cas9 RNP can be delivered into the cell cytoplasm and nucleus using arginine-functionalized gold nanoparticles. *See, e.g.* Mout et al., "Cytosolic and Nuclear Delivery of CRISPR/Cas9-ribonucleoprotein for Gene Editing Using Arginine Functionalized Gold Nanoparticles" Bio Protoc. 2017 Oct 20;7(20).

[0556] In one embodiment, particles/nanoparticles based on self-assembling bioadhesive polymers are contemplated, which may be applied to oral delivery of peptides, intravenous delivery of peptides and nasal delivery of peptides, all to the brain. Other embodiments, such as oral absorption and ocular delivery of hydrophobic drugs are also contemplated. The molecular envelope technology involves an engineered polymer envelope which is protected and delivered to the site of the disease (see, e.g., Mazza, M. et al. ACSNano, 2013. 7(2): 1016-1026; Siew, A., et al. Mol Pharm, 2012. 9(1):14-28; Lalatsa, A., et al. J Contr Rel, 2012. 161(2):523-36; Lalatsa, A., et al., Mol Pharm, 2012. 9(6):1665-80; Lalatsa, A., et al. Mol Pharm, 2012. 9(6):1764-74; Garrett, N.L., et al. J Biophotonics, 2012. 5(5-6):458-68; Garrett, N.L., et al. J Raman Spect, 2012. 43(5):681-688; Ahmad, S., et al. J Royal Soc Interface 2010. 7:S423-33; Uchegbu, I.F. Expert Opin Drug Deliv, 2006. 3(5):629-40; Qu, X.,et al. Biomacromolecules, 2006. 7(12):3452-9 and Uchegbu, I.F., et al. Int J Pharm, 2001. 224:185-199). Doses of about 5 mg/kg are contemplated, with single or multiple doses, depending on the target tissue.

[0557] In one embodiment, particles/nanoparticles that can deliver RNA to a cancer cell to stop tumor growth developed by Dan Anderson's lab at MIT may be used/and or adapted to the CRISPR Cas system of the present invention. In particular, the Anderson lab developed fully automated, combinatorial systems for the synthesis, purification, characterization, and

formulation of new biomaterials and nanoformulations. See, e.g., Alabi et al., Proc Natl Acad Sci U S A. 2013 Aug 6;110(32):12881-6; Zhang et al., Adv Mater. 2013 Sep 6;25(33):4641-5; Jiang et al., Nano Lett. 2013 Mar 13;13(3):1059-64; Karagiannis et al., ACS Nano. 2012 Oct 23;6(10):8484-7; Whitehead et al., ACS Nano. 2012 Aug 28;6(8):6922-9 and Lee et al., Nat Nanotechnol. 2012 Jun 3;7(6):389-93.

[0558] The lipid particles developed by the Qiaobing Xu's lab at Tufts University may be used/adapted to the present delivery system for cancer therapy. See 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]; Altınoğlu 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., Agnew Chem Int Ed Engl., 53(11):2893-8, Mar. 10, 2014.

[0559] US patent application 20110293703 relates to lipidoid compounds are also particularly useful in the administration of polynucleotides, which may be applied to deliver the CRISPR Cas system of the present invention. In one aspect, the aminoalcohol lipidoid compounds are combined with an agent to be delivered to a cell or a subject to form microparticles, nanoparticles, liposomes, or micelles. The agent to be delivered by the particles, liposomes, or micelles may be in the form of a gas, liquid, or solid, and the agent may be a polynucleotide, protein, peptide, or small molecule. The aminoalcohol lipidoid compounds may be combined with other aminoalcohol lipidoid compounds, polymers (synthetic or natural), surfactants, cholesterol, carbohydrates, proteins, lipids, etc. to form the particles. These particles may then optionally be combined with a pharmaceutical excipient to form a pharmaceutical composition.

[0560] US Patent Publication No. 20110293703 also provides methods of preparing the aminoalcohol lipidoid compounds. One or more equivalents of an amine are allowed to react with one or more equivalents of an epoxide-terminated compound under suitable conditions to form an aminoalcohol lipidoid compound of the present invention. In certain embodiments, all the amino groups of the amine are fully reacted with the epoxide-terminated compound to form tertiary amines. In other embodiments, all the amino groups of the amine are not fully reacted with the epoxide-terminated compound to form tertiary amines thereby resulting in primary or secondary amines in the aminoalcohol lipidoid compound. These primary or secondary amines

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

[0561] US Patent Publication No. 20110293703 also provides libraries of aminoalcohol lipidoid compounds prepared by the inventive methods. These aminoalcohol lipidoid compounds may be prepared and/or screened using high-throughput techniques involving liquid handlers, robots, microtiter plates, computers, etc. In certain embodiments, the aminoalcohol lipidoid compounds are screened for their ability to transfect polynucleotides or other agents (e.g., proteins, peptides, small molecules) into the cell.

[0562] US Patent Publication No. 20130302401 relates to a class of poly(beta-amino alcohols) (PBAAs) has been prepared using combinatorial polymerization. The inventive PBAAs may be used in biotechnology and biomedical applications as coatings (such as coatings of films or multilayer films for medical devices or implants), additives, materials, excipients, non-biofouling agents, micropatterning agents, and cellular encapsulation agents. When used as surface coatings, these PBAAs elicited different levels of inflammation, both *in vitro* and *in vivo*, depending on their chemical structures. The large chemical diversity of this

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

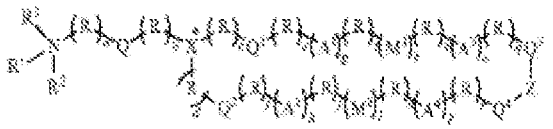
[0563] In another embodiment, lipid nanoparticles (LNPs) are contemplated. An antitransthyretin small interfering RNA has been encapsulated in lipid nanoparticles and delivered to humans (see, e.g., Coelho et al., N Engl J Med 2013;369:819-29), and such a system may be adapted and applied to the CRISPR Cas system of the present invention. Doses of about 0.01 to about 1 mg per kg of body weight administered intravenously are contemplated. Medications to reduce the risk of infusion-related reactions are contemplated, such as dexamethasone, acetaminophen, diphenhydramine or cetirizine, and ranitidine are contemplated. Multiple doses of about 0.3 mg per kilogram every 4 weeks for five doses are also contemplated.

[0564] Zhu et al. (US20140348900) provides for a process for preparing liposomes, lipid discs, and other lipid nanoparticles using a multi-port manifold, wherein the lipid solution stream, containing an organic solvent, is mixed with two or more streams of aqueous solution (e.g., buffer). In some aspects, at least some of the streams of the lipid and aqueous solutions are not directly opposite of each other. Thus, the process does not require dilution of the organic solvent as an additional step. In some embodiments, one of the solutions may also contain an active pharmaceutical ingredient (API). This invention provides a robust process of liposome manufacturing with different lipid formulations and different payloads. Particle size, morphology, and the manufacturing scale can be controlled by altering the port size and number of the manifold ports, and by selecting the flow rate or flow velocity of the lipid and aqueous solutions.

[0565] Cullis et al. (US 20140328759) provides limit size lipid nanoparticles with a diameter from 10-100 nm, in particular comprising a lipid bilayer surrounding an aqueous core. Methods and apparatus for preparing such limit size lipid nanoparticles are also disclosed.

[0566] Manoharan et al. (US 20140308304) provides cationic lipids of formula (I)

Formula (I)



or a salt thereof, wherein X is N or P; R' is absent, hydrogen, or alkyl; with respect to R^1 and R^2, (i) R^1 and R^2 are each, independently, optionally substituted alkyl, alkenyl, alkynyl, cycloalkyl, cycloalkylalkyl, heterocycle or R^{10}; (ii) R^1 and R^2, together with the nitrogen atom to which they are attached, form an optionally substituted heterocylic ring; or (iii) one of R^1 and R^2 is optionally substituted alkyl, alkenyl, alkynyl, cycloalkyl, cycloalkylalkyl, or heterocycle, and the other forms a 4-10 member heterocyclic ring or heteroaryl with (a) the adjacent nitrogen atom and (b) the $(R)_a$ group adjacent to the nitrogen atom; each occurrence of R is, independently, --(CR^3R^4)--; each occurrence of R^3 and R^4 are, independently H, halogen, OH, alkyl, alkoxy, --NH.sub.2, alkylamino, or dialkylamino; or R^3 and R^4, together with the carbon atom to which they are directly attached, form a cycloalkyl group, wherein no more than three R groups in each chain attached to the atom X* are cycloalkyl; each occurrence of R.sup.10 is independently selected from PEG and polymers based on poly(oxazoline), poly(ethylene oxide), poly(vinyl alcohol), poly(glycerol), poly(N-vinylpyrrolidone), poly[N-(2-hydroxypropyl)methacrylamide] and poly(amino acid)s, wherein (i) the PEG or polymer is linear or branched, (ii) the PEG or polymer is polymerized by n subunits, (iii) n is a number-averaged degree of polymerization between 10 and 200 units, and (iv) wherein the compound of formula has at most two R^{10} groups; Q is absent or is --O--, --NH--, --S--, --C(O)O--, --OC(O)--, --C(O)N(R^4)--, --N(R^5)C(O)--, --S--S--, --OC(O)O--, --O--N.dbd.C(R^5)--, --C(R^5).dbd.N--O--, --OC(O)N(R^5)--, --N(R^5)C(O)N(R^5)--, --N(R^5)C(O)O--, --C(O)S--, --C(S)O-- or --C(R^5).dbd.N--O--C(O)--; Q^1 and Q^2 are each, independently, absent, --O--, --S--, --OC(O)--, --C(O)O--, --SC(O)--, --C(O)S--, --OC(S)--, --C(S)O--, --S--S--, --C(O)(NR^5)--, --N(R^5)C(O)--, --C(S)(NR^5)--, --N(R^5)C(O)--, --N(R^5)C(O)N(R^5)--, or --OC(O)O--; Q^3 and Q^4 are each, independently, H, --(CR^3R^4)--, aryl, or a cholesterol moiety; each occurrence of A^1, A^2, A^3 and A^4 is, independently, --(CR^5R^5--CR^5.dbd.CR^5)--; each occurrence of R^5 is, independently, H or alkyl; M^1 and M^2 are each, independently, a biodegradable group (e.g., --OC(O)--, --C(O)O--, --SC(O)--, --C(O)S--, --OC(S)--, --C(S)O--, --S--S--, --C(R^5).dbd.N--, --N.dbd.C(R^5)--, --C(R^5).dbd.N--O--, --O--N.dbd.C(R5)--, --C(O)(NR5)--, --N(R5)C(O)--, --C(S)(NR^5)--, --N(R^5)C(O)--, --N(R^5)C(O)N(R^5)--, --OC(O)O--, --OSi(R^5).sub.2O--, --

C(O)(CR^3R^4)C(O)O--, or --OC(O)(CR^3R^4)C(O)--); Z is absent, alkylene or --O--P(O)(OH)--O--; each ------ attached to Z is an optional bond, such that when Z is absent, Q^3 and Q^4 are not directly covalently bound together; a is 1, 2, 3, 4, 5 or 6; b is 0, 1, 2, or 3; c, d, e, f, i, j, m, n, q and r are each, independently, 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10; g and h are each, independently, 0, 1 or 2; k and l are each, independently, 0 or 1, where at least one of k and l is 1; and o and p are each, independently, 0, 1 or 2, wherein Q^3 and Q^4 are each, independently, separated from the tertiary atom marked with an asterisk (X*) by a chain of 8 or more atoms. The cationic lipid can be used with other lipid components such as cholesterol and PEG-lipids to form lipid nanoparticles with oligonucleotides, to facilitate the cellular uptake and endosomal escape, and to knockdown target mRNA both *in vitro* and *in vivo*.

[0567] LNPs have been shown to be highly effective in delivering siRNAs to the liver (see, e.g., Tabernero et al., Cancer Discovery, April 2013, Vol. 3, No. 4, pages 363-470) and are therefore contemplated for delivering RNA encoding CRISPR Cas to the liver. A dosage of about four doses of 6 mg/kg of the LNP every two weeks may be contemplated. Tabernero et al. demonstrated that tumor regression was observed after the first 2 cycles of LNPs dosed at 0.7 mg/kg, and by the end of 6 cycles the patient had achieved a partial response with complete regression of the lymph node metastasis and substantial shrinkage of the liver tumors. A complete response was obtained after 40 doses in this patient, who has remained in remission and completed treatment after receiving doses over 26 months. Two patients with RCC and extrahepatic sites of disease including kidney, lung, and lymph nodes that were progressing following prior therapy with VEGF pathway inhibitors had stable disease at all sites for approximately 8 to 12 months, and a patient with PNET and liver metastases continued on the extension study for 18 months (36 doses) with stable disease.

[0568] However, the charge of the LNP must be taken into consideration. As cationic lipids combined with negatively charged lipids to induce nonbilayer structures that facilitate intracellular delivery. Because charged LNPs are rapidly cleared from circulation following intravenous injection, ionizable cationic lipids with pKa values below 7 were developed (see, e.g., Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). Negatively charged polymers such as RNA may be loaded into LNPs at low pH values (e.g., pH 4) where the ionizable lipids display a positive charge. However, at physiological pH values, the LNPs exhibit a low surface charge compatible with longer circulation times. Four species of ionizable cationic lipids have been focused upon, namely 1,2-dilineoyl-3-dimethylammonium-propane

(DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxy-keto-N,N-dimethyl-3-aminopropane (DLinKDMA), and 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA). It has been shown that LNP siRNA systems containing these lipids exhibit remarkably different gene silencing properties in hepatocytes *in vivo*, with potencies varying according to the series DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP employing a Factor VII gene silencing model (see, e.g., Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). A dosage of 1 μg/ml of LNP or CRISPR-Cas RNA in or associated with the LNP may be contemplated, especially for a formulation containing DLinKC2-DMA.

[0569] Preparation of LNPs and CRISPR Cas encapsulation may be used/and or adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). The cationic lipids 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA), (3-o-[2″-(methoxypolyethyleneglycol 2000) succinoyl]-1,2-dimyristoyl-sn-glycol (PEG-S-DMG), and R-3-[(ω-methoxy-poly(ethylene glycol)2000) carbamoyl]-1,2-dimyristyloxlpropyl-3-amine (PEG-C-DOMG) may be provided by Tekmira Pharmaceuticals (Vancouver, Canada) or synthesized. Cholesterol may be purchased from Sigma (St Louis, MO). The specific CRISPR Cas RNA may be encapsulated in LNPs containing DLinDAP, DLinDMA, DLinK-DMA, and DLinKC2-DMA (cationic lipid:DSPC:CHOL: PEGS-DMG or PEG-C-DOMG at 40:10:40:10 molar ratios). When required, 0.2% SP-DiOC18 (Invitrogen, Burlington, Canada) may be incorporated to assess cellular uptake, intracellular delivery, and biodistribution. Encapsulation may be performed by dissolving lipid mixtures comprised of cationic lipid:DSPC:cholesterol:PEG-c-DOMG (40:10:40:10 molar ratio) in ethanol to a final lipid concentration of 10 mmol/l. This ethanol solution of lipid may be added drop-wise to 50 mmol/l citrate, pH 4.0 to form multilamellar vesicles to produce a final concentration of 30% ethanol vol/vol. Large unilamellar vesicles may be formed following extrusion of multilamellar vesicles through two stacked 80 nm Nuclepore polycarbonate filters using the Extruder (Northern Lipids, Vancouver, Canada). Encapsulation may be achieved by adding RNA dissolved at 2 mg/ml in 50 mmol/l citrate, pH 4.0 containing 30% ethanol vol/vol drop-wise to extruded preformed large unilamellar vesicles and incubation at 31 °C for 30 minutes with constant mixing to a final RNA/lipid weight ratio of 0.06/1 wt/wt. Removal of ethanol and

neutralization of formulation buffer were performed by dialysis against phosphate-buffered saline (PBS), pH 7.4 for 16 hours using Spectra/Por 2 regenerated cellulose dialysis membranes. Nanoparticle size distribution may be determined by dynamic light scattering using a NICOMP 370 particle sizer, the vesicle/intensity modes, and Gaussian fitting (Nicomp Particle Sizing, Santa Barbara, CA). The particle size for all three LNP systems may be ~70 nm in diameter. RNA encapsulation efficiency may be determined by removal of free RNA using VivaPureD MiniH columns (Sartorius Stedim Biotech) from samples collected before and after dialysis. The encapsulated RNA may be extracted from the eluted nanoparticles and quantified at 260 nm. RNA to lipid ratio was determined by measurement of cholesterol content in vesicles using the Cholesterol E enzymatic assay from Wako Chemicals USA (Richmond, VA). In conjunction with the herein discussion of LNPs and PEG lipids, PEGylated liposomes or LNPs are likewise suitable for delivery of a CRISPR-Cas system or components thereof.

[0570] Preparation of large LNPs may be used/and or adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011. A lipid premix solution (20.4 mg/ml total lipid concentration) may be prepared in ethanol containing DLinKC2-DMA, DSPC, and cholesterol at 50:10:38.5 molar ratios. Sodium acetate may be added to the lipid premix at a molar ratio of 0.75:1 (sodium acetate:DLinKC2-DMA). The lipids may be subsequently hydrated by combining the mixture with 1.85 volumes of citrate buffer (10 mmol/l, pH 3.0) with vigorous stirring, resulting in spontaneous liposome formation in aqueous buffer containing 35% ethanol. The liposome solution may be incubated at 37 °C to allow for time-dependent increase in particle size. Aliquots may be removed at various times during incubation to investigate changes in liposome size by dynamic light scattering (Zetasizer Nano ZS, Malvern Instruments, Worcestershire, UK). Once the desired particle size is achieved, an aqueous PEG lipid solution (stock = 10 mg/ml PEG-DMG in 35% (vol/vol) ethanol) may be added to the liposome mixture to yield a final PEG molar concentration of 3.5% of total lipid. Upon addition of PEG-lipids, the liposomes should their size, effectively quenching further growth. RNA may then be added to the empty liposomes at an RNA to total lipid ratio of approximately 1:10 (wt:wt), followed by incubation for 30 minutes at 37 °C to form loaded LNPs. The mixture may be subsequently dialyzed overnight in PBS and filtered with a 0.45-μm syringe filter.

[0571] Preassembled recombinant complexes comprising Type VII effector and crRNA may be transfected, for example by electroporation, resulting in high mutation rates and

absence of detectable off-target mutations. Hur, J.K. et al, Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3596. [Epub ahead of print]

[0572] In terms of local delivery to the brain, this can be achieved in various ways. For instance, material can be delivered intrastriatally e.g. by injection. Injection can be performed stereotactically via a craniotomy.

[0573] Enhancing NHEJ or HR efficiency is also helpful for delivery. It is preferred that NHEJ efficiency is enhanced by co-expressing end-processing enzymes such as Trex2 (Dumitrache et al. Genetics. 2011 August; 188(4): 787–797). It is preferred that HR efficiency is increased by transiently inhibiting NHEJ machineries such as Ku70 and Ku86. HR efficiency can also be increased by co-expressing prokaryotic or eukaryotic homologous recombination enzymes such as RecBCD, RecA.

[0574] In some embodiments, sugar-based particles may be used, for example GalNAc, as described herein and with reference to WO2014118272 (incorporated herein by reference) and Nair, JK et al., 2014, Journal of the American Chemical Society 136 (49), 16958-16961) and the teaching herein, especially in respect of delivery applies to all particles unless otherwise apparent. This may be considered to be a sugar-based particle and further details on other particle delivery systems and/or formulations are provided herein. GalNAc can therefore be considered to be a particle in the sense of the other particles described herein, such that general uses and other considerations, for instance delivery of said particles, apply to GalNAc particles as well. A solution-phase conjugation strategy may for example be used to attach triantennary GalNAc clusters (mol. wt. ~2000) activated as PFP (pentafluorophenyl) esters onto 5′-hexylamino modified oligonucleotides (5′-HA ASOs, mol. wt. ~8000 Da; Østergaard et al., Bioconjugate Chem., 2015, 26 (8), pp 1451–1455). Similarly, poly(acrylate) polymers have been described for *in vivo* nucleic acid delivery (see WO2013158141 incorporated herein by reference). In further alternative embodiments, pre-mixing CRISPR nanoparticles (or protein complexes) with naturally occurring serum proteins may be used in order to improve delivery (Akinc A et al, 2010, Molecular Therapy vol. 18 no. 7, 1357–1364).

[0575] Additional nanoparticle configurations are also contemplated In another embodiment, Cas9 protein and sgRNA can be delivered into cells using"CRISPR-delivery particles," (CriPs), composed of nano-size complexes of Cas9 protein and sgRNA that are coated with an amphipathic peptide called Endo-Porter (EP) an amphipathic α-helical peptide

composed of leucine and histidine residues that mediates entry into cells. *See, e.g.* Shen et al., "CRISPR-delivery particles targeting nuclear receptor-interacting protein 1 (Nrip1) in adipose cells to enhance energy expenditure" J Biol Chem. 2018 Nov 2;293(44):17291-17305. In another embodiment, Cas9 plasmids and synthesized sgRNAs can be delivered into cells encapsulated in PEGylated nanoparticles (named P-HNPs) based on the cationic α-helical polypeptide poly(γ-4-((2-(piperidin-1-yl)ethyl)aminomethyl)benzyl-l-glutamate). *See, e.g.* Wang et al., "Nonviral gene editing via CRISPR/Cas9 delivery by membrane-disruptive and endosomolytic helical polypeptide" Proc Natl Acad Sci U S A. 2018 May 8; 115(19): 4903–4908. In another embodiment, a CRISPR-Cas9 plasmid can be incorporated into a self-assembled nanoparticle. The plasmid can be complexed with protamine sulfate and the resulting complex decorated by a multi-functional outer layer composed of an endosomolytic peptide (KALA) and aptamer AS1411 incorporated carboxymethyl chitosan. *See, e.g.* Liu et al., "Tumor targeted genome editing mediated by a multi-functional gene vector for regulating cell behaviors." J Control Release. 2018 Dec 10;291:90-98. In another embodiment, spCas9 and sgRNA RNP complexes can be delivered via a vesicle decorated with the fusogenic glycoprotein of the vesicular stomatitis virus (VSV-G). *See, e.g.* "VSV-G-Enveloped Vesicles for Traceless Delivery of CRISPR-Cas9" Montagna et al., Sci Rep. 2018 Nov 2;8(1):16304. In another embodiment, Cas9 RNP can be delivered by encapsulation in a nanoscale zeolitic imidazole framework (ZIF), in which enhanced endosomal escape is promoted by the protonated imidazole moieties. *See, e.g.* Alsaiari et al., "Endosomal Escape and Delivery of CRISPR/Cas9 Genome Editing Machinery Enabled by Nanoscale Zeolitic Imidazolate Framework" J Am Chem Soc. 2018 Jan 10;140(1):143-146.

[0576] Also contemplated are filtration-based systems. In one embodiment, Cas9 and sgRNA RNP complexes can be delivered into patient-derived hematopoietic stem and progenitor cells (HSPCs) *ex vivo* using TRansmembrane Internalization Assisted by Membrane Filtration (TRIAMF), a method to deliver RNPs into HSPCs by passing a RNP and cell mixture through a filter membrane. *See, e.g.* Yen et al., "TRIAMF: A New Method for Delivery of Cas9 Ribonucleoprotein Complex to Human Hematopoietic Stem Cells" Sci Rep. 2018 Nov 2;8(1):16304.

Options for DNA/RNA or DNA/DNA or RNA/RNA or protein/RNA

[0577] In some embodiments, the components of the CRISPR system may be delivered in various form, such as combinations of DNA/RNA or RNA/RNA or protein RNA. For example,

the Type VII effector may be delivered as a DNA-coding polynucleotide or an RNA--coding polynucleotide or as a protein. The guide may be delivered may be delivered as a DNA-coding polynucleotide or an RNA. All possible combinations are envisioned, including mixed forms of delivery.

[0578] In some embodiments, all such combinations (DNA/RNA or DNA/DNA or RNA/RNA or protein/RNA).

[0579] In some embodiment, when the Type VII effector is delivered in protein form, it is possible to pre-assemble same with one or more guide/s.

Nanoclews

[0580] Further, the CRISPR system may be delivered using nanoclews, for example as described in Sun W et al, *Cocoon-like self-degradable DNA nanoclew for anticancer drug delivery.,* J Am Chem Soc. 2014 Oct 22;136(42):14722-5. doi: 10.1021/ja5088024. Epub 2014 Oct 13. ; or in Sun W et al, *Self-Assembled DNA Nanoclews for the Efficient Delivery of CRISPR-Cas9 for Genome Editing.,* Angew Chem Int Ed Engl. 2015 Oct 5;54(41):12029-33. doi: 10.1002/anie.201506030. Epub 2015 Aug 27.

LNP

[0581] In some embodiments, delivery is by encapsulation of the Type VII effector protein or mRNA form in a lipid particle such as an LNP. In some embodiments, therefore, lipid nanoparticles (LNPs) are contemplated. An antitransthyretin small interfering RNA has been encapsulated in lipid nanoparticles and delivered to humans (see, e.g., Coelho et al., N Engl J Med 2013;369:819-29), and such a system may be adapted and applied to the CRISPR Cas system of the present invention. Doses of about 0.01 to about 1 mg per kg of body weight administered intravenously are contemplated. Medications to reduce the risk of infusion-related reactions are contemplated, such as dexamethasone, acetaminophen, diphenhydramine or cetirizine, and ranitidine are contemplated. Multiple doses of about 0.3 mg per kilogram every 4 weeks for five doses are also contemplated.

[0582] LNPs have been shown to be highly effective in delivering siRNAs to the liver (see, e.g., Tabernero et al., Cancer Discovery, April 2013, Vol. 3, No. 4, pages 363-470) and are therefore contemplated for delivering RNA encoding CRISPR Cas to the liver. A dosage of about four doses of 6 mg/kg of the LNP every two weeks may be contemplated. Tabernero et al. demonstrated that tumor regression was observed after the first 2 cycles of LNPs dosed at 0.7 mg/kg, and by the end of 6 cycles the patient had achieved a partial response with complete

regression of the lymph node metastasis and substantial shrinkage of the liver tumors. A complete response was obtained after 40 doses in this patient, who has remained in remission and completed treatment after receiving doses over 26 months. Two patients with RCC and extrahepatic sites of disease including kidney, lung, and lymph nodes that were progressing following prior therapy with VEGF pathway inhibitors had stable disease at all sites for approximately 8 to 12 months, and a patient with PNET and liver metastases continued on the extension study for 18 months (36 doses) with stable disease.

[0583] However, the charge of the LNP must be taken into consideration. As cationic lipids combined with negatively charged lipids to induce nonbilayer structures that facilitate intracellular delivery. Because charged LNPs are rapidly cleared from circulation following intravenous injection, ionizable cationic lipids with pKa values below 7 were developed (see, e.g., Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). Negatively charged polymers such as RNA may be loaded into LNPs at low pH values (e.g., pH 4) where the ionizable lipids display a positive charge. However, at physiological pH values, the LNPs exhibit a low surface charge compatible with longer circulation times. Four species of ionizable cationic lipids have been focused upon, namely 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxy-keto-N,N-dimethyl-3-aminopropane (DLinKDMA), and 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA). It has been shown that LNP siRNA systems containing these lipids exhibit remarkably different gene silencing properties in hepatocytes *in vivo*, with potencies varying according to the series DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP employing a Factor VII gene silencing model (see, e.g., Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). A dosage of 1 μg/ml of LNP or CRISPR-Cas RNA in or associated with the LNP may be contemplated, especially for a formulation containing DLinKC2-DMA.

[0584] Preparation of LNPs and CRISPR Cas encapsulation may be used/and or adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). The cationic lipids 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA), (3-o-[2″-(methoxypolyethyleneglycol 2000) succinoyl]-1,2-dimyristoyl-sn-glycol (PEG-S-DMG), and R-3-[(ω-methoxy-poly(ethylene glycol)2000) carbamoyl]-1,2-

dimyristyloxlpropyl-3-amine (PEG-C-DOMG) may be provided by Tekmira Pharmaceuticals (Vancouver, Canada) or synthesized. Cholesterol may be purchased from Sigma (St Louis, MO). The specific CRISPR Cas RNA may be encapsulated in LNPs containing DLinDAP, DLinDMA, DLinK-DMA, and DLinKC2-DMA (cationic lipid:DSPC:CHOL: PEGS-DMG or PEG-C-DOMG at 40:10:40:10 molar ratios). When required, 0.2% SP-DiOC18 (Invitrogen, Burlington, Canada) may be incorporated to assess cellular uptake, intracellular delivery, and biodistribution. Encapsulation may be performed by dissolving lipid mixtures comprised of cationic lipid:DSPC:cholesterol:PEG-c-DOMG (40:10:40:10 molar ratio) in ethanol to a final lipid concentration of 10 mmol/l. This ethanol solution of lipid may be added drop-wise to 50 mmol/l citrate, pH 4.0 to form multilamellar vesicles to produce a final concentration of 30% ethanol vol/vol. Large unilamellar vesicles may be formed following extrusion of multilamellar vesicles through two stacked 80 nm Nuclepore polycarbonate filters using the Extruder (Northern Lipids, Vancouver, Canada). Encapsulation may be achieved by adding RNA dissolved at 2 mg/ml in 50 mmol/l citrate, pH 4.0 containing 30% ethanol vol/vol drop-wise to extruded preformed large unilamellar vesicles and incubation at 31 °C for 30 minutes with constant mixing to a final RNA/lipid weight ratio of 0.06/1 wt/wt. Removal of ethanol and neutralization of formulation buffer were performed by dialysis against phosphate-buffered saline (PBS), pH 7.4 for 16 hours using Spectra/Por 2 regenerated cellulose dialysis membranes. Nanoparticle size distribution may be determined by dynamic light scattering using a NICOMP 370 particle sizer, the vesicle/intensity modes, and Gaussian fitting (Nicomp Particle Sizing, Santa Barbara, CA). The particle size for all three LNP systems may be ~70 nm in diameter. RNA encapsulation efficiency may be determined by removal of free RNA using VivaPureD MiniH columns (Sartorius Stedim Biotech) from samples collected before and after dialysis. The encapsulated RNA may be extracted from the eluted nanoparticles and quantified at 260 nm. RNA to lipid ratio was determined by measurement of cholesterol content in vesicles using the Cholesterol E enzymatic assay from Wako Chemicals USA (Richmond, VA). In conjunction with the herein discussion of LNPs and PEG lipids, PEGylated liposomes or LNPs are likewise suitable for delivery of a CRISPR-Cas system or components thereof.

[0585] Preparation of large LNPs may be used/and or adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011. A lipid premix solution (20.4 mg/ml total lipid concentration) may be prepared in ethanol containing DLinKC2-DMA, DSPC, and cholesterol at 50:10:38.5 molar ratios. Sodium acetate may be added to the lipid premix at a

molar ratio of 0.75:1 (sodium acetate:DLinKC2-DMA). The lipids may be subsequently hydrated by combining the mixture with 1.85 volumes of citrate buffer (10 mmol/l, pH 3.0) with vigorous stirring, resulting in spontaneous liposome formation in aqueous buffer containing 35% ethanol. The liposome solution may be incubated at 37 °C to allow for time-dependent increase in particle size. Aliquots may be removed at various times during incubation to investigate changes in liposome size by dynamic light scattering (Zetasizer Nano ZS, Malvern Instruments, Worcestershire, UK). Once the desired particle size is achieved, an aqueous PEG lipid solution (stock = 10 mg/ml PEG-DMG in 35% (vol/vol) ethanol) may be added to the liposome mixture to yield a final PEG molar concentration of 3.5% of total lipid. Upon addition of PEG-lipids, the liposomes should their size, effectively quenching further growth. RNA may then be added to the empty liposomes at an RNA to total lipid ratio of approximately 1:10 (wt:wt), followed by incubation for 30 minutes at 37 °C to form loaded LNPs. The mixture may be subsequently dialyzed overnight in PBS and filtered with a 0.45-μm syringe filter.

[0586] Spherical Nucleic Acid (SNA™) constructs and other nanoparticles (particularly gold nanoparticles) are also contemplated as a means to delivery CRISPR-Cas system to intended targets. Significant data show that AuraSense Therapeutics' Spherical Nucleic Acid (SNA™) constructs, based upon nucleic acid-functionalized gold nanoparticles, are useful.

[0587] Literature that may be employed in conjunction with herein teachings include: Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang et al., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi et al., Proc. Natl. Acad. Sci. USA. 2013 110(19):7625-7630, Jensen et al., Sci. Transl. Med. 5, 209ra152 (2013) and Mirkin, et al., Small, 10:186-192.

[0588] Self-assembling nanoparticles with RNA may be constructed with polyethyleneimine (PEI) that is PEGylated with an Arg-Gly-Asp (RGD) peptide ligand attached at the distal end of the polyethylene glycol (PEG). This system has been used, for example, as a means to target tumor neovasculature expressing integrins and deliver siRNA inhibiting vascular endothelial growth factor receptor-2 (VEGF R2) expression and thereby achieve tumor angiogenesis (see, e.g., Schiffelers et al., Nucleic Acids Research, 2004, Vol.

32, No. 19). Nanoplexes may be prepared by mixing equal volumes of aqueous solutions of cationic polymer and nucleic acid to give a net molar excess of ionizable nitrogen (polymer) to phosphate (nucleic acid) over the range of 2 to 6. The electrostatic interactions between cationic polymers and nucleic acid resulted in the formation of polyplexes with average particle size distribution of about 100 nm, hence referred to here as nanoplexes. A dosage of about 100 to 200 mg of CRISPR Cas is envisioned for delivery in the self-assembling nanoparticles of Schiffelers et al.

[0589] The nanoplexes of Bartlett et al. (PNAS, September 25, 2007,vol. 104, no. 39) may also be applied to the present invention. The nanoplexes of Bartlett et al. are prepared by mixing equal volumes of aqueous solutions of cationic polymer and nucleic acid to give a net molar excess of ionizable nitrogen (polymer) to phosphate (nucleic acid) over the range of 2 to 6. The electrostatic interactions between cationic polymers and nucleic acid resulted in the formation of polyplexes with average particle size distribution of about 100 nm, hence referred to here as nanoplexes. The DOTA-siRNA of Bartlett et al. was synthesized as follows: 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid mono(N-hydroxysuccinimide ester) (DOTA-NHSester) was ordered from Macrocyclics (Dallas, TX). The amine modified RNA sense strand with a 100-fold molar excess of DOTA-NHS-ester in carbonate buffer (pH 9) was added to a microcentrifuge tube. The contents were reacted by stirring for 4 h at room temperature. The DOTA-RNAsense conjugate was ethanol-precipitated, resuspended in water, and annealed to the unmodified antisense strand to yield DOTA-siRNA. All liquids were pretreated with Chelex-100 (Bio-Rad, Hercules, CA) to remove trace metal contaminants. Tf-targeted and nontargeted siRNA nanoparticles may be formed by using cyclodextrin-containing polycations. Typically, nanoparticles were formed in water at a charge ratio of 3 (+/-) and an siRNA concentration of 0.5 g/liter. One percent of the adamantane-PEG molecules on the surface of the targeted nanoparticles were modified with Tf (adamantane-PEG-Tf). The nanoparticles were suspended in a 5% (wt/vol) glucose carrier solution for injection.

[0590] Davis et al. (Nature, Vol 464, 15 April 2010) conducts a RNA clinical trial that uses a targeted nanoparticle-delivery system (clinical trial registration number NCT00689065). Patients with solid cancers refractory to standard-of-care therapies are administered doses of targeted nanoparticles on days 1, 3, 8 and 10 of a 21-day cycle by a 30-min intravenous infusion. The nanoparticles consist of a synthetic delivery system containing: (1) a linear, cyclodextrin-based polymer (CDP), (2) a human transferrin protein (TF) targeting ligand

displayed on the exterior of the nanoparticle to engage TF receptors (TFR) on the surface of the cancer cells, (3) a hydrophilic polymer (polyethylene glycol (PEG) used to promote nanoparticle stability in biological fluids), and (4) siRNA designed to reduce the expression of the RRM2 (sequence used in the clinic was previously denoted siR2B+5). The TFR has long been known to be upregulated in malignant cells, and RRM2 is an established anti-cancer target. These nanoparticles (clinical version denoted as CALAA-01) have been shown to be well tolerated in multi-dosing studies in non-human primates. Although a single patient with chronic myeloid leukemia has been administered siRNA by liposomal delivery, Davis et al.'s clinical trial is the initial human trial to systemically deliver siRNA with a targeted delivery system and to treat patients with solid cancer. To ascertain whether the targeted delivery system can provide effective delivery of functional siRNA to human tumors, Davis et al. investigated biopsies from three patients from three different dosing cohorts; patients A, B and C, all of whom had metastatic melanoma and received CALAA-01 doses of 18, 24 and 30 mg m-2 siRNA, respectively. Similar doses may also be contemplated for the CRISPR Cas system of the present invention. The delivery of the invention may be achieved with nanoparticles containing a linear, cyclodextrin-based polymer (CDP), a human transferrin protein (TF) targeting ligand displayed on the exterior of the nanoparticle to engage TF receptors (TFR) on the surface of the cancer cells and/or a hydrophilic polymer (for example, polyethylene glycol (PEG) used to promote nanoparticle stability in biological fluids).

[0591] US Patent No. 8,709,843, incorporated herein by reference, provides a drug delivery system for targeted delivery of therapeutic agent-containing particles to tissues, cells, and intracellular compartments. The invention provides targeted particles comprising polymer conjugated to a surfactant, hydrophilic polymer or lipid.

[0592] US Patent No. 6,007,845, incorporated herein by reference, provides particles which have a core of a multiblock copolymer formed by covalently linking a multifunctional compound with one or more hydrophobic polymers and one or more hydrophilic polymers, and contain a biologically active material.

[0593] US Patent No. 5,855,913, incorporated herein by reference, provides a particulate composition having aerodynamically light particles having a tap density of less than 0.4 g/cm3 with a mean diameter of between 5 μm and 30 μ m, incorporating a surfactant on the surface thereof for drug delivery to the pulmonary system.

[0594] US Patent No. 5,985,309, incorporated herein by reference, provides particles incorporating a surfactant and/or a hydrophilic or hydrophobic complex of a positively or negatively charged therapeutic or diagnostic agent and a charged molecule of opposite charge for delivery to the pulmonary system.

[0595] US. Patent No. 5,543,158, incorporated herein by reference, provides biodegradable injectable particles having a biodegradable solid core containing a biologically active material and poly(alkylene glycol) moieties on the surface.

[0596] WO2012135025 (also published as US20120251560), incorporated herein by reference, describes conjugated polyethyleneimine (PEI) polymers and conjugated aza-macrocycles (collectively referred to as "conjugated lipomer" or "lipomers"). In certain embodiments, it was envisioned that such conjugated lipomers can be used in the context of the CRISPR-Cas system to achieve *in vitro*, *ex vivo* and *in vivo* genomic perturbations to modify gene expression, including modulation of protein expression.

[0597] In one embodiment, the nanoparticle may be epoxide-modified lipid–polymer, advantageously 7C1 (see, e.g., James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84). C71 was synthesized by reacting C15 epoxide-terminated lipids with PEI600 at a 14:1 molar ratio, and was formulated with C14PEG2000 to produce nanoparticles (diameter between 35 and 60 nm) that were stable in PBS solution for at least 40 days.

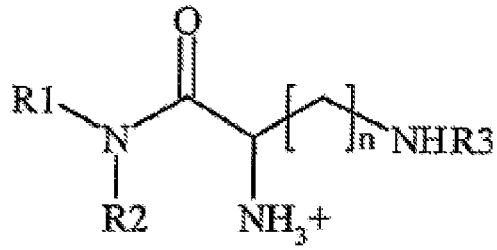
[0598] An epoxide-modified lipid-polymer may be utilized to deliver the CRISPR-Cas system of the present invention to pulmonary, cardiovascular or renal cells, however, one of skill in the art may adapt the system to deliver to other target organs. Dosage ranging from about 0.05 to about 0.6 mg/kg are envisioned. Dosages over several days or weeks are also envisioned, with a total dosage of about 2 mg/kg.

[0599] In some embodiments, the LNP for delivering the RNA molecules is prepared by methods known in the art, such as those described in, for example, International Patent Publication Nos. WO 2005/105152 (PCT/EP2005/004920), WO 2006/069782 (PCT/EP2005/014074), WO 2007/121947 (PCT/EP2007/003496), and WO 2015/082080 (PCT/EP2014/003274), which are incorporated herein by reference. LNPs aimed specifically at the enhanced and improved delivery of siRNA into mammalian cells are described in, for example, 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), which are herein incorporated by reference and may be applied to the present technology.

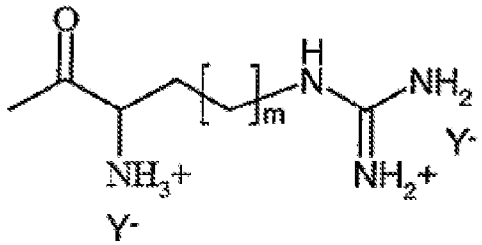
[0600] In some embodiments, the LNP includes any LNP disclosed in International Patent Publication Nos. WO 2005/105152 (PCT/EP2005/004920), WO 2006/069782 (PCT/EP2005/014074), WO 2007/121947 (PCT/EP2007/003496), and WO 2015/082080 (PCT/EP2014/003274).

[0601] In some embodiments, the LNP includes at least one lipid having Formula I:



(Formula I),

[0602] wherein R1 and R2 are each and independently selected from the group comprising alkyl, n is any integer between 1 and 4, and R3 is an acyl selected from the group comprising lysyl, ornithyl, 2,4-diaminobutyryl, histidyl and an acyl moiety according to Formula II:



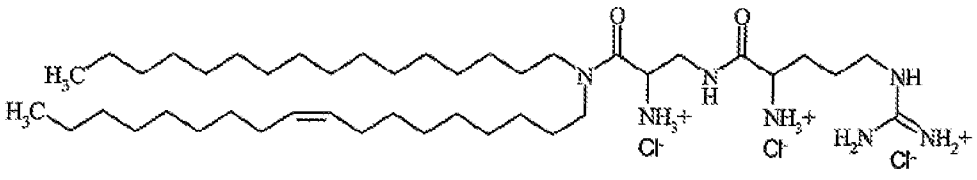
(Formula II),

[0603] wherein m is any integer from 1 to 3 and Y^- is a pharmaceutically acceptable anion. In some embodiments, a lipid according to Formula I includes at least two asymmetric C atoms. In some embodiments, enantiomers of Formula I include, but are not limited to, R-R; S-S; R-S and S-R enantiomer.

[0604] In some embodiments, R1 is lauryl and R2 is myristyl. In another embodiment, R1 is palmityl and R2 is oleyl. In some embodiments, m is 1 or 2. In some embodiments, Y- is selected from halogenids, acetate or trifluoroacetate.

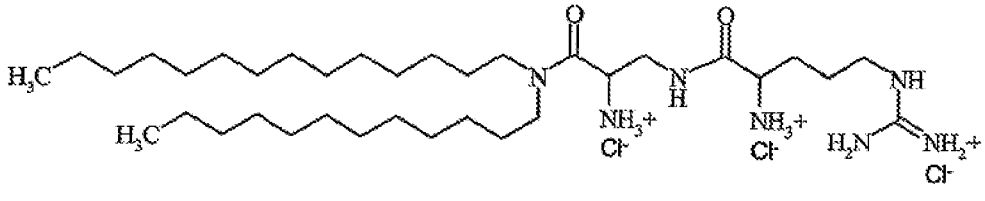
[0605] In some embodiments, the LNP comprises one or more lipids select from:

[0606] β-arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrochloride (Formula III):



(Formula III);

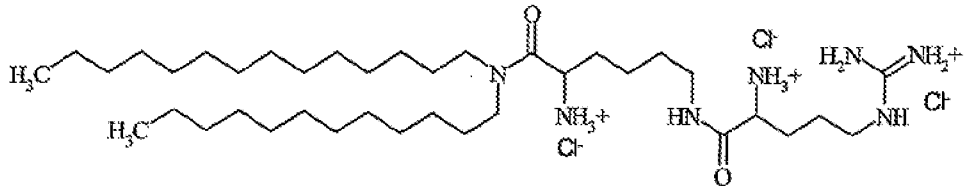
[0607] β-arginyl-2,3-diamino propionic acid-N-lauryl-N-myristyl-amide trihydrochloride (Formula IV):



(Formula IV);

and

[0608] ε-arginyl-lysine-N-lauryl-N-myristyl-amide trihydrochloride (Formula V):



(Formula V).

[0609] In some embodiments, the LNP also includes a constituent. By way of example, but not by way of limitation, in some embodiments, the constituent is selected from peptides, proteins, oligonucleotides, polynucleotides, nucleic acids, or a combination thereof. In some embodiments, the constituent is an antibody, e.g., a monoclonal antibody. In some embodiments, the constituent is a nucleic acid selected from, e.g., ribozymes, aptamers, spiegelmers, DNA, RNA, PNA, LNA, or a combination thereof. In some embodiments, the nucleic acid is gRNA and/or mRNA.

[0610] In some embodiments, the constituent of the LNP comprises an mRNA encoding a CRIPSR effector protein. In some embodiments, the constituent of the LNP comprises an mRNA encoding a Type-II, Type-V, or Type-VI CRIPSR effector protein. In some embodiments, the constituent of the LNP comprises an mRNA encoding an RNA-guided DNA binding protein. In some embodiments, the constituent of the LNP comprises an mRNA encoding an RNA-guided RNA binding protein.

[0611] In some embodiments, the constituent of the LNP further comprises one or more guide RNA. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to vascular endothelium. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to pulmonary endothelium. In some

embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to liver. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to lung. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to hearts. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to spleen. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to kidney. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to pancrea. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to brain. In some embodiments, the LNP is configured to deliver the aforementioned mRNA and guide RNA to macrophages.

[0612] In some embodiments, the LNP also includes at least one helper lipid. In some embodiments, the helper lipid is selected from phospholipids and steroids. In some embodiments, the phospholipids are di- and /or monoester of the phosphoric acid. In some embodiments, the phospholipids are phosphoglycerides and /or sphingolipids. In some embodiments, the steroids are naturally occurring and/or synthetic compounds based on the partially hydrogenated cyclopenta[a]phenanthrene. In some embodiments, the steroids contain 21 to 30 C atoms. In some embodiments, the steroid is cholesterol. In some embodiments, the helper lipid is selected from l,2-diphytanoyl-sn-glycero-3-phosphoethanolamine (DPhyPE), ceramide, and 1,2-dioleylsn-glycero-3-phosphoethanolamine (DOPE).

[0613] In some embodiments, the at least one helper lipid comprises a moiety selected from the group comprising a PEG moiety, a HEG moiety, a polyhydroxyethyl starch (polyHES) moiety and a polypropylene moiety. In some embodiments, the moiety has a molecule weight between about 500 to 10,000 Da or between about 2,000 to 5,000 Da. In some embodiments, the PEG moiety is selected from l,2-distearoyl-sn-glycero-3 phosphoethanolamine, l,2-dialkyl-sn-glycero-3-phosphoethanolamine, and Ceramide-PEG. In some embodiments, the PEG moiety has a molecular weight between about 500 to 10,000 Da or between about 2,000 to 5,000 Da. In some embodiments, the PEG moiety has a molecular weight of 2,000 Da.

[0614] 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.

[0615] In some embodiments, the LNP includes any of β-L-3-arginyl-2,3-diaminopropionic acid-N-palmityl-N-oleyl-amide trihydrochloride, β-L-arginyl-2,3-diaminopropionic acid-N-lauryl-N-myristyl-amide trihydrochloride or arginyl-lysine-N-lauryl-N-myristyl-amide trihydrochloride in combination with DPhyPE, wherein the content of DPhyPE is about 80 mol %, 65 mol %, 50 mol % and 35 mol % of the overall lipid content of the LNP. In some embodiments, the LNP includes β-L-arginyl-2,3-diamino propionic acid-N-pahnityl-N-oleyl-amide trihydrochloride (lipid) and 1,2-diphytanoyl-sn-glycero-3-phosphoethanolamine (helper lipid). In some embodiments, the LNP includes β-L-arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrochloride (lipid), l,2-diphytanoyl-sn-glycero-3-phosphoethanolamine (first helper lipid), and l,2-disteroyl-sn-glycero-3-phosphoethanolamine-PEG2000 (second helper lipid).

[0616] In some embodiments, the second helper lipid is between about 0.05 mol% to 4.9 mol% or between about 1 mol% to 3 mol% of the total lipid content. In some embodiments, the LNP includes lipids at between about 45 mol% to 50 mol% of the total lipid content, a first helper lipid between about 45 mol% to 50 mol% of the total lipid content, under the proviso that there is a PEGylated second helper lipid between about 0.1 mol% to 5 mol %, between about 1 mol% to 4 mol%, or at about 2 mol% of the total lipid content, wherein the sum of the content of the lipids, the first helper lipid, and of the second helper lipid is 100 mol% of the total lipid content and wherein the sum of the first helper lipid and the second helper lipid is 50 mol% of the total lipid content. In some embodiments, the LNP comprises: (a) 50 mol% of arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrochloride, 48 mol% of l,2-diphytanoyl-sn-glycero-3-phosphoethanolamine; and 2 mol% l,2-distearoyl-sn-glycero-3-phosphoethanolamine-PEG2000; or (b) 50 mol% of β-L-arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrocloride, 49 mol% l,2-diphytanoyl-sn-glycero-3-phosphoethanolamine; and 1 mol% N(Carbonyl-methoxypolyethylenglycol-2000)-1,2-distearoyl-sn-glycero3-phosphoethanolamine, or a sodium salt thereof.

[0617] 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.

[0618] 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

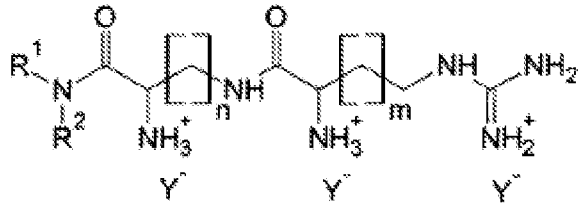
compound does not carry any charge on its surface or on the molecule as such. In some embodiments, the shielding compounds are polyethylenglycoles (PEGs), hydroxyethylglucose (HEG) based polymers, polyhydroxyethyl starch (polyHES) and polypropylene. In some embodiments, the PEG, HEG, polyHES, and a polypropylene 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.

[0619] In some embodiments, the LNP includes at least one lipid, a first helper lipid, and a shielding compound that is removable from the lipid composition under *in vivo* conditions. In some embodiments, the LNP also includes a second helper lipid. In some embodiments, the first helper lipid is ceramide. In some embodiments, the second helper lipid is ceramide. In some embodiments, the ceramide comprises at least one short carbon chain substituent of from 6 to 10 carbon atoms. In some embodiments, the ceramide comprises 8 carbon atoms. In some embodiments, the shielding compound is attached to a ceramide. In some embodiments, the shielding compound is attached to a ceramide. In some embodiments, the shielding compound is covalently attached to the ceramide. In some embodiments, the shielding compound is attached to a nucleic acid in the LNP. In some embodiments, the shielding compound is covalently attached to the nucleic acid. In some embodiments, the shielding compound is attached to the nucleic acid by a linker. In some embodiments, the linker is cleaved under physiological conditions. In some embodiments, the linker is selected from ssRNA, ssDNA, dsRNA, dsDNA, peptide, S-S-linkers and pH sensitive linkers. In some embodiments, the linker moiety is attached to the 3' end of the sense strand of the nucleic acid. In some embodiments, the shielding compound comprises a pH-sensitive linker or a pH-sensitive moiety. In some embodiments, the pH-sensitive linker or pH-sensitive moiety is an anionic linker or an anionic moiety. In some embodiments, the anionic linker or anionic moiety is less anionic or neutral in an acidic environment. In some embodiments, the pH-sensitive linker or the pH-sensitive moiety is selected from the oligo (glutamic acid), oligophenolate(s) and diethylene triamine penta acetic acid.

[0620] In any of the LNP embodiments in the previous paragraph, the LNP can have an osmolality between about 50 to 600 mosmole/kg, between about 250 to 350 mosmole/kg, or between about 280 to 320 mosmole/kg, and/or wherein the LNP formed by the lipid and/or one or two helper lipids and the shielding compound have a particle size between about 20 to 200 nm, between about 30 to 100 nm, or between about 40 to 80 nm.

[0621] In some embodiments, the shielding compound provides for a longer circulation time *in vivo* and allows for a better biodistribution of the nucleic acid containing LNP. In some embodiments, the shielding compound prevents immediate interaction of the LNP with serum compounds or compounds of other bodily fluids or cytoplasma membranes, e.g., cytoplasma membranes of the endothelial lining of the vasculature, into which the LNP is administered. Additionally or alternatively, in some embodiments, the shielding compounds also prevent elements of the immune system from immediately interacting with the LNP. Additionally or alternatively, in some embodiments, the shielding compound acts as an anti-opsonizing compound. Without wishing to be bound by any mechanism or theory, in some embodiments, the shielding compound forms a cover or coat that reduces the surface area of the LNP available for interaction with its environment. Additionally or alternatively, in some embodiments, the shielding compound shields the overall charge of the LNP.

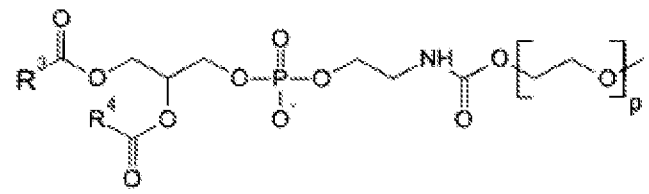
[0622] In another embodiment, the LNP includes at least one cationic lipid having Formula VI:



(Formula VI),

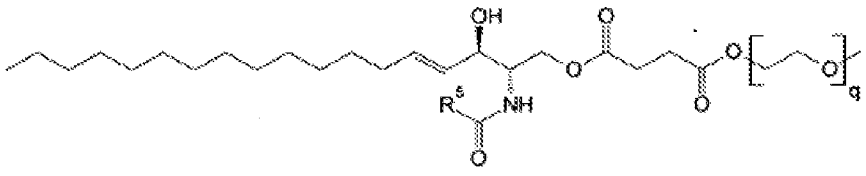
[0623] wherein n is 1, 2, 3, or 4, wherein m is 1, 2, or 3, wherein Y^- is anion, wherein each of R^1 and R^2 is individually and independently selected from the group consisting of linear C12-C18 alkyl and linear C12-C18 alkenyl, a sterol compound, wherein the sterol compound is selected from the group consisting of cholesterol and stigmasterol, and a PEGylated lipid, wherein the PEGylated lipid comprises a PEG moiety, wherein the PEGylated lipid is selected from the group consisting of:

[0624] a PEGylated phosphoethanolamine of Formula VII:



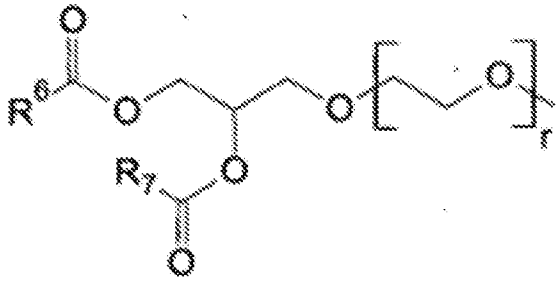
(Formula VII),

[0625] wherein R^3 and R^4 are individually and independently linear C13-C17 alkyl, and p is any integer between 15 to 130; a PEGylated ceramide of Formula VIII:



(Formula VIII);

[0626] wherein R^5 is linear C7-C15 alkyl, and q is any number between 15 to 130; and a PEGylated diacylglycerol of Formula IX:

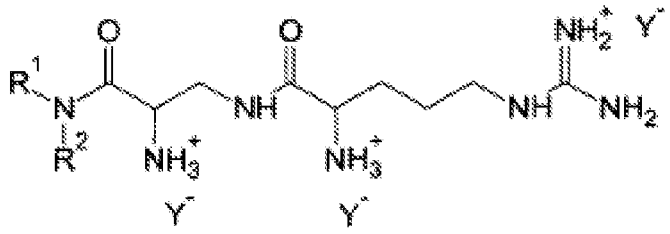


(Formula IX),

[0627] wherein each of R^6 and R^7 is individually and independently linear C11-C17 alkyl, and r is any integer from 15 to 130.

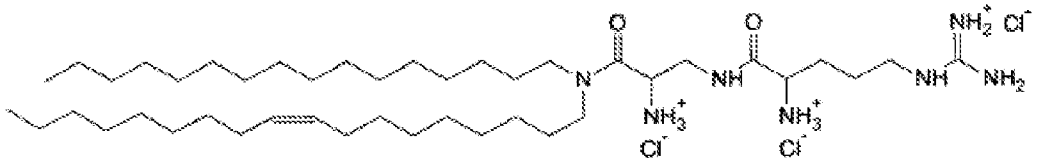
[0628] In some embodiments, R^1 and R^2 are different from each other. In some embodiments, R^1 is palmityl and R^2 is oleyl. In some embodiments, R^1 is lauryl and R^2 is myristyl. In some embodiments, R^1 and R^2 are the same. In some embodiments, each of R^1 and R^2 is individually and independently selected from the group consisting of C12 alkyl, C14 alkyl, Cl6 alkyl, C18 alkyl, Cl2 alkenyl, C14 alkenyl, Cl6 alkenyl and Cl8 alkenyl. In some embodiments, each of C12 alkenyl, C14 alkenyl, Cl6 alkenyl and Cl 8 alkenyl comprises one or two double bonds. In some embodiments, Cl8 alkenyl is Cl8 alkenyl with one double bond between C9 and C10. In some embodiments, C18 alkenyl is cis-9-octadecyl.

[0629] In some embodiments, the cationic lipid is a compound of Formula X:



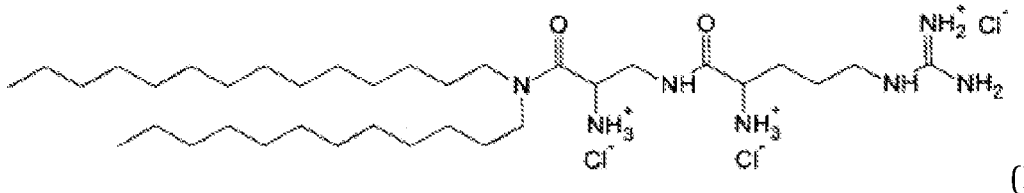
(Formula X).

[0630] In some embodiments, Y^- is selected from halogenids, acetate and trifluoroacetate. In some embodiments, the cationic lipid is β-arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrochloride of Formula III:



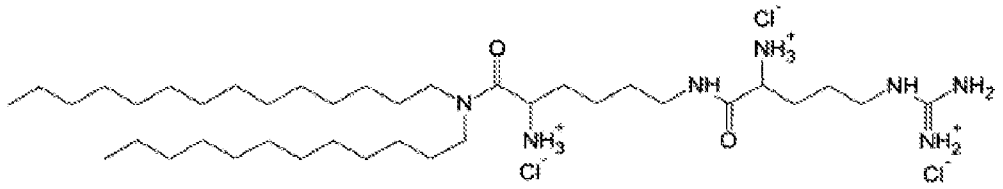
(Formula III).

[0631] In some embodiments, the cationic lipid is β-arginyl-2,3-diamino propionic acid-N-lauryl-N-myristyl-amide trihydrochloride of Formula IV:



(Formula IV).

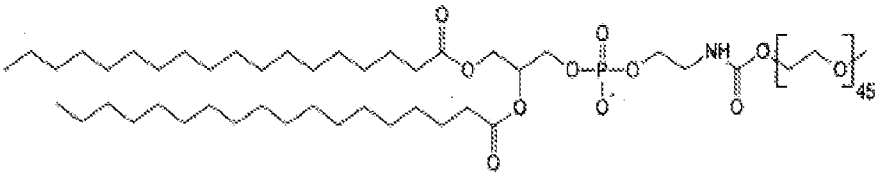
[0632] In some embodiments, the cationic lipid is ε-arginyl-lysine-N-lauryl-N-myristyl-amide trihydrochloride of Formula V:



(Formula V).

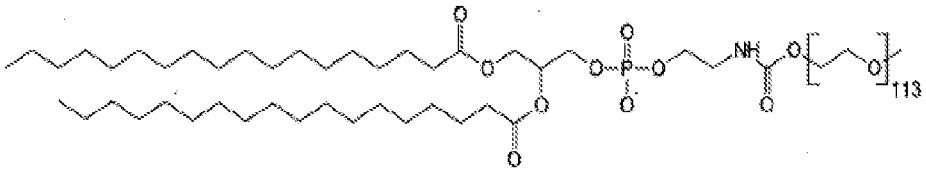
[0633] In some embodiments, the sterol compound is cholesterol. In some embodiments, the sterol compound is stigmasterin.

[0634] In some embodiments, the PEG moiety of the PEGylated lipid has a molecular weight from about 800 to 5,000 Da. In some embodiments, the molecular weight of the PEG moiety of the PEGylated lipid is about 800 Da. In some embodiments, the molecular weight of the PEG moiety of the PEGylated lipid is about 2,000 Da. In some embodiments, the molecular weight of the PEG moiety of the PEGylated lipid is about 5,000 Da. In some embodiments, the PEGylated lipid is a PEGylated phosphoethanolamine of Formula VII, wherein each of R^3 and R^4 is individually and independently linear C13-C17 alkyl, and p is any integer from 18, 19 or 20, or from 44, 45 or 46 or from 113, 114 or 115. In some embodiments, R^3 and R^4 are the same. In some embodiments, R^3 and R^4 are different. In some embodiments, each of R^3 and R^4 is individually and independently selected from the group consisting of C13 alkyl, Cl5 alkyl and Cl7 alkyl. In some embodiments, the PEGylated phosphoethanolamine of Formula VII is l,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-2000] (ammonium salt):



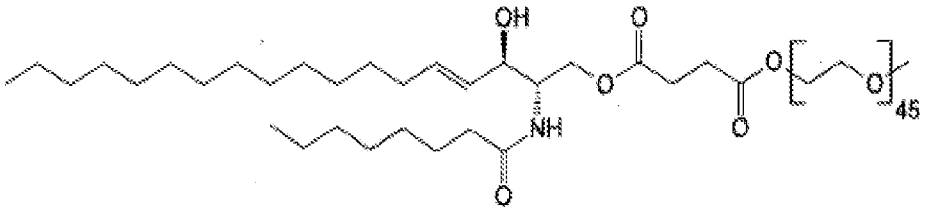
(Formula XI).

[0635] In some embodiments, the PEGylated phosphoethanolamine of Formula VII is 1,2-distearoyl-*sn*-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-5000] (ammonium salt):



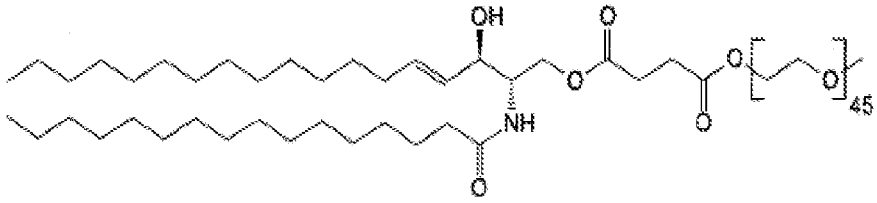
(Formula XII).

[0636] In some embodiments, the PEGylated lipid is a PEGylated ceramide of Formula VIII, wherein R^5 is linear C7-C15 alkyl, and q is any integer from 18, 19 or 20, or from 44, 45 or 46 or from 113, 114 or 115. In some embodiments, R^5 is linear C7 alkyl. In some embodiments, R^5 is linear C15 alkyl. In some embodiments, the PEGylated ceramide of Formula VIII is N-octanoyl-sphingosine-1- {succinyl[methoxy(polyethylene glycol)2000]}:



(Formula XIII).

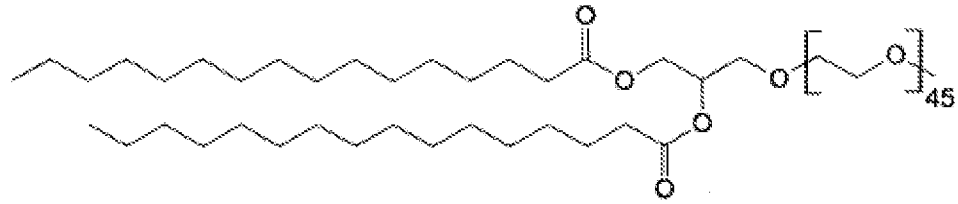
[0637] In some embodiments, the PEGylated ceramide of Formula VIII is N-palmitoyl-sphingosine-1- {succinyl[methoxy(polyethylene glycol)2000]}



(Formula XIV).

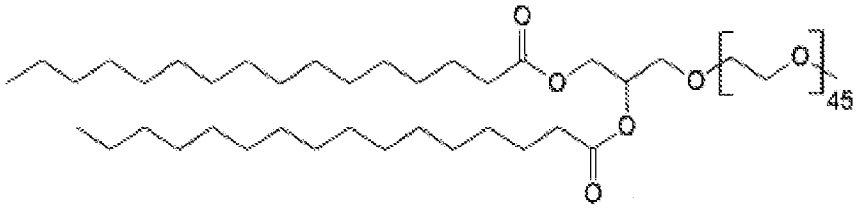
[0638] In some embodiments, the PEGylated lipid is a PEGylated diacylglycerol of Formula IX, wherein each of R^6 and R^7 is individually and independently linear C11-C17 alkyl, and r is any integer from 18, 19 or 20, or from 44, 45 or 46 or from 113, 114 or 115. In some embodiments, R^6 and R^7 are the same. In some embodiments, R^6 and R^7 are different. In some embodiments, each of R^6 and R^7 is individually and independently selected from the group consisting of linear C17 alkyl, linear C15 alkyl and linear C13 alkyl. In some embodiments, the

PEGylated diacylglycerol of Formula IX 1,2-Distearoyl-sn-glycerol [methoxy(polyethylene glycol)2000]:



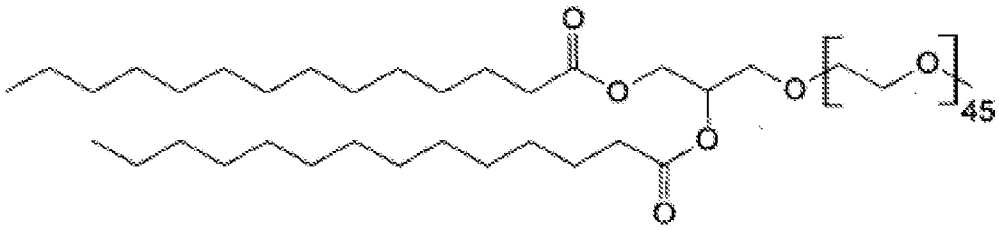
(Formula XV).

[0639] In some embodiments, the PEGylated diacylglycerol of Formula IX is 1,2-Dipalmitoyl-sn-glycerol [methoxy(polyethylene glycol)2000]:



(Formula XVI).

[0640] In some embodiments, the PEGylated diacylglycerol of Formula IX is:



(Formula XVII).

[0641] In some embodiments, the LNP includes at least one cationic lipid selected from of Formulas III, IV, and V, at least one sterol compound selected from a cholesterol and stigmasterin, and wherein the PEGylated lipid is at least one selected from Formulas XI and XII. In some embodiments, the LNP includes at least one cationic lipid selected from Formulas III, IV, and V, at least one sterol compound selected from a cholesterol and stigmasterin, and wherein the PEGylated lipid is at least one selected from Formulas XIII and XIV. In some embodiments, the LNP includes at least one cationic lipid selected from Formulas III, IV, and V, at least one sterol compound selected from a cholesterol and stigmasterin, and wherein the PEGylated lipid is at least one selected from Formulas XV and XVI. In some embodiments, the LNP includes a cationic lipid of Formula III, a cholesterol as the sterol compound, and wherein the PEGylated lipid is Formula XI.

[0642] In any of the LNP embodiments in the previous paragraph, wherein the content of the cationic lipid composition is between about 65 mole% to 75 mole%, the content of the

sterol compound is between about 24 mole% to 34 mole% and the content of the PEGylated lipid is between about 0.5 mole% to 1.5 mole%, wherein the sum of the content of the cationic lipid, of the sterol compound and of the PEGylated lipid for the lipid composition is 100 mole%. In some embodiments, the cationic lipid is about 70 mole%, the content of the sterol compound is about 29 mole% and the content of the PEGylated lipid is about 1 mole%. In some embodiments, the LNP is 70 mole% of Formula III, 29 mole% of cholesterol, and 1 mole% of Formula XI.

Exosomes

[0643] Exosomes are endogenous nano-vesicles that transport RNAs and proteins, and which can deliver RNA to the brain and other target organs. To reduce immunogenicity, Alvarez-Erviti et al. (2011, Nat Biotechnol 29: 341) used self-derived dendritic cells for exosome production. Targeting to the brain was achieved by engineering the dendritic cells to express Lamp2b, an exosomal membrane protein, fused to the neuron-specific RVG peptide. Purified exosomes were loaded with exogenous RNA by electroporation. Intravenously injected RVG-targeted exosomes delivered GAPDH siRNA specifically to neurons, microglia, oligodendrocytes in the brain, resulting in a specific gene knockdown. Pre-exposure to RVG exosomes did not attenuate knockdown, and non-specific uptake in other tissues was not observed. The therapeutic potential of exosome-mediated siRNA delivery was demonstrated by the strong mRNA (60%) and protein (62%) knockdown of BACE1, a therapeutic target in Alzheimer's disease.

[0644] To obtain a pool of immunologically inert exosomes, Alvarez-Erviti et al. harvested bone marrow from inbred C57BL/6 mice with a homogenous major histocompatibility complex (MHC) haplotype. As immature dendritic cells produce large quantities of exosomes devoid of T-cell activators such as MHC-II and CD86, Alvarez-Erviti et al. selected for dendritic cells with granulocyte/macrophage-colony stimulating factor (GM-CSF) for 7 d. Exosomes were purified from the culture supernatant the following day using well-established ultracentrifugation protocols. The exosomes produced were physically homogenous, with a size distribution peaking at 80 nm in diameter as determined by nanoparticle tracking analysis (NTA) and electron microscopy. Alvarez-Erviti et al. obtained 6-12 μg of exosomes (measured based on protein concentration) per 10^6 cells.

[0645] Next, Alvarez-Erviti et al. investigated the possibility of loading modified exosomes with exogenous cargoes using electroporation protocols adapted for nanoscale

applications. As electroporation for membrane particles at the nanometer scale is not well-characterized, nonspecific Cy5-labeled RNA was used for the empirical optimization of the electroporation protocol. The amount of encapsulated RNA was assayed after ultracentrifugation and lysis of exosomes. Electroporation at 400 V and 125 μF resulted in the greatest retention of RNA and was used for all subsequent experiments.

[0646] Alvarez-Erviti et al. administered 150 μg of each BACE1 siRNA encapsulated in 150 μg of RVG exosomes to normal C57BL/6 mice and compared the knockdown efficiency to four controls: untreated mice, mice injected with RVG exosomes only, mice injected with BACE1 siRNA complexed to an *in vivo* cationic liposome reagent and mice injected with BACE1 siRNA complexed to RVG-9R, the RVG peptide conjugated to 9 D-arginines that electrostatically binds to the siRNA. Cortical tissue samples were analyzed 3 d after administration and a significant protein knockdown (45%, $P < 0.05$, versus 62%, $P < 0.01$) in both siRNA-RVG-9R-treated and siRNARVG exosome-treated mice was observed, resulting from a significant decrease in BACE1 mRNA levels (66% [+ or -] 15%, $P < 0.001$ and 61% [+ or -] 13% respectively, $P < 0.01$). Moreover, a significant decrease (55%, $P < 0.05$) in the total [beta]-amyloid 1-42 levels is demonstrated, a main component of the amyloid plaques in Alzheimer's pathology, in the RVG-exosome-treated animals. The decrease observed was greater than the β-amyloid 1-40 decrease demonstrated in normal mice after intraventricular injection of BACE1 inhibitors. Alvarez-Erviti et al. carried out 5'-rapid amplification of cDNA ends (RACE) on BACE1 cleavage product, which provided evidence of RNAi-mediated knockdown by the siRNA.

[0647] Finally, Alvarez-Erviti et al. investigated whether RNA-RVG exosomes induced immune responses *in vivo* by assessing IL-6, IP-10, TNFα and IFN-α serum concentrations. Following exosome treatment, nonsignificant changes in all cytokines were registered similar to siRNA-transfection reagent treatment in contrast to siRNA-RVG-9R, which potently stimulated IL-6 secretion, confirming the immunologically inert profile of the exosome treatment. Given that exosomes encapsulate only 20% of siRNA, delivery with RVG-exosome appears to be more efficient than RVG-9R delivery as comparable mRNA knockdown and greater protein knockdown was achieved with fivefold less siRNA without the corresponding level of immune stimulation. This experiment demonstrated the therapeutic potential of RVG-exosome technology, which is potentially suited for long-term silencing of genes related to neurodegenerative diseases. The exosome delivery system of Alvarez-Erviti et al. may be

applied to deliver the CRISPR-Cas system of the present invention to therapeutic targets, especially neurodegenerative diseases. A dosage of about 100 to 1000 mg of CRISPR Cas encapsulated in about 100 to 1000 mg of RVG exosomes may be contemplated for the present invention.

[0648] El-Andaloussi et al. (Nature Protocols 7,2112–2126(2012)) discloses how exosomes derived from cultured cells can be harnessed for delivery of RNA *in vitro* and *in vivo*. This protocol first describes the generation of targeted exosomes through transfection of an expression vector, comprising an exosomal protein fused with a peptide ligand. Next, El-Andaloussi et al. explain how to purify and characterize exosomes from transfected cell supernatant. Next, El-Andaloussi et al. detail crucial steps for loading RNA into exosomes. Finally, El-Andaloussi et al. outline how to use exosomes to efficiently deliver RNA *in vitro* and *in vivo* in mouse brain. Examples of anticipated results in which exosome-mediated RNA delivery is evaluated by functional assays and imaging are also provided. The entire protocol takes ~3 weeks. Delivery or administration according to the invention may be performed using exosomes produced from self-derived dendritic cells. From the herein teachings, this can be employed in the practice of the invention.

[0649] In another embodiment, the plasma exosomes of Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130) are contemplated. Exosomes are nano-sized vesicles (30–90nm in size) produced by many cell types, including dendritic cells (DC), B cells, T cells, mast cells, epithelial cells and tumor cells. These vesicles are formed by inward budding of late endosomes and are then released to the extracellular environment upon fusion with the plasma membrane. Because exosomes naturally carry RNA between cells, this property may be useful in gene therapy, and from this disclosure can be employed in the practice of the instant invention.

[0650] Exosomes from plasma can be prepared by centrifugation of buffy coat at 900g for 20 min to isolate the plasma followed by harvesting cell supernatants, centrifuging at 300g for 10 min to eliminate cells and at 16 500g for 30 min followed by filtration through a 0.22 mm filter. Exosomes are pelleted by ultracentrifugation at 120 000g for70 min. Chemical transfection of siRNA into exosomes is carried out according to the manufacturer's instructions in RNAi Human/Mouse Starter Kit (Quiagen, Hilden, Germany). siRNA is added to 100 ml PBS at a final concentration of 2 mmol/ml. After adding HiPerFect transfection reagent, the mixture is incubated for 10 min at RT. In order to remove the excess of micelles, the exosomes

are re-isolated using aldehyde/sulfate latex beads. The chemical transfection of CRISPR Cas into exosomes may be conducted similarly to siRNA. The exosomes may be co-cultured with monocytes and lymphocytes isolated from the peripheral blood of healthy donors. Therefore, it may be contemplated that exosomes containing CRISPR Cas may be introduced to monocytes and lymphocytes of and autologously reintroduced into a human. Accordingly, delivery or administration according to the invention may be performed using plasma exosomes.

Liposomes

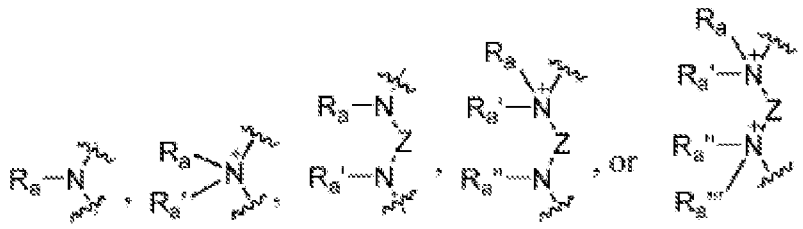
[0651] In the case where a non-viral delivery system is utilized, an exemplary delivery vehicle is a liposome. The use of lipid formulations is contemplated for the introduction of the nucleic acids into a host cell (*in vitro*, *ex vivo* or in vivo). In another aspect, the nucleic acid may be associated with a lipid. The nucleic acid associated with a lipid may be encapsulated in the aqueous interior of a liposome, interspersed within the lipid bilayer of a liposome, attached to a liposome via a linking molecule that is associated with both the liposome and the oligonucleotide, entrapped in a liposome, complexed with a liposome, dispersed in a solution containing a lipid, mixed with a lipid, combined with a lipid, contained as a suspension in a lipid, contained or complexed with a micelle, or otherwise associated with a lipid. Lipid, lipid/DNA or lipid/expression vector associated compositions are not limited to any particular structure in solution. For example, they may be present in a bilayer structure, as micelles, or with a "collapsed" structure. They may also simply be interspersed in a solution, possibly forming aggregates that are not uniform in size or shape. Lipids are fatty substances which may be naturally occurring or synthetic lipids. For example, lipids include the fatty droplets that naturally occur in the cytoplasm as well as the class of compounds which contain long-chain aliphatic hydrocarbons and their derivatives, such as fatty acids, alcohols, amines, amino alcohols, and aldehydes.

[0652] Lipids suitable for use can be obtained from commercial sources. For example, dimyristyl phosphatidylcholine ("DMPC") can be obtained from Sigma, St. Louis, Mo.; dicetyl phosphate ("DCP") can be obtained from K & K Laboratories (Plainview, N.Y.); cholesterol ("Choi") can be obtained from Calbiochem-Behring; dimyristyl phosphatidylglycerol ("DMPG") and other lipids may be obtained from Avanti Polar Lipids, Inc. (Birmingham, Ala.). Stock solutions of lipids in chloroform or chloroform/methanol can be stored at about −20° C. Chloroform is used as the only solvent since it is more readily evaporated than

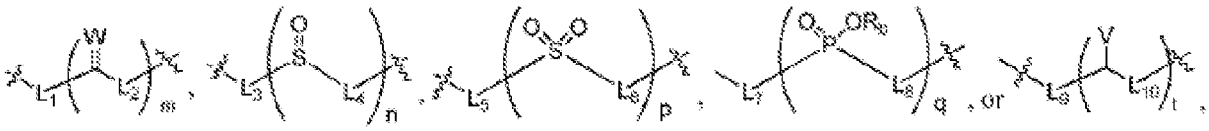
methanol. "Liposome" is a generic term encompassing a variety of single and multilamellar lipid vehicles formed by the generation of enclosed lipid bilayers or aggregates. Liposomes can be characterized as having vesicular structures with a phospholipid bilayer membrane and an inner aqueous medium. Multilamellar liposomes have multiple lipid layers separated by aqueous medium. They form spontaneously when phospholipids are suspended in an excess of aqueous solution. The lipid components undergo self-rearrangement before the formation of closed structures and entrap water and dissolved solutes between the lipid bilayers (Ghosh et al., 1991 Glycobiology 5: 505-10). However, compositions that have different structures in solution than the normal vesicular structure are also encompassed. For example, the lipids may assume a micellar structure or merely exist as nonuniform aggregates of lipid molecules. Also contemplated are lipofectamine-nucleic acid complexes.

[0653] The lipid, lipid particle, or lipid bilayer or lipid entity of the invention can be prepared by methods well known in the art. See Wang et al., ACS Synthetic Biology, 1, 403-07 (2012); Wang et al., PNAS, 113(11) 2868-2873 (2016); Manoharan, et al., International Patent Publication No. WO 2008/042973; Zugates et al., US Pat. No. 8,071,082; Xu et al., International Patent Publication No. WO 2014/186366 A1 (US20160082126). Xu et provides a way to make a nanocomplex for the delivery of saporin wherein the nanocomplex comprising saporin and a lipid-like compound, and wherein the nanocomplex has a particle size of 50 nm to 1000 nm; the saporin binds to the lipid-like compound via non-covalent interaction or covalent bonding; and the lipid-like compound has a hydrophilic moiety, a hydrophobic moiety, and a linker joining the hydrophilic moiety and the hydrophobic moiety, the hydrophilic moiety being optionally charged and the hydrophobic moiety having 8 to 24 carbon atoms. Xu et al., International Patent Publication No. WO 2014/186348 (US20160129120) provides examples of nanocomplexes of modified peptides or proteins comprising a cationic delivery agent and an anionic pharmaceutical agent, wherein the nanocomplex has a particle size of 50 to 1000 nm, the cationic delivery agent binds to the anionic pharmaceutical agent, and the anionic pharmaceutical agent is a modified peptide or protein formed of a peptide and a protein and an added chemical moiety that contains an anionic group. The added chemical moiety is linked to the peptide or protein via an amide group, an ester group, an ether group, a thioether group, a disulfide group, a hydrazone group, a sulfenate ester group, an amidine group, a urea group, a carbamate group, an imidoester group, or a carbonate group. More particularly these documents provide examples of lipid or lipid-like compounds that can be

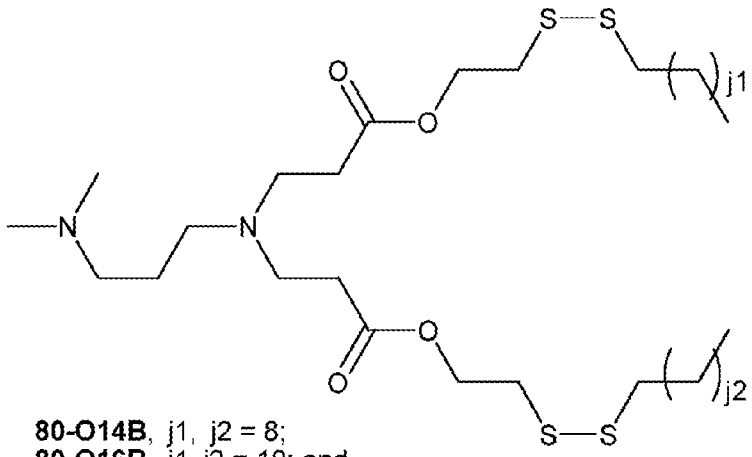
used to make the particle delivery system of the present invention, including compounds of the formula B_1-K_1-A-K_2-B_2, in which A, the hydrophilic moiety, is



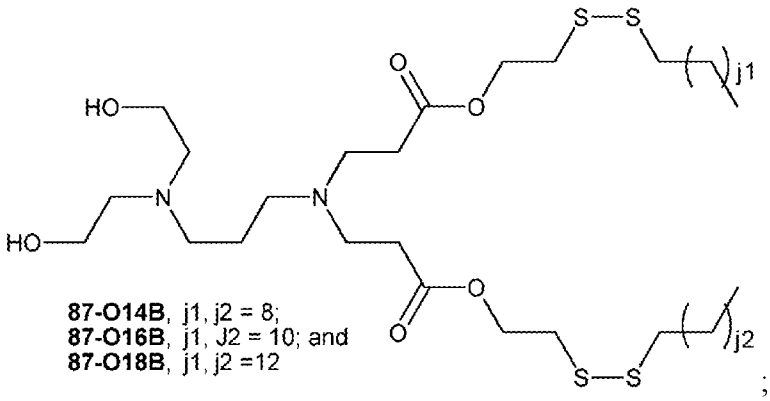
, each of R_a, R_a', R_a'', and R_a''', independently, being a C_1-C_{20} monovalent aliphatic radical, a C_1-C_{20} $_0$ monovalent heteroaliphatic radical, a monovalent aryl radical, or a monovalent heteroaryl radical; and Z being a C_1-C_{20} bivalent aliphatic radical, a C_1-C_{20} bivalent heteroaliphatic radical, a bivalent aryl radical, or a bivalent heteroaryl radical; each of B_1, the hydrophobic moiety, and B_2, also the hydrophobic moiety, independently, is a C_{12}-$_{20}$ aliphatic radical or a $_{C12-20}$ heteroaliphatic radical; and each of K_1, the linker, and K_2, also the linker, independently, is O, S, Si, C_1-C_6 alkylene



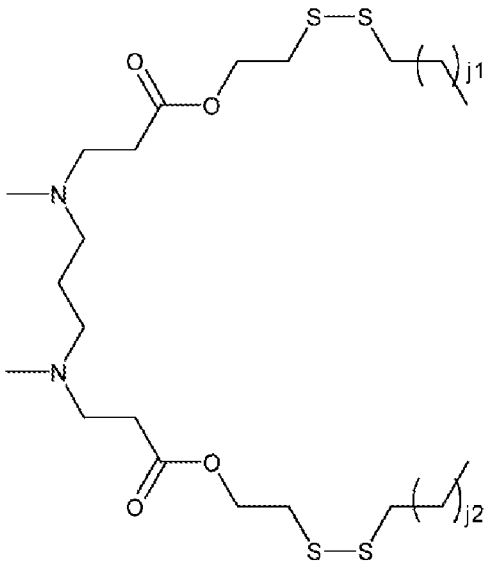
in which each of m, n, p, q, and t, independently, is 1-6; W is O, S, or NR_C; each of L_1, L_3, L_5, L_7, and L_9, independently, is a bond, O, S, or NR_d; each of L2, L_4, L_6, L_8, and L_{10}, independently, is a bond, O, S, or NR_e; and V is OR_f, SR_g, or NR_hR_i, each of R_b , R_c, R_d, R_e, R_f, R_g, R_h, and R_i, independently, being H, OH, a C_1-C_{10} oxyaliphatic radical, a C_1-C_{10} monovalent aliphatic radical, a C_1-C_{10} monovalent heteroaliphatic radical, a monovalent aryl radical, or a monovalent heteroaryl radical and specific compounds:

[0654] 

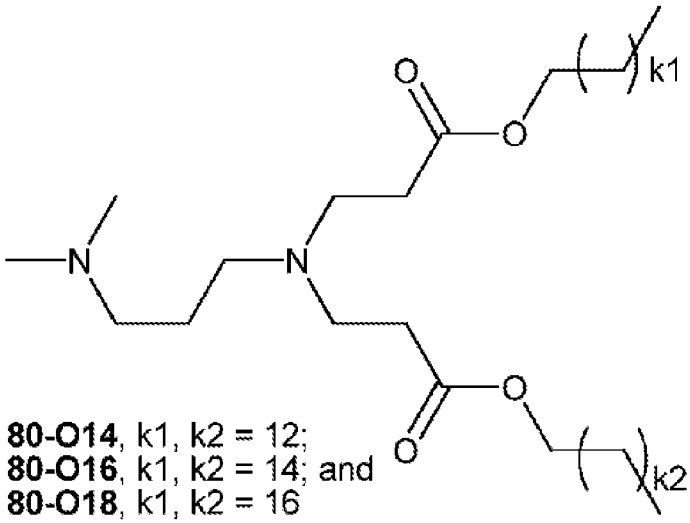
80-O14B, j1, j2 = 8;
80-O16B, j1, j2 = 10; and
80-O18B, j1, j2 = 12 ;

[0655] 

87-O14B, j1, j2 = 8;
87-O16B, j1, J2 = 10; and
87-O18B, j1, j2 =12 ;

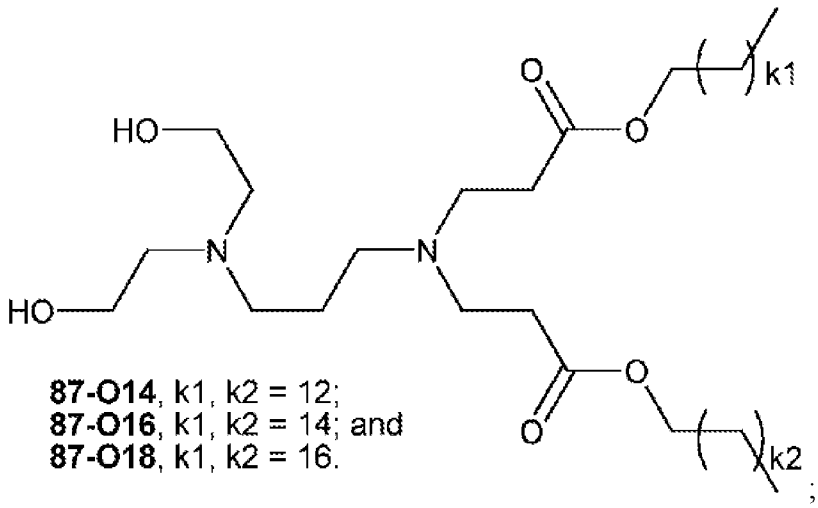
[0656] 

1-O16B, j1, j2 = 10; and
1-O18B, j1, j2 = 12 ;



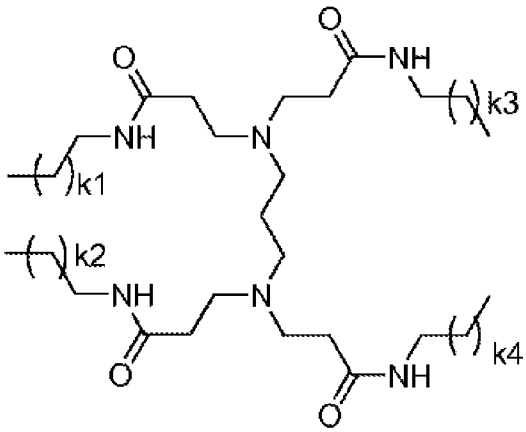
80-O14, k1, k2 = 12;
80-O16, k1, k2 = 14; and
80-O18, k1, k2 = 16

[0657] ;



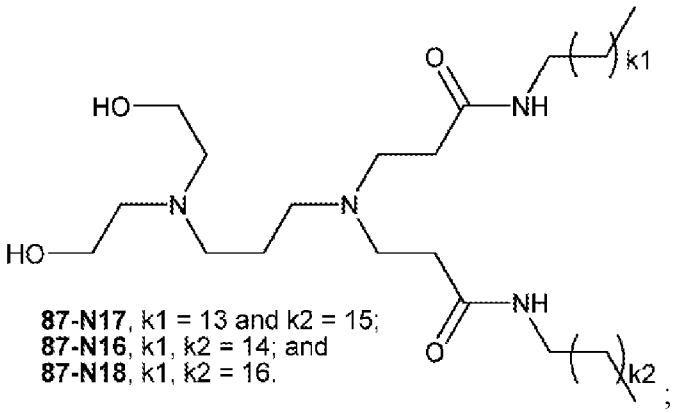
87-O14, k1, k2 = 12;
87-O16, k1, k2 = 14; and
87-O18, k1, k2 = 16.

[0658] ;

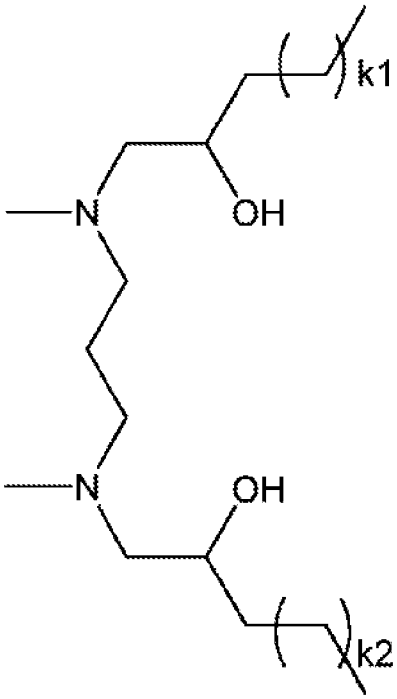


1-N16, k1, k2, k3, k4 = 14
1-N18, k1 = 12, k2 = 13, k3 = 15, and k4 = 16;

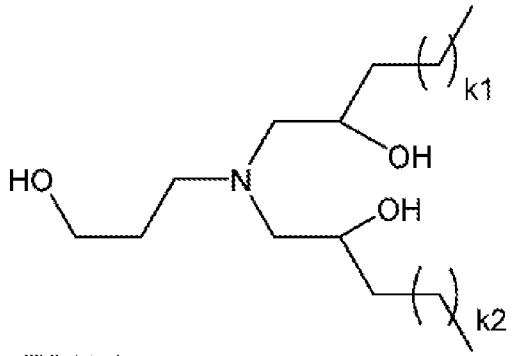
[0659]

[0660] 

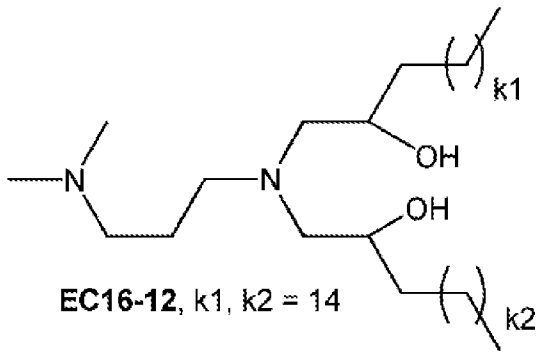
87-N17, k1 = 13 and k2 = 15;
87-N16, k1, k2 = 14; and
87-N18, k1, k2 = 16.

[0661] 

EC16-1, k1, k2 = 14 ;

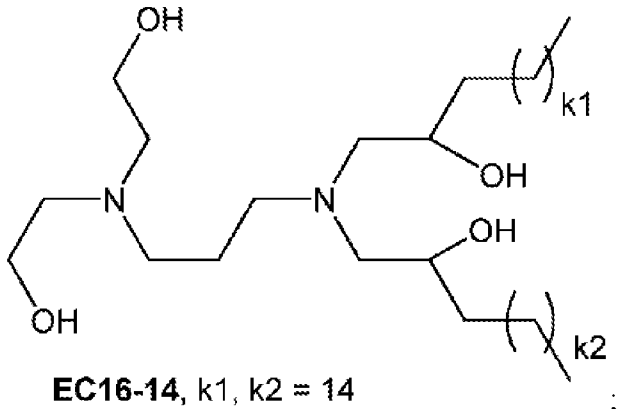
[0662] 

EC16-3, k1, k2 = 14 ;

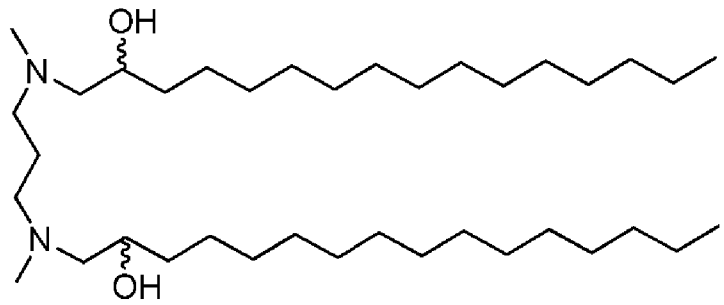


EC16-12, k1, k2 = 14

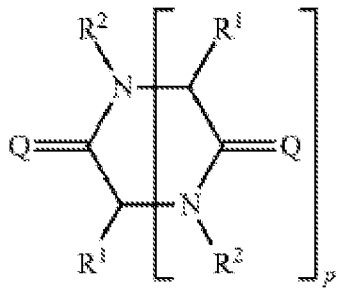
[0663] ;



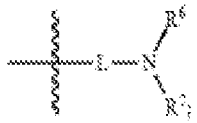
[0664] **EC16-14**, k1, k2 = 14 ; and



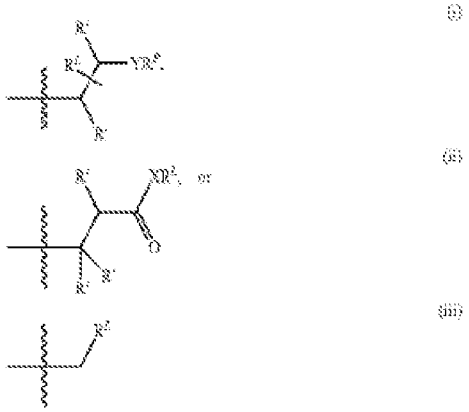
[0665] EC16-63

[0666] Additional examples of cationic lipid that can be used to make the particle delivery system of the invention can be found in US20150140070, wherein the cationic lipid has the formula  , wherein p is an integer between 1 and 9, inclusive; each

instance of Q is independently O, S, or NR^Q; R^Q is hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, a nitrogen protecting group, or a group of the formula (i), (ii) or (iii); each instance of R^1 is independently hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, halogen, $—OR^{A1}$, $—N(R^{A1})_2$, $—SR^{A1}$, or a group of formula:

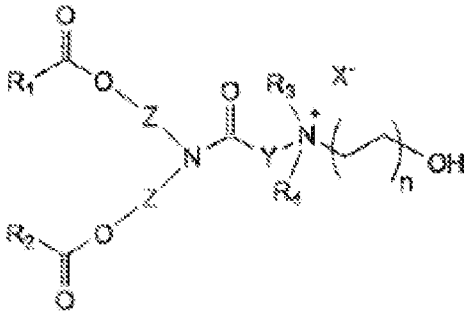


, L is an optionally substituted alkylene, optionally substituted alkenylene, optionally substituted alkynylene, optionally substituted heteroalkylene, optionally substituted heteroalkenylene, optionally substituted heteroalkynylene, optionally substituted carbocyclylene, optionally substituted heterocyclylene, optionally substituted arylene, or optionally substituted heteroarylene, or combination thereof, and each of R^6 and R^7 is independently hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, a nitrogen protecting group, or a group of formula (i), (ii) or (iii); each occurrence of R^{A1} is independently hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, an oxygen protecting group when attached to an oxygen atom, a sulfur protecting group when attached to an sulfur atom, a nitrogen protecting group when attached to a nitrogen atom, or two R^{A1} groups, together with the nitrogen atom to which they are attached, are joined to form an optionally substituted heterocyclic or optionally substituted heteroaryl ring; each instance of R^2 is independently hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, a nitrogen protecting group, or a group of the formula (i), (ii),

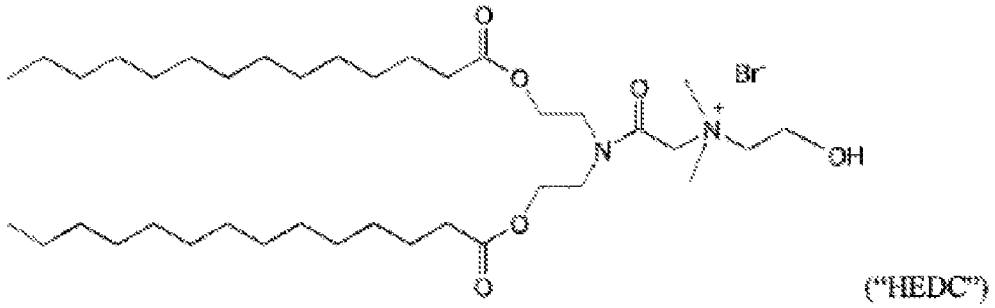


or (iii); Formulae (i), (ii), and (iii) are: , each instance of R′ is independently hydrogen or optionally substituted alkyl; X is O, S, or NR^X; R^X is hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, or a nitrogen protecting group; Y is O, S, or NR^Y; R^Y is hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, or a nitrogen protecting group; R^P is hydrogen, optionally substituted alkyl, optionally substituted alkenyl, optionally substituted alkynyl, optionally substituted carbocyclyl, optionally substituted heterocyclyl, optionally substituted aryl, optionally substituted heteroaryl, an oxygen protecting group when attached to an oxygen atom, a sulfur protecting group when attached to a sulfur atom, or a nitrogen protecting group when attached to a nitrogen atom; R^L is optionally substituted C_{1-50} alkyl, optionally substituted C_{2-50} alkenyl, optionally substituted C_{2-50} alkynyl, optionally substituted heteroC_{1-50} alkyl, optionally substituted heteroC_{2-50} alkenyl, optionally substituted heteroC_{2-50} alkynyl, or a polymer; provided that at least one instance of R^Q, R^2, R^6, or R^7 is a group of the formula (i), (ii), or (iii); in Liu et al., (US 20160200779, US 20150118216, US 20150071903, and US 20150071903), which provide examples of cationic lipids to include polyethylenimine, polyamidoamine (PAMAM) starburst dendrimers, 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.). Exemplary cationic liposomes can be made from N-[1-(2,3-

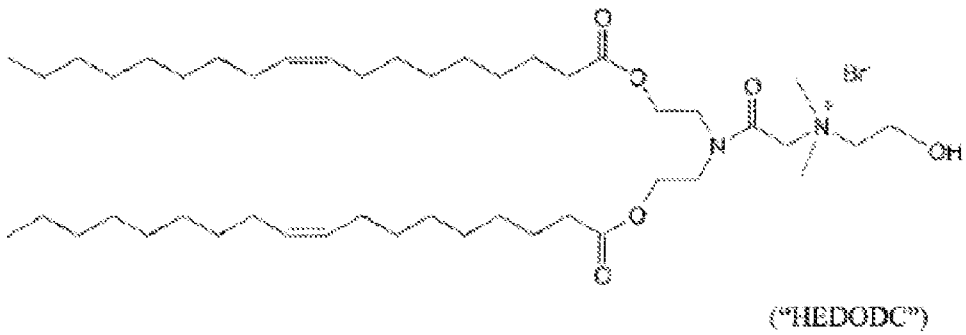
dioleoloxy)-propyl]-N,N,N-trimethylammonium chloride (DOTMA), N-[1-(2,3-dioleoloxy)-propyl]-N,N,N-trimethylammonium methylsulfate (DOTAP), 3.beta.-[N--(N',N'-dimethylaminoethane)carbamoyl]cholesterol (DC-Chol), 2,3,-dioleyloxy-N-[2(sperminecarboxamido)ethyl]-N,N-dimethyl-1-propanamin- ium trifluoroacetate (DOSPA), 1,2-dimyristyloxypropyl-3-dimethyl-hydroxyethyl ammonium bromide; and dimethyldioctadecylammonium bromide (DDAB); in WO2013/093648 which provides



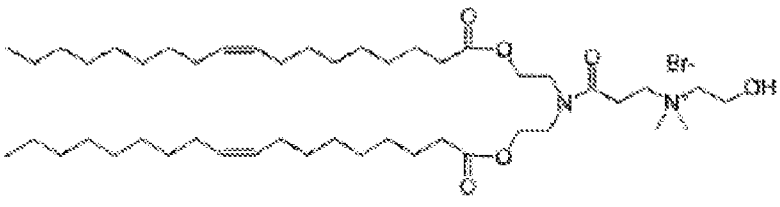
cationic lipids of formula in which Z = an alkyl linker, C_2-C_4 alkyl, Y = an alkyl linker, C_1-C_6 alkyl, R_1 and R_2 are each independently C_{10}-C_{30}alkyl, C_{10}-C_{30}alkenyl, or C_{10}-C_{30}alkynyl, C_{10}- C_{30}alkyl, C_{10}- C_{20}alkyl, C_{12}- C_{18}alkyl, C_{13}- C_{17}alkyl, C_{13}alkyl, C_{10}-C_{30}alkenyl, C_{10}-C_{20}oalkenyl. C_{12}-C_{18}alkenyl, C_{13}-C_{17}alkenyl, C_{17}alkenyl; R3 and R4 are each independently hydrogen, C_1-C_6 alkyl, or $–CH_2CH_2OH$, C_1-C_6 alkyl, C_1-C_3alkyl; n is 1 -6; and X is a counterion, including any nitrogen counterion, as that term is readily understood in the art, and specific cationic lipids including



("HEDC"),



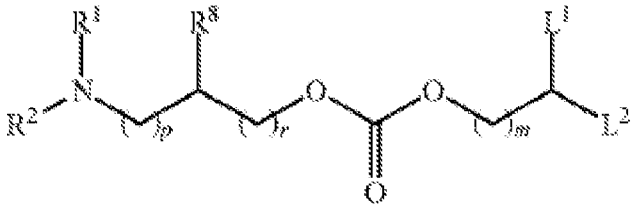
("HEDODC") and



("HE-Et-DODC")

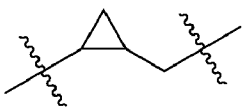
,

WO2013/093648 also provides examples of other cationic charged lipids at physiological pH including N,N-dioleyl-N,N-dimethylammonium chloride (DODAC), N,N-distearyl-N,N-dimethylammonium bromide (DDAB); N-(1,2-dimyristyloxyprop-3-yl)-N,N-dimethyl-N-hydroxyethylammonium bromide (DMRIE) and dioctadecylamidoglycyl carboxyspermidine (DOGS);in US 20160257951, which provides cationic lipids with a general formula

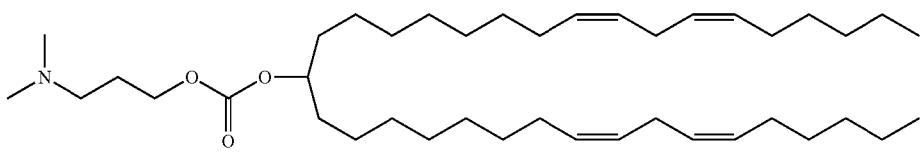


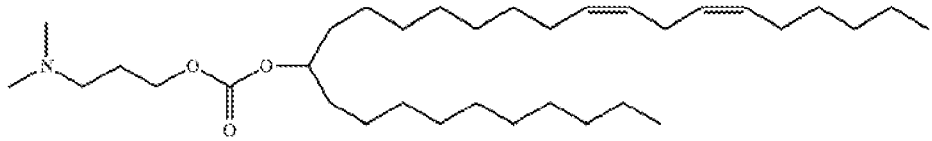
or a pharmacologically acceptable salt thereof, wherein R^1 and R^2 are each independently a hydrogen atom, a C_1-C_6 alkyl group optionally substituted with one or more substituents selected from substituent group α, a C_2-C_6 alkenyl group optionally substituted with one or more substituents selected from substituent group α, a C_2-C_6 alkynyl group optionally substituted with one or more substituents selected from substituent group α, or a C_3-C_7 cycloalkyl group optionally substituted with one or more substituents selected from substituent group α, or R^1 and R^2 form a 3- to 10-membered heterocyclic ring together with the nitrogen atom bonded thereto, wherein the heterocyclic ring is optionally substituted with one or more substituents selected from substituent group α and optionally contains one or more atoms selected from a nitrogen atom, an oxygen atom, and a sulfur atom, in addition to the nitrogen atom bonded to R^1 and R^2, as atoms constituting the heterocyclic ring; R^8 is a hydrogen atom or a C_1-C_6 alkyl group optionally substituted with one or more substituents selected from substituent group α; or R^1 and R^8 together are the group —$(CH_2)_q$—; substituent group α consists of a halogen atom, an oxo group, a hydroxy group, a sulfanyl group, an amino group, a cyano group, a C_1-C_6 alkyl group, a C_1-C_6 halogenated alkyl group, a C_1-C_6 alkoxy group, a C_1-C_6 alkylsulfanyl group, a C_1-C_6 alkylamino group, and a C_1-C_7 alkanoyl group; L^1 is a C_{10}-C_{24} alkyl group optionally substituted with one or more substituents selected from substituent group β1, a C_{10}-C_{24} alkenyl group optionally substituted

with one or more substituents selected from substituent group β1, a C_3-C_{24} alkynyl group optionally substituted with one or more substituents selected from substituent group β1, or a (C_1-C_{10} alkyl)-$(Q)_k$-(C_1-C_{10} alkyl) group optionally substituted with one or more substituents selected from substituent group β1; L^2 is, independently of L^1, a C_{10}-C_{24} alkyl group optionally substituted with one or more substituents selected from substituent group β1, a C_{10}-C_{24} alkenyl group optionally substituted with one or more substituents selected from substituent group β1, a C_3-C_{24} alkynyl group optionally substituted with one or more substituents selected from substituent group β1, a (C_1-C_{10} alkyl)-$(Q)_k$-(C_1-C_{10} alkyl) group optionally substituted with having one or more substituents selected from substituent group β1, a (C_{10}-C_{24} alkoxy)methyl group optionally substituted with one or more substituents selected from substituent group β1, a (C_{10}-C_{24} alkenyl)oxymethyl group optionally substituted with one or more substituents selected from substituent group β1, a (C_3-C_{24} alkynyl)oxymethyl group optionally substituted with one or more substituents selected from substituent group β1, or a (C_1-C_{10} alkyl)-$(Q)_k$-(C_1-C_{10} alkoxy)methyl group optionally substituted with one or more substituents selected from substituent group β1; substituent group β1 consists of a halogen atom, an oxo group, a cyano group, a C_1-C_6 alkyl group, a C_1-C_6 halogenated alkyl group, a C_1-C_6 alkoxy group, a C_1-C_6 alkylsulfanyl group, a C_1-C_7 alkanoyl group, a C_1-C_7 alkanoyloxy group, a C_3-C_7 alkoxyalkoxy group, a (C_1-C_6 alkoxy)carbonyl group, a (C_1-C_6 alkoxy)carboxyl group, a (C_1-C_6 alkoxy)carbamoyl group, and a (C_1-C_6 alkylamino)carboxyl group; Q is a group of formula:

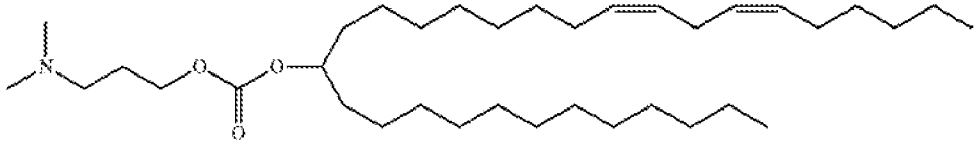


when L^1 and L^2 are each substituted with one or more substituents selected from substituent group β1 and substituent group β1 is a C_1-C_6 alkyl group, a C_1-C_6 alkoxy group, a C_1-C_6 alkylsulfanyl group, a C_1-C_7 alkanoyl group, or a C_1-C_7 alkanoyloxy group, the substituent or substituents selected from substituent group β1 in L^1 and the substituent or substituents selected from substituent group β1 in L^2 optionally bind to each other to form a cyclic structure; k is 1, 2, 3, 4, 5, 6, or 7; m is 0 or 1; p is 0, 1, or 2; q is 1, 2, 3, or 4; and r is 0, 1, 2, or 3, provided that p+r is 2 or larger, or q+r is 2 or larger, and specific cationic lipids including

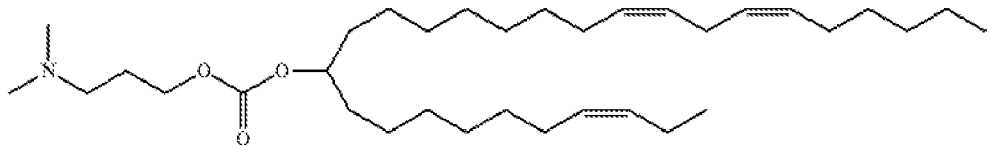




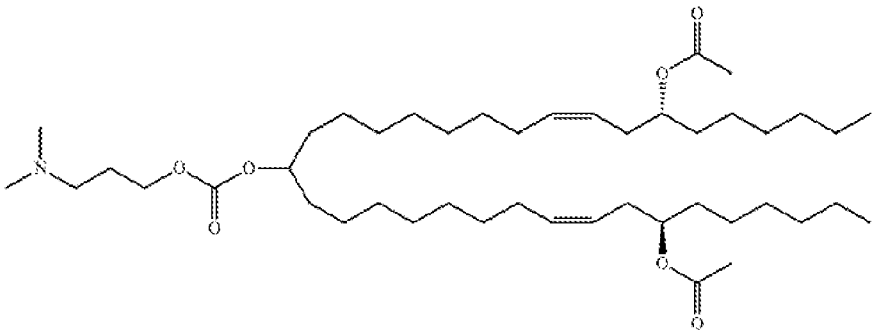
,



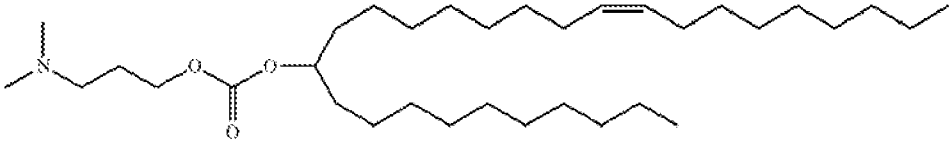
,



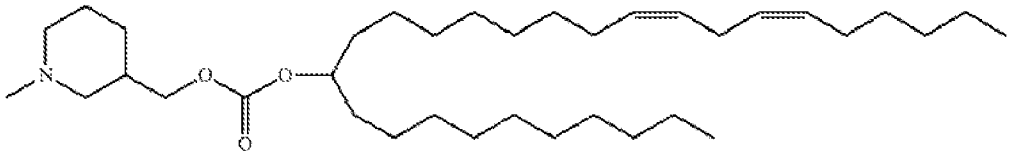
,



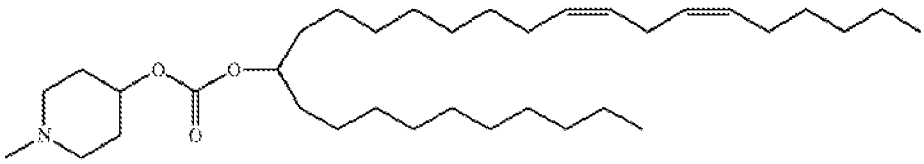
,



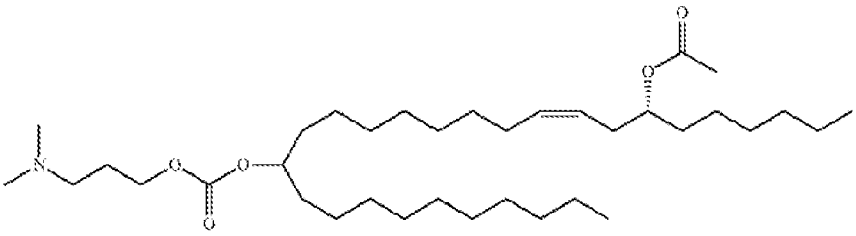
,



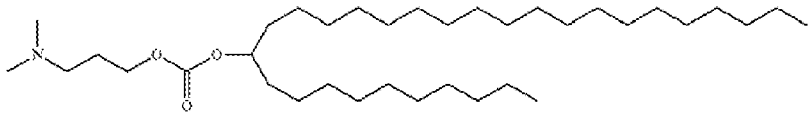
,



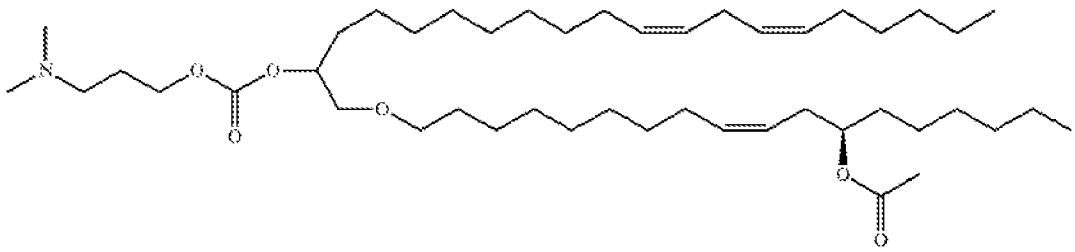
,



,



,



and

; and in US 20160244761, which provides cationic lipids that include 1,2-distearyloxy-N,N-dimethyl-3-aminopropane (DSDMA), 1,2-dioleyloxy-N,N-dimethyl-3-aminopropane (DODMA), 1,2-dilinoleyloxy-N,N-dimethyl-3-aminopropane (DLinDMA), 1,2-dilinolenyloxy-N,N-dimethyl-3-aminopropane (DLenDMA), 1,2-di-.gamma.-linolenyloxy-N,N-dimethylaminopropane (.gamma.-DLenDMA), 1,2-dilinoleyloxy-keto-N,N-dimethyl-3-aminopropane (DLin-K-DMA), 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLin-K--C2-DMA) (also known as DLin-C2K-DMA, XTC2, and C2K), 2,2-dilinoleyl-4-(3-dimethylaminopropyl)-[1,3]-dioxolane (DLin-K-C3-DMA), 2,2-dilinoleyl-4-(4-dimethylaminobutyl)-[1,3]-dioxolane (DLin-K--C4-DMA), 1,2-dilinolenyloxy-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLen-C2K-DMA), 1,2-di-.gamma.-linolenyloxy-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (.gamma.-DLen-C2K-DMA), dilinoleylmethyl-3-dimethylaminopropionate (DLin-M-C2-DMA) (also known as MC2), (6Z,9Z,28Z,31Z)-heptatriaconta-6,9,28,31-tetraen-19-yl 4-(dimethylamino) butanoate (DLin-M-C3-DMA) (also known as MC3) and 3-(dilinoleylmethoxy)-N,N-dimethylpropan-1-amine (DLin-MP-DMA) (also known as 1-Bl 1).

[0667] In one embodiment, the lipid compound is preferably a bio-reducible material, e.g., a bio-reducible polymer and a bio-reducible lipid-like compound.

[0668] In embodiment, the lipid compound comprises a hydrophilic head, and a hydrophobic tail, and optionally a linker.

[0669] In one embodiment, the hydrophilic head contains one or more hydrophilic functional groups, e.g., hydroxyl, carboxyl, amino, sulfhydryl, phosphate, amide, ester, ether, carbamate, carbonate, carbamide and phosphodiester. These groups can form hydrogen bonds and are optionally positively or negatively charged, in particular at physiological conditions such as physiological pH.

[0670] In one embodiment, the hydrophobic tail is a saturated or unsaturated, linear or branched, acyclic or cyclic, aromatic or nonaromatic hydrocarbon moiety, wherein the saturated or unsaturated, linear or branched, acyclic or cyclic, aromatic or nonaromatic hydrocarbon moiety optionally contains a disulfide bond and/or 8-24 carbon atoms. One or more of the carbon atoms can be replaced with a heteroatom, such as N, O, P, B, S, Si, Sb, Al, Sn, As, Se, and Ge. The lipid or lipid-like compounds containing disulfide bond can be bioreducible.

[0671] In one embodiment, the linker of the lipid or lipid-like compound links the hydrophilic head and the hydrophobic tail. The linker can be any chemical group that is hydrophilic or hydrophobic, polar or non-polar, e.g., O, S, Si, amino, alkylene, ester, amide, carbamate, carbamide, carbonate phosphate, phosphite, sulfate, sulfite, and thiosulfate.

[0672] The lipid or lipid-like compounds described above include the compounds themselves, as well as their salts and solvates, if applicable. A salt, for example, can be formed between an anion and a positively charged group (e.g., amino) on a lipid-like compound. Suitable anions include chloride, bromide, iodide, sulfate, nitrate, phosphate, citrate, methanesulfonate, trifluoroacetate, acetate, malate, tosylate, tartrate, fumurate, glutamate, glucuronate, lactate, glutarate, and maleate. Likewise, a salt can also be formed between a cation and a negatively charged group (e.g., carboxylate) on a lipid-like compound. Suitable cations include sodium ion, potassium ion, magnesium ion, calcium ion, and an ammonium cation such as tetramethylammonium ion. The lipid-like compounds also include those salts containing quaternary nitrogen atoms. A solvate refers to a complex formed between a lipid-like compound and a pharmaceutically acceptable solvent. Examples of pharmaceutically acceptable solvents include water, ethanol, isopropanol, ethyl acetate, acetic acid, and ethanolamine.

[0673] Delivery or administration according to the invention can be performed with liposomes. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes have gained considerable attention as drug delivery carriers because they are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB) (see, e.g., Spuch and

Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0674] Liposomes can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes as drug carriers. Although liposome formation is spontaneous when a lipid film is mixed with an aqueous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0675] Several other additives may be added to liposomes in order to modify their structure and properties. For instance, either cholesterol or sphingomyelin may be added to the liposomal mixture in order to help stabilize the liposomal structure and to prevent the leakage of the liposomal inner cargo. Further, liposomes are prepared from hydrogenated egg phosphatidylcholine or egg phosphatidylcholine, cholesterol, and dicetyl phosphate, and their mean vesicle sizes were adjusted to about 50 and 100 nm. (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0676] A liposome formulation may be mainly comprised of natural phospholipids and lipids such as 1,2-distearoryl-sn-glycero-3-phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines and monosialoganglioside. Since this formulation is made up of phospholipids only, liposomal formulations have encountered many challenges, one of the ones being the instability in plasma. Several attempts to overcome these challenges have been made, specifically in the manipulation of the lipid membrane. One of these attempts focused on the manipulation of cholesterol. Addition of cholesterol to conventional formulations reduces rapid release of the encapsulated bioactive compound into the plasma or 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) increases the stability (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0677] In a particularly advantageous embodiment, Trojan Horse liposomes (also known as Molecular Trojan Horses) are desirable and protocols may be found at http://cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long. These particles allow delivery of a transgene to the entire brain after an intravascular injection. Without being bound by limitation, it is believed that neutral lipid particles with specific antibodies conjugated to

surface allow crossing of the blood brain barrier via endocytosis. In some embodiments, Trojan Horse Liposomes can be used to deliver the CRISPR family of nucleases to the brain via an intravascular injection, which would allow whole brain transgenic animals without the need for embryonic manipulation. About 1-5 g of DNA or RNA may be contemplated for *in vivo* administration in liposomes.

[0678] In another embodiment, the CRISPR Cas system or components thereof may be administered in liposomes, such as a stable nucleic-acid-lipid particle (SNALP) (see, e.g., Morrissey et al., Nature Biotechnology, Vol. 23, No. 8, August 2005). Daily intravenous injections of about 1, 3 or 5 mg/kg/day of a specific CRISPR Cas targeted in a SNALP are contemplated. The daily treatment may be over about three days and then weekly for about five weeks. In another embodiment, a specific CRISPR Cas encapsulated SNALP) administered by intravenous injection to at doses of about 1 or 2.5 mg/kg are also contemplated (see, e.g., Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006). The SNALP formulation may contain the lipids 3-N-[(wmethoxypoly(ethylene glycol) 2000) carbamoyl] -1,2-dimyristyloxy-propylamine (PEG-C-DMA), 1,2-dilinoleyloxy-N,N-dimethyl-3-aminopropane (DLinDMA), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC) and cholesterol, in a 2:40:10:48 molar per cent ratio (see, e.g., Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006).

[0679] In another embodiment, stable nucleic-acid-lipid particles (SNALPs) have proven to be effective delivery molecules to highly vascularized HepG2-derived liver tumors but not in poorly vascularized HCT-116 derived liver tumors (see, e.g., Li, Gene Therapy (2012) 19, 775–780). The SNALP liposomes may be prepared by formulating D-Lin-DMA and PEG-C-DMA with distearoylphosphatidylcholine (DSPC), Cholesterol and siRNA using a 25:1 lipid/siRNA ratio and a 48/40/10/2 molar ratio of Cholesterol/D-Lin-DMA/DSPC/PEG-C-DMA. The resulted SNALP liposomes are about 80–100 nm in size.

[0680] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich, St Louis, MO, USA), dipalmitoylphosphatidylcholine (Avanti Polar Lipids, Alabaster, AL, USA), 3-N-[(w-methoxy poly(ethylene glycol)2000)carbamoyl]-1,2-dimyrestyloxypropylamine, and cationic 1,2-dilinoleyloxy-3-N,Ndimethylaminopropane (see, e.g., Geisbert et al., Lancet 2010; 375: 1896-905). A dosage of about 2 mg/kg total CRISPR Cas per dose administered as, for example, a bolus intravenous infusion may be contemplated.

[0681] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC; Avanti Polar Lipids Inc.), PEG-

cDMA, and 1,2-dilinoleyloxy-3-(N;N-dimethyl)aminopropane (DLinDMA) (see, e.g., Judge, J. Clin. Invest. 119:661-673 (2009)). Formulations used for *in vivo* studies may comprise a final lipid/RNA mass ratio of about 9:1.

[0682] The safety profile of RNAi nanomedicines has been reviewed by Barros and Gollob of Alnylam Pharmaceuticals (see, e.g., Advanced Drug Delivery Reviews 64 (2012) 1730–1737). The stable nucleic acid lipid particle (SNALP) is comprised of four different lipids — an ionizable lipid (DLinDMA) that is cationic at low pH, a neutral helper lipid, cholesterol, and a diffusible polyethylene glycol (PEG)-lipid. The particle is approximately 80 nm in diameter and is charge-neutral at physiologic pH. During formulation, the ionizable lipid serves to condense lipid with the anionic RNA during particle formation. When positively charged under increasingly acidic endosomal conditions, the ionizable lipid also mediates the fusion of SNALP with the endosomal membrane enabling release of RNA into the cytoplasm. The PEG-lipid stabilizes the particle and reduces aggregation during formulation, and subsequently provides a neutral hydrophilic exterior that improves pharmacokinetic properties.

[0683] To date, two clinical programs have been initiated using SNALP formulations with RNA. Tekmira Pharmaceuticals recently completed a phase I single-dose study of SNALP-ApoB in adult volunteers with elevated LDL cholesterol. ApoB is predominantly expressed in the liver and jejunum and is essential for the assembly and secretion of VLDL and LDL. Seventeen subjects received a single dose of SNALP-ApoB (dose escalation across 7 dose levels). There was no evidence of liver toxicity (anticipated as the potential dose-limiting toxicity based on preclinical studies). One (of two) subjects at the highest dose experienced flu-like symptoms consistent with immune system stimulation, and the decision was made to conclude the trial.

[0684] Alnylam Pharmaceuticals has similarly advanced ALN-TTR01, which employs the SNALP technology described above and targets hepatocyte production of both mutant and wild-type TTR to treat TTR amyloidosis (ATTR). Three ATTR syndromes have been described: familial amyloidotic polyneuropathy (FAP) and familial amyloidotic cardiomyopathy (FAC) — both caused by autosomal dominant mutations in TTR; and senile systemic amyloidosis (SSA) cause by wildtype TTR. A placebo-controlled, single dose-escalation phase I trial of ALN-TTR01 was recently completed in patients with ATTR. ALN-TTR01 was administered as a 15-minute IV infusion to 31 patients (23 with study drug and 8 with placebo) within a dose range of 0.01 to 1.0 mg/kg (based on siRNA). Treatment was well

tolerated with no significant increases in liver function tests. Infusion-related reactions were noted in 3 of 23 patients at≥0.4 mg/kg; all responded to slowing of the infusion rate and all continued on study. Minimal and transient elevations of serum cytokines IL-6, IP-10 and IL-1ra were noted in two patients at the highest dose of 1 mg/kg (as anticipated from preclinical and NHP studies). Lowering of serum TTR, the expected pharmacodynamics effect of ALN-TTR01, was observed at 1 mg/kg.

[0685] In yet another embodiment, a SNALP may be made by solubilizing a cationic lipid, DSPC, cholesterol and PEG-lipid e.g., in ethanol, e.g., at a molar ratio of 40:10:40:10, respectively (see, Semple et al., Nature Niotechnology, Volume 28 Number 2 February 2010, pp. 172-177). The lipid mixture was added to an aqueous buffer (50 mM citrate, pH 4) with mixing to a final ethanol and lipid concentration of 30% (vol/vol) and 6.1 mg/ml, respectively, and allowed to equilibrate at 22 °C for 2 min before extrusion. The hydrated lipids were extruded through two stacked 80 nm pore-sized filters (Nuclepore) at 22 °C using a Lipex Extruder (Northern Lipids) until a vesicle diameter of 70–90 nm, as determined by dynamic light scattering analysis, was obtained. This generally required 1–3 passes. The siRNA (solubilized in a 50 mM citrate, pH 4 aqueous solution containing 30% ethanol) was added to the pre-equilibrated (35 °C) vesicles at a rate of ~5 ml/min with mixing. After a final target siRNA/lipid ratio of 0.06 (wt/wt) was reached, the mixture was incubated for a further 30 min at 35 °C to allow vesicle reorganization and encapsulation of the siRNA. The ethanol was then removed and the external buffer replaced with PBS (155 mM NaCl, 3 mM Na_2HPO_4, 1 mM KH_2PO_4, pH 7.5) by either dialysis or tangential flow diafiltration. siRNA were encapsulated in SNALP using a controlled step-wise dilution method process. The lipid constituents of KC2-SNALP were DLin-KC2-DMA (cationic lipid), dipalmitoylphosphatidylcholine (DPPC; Avanti Polar Lipids), synthetic cholesterol (Sigma) and PEG-C-DMA used at a molar ratio of 57.1:7.1:34.3:1.4. Upon formation of the loaded particles, SNALP were dialyzed against PBS and filter sterilized through a 0.2 μm filter before use. Mean particle sizes were 75–85 nm and 90–95% of the siRNA was encapsulated within the lipid particles. The final siRNA/lipid ratio in formulations used for *in vivo* testing was ~0.15 (wt/wt). LNP-siRNA systems containing Factor VII siRNA were diluted to the appropriate concentrations in sterile PBS immediately before use and the formulations were administered intravenously through the lateral tail vein in a total volume of 10 ml/kg. This method and these delivery systems may be extrapolated to the CRISPR Cas system of the present invention.

Other Lipids

[0686] Other cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA) may be utilized to encapsulate CRISPR Cas or components thereof or nucleic acid molecule(s) coding therefor e.g., similar to SiRNA (see, e.g., Jayaraman, Angew. Chem. Int. Ed. 2012, 51, 8529 –8533), and hence may be employed in the practice of the invention. A preformed vesicle with the following lipid composition may be contemplated: amino lipid, distearoylphosphatidylcholine (DSPC), cholesterol and (R)-2,3-bis(octadecyloxy) propyl-1-(methoxy poly(ethylene glycol)2000)propylcarbamate (PEG-lipid) in the molar ratio 40/10/40/10, respectively, and a FVII siRNA/total lipid ratio of approximately 0.05 (w/w). To ensure a narrow particle size distribution in the range of 70–90 nm and a low polydispersity index of 0.11+0.04 (n=56), the particles may be extruded up to three times through 80 nm membranes prior to adding the guide RNA. Particles containing the highly potent amino lipid 16 may be used, in which the molar ratio of the four lipid components 16, DSPC, cholesterol and PEG-lipid (50/10/38.5/1.5) which may be further optimized to enhance *in vivo* activity.

[0687] Michael S D Kormann et al. ("Expression of therapeutic proteins after delivery of chemically modified mRNA in mice: Nature Biotechnology, Volume:29, Pages: 154–157 (2011)) describes the use of lipid envelopes to deliver RNA. Use of lipid envelopes is also preferred in the present invention.

[0688] In another embodiment, lipids may be formulated with the CRISPR Cas system of the present invention or component(s) thereof or nucleic acid molecule(s) coding therefor to form lipid nanoparticles (LNPs). Lipids include, but are not limited to, DLin-KC2-DMA4, C12-200 and colipids disteroylphosphatidyl choline, cholesterol, and PEG-DMG may be formulated with CRISPR Cas instead of siRNA (see, e.g., Novobrantseva, Molecular Therapy–Nucleic Acids (2012) 1, e4; doi:10.1038/mtna.2011.3) using a spontaneous vesicle formation procedure. The component molar ratio may be about 50/10/38.5/1.5 (DLin-KC2-DMA or C12-200/disteroylphosphatidyl choline/cholesterol/PEG-DMG). The final lipid:siRNA weight ratio may be ~12:1 and 9:1 in the case of DLin-KC2-DMA and C12-200 lipid nanoparticles (LNPs), respectively. The formulations may have mean particle diameters of ~80 nm with >90% entrapment efficiency. A 3 mg/kg dose may be contemplated.

[0689] Tekmira has a portfolio of approximately 95 patent families, in the U.S. and abroad, that are directed to various aspects of LNPs and LNP formulations (see, e.g., U.S. Pat. Nos. 7,982,027; 7,799,565; 8,058,069; 8,283,333; 7,901,708; 7,745,651; 7,803,397; 8,101,741;

8,188,263; 7,915,399; 8,236,943 and 7,838,658 and European Pat. Nos 1766035; 1519714; 1781593 and 1664316), all of which may be used and/or adapted to the present invention.

[0690] The CRISPR Cas system or components thereof or nucleic acid molecule(s) coding therefor may be delivered encapsulated in PLGA Microspheres such as that further described in US published applications 20130252281 and 20130245107 and 20130244279 (assigned to Moderna Therapeutics) which relate to aspects of formulation of compositions comprising modified nucleic acid molecules which may encode a protein, a protein precursor, or a partially or fully processed form of the protein or a protein precursor. The formulation may have a molar ratio 50:10:38.5:1.5-3.0 (cationic lipid:fusogenic lipid:cholesterol:PEG lipid). The PEG lipid may be selected from, but is not limited to PEG-c-DOMG, PEG-DMG. The fusogenic lipid may be DSPC. See also, Schrum et al., Delivery and Formulation of Engineered Nucleic Acids, US published application 20120251618.

[0691] Nanomerics' technology addresses bioavailability challenges for a broad range of therapeutics, including low molecular weight hydrophobic drugs, peptides, and nucleic acid based therapeutics (plasmid, siRNA, miRNA). Specific administration routes for which the technology has demonstrated clear advantages include the oral route, transport across the blood-brain-barrier, delivery to solid tumors, as well as to the eye. See, e.g., Mazza et al., 2013, ACS Nano. 2013 Feb 26;7(2):1016-26; Uchegbu and Siew, 2013, J Pharm Sci. 102(2):305-10 and Lalatsa et al., 2012, J Control Release. 2012 Jul 20; 161(2):523-36.

[0692] US Patent Publication No. 20050019923 describes cationic dendrimers for delivering bioactive molecules, such as polynucleotide molecules, peptides and polypeptides and/or pharmaceutical agents, to a mammalian body. The dendrimers are suitable for targeting the delivery of the bioactive molecules to, for example, the liver, spleen, lung, kidney or heart (or even the brain). Dendrimers are synthetic 3-dimensional macromolecules that are prepared in a step-wise fashion from simple branched monomer units, the nature and functionality of which can be easily controlled and varied. Dendrimers are synthesized from the repeated addition of building blocks to a multifunctional core (divergent approach to synthesis), or towards a multifunctional core (convergent approach to synthesis) and each addition of a 3-dimensional shell of building blocks leads to the formation of a higher generation of the dendrimers. Polypropylenimine dendrimers start from a diaminobutane core to which is added twice the number of amino groups by a double Michael addition of acrylonitrile to the primary amines followed by the hydrogenation of the nitriles. This results in a doubling of the amino

groups. Polypropylenimine dendrimers contain 100% protonable nitrogens and up to 64 terminal amino groups (generation 5, DAB 64). Protonable groups are usually amine groups which are able to accept protons at neutral pH. The use of dendrimers as gene delivery agents has largely focused on the use of the polyamidoamine. and phosphorous containing compounds with a mixture of amine/amide or N--P(O_2)S as the conjugating units respectively with no work being reported on the use of the lower generation polypropylenimine dendrimers for gene delivery. Polypropylenimine dendrimers have also been studied as pH sensitive controlled release systems for drug delivery and for their encapsulation of guest molecules when chemically modified by peripheral amino acid groups. The cytotoxicity and interaction of polypropylenimine dendrimers with DNA as well as the transfection efficacy of DAB 64 has also been studied.

[0693] US Patent Publication No. 20050019923 is based upon the observation that, contrary to earlier reports, cationic dendrimers, such as polypropylenimine dendrimers, display suitable properties, such as specific targeting and low toxicity, for use in the targeted delivery of bioactive molecules, such as genetic material. In addition, derivatives of the cationic dendrimer also display suitable properties for the targeted delivery of bioactive molecules. See also, Bioactive Polymers, US published application 20080267903, which discloses "Various polymers, including cationic polyamine polymers and dendrimeric polymers, are shown to possess anti-proliferative activity, and may therefore be useful for treatment of disorders characterized by undesirable cellular proliferation such as neoplasms and tumors, inflammatory disorders (including autoimmune disorders), psoriasis and atherosclerosis. The polymers may be used alone as active agents, or as delivery vehicles for other therapeutic agents, such as drug molecules or nucleic acids for gene therapy. In such cases, the polymers' own intrinsic anti-tumor activity may complement the activity of the agent to be delivered." The disclosures of these patent publications may be employed in conjunction with herein teachings for delivery of CRISPR Cas system(s) or component(s) thereof or nucleic acid molecule(s) coding therefor.

[0694] Additional delivery system formulations and particle types are also contemplated. In one embodiment, plasmids containing Cas9 and gRNA sequences can be delivered in a formulation that includes polyethyleneimine (PEI), specifically branched PEI 25 kD. *See, e.g.* Ryu et al., "Effective PEI-mediated delivery of CRISPR-Cas9 complex for targeted gene therapy" Nanomedicine. 2018 Oct;14(7):2095-2102. In another embodiment, plasmids can be

delivered in a formulation in which the plasmids are complexed with stearyl polyethylenimine as the core of human serum albumin nanoparticles noncovalently bound to CRISPR/Cas9 plasmids or siRNA for disrupting or silencing PD-L1 expression for immunotherapy. *See, e.g.* Cheng et al., Int J Nanomedicine. 2018 Nov 2;13:7079-7094. 2018. In another embodiment, plasmids containing Cas9 and gRNA sequences can be delivered in a hybrid nanoparticle produced by encapsulating the plasmid in an exosome, which is then fused with a liposome, to allow uptake by cell types that are not efficiently transfected using liposomes. *See, e.g.* Lin et al., "Exosome–Liposome Hybrid Nanoparticles Deliver CRISPR/Cas9 System in MSCs" Adv Sci (Weinh). 2018 Apr; 5(4): 1700611. In another embodiment, a Type VII effector plasmid can be delivered in a formulation that includes a novel fluorinated acid-labile branched hydroxyl-rich polycation (ARP-F). *See, e.g.* Qi et al., "Fluorinated Acid-Labile Branched Hydroxyl-Rich Nanosystems for Flexible and Robust Delivery of Plasmids." Small. 2018 Oct;14(42):e1803061.

Supercharged proteins

[0695] Supercharged proteins are a class of engineered or naturally occurring proteins with unusually high positive or negative net theoretical charge and may be employed in delivery of CRISPR Cas system(s) or component(s) thereof or nucleic acid molecule(s) coding therefor. Both supernegatively and superpositively charged proteins exhibit a remarkable ability to withstand thermally or chemically induced aggregation. Superpositively charged proteins are also able to penetrate mammalian cells. Associating cargo with these proteins, such as plasmid DNA, RNA, or other proteins, can enable the functional delivery of these macromolecules into mammalian cells both *in vitro* and *in vivo*. David Liu's lab reported the creation and characterization of supercharged proteins in 2007 (Lawrence et al., 2007, Journal of the American Chemical Society 129, 10110–10112).

[0696] The nonviral delivery of RNA and plasmid DNA into mammalian cells are valuable both for research and therapeutic applications (Akinc et al., 2010, Nat. Biotech. 26, 561–569). Purified +36 GFP protein (or other superpositively charged protein) is mixed with RNAs in the appropriate serum-free media and allowed to complex prior addition to cells. Inclusion of serum at this stage inhibits formation of the supercharged protein-RNA complexes and reduces the effectiveness of the treatment. The following protocol has been found to be effective for a variety of cell lines (McNaughton et al., 2009, Proc. Natl. Acad. Sci. USA 106, 6111–6116) (However, pilot experiments varying the dose of protein and RNA should be performed to

optimize the procedure for specific cell lines): (1) One day before treatment, plate 1 x 10^5 cells per well in a 48-well plate. (2) On the day of treatment, dilute purified +36 GFP protein in serumfree media to a final concentration 200nM. Add RNA to a final concentration of 50nM. Vortex to mix and incubate at room temperature for 10min. (3) During incubation, aspirate media from cells and wash once with PBS. (4) Following incubation of +36 GFP and RNA, add the protein-RNA complexes to cells. (5) Incubate cells with complexes at 37 °C for 4h. (6) Following incubation, aspirate the media and wash three times with 20 U/mL heparin PBS. Incubate cells with serum-containing media for a further 48h or longer depending upon the assay for activity. (7) Analyze cells by immunoblot, qPCR, phenotypic assay, or other appropriate method.

[0697] David Liu's lab has further found +36 GFP to be an effective plasmid delivery reagent in a range of cells. As plasmid DNA is a larger cargo than siRNA, proportionately more +36 GFP protein is required to effectively complex plasmids. For effective plasmid delivery a variant of +36 GFP bearing a C-terminal HA2 peptide tag can be used, a known endosome-disrupting peptide derived from the influenza virus hemagglutinin protein. The following protocol has been effective in a variety of cells, but as above it is advised that plasmid DNA and supercharged protein doses be optimized for specific cell lines and delivery applications: (1) One day before treatment, plate 1 x 10^5 per well in a 48-well plate. (2) On the day of treatment, dilute purified þ36 GFP protein in serum-free media to a final concentration 2 mM. Add 1mg of plasmid DNA. Vortex to mix and incubate at room temperature for 10min. (3) During incubation, aspirate media from cells and wash once with PBS. (4) Following incubation of þ36 GFP and plasmid DNA, gently add the protein-DNA complexes to cells. (5) Incubate cells with complexes at 37 C for 4h. (6) Following incubation, aspirate the media and wash with PBS. Incubate cells in serum-containing media and incubate for a further 24–48h. (7) Analyze plasmid delivery (e.g., by plasmid-driven gene expression) as appropriate.

[0698] See also, e.g., McNaughton et al., Proc. Natl. Acad. Sci. USA 106, 6111-6116 (2009); Cronican et al., ACS Chemical Biology 5, 747-752 (2010); Cronican et al., Chemistry & Biology 18, 833-838 (2011); Thompson et al., Methods in Enzymology 503, 293-319 (2012); Thompson, D.B., et al., Chemistry & Biology 19 (7), 831-843 (2012). The methods of the super charged proteins may be used and/or adapted for delivery of the CRISPR Cas system of the present invention. These systems of Dr. Lui and documents herein in conjunction with

herein teaching can be employed in the delivery of CRISPR Cas system(s) or component(s) thereof or nucleic acid molecule(s) coding therefor.

Cell Penetrating Peptides (CPPs)

[0699] In yet another embodiment, cell penetrating peptides (CPPs) are contemplated for the delivery of the CRISPR Cas system. CPPs are short peptides that facilitate cellular uptake of various molecular cargo (from nanosize particles to small chemical molecules and large fragments of DNA). The term "cargo" as used herein includes but is not limited to the group consisting of therapeutic agents, diagnostic probes, peptides, nucleic acids, antisense oligonucleotides, plasmids, proteins, particles, including nanoparticles, liposomes, chromophores, small molecules and radioactive materials. In aspects of the invention, the cargo may also comprise any component of the CRISPR Cas system or the entire functional CRISPR Cas system. Aspects of the present invention further provide methods for delivering a desired cargo into a subject comprising: (a) preparing a complex comprising the cell penetrating peptide of the present invention and a desired cargo, and (b) orally, intraarticularly, intraperitoneally, intrathecally, intrarterially, intranasally, intraparenchymally, subcutaneously, intramuscularly, intravenously, dermally, intrarectally, or topically administering the complex to a subject. The cargo is associated with the peptides either through chemical linkage via covalent bonds or through non-covalent interactions.

[0700] The function of the CPPs are to deliver the cargo into cells, a process that commonly occurs through endocytosis with the cargo delivered to the endosomes of living mammalian cells. Cell-penetrating peptides are of different sizes, amino acid sequences, and charges but all CPPs have one distinct characteristic, which is the ability to translocate the plasma membrane and facilitate the delivery of various molecular cargoes to the cytoplasm or an organelle. CPP translocation may be classified into three main entry mechanisms: direct penetration in the membrane, endocytosis-mediated entry, and translocation through the formation of a transitory structure. CPPs have found numerous applications in medicine as drug delivery agents in the treatment of different diseases including cancer and virus inhibitors, as well as contrast agents for cell labeling. Examples of the latter include acting as a carrier for GFP, MRI contrast agents, or quantum dots. CPPs hold great potential as *in vitro* and *in vivo* delivery vectors for use in research and medicine. CPPs typically have an amino acid composition that either contains a high relative abundance of positively charged amino acids such as lysine or arginine or has sequences that contain an alternating pattern of polar/charged

amino acids and non-polar, hydrophobic amino acids. These two types of structures are referred to as polycationic or amphipathic, respectively. A third class of CPPs are the hydrophobic peptides, containing only apolar residues, with low net charge or have hydrophobic amino acid groups that are crucial for cellular uptake. One of the initial CPPs discovered was the trans-activating transcriptional activator (Tat) from Human Immunodeficiency Virus 1 (HIV-1) which was found to be efficiently taken up from the surrounding media by numerous cell types in culture. Since then, the number of known CPPs has expanded considerably and small molecule synthetic analogues with more effective protein transduction properties have been generated. CPPs include but are not limited to Penetratin, Tat (48-60), Transportan, and (R-AhX-R4) (Ahx=aminohexanoyl).

[0701] US Patent 8,372,951, provides a CPP derived from eosinophil cationic protein (ECP) which exhibits highly cell-penetrating efficiency and low toxicity. Aspects of delivering the CPP with its cargo into a vertebrate subject are also provided. Further aspects of CPPs and their delivery are described in U. S. patents 8,575,305; 8;614,194 and 8,044,019. CPPs can be used to deliver the CRISPR-Cas system or components thereof. That CPPs can be employed to deliver the CRISPR-Cas system or components thereof is also provided in the manuscript "Gene disruption by cell-penetrating peptide-mediated delivery of Cas9 protein and guide RNA", by Suresh Ramakrishna, Abu-Bonsrah Kwaku Dad, Jagadish Beloor, et al. Genome Res. 2014 Apr 2. [Epub ahead of print], incorporated by reference in its entirety, wherein it is demonstrated that treatment with CPP-conjugated recombinant Cas9 protein and CPP-complexed guide RNAs lead to endogenous gene disruptions in human cell lines. In the paper the Cas9 protein was conjugated to CPP via a thioether bond, whereas the guide RNA was complexed with CPP, forming condensed, positively charged particles. It was shown that simultaneous and sequential treatment of human cells, including embryonic stem cells, dermal fibroblasts, HEK293T cells, HeLa cells, and embryonic carcinoma cells, with the modified Cas9 and guide RNA led to efficient gene disruptions with reduced off-target mutations relative to plasmid transfections.

Implantable devices

[0702] In another embodiment, implantable devices are also contemplated for delivery of the CRISPR Cas system or component(s) thereof or nucleic acid molecule(s) coding therefor. For example, US Patent Publication 20110195123 discloses an implantable medical device which elutes a drug locally and in prolonged period is provided, including several types of such

a device, the treatment modes of implementation and methods of implantation. The device comprising of polymeric substrate, such as a matrix for example, that is used as the device body, and drugs, and in some cases additional scaffolding materials, such as metals or additional polymers, and materials to enhance visibility and imaging. An implantable delivery device can be advantageous in providing release locally and over a prolonged period, where drug is released directly to the extracellular matrix (ECM) of the diseased area such as tumor, inflammation, degeneration or for symptomatic objectives, or to injured smooth muscle cells, or for prevention. One kind of drug is RNA, as disclosed above, and this system may be used/and or adapted to the CRISPR Cas system of the present invention. The modes of implantation in some embodiments are existing implantation procedures that are developed and used today for other treatments, including brachytherapy and needle biopsy. In such cases the dimensions of the new implant described in this invention are similar to the original implant. Typically a few devices are implanted during the same treatment procedure.

[0703] US Patent Publication 20110195123, provides a drug delivery implantable or insertable system, including systems applicable to a cavity such as the abdominal cavity and/or any other type of administration in which the drug delivery system is not anchored or attached, comprising a biostable and/or degradable and/or bioabsorbable polymeric substrate, which may for example optionally be a matrix. It should be noted that the term "insertion" also includes implantation. The drug delivery system is preferably implemented as a "Loder" as described in US Patent Publication 20110195123.

[0704] The polymer or plurality of polymers are biocompatible, incorporating an agent and/or plurality of agents, enabling the release of agent at a controlled rate, wherein the total volume of the polymeric substrate, such as a matrix for example, in some embodiments is optionally and preferably no greater than a maximum volume that permits a therapeutic level of the agent to be reached. As a non-limiting example, such a volume is preferably within the range of 0.1 m^3 to 1000 mm^3, as required by the volume for the agent load. The Loder may optionally be larger, for example when incorporated with a device whose size is determined by functionality, for example and without limitation, a knee joint, an intra-uterine or cervical ring and the like.

[0705] The drug delivery system (for delivering the composition) is designed in some embodiments to preferably employ degradable polymers, wherein the main release mechanism is bulk erosion; or in some embodiments, non-degradable, or slowly degraded polymers are

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

[0706] The drug delivery system optionally and preferably is designed to shield the nucleotide based therapeutic agent from degradation, whether chemical in nature or due to attack from enzymes and other factors in the body of the subject.

[0707] The drug delivery system of US Patent Publication 20110195123 is optionally associated with sensing and/or activation appliances that are operated at and/or after implantation of the device, by non- and/or minimally invasive methods of activation and/or acceleration/deceleration, for example optionally including but not limited to thermal heating and cooling, laser beams, and ultrasonic, including focused ultrasound and/or RF (radiofrequency) methods or devices.

[0708] According to some embodiments of US Patent Publication 20110195123, the site for local delivery may optionally include target sites characterized by high abnormal proliferation of cells, and suppressed apoptosis, including tumors, active and or chronic inflammation and infection including autoimmune diseases states, degenerating tissue including muscle and nervous tissue, chronic pain, degenerative sites, and location of bone fractures and other wound locations for enhancement of regeneration of tissue, and injured cardiac, smooth and striated muscle.

[0709] The site for implantation of the composition, or target site, preferably features a radius, area and/or volume that is sufficiently small for targeted local delivery. For example, the target site optionally has a diameter in a range of from about 0.1 mm to about 5 cm.

[0710] The location of the target site is preferably selected for maximum therapeutic efficacy. For example, the composition of the drug delivery system (optionally with a device for implantation as described above) is optionally and preferably implanted within or in the proximity of a tumor environment, or the blood supply associated thereof.

[0711] For example, the composition (optionally with the device) is optionally implanted within or in the proximity to pancreas, prostate, breast, liver, via the nipple, within the vascular system and so forth.

[0712] The target location is optionally selected from the group comprising, consisting essentially of, or consisting of (as non-limiting examples only, as optionally any site within the body may be suitable for implanting a Loder): 1. brain at degenerative sites like in Parkinson or Alzheimer disease at the basal ganglia, white and gray matter; 2. spine as in the case of amyotrophic lateral sclerosis (ALS); 3. uterine cervix to prevent HPV infection; 4. active and chronic inflammatory joints; 5. dermis as in the case of psoriasis; 6. sympathetic and sensoric nervous sites for analgesic effect; 7. Intra osseous implantation; 8. acute and chronic infection sites; 9. Intra vaginal; 10. Inner ear--auditory system, labyrinth of the inner ear, vestibular system; 11. Intra tracheal; 12. Intra-cardiac; coronary, epicardiac; 13. urinary bladder; 14. biliary system; 15. parenchymal tissue including and not limited to the kidney, liver, spleen; 16. lymph nodes; 17. salivary glands; 18. dental gums; 19. Intra-articular (into joints); 20. Intra-ocular; 21. Brain tissue; 22. Brain ventricles; 23. Cavities, including abdominal cavity (for example but without limitation, for ovary cancer); 24. Intra esophageal and 25. Intra rectal.

[0713] Optionally insertion of the system (for example a device containing the composition) is associated with injection of material to the ECM at the target site and the vicinity of that site to affect local pH and/or temperature and/or other biological factors affecting the diffusion of the drug and/or drug kinetics in the ECM, of the target site and the vicinity of such a site.

[0714] Optionally, according to some embodiments, the release of said agent could be associated with sensing and/or activation appliances that are operated prior and/or at and/or after insertion, by non- and/or minimally invasive and/or else methods of activation and/or acceleration/deceleration, including laser beam, radiation, thermal heating and cooling, and ultrasonic, including focused ultrasound and/or RF (radiofrequency) methods or devices, and chemical activators.

[0715] According to other embodiments of US Patent Publication 20110195123, the drug preferably comprises a RNA, for example for localized cancer cases in breast, pancreas, brain, kidney, bladder, lung, and prostate as described below. Although exemplified with RNAi, many drugs are applicable to be encapsulated in Loder, and can be used in association with this invention, as long as such drugs can be encapsulated with the Loder substrate, such as a matrix for example, and this system may be used and/or adapted to deliver the CRISPR Cas system of the present invention.

[0716] As another example of a specific application, neuro and muscular degenerative diseases develop due to abnormal gene expression. Local delivery of RNAs may have therapeutic properties for interfering with such abnormal gene expression. Local delivery of anti-apoptotic, anti-inflammatory and anti-degenerative drugs including small drugs and macromolecules may also optionally be therapeutic. In such cases the Loder is applied for prolonged release at constant rate and/or through a dedicated device that is implanted separately. All of this may be used and/or adapted to the CRISPR Cas system of the present invention.

[0717] As yet another example of a specific application, psychiatric and cognitive disorders are treated with gene modifiers. Gene knockdown is a treatment option. Loders locally delivering agents to central nervous system sites are therapeutic options for psychiatric and cognitive disorders including but not limited to psychosis, bi-polar diseases, neurotic disorders and behavioral maladies. The Loders could also deliver locally drugs including small drugs and macromolecules upon implantation at specific brain sites. All of this may be used and/or adapted to the CRISPR Cas system of the present invention.

[0718] As another example of a specific application, silencing of innate and/or adaptive immune mediators at local sites enables the prevention of organ transplant rejection. Local delivery of RNAs and immunomodulating reagents with the Loder implanted into the transplanted organ and/or the implanted site renders local immune suppression by repelling immune cells such as CD8 activated against the transplanted organ. All of this may be used/and or adapted to the CRISPR Cas system of the present invention.

[0719] As another example of a specific application, vascular growth factors including VEGFs and angiogenin and others are essential for neovascularization. Local delivery of the factors, peptides, peptidomimetics, or suppressing their repressors is an important therapeutic modality; silencing the repressors and local delivery of the factors, peptides, macromolecules

and small drugs stimulating angiogenesis with the Loder is therapeutic for peripheral, systemic and cardiac vascular disease.

[0720] The method of insertion, such as implantation, may optionally already be used for other types of tissue implantation and/or for insertions and/or for sampling tissues, optionally without modifications, or alternatively optionally only with non-major modifications in such methods. Such methods optionally include but are not limited to brachytherapy methods, biopsy, endoscopy with and/or without ultrasound, such as ERCP, stereotactic methods into the brain tissue, Laparoscopy, including implantation with a laparoscope into joints, abdominal organs, the bladder wall and body cavities.

[0721] Implantable devices may also include cells, such as epidermal progenitor cells that have been edited or modified to express the CRISPR-Cas systems disclosed herein and embedded with an implantable device, such as a patch.. *See*. Yue et al. "Engineered Epidermal Progenitor Cells Can Correct Diet-Induced Obesity and Diabetes" Cell Stem Cell (2017) 21(2):256-263.

[0722] Implantable device technology herein discussed can be employed with herein teachings and hence by this disclosure and the knowledge in the art, CRISPR-Cas system or components thereof or nucleic acid molecules thereof or encoding or providing components may be delivered via an implantable device.

Aerosol delivery

[0723] 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.

Hybrid Viral Capsid Delivery Systems

[0724] In one aspect, the invention provides a particle delivery system comprising a hybrid virus capsid protein or hybrid viral outer protein, wherein the hybrid virus capsid or outer protein comprises a virus capsid or outer protein attached to at least a portion of a non-capsid protein or peptide. The genetic material of a virus is stored within a viral structure called the capsid. The capsid of certain viruses are enclosed in a membrane called the viral envelope. The viral envelope is made up of a lipid bilayer embedded with viral proteins including viral

glycoproteins. 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. Non-limiting examples of outer or envelope proteins include, without limit, gp41 and gp120 of HIV, hemagglutinin, neuraminidase and M2 proteins of influenza virus.

[0725] In one example embodiment of the delivery system, the non-capsid protein or peptide has a molecular weight of up to a megadalton, or has a molecular weight in the range of 110 to 160 kDa, 160 to 200 kDa, 200 to 250 kDa, 250 to 300 kDa, 300 to 400 kDa, or 400 to 500 kDa, the non-capsid protein or peptide comprises a CRISPR protein.

[0726] The present application provides a vector for delivering an effector protein and at least one CRISPR guide RNA to a cell comprising 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 virus is an adeno-associated virus (AAV) or an adenovirus.

[0727] In a related aspect, the invention provides a lentiviral vector for delivering an effector protein and at least one CRISPR guide RNA to a cell comprising a promoter operably linked to a polynucleotide sequence encoding a Type VII effector and a second promoter operably linked to a polynucleotide sequence encoding at least one guide RNA, wherein the polynucleotide sequences are in reverse orientation.

[0728] In an embodiment of the delivery system, the virus is lentivirus or murine leukemia virus (MuMLV). In an embodiment of the delivery system, the virus is an Adenoviridae or a Parvoviridae or a retrovirus or a Rhabdoviridae or an enveloped virus having a glycoprotein protein (G protein). In an embodiment of the delivery system, the virus is VSV or rabies virus. In an embodiment of the delivery system, the capsid or outer protein comprises a capsid protein having VP1, VP2 or VP3. In an embodiment of the delivery system, the capsid protein is VP3, and the non-capsid protein is inserted into or attached to VP3 loop 3 or loop 6.

[0729] In an embodiment of the delivery system, the virus is delivered to the interior of a cell. In an embodiment of the delivery system, the capsid or outer protein and the non-capsid protein can dissociate after delivery into a cell.

[0730] In an embodiment of the delivery system, the capsid or outer protein is attached to the protein by a linker. In an embodiment of the delivery system, the linker comprises amino acids. In an embodiment of the delivery system, the linker is a chemical linker. In an embodiment of the delivery system, the linker is cleavable. In an embodiment of the delivery system, the linker is biodegradable. In an embodiment of the delivery system, the linker comprises (GGGGS)1-3 (**SEQ ID NOs: 21 - 23**), ENLYFQG (**SEQ ID NO: 24**), or a disulfide.

[0731] In an embodiment, the delivery system comprises a protease or nucleic acid molecule(s) encoding a protease that is expressed, said protease being capable of cleaving the linker, whereby there can be cleavage of the linker. In an embodiment of the invention, a protease is delivered with a particle component of the system, for example packaged, mixed with, or enclosed by lipid and or capsid. Entry of the particle into a cell is thereby accompanied or followed by cleavage and dissociation of payload from particle. In certain embodiments, an expressible nucleic acid encoding a protease is delivered, whereby at entry or following entry of the particle into a cell, there is protease expression, linker cleavage, and dissociation of payload from capsid. In certain embodiments, dissociation of payload occurs with viral replication. In certain embodiments, dissociation of payload occurs in the absence of productive virus replication.

[0732] In an embodiment of the delivery system, each terminus of a CRISPR protein is attached to the capsid or outer protein by a linker. In an embodiment of the delivery system, the non-capsid protein is attached to the exterior portion of the capsid or outer protein. In an embodiment of the delivery system, the non-capsid protein is attached to the interior portion of the capsid or outer protein. In an embodiment of the delivery system, the capsid or outer protein and the non-capsid protein are a fusion protein. In an embodiment of the delivery system, the non-capsid protein is encapsulated by the capsid or outer protein. In an embodiment of the delivery system, the non-capsid protein is attached to a component of the capsid protein or a component of the outer protein prior to formation of the capsid or the outer protein. In an embodiment of the delivery system, the protein is attached to the capsid or outer protein after formation of the capsid or outer protein.

[0733] In an embodiment, the delivery system 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.

[0734] 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/nnm.13.118 (2013), and the documents it cites, all of which are incorporated herein by reference. 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. 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.

[0735] 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 aspects 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.

[0736] 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 non-internalizing 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 adenovirous 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 is 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.

[0737] 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. 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 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. 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. αβ-integrins or integrins are a group of transmembrane glycoprotein receptors that mediate attachment between a cell and its surrounding tissues or extracellular matrix. Integrins contain two distinct chains (heterodimers) called α- and β-subunits. The tumor tissue-specific expression of integrin receptors can be been utilized for targeted delivery in the invention, e.g., whereby the targeting moiety can be an RGD peptide such as a cyclic RGD. Aptamers are ssDNA or RNA oligonucleotides that impart high affinity and specific recognition of the target molecules by electrostatic interactions, hydrogen bonding and hydro phobic 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 as a 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). The targeting moiety 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)). 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. 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. 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) can be incorporated into a linker, and can have antibody targeting, e.g., antibody 2C5. 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 Fe_3O_4 or γ-Fe_2O_3, e.g., those that are less than 10 nm in size. Targeted delivery can be then by exposure to a magnetic field.

[0738] 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, 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.

[0739] As to active targeting, cell-penetrating peptides (CPPs) facilitate uptake of macromolecules through cellular membranes and, thus, enhance the delivery of CPP-modified molecules inside the cell. CPPs can be split into two classes: amphipathic helical peptides, such as transportan and MAP, where lysine residue are major contributors to the positive charge; and Arg-rich peptides, such as TATp, Antennapedia or penetratin. TATp is a transcription-activating factor with 86 amino acids that contains a highly basic (two Lys and six Arg among nine residues) protein transduction domain, which brings about nuclear localization and RNA binding. Other CPPs that have been used for the modification of liposomes include the following: the minimal protein transduction domain of Antennapedia, a Drosophilia homeoprotein, called penetratin, which is a 16-mer peptide (residues 43–58) present in the third helix of the homeodomain; a 27-amino acid-long chimeric CPP, containing the peptide sequence from the amino terminus of the neuropeptide galanin bound via the Lys residue, mastoparan, a wasp venom peptide; VP22, a major structural component of HSV-1 facilitating intracellular transport and transportan (18-mer) amphipathic model peptide that translocates plasma membranes of mast cells and endothelial cells by both energy-dependent and -independent mechanisms. 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 delivers 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 enhances 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.

[0740] An embodiment of the invention includes the delivery system comprising an actively targeting lipid particle or nanoparticle or liposome or lipid bilayer delivery system; or comprising a lipid particle or nanoparticle or liposome or lipid bilayer comprising a targeting moiety whereby there is active targeting or wherein the targeting moiety is an actively targeting moiety. A targeting moiety can be one or more targeting moieties, and a targeting moiety can be for any desired type of targeting such as, e.g., to target a cell such as any herein-mentioned; or to target an organelle such as any herein-mentioned; or for targeting a response such as to a physical condition such as heat, energy, ultrasound, light, pH, chemical such as enzymatic, or magnetic stimuli; or to target to achieve a particular outcome such as delivery of payload to a particular location, such as by cell penetration.

[0741] It should be understood that as to each possible targeting or active targeting moiety herein-discussed, there is an aspect of the invention wherein the delivery system comprises such a targeting or active targeting moiety. Likewise, **Table 5** below provides exemplary targeting moieties that can be used in the practice of the invention an as to each an aspect of the invention provides a delivery system that comprises such a targeting moiety.

Table 5

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- *HER2* antibody	*HER2*	*HER2* -overexpressing tumors
anti-*GD2*	*GD2*	neuroblastoma, melanoma
anti-EGFR	*EGFR*	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)	avβ3	glioblastoma cells, human umbilical vein endothelial cells, tumor angiogenesis
ASSHN 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 (P6.1 peptide)	HER-2 receptor	cancer cells
affinity peptide LN (YEVGHRC)	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

[0742] 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 aspect 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).

[0743] Moreover, in view of the teachings herein the skilled artisan can readily select and apply a desired targeting moiety in the practice of the invention as to a lipid entity of the invention. The invention comprehends an embodiment wherein the delivery system comprises a lipid entity having a targeting moiety.

[0744]

Dosages

Dosage of vectors

[0745] In some embodiments, the vector, e.g., plasmid or viral vector is delivered to the tissue of interest by, for example, an intramuscular injection, while other times the delivery is via intravenous, transdermal, intranasal, oral, mucosal, or other delivery methods. Such delivery may be either via a single dose, or multiple doses. One skilled in the art understands that the actual dosage to be delivered herein may vary greatly depending upon a variety of factors, such as the vector choice, the target cell, organism, or tissue, the general condition of the subject to be treated, the degree of transformation/modification sought, the administration route, the administration mode, the type of transformation/modification sought, etc.

[0746] Such a dosage may further contain, for example, a carrier (water, saline, ethanol, glycerol, lactose, sucrose, calcium phosphate, gelatin, dextran, agar, pectin, peanut oil, sesame oil, etc.), a diluent, a pharmaceutically-acceptable carrier (e.g., phosphate-buffered saline), a

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

[0747] In an embodiment herein the delivery is via an adenovirus, which may be at a single dose or booster dose containing at least 1 x 10^5 particles (also referred to as particle units, pu) of adenoviral vector. In an embodiment herein, the dose preferably is at least about 1 x 10^6 particles (for example, about 1 x 10^6-1 x 10^{12} particles), more preferably at least about 1 x 10^7 particles, more preferably at least about 1 x 10^8 particles (e.g., about 1 x 10^8-1 x 10^{11} particles or about 1 x 10^8-1 x 10^{12} particles), and most preferably at least about 1 x 10^0 particles (e.g., about 1 x 10^9-1 x 10^{10} particles or about 1 x 10^9-1 x 10^{12} particles), or even at least about 1 x 10^{10} particles (e.g., about 1 x 10^{10}-1 x 10^{12} particles) of the adenoviral vector. Alternatively, the dose comprises no more than about 1 x 10^{14} particles, preferably no more than about 1 x 10^{13} particles, even more preferably no more than about 1 x 10^{12} particles, even more preferably no more than about 1 x 10^{11} particles, and most preferably no more than about 1 x 10^{10} particles (e.g., no more than about 1 x 10^9 articles). Thus, the dose may contain a single dose of adenoviral vector with, for example, about 1 x 10^6 particle units (pu), about 2 x 10^6 pu, about 4 x 10^6 pu, about 1 x 10^7 pu, about 2 x 10^7 pu, about 4 x 10^7 pu, about 1 x 10^8 pu, about 2 x 10^8 pu, about 4 x 10^8 pu, about 1 x 10^9 pu, about 2 x 10^9 pu, about 4 x 10^9 pu, about 1 x 10^{10} pu, about 2 x 10^{10} pu, about 4 x 10^{10} pu, about 1 x 10^{11} pu, about 2 x 10^{11} pu, about 4 x 10^{11} pu,

about 1 x 10^{12} pu, about 2 x 10^{12} pu, or about 4 x 10^{12} pu of adenoviral vector. See, for example, the adenoviral vectors in U.S. Patent No. 8,454,972 B2 to Nabel, et. al., granted on June 4, 2013; incorporated by reference herein, and the dosages at col 29, lines 36-58 thereof. In an embodiment herein, the adenovirus is delivered via multiple doses.

[0748] In an embodiment herein, the delivery is via an AAV. A therapeutically effective dosage for *in vivo* delivery of the AAV to a human is believed to be in the range of from about 20 to about 50 ml of saline solution containing from about 1 x 10^{10} to about 1 x 10^{10} functional AAV/ml solution. The dosage may be adjusted to balance the therapeutic benefit against any side effects. In an embodiment herein, the AAV dose is generally in the range of concentrations of from about 1 x 10^{5} to 1 x 10^{50} genomes AAV, from about 1 x 10^{8} to 1 x 10^{20} genomes AAV, from about 1 x 10^{10} to about 1 x 10^{16} genomes, or about 1 x 10^{11} to about 1 x 10^{16} genomes AAV. A human dosage may be about 1 x 10^{13} genomes AAV. Such concentrations may be delivered in from about 0.001 ml to about 100 ml, about 0.05 to about 50 ml, or about 10 to about 25 ml of a carrier solution. Other effective dosages can be readily established by one of ordinary skill in the art through routine trials establishing dose response curves. See, for example, U.S. Patent No. 8,404,658 B2 to Hajjar, et al., granted on March 26, 2013, at col. 27, lines 45-60.

[0749] In an embodiment herein the delivery is via a plasmid. In such plasmid compositions, the dosage should be a sufficient amount of plasmid to elicit a response. For instance, suitable quantities of plasmid DNA in plasmid compositions can be from about 0.1 to about 2 mg, or from about 1 μg to about 10 μg per 70 kg individual. Plasmids of the invention will generally comprise (i) a promoter; (ii) a sequence encoding a CRISPR enzyme, operably linked to said promoter; (iii) a selectable marker; (iv) an origin of replication; and (v) a transcription terminator downstream of and operably linked to (ii). The plasmid can also encode the RNA components of a CRISPR complex, but one or more of these may instead be encoded on a different vector.

[0750] The doses herein are based on an average 70 kg individual. The frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), or scientist skilled in the art. It is also noted that mice used in experiments are typically about 20g and from mice experiments one can scale up to a 70 kg individual.

[0751] The dosage used for the compositions provided herein include dosages for repeated administration or repeat dosing. In particular embodiments, the administration is repeated

within a period of several weeks, months, or years. Suitable assays can be performed to obtain an optimal dosage regime. Repeated administration can allow the use of lower dosage, which can positively affect off-target modifications.

Optimizing Dosages

[0752] In some aspects the dosage of the CRISPR-Cas system, component thereof, formulation thereof, vector, particle, cell, and the like described elsewhere herein can be optimized to minimize toxicity and/or maximize effect and/or specificity.

Identifying optimal dosage to minimize toxicity and maximize specificity.

[0753] It is generally accepted that the dosage of CRISPR components will be relevant to toxicity and specificity of the system (Pattanayak et al. Nat Biotechnol. 2013 Sep; 31(9): 839–843). Hsu et al. (Nat Biotechnol. 2013 Sep; 31(9): 827–832) demonstrated that the dosage of SpCas9 and sgRNA can be titrated to address these issues. In certain example embodiments, toxicity is minimized by saturating complex with guide by either pre-forming complex, putting guide under control of a strong promoter, or via timing of delivery to ensure saturating conditions available during expression of the effector protein.

Identifying appropriate delivery vector

[0754] In some embodiments, the components of the CRISPR system may be delivered in various form, such as combinations of DNA/RNA or RNA/RNA or protein/RNA. For example, the Type VII effector may be delivered as a DNA-coding polynucleotide or an RNA--coding polynucleotide or as a protein. The guide may be delivered as a DNA-coding polynucleotide or an RNA. All possible combinations are envisioned, including mixed forms of delivery.

[0755] In some aspects, the invention provides methods comprising delivering one or more polynucleotides, such as or one or more vectors as described herein, one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell.

CELLS AND ORGANISMS

[0756] Described herein are modified cells, cell populations, and organisms that can be modified by and/or modified to include a CRISPR-Cas system or component thereof described herein (e.g. a Type VII system and/or component thereof described herein). The modified cells, cell populations, and organisms can have an insertion of one or more polynucleotides, deletion of one or more polynucleotides, mutation of one or more polynucleotides, or a combination thereof. The modification can result in insertion, deletion, activation, inactivation, modulations, or a combination thereof of one or more polynucleotides and/or genes. Cells,

including cells of an organism, can be modified *in vitro*, in situ, *ex vivo*, or *in vivo*. The organism can be eukaryotic or prokaryotic. In some embodiments the eukaryotic cell is a human cell. In some embodiments, the eukaryotic cell is a non-human animal cell. In some embodiments, the cell can be a plant cell. Other suitable cells are described elsewhere herein. In some embodiments, the modification is insertion, deletion, modification, modulation or a combination thereof of a polynucleotide, gene, or allele of interest. In some embodiments, the polynucleotide, gene, or allele of interest is a pathogenic polynucleotide, gene, or allele. In some embodiments, the polynucleotide, gene, or allele of interest is or confers a commercially and/or therapeutically important trait, function, activity, and/or phenotype. Exemplary genes of interest are described in greater detail elsewhere herein.

Modified Cells

[0757] Described herein are modified cells that are modified to contain and/or express on e or more Type VII CRISPR-Cas systems or components thereof described herein. Also described herein are modified cells and cell populations that can be modified by an embodiment of a CRISPR-Cas system described in greater detail elsewhere herein. In some embodiments, the eukaryotic cell is a mammalian cell. In some embodiments, the eukaryotic cell is a non-human mammalian cell. In some embodiments, the cell is a human cell. In some embodiments, the cell is a plant cell. In some embodiments, the cell is a fungal cell. In some embodiments, the cell is a prokaryotic cell. The cells can be modified *in vitro*, *ex vivo*, or *in vivo*. The cells can be modified by delivering a polynucleotide modifying agent or system described in greater detail elsewhere herein or a component thereof into a cell by a suitable delivery mechanism. Suitable delivery methods and techniques include but are not limited to, transfection via a vector, transduction with viral particles, electroporation, endocytic methods, and others, which are described elsewhere herein and will be appreciated by those of ordinary skill in the art in view of this disclosure.

[0758] The modified cells can be further optionally cultured and/or expanded *in vitro* or *ex vivo* using any suitable cell culture techniques or conditions, which unless specified otherwise herein, will be appreciated by one of ordinary skill in the art in view of this disclosure. In some embodiments, the cells can be modified, optionally cultured and/or expanded, and administered to a subject in need thereof. In some embodiments, cells can be isolated from a subject, subsequently modified and optionally cultured and/or expanded, and administered back to the subject. Such administration can be referred to as autologous administration. In some

embodiments, cells can be isolated from a first subject, subsequently modified, optionally cultured and/or expanded, and administered to a second subject, where the first subject and the second subject are different. Such administration can be referred to as non-autologous administration.

[0759] In certain embodiments, the methods as described herein may comprise providing a Cas transgenic cell in which one or more nucleic acids encoding one or more guide RNAs are provided or introduced operably connected in the cell with a regulatory element comprising a promoter of one or more gene of interest. As used herein, the term "Cas transgenic cell" refers to a cell, such as a eukaryotic cell, in which a Cas gene has been genomically integrated. The nature, type, or origin of the cell are not particularly limiting according to the present invention. Also, the way the Cas transgene is introduced in the cell may vary and can be any method as is known in the art. In certain embodiments, the Cas transgenic cell is obtained by introducing the Cas transgene in an isolated cell. In certain other embodiments, the Cas transgenic cell is obtained by isolating cells from a Cas transgenic organism. By means of example, and without limitation, the Cas transgenic cell as referred to herein may be derived from a Cas transgenic eukaryote, such as a Cas knock-in eukaryote. Reference is made to International Patent Publication No. WO 2014/093622 (PCT/US13/74667), incorporated herein by reference. Methods of US Patent Publication Nos. 20120017290 and 20110265198 assigned to Sangamo BioSciences, Inc. directed to targeting the Rosa locus may be modified to utilize the CRISPR Cas system of the present invention. Methods of US Patent Publication No. 20130236946 assigned to Cellectis directed to targeting the Rosa locus may also be modified to utilize the CRISPR Cas system of the present invention. By means of further example reference is made to Platt et. al. (Cell; 159(2):440-455 (2014)), describing a Cas9 knock-in mouse, which is incorporated herein by reference. The Cas transgene can further comprise a Lox-Stop-polyA-Lox(LSL) cassette thereby rendering Cas expression inducible by Cre recombinase. Alternatively, the Cas transgenic cell may be obtained by introducing the Cas transgene in an isolated cell. Delivery systems for transgenes are well known in the art. By means of example, the Cas transgene may be delivered in for instance eukaryotic cell by means of vector (e.g., AAV, adenovirus, lentivirus) and/or particle and/or nanoparticle delivery, as also described herein elsewhere.

[0760] It will be understood by the skilled person that the cell, such as the Cas transgenic cell, as referred to herein may comprise further genomic alterations besides having an

integrated Cas gene or the mutations arising from the sequence specific action of Cas when complexed with RNA capable of guiding Cas to a target locus.

[0761] In some embodiments the modified cell, is generated by delivering a CRISPR-Cas system or component thereof via any suitable method (described in greater detail elsewhere herein) to a cell to be modified. In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject optionally to be reintroduced therein. In some embodiments, a cell that is transfected is taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/ 3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr -/-, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN / OPCT cell lines, Peer, PNT-1A / PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In some embodiments, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more

vector-derived sequences. In some embodiments, a cell transiently transfected with the components of a CRISPR system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a CRISPR complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[0762] A further aspect relates to an isolated cell obtained or obtainable from the methods described herein comprising the composition described herein or progeny of said modified cell, preferably wherein said cell comprises a hypoxanthine or a guanine in replace of said Adenine in said target RNA of interest compared to a corresponding cell not subjected to the method. In particular embodiments, the cell is a eukaryotic cell, preferably a human or non-human animal cell, optionally a therapeutic T cell or an antibody-producing B-cell.

[0763] In some embodiments, the modified cell is a therapeutic T cell, such as a T cell suitable for adoptive cell transfer therapies (e.g., CAR-T therapies). The modification may result in one or more desirable traits in the therapeutic T cell, as described further herein.

Organisms

[0764] Also described herein are modified organisms. In some embodiments, the modified organisms can include one or more modified cells as that are described elsewhere herein. In some embodiments, one or more cells of an organism can be modified using a CRISPR-Cas system or component thereof, such as a Type VII CRISPR-Cas system described herein. Modification can be *ex vivo* or *in vivo*. In some embodiments, the modified organism is a non-human mammal. In some embodiments, the modified organism is a modified plant. In some embodiments, the modified organism is an insect. In some embodiments, the modified organism is a fungus. In some embodiments, the modified organism is a fungus. The modified organisms can be generated using a that can be modified by an embodiment of the engineered or non-natural guided excision-transposition system described herein. Methods of making modified organisms are described in greater detail elsewhere herein.

[0765] The systems and methods described herein can be used in non-animal organisms, e.g., plants, fungi to generated modified non-animal organisms. The system and methods described can be used to generate non-human animal organisms. The system and methods described herein can be used to modify non-germline cells in a human. In some embodiments,

the modification is expression of a polynucleotide of interest, gene of interest, and/or allele of interest.

Non-Animal Organisms

[0766] The CRISPR-Cas systems (e.g. Type VII CRISPR-Cas systems) described herein can be used to modify non-animal organisms such as plants, yeast, etc. In general, the term "plant" relates to any various photosynthetic, eukaryotic, unicellular or multicellular organism of the kingdom Plantae characteristically growing by cell division, containing chloroplasts, and having cell walls comprised of cellulose. The term plant encompasses monocotyledonous and dicotyledonous plants. Specifically, the plants are intended to comprise without limitation angiosperm and gymnosperm plants such as acacia, alfalfa, amaranth, apple, apricot, artichoke, ash tree, asparagus, avocado, banana, barley, beans, beet, birch, beech, blackberry, blueberry, broccoli, Brussel's sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, cedar, a cereal, celery, chestnut, cherry, Chinese cabbage, citrus, clementine, clover, coffee, corn, cotton, cowpea, cucumber, cypress, eggplant, elm, endive, eucalyptus, fennel, figs, fir, geranium, grape, grapefruit, groundnuts, ground cherry, gum hemlock, hickory, kale, kiwifruit, kohlrabi, larch, lettuce, leek, lemon, lime, locust, pine, maidenhair, maize, mango, maple, melon, millet, mushroom, mustard, nuts, oak, oats, oil palm, okra, onion, orange, an ornamental plant or flower or tree, papaya, palm, parsley, parsnip, pea, peach, peanut, pear, peat, pepper, persimmon, pigeon pea, pine, pineapple, plantain, plum, pomegranate, potato, pumpkin, radicchio, radish, rapeseed, raspberry, rice, rye, sorghum, safflower, sallow, soybean, spinach, spruce, squash, strawberry, sugar beet, sugarcane, sunflower, sweet potato, sweet corn, tangerine, tea, tobacco, tomato, trees, triticale, turf grasses, turnips, vine, walnut, watercress, watermelon, wheat, yams, yew, and zucchini. The term plant also encompasses algae, which are mainly photoautotrophs unified primarily by their lack of roots, leaves and other organs that characterize higher plants.

[0767] The CRISPR-Cas systems (e.g. Type VII CRISPR-Cas systems) as described herein can be used to confer desired traits on essentially any plant. A wide variety of plants and plant cell systems may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation methods mentioned above. In preferred embodiments, target plants and plant cells for engineering include, but are not limited to, those monocotyledonous and dicotyledonous plants, such as crops including grain crops (e.g., wheat, maize, rice, millet,

barley), fruit crops (e.g., tomato, apple, pear, strawberry, orange), forage crops (e.g., alfalfa), root vegetable crops (e.g., carrot, potato, sugar beets, yam), leafy vegetable crops (e.g., lettuce, spinach); flowering plants (e.g., petunia, rose, chrysanthemum), conifers and pine trees (e.g., pine fir, spruce); plants used in phytoremediation (e.g., heavy metal accumulating plants); oil crops (e.g., sunflower, rape seed) and plants used for experimental purposes (e.g., Arabidopsis). Plant cells and tissues for engineering include, without limitation, roots, stems, leaves, flowers, and reproductive structures, undifferentiated meristematic cells, parenchyma, collenchyma, sclerenchyma, xylem, phloem, epidermis, and germplasm. Thus, the methods and modifying agents and systems described herein can be used over a broad range of plants, such as for example with dicotyledonous plants belonging to the orders Magniolales, Illiciales, Laurales, Piperales, Aristochiales, Nymphaeales, Ranunculales, Papeverales, Sarraceniaceae, 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, San tales, Rafflesiales, Celastrales, Euphorbiales, Rhamnales, Sapindales, Juglandales, Geraniales, Polygalales, Umbellales, Gentianales, Polemoniales, Lamiales, Plantaginales, Scrophulariales, Campanulales, Rubiales, Dipsacales, and Asterales; the methods and CRISPR-Cas systems can be used with monocotyledonous plants such as those belonging to the orders Alismatales, Hydrocharitales, Najadales, Triuridales, Commelinales, Eriocaulales, Restionales, Poales, Juncales, Cyperales, Typhales, Bromeliales, Zingiberales, Arecales, Cyclanthales, Pandanales, Arales, Lilliales, and Orchid ales, or with plants belonging to Gymnospermae, e.g those belonging to the orders Pinales, Ginkgoales, Cycadales, Araucariales, Cupressales and Gnetales.

[0768] The CRISPR-Cas systems (e.g. Type VII CRISPR-Cas systems) and methods of use described herein can be used over a broad range of plant species, included in the non-limitative list of dicot, monocot or gymnosperm genera hereunder: Atropa, Alseodaphne, Anacardium, Arachis, Beilschmiedia, Brassica, Carthamus, Cocculus, Croton, Cucumis, Citrus, Citrullus, Capsicum, Catharanthus, Cocos, Coffea, Cucurbita, Daucus, Duguetia, Eschscholzia, Ficus, Fragaria, Glaucium, Glycine, Gossypium, Helianthus, Hevea, Hyoscyamus, Lactuca, Landolphia, Linum, 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, Vilis, and Vigna; and the genera Allium, Andropogon, Aragrostis, Asparagus, Avena, Cynodon, Elaeis, Festuca, Festulolium, Heterocallis, Hordeum, Lemna, Lolium, Musa, Oryza, Panicum, Pannesetum, Phleum, Poa, Secale, Sorghum, Triticum, Zea, Abies, Cunninghamia, Ephedra, Picea, Pinus, and Pseudotsuga.

[0769] The CRISPR-Cas systems (e.g. Type VII CRISPR-Cas systems) and methods of use described herein can also be used over a broad range of "algae" or "algae cells"; including for example algea selected from several eukaryotic phyla, including the Rhodophyta (red algae), Chlorophyta (green algae), Phaeophyta (brown algae), Bacillariophyta (diatoms), Eustigmatophyta and dinoflagellates as well as the prokaryotic phylum Cyanobacteria (blue-green algae). The term "algae" includes for example algae selected from : Amphora, Anabaena, Anikstrodesmis, Botryococcus, Chaetoceros, Chlamydomonas, Chlorella, Chlorococcum, Cyclotella, Cylindrotheca, Dunaliella, Emiliana, Euglena, Hematococcus, Isochrysis, Monochrysis, Monoraphidium, Nannochloris, Nannnochloropsis, Navicula, Nephrochloris, Nephroselmis, Nitzschia, Nodularia, Nostoc, Oochromonas, Oocystis, Oscillartoria, Pavlova, Phaeodactylum, Playtmonas, Pleurochrysis, Porhyra, Pseudoanabaena, Pyramimonas, Stichococcus, Synechococcus, Synechocystis, Tetraselmis, Thalassiosira, and Trichodesmium.

[0770] A part of a plant, e.g., a "plant tissue" may be treated according to the methods of the present invention to produce an improved plant. Plant tissue also encompasses plant cells. The term “plant cell” as used herein refers to individual units of a living plant, either in an intact whole plant or in an isolated form grown in *in vitro* tissue cultures, on media or agar, in suspension in a growth media or buffer or as a part of higher organized unites, such as, for example, plant tissue, a plant organ, or a whole plant.

[0771] 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.

[0772] The term "transformation" broadly refers to the process by which a plant host is genetically modified by the introduction of DNA by means of Agrobacteria or one of a variety of chemical or physical methods. As used herein, the term "plant host" refers to plants, including any cells, tissues, organs, or progeny of the plants. Many suitable plant tissues or

plant cells can be transformed and include, but are not limited to, protoplasts, somatic embryos, pollen, leaves, seedlings, stems, calli, stolons, microtubers, and shoots. A plant tissue also refers to any clone of such a plant, seed, progeny, propagule whether generated sexually or asexually, and descendants of any of these, such as cuttings or seed.

[0773] The term "transformed" as used herein, refers to a cell, tissue, organ, or organism into which a foreign DNA molecule, such as a construct, has been introduced. The introduced DNA molecule may be integrated into the genomic DNA of the recipient cell, tissue, organ, or organism such that the introduced DNA molecule is transmitted to the subsequent progeny. In these embodiments, the "transformed" or "transgenic" cell or plant may also include progeny of the cell or plant and progeny produced from a breeding program employing such a transformed plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of the introduced DNA molecule. Preferably, the transgenic plant is fertile and capable of transmitting the introduced DNA to progeny through sexual reproduction.

[0774] The term "progeny", such as the progeny of a transgenic plant, is one that is born of, begotten by, or derived from a plant or the transgenic plant. The introduced DNA molecule may also be transiently introduced into the recipient cell such that the introduced DNA molecule is not inherited by subsequent progeny and thus not considered "transgenic". Accordingly, as used herein, a "non-transgenic" plant or plant cell is a plant which does not contain a foreign DNA stably integrated into its genome.

[0775] The term "plant promoter" as used herein is a promoter capable of initiating transcription in plant cells, whether or not its origin is a plant cell. Exemplary suitable plant promoters include, but are not limited to, those that are obtained from plants, plant viruses, and bacteria such as Agrobacterium or Rhizobium which comprise genes expressed in plant cells.

[0776] As used herein, a "fungal cell" refers to any type of eukaryotic cell within the kingdom of fungi. Phyla within the kingdom of fungi include Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota, Glomeromycota, Microsporidia, and Neocallimastigomycota. Fungal cells may include yeasts, molds, and filamentous fungi. In some embodiments, the fungal cell is a yeast cell.

[0777] As used herein, the term "yeast cell" refers to any fungal cell within the phyla Ascomycota and Basidiomycota. Yeast cells may include budding yeast cells, fission yeast cells, and mold cells. Without being limited to these organisms, many types of yeast used in laboratory and industrial settings are part of the phylum Ascomycota. In some embodiments,

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

[0778] In some embodiments, the fungal cell is an industrial strain. As used herein, "industrial strain" refers to any strain of fungal cell used in or isolated from an industrial process, e.g., production of a product on a commercial or industrial scale. Industrial strain may refer to a fungal species that is typically used in an industrial process, or it may refer to an isolate of a fungal species that may be also used for non-industrial purposes (e.g., laboratory research). Examples of industrial processes may include fermentation (e.g., in production of food or beverage products), distillation, biofuel production, production of a compound, and production of a polypeptide. Examples of industrial strains may include, without limitation, JAY270 and ATCC4124.

[0779] In some embodiments, the fungal cell is a polyploid cell. As used herein, a "polyploid" cell may refer to any cell whose genome is present in more than one copy. A polyploid cell may refer to a type of cell that is naturally found in a polyploid state, or it may refer to a cell that has been induced to exist in a polyploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A polyploid cell may refer to a cell whose entire genome is polyploid, or it may refer to a cell that is polyploid in a particular genomic locus of interest. Without wishing to be bound to theory, it is thought that the abundance of guideRNA may more often be a rate-limiting component in genome engineering of polyploidy cells than in haploid cells, and thus the methods using the systems described herein may take advantage of using a certain fungal cell type.

[0780] In some embodiments, the fungal cell is a diploid cell. As used herein, a "diploid" cell may refer to any cell whose genome is present in two copies. A diploid cell may refer to a type of cell that is naturally found in a diploid state, or it may refer to a cell that has been induced to exist in a diploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the S. cerevisiae strain S228C may be maintained in a haploid or diploid state. A diploid cell may refer to a cell whose entire genome is diploid, or it may refer to a cell that is diploid in a particular genomic locus of interest. In some embodiments, the fungal cell is a haploid cell. As used herein, a "haploid" cell may refer to any cell whose genome is present in one copy. A haploid cell may refer to a type of cell that is naturally found in a haploid state, or it may refer to a cell that has been induced to exist in a haploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the S. cerevisiae strain S228C may be maintained in a haploid or diploid state. A haploid cell may refer to a cell whose entire genome is haploid, or it may refer to a cell that is haploid in a particular genomic locus of interest.

[0781] 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 may contain, without limitation, a centromeric (CEN) sequence, an autonomous replication sequence (ARS), a promoter, such as an RNA Polymerase III promoter, operably linked to a sequence or gene of interest, a terminator such as an RNA polymerase III terminator, an origin of replication, and a marker gene (e.g., auxotrophic, antibiotic, or other selectable markers). Examples of expression vectors for use in yeast may include plasmids, yeast artificial chromosomes, 2μ plasmids, yeast integrative plasmids, yeast replicative plasmids, shuttle vectors, and episomal plasmids.

[0782] Described herein are plants and/or plant cells that can be produced by one or more of the methods described herein, 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 plant. This is described in greater detail herein.

[0783] Also described herein are gametes, seeds, germplasm, embryos, either zygotic or somatic, progeny or hybrids of plants comprising the genetic modification, which are produced by traditional breeding methods, are also included within the scope of the present invention. Such plants may contain a heterologous or foreign DNA sequence inserted at or instead of a target sequence. Alternatively, such plants may contain only an alteration (mutation, deletion, insertion, substitution) in one or more nucleotides. As such, such plants will only be different from their progenitor plants by the presence of the particular modification.

[0784] The polynucleotide modifying agent(s) and/or systems described herein can be used to confer desired traits on essentially any plant, algae, fungus, yeast, etc. A wide variety of plants, algae, fungus, yeast, etc. and plant algae, fungus, yeast cell or tissue systems may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation methods mentioned above.

[0785] In particular embodiments, the methods described herein are used to modify endogenous genes or to modify their expression without the permanent introduction into the genome of the plant, algae, fungus, yeast, etc. of any foreign gene, including those encoding CRISPR components, 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.

[0786] Also described herein are modified non-animal organisms (plants, yeast, algae, and other microorganisms) that can express one or more polynucleotides, genes or alleles of interest. Exemplary genes of interest are described in greater detail elsewhere herein.

Stable integration in the genome of plants and plant cells

[0787] In particular embodiments, the polynucleotides encoding the polynucleotide modifying agents or systems thereof are introduced for stable integration into the genome of a plant cell. In these embodiments, the design of the transformation vector or the expression system can be adjusted depending on for when, where and under what conditions the polynucleotide modifying agents or systems thereof are expressed. Suitable vectors and delivery are described in greater detail elsewhere herein.

[0788] In particular embodiments, the polynucleotide modifying agents or systems thereof are stably introduced into the genomic DNA of a plant cell. In particular embodiments, the polynucleotide modifying agents or systems thereof are introduced for stable integration into the DNA of a plant organelle such as, but not limited to a plastid, e mitochondrion or a chloroplast. In some embodiments, the expression system for stable integration into the genome of a plant cell can contain one or more of the following elements: a promoter element that can be used to express a polynucleotide modifying agent(s) or a system thereof 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 polynucleotide modifying agent(s) or a system thereof and other desired elements; and a 3' untranslated region to provide for efficient termination of the expressed transcript. 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.

[0789] DNA construct(s) containing the components of the systems, and, where applicable, template sequence may be introduced into the genome of a plant, plant part, or plant cell by a variety of conventional techniques. The process generally comprises the steps of selecting a suitable host cell or host tissue, introducing the construct(s) into the host cell or host tissue.

[0790] In particular embodiments, the DNA construct may be introduced into the plant cell using techniques such as but not limited to electroporation, microinjection, aerosol beam injection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using biolistic methods, such as DNA particle bombardment (see also Fu et al., Transgenic Res. 2000 Feb;9(1):11-9). The basis of particle bombardment is the acceleration of particles coated with gene/s of interest toward cells, resulting in the penetration of the protoplasm by the particles and typically stable integration into the genome. (see e.g. Klein et al, Nature (1987), Klein et ah, Bio/Technology (1992), Casas et ah, Proc. Natl. Acad. Sci. USA (1993).).

[0791] In particular embodiments, the DNA constructs containing components of the systems may be introduced into the plant by Agrobacterium-mediated transformation. The DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional Agrobacterium tumefaciens host vector. The foreign DNA can be incorporated into the genome of plants by infecting the plants or by incubating plant protoplasts with

Agrobacterium bacteria, containing one or more Ti (tumor-inducing) plasmids. (see e.g. Fraley et al., (1985), Rogers et al., (1987) and U.S. Pat. No. 5,563,055).

Transient expression of in plants and plant cells

[0792] In some embodiments, the polynucleotide modifying agent(s) and/or systems can be transiently expressed in the plant cell. In these embodiments, the system can ensure modification of a target gene only when all the required components of the system (e.g. in the context of a typical CRISPR-Cas system, the Cas enzyme(s) and guide RNA(s)) are present in a cell, such that polynucleotide modification can further be controlled. As the expression of the necessary components of the modification agent and/or system is transient, plants regenerated from such plant cells typically contain no foreign DNA. It will be appreciated that not all components must be expressed transiently for modification to be controlled by transient expression. In some embodiments where multiple components are necessary for modification to occur, one or more components of the modification system are expressed transiently and one or more components of the system are stably expressed. In some embodiments where a CRISPR-Cas system is employed, the Cas enzyme is stably expressed by the plant cell and the guide sequence is transiently expressed. In some embodiments where a CRISPR-Cas system is employed, the Cas enzyme is transiently expressed by the plant cell and the guide sequence is stably expressed.

[0793] In particular embodiments, the polynucleotide modifying agent(s) and/or system components can be transiently introduced in the plant cells using a plant viral vector (Scholthof et al. 1996, Annu Rev Phytopathol. 1996;34:299-323). In further particular embodiments, said viral vector is a vector from a DNA virus. For example, geminivirus (e.g., cabbage leaf curl virus, bean yellow dwarf virus, wheat dwarf virus, tomato leaf curl virus, maize streak virus, tobacco leaf curl virus, or tomato golden mosaic virus) or nanovirus (e.g., Faba bean necrotic yellow virus). In other particular embodiments, said viral vector is a vector from an RNA virus. For example, tobravirus (e.g., tobacco rattle virus, tobacco mosaic virus), potexvirus (e.g., potato virus X), or hordeivirus (e.g., barley stripe mosaic virus). The replicating genomes of plant viruses are non-integrative vectors.

[0794] In particular embodiments, the vector used for transient expression of constructs is for instance a pEAQ vector, which is tailored for Agrobacterium-mediated transient expression (Sainsbury F. et al., Plant Biotechnol J. 2009 Sep;7(7):682-93) in the protoplast. Precise targeting of genomic locations was demonstrated using a modified Cabbage Leaf Curl virus

(CaLCuV) vector to express gRNAs in stable transgenic plants expressing a CRISPR enzyme (Scientific Reports 5, Article number: 14926 (2015), doi:10.1038/srep14926).

[0795] In particular embodiments, double-stranded DNA fragments encoding the polynucleotide modifying agent(s) and/or system component(s) (e.g. where a CRISPR-Cas system is employed, a guide RNA and/or the Cas gene) can be transiently introduced into the plant cell. In such embodiments, the introduced double-stranded DNA fragments are provided in sufficient quantity to modify the cell but do not persist after a contemplated period of time has passed or after one or more cell divisions. Methods for direct DNA transfer in plants are known by the skilled artisan (see for instance Davey et al. Plant Mol Biol. 1989 Sep;13(3):273-85.)

[0796] In other embodiments, an RNA polynucleotide encoding the a protein polynucleotide modifying agent or system component (e.g. where a CRISPR-Cas system is employed, a Cas protein) is introduced into the plant cell, which is then translated and processed by the host cell generating the protein in sufficient quantity to modify the cell (in the presence of at least one guide RNA) but which does not persist after a contemplated period of time has passed or after one or more cell divisions. Methods for introducing mRNA to plant protoplasts for transient expression are known by the skilled artisan (see for instance in Gallie, Plant Cell Reports (1993), 13;119-122).

[0797] In some embodiments, a combination of the different methods described above can be used.

Plant promoters

[0798] In some embodiments, the polynucleotide modifying agent(s) or systems thereof described elsewhere herein can be placed under control of a suitable plant promoter, i.e. a promoter operable in plant cells. The use of different types of promoters is envisaged. Plant promoters can be constitutive, inducible, and/or tissue specific.

[0799] 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. "Regulated promoter" refers to promoters that direct gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and includes tissue-specific, tissue-preferred and inducible promoters. Different promoters may direct the expression of a gene in different

tissues or cell types, or at different stages of development, or in response to different environmental conditions. In particular embodiments, one or more of the Type VII CRISPR-Cas system or component thereof 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 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.

[0800] Examples of promoters that are inducible and that allow for spatiotemporal control of gene editing or gene expression may use a form of energy. The form of energy may include but is not limited to sound energy, electromagnetic radiation, chemical energy and/or thermal energy. Examples of inducible systems include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc), or light inducible systems (Phytochrome, LOV domains, or cryptochrome)., such as a Light Inducible Transcriptional Effector (LITE) that direct changes in transcriptional activity in a sequence-specific manner. The components of a light inducible system may include one or more gene modifying agents, a light-responsive cytochrome heterodimer (e.g. from Arabidopsis thaliana), and a transcriptional activation/repression domain. Further examples of inducible DNA binding proteins and methods for their use are provided in US 61/736465 and US 61/721,283, which is hereby incorporated by reference in its entirety.

[0801] In particular embodiments, transient or inducible expression can be achieved by using, for example, chemical-regulated promotors, i.e. whereby the application of an exogenous chemical induces gene expression. Modulating of gene expression can also be obtained by a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters include, but are not limited to, the maize 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.

Translocation to and/or expression in specific plant organelles

[0802] The system may comprise elements for translocation to and/or expression in a specific plant organelle. In some embodiments, a tissue specific promoter can be included in the expression construct. In some embodiments, a tissue localization or organelle localization sequence or signal can be incorporated into the expression constructs. Such promoters and localization signals are described in greater detail elsewhere herein and/or will be appreciated by one of ordinary skill in the art.

Chloroplast targeting

[0803] In some embodiments, the polynucleotide modifying system can specifically modify chloroplast genes or to ensure expression in the chloroplast. In some embodiments, chloroplast transformation methods or compartmentalization of the system components to the chloroplast. For instance, the introduction of genetic modifications in the plastid genome can reduce biosafety issues such as gene flow through pollen.

[0804] Methods of chloroplast transformation are known in the art and include Particle bombardment, PEG treatment, and microinjection. Additionally, methods involving the translocation of transformation cassettes from the nuclear genome to the plastid can be used as described in WO2010061186.

[0805] In some embodiments, one or more of the polynucleotide modifying system components can be targeted to the plant chloroplast. This can be achieved by incorporating in the expression construct a sequence encoding a chloroplast transit peptide (CTP) or plastid transit peptide, operably linked to the 5' region of the sequence encoding the Cas protein. The CTP is removed in a processing step during translocation into the chloroplast. Chloroplast targeting of expressed proteins is well known to the skilled artisan (see for instance Protein Transport into Chloroplasts, 2010, Annual Review of Plant Biology,Vol. 61: 157-180) . In such embodiments it is also can be desirable to target the guide RNA to the plant chloroplast. Methods and constructs which can be used for translocating guide RNA into the chloroplast by means of a chloroplast localization sequence are described, for instance, in US 20040142476, incorporated herein by reference. Such variations of constructs can be incorporated into the expression systems of the invention to efficiently translocate the Cas-guide RNA.

Introduction of polynucleotides in Algal cells.

[0806] In some embodiments, the modified organism is algae. Modified algae (or other plants such as rape) can be useful in a variety of situations, such as in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol) or other products. In some embodiments, such organisms can be engineered to express or overexpress high levels of a useful product. For example, they can be modified to produce oil and/or alcohols for use in the oil or biofuel industries.

[0807] Algae modification using polynucleotide modifying agents has been described in, for example U.S. Pat. No. 8,945,839 and International Patent Publication No. WO 2015086795, which can be adapted to modifying algae and similar organisms with the polynucleotide modifying agents and systems described herein. In some embodiments, the polynucleotide modifying agent(s) or system thereof can be introduced to the algae using a vector that expresses the polynucleotide modifying agent(s) or system thereof under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2 -tubulin. Some components of the polynucleotide modifying system (such as a guide RNA or other RNAs) can be optionally delivered using a vector containing T7 promoter. In some embodiments, a polynucleotide modifying agent and/or other components of the polynucleotide modifying system mRNA can be expressed and *in vitro* transcribed guide RNA can be delivered to algal cells. In some embodiments, delivery can be via electroporation. Electroporation protocols are available to the skilled person such as the standard recommended protocol from the GeneArt Chlamydomonas Engineering kit.

[0808] In particular embodiments, the endonuclease used herein is a split Cas enzyme. Split Cas enzymes used in Algae for targeted genome modification as has been described for Cas9 in International Patent Publication No. WO 2015/086795. Use of the Cas split system is suitable for an inducible method of genome targeting and can avoid or mitigate the potential toxic effect of the Cas overexpression within the algae cell. In particular embodiments, said Cas split domains (RuvC and HNH domains in the case of Cas9) can be simultaneously or sequentially introduced into the cell such that said split Cas domain(s) process the target nucleic acid sequence in the algae cell. The reduced size of the split Cas compared to the wild type Cas allows other methods of delivery of the systems to the cells, such as the use of cell penetrating peptides as described herein.

Introduction of polynucleotides in yeast cells

[0809] In some embodiments, a yeast cell can be modified using the polynucleotide modifying agents and/or systems described herein. Methods for transforming yeast cells which can be used to introduce polynucleotides encoding the systems components are well known to the artisan and are reviewed by Kawai et al., 2010, Bioeng Bugs. 2010 Nov-Dec; 1(6): 395–403). Non-limiting examples include transformation of yeast cells by lithium acetate treatment (which may further include carrier DNA and PEG treatment), bombardment or by electroporation.

Delivery to the plant cell

[0810] In particular embodiments, it is of interest to deliver one or more polynucleotide modifying agent(s) or components of the system directly to the plant cell. In particular embodiments, one or more of the polynucleotide modifying agent(s) or components of the system can be prepared outside the plant or plant cell and delivered to the cell. In some embodiments, the protein polynucleotide modifying agent (e.g. where a CRISPR-Cas system is used, a Cas protein) or system component is prepared *in vitro* prior to introduction to the plant cell. Proteins can be prepared by various methods known by one of skill in the art and include recombinant production and de novo synthesis. After expression, the protein can be isolated, refolded if needed, purified and optionally treated to remove any purification tags, such as a His-tag. Once crude, partially purified, or more completely purified protein is obtained, the protein may be introduced to the plant cell.

[0811] In some embodiments where a CRISPR-Cas or RNA guided system is employed, , the Cas or other protein(s) can be mixed with guide RNA(s) targeting the gene(s) of interest to form a pre-assembled ribonucleoprotein.

[0812] The individual components or pre-assembled ribonucleoprotein can be introduced into the plant cell via electroporation, by bombardment with Cas-associated gene product coated particles, by chemical transfection or by some other means of transport across a cell membrane. For instance, transfection of a plant protoplast with a pre-assembled CRISPR ribonucleoprotein has been demonstrated to ensure targeted modification of the plant genome (as described by Woo et al. Nature Biotechnology, 2015; DOI: 10.1038/nbt.3389), which can be adapted for use with the present invention.

[0813] In particular embodiments, the system components are introduced into the plant cells using nanoparticles. The components, either as protein or nucleic acid or in a combination

thereof, can be uploaded onto or packaged in nanoparticles and applied to the plants (such as for instance described in International Patent Publication No. WO 2008042156 and US Patent Publication No. 20130185823). In particular, embodiments of the invention comprise nanoparticles uploaded with or packed with DNA molecule(s) encoding the Cas protein, DNA molecules encoding the guide RNA and/or isolated guide RNA as described in WO2015089419.

[0814] In some embodiments, the polynucleotide modifying agent(s) or one or more components of the system to the plant cell is by using cell penetrating peptides (CPP). In some embodiments, the cell penetrating peptide can be linked to a protein polynucleotide modifying agent or other component of a polynucleotide modifying agent or system thereof.

[0815] In some embodiments where a CRISPR-Cas system is employed, the Cas protein and/or guide RNA is coupled to one or more CPPs to effectively transport them inside plant protoplasts; see also Ramakrishna (20140Genome Res. 2014 Jun;24(6):1020-7 for Cas9 in human cells). In other embodiments, the Cas gene and/or guide RNA are encoded by one or more circular or non-circular DNA molecule(s) which are coupled to one or more CPPs for plant protoplast delivery. The plant protoplasts can then regenerate to produce plant cells and further to plants.

[0816] CPPs are generally described as short peptides of fewer than 35 amino acids either derived from proteins or from chimeric sequences which are capable of transporting biomolecules across cell membrane in a receptor independent manner. CPP can be cationic peptides, peptides having hydrophobic sequences, amphipatic peptides, peptides having proline-rich and anti-microbial sequence, and chimeric or bipartite peptides (Pooga and Langel 2005). CPPs are able to penetrate biological membranes and as such trigger the movement of various biomolecules across cell membranes into the cytoplasm and to improve their intracellular routing, and hence facilitate interaction of the biomolecule with the target. Examples of CPP include amongst others: Tat, a nuclear transcriptional activator protein required for viral replication by HIV type1, penetratin, Kaposi fibroblast growth factor (FGF) signal peptide sequence, integrin β3 signal peptide sequence; polyarginine peptide Args sequence, Guanine rich-molecular transporters, sweet arrow peptide, etc.

Making genetically modified non-transgenic plants

[0817] In particular embodiments, the systems and methods described herein are used to modify endogenous genes or to modify their expression without the permanent introduction

into the genome of the plant of any foreign gene, including those encoding polynucleotide modifying agent(s) or components of a polynucleotide modifying system, 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.

[0818] In particular embodiments, this can be achieved by transient expression of the system components. In particular embodiments, one or more of the systems components are expressed on one or more viral vectors which produce sufficient components of the systems to consistently steadily ensure modification of a gene of interest according to a method described herein. In particular embodiments, transient expression of constructs is ensured in plant protoplasts and thus not integrated into the genome. The limited window of expression can be sufficient to allow the system to ensure modification of the target gene(s) as described herein.

[0819] In particular embodiments, different components of the system are introduced in the plant cell, protoplast or plant tissue either separately or in mixture, with the aid of particulate delivering molecules such as nanoparticles or CPP molecules as described herein above.

[0820] The expression of the components of the systems herein can induce targeted modification of the genome, either by direct activity of the polynucleotide modifying agent (e.g. when a CRISPR-Cas system is employed, a Cas protein) and optionally introduction of template DNA or by modification of genes targeted using the system as described herein. The different strategies described herein above can allow targeted genome editing without requiring the introduction of the components into the plant genome. Components which are transiently introduced into the plant cell can be, in some embodiments, removed upon crossing.

[0821] Protocols for targeted plant genome editing via CRISPR-Cas are also available based on those disclosed for the CRISPR-Cas9 system in volume 1284 of the series Methods in Molecular Biology pp 239-255 10 February 2015. A detailed procedure to design, construct, and evaluate dual gRNAs for plant codon optimized Cas9 (pcoCas9) mediated genome editing using Arabidopsis thaliana and Nicotiana benthamiana protoplasts s model cellular systems are described. Strategies to apply the CRISPR-Cas9 system to generating targeted genome modifications in whole plants are also discussed. The protocols described in the chapter can be applied to the polynucleotide modifying agent(s) and systems described herein.

[0822] Sugano et al. (Plant Cell Physiol. 2014 Mar;55(3):475-81. doi: 10.1093/pcp/pcu014. Epub 2014 Jan 18) reports the application of CRISPR-Cas9 to targeted

mutagenesis in the liverwort Marchantia polymorpha L., which has emerged as a model species for studying land plant evolution. The U6 promoter of M. polymorpha was identified and cloned to express the gRNA. The target sequence of the gRNA was designed to disrupt the gene encoding auxin response factor 1 (ARF1) in M. polymorpha. Using Agrobacterium-mediated transformation, Sugano et al. isolated stable mutants in the gametophyte generation of M. polymorpha. CRISPR-Cas9-based site-directed mutagenesis *in vivo* was achieved using either the Cauliflower mosaic virus 35S or M. polymorpha EF1α promoter to express Cas9. Isolated mutant individuals showing an auxin-resistant phenotype were not chimeric. Moreover, stable mutants were produced by asexual reproduction of T1 plants. Multiple arf1 alleles were easily established using CRIPSR-Cas9-based targeted mutagenesis. The methods of Sugano et al. can be applied to the polynucleotide modifying agent(s) and systems described herein.

[0823] Lowder et al. (Plant Physiol. 2015 Aug 21. pii: pp.00636.2015) also developed a CRISPR-Cas9 toolbox enables multiplex genome editing and transcriptional regulation of expressed, silenced or non-coding genes in plants. This toolbox provides researchers with a protocol and reagents to quickly and efficiently assemble functional CRISPR-Cas9 T-DNA constructs for monocots and dicots using Golden Gate and Gateway cloning methods. It comes with a full suite of capabilities, including multiplexed gene editing and transcriptional activation or repression of plant endogenous genes. T-DNA based transformation technology is fundamental to modern plant biotechnology, genetics, molecular biology and physiology. As such, a method for the assembly of Cas (WT, nickase or dCas) and gRNA(s) into a T-DNA destination-vector of interest can be used. The assembly method is based on both Golden Gate assembly and MultiSite Gateway recombination. Three modules are required for assembly. The first module is a Cas entry vector, which contains promoterless Cas or its derivative genes flanked by attL1 and attR5 sites. The second module is a gRNA entry vector which contains entry gRNA expression cassettes flanked by attL5 and attL2 sites. The third module includes attR1-attR2-containing destination T-DNA vectors that provide promoters of choice for Cas expression. The toolbox of Lowder et al. can be applied to the polynucleotide modifying agent(s) and systems described herein.

[0824] Wang et al. (bioRxiv 051342; doi: doi.org/10.1101/051342; Epub. May 12, 2016) demonstrate editing of homoeologous copies of four genes affecting important agronomic traits in hexaploid wheat using a multiplexed gene editing construct with several gRNA-tRNA units

under the control of a single promoter. The methods of Wang et al. can be applied to the polynucleotide modifying agent(s) and systems described herein.

[0825] In an advantageous embodiment, the plant may be a tree. The present invention may also utilize the herein disclosed systems for herbaceous systems (see, e.g., Belhaj et al., Plant Methods 9: 39 and Harrison et al., Genes & Development 28: 1859–1872). In a particularly advantageous embodiment, the polynucleotide modifying agent(s) and systems thereof can target single nucleotide polymorphisms (SNPs) in trees (see, e.g., Zhou et al., New Phytologist, Volume 208, Issue 2, pages 298–301, October 2015). In the Zhou et al. study, the authors applied a systems in the woody perennial Populus using the 4-coumarate:CoA ligase (4CL) gene family as a case study and achieved 100% mutational efficiency for two 4CL genes targeted, with every transformant examined carrying biallelic modifications. In the Zhou et al., study, the CRISPR-Cas9 system was highly sensitive to single nucleotide polymorphisms (SNPs), as cleavage for a third 4CL gene was abolished due to SNPs in the target sequence. These methods Wang et al. (bioRxiv 051342; doi: doi.org/10.1101/051342; Epub. May 12, 2016) demonstrate editing of homoeologous copies of four genes affecting important agronomic traits in hexaploid wheat using a multiplexed gene editing construct with several gRNA-tRNA units under the control of a single promoter. These techniques and methods can be applied to the polynucleotide modifying agent(s) and systems described herein.

[0826] In particular embodiments, the polynucleotide modification systems described herein, can be used for self-cleavage. In these embodiments, the promotor of the Cas enzyme and gRNA can be a constitutive promotor and a second gRNA is introduced in the same transformation cassette, but controlled by an inducible promoter. This second gRNA can be designated to induce site-specific cleavage in the Cas gene in order to create a non-functional Cas. In a further particular embodiment, the second gRNA induces cleavage on both ends of the transformation cassette, resulting in the removal of the cassette from the host genome. This system offers a controlled duration of cellular exposure to the Cas enzyme and further minimizes off-target editing. Furthermore, cleavage of both ends of a CRISPR/Cas cassette can be used to generate transgene-free T0 plants with bi-allelic mutations (as described for Cas9 e.g. Moore et al., Nucleic Acids Research, 2014; Schaeffer et al., Plant Science, 2015). The methods of Moore et al. can be applied to the polynucleotide modifying agent(s) and systems described herein.

[0827] Kabadi et al. (Nucleic Acids Res. 2014 Oct 29;42(19):e147. doi: 10.1093/nar/gku749. Epub 2014 Aug 13) developed a single lentiviral system to express a Cas9 variant, a reporter gene and up to four sgRNAs from independent RNA polymerase III promoters that are incorporated into the vector by a convenient Golden Gate cloning method. Each sgRNA was efficiently expressed and can mediate multiplex gene editing and sustained transcriptional activation in immortalized and primary human cells. The methods of Kabadi et al. may be applied to the Cas effector protein system of the present invention.

[0828] Ling et al. (BMC Plant Biology 2014, 14:327) developed a CRISPR-Cas9 binary vector set based on the pGreen or pCAMBIA backbone, as well as a gRNA This toolkit requires no restriction enzymes besides BsaI to generate final constructs harboring maize-codon optimized Cas9 and one or more gRNAs with high efficiency in as little as one cloning step. The toolkit was validated using maize protoplasts, transgenic maize lines, and transgenic Arabidopsis lines and was shown to exhibit high efficiency and specificity. More importantly, using this toolkit, targeted mutations of three Arabidopsis genes were detected in transgenic seedlings of the T1 generation. Moreover, the multiple-gene mutations could be inherited by the next generation. (guide RNA) module vector set, as a toolkit for multiplex genome editing in plants. The toolbox of Lin et al. can be applied to the polynucleotide modifying agent(s) and systems described herein.

[0829] The methods of Zhou et al. (New Phytologist, Volume 208, Issue 2, pages 298–301, October 2015) may be applied to the present invention as follows. Two 4CL genes, 4CL1 and 4CL2, associated with lignin and flavonoid biosynthesis, respectively are targeted for CRISPR-Cas9 editing. The Populus tremula × alba clone 717-1B4 routinely used for transformation is divergent from the genome-sequenced Populus trichocarpa. Therefore, the 4CL1 and 4CL2 gRNAs designed from the reference genome are interrogated with in-house 717 RNA-Seq data to ensure the absence of SNPs which could limit Cas efficiency. A third gRNA designed for 4CL5, a genome duplicate of 4CL1, is also included. The corresponding 717 sequence harbors one SNP in each allele near/within the PAM, both of which are expected to abolish targeting by the 4CL5-gRNA. All three gRNA target sites are located within the first exon. For 717 transformation, the gRNA is expressed from the Medicago U6.6 promoter, along with a human codon-optimized Cas under control of the CaMV 35S promoter in a binary vector. Transformation with the Cas-only vector can serve as a control. Randomly selected 4CL1 and 4CL2 lines are subjected to amplicon-sequencing. The data is then processed and

biallelic mutations are confirmed in all cases. These methods can be applied to the polynucleotide modifying agent(s) and systems described herein.

[0830] The following table (**Table 6**) provides additional references and related fields for which the systems, complexes, modified effector proteins, systems, and methods of optimization may be used to generate modified non-animal organisms.

Table 6

2-17-2014	PCT/US15/63434 (WO2016/099887)	Compositions and methods for efficient gene editing in E. coli using guide RNA/Cas endonuclease systems in combination with circular polynucleotide modification templates.
08-13-2014	PCT/US15/41256 (WO2016/025131)	Genetic targeting in non-conventional yeast using an RNA-guided endonuclease.
11-06-2014	PCT/US15/58760 (WO2016/073433)	Peptide-mediated delivery of RNA-guided endonuclease into cells.
10-12-2015	PCT/US16/56404 (WO2017/066175)	Protected DNA templates for gene modification and increased homologous recombination in cells and methods of use.
12-11-2015	PCT/US16/65070 (WO2017/100158)	Methods and compositions for enhanced nuclease-mediated genome modification and reduced off-target site effects.
12-18-2015	PCT/US16/65537 (WO 2017/105991)	Methods and compositions for T-RNA based guide RNA expression.
12-18-2015	PCT/US16/66772 (WO2017/106414)	Methods and compositions for polymerase II (Pol-II) based guide RNA expression.
12-16-2014	PCT/US15/65693 (WO2016/100272)	Fungal genome modification systems and methods of use.
12-16-2014	PCT/US15/66195 (WO2016/100571)	Fungal genome modification systems and methods of use
12-16-2014	PCT/US15/66192 (WO 2016/100568)	Fungal genome modification systems and methods of use.
12-16-2014	PCT/US15/66178 (WO 2016/100562)	Use of a helper strain with silenced NHEJ to improve homologous integration of targeted DNA cassettes in Trichoderma reesei.
07-28-2015	PCT/US16/44489 (WO 2017/019867)	Genome editing systems and methods of use.

Detecting modifications in the plant genome- selectable markers

[0831] In particular embodiments, a selectable marker can be included or introduced to allow for identification of modified cells. Selectable markers can be advantageous for many

situations, such as when the modification is made to an endogenous target gene of the plant genome. Any suitable method can be used to determine, after the plant, plant part or plant cell is infected or transfected with the system, whether gene targeting or targeted mutagenesis has occurred at the target site.

[0832] Where the method involves introduction of a transgene, a transformed plant cell, callus, tissue or plant may be identified and isolated by selecting or screening the engineered plant material for the presence of the transgene or for traits encoded by the transgene. Physical and biochemical methods may be used to identify plant or plant cell transformants containing inserted gene constructs or an endogenous DNA modification. These methods include but are not limited to: 1) Southern analysis or PCR amplification for detecting and determining the structure of the recombinant DNA insert or modified endogenous genes; 2) Northern blot, S1 RNase protection, primer-extension or reverse transcriptase-PCR amplification for detecting and examining RNA transcripts of the gene constructs; 3) enzymatic assays for detecting enzyme or ribozyme activity, where such gene products are encoded by the gene construct or expression is affected by the genetic modification; 4) protein gel electrophoresis, Western blot techniques, immunoprecipitation, or enzyme-linked immunoassays, where the gene construct or endogenous gene products are proteins. Additional techniques, such as in situ hybridization, enzyme staining, and immunostaining, also may be used to detect the presence or expression of the recombinant construct or detect a modification of endogenous gene in specific plant organs and tissues. The methods for doing all these assays are well known to those skilled in the art.

[0833] In some embodiments, the expression system encoding the polynucleotide modifying agent and/or system components can be designed to comprise one or more selectable or detectable markers that provide a means to isolate or efficiently select cells that contain and/or have been modified by the system at an early stage and on a large scale.

[0834] In the case of Agrobacterium-mediated transformation, the marker cassette may be adjacent to or between flanking T-DNA borders and contained within a binary vector. In another embodiment, the marker cassette may be outside of the T-DNA. A selectable marker cassette may also be within or adjacent to the same T-DNA borders as the expression cassette or may be somewhere else within a second T-DNA on the binary vector (e.g., a 2 T-DNA system).

[0835] For particle bombardment or with protoplast transformation, the expression system can include one or more isolated linear fragments or may be part of a larger construct that might contain bacterial replication elements, bacterial selectable markers or other detectable elements. The expression cassette(s) comprising the polynucleotide(s) encoding the polynucleotide modifying agents(s), system component(s), or system can be physically linked to a marker cassette or may be mixed with a second nucleic acid molecule encoding a marker cassette. The marker cassette can include the necessary elements to express a detectable or selectable marker that allows for efficient selection of transformed cells. Such elements will be appreciated by one of ordinary skill in the art.

[0836] The selection procedure for the cells based on the selectable marker will depend on the nature of the marker gene. In particular embodiments, use is made of a selectable marker, i.e. a marker which allows a direct selection of the cells based on the expression of the marker. A selectable marker can confer positive or negative selection and is conditional or non-conditional on the presence of external substrates (Miki et al. 2004, 107(3): 193–232). Most commonly, antibiotic or herbicide resistance genes are used as a marker, whereby selection is be performed by growing the engineered plant material on media containing an inhibitory amount of the antibiotic or herbicide to which the marker gene confers resistance. Examples of such genes are genes 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),

[0837] Transformed plants and plant cells can also be identified by screening for the activities of a visible marker, typically an enzyme capable of processing a colored substrate (e.g., the β-glucuronidase, luciferase, B or C1 genes). Such selection and screening methodologies are well known to those skilled in the art.

Plant cultures and regeneration

[0838] In particular embodiments, plant cells which have a modified genome and that are produced or obtained by any of the methods described herein, can be cultured to regenerate a whole plant which possesses the transformed or modified genotype and thus the desired phenotype. Conventional regeneration techniques are well known to those skilled in the art. Particular examples of such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, and typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. In

further particular embodiments, plant regeneration is obtained from cultured protoplasts, plant callus, explants, organs, pollens, embryos or parts thereof (see e.g. Evans et al. (1983), Handbook of Plant Cell Culture, Klee et al (1987) Ann. Rev. of Plant Phys.).

[0839] In particular embodiments, transformed or improved plants as described herein can be self-pollinated to provide seed for homozygous improved plants of the invention (homozygous for the DNA modification) or crossed with non-transgenic plants or different improved plants to provide seed for heterozygous plants. Where a recombinant DNA was introduced into the plant cell, the resulting plant of such a crossing is a plant which is heterozygous for the recombinant DNA molecule. Both such homozygous and heterozygous plants obtained by crossing from the improved plants and comprising the genetic modification (which can be a recombinant DNA) are referred to herein as "progeny". Progeny plants are plants descended from the original transgenic plant and containing the genome modification or recombinant DNA molecule introduced by the methods provided herein. Alternatively, genetically modified plants can be obtained by one of the methods described supra using the Cfp1 enzyme whereby no foreign DNA is incorporated into the genome. Progeny of such plants, obtained by further breeding may also contain the genetic modification. Breedings are performed by any breeding methods that are commonly used for different crops (e.g., Allard, Principles of Plant Breeding, John Wiley & Sons, NY, U. of CA, Davis, CA, 50-98 (1960).

Non-human Animals

[0840] The systems and methods may be used to generate modified non-human animals and cells thereof. In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example, a mammal. Also, the organism may be an arthropod such as an insect. The present invention may also be extended to other agricultural applications such as, for example, farm and production animals. For example, pigs have many features that make them attractive as biomedical models, especially in regenerative medicine. In particular, pigs with severe combined immunodeficiency (SCID) may provide useful models for regenerative medicine, xenotransplantation (discussed also elsewhere herein), and tumor

development and will aid in developing therapies for human SCID patients. Lee et al., (Proc Natl Acad Sci U S A. 2014 May 20;111(20):7260-5) utilized a reporter-guided transcription activator-like effector nuclease (TALEN) system to generated targeted modifications of recombination activating gene (RAG) 2 in somatic cells at high efficiency, including some that affected both alleles. Such techniques and modifications can be adapted for and used with the modifying agent(s) and systems thereof described herein to generate a modified non-human animal or cell thereof.

[0841] The methods of Lee et al., (Proc Natl Acad Sci U S A. 2014 May 20;111(20):7260-5) may be applied to the present invention analogously as follows. Mutated pigs are produced by targeted insertion for example in RAG2 in fetal fibroblast cells followed by SCNT and embryo transfer. Constructs coding for CRISPR Cas and a reporter are electroporated into fetal-derived fibroblast cells. After 48 h, transfected cells expressing the green fluorescent protein are sorted into individual wells of a 96-well plate at an estimated dilution of a single cell per well. Targeted modification of RAG2 are screened by amplifying a genomic DNA fragment flanking any CRISPR Cas cutting sites followed by sequencing the PCR products. After screening and ensuring lack of off-site mutations, cells carrying targeted modification of RAG2 are used for SCNT. The polar body, along with a portion of the adjacent cytoplasm of oocyte, presumably containing the metaphase II plate, are removed, and a donor cell are placed in the perivitelline. The reconstructed embryos are then electrically porated to fuse the donor cell with the oocyte and then chemically activated. The activated embryos are incubated in Porcine Zygote Medium 3 (PZM3) with 0.5 μM Scriptaid (S7817; Sigma-Aldrich) for 14–16 h. Embryos are then washed to remove the Scriptaid and cultured in PZM3 until they were transferred into the oviducts of surrogate pigs. Such techniques and modifications can be adapted for and used with the modifying agent(s) and systems thereof described herein to generate a modified non-human animal or cell thereof.

[0842] The modified non-human animals described herein can be a platform to model a disease or disorder of an animal, including but not limited to mammals. In some of these embodiments, the mammal can be a human. In certain embodiments, such models and platforms are rodent based, in non-limiting examples rat or mouse. Such models and platforms can take advantage of distinctions among and comparisons between inbred rodent strains. In certain embodiments, such models and platforms primate, horse, cattle, sheep, goat, swine, dog, cat or bird-based, for example to directly model diseases and disorders of such animals or

to create modified and/or improved lines of such animals. Advantageously, in certain embodiments, an animal-based platform or model is created to mimic a human disease or disorder. For example, the similarities of swine to humans make swine an ideal platform for modeling human diseases. Compared to rodent models, development of swine models has been costly and time intensive. On the other hand, swine and other animals are much more similar to humans genetically, anatomically, physiologically and pathophysiologically. The present invention provides a high efficiency platform for targeted gene and genome editing, gene and genome modification and gene and genome regulation to be used in such animal platforms and models. Though ethical standards block development of human models and in many cases models based on non-human primates, the present invention is used with *in vitro* systems, including but not limited to cell culture systems, three dimensional models and systems, and organoids to mimic, model, and investigate genetics, anatomy, physiology and pathophysiology of structures, organs, and systems of humans. The platforms and models provide manipulation of single or multiple targets.

[0843] In certain embodiments, the present invention is applicable to disease models like that of Schomberg et al. (FASEB Journal, April 2016; 30(1):Suppl 571.1). To model the inherited disease neurofibromatosis type 1 (NF-1) Schomberg used CRISPR-Cas9 to introduce mutations in the swine neurofibromin 1 gene by cytosolic microinjection of CRISPR/Cas9 components into swine embryos. CRISPR guide RNAs (gRNA) were created for regions targeting sites both upstream and downstream of an exon within the gene for targeted cleavage by Cas9 and repair was mediated by a specific single-stranded oligodeoxynucleotide (ssODN) template to introduce a 2500 bp deletion. The systems were also used to engineer swine with specific NF-1 mutations or clusters of mutations, and further can be used to engineer mutations that are specific to or representative of a given human individual. Such techniques and modifications can be adapted for and used with the modifying agent(s) and systems thereof described herein to generate a modified non-human animal or cell thereof. In some embodiments, the polynucleotide modifying agent(s) or systems thereof can be similarly used to develop animal models, including but not limited to swine models, of human multigenic diseases. In some embodiments, multiple genetic loci in one gene or in multiple genes are simultaneously targeted using multiplexed guides and optionally one or multiple templates.

[0844] SNPs of other animals, such as cows can also be modified or generated using one or more polynucleotide modifying agents or systems described herein. Tan et al. (Proc Natl

Acad Sci U S A. 2013 Oct 8; 110(41): 16526–16531) expanded the livestock gene editing toolbox to include transcription activator-like (TAL) effector nuclease (TALEN)- and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9- stimulated homology-directed repair (HDR) using plasmid, rAAV, and oligonucleotide templates. Gene specific gRNA sequences were cloned into the Church lab gRNA vector (Addgene ID: 41824) according to their methods (Mali P, et al. (2013) RNA-Guided Human Genome Engineering via Cas9. Science 339(6121):823-826). The Cas9 nuclease was provided either by co-transfection of the hCas9 plasmid (Addgene ID: 41815) or mRNA synthesized from RCIScript-hCas9. This RCIScript-hCas9 was constructed by sub-cloning the XbaI-AgeI fragment from the hCas9 plasmid (encompassing the hCas9 cDNA) into the RCIScript plasmid. Such techniques and modifications can be adapted for and used with the modifying agent(s) and systems thereof described herein to generate a modified non-human animal or cell thereof.

[0845] Heo et al. (Stem Cells Dev. 2015 Feb 1;24(3):393-402. doi: 10.1089/scd.2014.0278. Epub 2014 Nov 3) reported highly efficient gene targeting in the bovine genome using bovine pluripotent cells and clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 nuclease. First, Heo et al. generate induced pluripotent stem cells (iPSCs) from bovine somatic fibroblasts by the ectopic expression of yamanaka factors and GSK3β and MEK inhibitor (2i) treatment. Heo et al. observed that these bovine iPSCs are highly similar to naïve pluripotent stem cells with regard to gene expression and developmental potential in teratomas. Moreover, CRISPR-Cas9 nuclease, which was specific for the bovine NANOG locus, showed highly efficient editing of the bovine genome in bovine iPSCs and embryos. Such techniques and modifications can be adapted for and used with the modifying agent(s) and systems thereof described herein to generate a modified non-human animal or cell thereof.

[0846] Igenity® provides a profile analysis of animals, such as cows, to perform and transmit traits of economic traits of economic importance, such as carcass composition, carcass quality, maternal and reproductive traits and average daily gain. The analysis of a comprehensive Igenity® profile begins with the discovery of DNA markers (most often single nucleotide polymorphisms or SNPs). All the markers behind the Igenity® profile were discovered by independent scientists at research institutions, including universities, research organizations, and government entities such as USDA. Markers are then analyzed at Igenity® in validation populations. Igenity® uses multiple resource populations that represent various production environments and biological types, often working with industry partners from the

seedstock, cow-calf, feedlot and/or packing segments of the beef industry to collect phenotypes that are not commonly available. Cattle genome databases are widely available, see, e.g., the NAGRP Cattle Genome Coordination Program (www.animalgenome.org/cattle/maps/db.html). Thus, the polynucleotide modifying agent(s) and/or systems described herein can be applied to target bovine SNPs. One of skill in the art may utilize the above protocols for targeting SNPs and apply them to bovine SNPs as described, for example, by Tan et al. or Heo et al.

[0847] Qingjian Zou et al. (Journal of Molecular Cell Biology Advance Access published October 12, 2015) demonstrated increased muscle mass in dogs by targeting the first exon of the dog Myostatin (MSTN) gene (a negative regulator of skeletal muscle mass). First, the efficiency of the sgRNA was validated, using cotransfection of the sgRNA targeting MSTN with a Cas9 vector into canine embryonic fibroblasts (CEFs). Thereafter, MSTN KO dogs were generated by micro-injecting embryos with normal morphology with a mixture of Cas9 mRNA and MSTN sgRNA and auto-transplantation of the zygotes into the oviduct of the same female dog. The knock-out puppies displayed an obvious muscular phenotype on thighs compared with its wild-type littermate sister. This can also be performed using the polynucleotide agent(s) and/or systems provided herein. Such techniques and modifications can be adapted for and used with the modifying agent(s) and systems thereof described herein to generate a modified non-human animal or cell thereof.

Livestock

[0848] Also described herein are modified pigs or cells that can express one or more polynucleotides, genes or alleles of interest. Exemplary genes of interest are described in greater detail elsewhere herein. As reported by Kristin M Whitworth and Dr Randall Prather et al. (Nature Biotech 3434 published online 07 December 2015) CD163 (a viral target) was targeted using CRISPR-Cas9 and the offspring of edited pigs were resistant when exposed to PRRSv. One founder male and one founder female, both of whom had mutations in exon 7 of CD163, were bred to produce offspring. The founder male possessed an 11-bp deletion in exon 7 on one allele, which results in a frameshift mutation and missense translation at amino acid 45 in domain 5 and a subsequent premature stop codon at amino acid 64. The other allele had a 2-bp addition in exon 7 and a 377-bp deletion in the preceding intron, which were predicted to result in the expression of the first 49 amino acids of domain 5, followed by a premature stop code at amino acid 85. The sow had a 7 bp addition in one allele that when translated was

predicted to express the first 48 amino acids of domain 5, followed by a premature stop codon at amino acid 70. The sow's other allele was unamplifiable. Selected offspring were predicted to be a null animal (CD163–/–), i.e. a CD163 knock out. Such techniques and modifications can be adapted for and used with the modifying agent(s) and systems thereof described herein to generate a modified pig that can express a polynucleotide of interest. Thus, also described herein are modified pigs their progeny that also express one or more copies of the gene or allele of interest. This may be for livestock, breeding or modelling purposes (i.e. a porcine model). Semen comprising the modification (e.g. polynucleotide of interest) is also provided.

Other Animals

[0849] Also described herein are other non-human animals that are modified to express one or more polynucleotides, genes or alleles of interest. Genes of interest are described elsewhere herein. Suitable polynucleotide modifying agent(s) and/or system thereof described elsewhere herein can be used to generate other non-human animals such as non-human primates, chickens (reviewed in Sid and Schusser et al 2018. Front. Genet. Doi.org/10.3389/fgene.2018.00456) and other avians (e.g. Scott et al. 2010. ILAR J. 51(4):353-361), cattle (Yum et al., 2016. Scientific Reports. 6:27185 and Tait-Burkard et al. 2018. Genome Biology. 19:2014.), sheep and goats (see e.g. Kalds et al., 2019. Front. Genet. Doi.org//10.3389/fgene.2019.00750), horses (see e.g. West and Gill. 2016. J. Equine Vet. Sci. 41:1-6), dogs (see e.g. D. Duan. Nature Biomedical Engineering. 2018. 2: 795-796), reptiles (see e.g. Rasys et al. 2019. Cell Reports. 28:2288-2292), fish (including but not limited to zebrafish, see e.g. Datsomor et al. 2019. Scientific Reports. 9:7533, Liu et al. 2019. Front. Cell. Dev. Biol. https://doi.org/10.3389/fcell.2019.00013), insects (see e.g. Kotwica-Rolinska et al. 2019. Front. Physiol. https://doi.org/10.3389/fphys.2019.00891; Gantz and Akbari. 2018. Curr. Opin. Insect. Sci. 28:66-72), rabbits (see e.g. Kawano and Honda. 2017. Methods Mol. Biol. 4630:109-120; Liu et al., 2018. Nature Commun. 9:2717; and Liu et al. 2018. Gene. https://doi.org/10.1016/j.gene.2018.01.044), mice (see e.g. Hall et al. 2018. Curr Protoc Cell Biol. 81(1): e57), rats (see e.g. Back et al. 2019. Neuron. 102(1):105-119), amphibians (see e.g. Nakayama et al. 2013. Genesis. 51(12):835-843), nematodes (see e.g. J.B. Lok. 2019. Front. Genet. https://doi.org/10.3389/fgene.2019.00656), molluscs (see e.g. Abe and Kuroda. 2019. Development. 146: dev175976 doi: 10.1242/dev.175976, geckos, shrimp and other crustaceans (see e.g. Gui et al. Genes Genomes Genetics: 6(11): 3757-3764), oysters (Yu et al. 2019; Mar. Biotechnol (NY) 21(3):301-309. doi: 10.1007/s10126-019-09885-y), and sponges

(see e.g. Revilla-i-Domingo et al. 2018. Genetics. 210(2)435-443), the teachings of which can be adapted for use with one or more of the modifying agent(s) and/or systems described herein to generate the modified non-human animal or cell thereof.

KITS

[0850] Also described herein are kits that contain one or more of the Type VII CRISPR-Cas system polypeptides, polynucleotides, vectors, cells, or other components described herein and combinations thereof and pharmaceutical formulations described herein. In aspects, one or more of the Type VII CRISPR-Cas polypeptides, polynucleotides, vectors, cells, modulating agents, neurodegenerative disease therapeutic, or other components described herein and combinations thereof and pharmaceutical formulations described herein can be presented as a combination kit. As used herein, the terms "combination kit" or "kit of parts" refers to the compounds, or formulations and additional components that are used to package, screen, test, sell, market, deliver, and/or administer the combination of elements or a single element, such as the active ingredient, contained therein. Such additional components include but are not limited to, packaging, syringes, blister packages, bottles, and the like. The combination kit can contain one or more of the components (e.g. one or more of the one or more of the Type VII CRISPR-Cas polypeptides, polynucleotides, vectors, cells, modulating agents, neurodegenerative disease therapeutic, or other components described herein and combinations thereof and pharmaceutical formulations described herein) can be provided in a single formulation (e.g. a liquid, lyophilized powder, etc.), or in separate formulations. The separate components or formulations can be contained in a single package or in separate packages within the kit. The kit can also include instructions in a tangible medium of expression that can contain information and/or directions regarding the content of the components and/or formulations contained therein, safety information regarding the content of the components(s) and/or formulation(s) contained therein, information regarding the amounts, dosages, indications for use, screening methods, component design recommendations and/or information, recommended treatment regimen(s) for the components(s) and/or formulations contained therein. As used herein, “tangible medium of expression” refers to a medium that is physically tangible or accessible and is not a mere abstract thought or an unrecorded spoken word. “Tangible medium of expression” includes, but is not limited to, words on a cellulosic or plastic material, or data stored in a suitable computer readable memory form. The data can

be stored on a unit device, such as a flash memory drive or CD-ROM or on a server that can be accessed by a user via, e.g. a web interface.

[0851] In one aspect, the invention provides a kit comprising one or more of the components described herein. In some embodiments, the kit comprises one or more of the Type VII CRISPR-Cas polypeptides, polynucleotides, vectors, cells, modulating agents, neurodegenerative disease therapeutic, or other components described herein and combinations thereof and pharmaceutical formulations described herein and instructions for using the kit. In some embodiments, the kit includes instructions for modulating or modifying a polynucleotide using, administering, and/or contacting the cell with a Type VII CRISPR-Cas system or component thereof, vector or vector system, and/or formulation described herein. In some aspects, the instructions provide that modulation occurs by exposing a cell to a Type VII CRISPR-Cas. In some aspects, the instructions provide that the administering, contacting, and/or exposure can occur *in vitro*, *in vivo*, *ex vivo*, or *in situ*. In some aspects, the instructions provide that a subject in need thereof can be treated by administering one or more of the kit components to the subject in need thereof. The instructions can provide that one or more components of the kit can be used to treat a subject in need thereof, where the subject in need thereof has a disease, infection, condition, or disorder. In some aspects, the instructions provide that the disease, infection, condition, or disorder is a genetic or epigenetic disease. Specific examples of diseases are described in greater detail elsewhere herein.

METHODS OF USING THE TYPE VII SYSTEMS

[0852] The Type VII CRISPR-Cas systems and components thereof can be used to modify a polynucleotide, for targeted-delivery of an effector molecule to a target polynucleotide, screening methods, as therapies and preventions, as well as to generate modified cells and organisms. Other applications and uses of the Type VII CIRPSR-Cas systems are described here and elsewhere herein.

Models of Genetic and Epigenetic Conditions

[0853] A method of the invention may be used to create a plant, an animal or cell that may be used to model and/or study genetic or epigenetic conditions of interest, such as a through a model of mutations of interest or a disease model. As used herein, "disease" refers to a disease, disorder, or indication in a subject. For example, a method of the invention may be used to create an animal or cell that comprises a modification in one or more nucleic acid sequences associated with a disease, or a plant, animal or cell in which the expression of one or more

nucleic acid sequences associated with a disease are altered. Such a nucleic acid sequence may encode a disease associated protein sequence or may be a disease associated control sequence. Accordingly, it is understood that in embodiments of the invention, a plant, subject, patient, organism or cell can be a non-human subject, patient, organism or cell. Thus, the invention provides a plant, animal or cell, produced by the present methods, or a progeny thereof. The progeny may be a clone of the produced plant or animal, or may result from sexual reproduction by crossing with other individuals of the same species to introgress further desirable traits into their offspring. The cell may be *in vivo* or *ex vivo* in the cases of multicellular organisms, particularly animals or plants. In the instance where the cell is in culture, 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.

[0854] 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.

[0855] 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.

[0856] In another embodiment, this invention provides a method of developing a biologically active agent that modulates a cell signaling event associated with a disease gene. The method comprises contacting a test compound with a cell comprising one or more vectors that drive expression of one or more of a CRISPR enzyme, and a direct repeat sequence linked to a guide sequence; and detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with, e.g., a mutation in a disease gene contained in the cell.

[0857] A cell model or animal model can be constructed in combination with the method of the invention for screening a cellular function change. Such a model may be used to study

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

[0858] 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.

[0859] To assay for an agent-induced alteration in the level of mRNA transcripts or corresponding polynucleotides, nucleic acid contained in a sample is first extracted according to standard methods in the art. For instance, mRNA can be isolated using various lytic enzymes or chemical solutions according to the procedures set forth in Sambrook et al. (1989), or extracted by nucleic-acid-binding resins following the accompanying instructions provided by the manufacturers. The mRNA contained in the extracted nucleic acid sample is then detected by amplification procedures or conventional hybridization assays (e.g. Northern blot analysis) according to methods widely known in the art or based on the methods exemplified herein.

[0860] 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.

[0861] Detection of the gene expression level can be conducted in real time in an amplification assay. In one aspect, the amplified products can be directly visualized with fluorescent DNA-binding agents including but not limited to DNA intercalators and DNA groove binders. Because the amount of the intercalators incorporated into the double-stranded DNA molecules is typically proportional to the amount of the amplified DNA products, one can conveniently determine the amount of the amplified products by quantifying the fluorescence of the intercalated dye using conventional optical systems in the art. DNA-binding dye suitable for this application include SYBR green, SYBR blue, DAPI, propidium iodine, Hoeste, SYBR gold, ethidium bromide, acridines, proflavine, acridine orange, acriflavine, fluorcoumanin, ellipticine, daunomycin, chloroquine, distamycin D, chromomycin, homidium, mithramycin, ruthenium polypyridyls, anthramycin, and the like.

[0862] 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.

[0863] In yet another aspect, conventional hybridization assays using hybridization probes that share sequence homology with sequences associated with a signaling biochemical pathway can be performed. Typically, probes are allowed to form stable complexes with the sequences associated with a signaling biochemical pathway contained within the biological sample derived from the test subject in a hybridization reaction. It will be appreciated by one of skill in the art that where antisense is used as the probe nucleic acid, the target polynucleotides provided in the sample are chosen to be complementary to sequences of the antisense nucleic acids. Conversely, where the nucleotide probe is a sense nucleic acid, the target polynucleotide is selected to be complementary to sequences of the sense nucleic acid.

[0864] Hybridization can be performed under conditions of various stringency. Suitable hybridization conditions for the practice of the present invention are such that the recognition interaction between the probe and sequences associated with a signaling biochemical pathway is both sufficiently specific and sufficiently stable. Conditions that increase the stringency of a hybridization reaction are widely known and published in the art. See, for example, (Sambrook,

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

[0865] 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, ß-galactosidase, urease, alkaline phosphatase or peroxidase, avidin/biotin complex.

[0866] 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 phosphoimager. 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.

[0867] 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.

[0868] 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.

[0869] 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.

[0870] 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.

[0871] A number of techniques for protein analysis based on the general principles outlined above are available in the art. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunofluorescent assays, and SDS-PAGE.

[0872] 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-2α). 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.

[0873] In practicing the subject method, it may be desirable to discern the expression pattern of an protein associated with a signaling biochemical pathway in different bodily tissue, in different cell types, and/or in different subcellular structures. These studies can be performed with the use of tissue-specific, cell-specific or subcellular structure specific antibodies capable of binding to protein markers that are preferentially expressed in certain tissues, cell types, or subcellular structures.

[0874] An altered expression of a gene associated with a signaling biochemical pathway can also be determined by examining a change in activity of the gene product relative to a control cell. The assay for an agent-induced change in the activity of a protein associated with a signaling biochemical pathway will dependent on the biological activity and/or the signal transduction pathway that is under investigation. For example, where the protein is a kinase, a change in its ability to phosphorylate the downstream substrate(s) can be determined by a variety of assays known in the art. Representative assays include but are not limited to immunoblotting and immunoprecipitation with antibodies such as anti-phosphotyrosine antibodies that recognize phosphorylated proteins. In addition, kinase activity can be detected by high throughput chemiluminescent assays such as AlphaScreen™ (available from Perkin Elmer) and eTag™ assay (Chan-Hui, et al. (2003) Clinical Immunology 111: 162-174).

[0875] Where the protein associated with a signaling biochemical pathway is part of a signaling cascade leading to a fluctuation of intracellular pH condition, pH sensitive molecules such as fluorescent pH dyes can be used as the reporter molecules. In another example where the protein associated with a signaling biochemical pathway is an ion channel, fluctuations in membrane potential and/or intracellular ion concentration can be monitored. A number of commercial kits and high-throughput devices are particularly suited for a rapid and robust screening for modulators of ion channels. Representative instruments include FLIPRTM

(Molecular Devices, Inc.) and VIPR (Aurora Biosciences). These instruments are capable of detecting reactions in over 1000 sample wells of a microplate simultaneously, and providing real-time measurement and functional data within a second or even a minisecond.

[0876] In practicing any of the methods disclosed herein, a suitable vector can be introduced to a cell or an embryo via one or more methods known in the art, including without limitation, microinjection, electroporation, sonoporation, biolistics, calcium phosphate-mediated transfection, cationic transfection, liposome transfection, dendrimer transfection, heat shock transfection, nucleofection transfection, magnetofection, lipofection, impalefection, optical transfection, proprietary agent-enhanced uptake of nucleic acids, and delivery via liposomes, immunoliposomes, virosomes, or artificial virions. In some methods, the vector is introduced into an embryo by microinjection. The vector or vectors may be microinjected into the nucleus or the cytoplasm of the embryo. In some methods, the vector or vectors may be introduced into a cell by nucleofection. Vectors, vector preparation, and delivery are described in greater detail elsewhere herein.

[0877] The target polynucleotide of a CRISPR complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA). CRISPR-Cas system target polynucleotides are described in greater detail elsewhere herein. In some embodiments, the target polynucleotide is a disease-associated polynucleotide (also referred to herein as a disease-associated polynucleotide gene). Without wishing to be bound by theory, it is believed that the target sequence should be associated with a PAM (protospacer adjacent motif); that is, a short sequence recognized by the CRISPR complex. The precise sequence and length requirements for the PAM differ depending on the CRISPR enzyme used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence) Examples of PAM sequences are given in the examples section below, and the skilled person will be able to identify further PAM sequences for use with a given CRISPR enzyme. Further, engineering of the PAM Interacting (PI) domain may allow programing of PAM specificity, improve target site recognition fidelity, and increase the versatility of the Cas, e.g. Cas9, genome engineering platform. Cas proteins, such as Cas9 proteins may be engineered to alter their PAM specificity,

for example as described in Kleinstiver BP et al. *Engineered CRISPR-Cas9 nucleases with altered PAM specificities*. Nature. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592.

[0878] 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.

[0879] The target polynucleotide of a CRISPR complex may include a number of disease-associated genes and polynucleotides as well as signaling biochemical pathway-associated genes and polynucleotides as listed in US provisional patent applications 61/736,527 and 61/748,427 having Broad reference BI-2011/008/WSGR Docket No. 44063-701.101 and BI-2011/008/WSGR Docket No. 44063-701.102 respectively, both entitled SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION filed on December 12, 2012 and January 2, 2013, respectively, and PCT Application PCT/US2013/074667, entitled DELIVERY, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION AND THERAPEUTIC APPLICATIONS, filed December 12, 2013, the contents of all of which are herein incorporated by reference in their entirety.

Treatment and/or Prevention of Pathogens

HIV

[0880] Cas-mediated genome editing might be used to introduce protective mutations in somatic tissues to combat nongenetic or complex diseases. For example, NHEJ-mediated inactivation of the CCR5 receptor in lymphocytes (Lombardo et al., Nat Biotechnol. 2007 Nov; 25(11):1298-306) may be a viable strategy for circumventing HIV infection, whereas deletion

of PCSK9 (Cohen et al., Nat Genet. 2005 Feb; 37(2):161-5) orangiopoietin (Musunuru et al., N Engl J Med. 2010 Dec 2; 363(23):2220-7) may provide therapeutic effects against statin-resistant hypercholesterolemia or hyperlipidemia. Although these targets may be also addressed using siRNA-mediated protein knockdown, a unique advantage of NHEJ-mediated gene inactivation is the ability to achieve permanent therapeutic benefit without the need for continuing treatment. As with all gene therapies, it will of course be important to establish that each proposed therapeutic use has a favorable benefit-risk ratio.

[0881] Hydrodynamic delivery of plasmid DNA encoding Cas9 and guide RNA along with a repair template into the liver of an adult mouse model of tyrosinemia was shown to be able to correct the mutant Fah gene and rescue expression of the wild-type Fah protein in ~1 out of 250 cells (Nat Biotechnol. 2014 Jun; 32(6):551-3). In addition, clinical trials successfully used ZF nucleases to combat HIV infection by *ex vivo* knockout of the CCR5 receptor. In all patients, HIV DNA levels decreased, and in one out of four patients, HIV RNA became undetectable (Tebas et al., N Engl J Med. 2014 Mar 6; 370(10):901-10). Both of these results demonstrate the promise of programmable nucleases as a new therapeutic platform.

[0882] In another embodiment, self-inactivating lentiviral vectors with an siRNA targeting a common exon shared by HIV tat/rev, a nucleolar-localizing TAR decoy, and an anti–CCR5-specific hammerhead ribozyme (see, e.g., DiGiusto et al. (2010) Sci Transl Med 2:36ra43) may be used/and or adapted to the CRISPR-Cas system of the present invention. A minimum of 2.5 × 106 CD34+ cells per kilogram patient weight may be collected and prestimulated for 16 to 20 hours in X-VIVO 15 medium (Lonza) containing 2 μmol/L-glutamine, stem cell factor (100 ng/ml), Flt-3 ligand (Flt-3L) (100 ng/ml), and thrombopoietin (10 ng/ml) (CellGenix) at a density of 2 × 106 cells/ml. Prestimulated cells may be transduced with lentiviral at a multiplicity of infection of 5 for 16 to 24 hours in 75-cm2 tissue culture flasks coated with fibronectin (25 mg/cm2) (RetroNectin,Takara Bio Inc.).

[0883] With the knowledge in the art and the teachings in this disclosure the skilled person can correct HSCs as to immunodeficiency condition such as HIV / AIDS comprising contacting an HSC with a Type VII CRISPR system that targets and knocks out CCR5. An guide RNA (and advantageously a dual guide approach, e.g., a pair of different guide RNAs; for instance, guide RNAs targeting of two clinically relevant genes, B2M and CCR5, in primary human CD4+ T cells and CD34+ hematopoietic stem and progenitor cells (HSPCs)) that targets and knocks out CCR5-and-Type VII effector containing particle is contacted with HSCs. The so

contacted cells can be administered; and optionally treated / expanded; cf. Cartier. See also Kiem, "Hematopoietic stem cell-based gene therapy for HIV disease," Cell Stem Cell. Feb 3, 2012; 10(2): 137–147; incorporated herein by reference along with the documents it cites; Mandal et al, "Efficient Ablation of Genes in Human Hematopoietic Stem and Effector Cells using CRISPR/Cas9," Cell Stem Cell, Volume 15, Issue 5, p643–652, 6 November 2014; incorporated herein by reference along with the documents it cites. Mention is also made of Ebina, "CRISPR/Cas9 system to suppress HIV-1 expression by editing HIV-1 integrated proviral DNA" SCIENTIFIC REPORTS | 3 : 2510 | DOI: 10.1038/srep02510, incorporated herein by reference along with the documents it cites, as another means for combatting HIV/AIDS using a CRISPR-Type VII effector system.

[0884] The rationale for genome editing for HIV treatment originates from the observation that individuals homozygous for loss of function mutations in CCR5, a cellular co-receptor for the virus, are highly resistant to infection and otherwise healthy, suggesting that mimicking this mutation with genome editing could be a safe and effective therapeutic strategy [Liu, R., et al. Cell 86, 367-377 (1996)]. This idea was clinically validated when an HIV infected patient was given an allogeneic bone marrow transplant from a donor homozygous for a loss of function CCR5 mutation, resulting in undetectable levels of HIV and restoration of normal CD4 T-cell counts [Hutter, G., et al. The New England journal of medicine 360, 692-698 (2009)]. Although bone marrow transplantation is not a realistic treatment strategy for most HIV patients, due to cost and potential graft vs. host disease, HIV therapies that convert a patient's own T-cells into CCR5 are desirable.

[0885] Early studies using ZFNs and NHEJ to knockout CCR5 in humanized mouse models of HIV showed that transplantation of CCR5 edited CD4 T cells improved viral load and CD4 T-cell counts [Perez, E.E., et al. Nature biotechnology 26, 808-816 (2008)]. Importantly, these models also showed that HIV infection resulted in selection for CCR5 null cells, suggesting that editing confers a fitness advantage and potentially allowing a small number of edited cells to create a therapeutic effect.

[0886] As a result of this and other promising preclinical studies, genome editing therapy that knocks out CCR5 in patient T cells has now been tested in humans [Holt, N., et al. Nature biotechnology 28, 839-847 (2010); Li, L., et al. Molecular therapy : the journal of the American Society of Gene Therapy 21, 1259-1269 (2013)]. In a recent phase I clinical trial, CD4+ T cells from patients with HIV were removed, edited with ZFNs designed to knockout the CCR5 gene,

and autologously transplanted back into patients [Tebas, P., et al. The New England journal of medicine 370, 901-910 (2014)].

[0887] In another study (Mandal et al., Cell Stem Cell, Volume 15, Issue 5, p643–652, 6 November 2014), CRISPR-Cas9 has targeted two clinically relevant genes, B2M and CCR5, in human CD4+ T cells and CD34+ hematopoietic stem and progenitor cells (HSPCs). Use of single RNA guides led to highly efficient mutagenesis in HSPCs but not in T cells. A dual guide approach improved gene deletion efficacy in both cell types. HSPCs that had undergone genome editing with CRISPR-Cas9 retained multilineage potential. Predicted on- and off-target mutations were examined via target capture sequencing in HSPCs and low levels of off-target mutagenesis were observed at only one site. These results demonstrate that CRISPR-Cas9 can efficiently ablate genes in HSPCs with minimal off-target mutagenesis, which have broad applicability for hematopoietic cell-based therapy.

[0888] Wang et al. (PLoS One. 2014 Dec 26;9(12):e115987. doi: 10.1371/journal.pone.0115987) silenced CCR5 via CRISPR associated protein 9 (Cas9) and single guided RNAs (guide RNAs) with lentiviral vectors expressing Cas9 and CCR5 guide RNAs. Wang et al. showed that a single round transduction of lentiviral vectors expressing Cas9 and CCR5 guide RNAs into HIV-1 susceptible human CD4+ cells yields high frequencies of CCR5 gene disruption. CCR5 gene-disrupted cells are not only resistant to R5-tropic HIV-1, including transmitted/founder (T/F) HIV-1 isolates, but also have selective advantage over CCR5 gene-undisrupted cells during R5-tropic HIV-1 infection. Genome mutations at potential off-target sites that are highly homologous to these CCR5 guide RNAs in stably transduced cells even at 84 days post transduction were not detected by a T7 endonuclease I assay.

[0889] Fine et al. (Sci Rep. 2015 Jul 1;5:10777. doi: 10.1038/srep10777) identified a two-cassette system expressing pieces of the S. pyogenes Cas9 (SpCas9) protein which splice together in cellula to form a functional protein capable of site-specific DNA cleavage. With specific CRISPR guide strands, Fine et al. demonstrated the efficacy of this system in cleaving the HBB and CCR5 genes in human HEK-293T cells as a single Cas9 and as a pair of Cas9 nickases. The trans-spliced SpCas9 (tsSpCas9) displayed ~35% of the nuclease activity compared with the wild-type SpCas9 (wtSpCas9) at standard transfection doses, but had substantially decreased activity at lower dosing levels. The greatly reduced open reading frame length of the tsSpCas9 relative to wtSpCas9 potentially allows for more complex and longer

genetic elements to be packaged into an AAV vector including tissue-specific promoters, multiplexed guide RNA expression, and effector domain fusions to SpCas9.

[0890] Li et al. (J Gen Virol. 2015 Aug;96(8):2381-93. doi: 10.1099/vir.0.000139. Epub 2015 Apr 8) demonstrated that CRISPR-Cas9 can efficiently mediate the editing of the CCR5 locus in cell lines, resulting in the knockout of CCR5 expression on the cell surface. Next-generation sequencing revealed that various mutations were introduced around the predicted cleavage site of CCR5. For each of the three most effective guide RNAs that were analyzed, no significant off-target effects were detected at the 15 top-scoring potential sites. By constructing chimeric Ad5F35 adenoviruses carrying CRISPR-Cas9 components, Li et al. efficiently transduced primary CD4+ T-lymphocytes and disrupted CCR5 expression, and the positively transduced cells were conferred with HIV-1 resistance.

[0891] One of skill in the art may utilize the above studies of, for example, Holt, N., et al. Nature biotechnology 28, 839-847 (2010), Li, L., et al. Molecular therapy : the journal of the American Society of Gene Therapy 21, 1259-1269 (2013), Mandal et al., Cell Stem Cell, Volume 15, Issue 5, p643–652, 6 November 2014, Wang et al. (PLoS One. 2014 Dec 26;9(12):e115987. doi: 10.1371/journal.pone.0115987), Fine et al. (Sci Rep. 2015 Jul 1;5:10777. doi: 10.1038/srep10777) and Li et al. (J Gen Virol. 2015 Aug;96(8):2381-93. doi: 10.1099/vir.0.000139. Epub 2015 Apr 8) for targeting CCR5 with the CRISPR Cas system of the present invention.

HBV

[0892] The present invention may also be applied to treat hepatitis B virus (HBV). However, the CRISPR Cas system must be adapted to avoid the shortcomings of RNAi, such as the risk of oversaturating endogenous small RNA pathways, by for example, optimizing dose and sequence (see, e.g., Grimm et al., Nature vol. 441, 26 May 2006). For example, low doses, such as about 1-10 x 1014 particles per human are contemplated. In another embodiment, the CRISPR Cas system directed against HBV may be administered in liposomes, such as a stable nucleic-acid-lipid particle (SNALP) (see, e.g., Morrissey et al., Nature Biotechnology, Vol. 23, No. 8, August 2005). Daily intravenous injections of about 1, 3 or 5 mg/kg/day of CRISPR Cas targeted to HBV RNA in a SNALP are contemplated. The daily treatment may be over about three days and then weekly for about five weeks. In another embodiment, the system of Chen et al. (Gene Therapy (2007) 14, 11–19) may be used/and or adapted for the CRISPR Cas system of the present invention. Chen et al. use a double-stranded

adenoassociated virus 8-pseudotyped vector (dsAAV2/8) to deliver shRNA. A single administration of dsAAV2/8 vector (1 x 1012 vector genomes per mouse), carrying HBV-specific shRNA, effectively suppressed the steady level of HBV protein, mRNA and replicative DNA in liver of HBV transgenic mice, leading to up to 2–3 log10 decrease in HBV load in the circulation. Significant HBV suppression sustained for at least 120 days after vector administration. The therapeutic effect of shRNA was target sequence dependent and did not involve activation of interferon. For the present invention, a CRISPR Cas system directed to HBV may be cloned into an AAV vector, such as a dsAAV2/8 vector and administered to a human, for example, at a dosage of about 1 x 1015 vector genomes to about 1 x 1016 vector genomes per human. In another embodiment, the method of Wooddell et al. (Molecular Therapy vol. 21 no. 5, 973–985 May 2013) may be used/and or adapted to the CRISPR Cas system of the present invention. Wooddell et al. show that simple coinjection of a hepatocyte-targeted, N-acetylgalactosamine-conjugated melittin-like peptide (NAG-MLP) with a liver-tropic cholesterol-conjugated siRNA (chol-siRNA) targeting coagulation factor VII (F7) results in efficient F7 knockdown in mice and nonhuman primates without changes in clinical chemistry or induction of cytokines. Using transient and transgenic mouse models of HBV infection, Wooddell et al. show that a single coinjection of NAG-MLP with potent chol-siRNAs targeting conserved HBV sequences resulted in multilog repression of viral RNA, proteins, and viral DNA with long duration of effect. Intravenous coinjections, for example, of about 6 mg/kg of NAG-MLP and 6 mg/kg of HBV specific CRISPR Cas may be envisioned for the present invention. In the alternative, about 3 mg/kg of NAG-MLP and 3 mg/kg of HBV specific CRISPR Cas may be delivered on day one, followed by administration of about about 2-3 mg/kg of NAG-MLP and 2-3 mg/kg of HBV specific CRISPR Cas two weeks later.

[0893] In some embodiments, the target sequence is an HBV sequence. In some embodiments, the target sequences is comprised in an episomal viral nucleic acid molecule which is not integrated into the genome of the organism to thereby manipulate the episomal viral nucleic acid molecule. In some embodiments, the episomal nucleic acid molecule is a double-stranded DNA polynucleotide molecule or is a covalently closed circular DNA (cccDNA). In some embodiments, the CRISPR complex is capable of reducing the amount of episomal viral nucleic acid molecule in a cell of the organism compared to the amount of episomal viral nucleic acid molecule in a cell of the organism in the absence of providing the complex, or is capable of manipulating the episomal viral nucleic acid molecule to promote

degradation of the episomal nucleic acid molecule. In some embodiments, the target HBV sequence is integrated into the genome of the organism. In some embodiments, when formed within the cell, the CRISPR complex is capable of manipulating the integrated nucleic acid to promote excision of all or part of the target HBV nucleic acid from the genome of the organism. In some embodiments, said at least one target HBV nucleic acid is comprised in a double-stranded DNA polynucleotide cccDNA molecule and/or viral DNA integrated into the genome of the organism and wherein the CRISPR complex manipulates at least one target HBV nucleic acid to cleave viral cccDNA and/or integrated viral DNA. In some embodiments, said cleavage comprises one or more double-strand break(s) introduced into the viral cccDNA and/or integrated viral DNA, optionally at least two double-strand break(s). In some embodiments, said cleavage is via one or more single-strand break(s) introduced into the viral cccDNA and/or integrated viral DNA, optionally at least two single-strand break(s). In some embodiments, said one or more double-strand break(s) or said one or more single-strand break(s) leads to the formation of one or more insertion or deletion mutations (INDELs) in the viral cccDNA sequences and/or integrated viral DNA sequences.

[0894] Lin et al. (Mol Ther Nucleic Acids. 2014 Aug 19;3:e186. doi: 10.1038/mtna.2014.38) designed eight gRNAs against HBV of genotype A. With the HBV-specific gRNAs, the CRISPR-Cas9 system significantly reduced the production of HBV core and surface proteins in Huh-7 cells transfected with an HBV-expression vector. Among eight screened gRNAs, two effective ones were identified. One gRNA targeting the conserved HBV sequence acted against different genotypes. Using a hydrodynamics-HBV persistence mouse model, Lin et al. further demonstrated that this system could cleave the intrahepatic HBV genome-containing plasmid and facilitate its clearance *in vivo*, resulting in reduction of serum surface antigen levels. These data suggest that the CRISPR-Cas9 system could disrupt the HBV-expressing templates both *in vitro* and *in vivo*, indicating its potential in eradicating persistent HBV infection.

[0895] Dong et al. (Antiviral Res. 2015 Jun;118:110-7. doi: 10.1016/j.antiviral.2015.03.015. Epub 2015 Apr 3) used the CRISPR-Cas9 system to target the HBV genome and efficiently inhibit HBV infection. Dong et al. synthesized four single-guide RNAs (guide RNAs) targeting the conserved regions of HBV. The expression of these guide RNAS with Cas9 reduced the viral production in Huh7 cells as well as in HBV-replication cell HepG2.2.15. Dong et al. further demonstrated that CRISPR-Cas9 direct cleavage and cleavage-

mediated mutagenesis occurred in HBV cccDNA of transfected cells. In the mouse model carrying HBV cccDNA, injection of guide RNA-Cas9 plasmids via rapid tail vein resulted in the low level of cccDNA and HBV protein.

[0896] Liu et al. (J Gen Virol. 2015 Aug;96(8):2252-61. doi: 10.1099/vir.0.000159. Epub 2015 Apr 22) designed eight guide RNAs (gRNAs) that targeted the conserved regions of different HBV genotypes, which could significantly inhibit HBV replication both *in vitro* and *in vivo* to investigate the possibility of using the CRISPR-Cas9 system to disrupt the HBV DNA templates. The HBV-specific gRNA/Type VII effector system could inhibit the replication of HBV of different genotypes in cells, and the viral DNA was significantly reduced by a single gRNA/Type VII effector system and cleared by a combination of different gRNA/Type VII effector systems.

[0897] Wang et al. (World J Gastroenterol. 2015 Aug 28;21(32):9554-65. doi: 10.3748/wjg.v21.i32.9554) designed 15 gRNAs against HBV of genotypes A-D. Eleven combinations of two above gRNAs (dual-gRNAs) covering the regulatory region of HBV were chosen. The efficiency of each gRNA and 11 dual-gRNAs on the suppression of HBV (genotypes A-D) replication was examined by the measurement of HBV surface antigen (HBsAg) or e antigen (HBeAg) in the culture supernatant. The destruction of HBV-expressing vector was examined in HuH7 cells co-transfected with dual-gRNAs and HBV-expressing vector using polymerase chain reaction (PCR) and sequencing method, and the destruction of cccDNA was examined in HepAD38 cells using KCl precipitation, plasmid-safe ATP-dependent DNase (PSAD) digestion, rolling circle amplification and quantitative PCR combined method. The cytotoxicity of these gRNAs was assessed by a mitochondrial tetrazolium assay. All of gRNAs could significantly reduce HBsAg or HBeAg production in the culture supernatant, which was dependent on the region in which gRNA against. All of dual gRNAs could efficiently suppress HBsAg and/or HBeAg production for HBV of genotypes A-D, and the efficacy of dual gRNAs in suppressing HBsAg and/or HBeAg production was significantly increased when compared to the single gRNA used alone. Furthermore, by PCR direct sequencing it has been confirmed that these dual gRNAs could specifically destroy HBV expressing template by removing the fragment between the cleavage sites of the two used gRNAs. Most importantly, gRNA-5 and gRNA-12 combination not only could efficiently suppress HBsAg and/or HBeAg production, but also destroy the cccDNA reservoirs in HepAD38 cells.

[0898] Karimova et al. (Sci Rep. 2015 Sep 3;5:13734. doi: 10.1038/srep13734) identified cross-genotype conserved HBV sequences in the S and X region of the HBV genome that were targeted for specific and effective cleavage by a Cas9 nickase. This approach disrupted not only episomal cccDNA and chromosomally integrated HBV target sites in reporter cell lines, but also HBV replication in chronically and de novo infected hepatoma cell lines.

[0899] One of skill in the art may utilize the above studies of, for example, Lin et al. (Mol Ther Nucleic Acids. 2014 Aug 19;3:e186. doi: 10.1038/mtna.2014.38), Dong et al. (Antiviral Res. 2015 Jun;118:110-7. doi: 10.1016/j.antiviral.2015.03.015. Epub 2015 Apr 3), Liu et al. (J Gen Virol. 2015 Aug;96(8):2252-61. doi: 10.1099/vir.0.000159. Epub 2015 Apr 22), Wang et al. (World J Gastroenterol. 2015 Aug 28;21(32):9554-65. doi: 10.3748/wjg.v21.i32.9554) and Karimova et al. (Sci Rep. 2015 Sep 3;5:13734. doi: 10.1038/srep13734) for targeting HBV with the CRISPR Cas system of the present invention.

[0900] Chronic hepatitis B virus (HBV) infection is prevalent, deadly, and seldom cured due to the persistence of viral episomal DNA (cccDNA) in infected cells. Ramanan et al. (Ramanan V, Shlomai A, Cox DB, Schwartz RE, Michailidis E, Bhatta A, Scott DA, Zhang F, Rice CM, Bhatia SN, .Sci Rep. 2015 Jun 2;5:10833. doi: 10.1038/srep10833, published online 2nd June 2015.) showed that the CRISPR/Cas9 system can specifically target and cleave conserved regions in the HBV genome, resulting in robust suppression of viral gene expression and replication. Upon sustained expression of Cas9 and appropriately chosen guide RNAs, they demonstrated cleavage of cccDNA by Cas9 and a dramatic reduction in both cccDNA and other parameters of viral gene expression and replication. Thus, they showed that directly targeting viral episomal DNA is a novel therapeutic approach to control the virus and possibly cure patients. This is also described in WO2015089465 A1, in the name of The Broad Institute et al., the contents of which are hereby incorporated by reference Targeting viral episomal DNA in HBV is preferred in some embodiments.

[0901] The present invention may also be applied to treat pathogens, e.g. bacterial, fungal and parasitic pathogens. Most research efforts have focused on developing new antibiotics, which once developed, would nevertheless be subject to the same problems of drug resistance. The invention provides novel CRISPR-based alternatives which overcome those difficulties. Furthermore, unlike existing antibiotics, CRISPR-based treatments can be made pathogen specific, inducing bacterial cell death of a target pathogen while avoiding beneficial bacteria.

[0902] The present invention may also be applied to treat hepatitis C virus (HCV). The methods of Roelvinki et al. (Molecular Therapy vol. 20 no. 9, 1737-1749 Sep 2012) may be applied to the CRISPR Cas system. For example, an AAV vector such as AAV8 may be a contemplated vector and for example a dosage of about 1.25 × 1011 to 1.25 × 1013 vector genomes per kilogram body weight (vg/kg) may be contemplated. The present invention may also be applied to treat pathogens, e.g. bacterial, fungal and parasitic pathogens. Most research efforts have focused on developing new antibiotics, which once developed, would nevertheless be subject to the same problems of drug resistance. The invention provides novel CRISPR-based alternatives which overcome those difficulties. Furthermore, unlike existing antibiotics, CRISPR-based treatments can be made pathogen specific, inducing bacterial cell death of a target pathogen while avoiding beneficial bacteria.

[0903] Jiang et al. ("RNA-guided editing of bacterial genomes using CRISPR-Cas systems," Nature Biotechnology vol. 31, p. 233-9, March 2013) used a CRISPR-Cas9 system to mutate or kill S. pneumoniae and E. coli. The work, which introduced precise mutations into the genomes, relied on dual-RNA:Cas9-directed cleavage at the targeted genomic site to kill unmutated cells and circumvented the need for selectable markers or counter-selection systems. CRISPR systems have be used to reverse antibiotic resistance and eliminate the transfer of resistance between strains. Bickard et al. showed that Cas9, reprogrammed to target virulence genes, kills virulent, but not avirulent, S. aureus. Reprogramming the nuclease to target antibiotic resistance genes destroyed staphylococcal plasmids that harbor antibiotic resistance genes and immunized against the spread of plasmid-borne resistance genes. (see, Bikard et al., "Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials," Nature Biotechnology vol. 32, 1146–1150, doi:10.1038/nbt.3043, published online 05 October 2014.) Bikard showed that CRISPR-Cas9 antimicrobials function *in vivo* to kill S. aureus in a mouse skin colonization model. Similarly, Yosef et al used a CRISPR system to target genes encoding enzymes that confer resistance to β-lactam antibiotics (see Yousef et al., "Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria," Proc. Natl. Acad. Sci. USA, vol. 112, p. 7267–7272, doi: 10.1073/pnas.1500107112 published online May 18, 2015).

[0904] CRISPR systems can be used to edit genomes of parasites that are resistant to other genetic approaches. For example, a CRISPR-Cas9 system was shown to introduce double-stranded breaks into the in the Plasmodium yoelii genome (see, Zhang et al., "Efficient Editing

of Malaria Parasite Genome Using the CRISPR/Cas9 System," mBio. vol. 5, e01414-14, Jul-Aug 2014). Ghorbal et al. ("Genome editing in the human malaria parasite Plasmodium falciparumusing the CRISPR-Cas9 system," Nature Biotechnology, vol. 32, p. 819-821, doi: 10.1038/nbt.2925, published online June 1, 2014) modified the sequences of two genes, orc1 and kelch13, which have putative roles in gene silencing and emerging resistance to artemisinin, respectively. Parasites that were altered at the appropriate sites were recovered with very high efficiency, despite there being no direct selection for the modification, indicating that neutral or even deleterious mutations can be generated using this system. CRISPR-Cas9 is also used to modify the genomes of other pathogenic parasites, including Toxoplasma gondii (see Shen et al., "Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9," mBio vol. 5:e01114-14, 2014; and Sidik et al., "Efficient Genome Engineering of Toxoplasma gondii Using CRISPR/Cas9," PLoS One vol. 9, e100450, doi: 10.1371/journal.pone.0100450, published online June 27, 2014).

[0905] Vyas et al. ("A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families," Science Advances, vol. 1, e1500248, DOI: 10.1126/sciadv.1500248, April 3, 2015) employed a CRISPR system to overcome long-standing obstacles to genetic engineering in C. albicans and efficiently mutate in a single experiment both copies of several different genes. In an organism where several mechanisms contribute to drug resistance, Vyas produced homozygous double mutants that no longer displayed the hyper-resistance to fluconazole or cycloheximide displayed by the parental clinical isolate Can90. Vyas also obtained homozygous loss-of-function mutations in essential genes of C. albicans by creating conditional alleles. Null alleles of DCR1, which is required for ribosomal RNA processing, are lethal at low temperature but viable at high temperature. Vyas used a repair template that introduced a nonsense mutation and isolated dcr1/dcr1 mutants that failed to grow at 16°C.

Treating Diseases with Genetic or Epigenetic Aspects

[0906] The CRISPR-Cas systems of the present invention can be used to correct genetic mutations that were previously attempted with limited success using TALEN and ZFN and have been identified as potential targets for Cas9 systems, including as in published applications of Editas Medicine describing methods to use Cas9 systems to target loci to therapeutically address diseases with gene therapy, including, International Patent Publication No. WO 2015/048577 CRISPR-RELATED METHODS AND COMPOSITIONS of

Gluckmann et al.; WO 2015/070083 CRISPR-RELATED METHODS AND COMPOSITIONS WITH GOVERNING gRNAS of Glucksmann et al.; In some embodiments, the treatment, prophylaxis or diagnosis of Primary Open Angle Glaucoma (POAG) is provided. The target is preferably the MYOC gene. This is described in WO2015153780, the disclosure of which is hereby incorporated by reference.

[0907] Mention is made of WO2015/134812 CRISPR/CAS-RELATED METHODS AND COMPOSITIONS FOR TREATING USHER SYNDROME AND RETINITIS PIGMENTOSA of Maeder et al. Through the teachings herein the invention comprehends methods and materials of these documents applied in conjunction with the teachings herein. In an aspect of ocular and auditory gene therapy, methods and compositions for treating Usher Syndrome and Retinis-Pigmentosa may be adapted to the CRISPR-Cas system of the present invention (see, e.g., International Patent Publication No. WO 2015/134812). In an embodiment, WO 2015/134812 involves a treatment or delaying the onset or progression of Usher Syndrome type IIA (USH2A, USH11A) and retinitis pigmentosa 39 (RP39) by gene editing, e.g., using CRISPR-Cas9 mediated methods to correct the guanine deletion at position 2299 in the USH2A gene (e.g., replace the deleted guanine residue at position 2299 in the USH2A gene).A similar effect can be achieved with a Type VII effector. In a related aspect, a mutation is targeted by cleaving with either one or more nuclease, one or more nickase, or a combination thereof, e.g., to induce HDR with a donor template that corrects the point mutation (e.g., the single nucleotide, e.g., guanine, deletion). The alteration or correction of the mutant USH2A gene can be mediated by any mechanism. Exemplary mechanisms that can be associated with the alteration (e.g., correction) of the mutant HSH2A gene include, but are not limited to, non-homologous end joining, microhomology-mediated end joining (MMEJ), homology-directed repair (e.g., endogenous donor template mediated), SDSA (synthesis dependent strand annealing), single-strand annealing or single strand invasion. In an embodiment, the method used for treating Usher Syndrome and Retinis-Pigmentosa can include acquiring knowledge of the mutation carried by the subject, e.g., by sequencing the appropriate portion of the USH2A gene.

[0908] Accordingly, in some embodiments, the treatment, prophylaxis or diagnosis of Retinitis Pigmentosa is provided. A number of different genes are known to be associated with or result in Retinitis Pigmentosa, such as RP1, RP2 and so forth. These genes are targeted in

some embodiments and either knocked out or repaired through provision of suitable a template. In some embodiments, delivery is to the eye by injection.

[0909] One or more Retinitis Pigmentosa genes can, in some embodiments, be selected from: RP1 (Retinitis pigmentosa-1), RP2 (Retinitis pigmentosa-2), RPGR (Retinitis pigmentosa-3), PRPH2 (Retinitis pigmentosa-7), RP9 (Retinitis pigmentosa-9), IMPDH1 (Retinitis pigmentosa-10), PRPF31 (Retinitis pigmentosa-11), CRB1 (Retinitis pigmentosa-12, autosomal recessive), PRPF8 (Retinitis pigmentosa-13), TULP1 (Retinitis pigmentosa-14), CA4 (Retinitis pigmentosa-17), HPRPF3 (Retinitis pigmentosa-18), ABCA4 (Retinitis pigmentosa-19), EYS (Retinitis pigmentosa-25), CERKL (Retinitis pigmentosa-26), FSCN2 (Retinitis pigmentosa-30), TOPORS (Retinitis pigmentosa-31), SNRNP200 (Retinitis pigmentosa 33), SEMA4A (Retinitis pigmentosa-35), PRCD (Retinitis pigmentosa-36), NR2E3 (Retinitis pigmentosa-37), MERTK (Retinitis pigmentosa-38), USH2A (Retinitis pigmentosa-39), PROM1 (Retinitis pigmentosa-41), KLHL7 (Retinitis pigmentosa-42), CNGB1 (Retinitis pigmentosa-45), BEST1 (Retinitis pigmentosa-50), TTC8 (Retinitis pigmentosa 51), C2orf71 (Retinitis pigmentosa 54), ARL6 (Retinitis pigmentosa 55), ZNF513 (Retinitis pigmentosa 58), DHDDS (Retinitis pigmentosa 59), BEST1 (Retinitis pigmentosa, concentric), PRPH2 (Retinitis pigmentosa, digenic), LRAT (Retinitis pigmentosa, juvenile), SPATA7 (Retinitis pigmentosa, juvenile, autosomal recessive), CRX (Retinitis pigmentosa, late-onset dominant), and/or RPGR (Retinitis pigmentosa, X-linked, and sinorespiratory infections, with or without deafness).

[0910] In some embodiments, the Retinitis Pigmentosa gene is MERTK (Retinitis pigmentosa-38) or USH2A (Retinitis pigmentosa-39).

[0911] Mention is also made of International Patent Publication No. WO 2015/138510 and through the teachings herein the invention (using a CRISPR-Cas9 system) comprehends providing a treatment or delaying the onset or progression of Leber's Congenital Amaurosis 10 (LCA 10). LCA 10 is caused by a mutation in the CEP290 gene, e.g., a c.2991+1655, adenine to guanine mutation in the CEP290 gene which gives rise to a cryptic splice site in intron 26. This is a mutation at nucleotide 1655 of intron 26 of CEP290, e.g., an A to G mutation. CEP290 is also known as: CT87; MKS4; POC3; rd16; BBS14; JBTS5; LCAJO; NPHP6; SLSN6; and 3H11Ag (see, e.g., WO 2015/138510). In an aspect of gene therapy, the invention involves introducing one or more breaks near the site of the LCA target position (e.g., c.2991 + 1655; A to G) in at least one allele of the CEP290 gene. Altering the LCA10

target position refers to (1) break-induced introduction of an indel (also referred to herein as NHEJ-mediated introduction of an indel) in close proximity to or including a LCA10 target position (e.g., c.2991+1655A to G), or (2) break-induced deletion (also referred to herein as NHEJ-mediated deletion) of genomic sequence including the mutation at a LCA10 target position (e.g., c.2991+1655A to G). Both approaches give rise to the loss or destruction of the cryptic splice site resulting from the mutation at the LCA 10 target position. Accordingly, the use of a Type VII CRISPR system in the treatment of LCA is specifically envisaged.

[0912] Researchers are contemplating whether gene therapies could be employed to treat a wide range of diseases. The CRISPR systems of the present invention based on Type VII effector protein are envisioned for such therapeutic uses, including, but noted limited to further exemplified targeted areas and with delivery methods as below. Some examples of conditions or diseases that might be usefully treated using the present system are included in the examples of genes and references included herein and are currently associated with those conditions are also provided there. The genes and conditions exemplified are not exhaustive.

Treating Diseases of the Circulatory System

[0913] The present invention also contemplates delivering the CRISPR-Cas system, specifically the Type VII CRISPR effector protein systems described herein, to the blood or hematopoetic stem cells. The plasma exosomes of Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130) were previously described and may be utilized to deliver the CRISPR Cas system to the blood. The nucleic acid-targeting system of the present invention is also contemplated to treat hemoglobinopathies, such as thalassemias and sickle cell disease. See, e.g., International Patent Publication No. WO 2013/126794 for potential targets that may be targeted by the CRISPR Cas system of the present invention.

[0914] 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, incorporated herein by reference along with the documents it cites, as if set out in full, discuss modifying HSCs using a lentivirus that delivers a gene for β-globin or γ-globin. In contrast to using lentivirus, with the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to β-Thalassemia using a CRISPR-Cas system that targets and corrects the mutation (e.g., with a suitable HDR template that delivers a coding sequence for β-globin or γ-globin, advantageously non-sickling β-globin or γ-globin); specifically, the guide RNA can target

mutation that give rise to β-Thalassemia, and the HDR can provide coding for proper expression of β-globin or γ-globin. An guide RNA that targets the mutation-and-Cas protein 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 γ-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. In this regard mention is made of: Cavazzana, "Outcomes of Gene Therapy for β-Thalassemia Major via Transplantation of Autologous Hematopoietic Stem Cells Transduced *Ex Vivo* with a Lentiviral βA-T87Q-Globin Vector." tif2014.org/abstractFiles/Jean%20Antoine%20Ribeil_Abstract.pdf; 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 (βA-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) http://www.genome.org/cgi/doi/10.1101/gr.173427.114 (Cold Spring Harbor Laboratory Press); that is the subject of Cavazzana work involving human β-thalassaemia and the subject of the Xie work, are all incorporated herein by reference, together with all documents cited therein or associated therewith. In the instant invention, the HDR template can provide for the HSC to express an engineered β-globin gene (e.g., βA-T87Q), or β-globin as in Xie.

[0915] Xu et al. (Sci Rep. 2015 Jul 9;5:12065. doi: 10.1038/srep12065) have designed TALENs and CRISPR-Cas9 to directly target the intron2 mutation site IVS2-654 in the globin gene. Xu et al. observed different frequencies of double-strand breaks (DSBs) at IVS2-654 loci using TALENs and CRISPR-Cas9, and TALENs mediated a higher homologous gene targeting efficiency compared to CRISPR-Cas9 when combined with the piggyBac transposon donor. In addition, more obvious off-target events were observed for CRISPR-Cas9 compared to TALENs. Finally, TALENs-corrected iPSC clones were selected for erythroblast differentiation using the OP9 co-culture system and detected relatively higher transcription of HBB than the uncorrected cells.

[0916] Song et al. (Stem Cells Dev. 2015 May 1;24(9):1053-65. doi: 10.1089/scd.2014.0347. Epub 2015 Feb 5) used CRISPR/ Cas9 to correct β-Thal iPSCs; gene-corrected cells exhibit normal karyotypes and full pluripotency as human embryonic stem cells (hESCs) showed no off-targeting effects. Then, Song et al. evaluated the differentiation efficiency of the gene-corrected β-Thal iPSCs. Song et al. found that during hematopoietic differentiation, gene-corrected β-Thal iPSCs showed an increased embryoid body ratio and various hematopoietic progenitor cell percentages. More importantly, the gene-corrected β-Thal iPSC lines restored HBB expression and reduced reactive oxygen species production compared with the uncorrected group. Song et al.'s study suggested that hematopoietic differentiation efficiency of β-Thal iPSCs was greatly improved once corrected by the CRISPR-Cas9 system. Similar methods may be performed utilizing the CRISPR-Cas systems described herein, e.g. systems comprising Type VII effector proteins.

[0917] Sickle cell anemia is an autosomal recessive genetic disease in which red blood cells become sickle-shaped. It is caused by a single base substitution in the β-globin gene, which is located on the short arm of chromosome 11. As a result, valine is produced instead of glutamic acid causing the production of sickle hemoglobin (HbS). This results in the formation of a distorted shape of the erythrocytes. Due to this abnormal shape, small blood vessels can be blocked, causing serious damage to the bone, spleen and skin tissues. This may lead to episodes of pain, frequent infections, hand-foot syndrome or even multiple organ failure. The distorted erythrocytes are also more susceptible to hemolysis, which leads to serious anemia.. As in the case of β-thalassaemia, sickle cell anemia can be corrected by modifying HSCs with the CRISPR-Cas system. The system allows the specific editing of the cell's genome by cutting its DNA and then letting it repair itself. The Cas protein is inserted and directed by a RNA guide 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 CRISPR-Cas allows the correction of the mutation in the previously obtained stem cells. With the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to sickle cell anemia using a CRISPR-Cas 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 RNA can target mutation that give rise to sickle cell anemia, and the HDR can provide coding for proper expression of β-globin. An guide RNA that targets the mutation-and-Cas protein

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., βA-T87Q), or β-globin as in Xie.

[0918] Williams, "Broadening the Indications for Hematopoietic Stem Cell Genetic Therapies," Cell Stem Cell 13:263-264 (2013), incorporated herein by reference along with the documents it cites, as if set out in full, report lentivirus-mediated gene transfer into HSC/P cells from patients with the lysosomal storage disease metachromatic leukodystrophy disease (MLD), a genetic disease caused by deficiency of arylsulfatase A (ARSA), resulting in nerve demyelination; and lentivirus-mediated gene transfer into HSCs of patients with Wiskott-Aldrich syndrome (WAS) (patients with defective WAS protein, an effector of the small GTPase CDC42 that regulates cytoskeletal function in blood cell lineages and thus suffer from immune deficiency with recurrent infections, autoimmune symptoms, and thrombocytopenia with abnormally small and dysfunctional platelets leading to excessive bleeding and an increased risk of leukemia and lymphoma). In contrast to using lentivirus, with the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to MLD (deficiency of arylsulfatase A (ARSA)) using a CRISPR-Cas system that targets and corrects the mutation (deficiency of arylsulfatase A (ARSA)) (e.g., with a suitable HDR template that delivers a coding sequence for ARSA); specifically, the guide RNA can target mutation that gives rise to MLD (deficient ARSA), and the HDR can provide coding for proper expression of ARSA. An guide RNA that targets the mutation-and-Cas protein 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 ARSA; 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. In contrast to using lentivirus, with the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to WAS using a CRISPR-Cas system that targets and corrects the mutation (deficiency of WAS protein) (e.g., with a suitable HDR template that delivers a coding sequence for WAS protein); specifically, the guide RNA can target mutation that gives rise to WAS (deficient WAS protein), and the HDR can provide coding for proper expression of WAS

protein. An guide RNA that targets the mutation-and-Type VII protein 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 WAS protein; 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.

[0919] Watts, "Hematopoietic Stem Cell Expansion and Gene Therapy" Cytotherapy 13(10):1164–1171. doi:10.3109/14653249.2011.620748 (2011), incorporated herein by reference along with the documents it cites, as if set out in full, discusses hematopoietic stem cell (HSC) gene therapy, e.g., virus-mediated HSC gene therapy, as an highly attractive treatment option for many disorders including hematologic conditions, immunodeficiencies including HIV/AIDS, and other genetic disorders like lysosomal storage diseases, including SCID-X1, ADA-SCID, β-thalassemia, X-linked CGD, Wiskott-Aldrich syndrome, Fanconi anemia, adrenoleukodystrophy (ALD), and metachromatic leukodystrophy (MLD).

[0920] US Patent Publication Nos. 20110225664, 20110091441, 20100229252, 20090271881 and 20090222937 assigned to Cellectis, relates to CREI variants , wherein at least one of the two I-CreI monomers has at least two substitutions, one in each of the two functional subdomains of the LAGLIDADG (see e.g. SEQ ID NO: 26 in references) core domain situated respectively from positions 26 to 40 and 44 to 77 of I-CreI, said variant being able to cleave a DNA target sequence from the human interleukin-2 receptor gamma chain (IL2RG) gene also named common cytokine receptor gamma chain gene or gamma C gene. The target sequences identified in US Patent Publication Nos. 20110225664, 20110091441, 20100229252, 20090271881 and 20090222937 may be utilized for the nucleic acid-targeting system of the present invention.

[0921] Severe Combined Immune Deficiency (SCID) results from a defect in lymphocytes T maturation, always associated with a functional defect in lymphocytes B (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). Overall incidence is estimated to 1 in 75 000 births. Patients with untreated SCID are subject to multiple opportunist micro-organism infections, and do generally not live beyond one year. SCID can be treated by allogenic hematopoietic stem cell transfer, from a familial donor. Histocompatibility with the donor can vary widely. In the case of Adenosine Deaminase (ADA) deficiency, one of the SCID forms, patients can be treated by injection of recombinant Adenosine Deaminase enzyme.

[0922] Since the ADA gene has been shown to be mutated in SCID patients (Giblett et al., Lancet, 1972, 2, 1067-1069), several other genes involved in SCID have been identified (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). There are four major causes for SCID: (i) the most frequent form of SCID, SCID-X1 (X-linked SCID or X-SCID), is caused by mutation in the IL2RG gene, resulting in the absence of mature T lymphocytes and NK cells. IL2RG encodes the gamma C protein (Noguchi, et al., Cell, 1993, 73, 147-157), a common component of at least five interleukin receptor complexes. These receptors activate several targets through the JAK3 kinase (Macchi et al., Nature, 1995, 377, 65-68), which inactivation results in the same syndrome as gamma C inactivation; (ii) mutation in the ADA gene results in a defect in purine metabolism that is lethal for lymphocyte precursors, which in turn results in the quasi absence of B, T and NK cells; (iii) V(D)J recombination is an essential step in the maturation of immunoglobulins and T lymphocytes receptors (TCRs). Mutations in Recombination Activating Gene 1 and 2 (RAG1 and RAG2) and Artemis, three genes involved in this process, result in the absence of mature T and B lymphocytes; and (iv) Mutations in other genes such as CD45, involved in T cell specific signaling have also been reported, although they represent a minority of cases (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). Since when their genetic bases have been identified, the different SCID forms have become a paradigm for gene therapy approaches (Fischer et al., Immunol. Rev., 2005, 203, 98-109) for two major reasons. First, as in all blood diseases, an *ex vivo* treatment can be envisioned. Hematopoietic Stem Cells (HSCs) can be recovered from bone marrow, and keep their pluripotent properties for a few cell divisions. Therefore, they can be treated *in vitro*, and then reinjected into the patient, where they repopulate the bone marrow. Second, since the maturation of lymphocytes is impaired in SCID patients, corrected cells have a selective advantage. Therefore, a small number of corrected cells can restore a functional immune system. This hypothesis was validated several times by (i) the partial restoration of immune functions associated with the reversion of mutations in SCID patients (Hirschhorn et al., Nat. Genet., 1996, 13, 290-295; Stephan et al., N. Engl. J. Med., 1996, 335, 1563-1567; Bousso et al., Proc. Natl., Acad. Sci. USA, 2000, 97, 274-278; Wada et al., Proc. Natl. Acad. Sci. USA, 2001, 98, 8697-8702; Nishikomori et al., Blood, 2004, 103, 4565-4572), (ii) the correction of SCID-X1 deficiencies *in vitro* in hematopoietic cells (Candotti et al., Blood, 1996, 87, 3097-3102; Cavazzana-Calvo et al., Blood, 1996, Blood, 88, 3901-3909; Taylor et al., Blood, 1996,

87, 3103-3107; Hacein-Bey et al., Blood, 1998, 92, 4090-4097), (iii) the correction of SCID-X1 (Soudais et al., Blood, 2000, 95, 3071-3077; Tsai et al., Blood, 2002, 100, 72-79), JAK-3 (Bunting et al., Nat. Med., 1998, 4, 58-64; Bunting et al., Hum. Gene Ther., 2000, 11, 2353-2364) and RAG2 (Yates et al., Blood, 2002, 100, 3942-3949) deficiencies *in vivo* in animal models and (iv) by the result of gene therapy clinical trials (Cavazzana-Calvo et al., Science, 2000, 288, 669-672; Aiuti et al., Nat. Med., 2002; 8, 423-425; Gaspar et al., Lancet, 2004, 364, 2181-2187).

[0923] US Patent Publication No. 20110182867 assigned to the Children's Medical Center Corporation and the President and Fellows of Harvard College relates to methods and uses of modulating fetal hemoglobin expression (HbF) in a hematopoietic progenitor cells via inhibitors of BCL11A expression or activity, such as RNAi and antibodies. The targets disclosed in US Patent Publication No. 20110182867, such as BCL11A, may be targeted by the CRISPR Cas system of the present invention for modulating fetal hemoglobin expression. See also Bauer et al. (Science 11 October 2013: Vol. 342 no. 6155 pp. 253-257) and Xu et al. (Science 18 November 2011: Vol. 334 no. 6058 pp. 993-996) for additional BCL11A targets.

[0924] With the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to a genetic hematologic disorder, e.g., β-Thalassemia, Hemophilia, or a genetic lysosomal storage disease.

[0925] HSC—Delivery to and Editing of Hematopoetic Stem Cells; and Particular Conditions.

[0926] The term "Hematopoetic Stem Cell" or "HSC" is meant to include 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 hematopoeitic 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: CD34lo/-, SCA-1+, Thy1.1+/lo, CD38+, C-kit+, lin-, and Human HSC markers: CD34+, CD59+, Thy1/CD90+, CD38lo/-, 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.

[0927] The CRISPR-Cas system may be engineered to target genetic locus or loci in HSCs. Cas protein, advantageously codon-optimized for a eukaryotic cell and especially a mammalian cell, e.g., a human cell, for instance, HSC, and sgRNA targeting a locus or loci in HSC, e.g., the gene EMX1, may be prepared. These may be delivered via particles. The particles may be formed by the Cas protein and the gRNA being admixed. The gRNA and Cas protein mixture may for example be admixed with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol, whereby particles containing the gRNA and Cas protein may be formed. The invention comprehends so making particles and particles from such a method as well as uses thereof.

[0928] More generally, particles may be formed using an efficient process. First, Cas Type VII effector protein and gRNA targeting the gene EMX1 or the control gene LacZ may be mixed together at a suitable, e.g.,3:1 to 1:3 or 2:1 to 1:2 or 1:1 molar ratio, at a suitable temperature, e.g., 15-30C, e.g., 20-25C, e.g., room temperature, for a suitable time, e.g., 15-45, such as 30 minutes, advantageously in sterile, nuclease free buffer, e.g., 1X PBS. Separately, particle components such as or comprising: a surfactant, e.g., cationic lipid, e.g., 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP); phospholipid, e.g., dimyristoylphosphatidylcholine (DMPC); biodegradable polymer, such as an ethylene-glycol polymer or PEG, and a lipoprotein, such as a low-density lipoprotein, e.g., cholesterol may be dissolved in an alcohol, advantageously a C1-6 alkyl alcohol, such as methanol, ethanol, isopropanol, e.g., 100% ethanol. The two solutions may be mixed together to form particles containing the Cas Type VII effector-gRNA complexes. In certain embodiments the particle can contain an HDR template. That can be a particle co-administered with gRNA+Cas protein-containing particle, or i.e., in addition to contacting an HSC with an gRNA+Cas protein-containing particle, the HSC is contacted with a particle containing an HDR template; or the HSC is contacted with a particle containing all of the gRNA, Cas and the HDR template. The HDR template can be administered by a separate vector, whereby in a first instance the particle

penetrates an HSC cell and the separate vector also penetrates the cell, wherein the HSC genome is modified by the gRNA+Cas and the HDR template is also present, whereby a genomic loci is modified by the HDR; for instance, this may result in correcting a mutation.

[0929] After the particles form, HSCs in 96 well plates may be transfected with 15ug Type VII effector protein per well. Three days after transfection, HSCs may be harvested, and the number of insertions and deletions (indels) at the EMX1 locus may be quantified.

[0930] This illustrates how HSCs can be modified using CRISPR-Cas targeting a genomic locus or loci of interest in the HSC. The HSCs that are to be modified can be *in vivo*, i.e., in an organism, for example a human or a non-human eukaryote, e.g., animal, such as fish, e.g., zebra fish, mammal, e.g., primate, e.g., ape, chimpanzee, macaque, rodent, e.g., mouse, rabbit, rat, canine or dog, livestock (cow / bovine, sheep / ovine, goat or pig), fowl or poultry, e.g., chicken. The HSCs that are to be modified can be *in vitro*, i.e., outside of such an organism. And, modified HSCs can be used *ex vivo*, i.e., one or more HSCs of such an organism can be obtained or isolated from the organism, optionally the HSC(s) can be expanded, the HSC(s) are modified by a composition comprising a CRISPR-Cas that targets a genetic locus or loci in the HSC, e.g., by contacting the HSC(s) with the composition, for instance, wherein the composition comprises a particle containing the CRISPR enzyme and one or more gRNA that targets the genetic locus or loci in the HSC, such as a particle obtained or obtainable from admixing an gRNA and Cas protein mixture with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol (wherein one or more gRNA targets the genetic locus or loci in the HSC), optionally expanding the resultant modified HSCs and administering to the organism the resultant modified HSCs. In some instances the isolated or obtained HSCs can be from a first organism, such as an organism from a same species as a second organism, and the second organism can be the organism to which the resultant modified HSCs are administered, e.g., the first organism can be a donor (such as a relative as in a parent or sibling) to the second organism. Modified HSCs can have genetic modifications to address or alleviate or reduce symptoms of a disease or condition state of an individual or subject or patient. Modified HSCs, e.g., in the instance of a first organism donor to a second organism, can have genetic modifications to have the HSCs have one or more proteins e.g. surface markers or proteins more like that of the second organism. Modified HSCs can have genetic modifications to simulate a disease or condition state of an individual or subject or patient and would be re-administered to a non-human

organism so as to prepare an animal model. Expansion of HSCs is within the ambit of the skilled person from this disclosure and knowledge in the art, see e.g., 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.

[0931] As indicated to improve activity, gRNA may be pre-complexed with the Cas protein, before formulating the entire complex in a particle. Formulations may be made with a different molar ratio of different components known to promote delivery of nucleic acids into cells (e.g. 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP), 1,2-ditetradecanoyl-sn-glycero-3-phosphocholine (DMPC), polyethylene glycol (PEG), and cholesterol) For example DOTAP : DMPC : PEG : Cholesterol Molar Ratios may be DOTAP 100, DMPC 0, PEG 0, Cholesterol 0; or DOTAP 90, DMPC 0, PEG 10, Cholesterol 0; or DOTAP 90, DMPC 0, PEG 5, Cholesterol 5. DOTAP 100, DMPC 0, PEG 0, Cholesterol 0. The invention accordingly comprehends admixing gRNA, Cas protein and components that form a particle; as well as particles from such admixing.

[0932] In a preferred embodiment, particles containing the Cas-gRNA complexes may be formed by mixing Cas protein and one or more gRNAs together, preferably at a 1:1 molar ratio, enzyme: guide RNA. Separately, the different components known to promote delivery of nucleic acids (e.g. DOTAP, DMPC, PEG, and cholesterol) are dissolved, preferably in ethanol. The two solutions are mixed together to form particles containing the Cas-gRNA complexes. After the particles are formed, Cas-gRNA complexes may be transfected into cells (e.g. HSCs). Bar coding may be applied. The particles, the Cas-9 and/or the gRNA may be barcoded.

[0933] The invention in an embodiment comprehends a method of preparing an gRNA-and-Cas protein containing particle comprising admixing an gRNA and Cas protein mixture with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol. An embodiment comprehends an gRNA-and-Cas protein containing particle from the method. The invention in an embodiment comprehends use of the particle in a method of modifying a genomic locus of interest, or an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest, comprising contacting a cell containing the genomic locus of interest with the particle wherein the gRNA targets the genomic locus of interest; or a method of modifying a genomic locus of interest, or an organism or a non-human organism by manipulation of a target

sequence in a genomic locus of interest, comprising contacting a cell containing the genomic locus of interest with the particle wherein the gRNA targets the genomic locus of interest. In these embodiments, the genomic locus of interest is advantageously a genomic locus in an HSC.

[0934] Considerations for Therapeutic Applications: A consideration in genome editing therapy is the choice of sequence-specific nuclease, such as a variant of a Type VII nuclease. 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. Thus far, two therapeutic editing approaches with nucleases have shown significant promise: gene disruption and gene correction. Gene disruption involves stimulation of NHEJ to create targeted indels in genetic elements, often resulting in loss of function mutations that are beneficial to patients. In contrast, gene correction uses HDR to directly reverse a disease causing mutation, restoring function while preserving physiological regulation of the corrected element. HDR may also be used to insert a therapeutic transgene into a defined 'safe harbor' locus in the genome to recover missing gene function. 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, for example in the treatment of SCID-X1, modified hematopoietic progenitor cells selectively expand relative to their unedited counterparts. SCID-X1 is a disease caused by mutations in the IL2RG gene, the function of which is required for proper development of the hematopoietic lymphocyte lineage [Leonard, W.J., et al. Immunological reviews 138, 61-86 (1994); Kaushansky, K. & Williams, W.J. Williams hematology, (McGraw-Hill Medical, New York, 2010)]. In clinical trials with patients who received viral gene therapy for SCID-X1, and a rare example of a spontaneous correction of SCID-X1 mutation, corrected hematopoietic progenitor cells may be able to overcome this developmental block and expand relative to their diseased counterparts to mediate therapy [Bousso, P., et al. Proceedings of the National Academy of Sciences of the United States of America 97, 274-278 (2000); 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)]. 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. In contrast, editing for other hematopoietic diseases, like chronic granulomatous disorder (CGD), would induce no change in fitness for edited hematopoietic progenitor cells, increasing the therapeutic modification threshold. CGD is caused by mutations in genes encoding phagocytic oxidase proteins, which are normally used by neutrophils to generate reactive oxygen species that kill pathogens [Mukherjee, S. & Thrasher, A.J. Gene 525, 174-181 (2013)]. As dysfunction of these genes does not influence hematopoietic progenitor cell fitness or development, but only the ability of a mature hematopoietic cell type to fight infections, there would be likely no preferential expansion of edited cells in this disease. Indeed, no selective advantage for gene corrected cells in CGD has been observed in gene therapy trials, leading to difficulties with long-term cell engraftment [Malech, H.L., et al. Proceedings of the National Academy of Sciences of the United States of America 94, 12133-12138 (1997); Kang, H.J., et al. Molecular therapy: the journal of the American Society of Gene Therapy 19, 2092-2101 (2011)]. As such, significantly higher levels of editing would be required to treat diseases like CGD, 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 latter class of diseases would be particularly difficult to treat with genome editing therapy.

[0935] In addition to cell fitness, the amount of gene product necessary to treat disease also influences the minimal level of therapeutic genome editing that must be achieved to reverse symptoms. Haemophilia B is one disease where a small change in gene product levels can result in significant changes in clinical outcomes. This disease is caused by mutations in the gene encoding factor IX, a protein normally secreted by the liver into the blood, where it functions as a component of the clotting cascade. Clinical severity of haemophilia B is related to the amount of factor IX activity. Whereas severe disease is associated with less than 1% of normal activity, milder forms of the diseases are associated with greater than 1% of factor IX activity [Kaushansky, K. & Williams, W.J. Williams hematology, (McGraw-Hill Medical, New York, 2010); Lofqvist, T., et al. Journal of internal medicine 241, 395-400 (1997)]. This suggests that editing therapies that can restore factor IX expression to even a small percentage of liver cells could have a large impact on clinical outcomes. A study using ZFNs to correct a

mouse model of haemophilia B shortly after birth demonstrated that 3-7% correction was sufficient to reverse disease symptoms, providing preclinical evidence for this hypothesis [Li, H., et al. Nature 475, 217-221 (2011)].

[0936] Disorders 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 given the current technology. Targeting these diseases has now resulted in successes with editing therapy at the preclinical level and a phase I clinical trial. Improvements in DSB repair pathway manipulation and nuclease delivery are needed to extend these promising results to diseases with a neutral fitness advantage for edited cells, or where larger amounts of gene product are needed for treatment. **Table 7** below shows some examples of applications of genome editing to therapeutic models, and the references of the below **Table 7** and the documents cited in those references are hereby incorporated herein by reference as if set out in full.

Table 7

Disease Type	Nuclease Platform Employed	Therapeutic Strategy	References
Hemophilia B	ZFN	HDR-mediated insertion of correct gene sequence	Li, H., et al. Nature 475, 217-221 (2011)
SCID	ZFN	HDR-mediated insertion of correct gene sequence	Genovese, P., et al. Nature 510, 235-240 (2014)
Hereditary tyrosinemia	CRISPR	HDR-mediated correction of mutation in liver	Yin, H., et al. Nature biotechnology 32, 551-553 (2014)

[0937] Addressing each of the conditions of the foregoing table, using the CRISPR-Cas system to target by either HDR-mediated correction of mutation, or HDR-mediated insertion of correct gene sequence, advantageously via a delivery system as herein, e.g., a particle delivery system, is within the ambit of the skilled person from this disclosure and the knowledge in the art. Thus, an embodiment comprehends contacting a Hemophilia B, SCID (e.g., SCID-X1, ADA-SCID) or Hereditary tyrosinemia mutation-carrying HSC with an gRNA-and-Cas protein containing particle targeting a genomic locus of interest as to Hemophilia B, SCID (e.g., SCID-X1, ADA-SCID) or Hereditary tyrosinemia (e.g., as in Li, Genovese or Yin). The particle also can contain a suitable HDR template to correct the

mutation; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. In this regard, it is mentioned that Haemophilia B is an X-linked recessive disorder caused by loss-of-function mutations in the gene encoding Factor IX, a crucial component of the clotting cascade. Recovering Factor IX activity to above 1% of its levels in severely affected individuals can transform the disease into a significantly milder form, as infusion of recombinant Factor IX into such patients prophylactically from a young age to achieve such levels largely ameliorates clinical complications. With the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to Haemophilia B using a CRISPR-Cas system that targets and corrects the mutation (X-linked recessive disorder caused by loss-of-function mutations in the gene encoding Factor IX) (e.g., with a suitable HDR template that delivers a coding sequence for Factor IX); specifically, the gRNA can target mutation that give rise to Haemophilia B, and the HDR can provide coding for proper expression of Factor IX. An gRNA that targets the mutation-and-Cas protein 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 Factor IX; 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, discussed herein.

[0938] In 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, incorporated herein by reference along with the documents it cites, as if set out in full, there is recognition that allogeneic hematopoietic stem cell transplantation (HSCT) was utilized to deliver normal lysosomal enzyme to the brain of a patient with Hurler's disease, and a discussion of HSC gene therapy to treat ALD. In two patients, peripheral CD34+cells were collected after granulocyte-colony stimulating factor (G-CSF) mobilization and transduced with an myeloproliferative sarcoma virus enhancer, negative control region deleted, dl587rev primer binding site substituted (MND)-ALD lentiviral vector. CD34+ cells from the patients were transduced with the MND-ALD vector during 16 h in the presence of cytokines at low concentrations. Transduced CD34+ cells were frozen after transduction to perform on 5% of cells various safety tests that included in particular three replication-competent lentivirus (RCL) assays. Transduction efficacy of CD34+ cells ranged from 35% to 50% with a mean number of

lentiviral integrated copy between 0.65 and 0.70. After the thawing of transduced CD34+ cells, the patients were reinfused with more than 4.106 transduced CD34+ cells/kg following full myeloablation with busulfan and cyclophos-phamide. The patient's HSCs were ablated to favor engraftment of the gene-corrected HSCs. Hematological recovery occurred between days 13 and 15 for the two patients. Nearly complete immunological recovery occurred at 12 months for the first patient, and at 9 months for the second patient. In contrast to using lentivirus, with the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to ALD using a CRISPR-Cas (Type VII) system that targets and corrects the mutation (e.g., with a suitable HDR template); specifically, the gRNA can target mutations in ABCD1, a gene located on the X chromosome that codes for ALD, a peroxisomal membrane transporter protein, and the HDR can provide coding for proper expression of the protein. An gRNA that targets the mutation-and-Cas (Type VII) protein containing particle is contacted with HSCs, e.g., CD34+ cells carrying the mutation as in Cartier. The particle also can contain a suitable HDR template to correct the mutation for expression of the peroxisomal membrane transporter protein; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. The so contacted cells optionally can be treated as in Cartier. The so contacted cells can be administered as in Cartier.

[0939] Mention is made of International Patent Publication No. WO 2015/148860, through the teachings herein the invention comprehends methods and materials of these documents applied in conjunction with the teachings herein. In an aspect of blood-related disease gene therapy, methods and compositions for treating beta thalassemia may be adapted to the CRISPR-Cas system of the present invention (see, e.g., WO 2015/148860). In an embodiment, WO 2015/148860 involves the treatment or prevention of beta thalassemia, or its symptoms, e.g., by altering the gene for B-cell CLL/lymphoma 11A (BCL11A). The BCL11A gene is also known as B-cell CLL/lymphoma 11A, BCL11A -L, BCL11A -S, BCL11AXL, CTIP 1, HBFQTL5 and ZNF. BCL11A encodes a zinc-finger protein that is involved in the regulation of globin gene expression. By altering the BCL11A gene (e.g., one or both alleles of the BCL11A gene), the levels of gamma globin can be increased. Gamma globin can replace beta globin in the hemoglobin complex and effectively carry oxygen to tissues, thereby ameliorating beta thalassemia disease phenotypes.

[0940] Mention is also made of International Patent Publication No. WO 2015/148863 and through the teachings herein the invention comprehends methods and materials of these

documents which may be adapted to the CRISPR-Cas system of the present invention. In an aspect of treating and preventing sickle cell disease, which is an inherited hematologic disease, WO 2015/148863 comprehends altering the BCL11A gene. By altering the BCL11A gene (e.g., one or both alleles of the BCL11A gene), the levels of gamma globin can be increased. Gamma globin can replace beta globin in the hemoglobin complex and effectively carry oxygen to tissues, thereby ameliorating sickle cell disease phenotypes.

[0941] In an aspect of the invention, methods and compositions which involve editing a target nucleic acid sequence, or modulating expression of a target nucleic acid sequence, and applications thereof in connection with cancer immunotherapy are comprehended by adapting the CRISPR-Cas system of the present invention. Reference is made to the application of gene therapy in International Patent Publication No. WO 2015/161276 which involves methods and compositions which can be used to affect T-cell proliferation, survival and/or function by altering one or more T-cell expressed genes, e.g., one or more of FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, TRAC and/or TRBC genes. In a related aspect, T-cell proliferation can be affected by altering one or more T -cell expressed genes, e.g., the CBLB and/or PTPN6 gene, FAS and/ or BID gene, CTLA4 and/or PDCDI and/or TRAC and/or TRBC gene.

[0942] Chimeric antigen receptor (CAR)19 T-cells exhibit anti-leukemic effects in patient malignancies. However, leukemia patients often do not have enough T-cells to collect, meaning that treatment must involve modified T cells from donors. Accordingly, there is interest in establishing a bank of donor T-cells. Qasim et al. ("First Clinical Application of Talen Engineered Universal CAR19 T Cells in B-ALL" ASH 57th Annual Meeting and Exposition, Dec. 5-8, 2015, Abstract 2046 (https://ash.confex.com/ash/2015/webprogram/Paper81653.html published online November 2015) discusses modifying CAR19 T cells to eliminate the risk of graft-versus-host disease through the disruption of T-cell receptor expression and CD52 targeting. Furthermore, CD52 cells were targeted such that they became insensitive to Alemtuzumab, and thus allowed Alemtuzumab to prevent host-mediated rejection of human leukocyte antigen (HLA) mismatched CAR19 T-cells. Investigators used third generation self-inactivating lentiviral vector encoding a 4g7 CAR19 (CD19 scFv-4-1BB-CD3ζ) linked to RQR8, then electroporated cells with two pairs of TALEN mRNA for multiplex targeting for both the T-cell receptor (TCR) alpha constant chain locus and the CD52 gene locus. Cells which were still expressing TCR following *ex vivo* expansion were depleted using CliniMacs α/β TCR depletion, yielding

a T-cell product (UCART19) with <1% TCR expression, 85% of which expressed CAR19, and 64% becoming CD52 negative. The modified CAR19 T cells were administered to treat a patient's relapsed acute lymphoblastic leukemia. The teachings provided herein provide effective methods for providing modified hematopoietic stem cells and progeny thereof, including but not limited to cells of the myeloid and lymphoid lineages of blood, including T cells, B cells, monocytes, macrophages, neutrophils, basophils, eosinophils, erythrocytes, dendritic cells, and megakaryocytes or platelets, and natural killer cells and their precursors and progenitors. Such cells can be modified by knocking out, knocking in, or otherwise modulating targets, for example to remove or modulate CD52 as described above, and other targets, such as, without limitation, CXCR4, and PD-1. Thus compositions, cells, and method of the invention can be used to modulate immune responses and to treat, without limitation, malignancies, viral infections, and immune disorders, in conjunction with modification of administration of T cells or other cells to patients.

[0943] Mention is made of International Patent Publication No. WO 2015/148670 and through the teachings herein the invention comprehends methods and materials of this document applied in conjunction with the teachings herein. In an aspect of gene therapy, methods and compositions for editing of a target sequence related to or in connection with Human Immunodeficiency Virus (HIV) and Acquired Immunodeficiency Syndrome (AIDS) are comprehended. In a related aspect, the invention described herein comprehends prevention and treatment of HIV infection and AIDS, by introducing one or more mutations in the gene for C-C chemokine receptor type 5 (CCR5). The CCR5 gene is also known as CKR5, CCR-5, CD195, CKR-5, CCCKR5, CMKBR5, IDDM22, and CC-CKR-5. In a further aspect, the invention described herein comprehends provide for prevention or reduction of HIV infection and/or prevention or reduction of the ability for HIV to enter host cells, e.g., in subjects who are already infected. Exemplary host cells for HIV include, but are not limited to, CD4 cells, T cells, gut associated lymphatic tissue (GALT), macrophages, dendritic cells, myeloid precursor cell, and microglia. Viral entry into the host cells requires interaction of the viral glycoproteins gp41 and gp120 with both the CD4 receptor and a co-receptor, e.g., CCR5. If a co-receptor, e.g., CCR5, is not present on the surface of the host cells, the virus cannot bind and enter the host cells. The progress of the disease is thus impeded. By knocking out or knocking down CCR5 in the host cells, e.g., by introducing a protective mutation (such as a CCR5 delta 32 mutation), entry of the HIV virus into the host cells is prevented.

[0944] X-linked Chronic granulomatous disease (CGD) is a hereditary disorder of host defense due to absent or decreased activity of phagocyte NADPH oxidase. Using a CRISPR-Cas system that targets and corrects the mutation (absent or decreased activity of phagocyte NADPH oxidase) (e.g., with a suitable HDR template that delivers a coding sequence for phagocyte NADPH oxidase), specifically, the gRNA can target mutation that gives rise to CGD (deficient phagocyte NADPH oxidase), and the HDR can provide coding for proper expression of phagocyte NADPH oxidase. An gRNA that targets the mutation-and-Cas protein 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 phagocyte NADPH oxidase; 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.

[0945] Fanconi anemia: Mutations in at least 15 genes (FANCA, FANCB, FANCC, FANCD1/BRCA2, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCJ/BACH1/BRIP1, FANCL/PHF9/POG, FANCM, FANCN/PALB2, FANCO/Rad51C, and FANCP/SLX4/BTBD12) can cause Fanconi anemia. Proteins produced from these genes are involved in a cell process known as the FA pathway. The FA pathway is turned on (activated) when the process of making new copies of DNA, called DNA replication, is blocked due to DNA damage. The FA pathway sends certain proteins to the area of damage, which trigger DNA repair so DNA replication can continue. The FA pathway is particularly responsive to a certain type of DNA damage known as interstrand cross-links (ICLs). ICLs occur when two DNA building blocks (nucleotides) on opposite strands of DNA are abnormally attached or linked together, which stops the process of DNA replication. ICLs can be caused by a buildup of toxic substances produced in the body or by treatment with certain cancer therapy drugs. Eight proteins associated with Fanconi anemia group together to form a complex known as the FA core complex. The FA core complex activates two proteins, called FANCD2 and FANCI. The activation of these two proteins brings DNA repair proteins to the area of the ICL so the cross-link can be removed and DNA replication can continue. the FA core complex. More in particular, the FA core complex is a nuclear multiprotein complex consisting of FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, and FANCM, functions as an E3 ubiquitin ligase and mediates the activation of the ID complex, which is a heterodimer composed of FANCD2 and FANCI. Once monoubiquitinated, it interacts with classical tumor

suppressors downstream of the FA pathway including FANCD1/BRCA2, FANCN/PALB2, FANCJ/BRIP1, and FANCO/Rad51C and thereby contributes to DNA repair via homologous recombination (HR). Eighty to 90 percent of FA cases are due to mutations in one of three genes, FANCA, FANCC, and FANCG. These genes provide instructions for producing components of the FA core complex. Mutations in such genes associated with the FA core complex will cause the complex to be nonfunctional and disrupt the entire FA pathway. As a result, DNA damage is not repaired efficiently and ICLs build up over time. Geiselhart, "Review Article, Disrupted Signaling through the Fanconi Anemia Pathway Leads to Dysfunctional Hematopoietic Stem Cell Biology: Underlying Mechanisms and Potential Therapeutic Strategies," Anemia Volume 2012 (2012), Article ID 265790, http://dx.doi.org/10.1155/2012/265790 discussed FA and an animal experiment involving intrafemoral injection of a lentivirus encoding the FANCC gene resulting in correction of HSCs *in vivo*. Using a CRISPR-Cas (Type VII) system that targets and one or more of the mutations associated with FA, for instance a CRISPR-Cas (Type VII) system having gRNA(s) and HDR template(s) that respectively targets one or more of the mutations of FANCA, FANCC, or FANCG that give rise to FA and provide corrective expression of one or more of FANCA, FANCC or FANCG; e.g., the gRNA can target a mutation as to FANCC, and the HDR can provide coding for proper expression of FANCC. An gRNA that targets the mutation(s) (e.g., one or more involved in FA, such as mutation(s) as to any one or more of FANCA, FANCC or FANCG)-and-Cas (Type VII) protein containing particle is contacted with HSCs carrying the mutation(s). The particle also can contain a suitable HDR template(s) to correct the mutation for proper expression of one or more of the proteins involved in FA, such as any one or more of FANCA, FANCC or FANCG; 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.

[0946] The particle in the herein discussion (e.g., as to containing gRNA(s) and Cas, optionally HDR template(s), or HDR template(s); for instance as to Hemophilia B, SCID, SCID-X1, ADA-SCID, Hereditary tyrosinemia, β-thalassemia, X-linked CGD, Wiskott-Aldrich syndrome, Fanconi anemia, adrenoleukodystrophy (ALD), metachromatic leukodystrophy (MLD), HIV/AIDS, Immunodeficiency disorder, Hematologic condition, or genetic lysosomal storage disease) is advantageously obtained or obtainable from admixing an gRNA(s) and Cas protein mixture (optionally containing HDR template(s) or such mixture

only containing HDR template(s) when separate particles as to template(s) is desired) with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol (wherein one or more gRNA targets the genetic locus or loci in the HSC).

[0947] Indeed, the invention is especially suited for treating hematopoietic genetic disorders with genome editing and immunodeficiency disorders, such as genetic immunodeficiency disorders, especially through using the particle technology herein-discussed. Genetic immunodeficiencies are diseases where genome editing interventions of the instant invention can be successful. One reason is because hematopoietic cells, of which immune cells are a subset, are therapeutically accessible. They can be removed from the body and transplanted autologously or allogenically. Further, certain genetic immunodeficiencies, e.g., severe combined immunodeficiency (SCID), create a proliferative disadvantage for immune cells. Correction of genetic lesions causing SCID by rare, spontaneous 'reverse' mutations indicates that correcting even one lymphocyte progenitor may be sufficient to recover immune function in patients.../../../Users/t_kowalski/AppData/Local/Microsoft/Windows/Temporary Internet Files/Content.Outlook/GA8VY8LK/Treating SCID for Ellen.docx - _ENREF_1 See Bousso, P., et al. Diversity, functionality, and stability of the T cell repertoire derived *in vivo* from a single human T cell precursor. Proceedings of the National Academy of Sciences of the United States of America 97, 274-278 (2000). The selective advantage for edited cells allows for even low levels of editing to result in a therapeutic effect. This effect of the instant invention can be seen in SCID, Wiskott-Aldrich Syndrome, and the other conditions mentioned herein, including other genetic hematopoietic disorders such as alpha- and beta- thalassemia, where hemoglobin deficiencies negatively affect the fitness of erythroid progenitors.

[0948] The activity of NHEJ and HDR DSB repair varies significantly 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)].

[0949] 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 requires the concurrent delivery of nucleases and repair templates. In practice, these constraints have so far led to low levels of HDR in therapeutically relevant cell types. Clinical translation has therefore largely focused on NHEJ strategies to treat disease, although proof-of-concept preclinical HDR treatments have now been described for mouse models of haemophilia B and hereditary tyrosinemia [Li, H., et al. Nature 475, 217-221 (2011); Yin, H., et al. Nature biotechnology 32, 551-553 (2014)].

[0950] Any given genome editing application may comprise combinations of proteins, small RNA molecules, and/or repair templates, making delivery of these multiple parts substantially more challenging than small molecule therapeutics. Two main strategies for delivery of genome editing tools have been developed: *ex vivo* and *in vivo*. In *ex vivo* treatments, diseased cells are removed from the body, edited and then transplanted back into the patient. *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.

[0951] There may be drawbacks with *ex vivo* approaches that limit application to a small number of diseases. For instance, target cells must be capable of surviving manipulation outside the body. For many tissues, like the brain, culturing cells outside the body is a major challenge, because cells either fail to survive, or lose properties necessary for their function *in vivo*. Thus, in view of this disclosure and the knowledge in the art, *ex vivo* therapy as to tissues with adult stem cell populations amenable to *ex vivo* culture and manipulation, such as the

hematopoietic system, by the CRISPR-Cas (Type VII) system are enabled. [Bunn, H.F. & Aster, J. Pathophysiology of blood disorders, (McGraw-Hill, New York, 2011)]

[0952] *In vivo* genome editing involves direct delivery of editing systems to cell types in their native tissues. *In vivo* editing allows diseases in which the affected cell population is not amenable to *ex vivo* manipulation to be treated. Furthermore, delivering nucleases to cells *in situ* allows for the treatment of multiple tissue and cell types. These properties probably allow *in vivo* treatment to be applied to a wider range of diseases than *ex vivo* therapies.

[0953] To date, *in vivo* editing has largely been achieved through the use of viral vectors with defined, tissue-specific tropism. Such vectors are currently limited in terms of cargo carrying capacity and tropism, restricting this mode of therapy to organ systems where transduction with clinically useful vectors is efficient, such as the liver, muscle and eye [Kotterman, M.A. & Schaffer, D.V. Nature reviews. Genetics 15, 445-451 (2014); Nguyen, T.H. & Ferry, N. Gene therapy 11 Suppl 1, S76-84 (2004); Boye, S.E., et al. Molecular therapy : the journal of the American Society of Gene Therapy 21, 509-519 (2013)].

[0954] A potential barrier for *in vivo* delivery is the immune response that may be created in response to the large amounts of virus necessary for treatment, but this phenomenon is not unique to genome editing and is observed with other virus based gene therapies [Bessis, N., et al. Gene therapy 11 Suppl 1, S10-17 (2004)]. It is also possible that peptides from editing nucleases themselves are presented on MHC Class I molecules to stimulate an immune response, although there is little evidence to support this happening at the preclinical level. Another major difficulty with this mode of therapy is controlling the distribution and consequently the dosage of genome editing nucleases *in vivo*, leading to off-target mutation profiles that may be difficult to predict. However, in view of this disclosure and the knowledge in the art, including the use of virus- and particle-based therapies being used in the treatment of cancers, *in vivo* modification of HSCs, for instance by delivery by either particle or virus, is within the ambit of the skilled person.

[0955] *Ex Vivo* Editing Therapy: The long standing clinical expertise with the purification, culture and transplantation of hematopoietic cells has made diseases affecting the blood system such as SCID, Fanconi anemia, Wiskott-Aldrich syndrome and sickle cell anemia the focus of *ex vivo* editing therapy. Another reason to focus on hematopoietic cells is that, thanks to previous efforts to design gene therapy for blood disorders, delivery systems of relatively high efficiency already exist. With these advantages, this mode of therapy can be applied to diseases

where edited cells possess a fitness advantage, so that a small number of engrafted, edited cells can expand and treat disease. One such disease is HIV, where infection results in a fitness disadvantage to CD4+ T cells.

[0956] *Ex vivo* editing therapy has been recently extended to include gene correction strategies. The barriers to HDR *ex vivo* were overcome in a recent paper from Genovese and colleagues, who achieved gene correction of a mutated IL2RG gene in hematopoietic stem cells (HSCs) obtained from a patient suffering from SCID-X1 [Genovese, P., et al. Nature 510, 235-240 (2014)]. Genovese et. al. accomplished gene correction in HSCs using a multimodal strategy. First, HSCs were transduced using integration-deficient lentivirus containing an HDR template encoding a therapeutic cDNA for IL2RG. Following transduction, cells were electroporated with mRNA encoding ZFNs targeting a mutational hotspot in IL2RG to stimulate HDR based gene correction. To increase HDR rates, culture conditions were optimized with small molecules to encourage HSC division. With optimized culture conditions, nucleases and HDR templates, gene corrected HSCs from the SCID-X1 patient were obtained in culture at therapeutically relevant rates. HSCs from unaffected individuals that underwent the same gene correction procedure could sustain long-term hematopoiesis in mice, the gold standard for HSC function. HSCs are capable of giving rise to all hematopoietic cell types and can be autologously transplanted, making them an extremely valuable cell population for all hematopoietic genetic disorders [Weissman, I.L. & Shizuru, J.A. Blood 112, 3543-3553 (2008)]. Gene corrected HSCs could, in principle, be used to treat a wide range of genetic blood disorders making this study an exciting breakthrough for therapeutic genome editing.

[0957] *In Vivo* Editing Therapy: *In vivo* editing can be used advantageously from this disclosure and the knowledge in the art. For organ systems where delivery is efficient, there have already been a number of exciting preclinical therapeutic successes. The first example of successful *in vivo* editing therapy was demonstrated in a mouse model of haemophilia B [Li, H., et al. Nature 475, 217-221 (2011)]. As noted earlier, Haemophilia B is an X-linked recessive disorder caused by loss-of-function mutations in the gene encoding Factor IX, a crucial component of the clotting cascade. Recovering Factor IX activity to above 1% of its levels in severely affected individuals can transform the disease into a significantly milder form, as infusion of recombinant Factor IX into such patients prophylactically from a young age to achieve such levels largely ameliorates clinical complications [Lofqvist, T., et al. Journal of internal medicine 241, 395-400 (1997)]. Thus, only low levels of HDR gene correction are

necessary to change clinical outcomes for patients. In addition, Factor IX is synthesized and secreted by the liver, an organ that can be transduced efficiently by viral vectors encoding editing systems.

[0958] Using hepatotropic adeno-associated viral (AAV) serotypes encoding ZFNs and a corrective HDR template, up to 7% gene correction of a mutated, humanized Factor IX gene in the murine liver was achieved [Li, H., et al. Nature 475, 217-221 (2011)]. This resulted in improvement of clot formation kinetics, a measure of the function of the clotting cascade, demonstrating for the first time that *in vivo* editing therapy is not only feasible, but also efficacious. As discussed herein, the skilled person is positioned from the teachings herein and the knowledge in the art, e.g., Li to address Haemophilia B with a particle-containing HDR template and a CRISPR-Cas system that targets the mutation of the X-linked recessive disorder to reverse the loss-of-function mutation.

[0959] Building on this study, other groups have recently used *in vivo* genome editing of the liver with CRISPR-Cas to successfully treat a mouse model of hereditary tyrosinemia and to create mutations that provide protection against cardiovascular disease. These two distinct applications demonstrate the versatility of this approach for disorders that involve hepatic dysfunction [Yin, H., et al. Nature biotechnology 32, 551-553 (2014); Ding, Q., et al. Circulation research 115, 488-492 (2014)]. Application of *in vivo* editing to other organ systems are necessary to prove that this strategy is widely applicable. Currently, efforts to optimize both viral and non-viral vectors are underway to expand the range of disorders that can be treated with this mode of therapy [Kotterman, M.A. & Schaffer, D.V. Nature reviews. Genetics 15, 445-451 (2014); Yin, H., et al. Nature reviews. Genetics 15, 541-555 (2014)]. As discussed herein, the skilled person is positioned from the teachings herein and the knowledge in the art, e.g., Yin to address hereditary tyrosinemia with a particle-containing HDR template and a CRISPR-Cas system that targets the mutation.

[0960] Targeted deletion, therapeutic applications: Targeted deletion of genes may be preferred. Preferred are, therefore, genes involved in immunodeficiency disorder, hematologic condition, or genetic lysosomal storage disease, e.g., Hemophilia B, SCID, SCID-X1, ADA-SCID, Hereditary tyrosinemia, β-thalassemia, X-linked CGD, Wiskott-Aldrich syndrome, Fanconi anemia, adrenoleukodystrophy (ALD), metachromatic leukodystrophy (MLD), HIV/AIDS, other metabolic disorders, genes encoding mis-folded proteins involved in diseases, genes leading to loss-of-function involved in diseases; generally, mutations that can

be targeted in an HSC, using any herein-discussed delivery system, with the particle system considered advantageous.

[0961] In the present invention, the immunogenicity of the CRISPR enzyme in particular may be reduced following the approach first set out in Tangri et al with respect to erythropoietin and subsequently developed. Accordingly, directed evolution or rational design may be used to reduce the immunogenicity of the CRISPR enzyme (for instance a Type VII effector) in the host species (human or other species).

[0962] Genome editing: The Type VII CRISPR/Cas systems of the present invention can be used to correct genetic mutations that were previously attempted with limited success using TALEN and ZFN and lentiviruses, including as herein discussed; see also WO2013163628.

Treating Disease of the Brain, Central Nervous and Immune Systems

[0963] The present invention also contemplates delivering the CRISPR-Cas system to the brain or neurons. The CRISPR-Cas systems and molecules thereof described herein can be used, such as via administration to a subject, to treat and/or prevent a disease of the brain, CNS, and/or immune system or a symptom thereof in a subject. For example, RNA interference (RNAi) offers therapeutic potential for this disorder by reducing the expression of HTT, the disease-causing gene of Huntington's disease (see, e.g., McBride et al., Molecular Therapy vol. 19 no. 12 Dec. 2011, pp. 2152-2162), therefore it may be used/and or adapted to the CRISPR-Cas system. The CRISPR-Cas system may be generated using an algorithm to reduce the off-targeting potential of antisense sequences. The CRISPR-Cas sequences may target either a sequence in exon 52 of mouse, rhesus or human huntingtin and expressed in a viral vector, such as AAV. Animals, including humans, may be injected with about three microinjections per hemisphere (six injections total): the first 1 mm rostral to the anterior commissure (12 µl) and the two remaining injections (12 µl and 10 µl, respectively) spaced 3 and 6 mm caudal to the first injection with 1e12 vg/ml of AAV at a rate of about 1 µl/minute, and the needle was left in place for an additional 5 minutes to allow the injectate to diffuse from the needle tip.

[0964] DiFiglia et al. (PNAS, October 23, 2007, vol. 104, no. 43, 17204–17209) observed that single administration into the adult striatum of an siRNA targeting Htt can silence mutant Htt, attenuate neuronal pathology, and delay the abnormal behavioral phenotype observed in a rapid-onset, viral transgenic mouse model of HD. DiFiglia injected mice intrastriatally with 2 µl of Cy3-labeled cc-siRNA-Htt or unconjugated siRNA-Htt at 10 µM. A similar dosage of

CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 5-10 ml of 10 μM CRISPR Cas targeted to Htt may be injected intrastriatally.

[0965] In another example, Boudreau et al. (Molecular Therapy vol. 17 no. 6 June 2009) injects 5 μl of recombinant AAV serotype 2/1 vectors expressing htt-specific RNAi virus (at 4 x 1012 viral genomes/ml) into the straiatum. A similar dosage of CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 10-20 ml of 4 x 1012 viral genomes/ml) CRISPR Cas targeted to Htt may be injected intrastriatally.

[0966] In another example, a CRISPR Cas targeted to HTT may be administered continuously (see, e.g., Yu et al., Cell 150, 895–908, August 31, 2012). Yu et al. utilizes osmotic pumps delivering 0.25 ml/hr (Model 2004) to deliver 300 mg/day of ss-siRNA or phosphate-buffered saline (PBS) (Sigma Aldrich) for 28 days, and pumps designed to deliver 0.5 μl/hr (Model 2002) were used to deliver 75 mg/day of the positive control MOE ASO for 14 days. Pumps (Durect Corporation) were filled with ss-siRNA or MOE diluted in sterile PBS and then incubated at 37 C for 24 or 48 (Model 2004) hours prior to implantation. Mice were anesthetized with 2.5% isofluorane, and a midline incision was made at the base of the skull. Using stereotaxic guides, a cannula was implanted into the right lateral ventricle and secured with Loctite adhesive. A catheter attached to an Alzet osmotic mini pump was attached to the cannula, and the pump was placed subcutaneously in the midscapular area. The incision was closed with 5.0 nylon sutures. A similar dosage of CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 500 to 1000 g/day CRISPR Cas targeted to Htt may be administered.

[0967] In another example of continuous infusion, Stiles et al. (Experimental Neurology 233 (2012) 463–471) implanted an intraparenchymal catheter with a titanium needle tip into the right putamen. The catheter was connected to a SynchroMed® II Pump (Medtronic Neurological, Minneapolis, MN) subcutaneously implanted in the abdomen. After a 7 day infusion of phosphate buffered saline at 6 μL/day, pumps were re-filled with test article and programmed for continuous delivery for 7 days. About 2.3 to 11.52 mg/d of siRNA were infused at varying infusion rates of about 0.1 to 0.5 μL/min. A similar dosage of CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 20 to 200 mg/day CRISPR Cas targeted to Htt may be administered. In another example, the methods of US Patent Publication No. 20130253040 assigned to Sangamo may also be also be

adapted from TALES to the nucleic acid-targeting system of the present invention for treating Huntington's Disease.

[0968] In another example, the methods of US Patent Publication No. 20130253040 (WO 2013/130824) assigned to Sangamo may also be also be adapted from TALES to the CRISPR Cas system of the present invention for treating Huntington's Disease.

[0969] International Patent Publication No. WO 2015/089354 A1 in the name of The Broad Institute et al., hereby incorporated by reference, describes a targets for Huntington's Disease (HP). Possible target genes of CRISPR complex in regard to Huntington's Disease: PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; and TGM2. Accordingly, one or more of PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; and TGM2 may be selected as targets for Huntington's Disease in some embodiments of the present invention.

[0970] Other trinucleotide repeat disorders. These may include any of the following: Category I includes Huntington's disease (HD) and the spinocerebellar ataxias; Category II expansions are phenotypically diverse with heterogeneous expansions that are generally small in magnitude, but also found in the exons of genes; and Category III includes fragile X syndrome, myotonic dystrophy, two of the spinocerebellar ataxias, juvenile myoclonic epilepsy, and Friedreich's ataxia.

[0971] A further aspect of the invention relates to utilizing the CRISPR-Cas system for correcting defects in the EMP2A and EMP2B genes that have been identified to be associated with Lafora disease. Lafora disease is an autosomal recessive condition which is characterized by progressive myoclonus epilepsy which may start as epileptic seizures in adolescence. A few cases of the disease may be caused by mutations in genes yet to be identified. The disease causes seizures, muscle spasms, difficulty walking, dementia, and eventually death. There is currently no therapy that has proven effective against disease progression. Other genetic abnormalities associated with epilepsy may also be targeted by the CRISPR-Cas system and the underlying genetics is further described in Genetics of Epilepsy and Genetic Epilepsies, edited by Giuliano Avanzini, Jeffrey L. Noebels, Mariani Foundation Paediatric Neurology:20; 2009).

[0972] The methods of US Patent Publication No. 20110158957 assigned to Sangamo BioSciences, Inc. involved in inactivating T cell receptor (TCR) genes may also be modified to the CRISPR Cas system of the present invention. In another example, the methods of US Patent Publication No. 20100311124 assigned to Sangamo BioSciences, Inc. and US Patent

Publication No. 20110225664 assigned to Cellectis, which are both involved in inactivating glutamine synthetase gene expression genes may also be modified to the CRISPR Cas system of the present invention.

[0973] Delivery options for the brain include encapsulation of CRISPR enzyme and guide RNA 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 CRISPR enzyme and guide RNA. 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.

[0974] Zhang et al. (Mol Ther. 2003 Jan;7(1):11-8.)) describe how expression plasmids encoding reporters such as luciferase were encapsulated in the interior of an "artificial virus" comprised of an 85 nm pegylated immunoliposome, which was targeted to the rhesus monkey brain *in vivo* with a monoclonal antibody (MAb) to the human insulin receptor (HIR). The HIRMAb enables the liposome carrying the exogenous gene to undergo transcytosis across the blood-brain barrier and endocytosis across the neuronal plasma membrane following intravenous injection. The level of luciferase gene expression in the brain was 50-fold higher in the rhesus monkey as compared to the rat. Widespread neuronal expression of the beta-galactosidase gene in primate brain was demonstrated by both histochemistry and confocal microscopy. The authors indicate that this approach makes feasible reversible adult transgenics in 24 hours. Accordingly, the use of immunoliposome is preferred. These may be used in conjunction with antibodies to target specific tissues or cell surface proteins.

Alzheimer's Disease

[0975] The CRISPR-Cas systems and molecules thereof described herein can be used, such as via administration to a subject, to treat and/or prevent Alzheimer's disease or a symptom thereof in a subject. US Patent Publication No. 20110023153, describes use of zinc finger

nucleases to genetically modify cells, animals and proteins associated with Alzheimer's Disease. Once modified cells and animals may be further tested using known methods to study the effects of the targeted mutations on the development and/or progression of AD using measures commonly used in the study of AD – such as, without limitation, learning and memory, anxiety, depression, addiction, and sensory motor functions as well as assays that measure behavioral, functional, pathological, metabolic and biochemical function.

[0976] The present disclosure comprises editing of any chromosomal sequences that encode proteins associated with AD. The AD-related proteins are typically selected based on an experimental association of the AD-related protein to an AD disorder. For example, the production rate or circulating concentration of an AD-related protein may be elevated or depressed in a population having an AD disorder relative to a population lacking the AD 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 AD-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).

[0977] Examples of Alzheimer's disease associated proteins may include the very low density lipoprotein receptor protein (VLDLR) encoded by the VLDLR gene, the ubiquitin-like modifier activating enzyme 1 (UBA1) encoded by the UBA1 gene, or the NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) encoded by the UBA3 gene, for example.

[0978] By way of non-limiting example, proteins associated with AD include but are not limited to the proteins listed as follows: Chromosomal Sequence Encoded Protein ALAS2 Delta-aminolevulinate synthase 2 (ALAS2) ABCA1 ATP-binding cassette transporter (ABCA1) ACE Angiotensin I-converting enzyme (ACE) APOE Apolipoprotein E precursor (APOE) APP amyloid precursor protein (APP) AQP1 aquaporin 1 protein (AQP1) BIN1 Myc box-dependent-interacting protein 1 or bridging integrator 1 protein (BIN1) BDNF brain-derived neurotrophic factor (BDNF) BTNL8 Butyrophilin-like protein 8 (BTNL8) C1ORF49 chromosome 1 open reading frame 49 CDH4 Cadherin-4 CHRNB2 Neuronal acetylcholine receptor subunit beta-2 CKLFSF2 CKLF-like MARVEL transmembrane domain- containing protein 2 (CKLFSF2) CLEC4E C-type lectin domain family 4, member e (CLEC4E) CLU clusterin protein (also known as apoplipoprotein J) CR1 Erythrocyte complement receptor 1

(CR1, also known as CD35, C3b/C4b receptor and immune adherence receptor) CR1L Erythrocyte complement receptor 1 (CR1L) CSF3R granulocyte colony-stimulating factor 3 receptor (CSF3R) CST3 Cystatin C or cystatin 3 CYP2C Cytochrome P450 2C DAPK1 Death-associated protein kinase 1 (DAPK1) ESR1 Estrogen receptor 1 FCAR Fc fragment of IgA receptor (FCAR, also known as CD89) FCGR3B Fc fragment of IgG, low affinity IIIb, receptor (FCGR3B or CD16b) FFA2 Free fatty acid receptor 2 (FFA2) FGA Fibrinogen (Factor I) GAB2 GRB2-associated-binding protein 2 (GAB2) GAB2 GRB2-associated-binding protein 2 (GAB2) GALP Galanin-like peptide GAPDHS Glyceraldehyde-3-phosphate dehydrogenase, spermatogenic (GAPDHS) GMPB GMBP HP Haptoglobin (HP) HTR7 5-hydroxytryptamine (serotonin) receptor 7 (adenylate cyclase-coupled) IDE Insulin degrading enzyme IF127 IF127 IFI6 Interferon, alpha-inducible protein 6 (IFI6) IFIT2 Interferon-induced protein with tetratricopeptide repeats 2 (IFIT2) IL1RN interleukin-1 receptor antagonist (IL-1RA) IL8RA Interleukin 8 receptor, alpha (IL8RA or CD181) IL8RB Interleukin 8 receptor, beta (IL8RB) JAG1 Jagged 1 (JAG1) KCNJ15 Potassium inwardly-rectifying channel, subfamily J, member 15 (KCNJ15) LRP6 Low-density lipoprotein receptor-related protein 6 (LRP6) MAPT microtubule-associated protein tau (MAPT) MARK4 MAP/microtubule affinity-regulating kinase 4 (MARK4) MPHOSPH1 M-phase phosphoprotein 1 MTHFR 5,10-methylenetetrahydrofolate reductase MX2 Interferon-induced GTP-binding protein Mx2 NBN Nibrin, also known as NBN NCSTN Nicastrin NIACR2 Niacin receptor 2 (NIACR2, also known as GPR109B) NMNAT3 nicotinamide nucleotide adenylyltransferase 3 NTM Neurotrimin (or HNT) ORM1 Orosmucoid 1 (ORM1) or Alpha-1-acid glycoprotein 1 P2RY13 P2Y purinoceptor 13 (P2RY13) PBEF1 Nicotinamide phosphoribosyltransferase (NAmPRTase or Nampt) also known as pre-B-cell colony-enhancing factor 1 (PBEF1) or visfatin PCK1 Phosphoenolpyruvate carboxykinase PICALM phosphatidylinositol binding clathrin assembly protein (PICALM) PLAU Urokinase-type plasminogen activator (PLAU) PLXNC1 Plexin C1 (PLXNC1) PRNP Prion protein PSEN1 presenilin 1 protein (PSEN1) PSEN2 presenilin 2 protein (PSEN2) PTPRA protein tyrosine phosphatase receptor type A protein (PTPRA) RALGPS2 Ral GEF with PH domain and SH3 binding motif 2 (RALGPS2) RGSL2 regulator of G-protein signaling like 2 (RGSL2) SELENBP1 Selenium binding protein 1 (SELNBP1) SLC25A37 Mitoferrin-1 SORL1 sortilin-related receptor L(DLR class) A repeats-containing protein (SORL1) TF Transferrin TFAM Mitochondrial transcription factor A TNF Tumor necrosis factor TNFRSF10C Tumor necrosis factor receptor superfamily

member 10C (TNFRSF10C) TNFSF10 Tumor necrosis factor receptor superfamily, (TRAIL) member 10a (TNFSF10) UBA1 ubiquitin-like modifier activating enzyme 1 (UBA1) UBA3 NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) UBB ubiquitin B protein (UBB) UBQLN1 Ubiquilin-1 UCHL1 ubiquitin carboxyl-terminal esterase L1 protein (UCHL1) UCHL3 ubiquitin carboxyl-terminal hydrolase isozyme L3 protein (UCHL3) VLDLR very low density lipoprotein receptor protein (VLDLR) .

[0979] In exemplary embodiments, the proteins associated with AD whose chromosomal sequence is edited may be the very low density lipoprotein receptor protein (VLDLR) encoded by the VLDLR gene, the ubiquitin-like modifier activating enzyme 1 (UBA1) encoded by the UBA1 gene, the NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) encoded by the UBA3 gene, the aquaporin 1 protein (AQP1) encoded by the AQP1 gene, the ubiquitin carboxyl-terminal esterase L1 protein (UCHL1) encoded by the UCHL1 gene, the ubiquitin carboxyl-terminal hydrolase isozyme L3 protein (UCHL3) encoded by the UCHL3 gene, the ubiquitin B protein (UBB) encoded by the UBB gene, the microtubule-associated protein tau (MAPT) encoded by the MAPT gene, the protein tyrosine phosphatase receptor type A protein (PTPRA) encoded by the PTPRA gene, the phosphatidylinositol binding clathrin assembly protein (PICALM) encoded by the PICALM gene, the clusterin protein (also known as apoplipoprotein J) encoded by the CLU gene, the presenilin 1 protein encoded by the PSEN1 gene, the presenilin 2 protein encoded by the PSEN2 gene, the sortilin-related receptor L(DLR class) A repeats-containing protein (SORL1) protein encoded by the SORL1 gene, the amyloid precursor protein (APP) encoded by the APP gene, the Apolipoprotein E precursor (APOE) encoded by the APOE gene, or the brain-derived neurotrophic factor (BDNF) encoded by the BDNF gene. In an exemplary embodiment, the genetically modified animal is a rat, and the edited chromosomal sequence encoding the protein associated with AD is as follows: APP amyloid precursor protein (APP) NM_019288 AQP1 aquaporin 1 protein (AQP1) NM_012778 BDNF Brain-derived neurotrophic factor NM_012513 CLU clusterin protein (also known as NM_053021 apoplipoprotein J) MAPT microtubule-associated protein NM_017212 tau (MAPT) PICALM phosphatidylinositol binding NM_053554 clathrin assembly protein (PICALM) PSEN1 presenilin 1 protein (PSEN1) NM_019163 PSEN2 presenilin 2 protein (PSEN2) NM_031087 PTPRA protein tyrosine phosphatase NM_012763 receptor type A protein (PTPRA) SORL1 sortilin-related receptor L(DLR NM_053519, class) A repeats-containing XM_001065506, protein (SORL1) XM_217115 UBA1 ubiquitin-like modifier

activating NM_001014080 enzyme 1 (UBA1) UBA3 NEDD8-activating enzyme E1 NM_057205 catalytic subunit protein (UBE1C) UBB ubiquitin B protein (UBB) NM_138895 UCHL1 ubiquitin carboxyl-terminal NM_017237 esterase L1 protein (UCHL1) UCHL3 ubiquitin carboxyl-terminal NM_001110165 hydrolase isozyme L3 protein (UCHL3) VLDLR very low density lipoprotein NM_013155 receptor protein (VLDLR)

[0980] The animal or cell may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9,10, 11, 12, 13, 14, 15 or more disrupted chromosomal sequences encoding a protein associated with AD and zero, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more chromosomally integrated sequences encoding a protein associated with AD.

[0981] The edited or integrated chromosomal sequence may be modified to encode an altered protein associated with AD. A number of mutations in AD-related chromosomal sequences have been associated with AD. For instance, the V7171 (i.e. valine at position 717 is changed to isoleucine) missense mutation in APP causes familial AD. Multiple mutations in the presenilin-1 protein, such as H163R (i.e., histidine at position 163 is changed to arginine), A246E (i.e. alanine at position 246 is changed to glutamate), L286V (i.e. leucine at position 286 is changed to valine) and C410Y (i.e. cysteine at position 410 is changed to tyrosine) cause familial Alzheimer's type 3. Mutations in the presenilin-2 protein, such as N141 I (i.e. asparagine at position 141 is changed to isoleucine), M239V (i.e. methionine at position 239 is changed to valine), and D439A (i.e. aspartate at position 439 is changed to alanine) cause familial Alzheimer's type 4. Other associations of genetic variants in AD-associated genes and disease are known in the art. See, for example, Waring et al. (2008) Arch. Neurol. 65:329-334, the disclosure of which is incorporated by reference herein in its entirety.

Secretase Disorders

[0982] The CRISPR-Cas systems and molecules thereof described herein can be used, such as via administration to a subject, to treat and/or prevent a secretase disorder or a symptom thereof in a subject. US Patent Publication No. 20110023146, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with secretase-associated disorders. Secretases are essential for processing pre-proteins into their biologically active forms. Defects in various components of the secretase pathways contribute to many disorders, particularly those with hallmark amyloidogenesis or amyloid plaques, such as Alzheimer's disease (AD).

[0983] A secretase disorder and the proteins associated with these disorders are a diverse set of proteins that effect susceptibility for numerous disorders, the presence of the disorder, the severity of the disorder, or any combination thereof. The present disclosure comprises editing of any chromosomal sequences that encode proteins associated with a secretase disorder. The proteins associated with a secretase disorder are typically selected based on an experimental association of the secretase--related proteins with the development of a secretase disorder. For example, the production rate or circulating concentration of a protein associated with a secretase disorder may be elevated or depressed in a population with a secretase disorder relative to a population without a secretase 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 protein associated with a secretase disorder 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).

[0984] By way of non-limiting example, proteins associated with a secretase disorder include PSENEN (presenilin enhancer 2 homolog (C. elegans)), CTSB (cathepsin B), PSEN1 (presenilin 1), APP (amyloid beta (A4) precursor protein), APH1B (anterior pharynx defective 1 homolog B (C. elegans)), PSEN2 (presenilin 2 (Alzheimer disease 4)), BACE1 (beta-site APP-cleaving enzyme 1), ITM2B (integral membrane protein 2B), CTSD (cathepsin D), NOTCH1 (Notch homolog 1, translocation-associated (Drosophila)), TNF (tumor necrosis factor (TNF superfamily, member 2)), INS (insulin), DYT10 (dystonia 10), ADAM17 (ADAM metallopeptidase domain 17), APOE (apolipoprotein E), ACE (angiotensin I converting enzyme (peptidyl-dipeptidase A) 1), STN (statin), TP53 (tumor protein p53), IL6 (interleukin 6 (interferon, beta 2)), NGFR (nerve growth factor receptor (TNFR superfamily, member 16)), IL1B (interleukin 1, beta), ACHE (acetylcholinesterase (Yt blood group)), CTNNB1 (catenin (cadherin-associated protein), beta 1, 88kDa), IGF1 (insulin-like growth factor 1 (somatomedin C)), IFNG (interferon, gamma), NRG1 (neuregulin 1), CASP3 (caspase 3, apoptosis-related cysteine peptidase), MAPK1 (mitogen-activated protein kinase 1), CDH1 (cadherin 1, type 1, E-cadherin (epithelial)), APBB1 (amyloid beta (A4) precursor protein-binding, family B, member 1 (Fe65)), HMGCR (3-hydroxy-3-methylglutaryl-Coenzyme A reductase), CREB1 (cAMP responsive element binding protein 1), PTGS2 (prostaglandin-endoperoxide synthase

2 (prostaglandin G/H synthase and cyclooxygenase)), HES1 (hairy and enhancer of split 1, (Drosophila)), CAT (catalase), TGFB1 (transforming growth factor, beta 1), ENO2 (enolase 2 (gamma, neuronal)), ERBB4 (v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)), TRAPPC10 (trafficking protein particle complex 10), MAOB (monoamine oxidase B), NGF (nerve growth factor (beta polypeptide)), MMP12 (matrix metallopeptidase 12 (macrophage elastase)), JAG1 (jagged 1 (Alagille syndrome)), CD40LG (CD40 ligand), PPARG (peroxisome proliferator-activated receptor gamma), FGF2 (fibroblast growth factor 2 (basic)), IL3 (interleukin 3 (colony-stimulating factor, multiple)), LRP1 (low density lipoprotein receptor-related protein 1), NOTCH4 (Notch homolog 4 (Drosophila)), MAPK8 (mitogen-activated protein kinase 8), PREP (prolyl endopeptidase), NOTCH3 (Notch homolog 3 (Drosophila)), PRNP (prion protein), CTSG (cathepsin G), EGF (epidermal growth factor (beta-urogastrone)), REN (renin), CD44 (CD44 molecule (Indian blood group)), SELP (selectin P (granule membrane protein 140 kDa, antigen CD62)), GHR (growth hormone receptor), ADCYAP1 (adenylate cyclase activating polypeptide 1 (pituitary)), INSR (insulin receptor), GFAP (glial fibrillary acidic protein), MMP3 (matrix metallopeptidase 3 (stromelysin 1, progelatinase)), MAPK10 (mitogen-activated protein kinase 10), SP1 (Sp1 transcription factor), MYC (v-myc myelocytomatosis viral oncogene homolog (avian)), CTSE (cathepsin E), PPARA (peroxisome proliferator-activated receptor alpha), JUN (jun oncogene), TIMP1 (TIMP metallopeptidase inhibitor 1), IL5 (interleukin 5 (colony-stimulating factor, eosinophil)), IL1A (interleukin 1, alpha), MMP9 (matrix metallopeptidase 9 (gelatinase B, 92 kDa gelatinase, 92 kDa type IV collagenase)), HTR4 (5-hydroxytryptamine (serotonin) receptor 4), HSPG2 (heparan sulfate proteoglycan 2), KRAS (v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), CYCS (cytochrome c, somatic), SMG1 (SMG1 homolog, phosphatidylinositol 3-kinase-related kinase (C. elegans)), IL1R1 (interleukin 1 receptor, type I), PROK1 (prokineticin 1), MAPK3 (mitogen-activated protein kinase 3), NTRK1 (neurotrophic tyrosine kinase, receptor, type 1), IL13 (interleukin 13), MME (membrane metallo-endopeptidase), TKT (transketolase), CXCR2 (chemokine (C-X-C motif) receptor 2), IGF1R (insulin-like growth factor 1 receptor), RARA (retinoic acid receptor, alpha), CREBBP (CREB binding protein), PTGS1 (prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)), GALT (galactose-1-phosphate uridylyltransferase), CHRM1 (cholinergic receptor, muscarinic 1), ATXN1 (ataxin 1), PAWR (PRKC, apoptosis, WT1, regulator), NOTCH2 (Notch homolog 2 (Drosophila)), M6PR (mannose-6-phosphate receptor

(cation dependent)), CYP46A1 (cytochrome P450, family 46, subfamily A, polypeptide 1), CSNK1 D (casein kinase 1, delta), MAPK14 (mitogen-activated protein kinase 14), PRG2 (proteoglycan 2, bone marrow (natural killer cell activator, eosinophil granule major basic protein)), PRKCA (protein kinase C, alpha), L1 CAM (L1 cell adhesion molecule), CD40 (CD40 molecule, TNF receptor superfamily member 5), NR1I2 (nuclear receptor subfamily 1, group I, member 2), JAG2 (jagged 2), CTNND1 (catenin (cadherin-associated protein), delta 1), CDH2 (cadherin 2, type 1, N-cadherin (neuronal)), CMA1 (chymase 1, mast cell), SORT1 (sortilin 1), DLK1 (delta-like 1 homolog (Drosophila)), THEM4 (thioesterase superfamily member 4), JUP (junction plakoglobin), CD46 (CD46 molecule, complement regulatory protein), CCL11 (chemokine (C-C motif) ligand 11), CAV3 (caveolin 3), RNASE3 (ribonuclease, RNase A family, 3 (eosinophil cationic protein)), HSPA8 (heat shock 70kDa protein 8), CASP9 (caspase 9, apoptosis-related cysteine peptidase), CYP3A4 (cytochrome P450, family 3, subfamily A, polypeptide 4), CCR3 (chemokine (C-C motif) receptor 3), TFAP2A (transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)), SCP2 (sterol carrier protein 2), CDK4 (cyclin-dependent kinase 4), HIF1A (hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)), TCF7L2 (transcription factor 7-like 2 (T-cell specific, HMG-box)), IL1R2 (interleukin 1 receptor, type II), B3GALTL (beta 1,3-galactosyltransferase-like), MDM2 (Mdm2 p53 binding protein homolog (mouse)), RELA (v-rel reticuloendotheliosis viral oncogene homolog A (avian)), CASP7 (caspase 7, apoptosis-related cysteine peptidase), IDE (insulin-degrading enzyme), FABP4 (fatty acid binding protein 4, adipocyte), CASK (calcium/calmodulin-dependent serine protein kinase (MAGUK family)), ADCYAP1R1 (adenylate cyclase activating polypeptide 1 (pituitary) receptor type I), ATF4 (activating transcription factor 4 (tax-responsive enhancer element B67)), PDGFA (platelet-derived growth factor alpha polypeptide), C21 or f33 (chromosome 21 open reading frame 33), SCG5 (secretogranin V (7B2 protein)), RNF123 (ring finger protein 123), NFKB1 (nuclear factor of kappa light polypeptide gene enhancer in B-cells 1), ERBB2 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)), CAV1 (caveolin 1, caveolae protein, 22 kDa), MMP7 (matrix metallopeptidase 7 (matrilysin, uterine)), TGFA (transforming growth factor, alpha), RXRA (retinoid X receptor, alpha), STX1A (syntaxin 1A (brain)), PSMC4 (proteasome (prosome, macropain) 26S subunit, ATPase, 4), P2RY2 (purinergic receptor P2Y, G-protein coupled, 2), TNFRSF21 (tumor necrosis factor receptor superfamily, member 21), DLG1 (discs, large

homolog 1 (Drosophila)), NUMBL (numb homolog (Drosophila)-like), SPN (sialophorin), PLSCR1 (phospholipid scramblase 1), UBQLN2 (ubiquilin 2), UBQLN1 (ubiquilin 1), PCSK7 (proprotein convertase subtilisin/kexin type 7), SPON1 (spondin 1, extracellular matrix protein), SILV (silver homolog (mouse)), QPCT (glutaminyl-peptide cyclotransferase), HESS (hairy and enhancer of split 5 (Drosophila)), GCC1 (GRIP and coiled-coil domain containing 1), and any combination thereof.

[0985] The genetically modified animal or cell may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more disrupted chromosomal sequences encoding a protein associated with a secretase disorder and zero, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more chromosomally integrated sequences encoding a disrupted protein associated with a secretase disorder.

ALS

[0986] The CRISPR-Cas systems and molecules thereof described herein can be used, such as via administration to a subject, to treat and/or prevent ALS or a symptom thereof in a subject. US Patent Publication No. 20110023144, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with amyotrophyic lateral sclerosis (ALS) disease. ALS is characterized by the gradual steady degeneration of certain nerve cells in the brain cortex, brain stem, and spinal cord involved in voluntary movement.

[0987] Motor neuron disorders and the proteins associated with these disorders are a diverse set of proteins that effect susceptibility for developing a motor neuron disorder, the presence of the motor neuron disorder, the severity of the motor neuron disorder or any combination thereof. The present disclosure comprises editing of any chromosomal sequences that encode proteins associated with ALS disease, a specific motor neuron disorder. The proteins associated with ALS are typically selected based on an experimental association of ALS--related proteins to ALS. For example, the production rate or circulating concentration of a protein associated with ALS may be elevated or depressed in a population with ALS relative to a population without ALS. 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 proteins associated with ALS 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).

[0988] By way of non-limiting example, proteins associated with ALS include but are not limited to the following proteins: SOD1 superoxide dismutase 1, ALS3 amyotrophic lateral soluble sclerosis 3 SETX senataxin ALS5 amyotrophic lateral sclerosis 5 FUS fused in sarcoma ALS7 amyotrophic lateral sclerosis 7 ALS2 amyotrophic lateral DPP6 Dipeptidyl-peptidase 6 sclerosis 2 NEFH neurofilament, heavy PTGS1 prostaglandin- polypeptide endoperoxide synthase 1 SLC1A2 solute carrier family 1 TNFRSF10B tumor necrosis factor (glial high affinity receptor superfamily, glutamate transporter), member 10b member 2 PRPH peripherin HSP90AA1 heat shock protein 90 kDa alpha (cytosolic), class A member 1 GRIA2 glutamate receptor, IFNG interferon, gamma ionotropic, AMPA 2 S100B S100 calcium binding FGF2 fibroblast growth factor 2 protein B AOX1 aldehyde oxidase 1 CS citrate synthase TARDBP TAR DNA binding protein TXN thioredoxin RAPH1 Ras association MAP3K5 mitogen-activated protein (RaIGDS/AF-6) and kinase 5 pleckstrin homology domains 1 NBEAL1 neurobeachin-like 1 GPX1 glutathione peroxidase 1 ICA1L islet cell autoantigen RAC1 ras-related C3 botulinum 1.69 kDa-like toxin substrate 1 MAPT microtubule-associated ITPR2 inositol 1,4,5- protein tau triphosphate receptor, type 2 ALS2CR4 amyotrophic lateral GLS glutaminase sclerosis 2 (juvenile) chromosome region, candidate 4 ALS2CR8 amyotrophic lateral CNTFR ciliary neurotrophic factor sclerosis 2 (juvenile) receptor chromosome region, candidate 8 ALS2CR11 amyotrophic lateral FOLH1 folate hydrolase 1 sclerosis 2 (juvenile) chromosome region, candidate 11 FAM117B family with sequence P4HB prolyl 4-hydroxylase, similarity 117, member B beta polypeptide CNTF ciliary neurotrophic factor SQSTM1 sequestosome 1 STRADB STE20-related kinase NAIP NLR family, apoptosis adaptor beta inhibitory protein YWHAQ tyrosine 3- SLC33A1 solute carrier family 33 monooxygenase/tryptoph (acetyl-CoA transporter), an 5-monooxygenase member 1 activation protein, theta polypeptide TRAK2 trafficking protein, homolog, SAC1 kinesin binding 2 lipid phosphatase domain containing NIF3L1 NIF3 NGG1 interacting INA internexin neuronal factor 3-like 1 intermediate filament protein, alpha PARD3B par-3 partitioning COX8A cytochrome c oxidase defective 3 homolog B subunit VIIIA CDK15 cyclin-dependent kinase HECW1 HECT, C2 and WW 15 domain containing E3 ubiquitin protein ligase 1 NOS1 nitric oxide synthase 1 MET met proto-oncogene SOD2 superoxide dismutase 2, HSPB1 heat shock 27 kDa mitochondrial protein 1 NEFL neurofilament, light CTSB cathepsin B polypeptide ANG angiogenin, HSPA8 heat shock 70 kDa ribonuclease, RNase A protein 8 family, 5 VAPB VAMP (vesicle- ESR1 estrogen receptor 1 associated membrane protein)-associated protein B

and C SNCA synuclein, alpha HGF hepatocyte growth factor CAT catalase ACTB actin, beta NEFM neurofilament, medium TH tyrosine hydroxylase polypeptide BCL2 B-cell CLL/lymphoma 2 FAS Fas (TNF receptor superfamily, member 6) CASP3 caspase 3, apoptosis- CLU clusterin related cysteine peptidase SMN1 survival of motor neuron G6PD glucose-6-phosphate 1, telomeric dehydrogenase BAX BCL2-associated X HSF1 heat shock transcription protein factor 1 RNF19A ring finger protein 19A JUN jun oncogene ALS2CR12 amyotrophic lateral HSPA5 heat shock 70 kDa sclerosis 2 (juvenile) protein 5 chromosome region, candidate 12 MAPK14 mitogen-activated protein IL10 interleukin 10 kinase 14 APEX1 APEX nuclease TXNRD1 thioredoxin reductase 1 (multifunctional DNA repair enzyme) 1 NOS2 nitric oxide synthase 2, TIMP1 TIMP metallopeptidase inducible inhibitor 1 CASP9 caspase 9, apoptosis- XIAP X-linked inhibitor of related cysteine apoptosis peptidase GLG1 golgi glycoprotein 1 EPO erythropoietin VEGFA vascular endothelial ELN elastin growth factor A GDNF glial cell derived NFE2L2 nuclear factor (erythroid- neurotrophic factor derived 2)-like 2 SLC6A3 solute carrier family 6 HSPA4 heat shock 70 kDa (neurotransmitter protein 4 transporter, dopamine), member 3 APOE apolipoprotein E PSMB8 proteasome (prosome, macropain) subunit, beta type, 8 DCTN1 dynactin 1 TIMP3 TIMP metallopeptidase inhibitor 3 KIFAP3 kinesin-associated SLC1A1 solute carrier family 1 protein 3 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1 SMN2 survival of motor neuron CCNC cyclin C 2, centromeric MPP4 membrane protein, STUB1 STIP1 homology and U- palmitoylated 4 box containing protein 1 ALS2 amyloid beta (A4) PRDX6 peroxiredoxin 6 precursor protein SYP synaptophysin CABIN1 calcineurin binding protein 1 CASP1 caspase 1, apoptosis- GART phosphoribosylglycinami related cysteine de formyltransferase, peptidase phosphoribosylglycinami de synthetase, phosphoribosylaminoimi dazole synthetase CDK5 cyclin-dependent kinase 5 ATXN3 ataxin 3 RTN4 reticulon 4 C1QB complement component 1, q subcomponent, B chain VEGFC nerve growth factor HTT huntingtin receptor PARK7 Parkinson disease 7 XDH xanthine dehydrogenase GFAP glial fibrillary acidic MAP2 microtubule-associated protein protein 2 CYCS cytochrome c, somatic FCGR3B Fc fragment of IgG, low affinity IIIb, CCS copper chaperone for UBL5 ubiquitin-like 5 superoxide dismutase MMP9 matrix metallopeptidase SLC18A3 solute carrier family 18 9 ((vesicular acetylcholine), member 3 TRPM7 transient receptor HSPB2 heat shock 27 kDa potential cation channel, protein 2 subfamily M, member 7 AKT1 v-akt murine thymoma DERL1 Der1-like domain family, viral oncogene homolog 1 member 1 CCL2 chemokine (C-

-C motif) NGRN neugrin, neurite ligand 2 outgrowth associated GSR glutathione reductase TPPP3 tubulin polymerization- promoting protein family member 3 APAF1 apoptotic peptidase BTBD10 BTB (POZ) domain activating factor 1 containing 10 GLUD1 glutamate CXCR4 chemokine (C--X--C motif) dehydrogenase 1 receptor 4 SLC1A3 solute carrier family 1 FLT1 fms-related tyrosine (glial high affinity glutamate transporter), member 3 kinase 1 PON1 paraoxonase 1 AR androgen receptor LIF leukemia inhibitory factor ERBB3 v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 LGALS1 lectin, galactoside- CD44 CD44 molecule binding, soluble, 1 TP53 tumor protein p53 TLR3 toll-like receptor 3 GRIA1 glutamate receptor, GAPDH glyceraldehyde-3- ionotropic, AMPA 1 phosphate dehydrogenase GRIK1 glutamate receptor, DES desmin ionotropic, kainate 1 CHAT choline acetyltransferase FLT4 fms-related tyrosine kinase 4 CHMP2B chromatin modifying BAG1 BCL2-associated protein 2B athanogene MT3 metallothionein 3 CHRNA4 cholinergic receptor, nicotinic, alpha 4 GSS glutathione synthetase BAK1 BCL2-antagonist/killer 1 KDR kinase insert domain GSTP1 glutathione S-transferase receptor (a type III pi 1 receptor tyrosine kinase) OGG1 8-oxoguanine DNA IL6 interleukin 6 (interferon, glycosylase beta 2).

[0989] The animal or cell may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more disrupted chromosomal sequences encoding a protein associated with ALS and zero, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more chromosomally integrated sequences encoding the disrupted protein associated with ALS. Preferred proteins associated with ALS include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

Autism

[0990] The CRISPR-Cas systems and molecules thereof described herein can be used, such as via administration to a subject, to treat and/or prevent autism or a symptom thereof in a subject. US Patent Publication No. 20110023145, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with autism spectrum disorders (ASD). Autism spectrum disorders (ASDs) are a group of disorders characterized by qualitative impairment in social interaction and communication, and restricted repetitive and stereotyped patterns of behavior, interests, and activities. The three disorders, autism, Asperger syndrome (AS) and pervasive developmental disorder-not otherwise specified (PDD-NOS) are a

continuum of the same disorder with varying degrees of severity, associated intellectual functioning and medical conditions. ASDs are predominantly genetically determined disorders with a heritability of around 90%.

[0991] US Patent Publication No. 20110023145 comprises editing of any chromosomal sequences that encode proteins associated with ASD which may be applied to the CRISPR Cas system of the present invention. The proteins associated with ASD are typically selected based on an experimental association of the protein associated with ASD to an incidence or indication of an ASD. For example, the production rate or circulating concentration of a protein associated with ASD may be elevated or depressed in a population having an ASD relative to a population lacking the ASD. 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 proteins associated with ASD 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).

[0992] Non-limiting examples of disease states or disorders that may be associated with proteins associated with ASD include autism, Asperger syndrome (AS), pervasive developmental disorder-not otherwise specified (PDD-NOS), Rett's syndrome, tuberous sclerosis, phenylketonuria, Smith-Lemli-Opitz syndrome and fragile X syndrome. By way of non-limiting example, proteins associated with ASD include but are not limited to the following proteins: ATP10C aminophospholipid- MET MET receptor transporting ATPase tyrosine kinase (ATP10C) BZRAP1 MGLUR5 (GRM5) Metabotropic glutamate receptor 5 (MGLUR5) CDH10 Cadherin-10 MGLUR6 (GRM6) Metabotropic glutamate receptor 6 (MGLUR6) CDH9 Cadherin-9 NLGN1 Neuroligin-1 CNTN4 Contactin-4 NLGN2 Neuroligin-2 CNTNAP2 Contactin-associated SEMA5A Neuroligin-3 protein-like 2 (CNTNAP2) DHCR7 7-dehydrocholesterol NLGN4X Neuroligin-4 X- reductase (DHCR7) linked DOC2A Double C2-like domain- NLGN4Y Neuroligin-4 Y- containing protein alpha linked DPP6 Dipeptidyl NLGN5 Neuroligin-5 aminopeptidase-like protein 6 EN2 engrailed 2 (EN2) NRCAM Neuronal cell adhesion molecule (NRCAM) MDGA2 fragile X mental retardation NRXN1 Neurexin-1 1 (MDGA2) FMR2 (AFF2) AF4/FMR2 family member 2 OR4M2 Olfactory receptor (AFF2) 4M2 FOXP2 Forkhead box protein P2 OR4N4 Olfactory

receptor (FOXP2) 4N4 FXR1 Fragile X mental OXTR oxytocin receptor retardation, autosomal (OXTR) homolog 1 (FXR1) FXR2 Fragile X mental PAH phenylalanine retardation, autosomal hydroxylase (PAH) homolog 2 (FXR2) GABRA1 Gamma-aminobutyric acid PTEN Phosphatase and receptor subunit alpha-1 tensin homologue (GABRA1) (PTEN) GABRA5 GABAA (.gamma.-aminobutyric PTPRZ1 Receptor-type acid) receptor alpha 5 tyrosine-protein subunit (GABRA5) phosphatase zeta (PTPRZ1) GABRB1 Gamma-aminobutyric acid RELN Reelin receptor subunit beta-1 (GABRB1) GABRB3 GABAA (.gamma.-aminobutyric RPL10 60S ribosomal acid) receptor .beta.3 subunit protein L10 (GABRB3) GABRG1 Gamma-aminobutyric acid SEMA5A Semaphorin-5A receptor subunit gamma-1 (SEMA5A) (GABRG1) HIRIP3 HIRA-interacting protein 3 SEZ6L2 seizure related 6 homolog (mouse)- like 2 HOXA1 Homeobox protein Hox-A1 SHANK3 SH3 and multiple (HOXA1) ankyrin repeat domains 3 (SHANK3) IL6 Interleukin-6 SHBZRAP1 SH3 and multiple ankyrin repeat domains 3 (SHBZRAP1) LAMB1 Laminin subunit beta-1 SLC6A4 Serotonin (LAMB1) transporter (SERT) MAPK3 Mitogen-activated protein TAS2R1 Taste receptor kinase 3 type 2 member 1 TAS2R1 MAZ Myc-associated zinc finger TSC1 Tuberous sclerosis protein protein 1 MDGA2 MAM domain containing TSC2 Tuberous sclerosis glycosylphosphatidylinositol protein 2 anchor 2 (MDGA2) MECP2 Methyl CpG binding UBE3A Ubiquitin protein protein 2 (MECP2) ligase E3A (UBE3A) MECP2 methyl CpG binding WNT2 Wingless-type protein 2 (MECP2) MMTV integration site family, member 2 (WNT2)

[0993] The identity of the protein associated with ASD whose chromosomal sequence is edited can and will vary. In preferred embodiments, the proteins associated with ASD whose chromosomal sequence is edited may be the benzodiazepine receptor (peripheral) associated protein 1 (BZRAP1) encoded by the BZRAP1 gene, the AF4/FMR2 family member 2 protein (AFF2) encoded by the AFF2 gene (also termed MFR2), the fragile X mental retardation autosomal homolog 1 protein (FXR1) encoded by the FXR1 gene, the fragile X mental retardation autosomal homolog 2 protein (FXR2) encoded by the FXR2 gene, the MAM domain containing glycosylphosphatidylinositol anchor 2 protein (MDGA2) encoded by the MDGA2 gene, the methyl CpG binding protein 2 (MECP2) encoded by the MECP2 gene, the metabotropic glutamate receptor 5 (MGLUR5) encoded by the MGLUR5-1 gene (also termed GRM5), the neurexin 1 protein encoded by the NRXN1 gene, or the semaphorin-5A protein (SEMA5A) encoded by the SEMA5A gene. In an exemplary embodiment, the genetically

modified animal is a rat, and the edited chromosomal sequence encoding the protein associated with ASD is as listed below: BZRAP1 benzodiazepine receptor XM_002727789, (peripheral) associated XM_213427, protein 1 (BZRAP1) XM_002724533, XM_001081125 AFF2 (FMR2) AF4/FMR2 family member 2 XM_219832, (AFF2) XM_001054673 FXR1 Fragile X mental NM_001012179 retardation, autosomal homolog 1 (FXR1) FXR2 Fragile X mental NM_001100647 retardation, autosomal homolog 2 (FXR2) MDGA2 MAM domain containing NM_199269 glycosylphosphatidylinositol anchor 2 (MDGA2) MECP2 Methyl CpG binding NM_022673 protein 2 (MECP2) MGLUR5 Metabotropic glutamate NM_017012 (GRM5) receptor 5 (MGLUR5) NRXN1 Neurexin-1 NM_021767 SEMA5A Semaphorin-5A (SEMA5A) NM_001107659.

Trinucleotide Repeat Expansion Disorders

[0994] The CRISPR-Cas systems and molecules thereof described herein can be used, such as via administration to a subject, to treat and/or prevent a trinucleotide repeat expansion disorder. US Patent Publication No. 20110016540 describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with trinucleotide repeat expansion disorders. Trinucleotide repeat expansion disorders are complex, progressive disorders that involve developmental neurobiology and often affect cognition as well as sensori-motor functions.

[0995] Trinucleotide repeat expansion proteins are a diverse set of proteins associated with susceptibility for developing a trinucleotide repeat expansion disorder, the presence of a trinucleotide repeat expansion disorder, the severity of a trinucleotide repeat expansion disorder or any combination thereof. Trinucleotide repeat expansion disorders are divided into two categories determined by the type of repeat. The most common repeat is the triplet CAG, which, when present in the coding region of a gene, codes for the amino acid glutamine (Q). Therefore, these disorders are referred to as the polyglutamine (polyQ) disorders and comprise the following diseases: Huntington Disease (HD); Spinobulbar Muscular Atrophy (SBMA); Spinocerebellar Ataxias (SCA types 1, 2, 3, 6, 7, and 17); and Dentatorubro-Pallidoluysian Atrophy (DRPLA). The remaining trinucleotide repeat expansion disorders either do not involve the CAG triplet or the CAG triplet is not in the coding region of the gene and are, therefore, referred to as the non-polyglutamine disorders. The non-polyglutamine disorders comprise Fragile X Syndrome (FRAXA); Fragile XE Mental Retardation (FRAXE); Friedreich

Ataxia (FRDA); Myotonic Dystrophy (DM); and Spinocerebellar Ataxias (SCA types 8, and 12).

[0996] The proteins associated with trinucleotide repeat expansion disorders are typically selected based on an experimental association of the protein associated with a trinucleotide repeat expansion disorder to a trinucleotide repeat expansion disorder. For example, the production rate or circulating concentration of a protein associated with a trinucleotide repeat expansion disorder may be elevated or depressed in a population having a trinucleotide repeat expansion disorder relative to a population lacking the trinucleotide repeat expansion 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 proteins associated with trinucleotide repeat expansion disorders 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).

[0997] Non-limiting examples of proteins associated with trinucleotide repeat expansion disorders include AR (androgen receptor), FMR1 (fragile X mental retardation 1), HTT (huntingtin), DMPK (dystrophia myotonica-protein kinase), FXN (frataxin), ATXN2 (ataxin 2), ATN1 (atrophin 1), FEN1 (flap structure-specific endonuclease 1), TNRC6A (trinucleotide repeat containing 6A), PABPN1 (poly(A) binding protein, nuclear 1), JPH3 (junctophilin 3), MED15 (mediator complex subunit 15), ATXN1 (ataxin 1), ATXN3 (ataxin 3), TBP (TATA box binding protein), CACNA1A (calcium channel, voltage-dependent, P/Q type, alpha 1A subunit), ATXN80S (ATXN8 opposite strand (non-protein coding)), PPP2R2B (protein phosphatase 2, regulatory subunit B, beta), ATXN7 (ataxin 7), TNRC6B (trinucleotide repeat containing 6B), TNRC6C (trinucleotide repeat containing 6C), CELF3 (CUGBP, Elav-like family member 3), MAB21L1 (mab-21-like 1 (C. elegans)), MSH2 (mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)), TMEM185A (transmembrane protein 185A), SIX5 (SIX homeobox 5), CNPY3 (canopy 3 homolog (zebrafish)), FRAXE (fragile site, folic acid type, rare, fra(X)(q28) E), GNB2 (guanine nucleotide binding protein (G protein), beta polypeptide 2), RPL14 (ribosomal protein L14), ATXN8 (ataxin 8), INSR (insulin receptor), TTR (transthyretin), EP400 (E1A binding protein p400), GIGYF2 (GRB10 interacting GYF protein 2), OGG1 (8-oxoguanine DNA glycosylase), STC1 (stanniocalcin 1), CNDP1 (carnosine

dipeptidase 1 (metallopeptidase M20 family)), C10orf2 (chromosome 10 open reading frame 2), MAML3 mastermind-like 3 (Drosophila), DKC1 (dyskeratosis congenita 1, dyskerin), PAXIP1 (PAX interacting (with transcription-activation domain) protein 1), CASK (calcium/calmodulin-dependent serine protein kinase (MAGUK family)), MAPT (microtubule-associated protein tau), SP1 (Sp1 transcription factor), POLG (polymerase (DNA directed), gamma), AFF2 (AF4/FMR2 family, member 2), THBS1 (thrombospondin 1), TP53 (tumor protein p53), ESR1 (estrogen receptor 1), CGGBP1 (CGG triplet repeat binding protein 1), ABT1 (activator of basal transcription 1), KLK3 (kallikrein-related peptidase 3), PRNP (prion protein), JUN (jun oncogene), KCNN3 (potassium intermediate/small conductance calcium-activated channel, subfamily N, member 3), BAX (BCL2-associated X protein), FRAXA (fragile site, folic acid type, rare, fra(X)(q27.3) A (macroorchidism, mental retardation)), KBTBD10 (kelch repeat and BTB (POZ) domain containing 10), MBNL1 (muscleblind-like (Drosophila)), RAD51 (RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)), NCOA3 (nuclear receptor coactivator 3), ERDA1 (expanded repeat domain, CAG/CTG 1), TSC1 (tuberous sclerosis 1), COMP (cartilage oligomeric matrix protein), GCLC (glutamate-cysteine ligase, catalytic subunit), RRAD (Ras-related associated with diabetes), MSH3 (mutS homolog 3 (E. coli)), DRD2 (dopamine receptor D2), CD44 (CD44 molecule (Indian blood group)), CTCF (CCCTC-binding factor (zinc finger protein)), CCND1 (cyclin D1), CLSPN (claspin homolog (Xenopus laevis)), MEF2A (myocyte enhancer factor 2A), PTPRU (protein tyrosine phosphatase, receptor type, U), GAPDH (glyceraldehyde-3-phosphate dehydrogenase), TRIM22 (tripartite motif-containing 22), WT1 (Wilms tumor 1), AHR (aryl hydrocarbon receptor), GPX1 (glutathione peroxidase 1), TPMT (thiopurine S-methyltransferase), NDP (Norrie disease (pseudoglioma)), ARX (aristaless related homeobox), MUS81 (MUS81 endonuclease homolog (S. cerevisiae)), TYR (tyrosinase (oculocutaneous albinism IA)), EGR1 (early growth response 1), UNG (uracil-DNA glycosylase), NUMBL (numb homolog (Drosophila)-like), FABP2 (fatty acid binding protein 2, intestinal), EN2 (engrailed homeobox 2), CRYGC (crystallin, gamma C), SRP14 (signal recognition particle 14 kDa (homologous Alu RNA binding protein)), CRYGB (crystallin, gamma B), PDCD1 (programmed cell death 1), HOXA1 (homeobox A1), ATXN2L (ataxin 2-like), PMS2 (PMS2 postmeiotic segregation increased 2 (S. cerevisiae)), GLA (galactosidase, alpha), CBL (Cas-Br-M (murine) ecotropic retroviral transforming sequence), FTH1 (ferritin, heavy polypeptide 1), IL12RB2 (interleukin 12 receptor, beta 2), OTX2 (orthodenticle homeobox 2), HOXA5

(homeobox A5), POLG2 (polymerase (DNA directed), gamma 2, accessory subunit), DLX2 (distal-less homeobox 2), SIRPA (signal-regulatory protein alpha), OTX1 (orthodenticle homeobox 1), AHRR (aryl-hydrocarbon receptor repressor), MANF (mesencephalic astrocyte-derived neurotrophic factor), TMEM158 (transmembrane protein 158 (gene/pseudogene)), and ENSG00000078687.

[0998] Preferred proteins associated with trinucleotide repeat expansion disorders include HTT (Huntingtin), AR (androgen receptor), FXN (frataxin), Atxn3 (ataxin), Atxn1 (ataxin), Atxn2 (ataxin), Atxn7 (ataxin), Atxn10 (ataxin), DMPK (dystrophia myotonica-protein kinase), Atn1 (atrophin 1), CBP (creb binding protein), VLDLR (very low density lipoprotein receptor), and any combination thereof.

Treating Hearing Diseases

[0999] The present invention also contemplates delivering the CRISPR-Cas system to one or both ears.

[01000] Researchers are looking into whether gene therapy could be used to aid current deafness treatments - namely, cochlear implants. 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.

[01001] 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 described herein can be administered by intratympanic injection (e.g., into the middle ear), and/or injections into the outer, middle, and/or inner 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).

[01002] In another mode of administration, the pharmaceutical composition can be administered in situ, via a catheter or pump. A catheter or pump can, for example, direct a pharmaceutical composition into the cochlear luminae or the round window of the ear and/or the lumen of the colon. Exemplary drug delivery apparatus and methods suitable for

administering one or more of the compounds described herein into an ear, e.g., a human ear, are described by McKenna et al., (U.S. Patent Publication No. 2006/0030837) and Jacobsen et al., (U.S. Pat. No. 7,206,639). In some embodiments, 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 some embodiments, 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.

[01003] Alternatively or in addition, one or more of the compounds described herein can be administered in combination with a mechanical device such as a cochlear implant or a hearing aid, which is worn in the outer ear. An exemplary cochlear implant that is suitable for use with the present invention is described by Edge et al., (U.S. Publication No. 2007/0093878).

[01004] In some embodiments, the modes of administration described above may be combined in any order and can be simultaneous or interspersed.

[01005] Alternatively or in addition, the present invention may be administered according to any of the Food and Drug Administration approved methods, for example, as described in CDER Data Standards Manual, version number 004 (which is available at fda.give/cder/dsm/DRG/drg00301.htm).

[01006] 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.

[01007] 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 described 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. Publication No. 2005/0287127) and Li et al., (U.S. patent Ser. 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.

[01008] The CRISPR Cas molecules of the present invention 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 some embodiments the pharmaceutical composition is applied to the ear canal. Delivery to the ear may also be referred to as aural or otic delivery.

[01009] In some embodiments the RNA molecules of the invention are delivered in liposome or lipofectin formulations and the like and can be prepared by methods well known to those skilled in the art. Such methods are described, for example, in U.S. Pat. Nos. 5,593,972, 5,589,466, and 5,580,859, which are herein incorporated by reference.

[01010] Delivery systems aimed specifically at the enhanced and improved delivery of siRNA into mammalian cells have been developed, (see, for example, Shen et al FEBS Let. 2003, 539:111-114; Xia et al., Nat. Biotech. 2002, 20:1006-1010; Reich et al., Mol. Vision. 2003, 9: 210-216; Sorensen et al., J. Mol. Biol. 2003, 327: 761-766; Lewis et al., Nat. Gen.

2002, 32: 107-108 and Simeoni et al., NAR 2003, 31, 11: 2717-2724) and may be applied to the present invention. siRNA has recently been successfully used for inhibition of gene expression in primates (see for example. Tolentino et al., Retina 24(4):660 which may also be applied to the present invention.

[01011] Qi et al. discloses methods for efficient siRNA transfection to the inner ear through the intact round window by a novel proteidic delivery technology which may be applied to the nucleic acid-targeting system of the present invention (see, e.g., Qi et al., Gene Therapy (2013), 1-9). In particular, a TAT double stranded RNA-binding domains (TAT-DRBDs), which can transfect Cy3-labeled siRNA into cells of the inner ear, including the inner and outer hair cells, crista ampullaris, macula utriculi and macula sacculi, through intact round-window permeation was successful for delivering double stranded siRNAs *in vivo* for treating various inner ear ailments and preservation of hearing function. About 40 μl of 10mM RNA may be contemplated as the dosage for administration to the ear.

[01012] 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 nucleic acid-targeting system of the present invention for delivery to the ear.

[01013] Mukherjea et al. (Antioxidants & Redox Signaling, Volume 13, Number 5, 2010) document that knockdown of NOX3 using short interfering (si) RNA abrogated cisplatin ototoxicity, as evidenced by protection of OHCs from damage and reduced threshold shifts in auditory brainstem responses (ABRs). Different doses of siNOX3 (0.3, 0.6, and 0.9 μg) were

administered to rats and NOX3 expression was evaluated by real time RT-PCR. The lowest dose of NOX3 siRNA used (0.3 μg) did not show any inhibition of NOX3 mRNA when compared to transtympanic administration of scrambled siRNA or untreated cochleae. However, administration of the higher doses of NOX3 siRNA (0.6 and 0.9 μg) reduced NOX3 expression compared to control scrambled siRNA. Such a system may be applied to the CRISPR Cas system of the present invention for transtympanic administration with a dosage of about 2 mg to about 4 mg of CRISPR Cas for administration to a human.

[01014] Jung et al. (Molecular Therapy, vol. 21 no. 4, 834–841 apr. 2013) demonstrate that Hes5 levels in the utricle decreased after the application of siRNA and that the number of hair cells in these utricles was significantly larger than following control treatment. The data suggest that siRNA technology may be useful for inducing repair and regeneration in the inner ear and that the Notch signaling pathway is a potentially useful target for specific gene expression inhibition. Jung et al. injected 8 μg of Hes5 siRNA in 2 μl volume, prepared by adding sterile normal saline to the lyophilized siRNA to a vestibular epithelium of the ear. Such a system may be applied to the nucleic acid-targeting system of the present invention for administration to the vestibular epithelium of the ear with a dosage of about 1 to about 30 mg of CRISPR Cas for administration to a human.

Gene Targeting in Non-Dividing Cells (Neurones & Muscle)

[01015] 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 CRISPR–Cas9-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.

[01016] Thus, reactivation of HR in cells, especially non-dividing, fully differentiated cell types is preferred, in some embodiments. In some embodiments, promotion of the BRCA1–PALB2 interaction is preferred in some embodiments. In some embodiments, the target ell is a non-dividing cell. In some embodiments, the target cell is a neuron or muscle cell. In some embodiments, the target cell is targeted *in vivo*. In some embodiments, the cell is in G1 and HR is suppressed. In some embodiments, 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 some embodiments, 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 some embodiments, a BRCA1–PALB2 complex may be delivered to the G1 cell. In some embodiments, 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

[01017] The present invention also contemplates delivering the CRISPR-Cas system to one or both eyes.

[01018] In particular embodiments of the invention, the CRISPR-Cas system may 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.

[01019] In some embodiments, the condition to be treated or targeted is an eye disorder. In some embodiments, the eye disorder may include glaucoma. In some embodiments, the eye disorder includes a retinal degenerative disease. In some embodiments, the retinal degenerative disease is selected from Stargardt disease, Bardet-Biedl Syndrome, Best disease, Blue Cone Monochromacy, Choroidermia, Cone-rod dystrophy, Congenital Stationary Night Blindness, Enhanced S-Cone Syndrome, Juvenile X-Linked Retinoschisis, Leber Congenital Amaurosis, Malattia Leventinesse, 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 some embodiments, the retinal degenerative disease is Leber Congenital Amaurosis (LCA) or Retinitis Pigmentosa. In some embodiments, the CRISPR system is delivered to the eye, optionally via intravitreal injection or subretinal injection.

[01020] For administration to the eye, lentiviral vectors, in particular equine infectious anemia viruses (EIAV) are particularly preferred.

[01021] In another embodiment, minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV) are also contemplated, especially for ocular gene therapy (see, e.g., Balagaan, J Gene Med 2006; 8: 275 – 285, Published online 21 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jgm.845). The vectors are contemplated to have cytomegalovirus (CMV) promoter driving expression of the target gene. Intracameral, subretinal, intraocular and intravitreal injections are all contemplated (see, e.g., 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). 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-µl 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 µl 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 titres of either 1.0–1.4 × 1010 or 1.0–1.4 × 109 transducing units (TU)/ml.

[01022] In another embodiment, RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostain and angiostatin that is delivered via a subretinal injection for the treatment of the web form of age-related macular degeneration is also contemplated (see, e.g., Binley et al., HUMAN GENE THERAPY 23:980–991 (September 2012)). Such a vector may be modified for the CRISPR-Cas system of the present invention. Each eye may be treated with either RetinoStat® at a dose of 1.1 x 105 transducing units per eye (TU/eye) in a total volume of 100 μl.

[01023] In another embodiment, an E1-, partial E3-, E4-deleted adenoviral vector may be contemplated for delivery to the eye. Twenty-eight patients with advanced neovascular agerelated macular degeneration (AMD) were given a single intravitreous injection of an E1-, partial E3-, E4-deleted adenoviral vector expressing human pigment ep- ithelium-derived factor (AdPEDF.ll) (see, e.g., Campochiaro et al., Human Gene Therapy 17:167-176 (February 2006)). Doses ranging from 106 to 109.5 particle units (PU) were investigated and there were no serious adverse events related to AdPEDF.ll and no dose-limiting toxicities (see, e.g., Campochiaro et al., Human Gene Therapy 17:167-176 (February 2006)). Adenoviral vectormediated ocular gene transfer appears to be a viable approach for the treatment of ocular disorders and could be applied to the CRISPR Cas system.

[01024] In another embodiment, the sd-rxRNA® system of RXi Pharmaceuticals may be used/and or adapted for delivering CRISPR Cas 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 nucleic acid-targeting system

of the present invention, contemplating a dose of about 3 to 20 mg of CRISPR administered to a human.

[01025] Millington-Ward et al. (Molecular Therapy, vol. 19 no. 4, 642–649 apr. 2011) describes adeno-associated virus (AAV) vectors to deliver an RNA interference (RNAi)-based rhodopsin suppressor and a codon-modified rhodopsin replacement gene resistant to suppression due to nucleotide alterations at degenerate positions over the RNAi target site. An injection of either 6.0 x 108 vp or 1.8 x 1010 vp AAV were subretinally injected into the eyes by Millington-Ward et al. The AAV vectors of Millington-Ward et al. may be applied to the CRISPR Cas system of the present invention, contemplating a dose of about 2 x 1011 to about 6 x 1013 vp administered to a human.

[01026] Dalkara et al. (Sci Transl Med 5, 189ra76 (2013)) also relates to *in vivo* directed evolution to fashion an AAV vector that delivers wild-type versions of defective genes throughout the retina after noninjurious injection into the eyes' vitreous humor. Dalkara describes a 7mer peptide display library and an AAV library constructed by DNA shuffling of cap genes from AAV1, 2, 4, 5, 6, 8, and 9. The rcAAV libraries and rAAV vectors expressing GFP under a CAG or Rho promoter were packaged and deoxyribonuclease-resistant genomic titers were obtained through quantitative PCR. The libraries were pooled, and two rounds of evolution were performed, each consisting of initial library diversification followed by three *in vivo* selection steps. In each such step, P30 rho-GFP mice were intravitreally injected with 2 ml of iodixanol-purified, phosphate-buffered saline (PBS)–dialyzed library with a genomic titer of about 1 × 1012 vg/ml. The AAV vectors of Dalkara et al. may be applied to the nucleic acid-targeting system of the present invention, contemplating a dose of about 1 x 1015 to about 1 x 1016 vg/ml administered to a human.

[01027] In a particular embodiment, the rhodopsin gene may be targeted for the treatment of retinitis pigmentosa (RP), wherein the system of US Patent Publication No. 20120204282 assigned to Sangamo BioSciences, Inc. may be modified in accordance of the CRISPR Cas system of the present invention.

[01028] In another embodiment, the methods of US Patent Publication No. 20130183282 assigned to Cellectis, which is directed to methods of cleaving a target sequence from the human rhodopsin gene, may also be modified to the nucleic acid-targeting system of the present invention.

[01029] US Patent Publication No. 20130202678 assigned to Academia Sinica relates to methods 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. 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 nucleic acid-targeting system of the present invention.

[01030] 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.

[01031] US Patent Publication No. 20120159653, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with macular degeneration (MD). Macular degeneration (MD) is the primary cause of visual impairment in the elderly, but is also a hallmark symptom of childhood diseases such as Stargardt disease, Sorsby fundus, and fatal childhood neurodegenerative diseases, with an age of onset as young as infancy. Macular degeneration results in a loss of vision in the center of the visual field (the macula) because of damage to the retina. Currently existing animal models do not recapitulate major hallmarks of the disease as it is observed in humans. The available animal models comprising mutant genes encoding proteins associated with MD also produce highly variable phenotypes, making translations to human disease and therapy development problematic.

[01032] 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 nucleic acid-targeting system of the present invention. The proteins associated with MD are typically selected based on an experimental association of the protein associated with MD to an MD disorder. For example, the production rate or circulating concentration of a protein associated with MD may be elevated or depressed in a population having an MD disorder relative to a population lacking the MD 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 proteins associated with MD 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).

[01033] By way of non-limiting example, proteins associated with MD include but are not limited to the following proteins: (ABCA4) ATP-binding cassette, sub-family A (ABC1), member 4 ACHM1 achromatopsia (rod monochromacy) 1 ApoE Apolipoprotein E (ApoE) C1QTNF5 (CTRP5) C1q and tumor necrosis factor related protein 5 (C1QTNF5) C2 Complement component 2 (C2) C3 Complement components (C3) CCL2 Chemokine (C-C motif) Ligand 2 (CCL2) CCR2 Chemokine (C-C motif) receptor 2 (CCR2) CD36 Cluster of Differentiation 36 CFB Complement factor B CFH Complement factor CFH H CFHR1 complement factor H-related 1 CFHR3 complement factor H-related 3 CNGB3 cyclic nucleotide gated channel beta 3 CP ceruloplasmin (CP) CRP C reactive protein (CRP) CST3 cystatin C or cystatin 3 (CST3) CTSD Cathepsin D (CTSD) CX3CR1 chemokine (C-X3-C motif) receptor 1 ELOVL4 Elongation of very long chain fatty acids 4 ERCC6 excision repair crosscomplementing rodent repair deficiency, complementation group 6 FBLN5 Fibulin-5 FBLN5 Fibulin 5 FBLN6 Fibulin 6 FSCN2 fascin (FSCN2) HMCN1 Hemicentrin 1 HMCN1 hemicentin 1 HTRA1 HtrA serine peptidase 1 (HTRA1) HTRA1 HtrA serine peptidase 1 IL-6 Interleukin 6 IL-8 Interleukin 8 LOC387715 Hypothetical protein PLEKHA1 Pleckstrin homology domain-containing family A member 1 (PLEKHA1) PROM1 Prominin 1(PROM1 or CD133) PRPH2 Peripherin-2 RPGR retinitis pigmentosa GTPase regulator SERPING1 serpin peptidase inhibitor, clade G, member 1 (C1- inhibitor) TCOF1 Treacle TIMP3 Metalloproteinase inhibitor 3 (TIMP3) TLR3 Toll-like receptor 3.

[01034] The identity of the protein associated with MD whose chromosomal sequence is edited can and will vary. In preferred embodiments, the proteins associated with MD whose chromosomal sequence is edited may be the ATP-binding cassette, sub-family A (ABC1) member 4 protein (ABCA4) encoded by the ABCR gene, the apolipoprotein E protein (APOE) encoded by the APOE gene, the chemokine (C-C motif) Ligand 2 protein (CCL2) encoded by the CCL2 gene, the chemokine (C-C motif) receptor 2 protein (CCR2) encoded by the CCR2 gene, the ceruloplasmin protein (CP) encoded by the CP gene, the cathepsin D protein (CTSD) encoded by the CTSD gene, or the metalloproteinase inhibitor 3 protein (TIMP3) encoded by the TIMP3 gene. In an exemplary embodiment, the genetically modified animal is a rat, and the edited chromosomal sequence encoding the protein associated with MD may be: (ABCA4)

ATPbinding cassette, NM_000350 sub-family A (ABC1), member 4 APOE Apolipoprotein E NM_138828 (APOE) CCL2 Chemokine (C-C NM_031530 motif) Ligand 2 (CCL2) CCR2 Chemokine (C-C NM_021866 motif) receptor 2 (CCR2) CP ceruloplasmin (CP) NM_012532 CTSD Cathepsin D (CTSD) NM_134334 TIMP3 Metalloproteinase NM_012886 inhibitor 3 (TIMP3) The animal or cell may comprise 1, 2, 3, 4, 5, 6, 7 or more disrupted chromosomal sequences encoding a protein associated with MD and zero, 1, 2, 3, 4, 5, 6, 7 or more chromosomally integrated sequences encoding the disrupted protein associated with MD.

[01035] The edited or integrated chromosomal sequence may be modified to encode an altered protein associated with MD. Several mutations in MD-related chromosomal sequences have been associated with MD. Non-limiting examples of mutations in chromosomal sequences associated with MD include those that may cause MD including in the ABCR protein, E471K (i.e. glutamate at position 471 is changed to lysine), R1129L (i.e. arginine at position 1129 is changed to leucine), T1428M (i.e. threonine at position 1428 is changed to methionine), R1517S (i.e. arginine at position 1517 is changed to serine), I1562T (i.e. isoleucine at position 1562 is changed to threonine), and G1578R (i.e. glycine at position 1578 is changed to arginine); in the CCR2 protein, V64I (i.e. valine at position 192 is changed to isoleucine); in CP protein, G969B (i.e. glycine at position 969 is changed to asparagine or aspartate); in TIMP3 protein, S156C (i.e. serine at position 156 is changed to cysteine), G166C (i.e. glycine at position 166 is changed to cysteine), G167C (i.e. glycine at position 167 is changed to cysteine), Y168C (i.e. tyrosine at position 168 is changed to cysteine), S170C (i.e. serine at position 170 is changed to cysteine), Y172C (i.e. tyrosine at position 172 is changed to cysteine) and S181C (i.e. serine at position 181 is changed to cysteine). Other associations of genetic variants in MD-associated genes and disease are known in the art.

[01036] CRISPR systems are useful to correct diseases resulting from autosomal dominant genes. For example, CRISPR/Cas9 was used to remove an autosomal dominant gene that causes receptor loss in the eye. Bakondi, B. et al., *In Vivo* CRISPR/Cas9 Gene Editing Corrects Retinal Dystrophy in the S334ter-3 Rat Model of Autosomal Dominant Retinitis Pigmentosa. Molecular Therapy, 2015; DOI: 10.1038/mt.2015.220.

[01037] Treating Circulatory and Muscular Diseases

[01038] The present invention also contemplates delivering the CRISPR-Cas system described herein, e.g. Type VII effector protein systems, to the heart. For the heart, a myocardium tropic adena-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 1-10 x 1014 vector genomes are contemplated for systemic administration. See also, e.g., Eulalio et al. (2012) Nature 492: 376 and Somasuntharam et al. (2013) Biomaterials 34: 7790.

[01039] For example, US Patent Publication No. 20110023139, 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).

[01040] By way of example, the chromosomal sequence may comprise, but is not limited to, IL1B (interleukin 1, beta), XDH (xanthine dehydrogenase), TP53 (tumor protein p53), PTGIS (prostaglandin 12 (prostacyclin) synthase), MB (myoglobin), IL4 (interleukin 4), ANGPT1 (angiopoietin 1), ABCG8 (ATP-binding cassette, sub-family G (WHITE), member 8), CTSK (cathepsin K), PTGIR (prostaglandin 12 (prostacyclin) receptor (IP)), KCNJ11 (potassium inwardly-rectifying channel, subfamily J, member 11), INS (insulin), CRP (C-reactive protein, pentraxin-related), PDGFRB (platelet-derived growth factor receptor, beta polypeptide), CCNA2 (cyclin A2), PDGFB (platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog)), KCNJ5 (potassium inwardly-rectifying channel, subfamily J, member 5), KCNN3 (potassium intermediate/small conductance calcium-activated channel, subfamily N, member 3), CAPN10 (calpain 10), PTGES (prostaglandin E synthase), ADRA2B (adrenergic, alpha-2B-, receptor), ABCG5 (ATP-

binding cassette, sub-family G (WHITE), member 5), PRDX2 (peroxiredoxin 2), CAPN5 (calpain 5), PARP14 (poly (ADP-ribose) polymerase family, member 14), MEX3C (mex-3 homolog C (C. elegans)), ACE angiotensin I converting enzyme (peptidyl-dipeptidase A) 1), TNF (tumor necrosis factor (TNF superfamily, member 2)), IL6 (interleukin 6 (interferon, beta 2)), STN (statin), SERPINE1 (serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1), ALB (albumin), ADIPOQ (adiponectin, C1Q and collagen domain containing), APOB (apolipoprotein B (including Ag(x) antigen)), APOE (apolipoprotein E), LEP (leptin), MTHFR (5,10-methylenetetrahydrofolate reductase (NADPH)), APOA1 (apolipoprotein A-I), EDN1 (endothelin 1), NPPB (natriuretic peptide precursor B), NOS3 (nitric oxide synthase 3 (endothelial cell)), PPARG (peroxisome proliferator-activated receptor gamma), PLAT (plasminogen activator, tissue), PTGS2 (prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)), CETP (cholesteryl ester transfer protein, plasma), AGTR1 (angiotensin II receptor, type 1), HMGCR (3-hydroxy-3-methylglutaryl-Coenzyme A reductase), IGF1 (insulin-like growth factor 1 (somatomedin C)), SELE (selectin E), REN (renin), PPARA (peroxisome proliferator-activated receptor alpha), PON1 (paraoxonase 1), KNG1 (kininogen 1), CCL2 (chemokine (C-C motif) ligand 2), LPL (lipoprotein lipase), VWF (von Willebrand factor), F2 (coagulation factor II (thrombin)), ICAM1 (intercellular adhesion molecule 1), TGFB1 (transforming growth factor, beta 1), NPPA (natriuretic peptide precursor A), IL10 (interleukin 10), EPO (erythropoietin), SOD1 (superoxide dismutase 1, soluble), VCAM1 (vascular cell adhesion molecule 1), IFNG (interferon, gamma), LPA (lipoprotein, Lp(a)), MPO (myeloperoxidase), ESR1 (estrogen receptor 1), MAPK1 (mitogen-activated protein kinase 1), HP (haptoglobin), F3 (coagulation factor III (thromboplastin, tissue factor)), CST3 (cystatin C), COG2 (component of oligomeric golgi complex 2), MMP9 (matrix metallopeptidase 9 (gelatinase B, 92 kDa gelatinase, 92 kDa type IV collagenase)), SERPINC1 (serpin peptidase inhibitor, clade C (antithrombin), member 1), F8 (coagulation factor VIII, procoagulant component), HMOX1 (heme oxygenase (decycling) 1), APOC3 (apolipoprotein C-III), IL8 (interleukin 8), PROK1 (prokineticin 1), CBS (cystathionine-beta-synthase), NOS2 (nitric oxide synthase 2, inducible), TLR4 (toll-like receptor 4), SELP (selectin P (granule membrane protein 140 kDa, antigen CD62)), ABCA1 (ATP-binding cassette, sub-family A (ABC1), member 1), AGT (angiotensinogen (serpin peptidase inhibitor, clade A, member 8)), LDLR (low density lipoprotein receptor), GPT (glutamic-pyruvate transaminase (alanine aminotransferase)),

VEGFA (vascular endothelial growth factor A), NR3C2 (nuclear receptor subfamily 3, group C, member 2), IL18 (interleukin 18 (interferon-gamma-inducing factor)), NOS1 (nitric oxide synthase 1 (neuronal)), NR3C1 (nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)), FGB (fibrinogen beta chain), HGF (hepatocyte growth factor (hepapoietin A; scatter factor)), IL1A (interleukin 1, alpha), RETN (resistin), AKT1 (v-akt murine thymoma viral oncogene homolog 1), LIPC (lipase, hepatic), HSPD1 (heat shock 60 kDa protein 1 (chaperonin)), MAPK14 (mitogen-activated protein kinase 14), SPP1 (secreted phosphoprotein 1), ITGB3 (integrin, beta 3 (platelet glycoprotein 111a, antigen CD61)), CAT (catalase), UTS2 (urotensin 2), THBD (thrombomodulin), F10 (coagulation factor X), CP (ceruloplasmin (ferroxidase)), TNFRSF11B (tumor necrosis factor receptor superfamily, member 11b), EDNRA (endothelin receptor type A), EGFR (epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)), MMP2 (matrix metallopeptidase 2 (gelatinase A, 72 kDa gelatinase, 72 kDa type IV collagenase)), PLG (plasminogen), NPY (neuropeptide Y), RHOD (ras homolog gene family, member D), MAPK8 (mitogen-activated protein kinase 8), MYC (v-myc myelocytomatosis viral oncogene homolog (avian)), FN1 (fibronectin 1), CMA1 (chymase 1, mast cell), PLAU (plasminogen activator, urokinase), GNB3 (guanine nucleotide binding protein (G protein), beta polypeptide 3), ADRB2 (adrenergic, beta-2-, receptor, surface), APOA5 (apolipoprotein A-V), SOD2 (superoxide dismutase 2, mitochondrial), F5 (coagulation factor V (proaccelerin, labile factor)), VDR (vitamin D (1,25-dihydroxyvitamin D3) receptor), ALOX5 (arachidonate 5-lipoxygenase), HLA-DRB1 (major histocompatibility complex, class II, DR beta 1), PARP1 (poly (ADP-ribose) polymerase 1), CD40LG (CD40 ligand), PON2 (paraoxonase 2), AGER (advanced glycosylation end product-specific receptor), IRS1 (insulin receptor substrate 1), PTGS1 (prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)), ECE1 (endothelin converting enzyme 1), F7 (coagulation factor VII (serum prothrombin conversion accelerator)), URN (interleukin 1 receptor antagonist), EPHX2 (epoxide hydrolase 2, cytoplasmic), IGFBP1 (insulin-like growth factor binding protein 1), MAPK10 (mitogen-activated protein kinase 10), FAS (Fas (TNF receptor superfamily, member 6)), ABCB1 (ATP-binding cassette, sub-family B (MDR/TAP), member 1), JUN (jun oncogene), IGFBP3 (insulin-like growth factor binding protein 3), CD14 (CD14 molecule), PDE5A (phosphodiesterase 5A, cGMP-specific), AGTR2 (angiotensin II receptor, type 2), CD40 (CD40 molecule, TNF receptor superfamily member 5), LCAT (lecithin-cholesterol

acyltransferase), CCR5 (chemokine (C-C motif) receptor 5), MMP1 (matrix metallopeptidase 1 (interstitial collagenase)), TIMP1 (TIMP metallopeptidase inhibitor 1), ADM (adrenomedullin), DYT10 (dystonia 10), STAT3 (signal transducer and activator of transcription 3 (acute-phase response factor)), MMP3 (matrix metallopeptidase 3 (stromelysin 1, progelatinase)), ELN (elastin), USF1 (upstream transcription factor 1), CFH (complement factor H), HSPA4 (heat shock 70 kDa protein 4), MMP12 (matrix metallopeptidase 12 (macrophage elastase)), MME (membrane metallo-endopeptidase), F2R (coagulation factor II (thrombin) receptor), SELL (selectin L), CTSB (cathepsin B), ANXA5 (annexin A5), ADRB1 (adrenergic, beta-1-, receptor), CYBA (cytochrome b-245, alpha polypeptide), FGA (fibrinogen alpha chain), GGT1 (gamma-glutamyltransferase 1), LIPG (lipase, endothelial), HIF1A (hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)), CXCR4 (chemokine (C-X-C motif) receptor 4), PROC (protein C (inactivator of coagulation factors Va and VIIIa)), SCARB1 (scavenger receptor class B, member 1), CD79A (CD79a molecule, immunoglobulin-associated alpha), PLTP (phospholipid transfer protein), ADD1 (adducin 1 (alpha)), FGG (fibrinogen gamma chain), SAA1 (serum amyloid A1), KCNH2 (potassium voltage-gated channel, subfamily H (eag-related), member 2), DPP4 (dipeptidyl-peptidase 4), G6PD (glucose-6-phosphate dehydrogenase), NPR1 (natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A)), VTN (vitronectin), KIAA0101 (KIAA0101), FOS (FBJ murine osteosarcoma viral oncogene homolog), TLR2 (toll-like receptor 2), PPIG (peptidylprolyl isomerase G (cyclophilin G)), IL1R1 (interleukin 1 receptor, type I), AR (androgen receptor), CYP1A1 (cytochrome P450, family 1, subfamily A, polypeptide 1), SERPINA1 (serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1), MTR (5-methyltetrahydrofolate-homocysteine methyltransferase), RBP4 (retinol binding protein 4, plasma), APOA4 (apolipoprotein A-IV), CDKN2A (cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)), FGF2 (fibroblast growth factor 2 (basic)), EDNRB (endothelin receptor type B), ITGA2 (integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)), CABIN1 (calcineurin binding protein 1), SHBG (sex hormone-binding globulin), HMGB1 (high-mobility group box 1), HSP90B2P (heat shock protein 90 kDa beta (Grp94), member 2 (pseudogene)), CYP3A4 (cytochrome P450, family 3, subfamily A, polypeptide 4), GJA1 (gap junction protein, alpha 1, 43 kDa), CAV1 (caveolin 1, caveolae protein, 22 kDa), ESR2 (estrogen receptor 2 (ER beta)), LTA (lymphotoxin alpha (TNF superfamily, member 1)), GDF15 (growth differentiation factor 15), BDNF (brain-

derived neurotrophic factor), CYP2D6 (cytochrome P450, family 2, subfamily D, polypeptide 6), NGF (nerve growth factor (beta polypeptide)), SP1 (Sp1 transcription factor), TGIF1 (TGFB-induced factor homeobox 1), SRC (v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)), EGF (epidermal growth factor (beta-urogastrone)), PIK3CG (phosphoinositide-3-kinase, catalytic, gamma polypeptide), HLA-A (major histocompatibility complex, class I, A), KCNQ1 (potassium voltage-gated channel, KQT-like subfamily, member 1), CNR1 (cannabinoid receptor 1 (brain)), FBN1 (fibrillin 1), CHKA (choline kinase alpha), BEST1 (bestrophin 1), APP (amyloid beta (A4) precursor protein), CTNNB1 (catenin (cadherin-associated protein), beta 1, 88 kDa), IL2 (interleukin 2), CD36 (CD36 molecule (thrombospondin receptor)), PRKAB1 (protein kinase, AMP-activated, beta 1 non-catalytic subunit), TPO (thyroid peroxidase), ALDH7A1 (aldehyde dehydrogenase 7 family, member A1), CX3CR1 (chemokine (C-X3-C motif) receptor 1), TH (tyrosine hydroxylase), F9 (coagulation factor IX), GH1 (growth hormone 1), TF (transferrin), HFE (hemochromatosis), IL17A (interleukin 17A), PTEN (phosphatase and tensin homolog), GSTM1 (glutathione S-transferase mu 1), DMD (dystrophin), GATA4 (GATA binding protein 4), F13A1 (coagulation factor XIII, A1 polypeptide), TTR (transthyretin), FABP4 (fatty acid binding protein 4, adipocyte), PON3 (paraoxonase 3), APOC1 (apolipoprotein C-I), INSR (insulin receptor), TNFRSF1B (tumor necrosis factor receptor superfamily, member 1B), HTR2A (5-hydroxytryptamine (serotonin) receptor 2A), CSF3 (colony stimulating factor 3 (granulocyte)), CYP2C9 (cytochrome P450, family 2, subfamily C, polypeptide 9), TXN (thioredoxin), CYP11B2 (cytochrome P450, family 11, subfamily B, polypeptide 2), PTH (parathyroid hormone), CSF2 (colony stimulating factor 2 (granulocyte-macrophage)), KDR (kinase insert domain receptor (a type III receptor tyrosine kinase)), PLA2G2A (phospholipase A2, group IIA (platelets, synovial fluid)), B2M (beta-2-microglobulin), THBS1 (thrombospondin 1), GCG (glucagon), RHOA (ras homolog gene family, member A), ALDH2 (aldehyde dehydrogenase 2 family (mitochondrial)), TCF7L2 (transcription factor 7-like 2 (T-cell specific, HMG-box)), BDKRB2 (bradykinin receptor B2), NFE2L2 (nuclear factor (erythroid-derived 2)-like 2), NOTCH1 (Notch homolog 1, translocation-associated (Drosophila)), UGT1A1 (UDP glucuronosyltransferase 1 family, polypeptide A1), IFNA1 (interferon, alpha 1), PPARD (peroxisome proliferator-activated receptor delta), SIRT1 (sirtuin (silent mating type information regulation 2 homolog) 1 (S. cerevisiae)), GNRH1 (gonadotropin-releasing hormone 1 (luteinizing-releasing hormone)), PAPPA (pregnancy-associated plasma protein A,

pappalysin 1), ARR3 (arrestin 3, retinal (X-arrestin)), NPPC (natriuretic peptide precursor C), AHSP (alpha hemoglobin stabilizing protein), PTK2 (PTK2 protein tyrosine kinase 2), IL13 (interleukin 13), MTOR (mechanistic target of rapamycin (serine/threonine kinase)), ITGB2 (integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)), GSTT1 (glutathione S-transferase theta 1), IL6ST (interleukin 6 signal transducer (gp130, oncostatin M receptor)), CPB2 (carboxypeptidase B2 (plasma)), CYP1A2 (cytochrome P450, family 1, subfamily A, polypeptide 2), HNF4A (hepatocyte nuclear factor 4, alpha), SLC6A4 (solute carrier family 6 (neurotransmitter transporter, serotonin), member 4), PLA2G6 (phospholipase A2, group VI (cytosolic, calcium-independent)), TNFSF11 (tumor necrosis factor (ligand) superfamily, member 11), SLC8A1 (solute carrier family 8 (sodium/calcium exchanger), member 1), F2RL1 (coagulation factor II (thrombin) receptor-like 1), AKR1A1 (aldo-keto reductase family 1, member A1 (aldehyde reductase)), ALDH9A1 (aldehyde dehydrogenase 9 family, member A1), BGLAP (bone gamma-carboxyglutamate (gla) protein), MTTP (microsomal triglyceride transfer protein), MTRR (5-methyltetrahydrofolate-homocysteine methyltransferase reductase), SULT1A3 (sulfotransferase family, cytosolic, 1A, phenol-preferring, member 3), RAGE (renal tumor antigen), C4B (complement component 4B (Chido blood group), P2RY12 (purinergic receptor P2Y, G-protein coupled, 12), RNLS (renalase, FAD-dependent amine oxidase), CREB1 (cAMP responsive element binding protein 1), POMC (proopiomelanocortin), RAC1 (ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)), LMNA (lamin NC), CD59 (CD59 molecule, complement regulatory protein), SCN5A (sodium channel, voltage-gated, type V, alpha subunit), CYP1B1 (cytochrome P450, family 1, subfamily B, polypeptide 1), MIF (macrophage migration inhibitory factor (glycosylation-inhibiting factor)), MMP13 (matrix metallopeptidase 13 (collagenase 3)), TIMP2 (TIMP metallopeptidase inhibitor 2), CYP19A1 (cytochrome P450, family 19, subfamily A, polypeptide 1), CYP21A2 (cytochrome P450, family 21, subfamily A, polypeptide 2), PTPN22 (protein tyrosine phosphatase, non-receptor type 22 (lymphoid)), MYH14 (myosin, heavy chain 14, non-muscle), MBL2 (mannose-binding lectin (protein C) 2, soluble (opsonic defect)), SELPLG (selectin P ligand), AOC3 (amine oxidase, copper containing 3 (vascular adhesion protein 1)), CTSL1 (cathepsin L1), PCNA (proliferating cell nuclear antigen), IGF2 (insulin-like growth factor 2 (somatomedin A)), ITGB1 (integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)), CAST (calpastatin), CXCL12 (chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)),

IGHE (immunoglobulin heavy constant epsilon), KCNE1 (potassium voltage-gated channel, Isk-related family, member 1), TFRC (transferrin receptor (p90, CD71)), COL1A1 (collagen, type I, alpha 1), COL1A2 (collagen, type I, alpha 2), IL2RB (interleukin 2 receptor, beta), PLA2G10 (phospholipase A2, group X), ANGPT2 (angiopoietin 2), PROCR (protein C receptor, endothelial (EPCR)), NOX4 (NADPH oxidase 4), HAMP (hepcidin antimicrobial peptide), PTPN11 (protein tyrosine phosphatase, non-receptor type 11), SLC2A1 (solute carrier family 2 (facilitated glucose transporter), member 1), IL2RA (interleukin 2 receptor, alpha), CCL5 (chemokine (C-C motif) ligand 5), IRF1 (interferon regulatory factor 1), CFLAR (CASP8 and FADD-like apoptosis regulator), CALCA (calcitonin-related polypeptide alpha), EIF4E (eukaryotic translation initiation factor 4E), GSTP1 (glutathione S-transferase pi 1), JAK2 (Janus kinase 2), CYP3A5 (cytochrome P450, family 3, subfamily A, polypeptide 5), HSPG2 (heparan sulfate proteoglycan 2), CCL3 (chemokine (C-C motif) ligand 3), MYD88 (myeloid differentiation primary response gene (88)), VIP (vasoactive intestinal peptide), SOAT1 (sterol O-acyltransferase 1), ADRBK1 (adrenergic, beta, receptor kinase 1), NR4A2 (nuclear receptor subfamily 4, group A, member 2), MMP8 (matrix metallopeptidase 8 (neutrophil collagenase)), NPR2 (natriuretic peptide receptor B/guanylate cyclase B (atrionatriuretic peptide receptor B)), GCH1 (GTP cyclohydrolase 1), EPRS (glutamyl-prolyl-tRNA synthetase), PPARGC1A (peroxisome proliferator-activated receptor gamma, coactivator 1 alpha), F12 (coagulation factor XII (Hageman factor)), PECAM1 (platelet/endothelial cell adhesion molecule), CCL4 (chemokine (C-C motif) ligand 4), SERPINA3 (serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3), CASR (calcium-sensing receptor), GJA5 (gap junction protein, alpha 5, 40 kDa), FABP2 (fatty acid binding protein 2, intestinal), TTF2 (transcription termination factor, RNA polymerase II), PROS1 (protein S (alpha)), CTF1 (cardiotrophin 1), SGCB (sarcoglycan, beta (43 kDa dystrophin-associated glycoprotein)), YME1L1 (YME1-like 1 (S. cerevisiae)), CAMP (cathelicidin antimicrobial peptide), ZC3H12A (zinc finger CCCH-type containing 12A), AKR1B1 (aldo-keto reductase family 1, member B1 (aldose reductase)), DES (desmin), MMP7 (matrix metallopeptidase 7 (matrilysin, uterine)), AHR (aryl hydrocarbon receptor), CSF1 (colony stimulating factor 1 (macrophage)), HDAC9 (histone deacetylase 9), CTGF (connective tissue growth factor), KCNMA1 (potassium large conductance calcium-activated channel, subfamily M, alpha member 1), UGT1A (UDP glucuronosyltransferase 1 family, polypeptide A complex locus), PRKCA (protein kinase C, alpha), COMT (catechol-.beta.-

methyltransferase), S100B (S100 calcium binding protein B), EGR1 (early growth response 1), PRL (prolactin), IL15 (interleukin 15), DRD4 (dopamine receptor D4), CAMK2G (calcium/calmodulin-dependent protein kinase II gamma), SLC22A2 (solute carrier family 22 (organic cation transporter), member 2), CCL11 (chemokine (C-C motif) ligand 11), PGF (B321 placental growth factor), THPO (thrombopoietin), GP6 (glycoprotein VI (platelet)), TACR1 (tachykinin receptor 1), NTS (neurotensin), HNF1A (HNF1 homeobox A), SST (somatostatin), KCND1 (potassium voltage-gated channel, Shal-related subfamily, member 1), LOC646627 (phospholipase inhibitor), TBXAS1 (thromboxane A synthase 1 (platelet)), CYP2J2 (cytochrome P450, family 2, subfamily J, polypeptide 2), TBXA2R (thromboxane A2 receptor), ADH1C (alcohol dehydrogenase 1C (class I), gamma polypeptide), ALOX12 (arachidonate 12-lipoxygenase), AHSG (alpha-2-HS-glycoprotein), BHMT (betaine-homocysteine methyltransferase), GJA4 (gap junction protein, alpha 4, 37 kDa), SLC25A4 (solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4), ACLY (ATP citrate lyase), ALOX5AP (arachidonate 5-lipoxygenase-activating protein), NUMA1 (nuclear mitotic apparatus protein 1), CYP27B1 (cytochrome P450, family 27, subfamily B, polypeptide 1), CYSLTR2 (cysteinyl leukotriene receptor 2), SOD3 (superoxide dismutase 3, extracellular), LTC4S (leukotriene C4 synthase), UCN (urocortin), GHRL (ghrelin/obestatin prepropeptide), APOC2 (apolipoprotein C-II), CLEC4A (C-type lectin domain family 4, member A), KBTBD10 (kelch repeat and BTB (POZ) domain containing 10), TNC (tenascin C), TYMS (thymidylate synthetase), SHCl (SHC (Src homology 2 domain containing) transforming protein 1), LRP1 (low density lipoprotein receptor-related protein 1), SOCS3 (suppressor of cytokine signaling 3), ADH1B (alcohol dehydrogenase 1B (class I), beta polypeptide), KLK3 (kallikrein-related peptidase 3), HSD11B1 (hydroxysteroid (11-beta) dehydrogenase 1), VKORC1 (vitamin K epoxide reductase complex, subunit 1), SERPINB2 (serpin peptidase inhibitor, clade B (ovalbumin), member 2), TNS1 (tensin 1), RNF19A (ring finger protein 19A), EPOR (erythropoietin receptor), ITGAM (integrin, alpha M (complement component 3 receptor 3 subunit)), PITX2 (paired-like homeodomain 2), MAPK7 (mitogen-activated protein kinase 7), FCGR3A (Fc fragment of IgG, low affinity 111a, receptor (CD16a)), LEPR (leptin receptor), ENG (endoglin), GPX1 (glutathione peroxidase 1), GOT2 (glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)), HRH1 (histamine receptor H1), NR112 (nuclear receptor subfamily 1, group I, member 2), CRH (corticotropin releasing hormone), HTR1A (5-hydroxytryptamine (serotonin) receptor 1A),

VDAC1 (voltage-dependent anion channel 1), HPSE (heparanase), SFTPD (surfactant protein D), TAP2 (transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)), RNF123 (ring finger protein 123), PTK2B (PTK2B protein tyrosine kinase 2 beta), NTRK2 (neurotrophic tyrosine kinase, receptor, type 2), IL6R (interleukin 6 receptor), ACHE (acetylcholinesterase (Yt blood group)), GLP1R (glucagon-like peptide 1 receptor), GHR (growth hormone receptor), GSR (glutathione reductase), NQO1 (NAD(P)H dehydrogenase, quinone 1), NR5A1 (nuclear receptor subfamily 5, group A, member 1), GJB2 (gap junction protein, beta 2, 26 kDa), SLC9A1 (solute carrier family 9 (sodium/hydrogen exchanger), member 1), MAOA (monoamine oxidase A), PCSK9 (proprotein convertase subtilisin/kexin type 9), FCGR2A (Fc fragment of IgG, low affinity IIa, receptor (CD32)), SERPINF1 (serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1), EDN3 (endothelin 3), DHFR (dihydrofolate reductase), GAS6 (growth arrest-specific 6), SMPD1 (sphingomyelin phosphodiesterase 1, acid lysosomal), UCP2 (uncoupling protein 2 (mitochondrial, proton carrier)), TFAP2A (transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)), C4BPA (complement component 4 binding protein, alpha), SERPINF2 (serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 2), TYMP (thymidine phosphorylase), ALPP (alkaline phosphatase, placental (Regan isozyme)), CXCR2 (chemokine (C-X-C motif) receptor 2), SLC39A3 (solute carrier family 39 (zinc transporter), member 3), ABCG2 (ATP-binding cassette, sub-family G (WHITE), member 2), ADA (adenosine deaminase), JAK3 (Janus kinase 3), HSPA1A (heat shock 70 kDa protein 1A), FASN (fatty acid synthase), FGF1 (fibroblast growth factor 1 (acidic)), F11 (coagulation factor XI), ATP7A (ATPase, Cu++ transporting, alpha polypeptide), CR1 (complement component (3b/4b) receptor 1 (Knops blood group)), GFAP (glial fibrillary acidic protein), ROCK1 (Rho-associated, coiled-coil containing protein kinase 1), MECP2 (methyl CpG binding protein 2 (Rett syndrome)), MYLK (myosin light chain kinase), BCHE (butyrylcholinesterase), LIPE (lipase, hormone-sensitive), PRDX5 (peroxiredoxin 5), ADORA1 (adenosine A1 receptor), WRN (Werner syndrome, RecQ helicase-like), CXCR3 (chemokine (C-X-C motif) receptor 3), CD81 (CD81 molecule), SMAD7 (SMAD family member 7), LAMC2 (laminin, gamma 2), MAP3K5 (mitogen-activated protein kinase kinase kinase 5), CHGA (chromogranin A (parathyroid secretory protein 1)), IAPP (islet amyloid polypeptide), RHO (rhodopsin), ENPP1 (ectonucleotide pyrophosphatase/phosphodiesterase 1), PTHLH (parathyroid hormone-like hormone), NRG1

(neuregulin 1), VEGFC (vascular endothelial growth factor C), ENPEP (glutamyl aminopeptidase (aminopeptidase A)), CEBPB (CCAAT/enhancer binding protein (C/EBP), beta), NAGLU (N-acetylglucosaminidase, alpha-), F2RL3 (coagulation factor II (thrombin) receptor-like 3), CX3CL1 (chemokine (C-X3-C motif) ligand 1), BDKRB1 (bradykinin receptor B1), ADAMTS13 (ADAM metallopeptidase with thrombospondin type 1 motif, 13), ELANE (elastase, neutrophil expressed), ENPP2 (ectonucleotide pyrophosphatase/phosphodiesterase 2), CISH (cytokine inducible SH2-containing protein), GAST (gastrin), MYOC (myocilin, trabecular meshwork inducible glucocorticoid response), ATP1A2 (ATPase, Na+/K+ transporting, alpha 2 polypeptide), NF1 (neurofibromin 1), GJB1 (gap junction protein, beta 1, 32 kDa), MEF2A (myocyte enhancer factor 2A), VCL (vinculin), BMPR2 (bone morphogenetic protein receptor, type II (serine/threonine kinase)), TUBB (tubulin, beta), CDC42 (cell division cycle 42 (GTP binding protein, 25 kDa)), KRT18 (keratin 18), HSF1 (heat shock transcription factor 1), MYB (v-myb myeloblastosis viral oncogene homolog (avian)), PRKAA2 (protein kinase, AMP-activated, alpha 2 catalytic subunit), ROCK2 (Rho-associated, coiled-coil containing protein kinase 2), TFPI (tissue factor pathway inhibitor (lipoprotein-associated coagulation inhibitor)), PRKG1 (protein kinase, cGMP-dependent, type I), BMP2 (bone morphogenetic protein 2), CTNND1 (catenin (cadherin-associated protein), delta 1), CTH (cystathionase (cystathionine gamma-lyase)), CTSS (cathepsin S), VAV2 (vav 2 guanine nucleotide exchange factor), NPY2R (neuropeptide Y receptor Y2), IGFBP2 (insulin-like growth factor binding protein 2, 36 kDa), CD28 (CD28 molecule), GSTA1 (glutathione S-transferase alpha 1), PPIA (peptidylprolyl isomerase A (cyclophilin A)), APOH (apolipoprotein H (beta-2-glycoprotein I)), S100A8 (S100 calcium binding protein A8), IL11 (interleukin 11), ALOX15 (arachidonate 15-lipoxygenase), FBLN1 (fibulin 1), NR1H3 (nuclear receptor subfamily 1, group H, member 3), SCD (stearoyl-CoA desaturase (delta-9-desaturase)), GIP (gastric inhibitory polypeptide), CHGB (chromogranin B (secretogranin 1)), PRKCB (protein kinase C, beta), SRD5A1 (steroid-5-alpha-reductase, alpha polypeptide 1 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 1)), HSD11B2 (hydroxysteroid (11-beta) dehydrogenase 2), CALCRL (calcitonin receptor-like), GALNT2 (UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-T2)), ANGPTL4 (angiopoietin-like 4), KCNN4 (potassium intermediate/small conductance calcium-activated channel, subfamily N, member 4), PIK3C2A (phosphoinositide-3-kinase, class 2, alpha polypeptide), HBEGF (heparin-binding EGF-like

growth factor), CYP7A1 (cytochrome P450, family 7, subfamily A, polypeptide 1), HLA-DRB5 (major histocompatibility complex, class II, DR beta 5), BNIP3 (BCL2/adenovirus E1B 19 kDa interacting protein 3), GCKR (glucokinase (hexokinase 4) regulator), S100A12 (S100 calcium binding protein A12), PADI4 (peptidyl arginine deiminase, type IV), HSPA14 (heat shock 70 kDa protein 14), CXCR1 (chemokine (C-X-C motif) receptor 1), H19 (H19, imprinted maternally expressed transcript (non-protein coding)), KRTAP19-3 (keratin associated protein 19-3), IDDM2 (insulin-dependent diabetes mellitus 2), RAC2 (ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)), RYR1 (ryanodine receptor 1 (skeletal)), CLOCK (clock homolog (mouse)), NGFR (nerve growth factor receptor (TNFR superfamily, member 16)), DBH (dopamine beta-hydroxylase (dopamine beta-monooxygenase)), CHRNA4 (cholinergic receptor, nicotinic, alpha 4), CACNA1C (calcium channel, voltage-dependent, L type, alpha 1C subunit), PRKAG2 (protein kinase, AMP-activated, gamma 2 non-catalytic subunit), CHAT (choline acetyltransferase), PTGDS (prostaglandin D2 synthase 21 kDa (brain)), NR1H2 (nuclear receptor subfamily 1, group H, member 2), TEK (TEK tyrosine kinase, endothelial), VEGFB (vascular endothelial growth factor B), MEF2C (myocyte enhancer factor 2C), MAPKAPK2 (mitogen-activated protein kinase-activated protein kinase 2), TNFRSF11A (tumor necrosis factor receptor superfamily, member 11a, NFKB activator), HSPA9 (heat shock 70 kDa protein 9 (mortalin)), CYSLTR1 (cysteinyl leukotriene receptor 1), MAT1A (methionine adenosyltransferase I, alpha), OPRL1 (opiate receptor-like 1), IMPA1 (inositol(myo)-1(or 4)-monophosphatase 1), CLCN2 (chloride channel 2), DLD (dihydrolipoamide dehydrogenase), PSMA6 (proteasome (prosome, macropain) subunit, alpha type, 6), PSMB8 (proteasome (prosome, macropain) subunit, beta type, 8 (large multifunctional peptidase 7)), CHI3L1 (chitinase 3-like 1 (cartilage glycoprotein-39)), ALDH1B1 (aldehyde dehydrogenase 1 family, member B1), PARP2 (poly (ADP-ribose) polymerase 2), STAR (steroidogenic acute regulatory protein), LBP (lipopolysaccharide binding protein), ABCC6 (ATP-binding cassette, sub-family C(CFTR/MRP), member 6), RGS2 (regulator of G-protein signaling 2, 24 kDa), EFNB2 (ephrin-B2), GJB6 (gap junction protein, beta 6, 30 kDa), APOA2 (apolipoprotein A-II), AMPD1 (adenosine monophosphate deaminase 1), DYSF (dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)), FDFT1 (farnesyl-diphosphate farnesyltransferase 1), EDN2 (endothelin 2), CCR6 (chemokine (C-C motif) receptor 6), GJB3 (gap junction protein, beta 3, 31 kDa), IL1RL1 (interleukin 1 receptor-like 1), ENTPD1 (ectonucleoside triphosphate

diphosphohydrolase 1), BBS4 (Bardet-Biedl syndrome 4), CELSR2 (cadherin, EGF LAG seven-pass G-type receptor 2 (flamingo homolog, Drosophila)), F11R (F11 receptor), RAPGEF3 (Rap guanine nucleotide exchange factor (GEF) 3), HYAL1 (hyaluronoglucosaminidase 1), ZNF259 (zinc finger protein 259), ATOX1 (ATX1 antioxidant protein 1 homolog (yeast)), ATF6 (activating transcription factor 6), KHK (ketohexokinase (fructokinase)), SAT1 (spermidine/spermine N1-acetyltransferase 1), GGH (gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase)), TIMP4 (TIMP metallopeptidase inhibitor 4), SLC4A4 (solute carrier family 4, sodium bicarbonate cotransporter, member 4), PDE2A (phosphodiesterase 2A, cGMP-stimulated), PDE3B (phosphodiesterase 3B, cGMP-inhibited), FADS1 (fatty acid desaturase 1), FADS2 (fatty acid desaturase 2), TMSB4X (thymosin beta 4, X-linked), TXNIP (thioredoxin interacting protein), LIMS1 (LIM and senescent cell antigen-like domains 1), RHOB (ras homolog gene family, member B), LY96 (lymphocyte antigen 96), FOXO1 (forkhead box O1), PNPLA2 (patatin-like phospholipase domain containing 2), TRH (thyrotropin-releasing hormone), GJC1 (gap junction protein, gamma 1, 45 kDa), SLC17A5 (solute carrier family 17 (anion/sugar transporter), member 5), FTO (fat mass and obesity associated), GJD2 (gap junction protein, delta 2, 36 kDa), PSRC1 (proline/serine-rich coiled-coil 1), CASP12 (caspase 12 (gene/pseudogene)), GPBAR1 (G protein-coupled bile acid receptor 1), PXK (PX domain containing serine/threonine kinase), IL33 (interleukin 33), TRIB1 (tribbles homolog 1 (Drosophila)), PBX4 (pre-B-cell leukemia homeobox 4), NUPR1 (nuclear protein, transcriptional regulator, 1), 15-Sep(15 kDa selenoprotein), CILP2 (cartilage intermediate layer protein 2), TERC (telomerase RNA component), GGT2 (gamma-glutamyltransferase 2), MT-CO1 (mitochondrially encoded cytochrome c oxidase I), and UOX (urate oxidase, pseudogene). Any of these sequences, may be a target for the CRISPR-Cas system, e.g., to address mutation.

[01041] In an additional embodiment, the chromosomal sequence may further be selected from Pon1 (paraoxonase 1), LDLR (LDL receptor), ApoE (Apolipoprotein E), Apo B-100 (Apolipoprotein B-100), ApoA (Apolipoprotein(a)), ApoA1 (Apolipoprotein A1), CBS (Cystathione B-synthase), Glycoprotein IIb/IIb, MTHRF (5,10-methylenetetrahydrofolate reductase (NADPH), and combinations thereof. In one iteration, the chromosomal sequences and proteins encoded by chromosomal sequences involved in cardiovascular disease may be chosen from Cacna1C, Sod1, Pten, Ppar(alpha), Apo E, Leptin, and combinations thereof as target(s) for the CRISPR-Cas system.

Treating Diseases of the Liver and Kidney

[01042] The present invention also contemplates delivering the CRISPR-Cas system described herein, e.g. Type VII effector protein systems, to the liver and/or kidney. Delivery strategies to induce cellular uptake of the therapeutic nucleic acid 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 acids were addressed to renal cells with hydrodynamic hig- 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: http://www.intechopen.com/books/gene-therapy-applications/delivery-methods-to-target-rnas-inthe-kidney). Delivery methods to the kidney may include those in Yuan et al. (Am J Physiol Renal Physiol 295: F605–F617, 2008) investigated whether *in vivo* delivery of small interfering RNAs (siRNAs) targeting the 12/15-lipoxygenase (12/15-LO) pathway of arachidonate acid metabolism can ameliorate renal injury and diabetic nephropathy (DN) in a streptozotocininjected mouse model of type 1 diabetes. To achieve greater *in vivo* access and siRNA expression in the kidney, Yuan et al. used double-stranded 12/15-LO siRNA oligonucleotides conjugated with cholesterol. About 400 μg of siRNA was injected subcutaneously into mice. The method of Yuang et al. may be applied to the CRISPR Cas system of the present invention contemplating a 1-2 g subcutaneous injection of CRISPR Cas conjugated with cholesterol to a human for delivery to the kidneys.

[01043] Molitoris et al. (J Am Soc Nephrol 20: 1754–1764, 2009) exploited proximal tubule cells (PTCs), as the site of oligonucleotide reabsorption within the kidney to test the efficacy of siRNA targeted to p53, a pivotal protein in the apoptotic pathway, to prevent kidney injury. Naked synthetic siRNA to p53 injected intravenously 4 h after ischemic injury maximally protected both PTCs and kidney function. Molitoris et al.'s data indicates that rapid delivery of siRNA to proximal tubule cells follows intravenous administration. For dose-response analysis, rats were injected with doses of siP53, 0.33; 1, 3, or 5mg/kg, given at the same four time points, resulting in cumulative doses of 1.32; 4, 12, and 20 mg/kg, respectively. All siRNA doses tested produced a SCr reducing effect on day one with higher doses being effective over approximately five days compared with PBS-treated ischemic control rats. The 12 and 20

mg/kg cumulative doses provided the best protective effect. The method of Molitoris et al. may be applied to the nucleic acid-targeting system of the present invention contemplating 12 and 20 mg/kg cumulative doses to a human for delivery to the kidneys.

[01044] Thompson et al. (Nucleic Acid Therapeutics, Volume 22, Number 4, 2012) reports the toxicological and pharmacokinetic properties of the synthetic, small interfering RNA I5NP following intravenous administration in rodents and nonhuman primates. I5NP is designed to act via the RNA interference (RNAi) pathway to temporarily inhibit expression of the pro-apoptotic protein p53 and is being developed to protect cells from acute ischemia/reperfusion injuries such as acute kidney injury that can occur during major cardiac surgery and delayed graft function that can occur following renal transplantation. Doses of 800mg/kg I5NP in rodents, and 1,000 mg/kg I5NP in nonhuman primates, were required to elicit adverse effects, which in the monkey were isolated to direct effects on the blood that included a sub-clinical activation of complement and slightly increased clotting times. In the rat, no additional adverse effects were observed with a rat analogue of I5NP, indicating that the effects likely represent class effects of synthetic RNA duplexes rather than toxicity related to the intended pharmacologic activity of I5NP. Taken together, these data support clinical testing of intravenous administration of I5NP for the preservation of renal function following acute ischemia/reperfusion injury. The no observed adverse effect level (NOAEL) in the monkey was 500 mg/kg. No effects on cardiovascular, respiratory, and neurologic parameters were observed in monkeys following i.v. administration at dose levels up to 25 mg/kg. Therefore, a similar dosage may be contemplated for intravenous administration of CRISPR Cas to the kidneys of a human.

[01045] Shimizu et al. (J Am Soc Nephrol 21: 622–633, 2010) developed a system to target delivery of siRNAs to glomeruli via poly(ethylene glycol)-poly(L-lysine)-based vehicles. The siRNA/nanocarrier complex was approximately 10 to 20 nm in diameter, a size that would allow it to move across the fenestrated endothelium to access to the mesangium. After intraperitoneal injection of fluorescence-labeled siRNA/nanocarrier complexes, Shimizu et al. detected siRNAs in the blood circulation for a prolonged time. Repeated intraperitoneal administration of a mitogen-activated protein kinase 1 (MAPK1) siRNA/nanocarrier complex suppressed glomerular MAPK1 mRNA and protein expression in a mouse model of glomerulonephritis. For the investigation of siRNA accumulation, Cy5-labeled siRNAs complexed with PIC nanocarriers (0.5 ml, 5 nmol of siRNA content), naked Cy5-labeled

siRNAs (0.5 ml, 5 nmol), or Cy5-labeled siRNAs encapsulated in HVJ-E (0.5 ml, 5 nmol of siRNA content) were administrated to BALBc mice. The method of Shimizu et al. may be applied to the nucleic acid-targeting system of the present invention contemplating a dose of about of 10-20 μmol CRISPR Cas complexed with nanocarriers in about 1-2 liters to a human for intraperitoneal administration and delivery to the kidneys.

[01046] Delivery methods to the kidney are summarized in **Table 8** below:

Table 8

Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
Hydrodynamic / Lipid	TransIT *In Vivo* Gene Delivery System, DOTAP	p85α	Acute renal injury	Ischemia-reperfusion	Uptake, biodistribution	Larson et al., Surgery, (Aug 2007), Vol. 142, No. 2, pp. (262-269)
Hydrodynamic / Lipid	Lipofectamine 2000	Fas	Acute renal injury	Ischemia-reperfusion	Blood urea nitrogen, Fas Immunohistochemistry, apoptosis, histological scoring	Hamar et al., Proc Natl Acad Sci, (Oct 2004), Vol. 101, No. 41, pp. (14883-14888)
Hydrodynamic	n.a.	Apoptosis cascade elements	Acute renal injury	Ischemia-reperfusion	n.a.	Zheng et al., Am J Pathol, (Oct 2008), Vol. 173, No. 4, pp. (973–980)
Hydrodynamic	n.a.	Nuclear factor kappa-b (NFkB)	Acute renal injury	Ischemia-reperfusion	n.a.	Feng et al., Transplantation, (May 2009), Vol. 87, No. 9, pp. (1283–1289)
Hydrodynamic /Viral	Lipofectamine 2000	Apoptosis antagonizing transcription factor (AATF)	Acute renal injury	Ischemia-reperfusion	Apoptosis, oxidative stress, caspase activation, membrane lipid peroxidation	Xie & Guo, Am Soc Nephrol, (Dec 2006), Vol. 17, No. 12, pp. (3336–3346)
Hydrodynamic	pBAsi mU6 Neo/ TransIT-EE	Gremlin	Diabetic nephropathy	Streptozotozin -	Proteinuria, serum creatinine,	Q. Zhang et al., PloS ONE, (Jul 2010),

Table 8

Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
	Hydrodynamic Delivery System			induced diabetes	glomerular and tubular diameter, collagen type IV/BMP7 expression	Vol. 5, No. 7, e11709, pp. (1-13)
Viral/Lipid	pSUPER vector/Lipofectamine	TGF-β type II receptor	Interstitial renal fibrosis	Unilateral urethral obstruction	α-SMA expression, collagen content,	Kushibikia et al., J Controlled Release, (Jul 2005), Vol. 105, No. 3, pp. (318-331)
Viral	Adeno-associated virus-2	Mineral corticoid receptor	Hyper-tension caused renal damage	Cold-induced hypertension	blood pressure, serum albumin, serum urea nitrogen, serum creatinine, kidney weight, urinary sodium	Wang et al., Gene Therapy, (Jul 2006), Vol. 13, No. 14, pp. (1097-1103)
Hydrodynamic /Viral	pU6 vector	Luciferase	n.a.	n.a.	uptake	Kobayashi et al., Journal of Pharmacology and Experimental Therapeutics, (Feb 2004), Vol. 308, No. 2, pp. (688-693)
Lipid	Lipoproteins, albumin	apoB1, apoM	n.a.	n.a.	Uptake, binding affinity to lipoproteins and albumin	Wolfrum et al., Nature Biotechnology, (Sep 2007), Vol. 25, No. 10, pp. (1149-1157)

Table 8

Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
Lipid	Lipofectamine 2000	p53	Acute renal injury	Ischemic and cisplatin-induced acute injury	Histological scoring, apoptosis	Molitoris et al., J Am Soc Nephrol, (Aug 2009), Vol. 20, No. 8, pp. (1754-1764)
Lipid	DOTAP/DOPE, DOTAP/DOPE/DOPE-PEG2000	COX-2	Breast adeno-carcinoma	MDA-MB-231 breast cancer xenograft-bearing mouse	Cell viability, uptake	Mikhaylova et al., Cancer Gene Therapy, (Mar 2011), Vol. 16, No. 3, pp. (217-226)
Lipid	Cholesterol	12/15-lipoxygenase	Diabetic nephro-pathy	Streptozotocin - induced diabetes	Albuminuria, urinary creatinine, histology, type I and IV collagen, TGF-β, fibronectin, plasminogen activator inhibitor 1	Yuan et al., Am J Physiol Renal Physiol, (Jun 2008), Vol. 295, pp. (F605-F617)
Lipid	Lipofectamine 2000	Mitochondrial membrane 44 (TIM44)	Diabetic nephro-pathy	Streptozotocin - induced diabetes	Cell proliferation and apoptosis, histology, ROS, mitochondrial import of Mn-SOD and glutathione peroxidase, cellular membrane polarization	Y. Zhang et al., J Am Soc Nephrol, (Apr 2006), Vol. 17, No. 4, pp. (1090–1101)
Hydrodynamic / Lipid	Proteolipo-some	RLIP76	Renal carcinoma	Caki-2 kidney cancer xenograft-bearing mouse	uptake	Singhal et al., Cancer Res, (May 2009), Vol. 69, No. 10, pp. (4244-4251)

Table 8

Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
Polymer	PEGylated PEI	Luciferase pGL3	n.a.	n.a.	Uptake, biodistribution, erythrocyte aggregation	Malek et al., Toxicology and Applied Pharmacology, (Apr 2009), Vol. 236, No. 1, pp. (97-108)
Polymer	PEGylated poly-L-lysine	MAPK1	Lupus glomerulo- nephritis	Glomerulo-nephritis	Proteinuria, glomerulosclerosis, TGF- β, fibronectin, plasminogen activator inhibitor 1	Shimizu et al., J Am Soc Nephrology, (Apr 2010), Vol. 21, No. 4, pp. (622-633)
Polymer/Nano particle	Hyaluronic acid/ Quantum dot/ PEI	VEGF	Kidney cancer/ melanoma	B16F1 melanoma tumor-bearing mouse	Biodistribution, citotoxicity, tumor volume, endocytosis	Jiang et al., Molecular Pharmaceutics, (May-Jun 2009), Vol. 6, No. 3, pp. (727-737)
Polymer/Nano particle	PEGylated polycapro-lactone nanofiber	GAPDH	n.a.	n.a.	cell viability, uptake	Cao et al, J Controlled Release, (Jun 2010), Vol. 144, No. 2, pp. (203-212)
Aptamer	Spiegelmer mNOX-E36	CC chemokine ligand 2	Glomerulo sclerosis	Uninephrecto- mized mouse	urinary albumin, urinary creatinine, histopathology, glomerular filtration rate, macrophage count, serum Ccl2, Mac-2+, Ki-67+	Ninichuk et al., Am J Pathol, (Mar 2008), Vol. 172, No. 3, pp. (628-637)
Aptamer	Aptamer NOX-F37	vasopressin (AVP)	Congestive heart failure	n.a.	Binding affinity to D-AVP, Inhibition of AVP Signaling,	Purschke et al., Proc Natl Acad Sci, (Mar 2006), Vol. 103, No.

Table 8						
Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
					Urine osmolality and sodium concentration,	13, pp. (5173-5178)

Targeting the Liver or Liver Cells

[01047] Targeting liver cells is provided. This may be *in vitro* or *in vivo*. Hepatocytes are preferred. Delivery of the CRISPR protein, such as a Type VII effector herein may be via viral vectors, especially AAV (and in particular AAV2/6) vectors. These may be administered by intravenous injection.

[01048] A preferred target for 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 donor template) to be achieved even if only a small fraction of hepatocytes are edited.

[01049] Intron 1 of albumin has been shown by Wechsler et al. (reported at the 57th Annual Meeting and Exposition of the American Society of Hematology - abstract available online at https://ash.confex.com/ash/2015/webprogram/Paper86495.html and presented on 6th December 2015) to be a suitable target site. Their work used Zn Fingers to cut the DNA at this target site, and suitable guide sequences can be generated to guide cleavage at the same site by a CRISPR protein.

[01050] The use of targets within highly–expressed genes (genes with highly active enhancers/promoters) such as albumin may also allow a promoterless donor template to be used, as reported by Wechsler et al. and this is also broadly applicable outside liver targeting. Other examples of highly-expressed genes are known.

Other disease of the liver

[01051] In particular embodiments, the CRISPR proteins of the present invention are used in the treatment of liver disorders such as transthyretin amyloidosis (ATTR), alpha-1 antitrypsin deficiency and other hepatic-based inborn errors of metabolism. FAP is caused by a mutation in the gene that encodes transthyretin (TTR). While it is an autosomal dominant disease, not all carriers develop the disease. There are over 100 mutations in the TTR gene

known to be associated with the disease. Examples of common mutations include V30M. The principle of treatment of TTR based on gene silencing has been demonstrated by studies with iRNA (Ueda et al. 2014 Transl Neurogener. 3:19). Wilson's Disease (WD) is caused by mutations in the gene encoding ATP7B, which is found exclusively in the hepatocyte. There are over 500 mutations associated with WD, with increased prevalence in specific regions such as East Asia. Other examples are A1ATD (an autosomal recessive disease caused by mutations in the SERPINA1 gene) and PKU (an autosomal recessive disease caused by mutations in the phenylalanine hydroxylase (PAH) gene).

Liver –Associated Blood Disorders, especially Hemophilia and in particular Hemophilia B

[01052] Successful gene editing of hepatocytes has been achieved in mice (both *in vitro* and *in vivo*) and in non-human primates (*in vivo*), showing that treatment of blood disorders through gene editing/genome engineering in hepatocytes is feasible. In particular, expression of the human F9 (hF9) gene in hepatocytes has been shown in non-human primates indicating a treatment for Hemophillia B in humans.

[01053] Wechsler et al. reported at the 57th Annual Meeting and Exposition of the American Society of Hematology (abstract presented 6th December 2015 and available online at https://ash.confex.com/ash/2015/webprogram/Paper86495.html) that they has successfully expressed human F9 (hF9) from hepatocytes in non-human primates through *in vivo* gene editing. This was achieved using 1) two zinc finger nucleases (ZFNs) targeting intron 1 of the albumin locus, and 2) a human F9 donor template construct. The ZFNs and donor template were encoded on separate hepatotropic adeno-associated virus serotype 2/6 (AAV2/6) vectors injected intravenously, resulting in targeted insertion of a corrected copy of the hF9 gene into the albumin locus in a proportion of liver hepatocytes.

[01054] The albumin locus was selected as a "safe harbor" as production of this most abundant plasma protein exceeds 10 g/day, and moderate reductions in those levels are well-tolerated. Genome edited hepatocytes produced normal hFIX (hF9) in therapeutic quantities, rather than albumin, driven by the highly active albumin enhancer/promoter. Targeted integration of the hF9 transgene at the albumin locus and splicing of this gene into the albumin transcript was shown.

[01055] Mice studies: C57BL/6 mice were administered vehicle (n=20) or AAV2/6 vectors (n=25) encoding mouse surrogate reagents at 1.0 x1013 vector genome (vg)/kg via tail vein

injection. ELISA analysis of plasma hFIX in the treated mice showed peak levels of 50-1053 ng/mL that were sustained for the duration of the 6-month study. Analysis of FIX activity from mouse plasma confirmed bioactivity commensurate with expression levels.

[01056] Non-human primate (NHP) studies: a single intravenous co-infusion of AAV2/6 vectors encoding the NHP targeted albumin-specific ZFNs and a human F9 donor at 1.2x1013 vg/kg (n=5/group) resulted in >50 ng/mL (>1% of normal) in this large animal model. The use of higher AAV2/6 doses (up to 1.5x1014 vg/kg) yielded plasma hFIX levels up to 1000 ng/ml (or 20% of normal) in several animals and up to 2000 ng/ml (or 50% of normal) in a single animal, for the duration of the study (3 months).

[01057] The treatment was well tolerated in mice and NHPs, with no significant toxicological findings related to AAV2/6 ZFN + donor treatment in either species at therapeutic doses. Sangamo (CA, USA) has since applied to the FDA, and been granted, permission to conduct the world's first human clinical trial for an *in vivo* genome editing application. This follows on the back of the EMEA's approval of the Glybera gene therapy treatment of lipoprotein lipase deficiency.

[01058] Accordingly, it is preferred, in some embodiments, that any or all of the following are used: AAV (especially AAV2/6) vectors, preferably administered by intravenous injection; Albumin as target for gene editing/insertion of transgene/template– especially at intron 1 of albumin; human F9 donor template; and/ora promoterless donor template.

Hemophilia B

[01059] Accordingly, in some embodiments, it is preferred that the present invention is used to treat Hemophilia B. As such it is preferred that F9 (Factor IX) is targeted through provision of a suitable guide RNA. The enzyme and the guide may ideally be targeted to the liver where F9 is produced, although they can be delivered together or separately. A template is provided, in some embodiments, and that this is the human F9 gene. It will be appreciated that the hF9 template comprises the wt or 'correct' version of hF9 so that the treatment is effective. In some embodiments, a two-vector system may be used– one vector for the Type VII effector and one vector for the repair template(s). The repair template may include two or more repair templates, for example, two F9 sequences from different mammalian species. In some embodiments, both a mouse and human F9 sequence are provided. This is may be delivered to mice. Yang Yang, John White, McMenamin Deirdre, and Peter Bell, PhD, presenting at 58th Annual American Society of Hematology Meeting (Nov 2016), report that this increases potency and accuracy.

The second vector inserted the human sequence of factor IX into the mouse genome. In some embodiments, the targeted insertion leads to the expression of a chimeric hyperactive factor IX protein. In some embodiments, this is under the control of the native mouse factor IX promoter. Injecting this two-component system (vector 1 and vector 2) into newborn and adult "knock-out" mice at increasing doses led to expression and activity of stable factor IX activity at normal (or even higher) levels for over four months. In the case of treating humans, a native human F9 promoter may be used instead. In some embodiments, the wt phenotype is restored.

[01060] In an alternative embodiment, the hemophilia B version of F9 may be delivered so as to create a model organism, cell or cell line (for example a murine or non-human primate model organism, cell or cell line), the model organism, cell or cell line having or carrying the Hemophilia B phenotype, i.e. an inability to produce wt F9.

Hemophilia A

[01061] In some embodiments, the F9 (factor IX) gene may be replaced by the F8 (factor VIII) gene described above, leading to treatment of Hemophilia A (through provision of a correct F8 gene) and/or creation of a Hemophilia A model organism, cell or cell line (through provision of an incorrect, Hemophilia A version of the F8 gene).

Hemophilia C

[01062] In some embodiments, the F9 (factor IX) gene may be replaced by the F11 (factor XI) gene described above, leading to treatment of Hemophilia C (through provision of a correct F11 gene) and/or creation of a Hemophilia C model organism, cell or cell line (through provision of an incorrect, Hemophilia C version of the F11 gene).

Transthyretin Amyloidosis

[01063] Transthyretin is a protein, mainly produced in the liver, present in the serum and CSF which carries thyroxin hormone and retinol binding protein bound to retinol (Vitamin A). Over 120 different mutations can cause Transthyretin amyloidosis (ATTR), a heritable genetic disorder wherein mutant forms of the protein aggregate in tissues, particularly the peripheral nervous system, causing polyneuropathy. Familial amyloid polyneuropathy (FAP) is the most common TTR disorder and, in 2014, was thought to affect 47 per 100,000 people in Europe. A mutation in the TTR gene of Val30Met is thought be the most common mutation, causing an estimated 50% of FAP cases. In the absence a liver transplant, the only known cure to date, the disease is usually fatal within a decade of diagnosis. The majority of cases are monogenic.

[01064] In mouse models of ATTR, the TTR gene may be edited in a dose dependent manner by the delivery of CRISPR/Cas9. In some embodiments, the Type VII effector is provided as mRNA. In some embodiments, Type VII effector mRNA and guide RNA are packaged in LNPs. A system comprising Type VII effector mRNA and guide RNA packaged in LNPs achieved up to 60% editing efficiency in the liver, with serum TTR levels being reduced by up to 80%. In some embodiments, therefore, Transthyretin is targeted, in particular correcting for the Val30Met mutation. In some embodiments, therefore, ATTR is treated.

Alpha-1 Antitrypsin Deficiency

[01065] Alpha-1 Antitrypsin (A1AT) is a protein produced in the liver which primarily functions to decrease the activity of neutrophil elastase, an enzyme which degrades connective tissue, in the lungs. Alpha-1 Antitrypsin Deficiency (ATTD) is a disease caused by mutation of the SERPINA1 gene, which encodes A1AT. Impaired production of A1AT leads to a gradual degradation of the connective tissue of the lung resulting in emphysema like symptoms.

[01066] Several mutations can cause ATTD, though the most common mutations are Glu342Lys (referred to as Z allele, wild-type is referred to as M) or Glu264Val (referred to as the S allele), and each allele contributes equally to the disease state, with two affected alleles resulting in more pronounced pathophysiology. These results not only resulted in degradation of the connective tissue of sensitive organs, such as the lung, but accumulation of the mutants in the liver can result in proteotoxicity. Current treatments focus on the replacement of A1AT by injection of protein retrieved from donated human plasma. In severe cases a lung and/or liver transplant may be considered.

[01067] The common variants of the disease are again monogenic. In some embodiments, the SERPINA1 gene is targeted. In some embodiments, the Glu342Lys mutation (referred to as Z allele, wild-type is referred to as M) or the Glu264Val mutation (referred to as the S allele) are corrected for. In some embodiments, therefore, the faulty gene would require replacement by the wild-type functioning gene. In some embodiments, a knockout and repair approach is required, so a repair template is provided. In the case of bi-allelic mutations, in some embodiments only one guide RNA would be required for homozygous mutations, but in the case of heterozygous mutations two guide RNAs may be required. Delivery is, in some embodiments, to the lung or liver.

Inborn errors of metabolism

[01068] Inborn errors of metabolism (IEMs) are an umbrella group of diseases which affect metabolic processes. In some embodiments, an IEM is to be treated. The majority of these diseases are monogenic in nature (e.g. phenylketonuria) and the pathophysiology results from either the abnormal accumulation of substances which are inherently toxic, or mutations which result in an inability to synthesize essential substances. Depending on the nature of the IEM, CRISPR/Type VII effector may be used to facilitate a knock-out alone, or in combination with replacement of a faulty gene via a repair template. Exemplary diseases that may benefit from CRISPR/Type VII effector technology are, in some embodiments: primary hyperoxaluria type 1 (PH1), argininosuccinic lyase deficiency, ornithine transcarbamylase deficiency, phenylketonuria, or PKU, and maple syrup urine disease.

Treating Epithelial and Lung Diseases

[01069] The present invention also contemplates delivering the CRISPR-Cas system described herein, e.g. Type VII effector protein systems, to one or both lungs.

[01070] Although AAV-2-based vectors were originally proposed for CFTR delivery to CF airways, other serotypes such as AAV-1, AAV-5, AAV-6, and AAV-9 exhibit improved gene transfer efficiency in a variety of models of the lung epithelium (see, e.g., Li et al., Molecular Therapy, vol. 17 no. 12, 2067-2077 Dec 2009). AAV-1 was demonstrated to be ~100-fold more efficient than AAV-2 and AAV-5 at transducing human airway epithelial cells *in vitro*,5 although AAV-1 transduced murine tracheal airway epithelia *in vivo* with an efficiency equal to that of AAV-5. Other studies have shown that AAV-5 is 50-fold more efficient than AAV-2 at gene delivery to human airway epithelium (HAE) *in vitro* and significantly more efficient in the mouse lung airway epithelium *in vivo*. AAV-6 has also been shown to be more efficient than AAV-2 in human airway epithelial cells *in vitro* and murine airways *in vivo*. The more recent isolate, AAV-9, was shown to display greater gene transfer efficiency than AAV-5 in murine nasal and alveolar epithelia *in vivo* with gene expression detected for over 9 months suggesting AAV may enable long-term gene expression *in vivo*, a desirable property for a CFTR gene delivery vector. Furthermore, it was demonstrated that AAV-9 could be readministered to the murine lung with no loss of CFTR expression and minimal immune consequences. CF and non- CF HAE cultures may be inoculated on the apical surface with 100 µl of AAV vectors for hours (see, e.g., Li et al., Molecular Therapy, vol. 17 no. 12, 2067-2077 Dec 2009). The MOI may vary from 1 × 103 to 4 × 105 vector genomes/cell, depending on

virus concentration and purposes of the experiments. The above cited vectors are contemplated for the delivery and/or administration of the invention.

[01071] Zamora et al. (Am J Respir Crit Care Med Vol 183. pp 531–538, 2011) reported an example of the application of an RNA interference therapeutic to the treatment of human infectious disease and also a randomized trial of an antiviral drug in respiratory syncytial virus (RSV)-infected lung transplant recipients. Zamora et al. performed a randomized, double-blind, placebo-controlled trial in LTX recipients with RSV respiratory tract infection. Patients were permitted to receive standard of care for RSV. Aerosolized ALN-RSV01 (0.6 mg/kg) or placebo was administered daily for 3 days. This study demonstrates that an RNAi therapeutic targeting RSV can be safely administered to LTX recipients with RSV infection. Three daily doses of ALN-RSV01 did not result in any exacerbation of respiratory tract symptoms or impairment of lung function and did not exhibit any systemic proinflammatory effects, such as induction of cytokines or CRP. Pharmacokinetics showed only low, transient systemic exposure after inhalation, consistent with preclinical animal data showing that ALN-RSV01, administered intravenously or by inhalation, is rapidly cleared from the circulation through exonuclease-mediated digestion and renal excretion. The method of Zamora et al. may be applied to the nucleic acid-targeting system of the present invention and an aerosolized CRISPR Cas, for example with a dosage of 0.6 mg/kg, may be contemplated for the present invention.

[01072] 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 Cas, U6 or H1 promoter for guide RNA),: A preferred arrangement is to use a CFTRdelta508 targeting guide, a repair template for deltaF508 mutation and a codon optimized Type VII enzyme, with optionally one or more nuclear localization signal or sequence(s) (NLS(s)), e.g., two (2) NLSs. Constructs without NLS are also envisaged.

Treating Diseases of the Muscular System

[01073] The present invention also contemplates delivering the CRISPR-Cas system described herein, e.g. Type VII effector protein systems, to muscle(s).

[01074] Bortolanza et al. (Molecular Therapy vol. 19 no. 11, 2055–2064 Nov. 2011) shows that systemic delivery of RNA interference expression cassettes in the FRG1 mouse, after the onset of facioscapulohumeral muscular dystrophy (FSHD), led to a dose-dependent long-term FRG1 knockdown without signs of toxicity. Bortolanza et al. found that a single intravenous injection of 5 × 1012 vg of rAAV6-sh1FRG1 rescues muscle histopathology and muscle function of FRG1 mice. In detail, 200 µl containing 2 × 1012 or 5 × 1012 vg of vector in physiological solution were injected into the tail vein using a 25-gauge Terumo syringe. The method of Bortolanza et al. may be applied to an AAV expressing CRISPR Cas and injected into humans at a dosage of about 2 × 1015 or 2 × 1016 vg of vector.

[01075] Dumonceaux et al. (Molecular Therapy vol. 18 no. 5, 881–887 May 2010) inhibit the myostatin pathway using the technique of RNA interference directed against the myostatin receptor AcvRIIb mRNA (sh-AcvRIIb). The restoration of a quasi-dystrophin was mediated by the vectorized U7 exon-skipping technique (U7-DYS). Adeno-associated vectors carrying either the sh-AcvrIIb construct alone, the U7-DYS construct alone, or a combination of both constructs were injected in the tibialis anterior (TA) muscle of dystrophic mdx mice. The injections were performed with 1011 AAV viral genomes. The method of Dumonceaux et al. may be applied to an AAV expressing CRISPR Cas and injected into humans, for example, at a dosage of about 1014 to about 1015 vg of vector.

[01076] Kinouchi et al. (Gene Therapy (2008) 15, 1126–1130) report the effectiveness of *in vivo* siRNA delivery into skeletal muscles of normal or diseased mice through nanoparticle formation of chemically unmodified siRNAs with atelocollagen (ATCOL). ATCOL-mediated local application of siRNA targeting myostatin, a negative regulator of skeletal muscle growth, in mouse skeletal muscles or intravenously, caused a marked increase in the muscle mass within a few weeks after application. These results imply that ATCOL-mediated application of siRNAs is a powerful tool for future therapeutic use for diseases including muscular atrophy. MstsiRNAs (final concentration, 10 mM) were mixed with ATCOL (final concentration for local administration, 0.5%) (AteloGene, Kohken, Tokyo, Japan) according to the manufacturer's instructions. After anesthesia of mice (20-week-old male C57BL/6) by Nembutal (25 mg/kg, i.p.), the Mst-siRNA/ATCOL complex was injected into the masseter and biceps femoris muscles. The method of Kinouchi et al. may be applied to CRISPR Cas and injected into a human, for example, at a dosage of about 500 to 1000 ml of a 40 µM solution into the muscle. Hagstrom et al. (Molecular Therapy Vol. 10, No. 2, August 2004) describe an

intravascular, nonviral methodology that enables efficient and repeatable delivery of nucleic acids to muscle cells (myofibers) throughout the limb muscles of mammals. The procedure involves the injection of naked plasmid DNA or siRNA into a distal vein of a limb that is transiently isolated by a tourniquet or blood pressure cuff. Nucleic acid delivery to myofibers is facilitated by its rapid injection in sufficient volume to enable extravasation of the nucleic acid solution into muscle tissue. High levels of transgene expression in skeletal muscle were achieved in both small and large animals with minimal toxicity. Evidence of siRNA delivery to limb muscle was also obtained. For plasmid DNA intravenous injection into a rhesus monkey, a three-way stopcock was connected to two syringe pumps (Model PHD 2000; Harvard Instruments), each loaded with a single syringe. Five minutes after a papaverine injection, pDNA (15.5 to 25.7 mg in 40 –100 ml saline) was injected at a rate of 1.7 or 2.0 ml/s. This could be scaled up for plasmid DNA expressing CRISPR Cas of the present invention with an injection of about 300 to 500 mg in 800 to 2000 ml saline for a human. For adenoviral vector injections into a rat, 2 x 109 infectious particles were injected in 3 ml of normal saline solution (NSS). This could be scaled up for an adenoviral vector expressing CRISPR Cas of the present invention with an injection of about 1 x 1013 infectious particles were injected in 10 liters of NSS for a human. For siRNA, a rat was injected into the great saphenous vein with 12.5 μg of a siRNA and a primate was injected into the great saphenous vein with 750 μg of a siRNA. This could be scaled up for a CRISPR Cas of the present invention, for example, with an injection of about 15 to about 50 mg into the great saphenous vein of a human.

[01077] See also, for example, WO2013163628 A2, Genetic Correction of Mutated Genes, published application of Duke University describes efforts to correct, for example, a frameshift mutation which causes a premature stop codon and a truncated gene product that can be corrected via nuclease mediated non-homologous end joining such as those responsible for Duchenne Muscular Dystrophy, ("DMD") a recessive, fatal, X-linked disorder that results in muscle degeneration due to mutations in the dystrophin gene. The majority of dystrophin mutations that cause DMD are deletions of exons that disrupt the reading frame and cause premature translation termination in the dystrophin gene. Dystrophin is a cytoplasmic protein that provides structural stability to the dystroglycan complex of the cell membrane that is responsible for regulating muscle cell integrity and function. The dystrophin gene or "DMD gene" as used interchangeably herein is 2.2 megabases at locus Xp21. The primary transcription measures about 2,400 kb with the mature mRNA being about 14 kb. 79 exons

code for the protein which is over 3500 amino acids. Exon 51 is frequently adjacent to frame-disrupting deletions in DMD patients and has been targeted in clinical trials for oligonucleotide-based exon skipping. A clinical trial for the exon 51 skipping compound eteplirsen recently reported a significant functional benefit across 48 weeks, with an average of 47% dystrophin positive fibers compared to baseline. Mutations in exon 51 are ideally suited for permanent correction by NHEJ-based genome editing.

[01078] The methods of US Patent Publication No. 20130145487 assigned to Cellectis, which relates to meganuclease variants to cleave a target sequence from the human dystrophin gene (DMD), may also be modified to for the nucleic acid-targeting system of the present invention.

Treating Diseases of the Skin

[01079] The present invention also contemplates delivering the CRISPR-Cas system described herein, e.g. Type VII effector protein systems, to the skin.

[01080] Hickerson et al. (Molecular Therapy—Nucleic Acids (2013) 2, e129) relates to a motorized microneedle array skin delivery device for delivering self-delivery (sd)-siRNA to human and murine skin. The primary challenge to translating siRNA-based skin therapeutics to the clinic is the development of effective delivery systems. Substantial effort has been invested in a variety of skin delivery technologies with limited success. In a clinical study in which skin was treated with siRNA, the exquisite pain associated with the hypodermic needle injection precluded enrollment of additional patients in the trial, highlighting the need for improved, more "patient-friendly" (i.e., little or no pain) delivery approaches. Microneedles represent an efficient way to deliver large charged cargos including siRNAs across the primary barrier, the stratum corneum, and are generally regarded as less painful than conventional hypodermic needles. Motorized "stamp type" microneedle devices, including the motorized microneedle array (MMNA) device used by Hickerson et al., have been shown to be safe in hairless mice studies and cause little or no pain as evidenced by (i) widespread use in the cosmetic industry and (ii) limited testing in which nearly all volunteers found use of the device to be much less painful than a flushot, suggesting siRNA delivery using this device will result in much less pain than was experienced in the previous clinical trial using hypodermic needle injections. The MMNA device (marketed as Triple-M or Tri-M by Bomtech Electronic Co, Seoul, South Korea) was adapted for delivery of siRNA to mouse and human skin. sd-siRNA solution (up to 300 μl of 0.1 mg/ml RNA) was introduced into the chamber of the disposable

Tri-M needle cartridge (Bomtech), which was set to a depth of 0.1 mm. For treating human skin, deidentified skin (obtained immediately following surgical procedures) was manually stretched and pinned to a cork platform before treatment. All intradermal injections were performed using an insulin syringe with a 28-gauge 0.5-inch needle. The MMNA device and method of Hickerson et al. could be used and/or adapted to deliver the CRISPR Cas of the present invention, for example, at a dosage of up to 300 µl of 0.1 mg/ml CRISPR Cas to the skin.

[01081] Leachman et al. (Molecular Therapy, vol. 18 no. 2, 442–446 Feb. 2010) relates to a phase Ib clinical trial for treatment of a rare skin disorder pachyonychia congenita (PC), an autosomal dominant syndrome that includes a disabling plantar keratoderma, utilizing the first short-interfering RNA (siRNA)-based therapeutic for skin. This siRNA, called TD101, specifically and potently targets the keratin 6a (K6a) N171K mutant mRNA without affecting wild-type K6a mRNA.

[01082] Zheng et al. (PNAS, July 24, 2012, vol. 109, no. 30, 11975–11980) show that spherical nucleic acid nanoparticle conjugates (SNA-NCs), gold cores surrounded by a dense shell of highly oriented, covalently immobilized siRNA, freely penetrate almost 100% of keratinocytes *in vitro*, mouse skin, and human epidermis within hours after application. Zheng et al. demonstrated that a single application of 25 nM epidermal growth factor receptor (EGFR) SNA-NCs for 60 h demonstrate effective gene knockdown in human skin. A similar dosage may be contemplated for CRISPR Cas immobilized in SNA-NCs for administration to the skin.

Cancer

[01083] In some embodiments, the treatment, prophylaxis or diagnosis of cancer is provided. The target is preferably one or more of the FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, TRAC or TRBC genes. The cancer may be one or more of lymphoma, chronic lymphocytic leukemia (CLL), B cell acute lymphocytic leukemia (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. This may be implemented with engineered chimeric antigen receptor (CAR) T cell. This is described in WO2015161276, the disclosure of which is hereby incorporated by reference and described herein below.

[01084] Target genes suitable for the treatment or prophylaxis of cancer may include, in some embodiments, those described in WO2015048577 the disclosure of which is hereby incorporated by reference.

Usher Syndrome or retinitis pigmentosa-39

[01085] In some embodiments, the treatment, prophylaxis or diagnosis of Usher Syndrome or retinitis pigmentosa-39 is provided. The target is preferably the USH2A gene. In some embodiments, correction of a G deletion at position 2299 (2299delG) is provided. This is described in WO2015134812A1, the disclosure of which is hereby incorporated by reference.

Autoimmune and inflammatory disorders

[01086] In some embodiments, autoimmune and inflammatory disorders are treated. These include Multiple Sclerosis (MS) or Rheumatoid Arthritis (RA), for example.

Cystic Fibrosis (CF)

[01087] In some embodiments, the treatment, prophylaxis or diagnosis of cystic fibrosis is provided. The target is preferably the SCNN1A or the CFTR gene. This is described in WO2015157070, the disclosure of which is hereby incorporated by reference.

[01088] Schwank et al. (Cell Stem Cell, 13:653–58, 2013) used CRISPR-Cas9 to correct a defect associated with cystic fibrosis in human stem cells. The team's target was the gene for an ion channel, cystic fibrosis transmembrane conductor receptor (CFTR). A deletion in CFTR causes the protein to misfold in cystic fibrosis patients. Using cultured intestinal stem cells developed from cell samples from two children with cystic fibrosis, Schwank et al. were able to correct the defect using CRISPR along with a donor plasmid containing the reparative sequence to be inserted. The researchers then grew the cells into intestinal "organoids," or miniature guts, and showed that they functioned normally. In this case, about half of clonal organoids underwent the proper genetic correction.

[01089] In some embodiments, Cystic fibrosis is treated, for example. Delivery to the lungs is therefore preferred. The F508 mutation (delta-F508, full name CFTRΔF508 or F508del-CFTR) is preferably corrected. In some embodiments, the targets may be ABCC7, CF or MRP7.

Duchenne's muscular dystrophy

[01090] Duchenne's muscular dystrophy (DMD) is a recessive, sex-linked muscle wasting disease that affects approximately 1 in 5000 males at birth. Mutations of the dystrophin gene result in an absence of dystrophin in skeletal muscle, where it normally functions to connect

the cytoskeleton of the muscle fiber to the basal lamina. The absence of dystrophin caused be these mutations results in excessive calcium entry into the soma which causes the mitochondria to rupture, destroying the cell. Current treatments are focused on easing the symptoms of DMD, and the average life expectancy is approximately 26 years.

[01091] CRISPR/Cas9 efficacy as a treatment for certain types of DMD has been demonstrated in mouse models. In one such study, the muscular dystrophy phenotype was partially corrected in the mouse by knocking-out a mutant exon resulting in a functional protein (see Nelson et al. (2016) Science, Long et al. (2016) Science, and Tabebordbar et al. (2016) Science).

[01092] In some embodiments, DMD is treated. In some embodiments, delivery is to the muscle by injection.

Glycogen Storage Diseases, including 1a

[01093] Glycogen Storage Disease 1a is a genetic disease resulting from deficiency of the enzyme glucose-6-phosphatase. The deficiency impairs the ability of the liver to produce free glucose from glycogen and from gluconeogenesis. In some embodiments, the gene encoding the glucose-6-phosphatase enzyme is targeted. In some embodiments, Glycogen Storage Disease 1a is treated. In some embodiments, delivery is to the liver by encapsulation of the Type VII effector (in protein or mRNA form) in a lipid particle, such as an LNP.

[01094] In some embodiments, Glycogen Storage Diseases, including 1a, are targeted and preferably treated, for example by targeting polynucleotides associated with the condition/disease/infection. The associated polynucleotides include DNA, which may include genes (where genes include any coding sequence and regulatory elements such as enhancers or promoters). In some embodiments, the associated polynucleotides may include the SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, or PFKM genes.

Hurler Syndrome

[01095] Hurler syndrome, also known as mucopolysaccharidosis type I (MPS I), Hurler's disease, is a genetic disorder that results in the buildup of glycosaminoglycans (formerly known as mucopolysaccharides) due to a deficiency of alpha-L iduronidase, an enzyme responsible for the degradation of mucopolysaccharides in lysosomes. Hurler syndrome is often classified as a lysosomal storage disease, and is clinically related to Hunter Syndrome. Hunter syndrome is X-linked while Hurler syndrome is autosomal recessive. MPS I is divided into three subtypes

based on severity of symptoms. All three types result from an absence of, or insufficient levels of, the enzyme α-L-iduronidase. MPS I H or Hurler syndrome is the most severe of the MPS I subtypes. The other two types are MPS I S or Scheie syndrome and MPS I H-S or Hurler-Scheie syndrome. Children born to an MPS I parent carry a defective IDUA gene, which has been mapped to the 4p16.3 site on chromosome 4. The gene is named IDUA because of its iduronidase enzyme protein product. As of 2001, 52 different mutations in the IDUA gene have been shown to cause Hurler syndrome. Successful treatment of the mouse, dog, and cat models of MPS I by delivery of the iduronidase gene through retroviral, lentiviral, AAV, and even nonviral vectors.

[01096] In some embodiments, the α-L-iduronidase gene is targeted and a repair template preferably provided.

HIV and AIDS

[01097] In some embodiments, the treatment, prophylaxis or diagnosis of HIV and AIDS is provided. The target is preferably the CCR5 gene in HIV. This is described in WO2015148670A1, the disclosure of which is hereby incorporated by reference.

Beta Thalassaemia

[01098] In some embodiments, the treatment, prophylaxis or diagnosis of Beta Thalassaemia is provided. The target is preferably the BCL11A gene. This is described in WO2015148860, the disclosure of which is hereby incorporated by reference.

Sickle Cell Disease (SCD)

[01099] In some embodiments, the treatment, prophylaxis or diagnosis of Sickle Cell Disease (SCD) is provided. The target is preferably the HBB or BCL11A gene. This is described in WO2015148863, the disclosure of which is hereby incorporated by reference.

Herpes Simplex Virus 1 and 2

[01100] Herpesviridae are a family of viruses composed of linear double-stranded DNA genomes with 75-200 genes. For the purposes of gene editing, the most commonly studied family member is Herpes Simplex Virus – 1 (HSV-1), a virus which has a distinct number of advantages over other viral vectors (reviewed in Vannuci et al. (2003)). Thus, in some embodiments, the viral vector is an HSV viral vector. In some embodiments, the HSV viral vector is HSV-1.

[01101] HSV-1 has a large genome of approximately 152 kb of double stranded DNA. This genome comprises of more than 80 genes, many of which can be replaced or removed, allowing

a gene insert of between 30-150 kb. The viral vectors derived from HSV-1 are generally separated into 3 groups: replication-competent attenuated vectors, replication-incompetent recombinant vectors, and defective helper-dependent vectors known as amplicons. Gene transfer using HSV-1 as a vector has been demonstrated previously, for instance for the treatment of neuropathic pain (see, e.g., Wolfe et al. (2009) Gene Ther) and rheumatoid arthritis (see e.g., Burton et al. (2001) Stem Cells).

[01102] Thus, in some embodiments, the viral vector is an HSV viral vector. In some embodiments, the HSV viral vector is HSV-1. In some embodiments, the vector is used for delivery of one or more CRISPR components. It may be particularly useful for delivery of the Type VII effector and one or more guide RNAs, for example 2 or more, 3 or more, or 4 or more guide RNAs. In some embodiments, the vector is therefore useful in a multiplex system. In some embodiments, this delivery is for the treatment of treatment of neuropathic pain or rheumatoid arthritis.

[01103] In some embodiments, the treatment, prophylaxis or diagnosis of HSV-1 (Herpes Simplex Virus 1) is provided. The target is preferably the UL19, UL30, UL48 or UL50 gene in HSV-1. This is described in WO2015153789, the disclosure of which is hereby incorporated by reference.

[01104] In other embodiments, the treatment, prophylaxis or diagnosis of HSV-2 (Herpes Simplex Virus 2) is provided. The target is preferably the UL19, UL30, UL48 or UL50 gene in HSV-2. This is described in WO2015153791, the disclosure of which is hereby incorporated by reference.

[01105] In some embodiments, the treatment, prophylaxis or diagnosis of Primary Open Angle Glaucoma (POAG) is provided. The target is preferably the MYOC gene. This is described in WO2015153780, the disclosure of which is hereby incorporated by reference.

Cell-Based Therapies

[01106] Also described herein are cell-based therapies where a modified cell, such as any of those described in greater detail elsewhere herein, are administered to a subject in need thereof. In some embodiments, the cell-based therapy when administered can treat or prevent a disease or a symptom thereof in the subject. In some embodiments, the subject is a human. In some embodiments, the subject is a non-human animal. In some embodiments, the subject is a non-animal organism, such as a plant. Cell-based therapies can be autologous, allogeneic, or xenogeneic.

Adoptive Cell Therapies

[01107] The present invention also contemplates use of the CRISPR-Cas system described herein, e.g. Type VII effector protein systems, to modify cells for adoptive therapies. Aspects of the invention accordingly involve the adoptive transfer of immune system cells, such as T cells, specific for selected antigens, such as tumor associated antigens (see 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; and, 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). 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).

[01108] 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 WO9215322). 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 CD8α hinge domain and a CD8α transmembrane domain, to the transmembrane and intracellular signaling domains of either CD3ζ or FcRγ (scFv-CD3ζ or scFv-FcRγ; 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-CD3ζ; 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 CD3ζ-chain, CD97, GDI la-CD18, CD2, ICOS, CD27, CD154, CDS, OX40, 4-1BB, or CD28 signaling domains (for example scFv-CD28-4-1BB-CD3ζ or scFv-CD28-OX40-CD3ζ; 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. WO2014134165; PCT Publication No. WO2012079000). 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 αβTCR, 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.

[01109] 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 CD3ζ and either CD28 or CD137. Viral vectors may for example include vectors based on HIV, SV40, EBV, HSV or BPV.

[01110] 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 threat tumor xenografts.

[01111] 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). 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.

[01112] In one embodiment, the treatment can be administrated into patients undergoing an immunosuppressive 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. Not being bound by a theory, the immunosuppressive treatment should help the selection and expansion of the immunoresponsive or T cells according to the invention within the patient.

[01113] The administration of the cells or population of cells according to the present invention 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, by intravenous or intralymphatic injection, or intraperitoneally. In one embodiment, the cell compositions of the present invention are preferably administered by intravenous injection.

[01114] 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 administrated in one or more doses. In another embodiment, the effective amount of cells are administrated as a single dose. In another embodiment, the effective amount of cells are administrated 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 administrated 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.

[01115] In another embodiment, the effective amount of cells or composition comprising those cells are administrated parenterally. The administration can be an intravenous administration. The administration can be directly done by injection within a tumor.

[01116] 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; PCT Patent Publication WO2011146862; PCT Patent Publication WO2014011987; PCT Patent Publication WO2013040371; 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)).

[01117] In a further refinement of adoptive therapies, genome editing with a CRISPR-Cas system as described herein 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). For example, immunoresponsive cells may be edited to delete expression of some or all of the class of HLA type II and/or type I molecules, or to knockout selected genes that may inhibit the desired immune response, such as the PD1 gene.

[01118] Cells may be edited using any CRISPR system and method of use thereof as described herein. CRISPR systems may be delivered to an immune cell by any method described 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 to eliminate potential alloreactive T-cell receptors (TCR), disrupt the target of a chemotherapeutic agent, block an immune checkpoint, activate a T cell, and/or increase the differentiation and/or proliferation of functionally exhausted or dysfunctional CD8+ T-cells (see PCT Patent Publications: WO2013176915, WO2014059173, WO2014172606, WO2014184744, and WO2014191128). Editing may result in inactivation of a gene.

[01119] 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 CRISPR 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.

[01120] 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 TCRα or TCRβ 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.

[01121] 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.

[01122] Immune checkpoints are inhibitory pathways that slow down or stop immune reactions and prevent excessive tissue damage from uncontrolled activity of immune cells. In certain embodiments, 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.

[01123] 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).

[01124] WO2014172606 relates to the use of MT1 and/or MT1 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 certain embodiments, metallothioneins are targeted by gene editing in adoptively transferred T cells.

[01125] In certain embodiments, 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, TGFBRII, TGFRBRI, 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 or TIM-3. 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.

[01126] In other embodiments, at least two genes are edited. Pairs of genes may include, but are not limited to PD1 and TCRα, PD1 and TCRβ, CTLA-4 and TCRα, CTLA-4 and TCRβ, LAG3 and TCRα, LAG3 and TCRβ, Tim3 and TCRα, Tim3 and TCRβ, BTLA and TCRα,

BTLA and TCRβ, BY55 and TCRα, BY55 and TCRβ, TIGIT and TCRα, TIGIT and TCRβ, B7H5 and TCRα, B7H5 and TCRβ, LAIR1 and TCRα, LAIR1 and TCRβ, SIGLEC10 and TCRα, SIGLEC10 and TCRβ, 2B4 and TCRα, 2B4 and TCRβ.

[01127] 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. Patents 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*.

[01128] The practice of the present invention employs, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See 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.); and ANIMAL CELL CULTURE (1987) (R.I. Freshney, ed.).

[01129] The practice of the present invention employs, unless otherwise indicated, conventional techniques for generation of genetically modified mice. See Marten H. Hofker and Jan van Deursen, TRANSGENIC MOUSE METHODS AND PROTOCOLS, 2nd edition (2011).

[01130] In some embodiments, the invention described herein relates to a method for adoptive immunotherapy, in which T cells are edited *ex vivo* by CRISPR to modulate at least one gene and subsequently administered to a patient in need thereof. In some embodiments, the CRISPR editing comprising knocking-out or knocking-down the expression of at least one target gene in the edited T cells. In some embodiments, in addition to modulating the target gene, the T cells are also edited *ex vivo* by CRISPR to (1) knock-in an exogenous gene encoding a chimeric antigen receptor (CAR) or a T-cell receptor (TCR), (2) knock-out or knock-down expression of an immune checkpoint receptor, (3) knock-out or knock-down expression of an

endogenous TCR, (4) knock-out or knock-down expression of a human leukocyte antigen class I (HLA-I) proteins, and/or (5) knock-out or knock-down expression of an endogenous gene encoding an antigen targeted by an exogenous CAR or TCR.

[01131] In some embodiments, the T cells are contacted *ex vivo* with an adeno-associated virus (AAV) vector encoding a CRISPR effector protein, and a guide molecule comprising a guide sequence hybridizable to a target sequence, a tracr mate sequence, and a tracr sequence hybridizable to the tracr mate sequence. In some embodiments, the T cells are contacted *ex vivo* (e.g., by electroporation) with a ribonucleoprotein (RNP) comprising a CRISPR effector protein complexed with a guide molecule, wherein the guide molecule comprising a guide sequence hybridizable to a target sequence, a tracr mate sequence, and a tracr sequence hybridizable to the tracr mate sequence. See Rupp et al., Scientific Reports 7:737 (2017); Liu et al., Cell Research 27:154-157 (2017). In some embodiments, the T cells are contacted *ex vivo* (e.g., by electroporation) with an mRNA encoding a CRISPR effector protein, and a guide molecule comprising a guide sequence hybridizable to a target sequence, a tracr mate sequence, and a tracr sequence hybridizable to the tracr mate sequence. See Eyquem et al., Nature 543:113-117 (2017). In some embodiments, the T cells are not contacted *ex vivo* with a lentivirus or retrovirus vector.

[01132] In some embodiments, the method comprises editing T cells *ex vivo* by CRISPR to knock-in an exogenous gene encoding a CAR, thereby allowing the edited T cells to recognize cancer cells based on the expression of specific proteins located on the cell surface. In some embodiments, T cells are edited *ex vivo* by CRISPR to knock-in an exogenous gene encoding a TCR, thereby allowing the edited T cells to recognize proteins derived from either the surface or inside of the cancer cells. In some embodiments, the method comprising providing an exogenous CAR-encoding or TCR-encoding sequence as a donor sequence, which can be integrated by homology-directed repair (HDR) into a genomic locus targeted by a CRISPR guide sequence. In some embodiments, targeting the exogenous CAR or TCR to an endogenous TCR α constant (TRAC) locus can reduce tonic CAR signaling and facilitate effective internalization and re-expression of the CAR following single or repeated exposure to antigen, thereby delaying effector T-cell differentiation and exhaustion. See Eyquem et al., Nature 543:113-117 (2017).

[01133] In some embodiments, the method comprises editing T cells *ex vivo* by CRISPR to block one or more immune checkpoint receptors to reduce immunosuppression by cancer cells.

In some embodiments, T cells are edited *ex vivo* by CRISPR to knock-out or knock-down an endogenous gene involved in the programmed death-1 (PD-1) signaling pathway, such as PD-1 and PD-L1. In some embodiments, T cells are edited *ex vivo* by CRISPR to mutate the Pdcd1 locus or the CD274 locus. In some embodiments, T cells are edited *ex vivo* by CRISPR using one or more guide sequences targeting the first exon of PD-1. See Rupp et al., Scientific Reports 7:737 (2017); Liu et al., Cell Research 27:154-157 (2017).

[01134] In some embodiments, the method comprises editing T cells *ex vivo* by CRISPR to eliminate potential alloreactive TCRs to allow allogeneic adoptive transfer. In some embodiments, T cells are edited *ex vivo* by CRISPR to knock-out or knock-down an endogenous gene encoding a TCR (e.g., an αβ TCR) to avoid graft-versus-host-disease (GVHD). In some embodiments, T cells are edited *ex vivo* by CRISPR to mutate the TRAC locus. In some embodiments, T cells are edited *ex vivo* by CRISPR using one or more guide sequences targeting the first exon of TRAC. See Liu et al., Cell Research 27:154-157 (2017). In some embodiments, the method comprises use of CRISPR to knock-in an exogenous gene encoding a CAR or a TCR into the TRAC locus, while simultaneously knocking-out the endogenous TCR (e.g., with a donor sequence encoding a self-cleaving P2A peptide following the CAR cDNA). See Eyquem et al., Nature 543:113-117 (2017). In some embodiments, the exogenous gene comprises a promoter-less CAR-encoding or TCR-encoding sequence which is inserted operably downstream of an endogenous TCR promoter.

[01135] In some embodiments, the method comprises editing T cells *ex vivo* by CRISPR to knock-out or knock-down an endogenous gene encoding an HLA-I protein to minimize immunogenicity of the edited T cells. In some embodiments, T cells are edited *ex vivo* by CRISPR to mutate the beta-2 microglobulin (B2M) locus. In some embodiments, T cells are edited *ex vivo* by CRISPR using one or more guide sequences targeting the first exon of B2M. See Liu et al., Cell Research 27:154-157 (2017). In some embodiments, the method comprises use of CRISPR to knock-in an exogenous gene encoding a CAR or a TCR into the B2M locus, while simultaneously knocking-out the endogenous B2M (e.g., with a donor sequence encoding a self-cleaving P2A peptide following the CAR cDNA). See Eyquem et al., Nature 543:113-117 (2017). In some embodiments, the exogenous gene comprises a promoter-less CAR-encoding or TCR-encoding sequence which is inserted operably downstream of an endogenous B2M promoter.

[01136] In some embodiments, the method comprises editing T cells *ex vivo* by CRISPR to knock-out or knock-down an endogenous gene encoding an antigen targeted by an exogenous CAR or TCR. In some embodiments, the T cells are edited *ex vivo* by CRISPR to knock-out or knock-down the expression of a tumor antigen selected from 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 or cyclin (DI) (see WO2016/011210). In some embodiments, the T cells are edited *ex vivo* by CRISPR to knock-out or knock-down the expression of an antigen selected from B cell maturation antigen (BCMA), transmembrane activator and CAML Interactor (TACI), or B-cell activating factor receptor (BAFF-R), CD38, CD138, CS-1, CD33, CD26, CD30, CD53, CD92, CD100, CD148, CD150, CD200, CD261, CD262, or CD362 (see WO2017/011804).

Gene Drives

[01137] The present invention also contemplates use of the CRISPR-Cas system described herein, e.g. Type VII effector protein systems, 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 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).

Xenotransplantation

[01138] The present invention also contemplates use of the CRISPR-Cas system described herein, e.g. Type VII effector protein systems, to provide RNA-guided DNA nucleases adapted to be used to provide modified tissues for transplantation. 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 α(l,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.

General Gene Therapy Considerations

[01139] Examples of disease-associated genes and polynucleotides and disease specific information is available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web.

[01140] Mutations in these genes and pathways can result in production of improper proteins or proteins in improper amounts which affect function. Further examples of genes, diseases and proteins are hereby incorporated by reference from US Provisional Application No. 61/736,527 filed December 12, 2012. Such genes, proteins and pathways may be the target polynucleotide of a CRISPR complex of the present invention. Examples of disease-associated genes and polynucleotides are listed in **Tables 9 and 10**. Examples of signaling biochemical pathway-associated genes and polynucleotides are listed in **Table 11**.

Table 9

DISEASE/DISORDERS	GENE(S)
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
Age-related Macular Degeneration	Abcr; Ccl2; Cc2; cp (ceruloplasmin); Timp3; cathepsinD; Vldlr; Ccr2
Schizophrenia Disorders	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)
Trinucleotide Repeat Disorders	HTT (Huntington's Dx); SBMA/SMAX1/AR (Kennedy's Dx); FXN/X25 (Friedrich's Ataxia); ATX3 (Machado-Joseph's Dx); ATXN1 and ATXN2 (spinocerebellar ataxias); DMPK (myotonic dystrophy); Atrophin-1 and Atn1 (DRPLA Dx); CBP (Creb-BP - global instability); VLDLR (Alzheimer's); Atxn7; Atxn10
Fragile X Syndrome	FMR2; FXR1; FXR2; mGLUR5

Secretase Related Disorders	APH-1 (alpha and beta); Presenilin (Psen1); nicastrin (Ncstn); PEN-2
Others	Nos1; Parp1; Nat1; Nat2
Prion - related disorders	Prp
ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a; VEGF-b; VEGF-c)
Drug addiction	Prkce (alcohol); Drd2; Drd4; ABAT (alcohol); GRIA2; Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1 (alcohol)
Autism	Mecp2; BZRAP1; MDGA2; Sema5A; Neurexin 1; Fragile X (FMR2 (AFF2); FXR1; FXR2; Mglur5)
Alzheimer's Disease	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterin; PS1; SORL1; CR1; Vldlr; Uba1; Uba3; CHIP28 (Aqp1, Aquaporin 1); Uchl1; Uchl3; APP
Inflammation	IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL-17b; IL-17c; IL-17d; IL-17f); II-23; Cx3cr1; ptpn22; TNFa; NOD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b); CTLA4; Cx3cl1
Parkinson's Disease	x-Synuclein; DJ-1; LRRK2; Parkin; PINK1

Table 10	
Blood and coagulation diseases and disorders	Anemia (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); Bare lymphocyte syndrome (TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5), Bleeding disorders (TBXA2R, P2RX1, P2X1); Factor H and factor H-like 1 (HF1, CFH, HUS); Factor V and factor VIII (MCFD2); Factor VII deficiency (F7); Factor X deficiency (F10); Factor XI deficiency (F11); Factor XII deficiency (F12, HAF); Factor XIIIA deficiency (F13A1, F13A); Factor XIIIB deficiency (F13B); Fanconi anemia (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCJ, PHF9, FANCL, FANCM, KIAA1596); Hemophagocytic lymphohistiocytosis disorders (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemophilia A (F8, F8C, HEMA); Hemophilia B (F9, HEMB), Hemorrhagic disorders (PI, ATT, F5); Leukocyde deficiencies and disorders (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); Sickle cell anemia (HBB); Thalassemia (HBA2, HBB, HBD, LCRB, HBA1).
Cell dysregulation and oncology diseases and disorders	B-cell non-Hodgkin lymphoma (BCL7A, BCL7); Leukemia (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).
Inflammation and immune related diseases and disorders	AIDS (KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1); Autoimmune lymphoproliferative syndrome (TNFRSF6, APT1, FAS, CD95, ALPS1A); Combined immunodeficiency, (IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228), HIV susceptibility or infection (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCKR5 (CCR5)); Immunodeficiencies (CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI); Inflammation (IL-10, IL-1 (IL-1a, IL-1b), IL-13, IL-17 (IL-17a (CTLA8), IL-17b, IL-17c, IL-17d, IL-17f), II-23, Cx3cr1, ptpn22, TNFa, NOD2/CARD15 for IBD, IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3cl1); Severe combined immunodeficiencies (SCIDs)(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).
Metabolic, liver, kidney and protein diseases and disorders	Amyloid neuropathy (TTR, PALB); Amyloidosis (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); Cirrhosis (KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); Cystic fibrosis (CFTR, ABCC7, CF, MRP7); Glycogen storage diseases (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Hepatic adenoma, 142330 (TCF1, HNF1A, MODY3), Hepatic failure, early onset, and neurologic disorder (SCOD1, SCO1), Hepatic lipase deficiency (LIPC), Hepatoblastoma, cancer and carcinomas (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; Medullary cystic kidney disease (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); Phenylketonuria (PAH, PKU1, QDPR, DHPR, PTS); Polycystic kidney and hepatic disease (FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63).
Muscular / Skeletal diseases and disorders	Becker muscular dystrophy (DMD, BMD, MYF6), Duchenne Muscular Dystrophy (DMD, BMD); Emery-Dreifuss muscular dystrophy (LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); Facioscapulohumeral muscular dystrophy (FSHMD1A, FSHD1A); Muscular dystrophy (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); Osteopetrosis (LRP5, BMND1, LRP7,

	LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); Muscular atrophy (VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1).
Neurological and neuronal diseases and disorders	ALS (SOD1, ALS2, STEX, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c); Alzheimer disease (APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLAU, URK, ACE, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); Autism (Mecp2, BZRAP1, MDGA2, Sema5A, Neurexin 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); Fragile X Syndrome (FMR2, FXR1, FXR2, mGLUR5); Huntington's disease and disease like disorders (HD, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); Parkinson disease (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK5, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); Rett syndrome (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Synuclein, DJ-1); Schizophrenia (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)); Secretase Related Disorders (APH-1 (alpha and beta), Presenilin (Psen1), nicastrin, (Ncstn), PEN-2, Nos1, Parp1, Nat1, Nat2); Trinucleotide Repeat Disorders (HTT (Huntington's Dx), SBMA/SMAX1/AR (Kennedy's Dx), FXN/X25 (Friedrich's Ataxia), ATX3 (Machado- Joseph's Dx), ATXN1 and ATXN2 (spinocerebellar ataxias), DMPK (myotonic dystrophy), Atrophin-1 and Atn1 (DRPLA Dx), CBP (Creb-BP - global instability), VLDLR (Alzheimer's), Atxn7, Atxn10).
Ocular diseases and disorders	Age-related macular degeneration (Abcr, Ccl2, Cc2, cp (ceruloplasmin), Timp3, cathepsinD, Vldlr, Ccr2); Cataract (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); Corneal clouding and dystrophy (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); Glaucoma (MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A); Leber congenital amaurosis (CRB1, RP12, CRX, CORD2, CRD, RPGRIP1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3);

	Macular dystrophy (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2).

Table 11

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	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
Actin Cytoskeleton Signaling	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	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
Aryl Hydrocarbon Receptor Signaling	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; TGFBR1; 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	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
Wnt & Beta catenin Signaling	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; TGFBR1; 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	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1;
Stress Response	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	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1;
Metabolism	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; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13; RELA; TLR4; MAPK14; IKBKG; 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; MTTP; 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	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9; PARK7; PSEN1; PARK2; APP; CASP3
Notch Signaling	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2; PSEN1; NOTCH3; NOTCH1; DLL4
Endoplasmic Reticulum Stress Pathway	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4; EIF2AK3; CASP3
Pyrimidine Metabolism	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

[01141] 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.

[01142] 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.

[01143] General Comments on Methods of Use of the CRISPR system

[01144] In particular embodiments, the methods described herein may involve targeting one or more polynucleotide targets of interest. The polynucleotide targets of interest may be targets which are relevant to a specific disease or the treatment thereof, relevant for the generation of a given trait of interest or relevant for the production of a molecule of interest. When referring to the targeting of a "polynucleotide target" this may include targeting one or more of a coding regions, an intron, a promoter and any other 5' or 3' regulatory regions such as termination regions, ribosome binding sites, enhancers, silencers etc. The gene may encode any protein or RNA of interest. Accordingly, the target may be a coding region which can be transcribed into mRNA, tRNA or rRNA, but also recognition sites for proteins involved in replication, transcription and regulation thereof.

[01145] In particular embodiments, the methods described herein may involve targeting one or more genes of interest, wherein at least one gene of interest encodes a long noncoding RNA (lncRNA). While lncRNAs have been found to be critical for cellular functioning. As the lncRNAs that are essential have been found to differ for each cell type (C.P. Fulco et al., 2016, Science, doi:10.1126/science.aag2445; N.E. Sanjana et al., 2016, Science, doi:10.1126/science.aaf8325), the methods provided herein may involve the step of determining the lncRNA that is relevant for cellular function for the cell of interest.

[01146] In an exemplary method for modifying a target polynucleotide by integrating an exogenous polynucleotide template, a double stranded break is introduced into the genome sequence by the CRISPR complex, the break is repaired via homologous recombination an exogenous polynucleotide template such that the template is integrated into the genome. The presence of a double-stranded break facilitates integration of the template.

[01147] In other embodiments, this invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. The method comprises increasing or decreasing

expression of a target polynucleotide by using a CRISPR complex that binds to the polynucleotide.

[01148] In some methods, a target polynucleotide can be inactivated to effect the modification of the expression in a cell. For example, upon the binding of a CRISPR complex to a target sequence in a cell, the target polynucleotide is inactivated such that the sequence is not transcribed, the coded protein is not produced, or the sequence does not function as the wild-type sequence does. For example, a protein or microRNA coding sequence may be inactivated such that the protein is not produced.

[01149] In some methods, a control sequence can be inactivated such that it no longer functions as a control sequence. As used herein, "control sequence" refers to any nucleic acid sequence that effects the transcription, translation, or accessibility of a nucleic acid sequence. Examples of a control sequence include, a promoter, a transcription terminator, and an enhancer are control sequences. The inactivated target sequence may include a deletion mutation (i.e., deletion of one or more nucleotides), an insertion mutation (i.e., insertion of one or more nucleotides), or a nonsense mutation (i.e., substitution of a single nucleotide for another nucleotide such that a stop codon is introduced). In some methods, the inactivation of a target sequence results in "knockout" of the target sequence.

[01150] Also provided herein are methods of functional genomics which involve identifying cellular interactions by introducing multiple combinatorial perturbations and correlating observed genomic, genetic, proteomic, epigenetic and/or phenotypic effects with the perturbation detected in single cells, also referred to as "perturb-seq". In one embodiment, these methods combine single-cell RNA sequencing (RNA-seq) and clustered regularly interspaced short palindromic repeats (CRISPR)-based perturbations (Dixit et al. 2016, Cell 167, 1853–1866; Adamson et al. 2016, Cell 167, 1867–1882). Generally, these methods involve introducing a number of combinatorial perturbations to a plurality of cells in a population of cells, wherein each cell in the plurality of the cells receives at least 1 perturbation, detecting genomic, genetic, proteomic, epigenetic and/or phenotypic differences in single cells compared to one or more cells that did not receive any perturbation, and detecting the perturbation(s) in single cells; and determining measured differences relevant to the perturbations by applying a model accounting for co-variates to the measured differences, whereby intercellular and/or intracellular networks or circuits are inferred. More particularly, the single cell sequencing comprises cell barcodes, whereby the cell-of-origin of each RNA is recorded. More

particularly, the single cell sequencing comprises unique molecular identifiers (UMI), whereby the capture rate of the measured signals, such as transcript copy number or probe binding events, in a single cell is determined.

[01151] These methods can be used for combinatorial probing of cellular circuits, for dissecting cellular circuitry, for delineating molecular pathways, and/or for identifying relevant targets for therapeutics development. More particularly, these methods may be used to identify groups of cells based on their molecular profiling. Similarities in gene-expression profiles between organic (e.g. disease) and induced (e.g. by small molecule) states may identify clinically-effective therapies.

[01152] Accordingly, in particular embodiments, therapeutic methods provided herein comprise, determining, for a population of cells isolated from a subject, optimal therapeutic target and/or therapeutic, using perturb-seq as described above.

[01153] In particular embodiments, pertub-seq methods as referred to herein elsewhere are used to determine, in an isolated cell or cell line, cellular circuits which may affect production of a molecule of interest.

Additional CRISPR-Cas Development and Use Considerations

[01154] The present invention may be further illustrated and extended based on aspects of CRISPR-Cas9 development and use as set forth in the following articles and particularly as relates to delivery of a CRISPR protein complex and uses of an RNA guided endonuclease in cells and organisms. The techniques and methods described therein can be adapted for use with the Type VII CRISPR-Cas systems and components thereof described herein. Such references are as follows: Multiplex genome engineering using CRISPR/Cas systems. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, F. Science Feb 15;339(6121):819-23 (2013);RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Jiang W., Bikard D., Cox D., Zhang F, Marraffini LA. Nat Biotechnol Mar;31(3):233-9 (2013); One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. Wang H., Yang H., Shivalila CS., Dawlaty MM., Cheng AW., Zhang F., Jaenisch R. Cell May 9;153(4):910-8 (2013); Optical control of mammalian endogenous transcription and epigenetic states. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. Nature. Aug 22;500(7463):472-6. doi: 10.1038/Nature12466. Epub 2013 Aug 23 (2013); Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing

Specificity. Ran, FA., Hsu, PD., Lin, CY., Gootenberg, JS., Konermann, S., Trevino, AE., Scott, DA., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. Cell Aug 28. pii: S0092-8674(13)01015-5 (2013-A); DNA targeting specificity of RNA-guided Cas9 nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. Nat Biotechnol doi:10.1038/nbt.2647 (2013); Genome engineering using the CRISPR-Cas9 system. Ran, FA., Hsu, PD., Wright, J., Agarwala, V., Scott, DA., Zhang, F. Nature Protocols Nov;8(11):2281-308 (2013-B); Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. Science Dec 12. (2013). [Epub ahead of print]; Crystal structure of cas9 in complex with guide RNA and target DNA. Nishimasu, H., Ran, FA., Hsu, PD., Konermann, S., Shehata, SI., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. Cell Feb 27, 156(5):935-49 (2014); Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Wu X., Scott DA., Kriz AJ., Chiu AC., Hsu PD., Dadon DB., Cheng AW., Trevino AE., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp PA. Nat Biotechnol. Apr 20. doi: 10.1038/nbt.2889 (2014); CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, Graham DB, Jhunjhunwala S, Heidenreich M, Xavier RJ, Langer R, Anderson DG, Hacohen N, Regev A, Feng G, Sharp PA, Zhang F. Cell 159(2): 440-455 DOI: 10.1016/j.cell.2014.09.014(2014); Development and Applications of CRISPR-Cas9 for Genome Engineering, Hsu PD, Lander ES, Zhang F., Cell. Jun 5;157(6):1262-78 (2014); Genetic screens in human cells using the CRISPR/Cas9 system, Wang T, Wei JJ, Sabatini DM, Lander ES., Science. January 3; 343(6166): 80–84. doi:10.1126/science.1246981 (2014); Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation, Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE., (published online 3 September 2014) Nat Biotechnol. Dec;32(12):1262-7 (2014); *In vivo* interrogation of gene function in the mammalian brain using CRISPR-Cas9, Swiech L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, Sur M, Zhang F., (published online 19 October 2014) Nat Biotechnol. Jan;33(1):102-6 (2015); Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F., Nature. Jan 29;517(7536):583-8 (2015); A

split-Cas9 architecture for inducible genome editing and transcription modulation, Zetsche B, Volz SE, Zhang F., (published online 02 February 2015) Nat Biotechnol. Feb;33(2):139-42 (2015); Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis, Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R, Lee H, Zhang F, Sharp PA. Cell 160, 1246–1260, March 12, 2015 (multiplex screen in mouse); *In vivo* genome editing using Staphylococcus aureus Cas9, Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F., (published online 01 April 2015), Nature. Apr 9;520(7546):186-91 (2015); Shalem et al., "High-throughput functional genomics using CRISPR-Cas9," Nature Reviews Genetics 16, 299-311 (May 2015); Xu et al., "Sequence determinants of improved CRISPR sgRNA design," Genome Research 25, 1147-1157 (August 2015); Parnas et al., "A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks," Cell 162, 675-686 (July 30, 2015); Ramanan et al., CRISPR/Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus," Scientific Reports 5:10833. doi: 10.1038/srep10833 (June 2, 2015); Nishimasu et al., Crystal Structure of Staphylococcus aureus Cas9," Cell 162, 1113-1126 (Aug. 27, 2015); BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis, Canver et al., Nature 527(7577):192-7 (Nov. 12, 2015) doi: 10.1038/nature15521. Epub 2015 Sep 16; Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System, Zetsche et al., Cell 163, 759-71 (Sep 25, 2015); Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems, Shmakov et al., Molecular Cell, 60(3), 385–397 doi: 10.1016/j.molcel.2015.10.008 Epub October 22, 2015; Rationally engineered Cas9 nucleases with improved specificity, Slaymaker et al., Science 2016 Jan 1 351(6268): 84-88 doi: 10.1126/science.aad5227. Epub 2015 Dec 1. [Epub ahead of print]; Gao et al, "Engineered Cpf1 Enzymes with Altered PAM Specificities," bioRxiv 091611; doi: http://dx.doi.org/10.1101/091611 (Dec. 4, 2016), each of which is incorporated herein by reference, may be considered in the practice of the instant invention, and discussed briefly below:

[01155] Cong et al. engineered type II CRISPR-Cas systems for use in eukaryotic cells based on both Streptococcus thermophilus Cas9 and also Streptococcus pyogenes Cas9 and demonstrated that Cas9 nucleases can be directed by short RNAs to induce precise cleavage of DNA in human and mouse cells. Their study further showed that Cas9 as converted into a nicking enzyme can be used to facilitate homology-directed repair in eukaryotic cells with

minimal mutagenic activity. Additionally, their study demonstrated that multiple guide sequences can be encoded into a single CRISPR array to enable simultaneous editing of several at endogenous genomic loci sites within the mammalian genome, demonstrating easy programmability and wide applicability of the RNA-guided nuclease technology. This ability to use RNA to program sequence specific DNA cleavage in cells defined a new class of genome engineering tools. These studies further showed that other CRISPR loci are likely to be transplantable into mammalian cells and can also mediate mammalian genome cleavage. Importantly, it can be envisaged that several aspects of the CRISPR-Cas system can be further improved to increase its efficiency and versatility.

[01156] Jiang et al. used the clustered, regularly interspaced, short palindromic repeats (CRISPR)–associated Cas9 endonuclease complexed with dual-RNAs to introduce precise mutations in the genomes of Streptococcus pneumoniae and Escherichia coli. The approach relied on dual-RNA:Cas9-directed cleavage at the targeted genomic site to kill unmutated cells and circumvents the need for selectable markers or counter-selection systems. The study reported reprogramming dual-RNA:Cas9 specificity by changing the sequence of short CRISPR RNA (crRNA) to make single- and multinucleotide changes carried on editing templates. The study showed that simultaneous use of two crRNAs enabled multiplex mutagenesis. Furthermore, when the approach was used in combination with recombineering, in S. pneumoniae, nearly 100% of cells that were recovered using the described approach contained the desired mutation, and in E. coli, 65% that were recovered contained the mutation.

[01157] Wang et al. (2013) used the CRISPR-Cas system for the one-step generation of mice carrying mutations in multiple genes which were traditionally generated in multiple steps by sequential recombination in embryonic stem cells and/or time-consuming intercrossing of mice with a single mutation. The CRISPR-Cas system will greatly accelerate the *in vivo* study of functionally redundant genes and of epistatic gene interactions.

[01158] Konermann et al. (2013) addressed the need in the art for versatile and robust technologies that enable optical and chemical modulation of DNA-binding domains based CRISPR Cas9 enzyme and also Transcriptional Activator Like Effectors

[01159] Ran et al. (2013-A) described an approach that combined a Cas9 nickase mutant with paired guide RNAs to introduce targeted double-strand breaks. This addresses the issue of the Cas9 nuclease from the microbial CRISPR-Cas system being targeted to specific genomic loci by a guide sequence, which can tolerate certain mismatches to the DNA target

and thereby promote undesired off-target mutagenesis. Because individual nicks in the genome are repaired with high fidelity, simultaneous nicking via appropriately offset guide RNAs is required for double-stranded breaks and extends the number of specifically recognized bases for target cleavage. The authors demonstrated that using paired nicking can reduce off-target activity by 50- to 1,500-fold in cell lines and to facilitate gene knockout in mouse zygotes without sacrificing on-target cleavage efficiency. This versatile strategy enables a wide variety of genome editing applications that require high specificity.

[01160] Hsu et al. (2013) characterized SpCas9 targeting specificity in human cells to inform the selection of target sites and avoid off-target effects. The study evaluated >700 guide RNA variants and SpCas9-induced indel mutation levels at >100 predicted genomic off-target loci in 293T and 293FT cells. The authors that SpCas9 tolerates mismatches between guide RNA and target DNA at different positions in a sequence-dependent manner, sensitive to the number, position and distribution of mismatches. The authors further showed that SpCas9-mediated cleavage is unaffected by DNA methylation and that the dosage of SpCas9 and gRNA can be titrated to minimize off-target modification. Additionally, to facilitate mammalian genome engineering applications, the authors reported providing a web-based software tool to guide the selection and validation of target sequences as well as off-target analyses.

[01161] Ran et al. (2013-B) described a set of tools for Cas9-mediated genome editing via non-homologous end joining (NHEJ) or homology-directed repair (HDR) in mammalian cells, as well as generation of modified cell lines for downstream functional studies. To minimize off-target cleavage, the authors further described a double-nicking strategy using the Cas9 nickase mutant with paired guide RNAs. The protocol provided by the authors experimentally derived guidelines for the selection of target sites, evaluation of cleavage efficiency and analysis of off-target activity. The studies showed that beginning with target design, gene modifications can be achieved within as little as 1–2 weeks, and modified clonal cell lines can be derived within 2–3 weeks.

[01162] Shalem et al. described a new way to interrogate gene function on a genome-wide scale. Their studies showed that delivery of a genome-scale CRISPR-Cas9 knockout (GeCKO) library targeted 18,080 genes with 64,751 unique guide sequences enabled both negative and positive selection screening in human cells. First, the authors showed use of the GeCKO library to identify genes essential for cell viability in cancer and pluripotent stem cells. Next, in a melanoma model, the authors screened for genes whose loss is involved in resistance to

vemurafenib, a therapeutic that inhibits mutant protein kinase BRAF. Their studies showed that the highest-ranking candidates included previously validated genes NF1 and MED12 as well as novel hits NF2, CUL3, TADA2B, and TADA1. The authors observed a high level of consistency between independent guide RNAs targeting the same gene and a high rate of hit confirmation, and thus demonstrated the promise of genome-scale screening with Cas9.

[01163] Nishimasu et al. reported the crystal structure of Streptococcus pyogenes Cas9 in complex with sgRNA and its target DNA at 2.5 A° resolution. The structure revealed a bilobed architecture composed of target recognition and nuclease lobes, accommodating the sgRNA:DNA heteroduplex in a positively charged groove at their interface. Whereas the recognition lobe is essential for binding sgRNA and DNA, the nuclease lobe contains the HNH and RuvC nuclease domains, which are properly positioned for cleavage of the complementary and non-complementary strands of the target DNA, respectively. The nuclease lobe also contains a carboxyl-terminal domain responsible for the interaction with the protospacer adjacent motif (PAM). This high-resolution structure and accompanying functional analyses have revealed the molecular mechanism of RNA-guided DNA targeting by Cas9, thus paving the way for the rational design of new, versatile genome-editing technologies.

[01164] Wu et al. mapped genome-wide binding sites of a catalytically inactive Cas9 (dCas9) from Streptococcus pyogenes loaded with single guide RNAs (sgRNAs) in mouse embryonic stem cells (mESCs). The authors showed that each of the four sgRNAs tested targets dCas9 to between tens and thousands of genomic sites, frequently characterized by a 5-nucleotide seed region in the sgRNA and an NGG protospacer adjacent motif (PAM). Chromatin inaccessibility decreases dCas9 binding to other sites with matching seed sequences; thus 70% of off-target sites are associated with genes. The authors showed that targeted sequencing of 295 dCas9 binding sites in mESCs transfected with catalytically active Cas9 identified only one site mutated above background levels. The authors proposed a two-state model for Cas9 binding and cleavage, in which a seed match triggers binding but extensive pairing with target DNA is required for cleavage.

[01165] Platt et al. established a Cre-dependent Cas9 knockin mouse. The authors demonstrated *in vivo* as well as *ex vivo* genome editing using adeno-associated virus (AAV)-, lentivirus-, or particle-mediated delivery of guide RNA in neurons, immune cells, and endothelial cells.

[01166] Hsu et al. (2014) is a review article that discusses generally CRISPR-Cas9 history from yogurt to genome editing, including genetic screening of cells.

[01167] Wang et al. (2014) relates to a pooled, loss-of-function genetic screening approach suitable for both positive and negative selection that uses a genome-scale lentiviral single guide RNA (sgRNA) library.

[01168] Doench et al. created a pool of sgRNAs, tiling across all possible target sites of a panel of six endogenous mouse and three endogenous human genes and quantitatively assessed their ability to produce null alleles of their target gene by antibody staining and flow cytometry. The authors showed that optimization of the PAM improved activity and also provided an on-line tool for designing sgRNAs.

[01169] Swiech et al. demonstrate that AAV-mediated SpCas9 genome editing can enable reverse genetic studies of gene function in the brain.

[01170] Konermann et al. (2015) discusses the ability to attach multiple effector domains, e.g., transcriptional activator, functional and epigenomic regulators at appropriate positions on the guide such as stem or tetraloop with and without linkers.

[01171] Zetsche et al. demonstrates that the Cas9 enzyme can be split into two and hence the assembly of Cas9 for activation can be controlled.

[01172] Chen et al. relates to multiplex screening by demonstrating that a genome-wide *in vivo* CRISPR-Cas9 screen in mice reveals genes regulating lung metastasis.

[01173] Ran et al. (2015) relates to SaCas9 and its ability to edit genomes and demonstrates that one cannot extrapolate from biochemical assays.

[01174] Shalem et al. (2015) described ways in which catalytically inactive Cas9 (dCas9) fusions are used to synthetically repress (CRISPRi) or activate (CRISPRa) expression, showing. advances using Cas9 for genome-scale screens, including arrayed and pooled screens, knockout approaches that inactivate genomic loci and strategies that modulate transcriptional activity.

[01175] Xu et al. (2015) assessed the DNA sequence features that contribute to single guide RNA (sgRNA) efficiency in CRISPR-based screens. The authors explored efficiency of CRISPR/Cas9 knockout and nucleotide preference at the cleavage site. The authors also found that the sequence preference for CRISPRi/a is substantially different from that for CRISPR/Cas9 knockout.

[01176] Parnas et al. (2015) introduced genome-wide pooled CRISPR-Cas9 libraries into dendritic cells (DCs) to identify genes that control the induction of tumor necrosis factor (Tnf) by bacterial lipopolysaccharide (LPS). Known regulators of Tlr4 signaling and previously unknown candidates were identified and classified into three functional modules with distinct effects on the canonical responses to LPS.

[01177] Ramanan et al (2015) demonstrated cleavage of viral episomal DNA (cccDNA) in infected cells. The HBV genome exists in the nuclei of infected hepatocytes as a 3.2kb double-stranded episomal DNA species called covalently closed circular DNA (cccDNA), which is a key component in the HBV life cycle whose replication is not inhibited by current therapies. The authors showed that sgRNAs specifically targeting highly conserved regions of HBV robustly suppresses viral replication and depleted cccDNA.

[01178] Nishimasu et al. (2015) reported the crystal structures of SaCas9 in complex with a single guide RNA (sgRNA) and its double-stranded DNA targets, containing the 5'-TTGAAT-3' PAM and the 5'-TTGGGT-3' PAM. A structural comparison of SaCas9 with SpCas9 highlighted both structural conservation and divergence, explaining their distinct PAM specificities and orthologous sgRNA recognition.

[01179] Canver et al. (2015) demonstrated a CRISPR-Cas9-based functional investigation of non-coding genomic elements. The authors developed pooled CRISPR-Cas9 guide RNA libraries to perform in situ saturating mutagenesis of the human and mouse BCL11A enhancers which revealed critical features of the enhancers.

[01180] Zetsche et al. (2015) reported characterization of Cpf1, a class 2 CRISPR nuclease from Francisella novicida U112 having features distinct from Cas9. Cpf1 is a single RNA-guided endonuclease lacking tracrRNA, utilizes a T-rich protospacer-adjacent motif, and cleaves DNA via a staggered DNA double-stranded break.

[01181] Shmakov et al. (2015) reported three distinct Class 2 CRISPR-Cas systems. Two system CRISPR enzymes (C2c1 and C2c3) contain RuvC-like endonuclease domains distantly related to Cpf1. Unlike Cpf1, C2c1 depends on both crRNA and tracrRNA for DNA cleavage. The third enzyme (C2c2) contains two predicted HEPN RNase domains and is tracrRNA independent.

[01182] Slaymaker et al (2016) reported the use of structure-guided protein engineering to improve the specificity of Streptococcus pyogenes Cas9 (SpCas9). The authors developed

"enhanced specificity" SpCas9 (eSpCas9) variants which maintained robust on-target cleavage with reduced off-target effects.

[01183] The methods and tools provided herein are exemplified for certain Type VII effectors. Further type VII nucleases with similar properties can be identified using methods described in the art (Shmakov et al. 2015, 60:385–397; Abudayeh et al. 2016, Science, 5;353(6299)) . In particular embodiments, such methods for identifying novel CRISPR effector proteins may comprise the steps of selecting sequences from the database encoding a seed which identifies the presence of a CRISPR Cas locus, identifying loci located within 10 kb of the seed comprising Open Reading Frames (ORFs) in the selected sequences, selecting therefrom loci comprising ORFs of which only a single ORF encodes a novel CRISPR effector having greater than 700 amino acids and no more than 90% homology to a known CRISPR effector. In particular embodiments, the seed is a protein that is common to the CRISPR-Cas system, such as Cas1. In further embodiments, the CRISPR array is used as a seed to identify new effector proteins.

[01184] Preassembled recombinant CRISPR-Type VII effector complexes comprising Type VII effector and crRNA may be transfected, for example by electroporation, resulting in high mutation rates and absence of detectable off-target mutations, as has been demonstrated for certain other CRISPR effectors. Hur, J.K. et al, Targeted mutagenesis in mice by electroporation of Cpf1 ribonucleoproteins, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3596. [Epub ahead of print]. Genome-wide analyses shows that Cpf1 is highly specific. By one measure, *in vitro* cleavage sites determined for SpCas9 in human HEK293T cells were significantly fewer that for SpCas9. Kim, D. et al., Genome-wide analysis reveals specificities of Cpf1 endonucleases in human cells, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3609. [Epub ahead of print]. An efficient multiplexed system employing Cpf1 has been demonstrated in Drosophila employing gRNAs processed from an array containing inventing tRNAs. Port, F. et al, Expansion of the CRISPR toolbox in an animal with tRNA-flanked Cas9 and Cpf1 gRNAs. doi: http://dx.doi.org/10.1101/046417.

[01185] Also, "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing", Shengdar Q. Tsai, Nicolas Wyvekens, Cyd Khayter, Jennifer A. Foden, Vishal Thapar, Deepak Reyon, Mathew J. Goodwin, Martin J. Aryee, J. Keith Joung Nature Biotechnology 32(6): 569-77 (2014), relates to dimeric RNA-guided FokI Nucleases that

recognize extended sequences and can edit endogenous genes with high efficiencies in human cells.

[01186] With respect to general information on CRISPR-Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, AAV, and making and using thereof, including as to amounts and formulations, all useful in the practice of the instant invention, reference is made to: US Patents Nos. 8,697,359, 8,771,945, 8,795,965, 8,865,406, 8,871,445, 8,889,356, 8,889,418, 8,895,308, 8,906,616, 8,932,814, 8,945,839, 8,993,233 and 8,999,641; US Patent Publication Nos. US 2014-0310830 A1 (US App. Ser. No. 14/105,031), US 2014-0287938 A1 (U.S. App. Ser. No. 14/213,991), US 2014-0273234 A1 (U.S. App. Ser. No. 14/293,674), US2014-0273232 A1 (U.S. App. Ser. No. 14/290,575), US 2014-027323 A1 (U.S. App. Ser. No. 14/259,420), US 2014-0256046 A1 (U.S. App. Ser. No. 14/226,274), US 2014-0248702 A1 (U.S. App. Ser. No. 14/258,458), US 2014-0242700 A1 (U.S. App. Ser. No. 14/222,930), US 2014-0242699 A1 (U.S. App. Ser. No. 14/183,512), US 2014-0242664 A1 (U.S. App. Ser. No. 14/104,990), US 2014-0234972 A1 (U.S. App. Ser. No. 14/183,471), US 2014-0227787 A1 (U.S. App. Ser. No. 14/256,912), US 2014-0189896 A1 (U.S. App. Ser. No. 14/105,035), US 2014-0186958 A1 (U.S. App. Ser. No. 14/105,017), US 2014-0186919 A1 (U.S. App. Ser. No. 14/104,977), US 2014-0186843 A1 (U.S. App. Ser. No. 14/104,900), US 2014-0179770 A1 (U.S. App. Ser. No. 14/104,837) and US 2014-0179006 A1 (U.S. App. Ser. No. 14/183,486), US 2014-0170753 A1 (US App Ser No 14/183,429); US 2015-0184139 A1 (U.S. App. Ser. No. 14/324,960); 14/054,414 European Patent Applications EP 2771468 (EP13818570.7), EP 2764103 (EP13824232.6), and EP 2784162 (EP14170383.5); and PCT Patent Publications WO 2014/093661 (PCT/US2013/074743), WO 2014/093694 (PCT/US2013/074790), WO 2014/093595 (PCT/US2013/074611), WO 2014/093718 (PCT/US2013/074825), WO 2014/093709 (PCT/US2013/074812), WO 2014/093622 (PCT/US2013/074667), WO 2014/093635 (PCT/US2013/074691), WO 2014/093655 (PCT/US2013/074736), WO 2014/093712 (PCT/US2013/074819), WO 2014/093701 (PCT/US2013/074800), WO 2014/018423 (PCT/US2013/051418), WO 2014/204723 (PCT/US2014/041790), WO 2014/204724 (PCT/US2014/041800), WO 2014/204725 (PCT/US2014/041803), WO 2014/204726 (PCT/US2014/041804), WO 2014/204727 (PCT/US2014/041806), WO 2014/204728 (PCT/US2014/041808), WO 2014/204729 (PCT/US2014/041809), WO 2015/089351 (PCT/US2014/069897), WO 2015/089354 (PCT/US2014/069902), WO 2015/089364

(PCT/US2014/069925), WO 2015/089427 (PCT/US2014/070068), WO 2015/089462 (PCT/US2014/070127), WO 2015/089419 (PCT/US2014/070057), WO 2015/089465 (PCT/US2014/070135), WO 2015/089486 (PCT/US2014/070175), PCT/US2015/051691, PCT/US2015/051830. Reference is also made to US Provisional Application Nos. 61/758,468; 61/802,174; 61/806,375; 61/814,263; 61/819,803 and 61/828,130, filed on January 30, 2013; March 15, 2013; March 28, 2013; April 20, 2013; May 6, 2013 and May 28, 2013 respectively. Reference is also made to US Provisional Application No. 61/836,123, filed on June 17, 2013. Reference is additionally made to US Provisional Application Nos. 61/835,931, 61/835,936, 61/835,973, 61/836,080, 61/836,101, and 61/836,127, each filed June 17, 2013. Further reference is made to US Provisional Application Nos. 61/862,468 and 61/862,355 filed on August 5, 2013; 61/871,301 filed on August 28, 2013; 61/960,777 filed on September 25, 2013 and 61/961,980 filed on October 28, 2013. Reference is yet further made to: PCT/US2014/62558 filed October 28, 2014, and US Provisional Patent Application Nos.: 61/915,148, 61/915,150, 61/915,153, 61/915,203, 61/915,251, 61/915,301, 61/915,267, 61/915,260, and 61/915,397, each filed December 12, 2013; 61/757,972 and 61/768,959, filed on January 29, 2013 and February 25, 2013; 62/010,888 and 62/010,879, both filed June 11, 2014; 62/010,329, 62/010,439 and 62/010,441, each filed June 10, 2014; 61/939,228 and 61/939,242, each filed February 12, 2014; 61/980,012, filed April 15, 2014; 62/038,358, filed August 17, 2014; 62/055,484, 62/055,460 and 62/055,487, each filed September 25, 2014; and 62/069,243, filed October 27, 2014. Reference is made to PCT application designating, inter alia, the United States, application No. PCT/US14/41806, filed June 10, 2014. Reference is made to US provisional patent application 61/930,214 filed on January 22, 2014. Reference is made to PCT application designating, inter alia, the United States, application No. PCT/US14/41806, filed June 10, 2014.

[01187] Mention is also made of US Provisional Application No. 62/180,709, filed 17-Jun-2015, PROTECTED GUIDE RNAS (PGRNAS); US Provisional Application No. 62/091,455, filed 12-Dec-2014, PROTECTED GUIDE RNAS (PGRNAS); US Provisional Application No. 62/096,708, filed 24-Dec-2014, PROTECTED GUIDE RNAS (PGRNAS); US Provisional Application Nos. 62/091,462, filed 12-Dec-2014, 62/096,324, filed 23-Dec-2014, 62/180,681, filed 17-Jun-2015, and 62/237,496, filed 5-Oct-2015, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US Provisional Application No. 62/091,456, filed 12-Dec-2014 and 62/180,692, filed 17-Jun-2015, ESCORTED AND FUNCTIONALIZED GUIDES

FOR CRISPR-CAS SYSTEMS; US Provisional Application No. 62/091,461, filed 12-Dec-2014, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs); US Provisional Application No. 62/094,903, filed 19-Dec-2014, UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE SEQUENCING; US Provisional Application No. 62/096,761, filed 24-Dec-14, ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION; US Provisional Application Nos. 62/098,059, filed 30-Dec-2014, 62/181,641, filed 18-Jun-2015, and 62/181,667, filed 18-Jun-2015, RNA-TARGETING SYSTEM; US Provisional Application No2. 62/096,656, filed 24-Dec-2014 and 62/181,151, filed 17-Jun-2015, CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS; US Provisional Application No. 62/096,697, filed 24-Dec-2014, CRISPR HAVING OR ASSOCIATED WITH AAV; US Provisional Application No. 62/098,158, filed 30-Dec-2014, ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS; US Provisional Application No. 62/151,052, filed 22-Apr-2015, CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING; US Provisional Application No. 62/054,490, filed 24-Sep-2014, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS; US Provisional Application No. 61/939,154, filed 12-Feb-2014, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US Provisional Application No. 62/055,484, filed 25-Sep-2014, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US Provisional Application No. 62/087,537, filed 4-Dec-2014, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US Provisional Application No. 62/054,651, filed 24-Sep-2014, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS *IN VIVO*; US Provisional Application No. 62/067,886, filed 23-Oct-2014, DELIVERY, USE AND THERAPEUTIC

APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS *IN VIVO*; US Provisional Application No. 62/054,675, filed 24-Sep-2014 and 62/181,002, filed 17-Jun-2015, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES; US Provisional Application No. 62/054,528, filed 24-Sep-2014, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS; US Provisional Application No. 62/055,454, filed 25-Sep-2014, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP); US Provisional Application No. 62/055,460, filed 25-Sep-2014, MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; US Provisional Application Nos. 62/087,475, filed 4-Dec-2014 and 62/181,690, filed 18-Jun-2015, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US Provisional Application No. 62/055,487, filed 25-Sep-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US Provisional Application Nos. 62/087,546, filed 4-Dec-2014 and 62/181,687, filed 18-Jun-2015, MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; and US Provisional Application No. 62/098,285, filed 30-Dec-2014, CRISPR MEDIATED *IN VIVO* MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS.

[01188] Mention is made of US Provisional Application Nos. 62/181,659, filed 18-Jun-2015 and 62/207,318, filed 19-Aug-2015, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS, ENZYME AND GUIDE SCAFFOLDS OF CAS9 ORTHOLOGS AND VARIANTS FOR SEQUENCE MANIPULATION. Mention is made of US Provisional Application Nos. 62/181,663, filed 18-Jun-2015 and 62/245,264, filed 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS, US Provisional Application Nos. 62/181,675, filed 18-Jun-2015, 62/285,349, filed 22-Oct-2015, 62/296,522, filed 17-Feb-2016, and 62/320,231, filed 8-Apr-2016, NOVEL CRISPR ENZYMES AND SYSTEMS, US Provisional Application No. 62/232,067, filed 24-Sep-2015, US Application No. 14/975,085, filed 18-Dec-2015, European Application No. 16150428.7, US Provisional Application No. 62/205,733, filed 16-

Aug-2015, US Provisional Application No. 62/201,542, filed 5-Aug-2015, US Provisional Application No. 62/193,507, filed 16-Jul-2015, and US Provisional Application No. 62/181,739, filed 18-Jun-2015, each entitled NOVEL CRISPR ENZYMES AND SYSTEMS and of US Provisional Application No. 62/245,270, filed 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS. Mention is also made of US Provisional Application No. 61/939,256, filed 12-Feb-2014, and International Patent Publication No. WO 2015/089473 (PCT/US2014/070152), 12-Dec-2014, each entitled ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED GUIDE COMPOSITIONS WITH NEW ARCHITECTURES FOR SEQUENCE MANIPULATION. Mention is also made of PCT/US2015/045504, 15-Aug-2015, US Provisional Application No. 62/180,699, filed 17-Jun-2015, and US Provisional Application No. 62/038,358, filed 17-Aug-2014, each entitled GENOME EDITING USING CAS9 NICKASES.

[01189] In addition, mention is made of PCT application PCT/US14/70057, Attorney Reference 47627.99.2060 and BI-2013/107 entitled "DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS (claiming priority from one or more or all of US Provisional Application Nos. 62/054,490, filed September 24, 2014; 62/010,441, filed June 10, 2014; and 61/915,118, 61/915,215 and 61/915,148, each filed on December 12, 2013) ("the Particle Delivery PCT"), incorporated herein by reference, and of PCT application PCT/US14/70127, Attorney Reference 47627.99.2091 and BI-2013/101 entitled "DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING " (claiming priority from one or more or all of US Provisional Application Nos. 61/915,176; 61/915,192; 61/915,215; 61/915,107, 61/915,145; 61/915,148; and 61/915,153 each filed December 12, 2013) ("the Eye PCT"), incorporated herein by reference, with respect to a method of preparing an sgRNA-and-Type VII effector protein containing particle comprising admixing a mixture comprising an sgRNA and Type VII effector protein (and optionally HDR template) with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol; and particles from such a process. For example, wherein Type VII effector protein and sgRNA were mixed together at a suitable, e.g., 3:1 to 1:3 or 2:1 to 1:2 or 1:1 molar ratio, at a suitable temperature, e.g., 15-30C, e.g., 20-25C, e.g., room temperature,

for a suitable time, e.g., 15-45, such as 30 minutes, advantageously in sterile, nuclease free buffer, e.g., 1X PBS. Separately, particle components such as or comprising: a surfactant, e.g., cationic lipid, e.g., 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP); phospholipid, e.g., dimyristoylphosphatidylcholine (DMPC); biodegradable polymer, such as an ethylene-glycol polymer or PEG, and a lipoprotein, such as a low-density lipoprotein, e.g., cholesterol were dissolved in an alcohol, advantageously a C1-6 alkyl alcohol, such as methanol, ethanol, isopropanol, e.g., 100% ethanol. The two solutions were mixed together to form particles containing the Cas9-sgRNA complexes. Accordingly, sgRNA may be pre-complexed with the Type VII effector protein, before formulating the entire complex in a particle. Formulations may be made with a different molar ratio of different components known to promote delivery of nucleic acids into cells (e.g. 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP), 1,2-ditetradecanoyl-sn-glycero-3-phosphocholine (DMPC), polyethylene glycol (PEG), and cholesterol) For example DOTAP : DMPC : PEG : Cholesterol Molar Ratios may be DOTAP 100, DMPC 0, PEG 0, Cholesterol 0; or DOTAP 90, DMPC 0, PEG 10, Cholesterol 0; or DOTAP 90, DMPC 0, PEG 5, Cholesterol 5. DOTAP 100, DMPC 0, PEG 0, Cholesterol 0. That application accordingly comprehends admixing sgRNA, Type VII effector protein and components that form a particle; as well as particles from such admixing. Aspects of the instant invention can involve particles; for example, particles using a process analogous to that of the Particle Delivery PCT or that of the Eye PCT, e.g., by admixing a mixture comprising sgRNA and/or Type VII effector as in the instant invention and components that form a particle, e.g., as in the Particle Delivery PCT or in the Eye PCT, to form a particle and particles from such admixing (or, of course, other particles involving sgRNA and/or Type VII effector as in the instant invention).

[01190] The subject invention may be used as part of a research program wherein there is transmission of results or data. A computer system (or digital device) may be used to receive, transmit, display and/or store results, analyze the data and/or results, and/or produce a report of the results and/or data and/or analysis. A computer system may be understood as a logical apparatus that can read instructions from media (e.g. software) and/or network port (e.g. from the internet), which can optionally be connected to a server having fixed media. A computer system may comprise one or more of a CPU, disk drives, input devices such as keyboard and/or mouse, and a display (e.g. a monitor). Data communication, such as transmission of instructions or reports, can be achieved through a communication medium to a server at a local

or a remote location. The communication medium can include any means of transmitting and/or receiving data. For example, the communication medium can be a network connection, a wireless connection, or an internet connection. Such a connection can provide for communication over the World Wide Web. It is envisioned that data relating to the present invention can be transmitted over such networks or connections (or any other suitable means for transmitting information, including but not limited to mailing a physical report, such as a print-out) for reception and/or for review by a receiver. The receiver can be but is not limited to an individual, or electronic system (e.g., one or more computers, and/or one or more servers). In some embodiments, the computer system comprises one or more processors. Processors may be associated with one or more controllers, calculation units, and/or other units of a computer system, or implanted in firmware as desired. If implemented in software, the routines may be stored in any computer readable memory such as in RAM, ROM, flash memory, a magnetic disk, a laser disk, or other suitable storage medium. Likewise, this software may be delivered to a computing device via any known delivery method including, for example, over a communication channel such as a telephone line, the internet, a wireless connection, etc., or via a transportable medium, such as a computer readable disk, flash drive, etc. The various steps may be implemented as various blocks, operations, tools, modules and techniques which, in turn, may be implemented in hardware, firmware, software, or any combination of hardware, firmware, and/or software. When implemented in hardware, some or all of the blocks, operations, techniques, etc. may be implemented in, for example, a custom integrated circuit (IC), an application specific integrated circuit (ASIC), a field programmable logic array (FPGA), a programmable logic array (PLA), etc. A client-server, relational database architecture can be used in embodiments of the invention. A client-server architecture is a network architecture in which each computer or process on the network is either a client or a server. Server computers are typically powerful computers dedicated to managing disk drives (file servers), printers (print servers), or network traffic (network servers). Client computers include PCs (personal computers) or workstations on which users run applications, as well as example output devices as disclosed herein. Client computers rely on server computers for resources, such as files, devices, and even processing power. In some embodiments of the invention, the server computer handles all of the database functionality. The client computer can have software that handles all the front-end data management and can also receive data input from users. A machine-readable medium comprising computer-executable code may

take many forms, including but not limited to, a tangible storage medium, a carrier wave medium or physical transmission medium. Non-volatile storage media include, for example, optical or magnetic disks, such as any of the storage devices in any computer(s) or the like, such as may be used to implement the databases, etc. shown in the drawings. Volatile storage media include dynamic memory, such as main memory of such a computer platform. Tangible transmission media include coaxial cables; copper wire and fiber optics, including the wires that comprise a bus within a computer system. Carrier-wave transmission media may take the form of electric or electromagnetic signals, or acoustic or light waves such as those generated during radio frequency (RF) and infrared (IR) data communications. Common forms of computer-readable media therefore include for example: a floppy disk, a flexible disk, hard disk, magnetic tape, any other magnetic medium, a CD-ROM, DVD or DVD-ROM, any other optical medium, punch cards paper tape, any other physical storage medium with patterns of holes, a RAM, a ROM, a PROM and EPROM, a FLASH-EPROM, any other memory chip or cartridge, a carrier wave transporting data or instructions, cables or links transporting such a carrier wave, or any other medium from which a computer may read programming code and/or data. Many of these forms of computer readable media may be involved in carrying one or more sequences of one or more instructions to a processor for execution. Accordingly, the invention comprehends performing any method herein-discussed and storing and/or transmitting data and/or results therefrom and/or analysis thereof, as well as products from performing any method herein-discussed, including intermediates.

[01191] Throughout this disclosure there has been mention of CRISPR or CRISPR-Cas complexes or systems. CRISPR systems or complexes can target nucleic acid molecules, e.g., CRISPR-Type VII effector complexes can target and cleave or nick or simply sit upon a target DNA molecule (depending if the Type VII effector has mutations that render it a nickase or "dead"). Such systems or complexes are amenable for achieving tissue-specific and temporally controlled targeted deletion of candidate disease genes. Examples include but are not limited to genes involved in cholesterol and fatty acid metabolism, amyloid diseases, dominant negative diseases, latent viral infections, among other disorders. Accordingly, target sequences for such systems or complexes can be in candidate disease genes, as shown for example in **Table 12** below.

Table 12

Disease	GENE	SPACER	PAM	Mechanism	References

Hypercholesterolemia	HMG-CR	GCCAAATTGGACGACCCTCG **(SEQ ID NO: 26)**	CGG	Knockout	Fluvastatin: a review of its pharmacology and use in the management of hypercholesterolaemia.(Plosker GL et al. Drugs 1996, 51(3):433-459)
Hypercholesterolemia	SQLE	CGAGGAGACCCCCGTTTCGG**(SEQ ID NO: 27)**	TGG	Knockout	Potential role of nonstatin cholesterol lowering agents (Trapani et al. IUBMB Life, Volume 63, Issue 11, pages 964–971, November 2011)
Hyperlipidemia	DGAT1	CCCGCCGCCGCCGTGGCTCG**(SEQ ID NO: 28)**	AGG	Knockout	DGAT1 inhibitors as anti-obesity and anti-diabetic agents. (Birch AM et al. Current Opinion in Drug Discovery & Development [2010, 13(4):489-496)
Leukemia	BCR-ABL	TGAGCTCTACGAGATCCACA**(SEQ ID NO: 29)**	AGG	Knockout	Killing of leukemic cells with a BCR/ABL fusion gene by RNA interference (RNAi).(Fuchs et al. Oncogene 2002, 21(37):5716-5724)

[01192] Thus, the present invention, with regard to CRISPR or CRISPR-Cas complexes contemplates correction of hematopoietic disorders. For example, Severe Combined Immune Deficiency (SCID) results from a defect in lymphocytes T maturation, always associated with a functional defect in lymphocytes B (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). In the case of Adenosine Deaminase (ADA) deficiency, one of the SCID forms, patients can be treated by injection of recombinant Adenosine Deaminase enzyme. Since the ADA gene has been shown to be mutated in SCID patients (Giblett et al., Lancet, 1972, 2, 1067-1069), several other genes involved in SCID have been identified (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). There are four major causes for SCID: (i) the most frequent form of SCID, SCID-X1 (X-linked SCID or X-SCID), is caused by mutation in the IL2RG gene, resulting in the absence of mature T lymphocytes and NK cells. IL2RG encodes the gamma C protein (Noguchi, et al., Cell, 1993, 73, 147-157), a common component of at least five interleukin receptor complexes. These receptors activate several targets through the JAK3 kinase (Macchi et al., Nature, 1995, 377, 65-68), which inactivation results in the same syndrome as gamma C inactivation; (ii) mutation in the ADA gene results in a defect in purine metabolism that is lethal for lymphocyte precursors, which in

turn results in the quasi absence of B, T and NK cells; (iii) V(D)J recombination is an essential step in the maturation of immunoglobulins and T lymphocytes receptors (TCRs). Mutations in Recombination Activating Gene 1 and 2 (RAG1 and RAG2) and Artemis, three genes involved in this process, result in the absence of mature T and B lymphocytes; and (iv) Mutations in other genes such as CD45, involved in T cell specific signaling have also been reported, although they represent a minority of cases (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). In aspect of the invention, relating to CRISPR or CRISPR-Cas complexes contemplates system, the invention contemplates that it may 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. Non-limiting examples of ocular defects to be corrected include macular degeneration (MD), retinitis pigmentosa (RP). Non-limiting examples of genes and proteins associated with ocular defects include but are not limited to the following proteins: (ABCA4) ATP-binding cassette, sub-family A (ABC1), member 4 ACHM1 achromatopsia (rod monochromacy) 1 ApoE Apolipoprotein E (ApoE) C1QTNF5 (CTRP5) C1q and tumor necrosis factor related protein 5 (C1QTNF5) C2 Complement component 2 (C2) C3 Complement components (C3) CCL2 Chemokine (C-C motif) Ligand 2 (CCL2) CCR2 Chemokine (C-C motif) receptor 2 (CCR2) CD36 Cluster of Differentiation 36 CFB Complement factor B CFH Complement factor CFH H CFHR1 complement factor H-related 1 CFHR3 complement factor H-related 3 CNGB3 cyclic nucleotide gated channel beta 3 CP ceruloplasmin (CP) CRP C reactive protein (CRP) CST3 cystatin C or cystatin 3 (CST3) CTSD Cathepsin D (CTSD) CX3CR1 chemokine (C-X3-C motif) receptor 1 ELOVL4 Elongation of very long chain fatty acids 4 ERCC6 excision repair cross- complementing rodent repair deficiency, complementation group 6 FBLN5 Fibulin-5 FBLN5 Fibulin 5 FBLN6 Fibulin 6 FSCN2 fascin (FSCN2) HMCN1 Hemicentrin 1 HMCN1 hemicentin 1 HTRA1 HtrA serine peptidase 1 (HTRA1) HTRA1 HtrA serine peptidase 1 IL-6 Interleukin 6 IL-8 Interleukin 8 LOC387715 Hypothetical protein PLEKHA1 Pleckstrin homology domain-containing family A member 1 (PLEKHA1) PROM1 Prominin 1(PROM1 or CD133) PRPH2 Peripherin-2 RPGR retinitis pigmentosa GTPase regulator SERPING1 serpin peptidase inhibitor, clade G, member 1 (C1- inhibitor) TCOF1 Treacle TIMP3 Metalloproteinase inhibitor 3 (TIMP3) TLR3 Toll-like receptor 3 The present invention, with regard to CRISPR or CRISPR-Cas complexes contemplates also contemplates delivering to the heart. For the

heart, a myocardium tropic adena-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). For example, US Patent Publication No. 20110023139, 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. By way of example, the chromosomal sequence may comprise, but is not limited to, IL1B (interleukin 1, beta), XDH (xanthine dehydrogenase), TP53 (tumor protein p53), PTGIS (prostaglandin 12 (prostacyclin) synthase), MB (myoglobin), IL4 (interleukin 4), ANGPT1 (angiopoietin 1), ABCG8 (ATP-binding cassette, sub-family G (WHITE), member 8), CTSK (cathepsin K), PTGIR (prostaglandin 12 (prostacyclin) receptor (IP)), KCNJ11 (potassium inwardly-rectifying channel, subfamily J, member 11), INS (insulin), CRP (C-reactive protein, pentraxin-related), PDGFRB (platelet-derived growth factor receptor, beta polypeptide), CCNA2 (cyclin A2), PDGFB (platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog)), KCNJ5 (potassium inwardly-rectifying channel, subfamily J, member 5), KCNN3 (potassium intermediate/small conductance calcium-activated channel, subfamily N, member 3), CAPN10 (calpain 10), PTGES (prostaglandin E synthase), ADRA2B (adrenergic, alpha-2B-, receptor), ABCG5 (ATP-binding cassette, sub-family G (WHITE), member 5), PRDX2 (peroxiredoxin 2), CAPN5 (calpain 5), PARP14 (poly (ADP-ribose) polymerase family, member 14), MEX3C (mex-3 homolog C (C. elegans)), ACE angiotensin I converting enzyme (peptidyl-dipeptidase A) 1), TNF (tumor necrosis factor (TNF superfamily, member 2)), IL6 (interleukin 6 (interferon, beta 2)), STN (statin), SERPINE1 (serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1), ALB (albumin), ADIPOQ (adiponectin, C1Q and collagen domain containing), APOB (apolipoprotein B (including Ag(x) antigen)), APOE (apolipoprotein E), LEP (leptin), MTHFR (5,10-methylenetetrahydrofolate reductase (NADPH)), APOA1 (apolipoprotein A-I), EDN1 (endothelin 1), NPPB (natriuretic peptide precursor B), NOS3 (nitric oxide synthase 3 (endothelial cell)), PPARG (peroxisome proliferator-activated receptor gamma), PLAT (plasminogen activator, tissue), PTGS2 (prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)), CETP (cholesteryl ester transfer protein, plasma), AGTR1 (angiotensin II receptor, type 1), HMGCR (3-hydroxy-3-methylglutaryl-Coenzyme A reductase), IGF1 (insulin-like growth factor 1 (somatomedin C)), SELE (selectin E), REN

(renin), PPARA (peroxisome proliferator-activated receptor alpha), PON1 (paraoxonase 1), KNG1 (kininogen 1), CCL2 (chemokine (C-C motif) ligand 2), LPL (lipoprotein lipase), VWF (von Willebrand factor), F2 (coagulation factor II (thrombin)), ICAM1 (intercellular adhesion molecule 1), TGFB1 (transforming growth factor, beta 1), NPPA (natriuretic peptide precursor A), IL10 (interleukin 10), EPO (erythropoietin), SOD1 (superoxide dismutase 1, soluble), VCAM1 (vascular cell adhesion molecule 1), IFNG (interferon, gamma), LPA (lipoprotein, Lp(a)), MPO (myeloperoxidase), ESR1 (estrogen receptor 1), MAPK1 (mitogen-activated protein kinase 1), HP (haptoglobin), F3 (coagulation factor III (thromboplastin, tissue factor)), CST3 (cystatin C), COG2 (component of oligomeric golgi complex 2), MMP9 (matrix metallopeptidase 9 (gelatinase B, 92 kDa gelatinase, 92 kDa type IV collagenase)), SERPINC1 (serpin peptidase inhibitor, clade C (antithrombin), member 1), F8 (coagulation factor VIII, procoagulant component), HMOX1 (heme oxygenase (decycling) 1), APOC3 (apolipoprotein C-III), IL8 (interleukin 8), PROK1 (prokineticin 1), CBS (cystathionine-beta-synthase), NOS2 (nitric oxide synthase 2, inducible), TLR4 (toll-like receptor 4), SELP (selectin P (granule membrane protein 140 kDa, antigen CD62)), ABCA1 (ATP-binding cassette, sub-family A (ABC1), member 1), AGT (angiotensinogen (serpin peptidase inhibitor, clade A, member 8)), LDLR (low density lipoprotein receptor), GPT (glutamic-pyruvate transaminase (alanine aminotransferase)), VEGFA (vascular endothelial growth factor A), NR3C2 (nuclear receptor subfamily 3, group C, member 2), IL18 (interleukin 18 (interferon-gamma-inducing factor)), NOS1 (nitric oxide synthase 1 (neuronal)), NR3C1 (nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)), FGB (fibrinogen beta chain), HGF (hepatocyte growth factor (hepapoietin A; scatter factor)), IL1A (interleukin 1, alpha), RETN (resistin), AKT1 (v-akt murine thymoma viral oncogene homolog 1), LIPC (lipase, hepatic), HSPD1 (heat shock 60 kDa protein 1 (chaperonin)), MAPK14 (mitogen-activated protein kinase 14), SPP1 (secreted phosphoprotein 1), ITGB3 (integrin, beta 3 (platelet glycoprotein 111a, antigen CD61)), CAT (catalase), UTS2 (urotensin 2), THBD (thrombomodulin), F10 (coagulation factor X), CP (ceruloplasmin (ferroxidase)), TNFRSF11B (tumor necrosis factor receptor superfamily, member 11b), EDNRA (endothelin receptor type A), EGFR (epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)), MMP2 (matrix metallopeptidase 2 (gelatinase A, 72 kDa gelatinase, 72 kDa type IV collagenase)), PLG (plasminogen), NPY (neuropeptide Y), RHOD (ras homolog gene family, member D), MAPK8 (mitogen-activated protein kinase 8), MYC (v-myc myelocytomatosis viral oncogene

homolog (avian)), FN1 (fibronectin 1), CMA1 (chymase 1, mast cell), PLAU (plasminogen activator, urokinase), GNB3 (guanine nucleotide binding protein (G protein), beta polypeptide 3), ADRB2 (adrenergic, beta-2-, receptor, surface), APOA5 (apolipoprotein A-V), SOD2 (superoxide dismutase 2, mitochondrial), F5 (coagulation factor V (proaccelerin, labile factor)), VDR (vitamin D (1,25-dihydroxyvitamin D3) receptor), ALOX5 (arachidonate 5-lipoxygenase), HLA-DRB1 (major histocompatibility complex, class II, DR beta 1), PARP1 (poly (ADP-ribose) polymerase 1), CD40LG (CD40 ligand), PON2 (paraoxonase 2), AGER (advanced glycosylation end product-specific receptor), IRS1 (insulin receptor substrate 1), PTGS1 (prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)), ECE1 (endothelin converting enzyme 1), F7 (coagulation factor VII (serum prothrombin conversion accelerator)), URN (interleukin 1 receptor antagonist), EPHX2 (epoxide hydrolase 2, cytoplasmic), IGFBP1 (insulin-like growth factor binding protein 1), MAPK10 (mitogen-activated protein kinase 10), FAS (Fas (TNF receptor superfamily, member 6)), ABCB1 (ATP-binding cassette, sub-family B (MDR/TAP), member 1), JUN (jun oncogene), IGFBP3 (insulin-like growth factor binding protein 3), CD14 (CD14 molecule), PDE5A (phosphodiesterase 5A, cGMP-specific), AGTR2 (angiotensin II receptor, type 2), CD40 (CD40 molecule, TNF receptor superfamily member 5), LCAT (lecithin-cholesterol acyltransferase), CCR5 (chemokine (C-C motif) receptor 5), MMP1 (matrix metallopeptidase 1 (interstitial collagenase)), TIMP1 (TIMP metallopeptidase inhibitor 1), ADM (adrenomedullin), DYT10 (dystonia 10), STAT3 (signal transducer and activator of transcription 3 (acute-phase response factor)), MMP3 (matrix metallopeptidase 3 (stromelysin 1, progelatinase)), ELN (elastin), USF1 (upstream transcription factor 1), CFH (complement factor H), HSPA4 (heat shock 70 kDa protein 4), MMP12 (matrix metallopeptidase 12 (macrophage elastase)), MME (membrane metallo-endopeptidase), F2R (coagulation factor II (thrombin) receptor), SELL (selectin L), CTSB (cathepsin B), ANXA5 (annexin A5), ADRB1 (adrenergic, beta-1-, receptor), CYBA (cytochrome b-245, alpha polypeptide), FGA (fibrinogen alpha chain), GGT1 (gamma-glutamyltransferase 1), LIPG (lipase, endothelial), HIF1A (hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)), CXCR4 (chemokine (C-X-C motif) receptor 4), PROC (protein C (inactivator of coagulation factors Va and VIIIa)), SCARB1 (scavenger receptor class B, member 1), CD79A (CD79a molecule, immunoglobulin-associated alpha), PLTP (phospholipid transfer protein), ADD1 (adducin 1 (alpha)), FGG (fibrinogen gamma chain), SAA1 (serum amyloid A1), KCNH2

(potassium voltage-gated channel, subfamily H (eag-related), member 2), DPP4 (dipeptidyl-peptidase 4), G6PD (glucose-6-phosphate dehydrogenase), NPR1 (natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A)), VTN (vitronectin), KIAA0101 (KIAA0101), FOS (FBJ murine osteosarcoma viral oncogene homolog), TLR2 (toll-like receptor 2), PPIG (peptidylprolyl isomerase G (cyclophilin G)), IL1R1 (interleukin 1 receptor, type I), AR (androgen receptor), CYP1A1 (cytochrome P450, family 1, subfamily A, polypeptide 1), SERPINA1 (serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1), MTR (5-methyltetrahydrofolate-homocysteine methyltransferase), RBP4 (retinol binding protein 4, plasma), APOA4 (apolipoprotein A-IV), CDKN2A (cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)), FGF2 (fibroblast growth factor 2 (basic)), EDNRB (endothelin receptor type B), ITGA2 (integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)), CABIN1 (calcineurin binding protein 1), SHBG (sex hormone-binding globulin), HMGB1 (high-mobility group box 1), HSP90B2P (heat shock protein 90 kDa beta (Grp94), member 2 (pseudogene)), CYP3A4 (cytochrome P450, family 3, subfamily A, polypeptide 4), GJA1 (gap junction protein, alpha 1, 43 kDa), CAV1 (caveolin 1, caveolae protein, 22 kDa), ESR2 (estrogen receptor 2 (ER beta)), LTA (lymphotoxin alpha (TNF superfamily, member 1)), GDF15 (growth differentiation factor 15), BDNF (brain-derived neurotrophic factor), CYP2D6 (cytochrome P450, family 2, subfamily D, polypeptide 6), NGF (nerve growth factor (beta polypeptide)), SP1 (Sp1 transcription factor), TGIF1 (TGFB-induced factor homeobox 1), SRC (v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)), EGF (epidermal growth factor (beta-urogastrone)), PIK3CG (phosphoinositide-3-kinase, catalytic, gamma polypeptide), HLA-A (major histocompatibility complex, class I, A), KCNQ1 (potassium voltage-gated channel, KQT-like subfamily, member 1), CNR1 (cannabinoid receptor 1 (brain)), FBN1 (fibrillin 1), CHKA (choline kinase alpha), BEST1 (bestrophin 1), APP (amyloid beta (A4) precursor protein), CTNNB1 (catenin (cadherin-associated protein), beta 1, 88 kDa), IL2 (interleukin 2), CD36 (CD36 molecule (thrombospondin receptor)), PRKAB1 (protein kinase, AMP-activated, beta 1 non-catalytic subunit), TPO (thyroid peroxidase), ALDH7A1 (aldehyde dehydrogenase 7 family, member A1), CX3CR1 (chemokine (C-X3-C motif) receptor 1), TH (tyrosine hydroxylase), F9 (coagulation factor IX), GH1 (growth hormone 1), TF (transferrin), HFE (hemochromatosis), IL17A (interleukin 17A), PTEN (phosphatase and tensin homolog), GSTM1 (glutathione S-transferase mu 1), DMD (dystrophin), GATA4 (GATA binding protein 4), F13A1 (coagulation

factor XIII, A1 polypeptide), TTR (transthyretin), FABP4 (fatty acid binding protein 4, adipocyte), PON3 (paraoxonase 3), APOC1 (apolipoprotein C-I), INSR (insulin receptor), TNFRSF1B (tumor necrosis factor receptor superfamily, member 1B), HTR2A (5-hydroxytryptamine (serotonin) receptor 2A), CSF3 (colony stimulating factor 3 (granulocyte)), CYP2C9 (cytochrome P450, family 2, subfamily C, polypeptide 9), TXN (thioredoxin), CYP11B2 (cytochrome P450, family 11, subfamily B, polypeptide 2), PTH (parathyroid hormone), CSF2 (colony stimulating factor 2 (granulocyte-macrophage)), KDR (kinase insert domain receptor (a type III receptor tyrosine kinase)), PLA2G2A (phospholipase A2, group IIA (platelets, synovial fluid)), B2M (beta-2-microglobulin), THBS1 (thrombospondin 1), GCG (glucagon), RHOA (ras homolog gene family, member A), ALDH2 (aldehyde dehydrogenase 2 family (mitochondrial)), TCF7L2 (transcription factor 7-like 2 (T-cell specific, HMG-box)), BDKRB2 (bradykinin receptor B2), NFE2L2 (nuclear factor (erythroid-derived 2)-like 2), NOTCH1 (Notch homolog 1, translocation-associated (Drosophila)), UGT1A1 (UDP glucuronosyltransferase 1 family, polypeptide A1), IFNA1 (interferon, alpha 1), PPARD (peroxisome proliferator-activated receptor delta), SIRT1 (sirtuin (silent mating type information regulation 2 homolog) 1 (S. cerevisiae)), GNRH1 (gonadotropin-releasing hormone 1 (luteinizing-releasing hormone)), PAPPA (pregnancy-associated plasma protein A, pappalysin 1), ARR3 (arrestin 3, retinal (X-arrestin)), NPPC (natriuretic peptide precursor C), AHSP (alpha hemoglobin stabilizing protein), PTK2 (PTK2 protein tyrosine kinase 2), IL13 (interleukin 13), MTOR (mechanistic target of rapamycin (serine/threonine kinase)), ITGB2 (integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)), GSTT1 (glutathione S-transferase theta 1), IL6ST (interleukin 6 signal transducer (gp130, oncostatin M receptor)), CPB2 (carboxypeptidase B2 (plasma)), CYP1A2 (cytochrome P450, family 1, subfamily A, polypeptide 2), HNF4A (hepatocyte nuclear factor 4, alpha), SLC6A4 (solute carrier family 6 (neurotransmitter transporter, serotonin), member 4), PLA2G6 (phospholipase A2, group VI (cytosolic, calcium-independent)), TNFSF11 (tumor necrosis factor (ligand) superfamily, member 11), SLC8A1 (solute carrier family 8 (sodium/calcium exchanger), member 1), F2RL1 (coagulation factor II (thrombin) receptor-like 1), AKR1A1 (aldo-keto reductase family 1, member A1 (aldehyde reductase)), ALDH9A1 (aldehyde dehydrogenase 9 family, member A1), BGLAP (bone gamma-carboxyglutamate (gla) protein), MTTP (microsomal triglyceride transfer protein), MTRR (5-methyltetrahydrofolate-homocysteine methyltransferase reductase), SULT1A3 (sulfotransferase family, cytosolic, 1A, phenol-preferring, member 3),

RAGE (renal tumor antigen), C4B (complement component 4B (Chido blood group), P2RY12 (purinergic receptor P2Y, G-protein coupled, 12), RNLS (renalase, FAD-dependent amine oxidase), CREB1 (cAMP responsive element binding protein 1), POMC (proopiomelanocortin), RAC1 (ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)), LMNA (lamin NC), CD59 (CD59 molecule, complement regulatory protein), SCN5A (sodium channel, voltage-gated, type V, alpha subunit), CYP1B1 (cytochrome P450, family 1, subfamily B, polypeptide 1), MIF (macrophage migration inhibitory factor (glycosylation-inhibiting factor)), MMP13 (matrix metallopeptidase 13 (collagenase 3)), TIMP2 (TIMP metallopeptidase inhibitor 2), CYP19A1 (cytochrome P450, family 19, subfamily A, polypeptide 1), CYP21A2 (cytochrome P450, family 21, subfamily A, polypeptide 2), PTPN22 (protein tyrosine phosphatase, non-receptor type 22 (lymphoid)), MYH14 (myosin, heavy chain 14, non-muscle), MBL2 (mannose-binding lectin (protein C) 2, soluble (opsonic defect)), SELPLG (selectin P ligand), AOC3 (amine oxidase, copper containing 3 (vascular adhesion protein 1)), CTSL1 (cathepsin L1), PCNA (proliferating cell nuclear antigen), IGF2 (insulin-like growth factor 2 (somatomedin A)), ITGB1 (integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)), CAST (calpastatin), CXCL12 (chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)), IGHE (immunoglobulin heavy constant epsilon), KCNE1 (potassium voltage-gated channel, Isk-related family, member 1), TFRC (transferrin receptor (p90, CD71)), COL1A1 (collagen, type I, alpha 1), COL1A2 (collagen, type I, alpha 2), IL2RB (interleukin 2 receptor, beta), PLA2G10 (phospholipase A2, group X), ANGPT2 (angiopoietin 2), PROCR (protein C receptor, endothelial (EPCR)), NOX4 (NADPH oxidase 4), HAMP (hepcidin antimicrobial peptide), PTPN11 (protein tyrosine phosphatase, non-receptor type 11), SLC2A1 (solute carrier family 2 (facilitated glucose transporter), member 1), IL2RA (interleukin 2 receptor, alpha), CCL5 (chemokine (C-C motif) ligand 5), IRF1 (interferon regulatory factor 1), CFLAR (CASP8 and FADD-like apoptosis regulator), CALCA (calcitonin-related polypeptide alpha), EIF4E (eukaryotic translation initiation factor 4E), GSTP1 (glutathione S-transferase pi 1), JAK2 (Janus kinase 2), CYP3A5 (cytochrome P450, family 3, subfamily A, polypeptide 5), HSPG2 (heparan sulfate proteoglycan 2), CCL3 (chemokine (C-C motif) ligand 3), MYD88 (myeloid differentiation primary response gene (88)), VIP (vasoactive intestinal peptide), SOAT1 (sterol O-acyltransferase 1), ADRBK1 (adrenergic, beta, receptor kinase 1), NR4A2 (nuclear receptor subfamily 4, group A, member 2), MMP8 (matrix metallopeptidase 8

(neutrophil collagenase)), NPR2 (natriuretic peptide receptor B/guanylate cyclase B (atrionatriuretic peptide receptor B)), GCH1 (GTP cyclohydrolase 1), EPRS (glutamyl-prolyl-tRNA synthetase), PPARGC1A (peroxisome proliferator-activated receptor gamma, coactivator 1 alpha), F12 (coagulation factor XII (Hageman factor)), PECAM1 (platelet/endothelial cell adhesion molecule), CCL4 (chemokine (C-C motif) ligand 4), SERPINA3 (serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3), CASR (calcium-sensing receptor), GJA5 (gap junction protein, alpha 5, 40 kDa), FABP2 (fatty acid binding protein 2, intestinal), TTF2 (transcription termination factor, RNA polymerase II), PROS1 (protein S (alpha)), CTF1 (cardiotrophin 1), SGCB (sarcoglycan, beta (43 kDa dystrophin-associated glycoprotein)), YME1L1 (YME1-like 1 (S. cerevisiae)), CAMP (cathelicidin antimicrobial peptide), ZC3H12A (zinc finger CCCH-type containing 12A), AKR1B1 (aldo-keto reductase family 1, member B1 (aldose reductase)), DES (desmin), MMP7 (matrix metallopeptidase 7 (matrilysin, uterine)), AHR (aryl hydrocarbon receptor), CSF1 (colony stimulating factor 1 (macrophage)), HDAC9 (histone deacetylase 9), CTGF (connective tissue growth factor), KCNMA1 (potassium large conductance calcium-activated channel, subfamily M, alpha member 1), UGT1A (UDP glucuronosyltransferase 1 family, polypeptide A complex locus), PRKCA (protein kinase C, alpha), COMT (catechol-.beta.-methyltransferase), S100B (S100 calcium binding protein B), EGR1 (early growth response 1), PRL (prolactin), IL15 (interleukin 15), DRD4 (dopamine receptor D4), CAMK2G (calcium/calmodulin-dependent protein kinase II gamma), SLC22A2 (solute carrier family 22 (organic cation transporter), member 2), CCL11 (chemokine (C-C motif) ligand 11), PGF (B321 placental growth factor), THPO (thrombopoietin), GP6 (glycoprotein VI (platelet)), TACR1 (tachykinin receptor 1), NTS (neurotensin), HNF1A (HNF1 homeobox A), SST (somatostatin), KCND1 (potassium voltage-gated channel, Shal-related subfamily, member 1), LOC646627 (phospholipase inhibitor), TBXAS1 (thromboxane A synthase 1 (platelet)), CYP2J2 (cytochrome P450, family 2, subfamily J, polypeptide 2), TBXA2R (thromboxane A2 receptor), ADH1C (alcohol dehydrogenase 1C (class I), gamma polypeptide), ALOX12 (arachidonate 12-lipoxygenase), AHSG (alpha-2-HS-glycoprotein), BHMT (betaine-homocysteine methyltransferase), GJA4 (gap junction protein, alpha 4, 37 kDa), SLC25A4 (solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4), ACLY (ATP citrate lyase), ALOX5AP (arachidonate 5-lipoxygenase-activating protein), NUMA1 (nuclear mitotic apparatus protein 1), CYP27B1 (cytochrome P450, family 27,

subfamily B, polypeptide 1), CYSLTR2 (cysteinyl leukotriene receptor 2), SOD3 (superoxide dismutase 3, extracellular), LTC4S (leukotriene C4 synthase), UCN (urocortin), GHRL (ghrelin/obestatin prepropeptide), APOC2 (apolipoprotein C-II), CLEC4A (C-type lectin domain family 4, member A), KBTBD10 (kelch repeat and BTB (POZ) domain containing 10), TNC (tenascin C), TYMS (thymidylate synthetase), SHCl (SHC (Src homology 2 domain containing) transforming protein 1), LRP1 (low density lipoprotein receptor-related protein 1), SOCS3 (suppressor of cytokine signaling 3), ADH1B (alcohol dehydrogenase 1B (class I), beta polypeptide), KLK3 (kallikrein-related peptidase 3), HSD11B1 (hydroxysteroid (11-beta) dehydrogenase 1), VKORC1 (vitamin K epoxide reductase complex, subunit 1), SERPINB2 (serpin peptidase inhibitor, clade B (ovalbumin), member 2), TNS1 (tensin 1), RNF19A (ring finger protein 19A), EPOR (erythropoietin receptor), ITGAM (integrin, alpha M (complement component 3 receptor 3 subunit)), PITX2 (paired-like homeodomain 2), MAPK7 (mitogen-activated protein kinase 7), FCGR3A (Fc fragment of IgG, low affinity 111a, receptor (CD16a)), LEPR (leptin receptor), ENG (endoglin), GPX1 (glutathione peroxidase 1), GOT2 (glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)), HRH1 (histamine receptor H1), NR112 (nuclear receptor subfamily 1, group I, member 2), CRH (corticotropin releasing hormone), HTR1A (5-hydroxytryptamine (serotonin) receptor 1A), VDAC1 (voltage-dependent anion channel 1), HPSE (heparanase), SFTPD (surfactant protein D), TAP2 (transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)), RNF123 (ring finger protein 123), PTK2B (PTK2B protein tyrosine kinase 2 beta), NTRK2 (neurotrophic tyrosine kinase, receptor, type 2), IL6R (interleukin 6 receptor), ACHE (acetylcholinesterase (Yt blood group)), GLP1R (glucagon-like peptide 1 receptor), GHR (growth hormone receptor), GSR (glutathione reductase), NQO1 (NAD(P)H dehydrogenase, quinone 1), NR5A1 (nuclear receptor subfamily 5, group A, member 1), GJB2 (gap junction protein, beta 2, 26 kDa), SLC9A1 (solute carrier family 9 (sodium/hydrogen exchanger), member 1), MAOA (monoamine oxidase A), PCSK9 (proprotein convertase subtilisin/kexin type 9), FCGR2A (Fc fragment of IgG, low affinity IIa, receptor (CD32)), SERPINF1 (serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1), EDN3 (endothelin 3), DHFR (dihydrofolate reductase), GAS6 (growth arrest-specific 6), SMPD1 (sphingomyelin phosphodiesterase 1, acid lysosomal), UCP2 (uncoupling protein 2 (mitochondrial, proton carrier)), TFAP2A (transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)), C4BPA (complement component 4 binding protein, alpha),

SERPINF2 (serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 2), TYMP (thymidine phosphorylase), ALPP (alkaline phosphatase, placental (Regan isozyme)), CXCR2 (chemokine (C-X-C motif) receptor 2), SLC39A3 (solute carrier family 39 (zinc transporter), member 3), ABCG2 (ATP-binding cassette, sub-family G (WHITE), member 2), ADA (adenosine deaminase), JAK3 (Janus kinase 3), HSPA1A (heat shock 70 kDa protein 1A), FASN (fatty acid synthase), FGF1 (fibroblast growth factor 1 (acidic)), F11 (coagulation factor XI), ATP7A (ATPase, Cu++ transporting, alpha polypeptide), CR1 (complement component (3b/4b) receptor 1 (Knops blood group)), GFAP (glial fibrillary acidic protein), ROCK1 (Rho-associated, coiled-coil containing protein kinase 1), MECP2 (methyl CpG binding protein 2 (Rett syndrome)), MYLK (myosin light chain kinase), BCHE (butyrylcholinesterase), LIPE (lipase, hormone-sensitive), PRDX5 (peroxiredoxin 5), ADORA1 (adenosine A1 receptor), WRN (Werner syndrome, RecQ helicase-like), CXCR3 (chemokine (C-X-C motif) receptor 3), CD81 (CD81 molecule), SMAD7 (SMAD family member 7), LAMC2 (laminin, gamma 2), MAP3K5 (mitogen-activated protein kinase kinase kinase 5), CHGA (chromogranin A (parathyroid secretory protein 1)), IAPP (islet amyloid polypeptide), RHO (rhodopsin), ENPP1 (ectonucleotide pyrophosphatase/phosphodiesterase 1), PTHLH (parathyroid hormone-like hormone), NRG1 (neuregulin 1), VEGFC (vascular endothelial growth factor C), ENPEP (glutamyl aminopeptidase (aminopeptidase A)), CEBPB (CCAAT/enhancer binding protein (C/EBP), beta), NAGLU (N-acetylglucosaminidase, alpha-), F2RL3 (coagulation factor II (thrombin) receptor-like 3), CX3CL1 (chemokine (C-X3-C motif) ligand 1), BDKRB1 (bradykinin receptor B1), ADAMTS13 (ADAM metallopeptidase with thrombospondin type 1 motif, 13), ELANE (elastase, neutrophil expressed), ENPP2 (ectonucleotide pyrophosphatase/phosphodiesterase 2), CISH (cytokine inducible SH2-containing protein), GAST (gastrin), MYOC (myocilin, trabecular meshwork inducible glucocorticoid response), ATP1A2 (ATPase, Na+/K+ transporting, alpha 2 polypeptide), NF1 (neurofibromin 1), GJB1 (gap junction protein, beta 1, 32 kDa), MEF2A (myocyte enhancer factor 2A), VCL (vinculin), BMPR2 (bone morphogenetic protein receptor, type II (serine/threonine kinase)), TUBB (tubulin, beta), CDC42 (cell division cycle 42 (GTP binding protein, 25 kDa)), KRT18 (keratin 18), HSF1 (heat shock transcription factor 1), MYB (v-myb myeloblastosis viral oncogene homolog (avian)), PRKAA2 (protein kinase, AMP-activated, alpha 2 catalytic subunit), ROCK2 (Rho-associated, coiled-coil containing protein kinase 2), TFPI (tissue factor pathway

inhibitor (lipoprotein-associated coagulation inhibitor)), PRKG1 (protein kinase, cGMP-dependent, type I), BMP2 (bone morphogenetic protein 2), CTNND1 (catenin (cadherin-associated protein), delta 1), CTH (cystathionase (cystathionine gamma-lyase)), CTSS (cathepsin S), VAV2 (vav 2 guanine nucleotide exchange factor), NPY2R (neuropeptide Y receptor Y2), IGFBP2 (insulin-like growth factor binding protein 2, 36 kDa), CD28 (CD28 molecule), GSTA1 (glutathione S-transferase alpha 1), PPIA (peptidylprolyl isomerase A (cyclophilin A)), APOH (apolipoprotein H (beta-2-glycoprotein I)), S100A8 (S100 calcium binding protein A8), IL11 (interleukin 11), ALOX15 (arachidonate 15-lipoxygenase), FBLN1 (fibulin 1), NR1H3 (nuclear receptor subfamily 1, group H, member 3), SCD (stearoyl-CoA desaturase (delta-9-desaturase)), GIP (gastric inhibitory polypeptide), CHGB (chromogranin B (secretogranin 1)), PRKCB (protein kinase C, beta), SRD5A1 (steroid-5-alpha-reductase, alpha polypeptide 1 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 1)), HSD11B2 (hydroxysteroid (11-beta) dehydrogenase 2), CALCRL (calcitonin receptor-like), GALNT2 (UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-T2)), ANGPTL4 (angiopoietin-like 4), KCNN4 (potassium intermediate/small conductance calcium-activated channel, subfamily N, member 4), PIK3C2A (phosphoinositide-3-kinase, class 2, alpha polypeptide), HBEGF (heparin-binding EGF-like growth factor), CYP7A1 (cytochrome P450, family 7, subfamily A, polypeptide 1), HLA-DRB5 (major histocompatibility complex, class II, DR beta 5), BNIP3 (BCL2/adenovirus E1B 19 kDa interacting protein 3), GCKR (glucokinase (hexokinase 4) regulator), S100A12 (S100 calcium binding protein A12), PADI4 (peptidyl arginine deiminase, type IV), HSPA14 (heat shock 70 kDa protein 14), CXCR1 (chemokine (C-X-C motif) receptor 1), H19 (H19, imprinted maternally expressed transcript (non-protein coding)), KRTAP19-3 (keratin associated protein 19-3), IDDM2 (insulin-dependent diabetes mellitus 2), RAC2 (ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)), RYR1 (ryanodine receptor 1 (skeletal)), CLOCK (clock homolog (mouse)), NGFR (nerve growth factor receptor (TNFR superfamily, member 16)), DBH (dopamine beta-hydroxylase (dopamine beta-monooxygenase)), CHRNA4 (cholinergic receptor, nicotinic, alpha 4), CACNA1C (calcium channel, voltage-dependent, L type, alpha 1C subunit), PRKAG2 (protein kinase, AMP-activated, gamma 2 non-catalytic subunit), CHAT (choline acetyltransferase), PTGDS (prostaglandin D2 synthase 21 kDa (brain)), NR1H2 (nuclear receptor subfamily 1, group H, member 2), TEK (TEK tyrosine kinase, endothelial), VEGFB (vascular endothelial

growth factor B), MEF2C (myocyte enhancer factor 2C), MAPKAPK2 (mitogen-activated protein kinase-activated protein kinase 2), TNFRSF11A (tumor necrosis factor receptor superfamily, member 11a, NFKB activator), HSPA9 (heat shock 70 kDa protein 9 (mortalin)), CYSLTR1 (cysteinyl leukotriene receptor 1), MAT1A (methionine adenosyltransferase I, alpha), OPRL1 (opiate receptor-like 1), IMPA1 (inositol(myo)-1(or 4)-monophosphatase 1), CLCN2 (chloride channel 2), DLD (dihydrolipoamide dehydrogenase), PSMA6 (proteasome (prosome, macropain) subunit, alpha type, 6), PSMB8 (proteasome (prosome, macropain) subunit, beta type, 8 (large multifunctional peptidase 7)), CHI3L1 (chitinase 3-like 1 (cartilage glycoprotein-39)), ALDH1B1 (aldehyde dehydrogenase 1 family, member B1), PARP2 (poly (ADP-ribose) polymerase 2), STAR (steroidogenic acute regulatory protein), LBP (lipopolysaccharide binding protein), ABCC6 (ATP-binding cassette, sub-family C(CFTR/MRP), member 6), RGS2 (regulator of G-protein signaling 2, 24 kDa), EFNB2 (ephrin-B2), GJB6 (gap junction protein, beta 6, 30 kDa), APOA2 (apolipoprotein A-II), AMPD1 (adenosine monophosphate deaminase 1), DYSF (dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)), FDFT1 (farnesyl-diphosphate farnesyltransferase 1), EDN2 (endothelin 2), CCR6 (chemokine (C-C motif) receptor 6), GJB3 (gap junction protein, beta 3, 31 kDa), IL1RL1 (interleukin 1 receptor-like 1), ENTPD1 (ectonucleoside triphosphate diphosphohydrolase 1), BBS4 (Bardet-Biedl syndrome 4), CELSR2 (cadherin, EGF LAG seven-pass G-type receptor 2 (flamingo homolog, Drosophila)), F11R (F11 receptor), RAPGEF3 (Rap guanine nucleotide exchange factor (GEF) 3), HYAL1 (hyaluronoglucosaminidase 1), ZNF259 (zinc finger protein 259), ATOX1 (ATX1 antioxidant protein 1 homolog (yeast)), ATF6 (activating transcription factor 6), KHK (ketohexokinase (fructokinase)), SAT1 (spermidine/spermine N1-acetyltransferase 1), GGH (gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase)), TIMP4 (TIMP metallopeptidase inhibitor 4), SLC4A4 (solute carrier family 4, sodium bicarbonate cotransporter, member 4), PDE2A (phosphodiesterase 2A, cGMP-stimulated), PDE3B (phosphodiesterase 3B, cGMP-inhibited), FADS1 (fatty acid desaturase 1), FADS2 (fatty acid desaturase 2), TMSB4X (thymosin beta 4, X-linked), TXNIP (thioredoxin interacting protein), LIMS1 (LIM and senescent cell antigen-like domains 1), RHOB (ras homolog gene family, member B), LY96 (lymphocyte antigen 96), FOXO1 (forkhead box O1), PNPLA2 (patatin-like phospholipase domain containing 2), TRH (thyrotropin-releasing hormone), GJC1 (gap junction protein, gamma 1, 45 kDa), SLC17A5 (solute carrier family 17 (anion/sugar transporter), member 5),

FTO (fat mass and obesity associated), GJD2 (gap junction protein, delta 2, 36 kDa), PSRC1 (proline/serine-rich coiled-coil 1), CASP12 (caspase 12 (gene/pseudogene)), GPBAR1 (G protein-coupled bile acid receptor 1), PXK (PX domain containing serine/threonine kinase), IL33 (interleukin 33), TRIB1 (tribbles homolog 1 (Drosophila)), PBX4 (pre-B-cell leukemia homeobox 4), NUPR1 (nuclear protein, transcriptional regulator, 1), 15-Sep(15 kDa selenoprotein), CILP2 (cartilage intermediate layer protein 2), TERC (telomerase RNA component), GGT2 (gamma-glutamyltransferase 2), MT-CO1 (mitochondrially encoded cytochrome c oxidase I), and UOX (urate oxidase, pseudogene). In an additional embodiment, the chromosomal sequence may further be selected from Pon1 (paraoxonase 1), LDLR (LDL receptor), ApoE (Apolipoprotein E), Apo B-100 (Apolipoprotein B-100), ApoA (Apolipoprotein(a)), ApoA1 (Apolipoprotein A1), CBS (Cystathione B-synthase), Glycoprotein IIb/IIb, MTHRF (5,10-methylenetetrahydrofolate reductase (NADPH), and combinations thereof. In one iteration, the chromosomal sequences and proteins encoded by chromosomal sequences involved in cardiovascular disease may be chosen from Cacna1C, Sod1, Pten, Ppar(alpha), Apo E, Leptin, and combinations thereof. The text herein accordingly provides exemplary targets as to CRISPR or CRISPR-Cas systems or complexes.

[01193] Each of these patents, patent publications, and applications, and all documents cited therein or during their prosecution ("appln cited documents") and all documents cited or referenced in the appln cited documents, together with any instructions, descriptions, product specifications, and product sheets for any products mentioned therein or in any document therein and incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention. All documents (e.g., these patents, patent publications and applications and the appln cited documents) are incorporated herein by reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

[01194] Kits

[01195] In another aspect, the invention is directed to kit and kit of parts. The terms "kit of parts" and "kit" as used throughout this specification refer to a product containing components necessary for carrying out the specified methods (e.g., methods for detecting, quantifying or isolating immune cells as taught herein), packed so as to allow their transport and storage. Materials suitable for packing the components comprised in a kit include crystal, plastic (e.g., polyethylene, polypropylene, polycarbonate), bottles, flasks, vials, ampules, paper, envelopes,

or other types of containers, carriers or supports. Where a kit comprises a plurality of components, at least a subset of the components (e.g., two or more of the plurality of components) or all of the components may be physically separated, e.g., comprised in or on separate containers, carriers or supports. The components comprised in a kit may be sufficient or may not be sufficient for carrying out the specified methods, such that external reagents or substances may not be necessary or may be necessary for performing the methods, respectively. Typically, kits are employed in conjunction with standard laboratory equipment, such as liquid handling equipment, environment (e.g., temperature) controlling equipment, analytical instruments, etc. In addition to the recited binding agents(s) as taught herein, such as for example, antibodies, hybridization probes, amplification and/or sequencing primers, optionally provided on arrays or microarrays, the present kits may also include some or all of solvents, buffers (such as for example but without limitation histidine-buffers, citrate-buffers, succinate-buffers, acetate-buffers, phosphate-buffers, formate buffers, benzoate buffers, TRIS (Tris(hydroxymethyl)-aminomethan) buffers or maleate buffers, or mixtures thereof), enzymes (such as for example but without limitation thermostable DNA polymerase), detectable labels, detection reagents, and control formulations (positive and/or negative), useful in the specified methods. Typically, the kits may also include instructions for use thereof, such as on a printed insert or on a computer readable medium. The terms may be used interchangeably with the term “article of manufacture”, which broadly encompasses any man-made tangible structural product, when used in the present context.

Use of Type VII Systems in Non-Animal Organisms

[01196] The Type VII CRISPR effector protein system(s) (e.g., single or multiplexed) can be used in conjunction with recent advances in crop genomics. The systems described herein can be used to perform efficient and cost-effective plant gene or genome interrogation or editing or manipulation—for instance, for rapid investigation and/or selection and/or interrogations and/or comparison and/or manipulations and/or transformation of plant genes or genomes; e.g., to create, identify, develop, optimize, or confer trait(s) or characteristic(s) to plant(s) or to transform a plant genome. Accordingly, there can be be improved production of plants, new plants with new combinations of traits or characteristics or new plants with enhanced traits. The CRISPR effector protein system(s) can be used with regard to plants in Site-Directed Integration (SDI) or Gene Editing (GE) or any Near Reverse Breeding (NRB) or Reverse Breeding (RB) techniques. Aspects of utilizing the herein described CRISPR effector

protein systems may be analogous to the use of the CRISPR-Cas (e.g. CRISPR-Cas9) system in plants, and mention is made of the University of Arizona website "CRISPR-PLANT" (http://www.genome.arizona.edu/crispr/) (supported by Penn State and AGI). Embodiments of the invention can be used with haploid induction. For example, a corn line capable of making pollen able to trigger haploid induction is transformed with a CRISPR system programmed to target genes related to desirable traits. The pollen is used to transfer the CRISPR system to other corn varieties otherwise resistant to CRISPR transfer. In certain embodiments, the CRISPR-carrying corn pollen can edit the DNA of wheat. Embodiments of the invention can be used in genome editing in plants or where RNAi or similar genome editing techniques have been used previously; see, e.g., Nekrasov, "Plant genome editing made easy: targeted mutagenesis in model and crop plants using the CRISPR-Cas system," Plant Methods 2013, 9:39 (doi:10.1186/1746-4811-9-39); Brooks, "Efficient gene editing in tomato in the first generation using the CRISPR-Cas9 system," Plant Physiology September 2014 pp 114.247577; Shan, "Targeted genome modification of crop plants using a CRISPR-Cas system," Nature Biotechnology 31, 686-688 (2013); Feng, "Efficient genome editing in plants using a CRISPR/Cas system," Cell Research (2013) 23:1229–1232. doi:10.1038/cr.2013.114; published online 20 August 2013; Xie, "RNA-guided genome editing in plants using a CRISPR-Cas system," Mol Plant. 2013 Nov;6(6):1975-83. doi: 10.1093/mp/sst119. Epub 2013 Aug 17; Xu, "Gene targeting using the Agrobacterium tumefaciens-mediated CRISPR-Cas system in rice," Rice 2014, 7:5 (2014), Zhou et al., "Exploiting SNPs for biallelic CRISPR mutations in the outcrossing woody perennial Populus reveals 4-coumarate: CoA ligase specificity and Redundancy," New Phytologist (2015) (Forum) 1-4 (available online only at www.newphytologist.com); Caliando et al, "Targeted DNA degradation using a CRISPR device stably carried in the host genome, NATURE COMMUNICATIONS 6:6989, DOI: 10.1038/ncomms7989, www.nature.com/naturecommunications DOI: 10.1038/ncomms7989; US Patent No. 6,603,061 - Agrobacterium-Mediated Plant Transformation Method; US Patent No. 7,868,149 - Plant Genome Sequences and Uses Thereof and US 2009/0100536 - Transgenic Plants with Enhanced Agronomic Traits, all the contents and disclosure of each of which are herein incorporated by reference in their entirety. In the practice of the invention, the contents and disclosure of Morrell et al "Crop genomics: advances and applications," Nat Rev Genet. 2011 Dec 29;13(2):85-96; each of which is incorporated by reference herein including as to how herein embodiments may be used as to plants. Accordingly, reference

herein to animal cells may also apply, mutatis mutandis, to plant cells unless otherwise apparent; and, the enzymes herein having reduced off-target effects and systems employing such enzymes can be used in plant applications, including those mentioned herein.

[01197] The CRISPR-Cas systems (e.g. Type VII CRISPR-Cas systems) can be used to modify a non-animal or cell thereof. Methods of delivering a Type VII CRISPR-Cas system or component thereof to a cell and/or organism are described in greater detail elsewhere herein. See e.g. "Modified Cells and Organisms herein" In some embodiments, the method can include delivering a Type VII CRISPR-Cas system or component thereof to a non-animal cell and expressing one or more components of said system therein. In some embodiments, the method further includes expanding modified cells and/or propagating the modified organisms.

[01198] The methods for genome editing using the Type VII CRISPR system as described herein can be used to confer desired traits on essentially any plant. A wide variety of plants and plant cell systems may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation methods mentioned above. In some embodiments, the non-animal organism can be modified to generate an exogenous product and/or increase or decrease production of an endogenous product. Target plants and plant cells for engineered are described in greater detail elsewhere herein (see e.g. "Modified Cells and Organisms").

[01199] In some embodiments, the method can include delivering a Type VII CRISPR-Cas expression system that includes at least: (a) a nucleotide sequence encoding a guide RNA (gRNA) that hybridizes with a target sequence in a plant, and wherein the guide RNA comprises a guide sequence and a direct repeat sequence, and(b) a nucleotide sequence encoding a Cas protein, where components (a) or (b) are located on the same or on different constructs, and whereby the different nucleotide sequences can be under control of the same or a different regulatory element operable in a plant cell or non-animal organism cell.

[01200] In some embodiments, the modified plants, algae, yeast or other non-animal organisms can be used to produce a desirable gene product. The desirable gene product can then be harvested after production and used accordingly.

[01201] In particular embodiments, the polynucleotide modifying agents and system can be used for visualization of genetic element dynamics. For example, CRISPR imaging can visualize either repetitive or non-repetitive genomic sequences, report telomere length change

and telomere movements and monitor the dynamics of gene loci throughout the cell cycle (Chen et al., Cell, 2013). These methods may also be applied to plants.

[01202] Other applications of the systems, and preferably the systems described herein, are the targeted gene disruption positive-selection screening *in vitro* and *in vivo* (Malina et al., Genes and Development, 2013). These methods may also be applied to plants.

[01203] In particular embodiments, fusion of inactive Cas endonucleases with histone-modifying enzymes can introduce custom changes in the complex epigenome (Rusk et al., Nature Methods, 2014). These methods may also be applied to plants.

[01204] In particular embodiments, the systems, and preferably the systems described herein, can be used to purify a specific portion of the chromatin and identify the associated proteins, thus elucidating their regulatory roles in transcription (Waldrip et al., Epigenetics, 2014). These methods may also be applied to plants.

[01205] In particular embodiments, present invention can be used as a therapy for virus removal in plant systems as it is able to cleave both viral DNA and RNA. Previous studies in human systems have demonstrated the success of utilizing CRISPR in targeting the single strand RNA virus, hepatitis C (A. Price, et al., Proc. Natl. Acad. Sci, 2015) as well as the double stranded DNA virus, hepatitis B (V. Ramanan, et al., Sci. Rep, 2015). These methods may also be adapted for using the systems in plants.

[01206] In particular embodiments, present invention could be used to alter genome complexity. In further particular embodiment, the systems, and preferably the systems described herein, can be used to disrupt or alter chromosome number and generate haploid plants, which only contain chromosomes from one parent. Such plants can be induced to undergo chromosome duplication and converted into diploid plants containing only homozygous alleles (Karimi-Ashtiyani et al., PNAS, 2015; Anton et al., Nucleus, 2014). These methods may also be applied to plants.

[01207] The polynucleotide modifying agent(s) and systems can be used to generate loss of function plants, algae, yeast, and other non-animal organisms, which can allow for functional analysis of genomic material. Ma et al. (Mol Plant. 2015 Aug 3;8(8):1274-84. doi: 10.1016/j.molp.2015.04.007) reports robust CRISPR-Cas9 vector system, utilizing a plant codon optimized Cas9 gene, for convenient and high-efficiency multiplex genome editing in monocot and dicot plants. Ma et al. designed PCR-based procedures to rapidly generate multiple sgRNA expression cassettes, which can be assembled into the binary CRISPR-Cas9

vectors in one round of cloning by Golden Gate ligation or Gibson Assembly. With this system, Ma et al. edited 46 target sites in rice with an average 85.4% rate of mutation, mostly in biallelic and homozygous status. Ma et al. provide examples of loss-of-function gene mutations in T0 rice and T1Arabidopsis plants by simultaneous targeting of multiple (up to eight) members of a gene family, multiple genes in a biosynthetic pathway, or multiple sites in a single gene. The methods of Ma et al. can be applied to the polynucleotide modifying agent(s) and systems described herein.

[01208] In plants, pathogens are often host-specific. For example, Fusarium oxysporum f. sp. lycopersici causes tomato wilt but attacks only tomato, and F. oxysporum f. dianthii Puccinia graminis f. sp. tritici attacks only wheat. Plants have existing and induced defenses to resist most pathogens. Mutations and recombination events across plant generations lead to genetic variability that gives rise to susceptibility, especially as pathogens reproduce with more frequency than plants. In plants there can be non-host resistance, e.g., the host and pathogen are incompatible. There can also be Horizontal Resistance, e.g., partial resistance against all races of a pathogen, typically controlled by many genes and Vertical Resistance, e.g., complete resistance to some races of a pathogen but not to other races, typically controlled by a few genes. In a Gene-for-Gene level, plants and pathogens evolve together, and the genetic changes in one balance changes in other. Accordingly, using Natural Variability, breeders combine most useful genes for Yield, Quality, Uniformity, Hardiness, Resistance. The sources of resistance genes include native or foreign Varieties, Heirloom Varieties, Wild Plant Relatives, and Induced Mutations, e.g., treating plant material with mutagenic agents. The polynucleotide modifying agents and systems can be used to induce mutations, to analyze the genome of sources of resistance genes, and in varieties having desired characteristics or traits employ the present invention to induce the rise of resistance genes, with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs. Further, the modifying agents and systems described herein can be used to generate plants with one or more disease resistant genes or alleles.

[01209] Similarly, the polynucleotide modifying agents and systems described herein can be used to induce mutations to allow for genome wide screening for mutations, alleles, and variants that have a desired characteristic (e.g. heat tolerance, cold tolerance, fast growth, pest resistance, etc.) and also used to generate plants with the identified and desired allele(s).

[01210] In some embodiments, the polynucleotide modifying agents and systems described herein can be used to generate non-animal organism model systems of animals. Modified non-animal organisms or cells thereof can be modified to express one or more heterologous genes, such as genes from a human or non-human animal. Such model systems can be used to determine response to environmental toxins, pharmaceutical agents, or other stimuli. Other uses for such model systems will be appreciated by those of ordinary skill in the art.

Improved Non-animal Organisms

[01211] The methods described herein can result in the generation of "improved plants, algae, fungi, yeast, etc." in that they have one or more desirable traits compared to the wildtype plant. In particular embodiments, the plants, algae, fungi, yeast, etc., cells or parts obtained are transgenic plants, comprising an exogenous DNA sequence incorporated into the genome of all or part of the cells. In particular embodiments, non-transgenic genetically modified plants, algae, fungi, yeast, etc., parts or cells are obtained, in that no exogenous DNA sequence is incorporated into the genome of any of the cells of the plant. In such embodiments, the improved plants, algae, fungi, yeast, etc. are non-transgenic. Where only the modification of an endogenous gene is ensured and no foreign genes are introduced or maintained in the plant, algae, fungi, yeast, etc. genome, the resulting genetically modified crops contain no foreign genes and can thus basically be considered non-transgenic. The different applications of the systems for plant, algae, fungi, yeast, etc. genome editing include, but are not limited to: introduction of one or more foreign genes to confer an agricultural trait of interest; editing of endogenous genes to confer an agricultural trait of interest; modulating of endogenous genes by the systems to confer an agricultural trait of interest. Exemplary genes conferring agronomic traits include, but are not limited to, genes that confer resistance to pests or diseases; genes involved in plant diseases, such as those listed in WO 2013046247; genes that confer resistance to herbicides, fungicides, or the like; genes involved in (abiotic) stress tolerance. Other aspects of the use of the systems include, but are not limited to: create (male) sterile plants; increasing the fertility stage in plants/algae etc.; generate genetic variation in a crop of interest; affect fruit-ripening; increasing storage life of plants/algae etc.; reducing allergen in plants/algae etc.; ensure a value-added trait (e.g. nutritional improvement); Screening methods for endogenous genes of interest; biofuel, fatty acid, organic acid, etc. production.

[01212] Also described here are modified non-animal organisms (e.g. plants, algae, and yeast cells) obtainable and obtained by the methods provided herein that can be improved in at

least one aspect as compared to an unmodified plant. The improved non-animal organisms obtained by the methods described herein may be useful in one or more fields (e.g. food or feed production) through expression of genes or alleles which, for instance ensure tolerance to infectious agents, pests, herbicides, drought, low or high temperatures, excessive water, toxins, etc.

[01213] The improved plants obtained by the methods described herein, especially crops and algae may be useful in food or feed production through expression of, for instance, higher protein, carbohydrate, nutrient or vitamin levels than would normally be seen in the wildtype. In this regard, improved plants, especially pulses and tubers are preferred.

[01214] Improved algae or other plants such as rape may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol), for instance. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[01215] Also described herein are improved parts of a plant. Plant parts include, but are not limited to, leaves, stems, roots, tubers, seeds, endosperm, ovule, and pollen. Plant parts as envisaged herein may be viable, nonviable, regeneratable, and/or non- regeneratable. The improved part of the plant can, for example, result in earlier fruit, higher content of one or more molecules involved in fruit taste, color, maturity, ripening, etc. or have other desired characteristics. In one embodiment, the method described in Soyk et al. (Nat Genet. 2017 Jan;49(1):162-168), which used CRISPR-Cas9 mediated mutation targeting flowering repressor SP5G in tomatoes to produce early yield tomatoes can be modified and adapted for use with the polynucleotide modifying agent(s) and systems thereof described herein.

[01216] The Cas based CRISPR systems provided herein can be used to introduce targeted double-strand or single-strand breaks and/or to introduce gene activator and or repressor systems and without being limitative, can be used for gene targeting, gene replacement, targeted mutagenesis, targeted deletions or insertions, targeted inversions and/or targeted translocations. By co-expression of multiple targeting RNAs directed to achieve multiple modifications in a single cell, multiplexed genome modification can be ensured. This technology can be used to high-precision engineering of plants with improved characteristics, including 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.

[01217] In particular embodiments, the Cas CRISPR system as described herein is used to introduce targeted double-strand breaks (DSB) in an endogenous DNA sequence. The DSB activates cellular DNA repair pathways, which can be harnessed to achieve desired DNA sequence modifications near the break site. This is of interest where the inactivation of endogenous genes can confer or contribute to a desired trait. In particular embodiments, homologous recombination with a template sequence is promoted at the site of the DSB, in order to introduce a gene of interest.

[01218] In particular embodiments, the Cas CRISPR system may be used as a generic nucleic acid binding protein with fusion to or being operably linked to a functional domain for activation and/or repression of endogenous plant genes. Exemplary functional domains may include but are not limited to translational initiator, translational activator, translational repressor, nucleases, in particular ribonucleases, a spliceosome, beads, a light inducible/controllable domain or a chemically inducible/controllable domain. Typically, in these embodiments, the Cas protein comprises at least one mutation, such that it has no more than 5% of the activity of the Cas protein not having the at least one mutation; the guide RNA comprises a guide sequence capable of hybridizing to a target sequence.

[01219] The methods described herein generally result in the generation of "improved plants" in that they have one or more desirable traits compared to the wildtype plant. In particular embodiments, the plants, plant cells or plant parts obtained are transgenic plants, comprising an exogenous DNA sequence incorporated into the genome of all or part of the cells of the plant. In particular embodiments, non-transgenic genetically modified plants, plant parts or cells are obtained, in that no exogenous DNA sequence is incorporated into the genome of any of the plant cells of the plant. In such embodiments, the improved plants are non-transgenic. Where only the modification of an endogenous gene is ensured and no foreign genes are introduced or maintained in the plant genome, the resulting genetically modified crops contain no foreign genes and can thus basically be considered non-transgenic. The different applications of the Cas CRISPR system for plant genome editing are described more in detail below and elsewhere herein. Such applications can include, without limitation, introduction of one or more foreign genes to confer an agricultural trait of interest; editing of endogenous genes to confer an agricultural trait of interest; modulating of endogenous genes by the Cas CRISPR system to confer an agricultural trait of interest. Exemplary genes conferring agronomic traits include, but are not limited to genes that confer resistance to pests

or diseases; genes involved in plant diseases, such as those listed in WO 2013046247; genes that confer resistance to herbicides, fungicides, or the like; genes involved in (abiotic) stress tolerance. Other aspects of the use of the CRISPR-Cas system include, but are not limited to: create (male) sterile plants; increasing the fertility stage in plants/algae etc; generate genetic variation in a crop of interest; affect fruit-ripening; increasing storage life of plants/algae etc; reducing allergen in plants/algae etc; ensure a value added trait (e.g. nutritional improvement); Screening methods for endogenous genes of interest; biofuel, fatty acid, organic acid, etc production.

Introduction of one or more foreign genes to confer an agricultural trait of interest

[01220] The invention provides methods of genome editing or modifying sequences associated with or at a target locus of interest wherein the method comprises introducing a Cas effector protein complex into a plant cell, whereby the Cas effector protein complex effectively functions to integrate a DNA insert, e.g. encoding a foreign gene of interest, into the genome of the plant cell. In preferred embodiments the integration of the DNA insert is facilitated by HR with an exogenously introduced DNA template or repair template. Typically, the exogenously introduced DNA template or repair template is delivered together with the Cas effector protein complex or one component or a polynucleotide vector for expression of a component of the complex.

[01221] The Cas CRISPR systems provided herein allow for targeted gene delivery. It has become increasingly clear that the efficiency of expressing a gene of interest is to a great extent determined by the location of integration into the genome. The present methods allow for targeted integration of the foreign gene into a desired location in the genome. The location can be selected based on information of previously generated events or can be selected by methods disclosed elsewhere herein.

[01222] In particular embodiments, the methods provided herein include (a) introducing into the cell a Cas CRISPR complex comprising a guide RNA, comprising a direct repeat and a guide sequence, wherein the guide sequence hybridizes to a target sequence that is endogenous to the plant cell; (b) introducing into the plant cell a Cas effector molecule which complexes with the guide RNA when the guide sequence hybridizes to the target sequence and induces a double strand break at or near the sequence to which the guide sequence is targeted; and (c) introducing into the cell a nucleotide sequence encoding an HDR repair template which encodes the gene of interest and which is introduced into the location of the DS break as a

result of HDR. In particular embodiments, the step of introducing can include delivering to the plant cell one or more polynucleotides encoding Cas effector protein, the guide RNA and the repair template. In particular embodiments, the polynucleotides are delivered into the cell by a DNA virus (e.g., a geminivirus) or an RNA virus (e.g., a tobravirus). In particular embodiments, the introducing steps include delivering to the plant cell a T-DNA containing one or more polynucleotide sequences encoding the Cas effector protein, the guide RNA and the repair template, where the delivering is via Agrobacterium. The nucleic acid sequence encoding the Cas effector protein can be operably linked to a promoter, such as a constitutive promoter (e.g., a cauliflower mosaic virus 35S promoter), or a cell specific or inducible promoter. In particular embodiments, the polynucleotide is introduced by microprojectile bombardment. In particular embodiments, the method further includes screening the plant cell after the introducing steps to determine whether the repair template i.e. the gene of interest has been introduced. In particular embodiments, the methods include the step of regenerating a plant from the plant cell. In further embodiments, the methods include cross breeding the plant to obtain a genetically desired plant lineage. Examples of foreign genes encoding a trait of interest are listed below.

Editing of endogenous genes to confer an agricultural trait of interest

[01223] The invention provides methods of genome editing or modifying sequences associated with or at a target locus of interest wherein the method comprises introducing a Cas effector protein complex into a plant cell, whereby the Cas complex modifies the expression of an endogenous gene of the plant. This can be achieved in different ways, In particular embodiments, the elimination of expression of an endogenous gene is desirable and the Cas CRISPR complex is used to target and cleave an endogenous gene so as to modify gene expression. In these embodiments, the methods provided herein include (a) introducing into the plant cell a Cas CRISPR complex comprising a guide RNA, comprising a direct repeat and a guide sequence, wherein the guide sequence hybridizes to a target sequence within a gene of interest in the genome of the plant cell; and (b) introducing into the cell a Cas effector protein, which upon binding to the guide RNA comprises a guide sequence that is hybridized to the target sequence, ensures a double strand break at or near the sequence to which the guide sequence is targeted; In particular embodiments, the step of introducing can include delivering to the plant cell one or more polynucleotides encoding Cas effector protein and the guide RNA.

[01224] In particular embodiments, the polynucleotides are delivered into the cell by a DNA virus (e.g., a geminivirus) or an RNA virus (e.g., a tobravirus). In particular embodiments, the introducing steps include delivering to the plant cell a T-DNA containing one or more polynucleotide sequences encoding the Cas effector protein and the guide RNA, where the delivering is via Agrobacterium. The polynucleotide sequence encoding the components of the Cas CRISPR system can be operably linked to a promoter, such as a constitutive promoter (e.g., a cauliflower mosaic virus 35S promoter), or a cell specific or inducible promoter. In particular embodiments, the polynucleotide is introduced by microprojectile bombardment. In particular embodiments, the method further includes screening the plant cell after the introducing steps to determine whether the expression of the gene of interest has been modified. In particular embodiments, the methods include the step of regenerating a plant from the plant cell. In further embodiments, the methods include cross breeding the plant to obtain a genetically desired plant lineage.

[01225] In particular embodiments of the methods described above, disease resistant crops are obtained by targeted mutation of disease susceptibility genes or genes encoding negative regulators (e.g. Mlo gene) of plant defense genes. In a particular embodiment, herbicide-tolerant crops are generated by targeted substitution of specific nucleotides in plant genes such as those encoding acetolactate synthase (ALS) and protoporphyrinogen oxidase (PPO). In particular embodiments drought and salt tolerant crops by targeted mutation of genes encoding negative regulators of abiotic stress tolerance, low amylose grains by targeted mutation of Waxy gene, rice or other grains with reduced rancidity by targeted mutation of major lipase genes in aleurone layer, etc. In particular embodiments. A more extensive list of endogenous genes encoding a traits of interest are listed below.

Modulating of endogenous genes by the CRISPR system to confer an agricultural trait of interest

[01226] Also provided herein are methods for modulating (i.e. activating or repressing) endogenous gene expression using the Cas protein provided herein. Such methods make use of distinct RNA sequence(s) which are targeted to the plant genome by the Cas complex. More particularly the distinct RNA sequence(s) bind to two or more adaptor proteins (e.g. aptamers) whereby each adaptor protein is associated with one or more functional domains and wherein at least one of the one or more functional domains associated with the adaptor protein have one or more activities comprising methylase activity, demethylase activity, transcription activation

activity, transcription repression activity, transcription release factor activity, histone modification activity, DNA integration activity RNA cleavage activity, DNA cleavage activity or nucleic acid binding activity; The functional domains are used to modulate expression of an endogenous plant gene so as to obtain the desired trait. Typically, in these embodiments, the Cas effector protein has one or more mutations such that it has no more than 5% of the nuclease activity.

[01227] In particular embodiments, the methods provided herein include the steps of (a) introducing into the cell a Cas CRISPR complex comprising a guide RNA, comprising a direct repeat and a guide sequence, wherein the guide sequence hybridizes to a target sequence that is endogenous to the plant cell; (b) introducing into the plant cell a Cas effector molecule which complexes with the guide RNA when the guide sequence hybridizes to the target sequence; and wherein either the guide RNA is modified to comprise a distinct RNA sequence (aptamer) binding to a functional domain and/or the Cas effector protein is modified in that it is linked to a functional domain. In particular embodiments, the step of introducing can include delivering to the plant cell one or more polynucleotides encoding the (modified) Cas effector protein and the (modified) guide RNA. The details the components of the Cas CRISPR system for use in these methods are described elsewhere herein.

[01228] In particular embodiments, the polynucleotides are delivered into the cell by a DNA virus (e.g., a geminivirus) or an RNA virus (e.g., a tobravirus). In particular embodiments, the introducing steps include delivering to the plant cell a T-DNA containing one or more polynucleotide sequences encoding the Cas effector protein and the guide RNA, where the delivering is via Agrobacterium. The nucleic acid sequence encoding the one or more components of the Cas CRISPR system can be operably linked to a promoter, such as a constitutive promoter (e.g., a cauliflower mosaic virus 35S promoter), or a cell specific or inducible promoter. In particular embodiments, the polynucleotide is introduced by microprojectile bombardment. In particular embodiments, the method further includes screening the plant cell after the introducing steps to determine whether the expression of the gene of interest has been modified. In particular embodiments, the methods include the step of regenerating a plant from the plant cell. In further embodiments, the methods include cross breeding the plant to obtain a genetically desired plant lineage. A more extensive list of endogenous genes encoding a traits of interest are listed below.

Use of a Type VII Cas system to modify polyploid plants

[01229] Many plants are polyploid, which means they carry duplicate copies of their genomes—sometimes as many as six, as in wheat. The methods according to the present invention, which make use of the Type VII Cas CRISPR effector protein and systems thereof can be "multiplexed" to affect all copies of a gene, or to target dozens of genes at once. For instance, in particular embodiments, the methods of the present invention are used to simultaneously ensure a loss of function mutation in different genes responsible for suppressing defenses against a disease. In particular embodiments, the methods of the present invention are used to simultaneously suppress the expression of the TaMLO-Al, TaMLO-Bl and TaMLO-Dl 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 (see also WO2015109752).

Exemplary genes conferring agronomic traits

[01230] As described herein above, in particular embodiments, the invention encompasses the use of the Cas CRISPR system as described herein for the insertion of a DNA of interest, including one or more plant expressible gene(s). In further particular embodiments, the invention encompasses methods and tools using the Cas system as described herein for partial or complete deletion of one or more plant expressed gene(s). In other further particular embodiments, the invention encompasses methods and tools using the Cas system as described herein to ensure modification of one or more plant-expressed genes by mutation, substitution, insertion of one of more nucleotides. In other particular embodiments, the invention encompasses the use of Cas CRISPR system as described herein to ensure modification of expression of one or more plant-expressed genes by specific modification of one or more of the regulatory elements directing expression of said genes.

[01231] In particular embodiments, the invention encompasses methods which involve the introduction of exogenous genes and/or the targeting of endogenous genes and their regulatory elements, such as listed below:

Genes that confer resistance to pests or disease:

[01232] In some embodiments the plants can be modified using the Type VII CRISPR-Cas systems or components thereof described herein to have one or more disease resistance genes. A plant can be transformed with cloned resistance genes to engineer plants that are resistant to specific pathogen strains. See, e.g., Jones et al., Science 266:789 (1994) (cloning of the tomato Cf- 9 gene for resistance to Cladosporium fulvum); Martin et al., Science 262:1432 (1993)

(tomato Pto gene for resistance to Pseudomonas syringae pv. tomato encodes a protein kinase); Mindrinos et al., Cell 78:1089 (1994) (Arabidopsmay be RSP2 gene for resistance to Pseudomonas syringae). A plant gene that is upregulated or down regulated during pathogen infection can be engineered for pathogen resistance. See, e.g., Thomazella et al., bioRxiv 064824; doi: https://doi.org/10.1101/064824 Epub. July 23, 2016 (tomato plants with deletions in the SlDMR6-1 which is normally upregulated during pathogen infection).

[01233] The following is a non-exhaustive list of genes that can be modified or introduced in a plant using the CRISPR-Cas systems and components thereof described herein that can confer disease resistance to a plant.

[01234] Genes conferring resistance to a pest, such as soybean cyst nematode. See e.g., PCT Application WO 96/30517; PCT Application WO 93/19181.

[01235] Bacillus thuringiensis proteins see, e.g., Geiser et al., Gene 48:109 (1986).

[01236] Lectins, see, for example, Van Damme et al., Plant Molec. Biol. 24:25 (1994.

[01237] Vitamin-binding protein, such as avidin, see PCT application US93/06487, teaching the use of avidin and avidin homologues as larvicides against insect pests.

[01238] Enzyme inhibitors such as protease or proteinase inhibitors or amylase inhibitors. See, e.g., Abe et al., J. Biol. Chem. 262:16793 (1987), Huub et al., Plant Molec. Biol. 21:985 (1993)), Sumitani et al., Biosci. Biotech. Biochem. 57:1243 (1993) and U.S. Pat. No. 5,494,813.

[01239] Insect-specific hormones or pheromones such as ecdysteroid or juvenile hormone, a variant thereof, a mimetic based thereon, or an antagonist or agonist thereof. See, for example Hammock et al., Nature 344:458 (1990).

[01240] Insect-specific peptides or neuropeptides which, upon expression, disrupts the physiology of the affected pest. For example Regan, J. Biol. Chem. 269:9 (1994) and Pratt et al., Biochem. Biophys. Res. Comm. 163:1243 (1989). See also U.S. Pat. No. 5,266,317.

[01241] Insect-specific venom produced in nature by a snake, a wasp, or any other organism. For example, see Pang et al., Gene 116: 165 (1992).

[01242] Enzymes responsible for a hyperaccumulation of a monoterpene, a sesquiterpene, a steroid, hydroxamic acid, a phenylpropanoid derivative or another nonprotein molecule with insecticidal activity.

[01243] Enzymes involved in the modification, including the post-translational modification, of a biologically active molecule; for example, 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. See PCT application WO93/02197, Kramer et al., Insect Biochem. Molec. Biol. 23:691 (1993) and Kawalleck et al., Plant Molec. Biol. 21 :673 (1993).

[01244] Molecules that stimulates signal transduction. For example, see Botella et al., Plant Molec. Biol. 24:757 (1994), and Griess et al., Plant Physiol. 104:1467 (1994).

[01245] Viral-invasive proteins or a complex toxin derived therefrom. See Beachy et al., Ann. rev. Phytopathol. 28:451 (1990).

[01246] Developmental-arrestive proteins produced in nature by a pathogen or a parasite. See Lamb et al., Bio/Technology 10:1436 (1992) and Toubart et al., Plant J. 2:367 (1992).

[01247] A developmental-arrestive protein produced in nature by a plant. For example, Logemann et al., Bio/Technology 10:305 (1992).

[01248] In plants, pathogens are often host-specific. For example, some Fusarium species will cause tomato wilt but attack only tomato, and other Fusarium species attack only wheat. Plants have existing and induced defenses to resist most pathogens. Mutations and recombination events across plant generations lead to genetic variability that gives rise to susceptibility, especially as pathogens reproduce with more frequency than plants. In plants there can be non-host resistance, e.g., the host and pathogen are incompatible 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. Such resistance is typically controlled by a few genes. Using methods and components of the CRISPR-Cas system, a new tool now exists to induce specific mutations in anticipation hereon. Accordingly, one can analyze the genome of sources of resistance genes, and in plants having desired characteristics or traits, use the method and components of the Cas CRISPR system to induce the rise of resistance genes. The present systems can do so with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs.

Genes involved in plant diseases, such as those listed in WO 2013046247:

[01249] In some embodiments, the Type VII CRISPR-Cas system described herein can modify a polynucleotide in one or more cells of a plant such that the plant and/or cell is resistant to or protected completely against a disease or pathogen. Such disease include, but are not limited to, any of the following diseases:

[01250] Rice diseases: Magnaporthe grisea, Cochliobolus miyabeanus, Rhizoctonia solani, Gibberella fujikuroi; Wheat diseases: Erysiphe graminis, Fusarium graminearum, F. avenaceum, F. culmorum, Microdochium nivale, Puccinia striiformis, P. graminis, P. recondita, Micronectriella nivale, Typhula sp., Ustilago tritici, Tilletia caries, Pseudocercosporella herpotrichoides, Mycosphaerella graminicola, Stagonospora nodorum, Pyrenophora tritici-repentis;Barley diseases: Erysiphe graminis, Fusarium graminearum, F. avenaceum, F. culmorum, Microdochium nivale, Puccinia striiformis, P. graminis, P. hordei, Ustilago nuda, Rhynchosporium secalis, Pyrenophora teres, Cochliobolus sativus, Pyrenophora graminea, Rhizoctonia solani;Maize diseases: Ustilago maydis, Cochliobolus heterostrophus, Gloeocercospora sorghi, Puccinia polysora, Cercospora zeae-maydis, Rhizoctonia solani;

[01251] Citrus diseases: Diaporthe citri, Elsinoe fawcetti, Penicillium digitatum, P. italicum, Phytophthora parasitica, Phytophthora citrophthora; Apple diseases: Monilinia mali, Valsa ceratosperma, Podosphaera leucotricha, Alternaria alternata apple pathotype, Venturia inaequalis, Colletotrichum acutatum, Phytophtora cactorum;

[01252] Pear diseases: Venturia nashicola, V. pirina, Alternaria alternata Japanese pear pathotype, Gymnosporangium haraeanum, Phytophtora cactorum;

[01253] Peach diseases: Monilinia fructicola, Cladosporium carpophilum, Phomopsis sp.;

[01254] Grape diseases: Elsinoe ampelina, Glomerella cingulata, Uninula necator, Phakopsora ampelopsidis, Guignardia bidwellii, Plasmopara viticola;

[01255] Persimmon diseases: Gloesporium kaki, Cercospora kaki, Mycosphaerela nawae;

[01256] Gourd diseases: Colletotrichum lagenarium, Sphaerotheca fuliginea, Mycosphaerella melonis, Fusarium oxysporum, Pseudoperonospora cubensis, Phytophthora sp., Pythium sp.;

[01257] Tomato diseases: Alternaria solani, Cladosporium fulvum, Phytophthora infestans; Pseudomonas syringae pv. Tomato; Phytophthora capsici; Xanthomonas

[01258] Eggplant diseases: Phomopsis vexans, Erysiphe cichoracearum; Brassicaceous vegetable diseases: Alternaria japonica, Cercosporella brassicae, Plasmodiophora brassicae, Peronospora parasitica;

[01259] Welsh onion diseases: Puccinia allii, Peronospora destructor;

[01260] Soybean diseases: Cercospora kikuchii, Elsinoe glycines, Diaporthe phaseolorum var. sojae, Septoria glycines, Cercospora sojina, Phakopsora pachyrhizi, Phytophthora sojae, Rhizoctonia solani, Corynespora casiicola, Sclerotinia sclerotiorum;

[01261] Kidney bean diseases: Colletrichum lindemthianum;

[01262] Peanut diseases: Cercospora personata, Cercospora arachidicola, Sclerotium rolfsii;

[01263] Pea diseases pea: Erysiphe pisi;

[01264] Potato diseases: Alternaria solani, Phytophthora infestans, Phytophthora erythroseptica, Spongospora subterranean, f. sp. Subterranean;

[01265] Strawberry diseases: Sphaerotheca humuli, Glomerella cingulata;

[01266] Tea diseases: Exobasidium reticulatum, Elsinoe leucospila, Pestalotiopsis sp., Colletotrichum theae-sinensis;

[01267] Tobacco diseases: Alternaria longipes, Erysiphe cichoracearum, Colletotrichum tabacum, Peronospora tabacina, Phytophthora nicotianae;

[01268] Rapeseed diseases: Sclerotinia sclerotiorum, Rhizoctonia solani;

[01269] Cotton diseases: Rhizoctonia solani;

[01270] Beet diseases: Cercospora beticola, Thanatephorus cucumeris, Thanatephorus cucumeris, Aphanomyces cochlioides;

[01271] Rose diseases: Diplocarpon rosae, Sphaerotheca pannosa, Peronospora sparsa;

[01272] Diseases of chrysanthemum and asteraceae: Bremia lactuca, Septoria chrysanthemi-indici, Puccinia horiana;

[01273] Diseases of various plants: Pythium aphanidermatum, Pythium debarianum, Pythium graminicola, Pythium irregulare, Pythium ultimum, Botrytis cinerea, Sclerotinia sclerotiorum;

[01274] Radish diseases: Alternaria brassicicola;

[01275] Zoysia diseases: Sclerotinia homeocarpa, Rhizoctonia solani;

[01276] Banana diseases: Mycosphaerella fijiensis, Mycosphaerella musicola;

[01277] Sunflower diseases: Plasmopara halstedii;

[01278] Seed diseases or diseases in the initial stage of growth of various plants caused by Aspergillus spp., Penicillium spp., Fusarium spp., Gibberella spp., Tricoderma spp., Thielaviopsis spp., Rhizopus spp., Mucor spp., Corticium spp., Rhoma spp., Rhizoctonia spp., Diplodia spp., or the like;

[01279] Virus diseases of various plants mediated by Polymixa spp., Olpidium spp., or the like.

Examples of genes that confer resistance to herbicides:

[01280] As part of normal horticulture practices, herbicides are applied and can negatively affect the desired plant as well as inhibit or kill weed plants. In some embodiments, the Type VII CRISPR-Cas systems and/or components thereof described herein can be used to modify a plant or cell thereof such that it has resistance to the herbicide and is not or less affected by the herbicide.

[01281] Resistance to herbicides that inhibit the growing point or meristem, such as an imidazolinone or a sulfonylurea, for example, by Lee et al., EMBO J. 7:1241 (1988), and Miki et al., Theor. Appl. Genet. 80:449 (1990), respectively.

[01282] Glyphosate tolerance (resistance conferred by, e.g., mutant 5-enolpyruvylshikimate-3- phosphate synthase (EPSPs) 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 proprionic acids and cyclohexones by ACCase inhibitor-encoding genes. See, for example, U.S. Pat. No. 4,940,835 and U.S. Pat. 6,248,876 , U.S. Pat. No. 4,769,061 , EP No. 0 333 033 and U.S. Pat No. 4,975,374. See also EP No. 0242246, DeGreef et al., Bio/Technology 7:61 (1989), Marshall et al., Theor. Appl. Genet. 83:435 (1992), WO 2005012515 to Castle et. al. and WO 2005107437.

[01283] Resistance to herbicides that inhibit photosynthesis, such as a triazine (psbA and gs+ genes) or a benzonitrile (nitrilase gene), and glutathione S-transferase in Przibila et al., Plant Cell 3:169 (1991), U.S. Pat. No. 4,810,648, and Hayes et al., Biochem. J. 285: 173 (1992).

[01284] Genes encoding enzymes detoxifying the herbicide or a mutant glutamine synthase enzyme that is resistant to inhibition, e.g. n U.S. patent application Ser. No. 11/760,602. Or a detoxifying enzyme is an enzyme encoding a phosphinothricin acetyltransferase (such as the bar or pat protein from Streptomyces species). Phosphinothricin acetyltransferases are for example described in U.S. Pat. Nos. 5,561,236; 5,648,477; 5,646,024; 5,273,894; 5,637,489; 5,276,268; 5,739,082; 5,908,810 and 7,112,665.

[01285] Hydroxyphenylpyruvatedioxygenases (HPPD) inhibitors, i.e., naturally occurring HPPD resistant enzymes, or genes encoding a mutated or chimeric HPPD enzyme as described

in WO 96/38567, WO 99/24585, and WO 99/24586, WO 2009/144079, WO 2002/046387, or U.S. Pat. No. 6,768,044.

[01286] Any one or more of these genes can be modified to confer herbicide resistance using the Type VII CRISPR-Cas systems and/or components thereof described herein.

Using Type VII CRISPR-Cas Systems to confer Abiotic stress tolerance

[01287] In some embodiments, the Type VII CRISPR-Cas systems and/or components thereof described herein can be used to modify a plant or cell thereof such that it has resistance to the abiotic stress and is not or less affected by the abiotic stress. Any one or more of the following genes can be modified or introduced to confer abiotic resistance using the Type VII CRISPR-Cas systems and/or components thereof described herein.

[01288] Transgene capable of reducing the expression and/or the activity of poly(ADP-ribose) polymerase (PARP) gene in the plant cells or plants as described in International Patent Publication Nos. WO 00/04173 or WO/2006/045633.

[01289] Transgenes capable of reducing the expression and/or the activity of the PARG encoding genes of the plants or plants cells, as described e.g. in WO 2004/090140.

[01290] Transgenes coding for a plant-functional enzyme of the nicotineamide adenine dinucleotide salvage synthesis pathway including nicotinamidase, nicotinate phosphoribosyltransferase, nicotinic acid mononucleotide adenyl transferase, nicotinamide adenine dinucleotide synthetase or nicotine amide phosphorybosyltransferase as described e.g. in EP 04077624.7, WO 2006/133827, PCT/EP07/002,433, EP 1999263, or WO 2007/107326.

[01291] Enzymes involved in carbohydrate biosynthesis include those described in e.g. EP 0571427, WO 95/04826, EP 0719338, WO 96/15248, WO 96/19581, WO 96/27674, WO 97/11188, WO 97/26362, WO 97/32985, WO 97/42328, WO 97/44472, WO 97/45545, WO 98/27212, WO 98/40503, WO99/58688, WO 99/58690, WO 99/58654, WO 00/08184, WO 00/08185, WO 00/08175, WO 00/28052, WO 00/77229, WO 01/12782, WO 01/12826, WO 02/101059, WO 03/071860, WO 2004/056999, WO 2005/030942, WO 2005/030941, WO 2005/095632, WO 2005/095617, WO 2005/095619, WO 2005/095618, WO 2005/123927, WO 2006/018319, WO 2006/103107, WO 2006/108702, WO 2007/009823, WO 00/22140, WO 2006/063862, WO 2006/072603, WO 02/034923, EP 06090134.5, EP 06090228.5, EP 06090227.7, EP 07090007.1, EP 07090009.7, WO 01/14569, WO 02/79410, WO 03/33540, WO 2004/078983, WO 01/19975, WO 95/26407, WO 96/34968, WO 98/20145, WO 99/12950, WO 99/66050, WO 99/53072, U.S. Pat. No. 6,734,341, WO 00/11192, WO

98/22604, WO 98/32326, WO 01/98509, WO 01/98509, WO 2005/002359, U.S. Pat. No. 5,824,790, U.S. Pat. No. 6,013,861, WO 94/04693, WO 94/09144, WO 94/11520, WO 95/35026 or WO 97/20936 or enzymes involved in the production of polyfructose, especially of the inulin and levan-type, as disclosed in EP 0663956, WO 96/01904, WO 96/21023, WO 98/39460, and WO 99/24593, the production of alpha-1,4-glucans as disclosed in WO 95/31553, US 2002031826, U.S. Pat. No. 6,284,479, U.S. Pat. No. 5,712,107, WO 97/47806, WO 97/47807, WO 97/47808 and WO 00/14249, the production of alpha-1,6 branched alpha-1,4-glucans, as disclosed in WO 00/73422, the production of alternan, as disclosed in e.g. WO 00/47727, WO 00/73422, EP 06077301.7, U.S. Pat. No. 5,908,975 and EP 0728213, the production of hyaluronan, as for example disclosed in WO 2006/032538, WO 2007/039314, WO 2007/039315, WO 2007/039316, JP 2006304779, and WO 2005/012529.

[01292] Genes that improve drought resistance. For example, WO 2013122472 discloses that the absence or reduced level of functional Ubiquitin Protein Ligase protein (UPL) protein, more specifically, UPL3, leads to a decreased need for water or improved resistance to drought of said plant. Other examples of transgenic plants with increased drought tolerance are disclosed in, for example, US Patent Publication Nos. 2009/0144850 and US 2007/0266453, and International Patent Publication WO 2002/083911. US2009/0144850 describes a plant displaying a drought tolerance phenotype due to altered expression of a DR02 nucleic acid. US 2007/0266453 describes a plant displaying a drought tolerance phenotype due to altered expression of a DR03 nucleic acid and WO 2002/08391 1 describes a plant having an increased tolerance to drought stress due to a reduced activity of an ABC transporter which is expressed in guard cells. Another example is the work by Kasuga and co-authors (1999), who describe that overexpression of cDNA encoding DREB1 A in transgenic plants activated the expression of many stress tolerance genes under normal growing conditions and resulted in improved tolerance to drought, salt loading, and freezing. However, the expression of DREB1A also resulted in severe growth retardation under normal growing conditions (Kasuga (1999) Nat Biotechnol 17(3) 287-291).

[01293] In further particular embodiments, crop plants can be improved by influencing specific plant traits. For example, by developing pesticide-resistant plants, improving disease resistance in plants, improving plant insect and nematode resistance, improving plant resistance against parasitic weeds, improving plant drought tolerance, improving plant nutritional value,

improving plant stress tolerance, avoiding self-pollination, plant forage digestibility biomass, grain yield etc. A few specific non-limiting examples are provided hereinbelow.

[01294] In addition to targeted mutation of single genes, Cas CRISPR complexes can be designed to allow targeted mutation of multiple genes, deletion of chromosomal fragment, site-specific integration of transgene, site-directed mutagenesis *in vivo*, and precise gene replacement or allele swapping in plants. Therefore, the methods described herein have broad applications in gene discovery and validation, mutational and cisgenic breeding, and hybrid breeding. These applications facilitate the production of a new generation of genetically modified crops with various improved agronomic traits such as herbicide resistance, disease resistance, abiotic stress tolerance, high yield, and superior quality.

Use of Type VII CRISPR-Cas systems to create male sterile plants

[01295] 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, genes have been identified which are important for plant fertility, more particularly male fertility. For instance, in maize, at least two genes have been identified which are important in fertility (Amitabh Mohanty International Conference on New Plant Breeding Molecular Technologies Technology Development and Regulation, Oct 9-10, 2014, Jaipur, India; Svitashev et al. Plant Physiol. 2015 Oct;169(2):931-45; Djukanovic et al. Plant J. 2013 Dec;76(5):888-99). The methods provided herein can be used to target genes required for male fertility so as to generate male sterile plants which can easily be crossed to generate hybrids. In particular embodiments, the Type VII Cas CRISPR systems or components thereof provided herein is used for targeted mutagenesis of the cytochrome P450-like gene (MS26) or the meganuclease gene (MS45) thereby conferring male sterility to the maize plant. Maize plants which are as such genetically altered can be used in hybrid breeding programs.

Increasing the fertility stage in plants

[01296] In some embodiments, one or more genes can be modified and/or inserted using the Type VII CRISRP-Cas system or component thereof described herein that can increase the fertility stage in a plant.

[01297] In particular embodiments, the methods provided herein can used to prolong the fertility stage of a plant, such as of a rice plant. 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 (as described in CN 104004782)

Use of Type VII CRISPR-Cas systems to generate genetic variation in a crop of interest

[01298] The availability of wild germplasm and genetic variations in crop plants is the key to crop improvement programs, but the available diversity in germplasms from crop plants is limited. The present invention envisages methods for generating a diversity of genetic variations in a germplasm of interest. In some embodiments, the Type VII Cas CRISPR system can include a library of guide RNAs targeting different locations in the plant genome is provided, which can be introduced into one or more plant cells together with a Cas effector protein (e.g. a Type VII Cas effector protein) such that a collection of genome-scale point mutations and gene knock-outs can be generated. In particular embodiments, the methods comprise generating a plant part or plant from the cells so obtained and screening the cells for a trait of interest. The target genes can include both coding and non-coding regions. In particular embodiments, the trait is stress tolerance and a method for the generation of stress-tolerant crop varieties.

Use of Type VII CRISPR-Cas systems to modulate fruit-ripening

[01299] In some embodiments, the Type VII CRISPR-Cas system can be used to modify one or more polynucleotides such that ripening of fruit is modulated. Ripening is a normal phase in the maturation process of fruits and vegetables. Only a few days after it starts it renders a fruit or vegetable inedible. This process brings significant losses to both farmers and consumers. In particular embodiments, the methods of the present invention are used to reduce ethylene production. This is ensured by ensuring one or more of the following: a. Suppression of ACC synthase gene expression. ACC (1-aminocyclopropane-1-carboxylic acid) synthase is the enzyme responsible for the conversion of S-adenosylmethionine (SAM) to ACC; the second to the last step in ethylene biosynthesis. Enzyme expression is hindered when an antisense ("mirror-image") or truncated copy of the synthase gene is inserted into the plant's genome; b. Insertion of the ACC deaminase gene. The gene coding for the enzyme is obtained from Pseudomonas chlororaphis, a common nonpathogenic soil bacterium. It converts ACC to a different compound thereby reducing the amount of ACC available for ethylene production; c. Insertion of the SAM hydrolase gene. This approach is similar to ACC deaminase wherein ethylene production is hindered when the amount of its precursor metabolite is reduced; in this case SAM is converted to homoserine. The gene coding for the enzyme is obtained from E. coli T3 bacteriophage and d. Suppression of ACC oxidase gene expression. ACC oxidase is the enzyme which catalyzes the oxidation of ACC to ethylene, the last step in the ethylene

biosynthetic pathway. Using the methods described herein, down regulation of the ACC oxidase gene results in the suppression of ethylene production, thereby delaying fruit ripening. In particular embodiments, additionally or alternatively to the modifications described above, the methods described herein are used to modify ethylene receptors, so as to interfere with ethylene signals obtained by the fruit. In particular embodiments, expression of the ETR1 gene, encoding an ethylene binding protein is modified, more particularly suppressed. In particular embodiments, additionally or alternatively to the modifications described above, the methods described herein are used to modify expression of the gene encoding Polygalacturonase (PG), which is the enzyme responsible for the breakdown of pectin, the substance that maintains the integrity of plant cell walls. Pectin breakdown occurs at the start of the ripening process resulting in the softening of the fruit. Accordingly, in particular embodiments, the methods described herein are used to introduce a mutation in the PG gene or to suppress activation of the PG gene in order to reduce the amount of PG enzyme produced thereby delaying pectin degradation.

[01300] Thus, in particular embodiments, the methods comprise the use of the Type VII Cas CRISPR system or component thereof described herein to ensure one or more modifications of the genome of a plant cell such as described above, and regenerating a plant therefrom. In particular embodiments, the plant is a tomato plant.

Use of Type VII CRISPR-Cas systems to increase storage life of plants

[01301] In particular embodiments, the Type VII CRISPR-Cas systems, components thereof, and methods of the present invention are used to modify genes involved in the production of compounds which affect storage life of the plant or plant part. More particularly, the modification is in one or more genes that can prevent 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 (Clasen et al. DOI: 10.1111/pbi.12370).

Use of the Type VII Cas CRISPR system to ensure a value added trait

[01302] In particular embodiments the Type VII Cas CRISPR system and/or component thereof is used to produce nutritionally improved agricultural crops. In particular embodiments,

the methods provided herein are adapted to generate "functional foods", i.e. a modified food or food ingredient that may provide a health benefit beyond the traditional nutrients it contains and or "nutraceutical", i.e. substances that may be considered a food or part of a food and provides health benefits, including the prevention and treatment of disease. In particular embodiments, the nutraceutical is useful in the prevention and/or treatment of one or more of cancer, diabetes, cardiovascular disease, and hypertension.

[01303] Examples of nutritionally improved crops include (Newell-McGloughlin, Plant Physiology, July 2008, Vol. 147, pp. 939–953): modified protein quality, content and/or amino acid composition, such as have been described for Bahiagrass (Luciani et al. 2005, Florida Genetics Conference Poster), Canola (Roesler et al., 1997, Plant Physiol 113 75–81), Maize (Cromwell et al, 1967, 1969 J Anim Sci 26 1325–1331, O'Quin et al. 2000 J Anim Sci 78 2144–2149, Yang et al. 2002, Transgenic Res 11 11–20, Young et al. 2004, Plant J 38 910–922), Potato (Yu J and Ao, 1997 Acta Bot Sin 39 329–334; Chakraborty et al. 2000, Proc Natl Acad Sci USA 97 3724–3729; Li et al. 2001) Chin Sci Bull 46 482–484, Rice (Katsube et al. 1999, Plant Physiol 120 1063–1074), Soybean (Dinkins et al. 2001, Rapp 2002, *In Vitro* Cell Dev Biol Plant 37 742–747), Sweet Potato (Egnin and Prakash 1997, *In Vitro* Cell Dev Biol 33 52A).

[01304] Essential amino acid content, such as has been described for Canola (Falco et al. 1995, Bio/Technology 13 577–582), Lupin (White et al. 2001, J Sci Food Agric 81 147–154), Maize (Lai and Messing, 2002, Agbios 2008 GM crop database (March 11, 2008)), Potato (Zeh et al. 2001, Plant Physiol 127 792–802), Sorghum (Zhao et al. 2003, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp 413–416), Soybean (Falco et al. 1995 Bio/Technology 13 577–582; Galili et al. 2002 Crit Rev Plant Sci 21 167–204).

[01305] Oils and Fatty acids such as for Canola (Dehesh et al. (1996) Plant J 9 167–172 [PubMed] ; Del Vecchio (1996) INFORM International News on Fats, Oils and Related Materials 7 230–243; Roesler et al. (1997) Plant Physiol 113 75–81 [PMC free article] [PubMed]; Froman and Ursin (2002, 2003) Abstracts of Papers of the American Chemical Society 223 U35; James et al. (2003) Am J Clin Nutr 77 1140–1145 [PubMed]; Agbios (2008, above); coton (Chapman et al. (2001) . J Am Oil Chem Soc 78 941–947; Liu et al. (2002) J Am Coll Nutr 21 205S–211S [PubMed]; O'Neill (2007) Australian Life Scientist. http://www.biotechnews.com.au/index.php/id;866694817;fp;4;fpid;2 (June 17, 2008), Linseed (Abbadi et al., 2004, Plant Cell 16: 2734–2748), Maize (Young et al., 2004, Plant J 38 910–

922), oil palm (Jalani et al. 1997, J Am Oil Chem Soc 74 1451–1455; Parveez, 2003, AgBiotechNet 113 1–8), Rice (Anai et al., 2003, Plant Cell Rep 21 988–992), Soybean (Reddy and Thomas, 1996, Nat Biotechnol 14 639–642; Kinney and Kwolton, 1998, Blackie Academic and Professional, London, pp 193–213), Sunflower (Arcadia, Biosciences 2008)

[01306] Carbohydrates, such as Fructans described for Chicory (Smeekens (1997) Trends Plant Sci 2 286–287, Sprenger et al. (1997) FEBS Lett 400 355–358, Sévenier et al. (1998) Nat Biotechnol 16 843–846), Maize (Caimi et al. (1996) Plant Physiol 110 355–363), Potato (Hellwege et al. ,1997 Plant J 12 1057–1065), Sugar Beet (Smeekens et al. 1997, above), Inulin, such as described for Potato (Hellewege et al. 2000, Proc Natl Acad Sci USA 97 8699–8704), Starch, such as described for Rice (Schwall et al. (2000) Nat Biotechnol 18 551–554, Chiang et al. (2005) Mol Breed 15 125–143),

[01307] Vitamins and carotenoids, such as described for Canola (Shintani and DellaPenna (1998) Science 282 2098–2100), Maize (Rocheford et al. (2002) . J Am Coll Nutr 21 191S–198S, Cahoon et al. (2003) Nat Biotechnol 21 1082–1087, Chen et al. (2003) Proc Natl Acad Sci USA 100 3525–3530), Mustardseed (Shewmaker et al. (1999) Plant J 20 401–412, Potato (Ducreux et al., 2005, J Exp Bot 56 81–89), Rice (Ye et al. (2000) Science 287 303–305, Strawberry (Agius et al. (2003), Nat Biotechnol 21 177–181), Tomato (Rosati et al. (2000) Plant J 24 413–419, Fraser et al. (2001) J Sci Food Agric 81 822–827, Mehta et al. (2002) Nat Biotechnol 20 613–618, Díaz de la Garza et al. (2004) Proc Natl Acad Sci USA 101 13720–13725, Enfissi et al. (2005) Plant Biotechnol J 3 17–27, DellaPenna (2007) Proc Natl Acad Sci USA 104 3675–3676.

[01308] Functional secondary metabolites, such as described for Apple (stilbenes, Szankowski et al. (2003) Plant Cell Rep 22: 141–149), Alfalfa (resveratrol, Hipskind and Paiva (2000) Mol Plant Microbe Interact 13 551–562), Kiwi (resveratrol, Kobayashi et al. (2000) Plant Cell Rep 19 904–910), Maize and Soybean (flavonoids, Yu et al. (2000) Plant Physiol 124 781–794), Potato (anthocyanin and alkaloid glycoside, Lukaszewicz et al. (2004) J Agric Food Chem 52 1526–1533), Rice (flavonoids & resveratrol, Stark-Lorenzen et al. (1997) Plant Cell Rep 16 668–673, Shin et al. (2006) Plant Biotechnol J 4 303–315), Tomato (+resveratrol, chlorogenic acid, flavonoids, stilbene; Rosati et al. (2000) above, Muir et al. (2001) Nature 19 470–474, Niggeweg et al. (2004) Nat Biotechnol 22 746–754, Giovinazzo et al. (2005) Plant Biotechnol J 3 57–69), wheat (caffeic and ferulic acids, resveratrol; United Press International (2002)); and

[01309] Mineral availabilities such as described for Alfalfa (phytase, Austin-Phillips et al. (1999) http://www.molecularfarming.com/nonmedical.html), Lettuse (iron, Goto et al. (2000) Theor Appl Genet 100 658–664), Rice (iron, Lucca et al. (2002) J Am Coll Nutr 21 184S–190S), Maize, Soybean and wheat (phytase, Drakakaki et al. (2005) Plant Mol Biol 59 869–880, Denbow et al. (1998) Poult Sci 77 878–881, Brinch-Pedersen et al. (2000) Mol Breed 6 195–206).

[01310] In particular embodiments, the value-added trait is related to the envisaged health benefits of the compounds present in the plant. For instance, in particular embodiments, the value-added crop is obtained by applying the methods of the invention to ensure the modification of or induce/increase the synthesis of one or more of the following compounds:

[01311] Carotenoids, such as α-Carotene present in carrots which Neutralizes free radicals that may cause damage to cells or β-Carotene present in various fruits and vegetables which neutralizes free radicals

[01312] Lutein present in green vegetables which contributes to maintenance of healthy vision

[01313] Lycopene present in tomato and tomato products, which is believed to reduce the risk of prostate cancer

[01314] Zeaxanthin, present in citrus and maize, which contributes to maintenance of healthy vision

[01315] Dietary fiber such as insoluble fiber present in wheat bran which may reduce the risk of breast and/or colon cancer and β-Glucan present in oat, soluble fiber present in Psylium and whole cereal grains which may reduce the risk of cardiovascular disease (CVD)

[01316] Fatty acids, such as ω-3 fatty acids which may reduce the risk of CVD and improve mental and visual functions, Conjugated linoleic acid, which may improve body composition, may decrease risk of certain cancers and GLA which may reduce inflammation risk of cancer and CVD, may improve body composition

[01317] Flavonoids such as hydroxycinnamates, present in wheat which have Antioxidant-like activities, may reduce risk of degenerative diseases, flavonols, catechins and tannins present in fruits and vegetables which neutralize free radicals and may reduce risk of cancer

[01318] Glucosinolates, indoles, isothiocyanates, such as Sulforaphane, present in Cruciferous vegetables (broccoli, kale), horseradish, which neutralize free radicals, may reduce risk of cancer

[01319] Phenolics, such as stilbenes present in grape which May reduce risk of degenerative diseases, heart disease, and cancer, may have longevity effect and caffeic acid and ferulic acid present in vegetables and citrus which have Antioxidant-like activities, may reduce risk of degenerative diseases, heart disease, and eye disease, and epicatechin present in cacao which has Antioxidant-like activities, may reduce risk of degenerative diseases and heart disease

[01320] Plant stanols/sterols present in maize, soy, wheat and wooden oils which May reduce risk of coronary heart disease by lowering blood cholesterol levels

[01321] Fructans, inulins, fructo-oligosaccharides present in Jerusalem artichoke, shallot, onion powder which may improve gastrointestinal health

[01322] Saponins present in soybean, which may lower LDL cholesterol .

[01323] Soybean protein present in soybean which may reduce risk of heart disease

[01324] Phytoestrogens such as isoflavones present in soybean which May reduce menopause symptoms, such as hot flashes, may reduce osteoporosis and CVD and lignans present in flax, rye and vegetables, which May protect against heart disease and some cancers, may lower LDL cholesterol, total cholesterol.

[01325] Sulfides and thiols such as diallyl sulphide present in onion, garlic, olive, leek and scallion and Allyl methyl trisulfide, dithiolthiones present in cruciferous vegetables which may lower LDL cholesterol, helps to maintain healthy immune system

[01326] Tannins, such as proanthocyanidins, present in cranberry, cocoa, which may improve urinary tract health, may reduce risk of CVD and high blood pressure.

[01327] Any one or more genes involved in the biochemical and biological processes within the plant and/or cell(s) to produce the desirable products described above can be modified such that an increased or optimized level of the products described above are produced to achieve the desired result (i.e. increased nutritional quality etc.)

[01328] In addition, the methods of the present invention also envisage modifying protein/starch functionality, shelf life, taste/aesthetics, fiber quality, and allergen, antinutrient, and toxin reduction traits.

[01329] Accordingly, the invention encompasses methods for producing plants with nutritional added value, said methods comprising introducing into a plant cell a gene encoding an enzyme involved in the production of a component of added nutritional value using the Type VII Cas CRISPR system as described herein and regenerating a plant from said plant cell, said plant characterized in an increase expression of said component of added nutritional value. In

particular embodiments, the Cas CRISPR system is 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. Methods for introducing a gene of interest into a plant cell and/or modifying an endogenous gene using the Cas CRISPR system are described herein above.

[01330] Some specific examples of modifications in plants that have been modified to confer value-added traits are: plants with modified fatty acid metabolism, for example, by transforming a plant with an antisense gene of stearyl-ACP desaturase to increase stearic acid content of the plant. See Knultzon et al., Proc. Natl. Acad. Sci. U.S.A. 89:2624 (1992). Another example involves decreasing phytate content, for example by cloning and then reintroducing DNA associated with the single allele which may be responsible for maize mutants characterized by low levels of phytic acid. See Raboy et al, Maydica 35:383 (1990).

[01331] Similarly, expression of the maize (Zea mays) Tfs C1 and R, which regulate the production of flavonoids in maize aleurone layers under the control of a strong promoter, resulted in a high accumulation rate of anthocyanins in Arabidopsis (Arabidopsis thaliana), presumably by activating the entire pathway (Bruce et al., 2000, Plant Cell 12:65–80). DellaPenna (Welsch et al., 2007 Annu Rev Plant Biol 57: 711–738) found that Tf RAP2.2 and its interacting partner SINAT2 increased carotenogenesis in Arabidopsis leaves. Expressing the Tf Dof1 induced the up-regulation of genes encoding enzymes for carbon skeleton production, a marked increase of amino acid content, and a reduction of the Glc level in transgenic Arabidopsis (Yanagisawa, 2004 Plant Cell Physiol 45: 386–391), and the DOF Tf AtDof1.1 (OBP2) up-regulated all steps in the glucosinolate biosynthetic pathway in Arabidopsis (Skirycz et al., 2006 Plant J 47: 10–24).

Use of Type VII CRISPR-Cas systems to reduce allergens in plants

[01332] In particular embodiments the Type VII CRISPR-Cas systems and/or components thereof and methods provided herein are used to generate plants with a reduced level of allergens, making them safer for the consumer. In particular embodiments, the methods comprise modifying expression of one or more genes responsible for the production of plant allergens. For instance, in particular embodiments, the methods comprise down-regulating expression of a Lol p5 gene in a plant cell, such as a ryegrass plant cell and regenerating a plant therefrom so as to reduce allergenicity of the pollen of said plant (Bhalla et al. 1999, Proc. Natl. Acad. Sci. USA Vol. 96: 11676–11680).

[01333] Peanut allergies and allergies to legumes generally are a real and serious health concern. The Cas effector protein system of the present invention can be used to identify and then edit or silence genes encoding allergenic proteins of such legumes. Without limitation as to such genes and proteins, Nicolaou et al. identifies allergenic proteins in peanuts, soybeans, lentils, peas, lupin, green beans, and mung beans. See, Nicolaou et al., Current Opinion in Allergy and Clinical Immunology 2011;11(3):222).

Screening methods for endogenous genes of interest

[01334] The methods provided herein further allow the identification of genes of value 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 in plants using the Type VII Cas CRISPR system as described herein, 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.

Further applications of the CRISPR system in plants and yeast

Use of Type VII CRISPR system in biofuel production

[01335] The term "biofuel" as used herein is an alternative fuel made from plant and plant-derived resources. Renewable biofuels can be extracted from organic matter whose energy has been obtained through a process of carbon fixation or are made through the use or conversion of biomass. This biomass can be used directly for biofuels or can be converted to convenient energy containing substances by thermal conversion, chemical conversion, and biochemical conversion. This biomass conversion can result in fuel in solid, liquid, or gas form. There are two types of biofuels: bioethanol and biodiesel. Bioethanol is mainly produced by the sugar fermentation process of cellulose (starch), which is mostly derived from maize and sugar cane. Biodiesel on the other hand is mainly produced from oil crops such as rapeseed, palm, and soybean. Biofuels are used mainly for transportation.

Enhancing plant properties for biofuel production

[01336] In particular embodiments, the methods using the Cas CRISPR system as described herein are used to alter the properties of the cell wall in order to facilitate access by key hydrolyzing agents for a more efficient release of sugars for fermentation. In particular

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

[01337] In particular embodiments, the methods described herein are used to produce plant mass that produces lower levels of acetic acid during fermentation (see also WO 2010096488). More particularly, the methods disclosed herein are used to generate mutations in homologs to CaslL to reduce polysaccharide acetylation.

Modifying yeast for Biofuel production

[01338] In particular embodiments, the Cas enzyme provided herein is used for bioethanol production by recombinant micro-organisms. For instance, Cas can be used to engineer micro-organisms, such as yeast, to generate biofuel or biopolymers from fermentable sugars and optionally to be able to degrade plant-derived lignocellulose derived from agricultural waste as a source of fermentable sugars. More particularly, the invention provides methods whereby the Cas CRISPR complex is used to introduce foreign genes required for biofuel production into micro-organisms and/or to modify endogenous genes why may interfere with the biofuel synthesis. More particularly the methods involve introducing into a micro-organism such as a yeast one or more nucleotide sequence encoding enzymes involved in the conversion of pyruvate to ethanol or another product of interest. In particular embodiments the methods ensure the introduction of one or more enzymes which allows the micro-organism to degrade cellulose, such as a cellulase. In yet further embodiments, the Cas CRISPR complex is used to modify endogenous metabolic pathways which compete with the biofuel production pathway.

[01339] Accordingly, in more particular embodiments, the methods described herein are used to modify a micro-organism as follows:

[01340] To introduce at least one heterologous nucleic acid or increase expression of at least one endogenous nucleic acid encoding a plant cell wall degrading enzyme, such that said micro-organism is capable of expressing said nucleic acid and of producing and secreting said plant cell wall degrading enzyme; to introduce at least one heterologous nucleic acid or increase expression of at least one endogenous nucleic acid encoding an enzyme that converts pyruvate to acetaldehyde optionally combined with at least one heterologous nucleic acid encoding an enzyme that converts acetaldehyde to ethanol such that said host cell is capable of expressing said nucleic acid; and/or to modify at least one nucleic acid encoding for an enzyme in a metabolic pathway in said host cell, wherein said pathway produces a metabolite other than acetaldehyde from pyruvate or ethanol from acetaldehyde, and wherein said modification results in a reduced production of said metabolite, or to introduce at least one nucleic acid encoding for an inhibitor of said enzyme.

Modifying Algae and plants for production of vegetable oils or biofuels

[01341] Transgenic algae or other plants such as rape may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol), for instance. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[01342] According to particular embodiments of the invention, the Cas CRISPR system is used to generate lipid-rich diatoms which are useful in biofuel production.

[01343] In particular embodiments it is envisaged to specifically modify genes that are involved in the modification of the quantity of lipids and/or the quality of the lipids produced by the algal cell. Examples of genes encoding enzymes involved in the pathways of fatty acid synthesis can encode proteins having for instance acetyl-CoA carboxylase, fatty acid synthase, 3-ketoacyl_acyl- carrier protein synthase III, glycerol-3-phospate deshydrogenase (G3PDH), Enoyl-acyl carrier protein reductase (Enoyl-ACP-reductase), glycerol-3-phosphate acyltransferase, lysophosphatidic acyl transferase or diacylglycerol acyltransferase, phospholipid:diacylglycerol acyltransferase, phoshatidate phosphatase, fatty acid thioesterase such as palmitoyi protein thioesterase, or malic enzyme activities. In further embodiments it is envisaged to generate diatoms that have increased lipid accumulation. This can be achieved by targeting genes that decrease lipid catabolisation. Of particular interest for use in the methods

of the present invention are genes involved in the activation of both triacylglycerol and free fatty acids, as well as genes directly involved in β-oxidation of fatty acids, such as acyl-CoA synthetase, 3-ketoacyl-CoA thiolase, acyl-CoA oxidase activity and phosphoglucomutase. The Cas CRISPR system and methods described herein can be used to specifically activate such genes in diatoms as to increase their lipid content.

[01344] Organisms such as microalgae are widely used for synthetic biology. Stovicek et al. (Metab. Eng. Comm., 2015; 2:13 describes genome editing of industrial yeast, for example, Saccharomyces cerevisae, to efficiently produce robust strains for industrial production. Stovicek used a CRISPR-Cas9 system codon-optimized for yeast to simultaneously disrupt both alleles of an endogenous gene and knock in a heterologous gene. Cas9 and gRNA were expressed from genomic or episomal 2μ-based vector locations. The authors also showed that gene disruption efficiency could be improved by optimization of the levels of Cas9 and gRNA expression. Hlavová et al. (Biotechnol. Adv. 2015) discusses development of species or strains of microalgae using techniques such as CRISPR to target nuclear and chloroplast genes for insertional mutagenesis and screening. The methods of Stovicek and Hlavová may be applied to the Cas effector protein system of the present invention.

[01345] US 8,945,839 describes a method for engineering Micro-Algae (Chlamydomonas reinhardtii cells) species) using Cas9 . Using similar tools, the methods of the Cas CRISPR system described herein can be applied on Chlamydomonas species and other algae. In particular embodiments, Cas and guide RNA are introduced in algae expressed using a vector that expresses Cas under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2 -tubulin. Guide RNA will be delivered using a vector containing T7 promoter. Alternatively, Cas mRNA and *in vitro* transcribed guide RNA can be delivered to algal cells. Electroporation protocol follows standard recommended protocol from the GeneArt Chlamydomonas Engineering kit.

Use of Type VII CRISPR-Cas systems to generate of microorganisms capable of fatty acid production

[01346] In particular embodiments, the Type VII CRISPR-Cas systems and molecules and methods of the invention are used for the generation of genetically engineered micro-organisms capable of the production of fatty esters, such as fatty acid methyl esters ("FAME") and fatty acid ethyl esters ("FAEE"),

[01347] Typically, host cells can be engineered to produce fatty esters from a carbon source, such as an alcohol, present in the medium, by expression or overexpression of a gene encoding a thioesterase, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase. Accordingly, the methods provided herein are used to modify a micro-organisms so as to overexpress or introduce a thioesterase gene, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase. In particular embodiments, the thioesterase gene is selected from tesA, 'tesA, tesB, fatB, fatB2,fatB3,fatAl, or fatA. In particular embodiments, the gene encoding an acyl-CoA synthase is selected from fadDJadK, BH3103, pfl-4354, EAV15023, fadDl, fadD2, RPC_4074,fadDD35, fadDD22, faa39, or an identified gene encoding an enzyme having the same properties. In particular embodiments, the gene encoding an ester synthase is a gene encoding a synthase/acyl-CoA:diacylglycerl acyltransferase from Simmondsia chinensis, Acinetobacter sp. ADP, Alcanivorax borkumensis, Pseudomonas aeruginosa, Fundibacter jadensis, Arabidopsis thaliana, or Alkaligenes eutrophus, or a variant thereof.

Additionally or alternatively, the methods provided herein are used to decrease expression in said micro-organism of at least one of a gene encoding an acyl-CoA dehydrogenase, a gene encoding an outer membrane protein receptor, and a gene encoding a transcriptional regulator of fatty acid biosynthesis. In particular embodiments one or more of these genes is inactivated, such as by introduction of a mutation.
In particular embodiments, the gene encoding an acyl-CoA dehydrogenase is fadE. In particular embodiments, the gene encoding a transcriptional regulator of fatty acid biosynthesis encodes a DNA transcription repressor, for example, fabR.

[01348] Additionally or alternatively, said microorganism is modified to reduce expression of at least one of a gene encoding a pyruvate formate lyase, a gene encoding a lactate dehydrogenase, or both. In particular embodiments, the gene encoding a pyruvate formate lyase is pflB. In particular embodiments, the gene encoding a lactate dehydrogenase is IdhA. In particular embodiments one or more of these genes is inactivated, such as by introduction of a mutation therein.

[01349] In particular embodiments, the microorganism is selected from the genus Escherichia, Bacillus, Lactobacillus, Rhodococcus, Synechococcus, Synechoystis, Pseudomonas, Aspergillus, Trichoderma, Neurospora, Fusarium, Humicola, Rhizomucor, Kluyveromyces, Pichia, Mucor, Myceliophtora, Penicillium, Phanerochaete, Pleurotus,

Trametes, Chrysosporium, Saccharomyces, Stenotrophamonas, Schizosaccharomyces, Yarrowia, or Streptomyces.

Use of Type VII CRISPR-Cas systems to generate microorganisms capable of organic acid production

[01350] The Type VII CRISPR-Cas systems and molecules and provided herein are further used to engineer micro-organisms capable of organic acid production, more particularly from pentose or hexose sugars. In particular embodiments, the methods comprise introducing into a micro-organism an exogenous LDH gene. In particular embodiments, the organic acid production in said micro-organisms is additionally or alternatively increased by inactivating endogenous genes encoding proteins involved in 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. In particular embodiments, the modification ensures that the production of the metabolite other than the organic acid of interest is reduced. According to particular embodiments, the methods are used to introduce at least one engineered gene deletion and/or inactivation of an endogenous pathway in which the organic acid is consumed or a gene encoding a product involved in an endogenous pathway which produces a metabolite other than the organic acid of interest. In particular embodiments, the at least one engineered gene deletion or inactivation is in one or more gene encoding an enzyme selected from the group consisting of pyruvate decarboxylase (pdc), fumarate reductase, alcohol dehydrogenase (adh), acetaldehyde dehydrogenase, phosphoenolpyruvate carboxylase (ppc), D-lactate dehydrogenase (d-ldh), L-lactate dehydrogenase (l-ldh), lactate 2-monooxygenase. In further embodiments the at least one engineered gene deletion and/or inactivation is in an endogenous gene encoding pyruvate decarboxylase (pdc).

[01351] In further embodiments, the micro-organism is engineered to produce lactic acid and the at least one engineered gene deletion and/or inactivation is in an endogenous gene encoding lactate dehydrogenase. Additionally or alternatively, the micro-organism comprises at least one engineered gene deletion or inactivation of an endogenous gene encoding a cytochrome-dependent lactate dehydrogenase, such as a cytochrome B2-dependent L-lactate dehydrogenase.

[01352] The use of Type VII Cas or system thereof in the generation of improved xylose or cellobiose utilizing yeasts strains

[01353] In particular embodiments, the Type VII CRISPR-Cas system may be applied to select for improved xylose or cellobiose utilizing yeast strains. Error-prone PCR can be used to amplify one (or more) genes involved in the xylose utilization or cellobiose utilization pathways. Examples of genes involved in xylose utilization pathways and cellobiose utilization pathways may include, without limitation, 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. Resulting libraries of double-stranded DNA molecules, each comprising a random mutation in such a selected gene could be co-transformed with the components of the Type VII CRISPR-Cas system into a yeast strain (for instance S288C) and strains can be selected with enhanced xylose or cellobiose utilization capacity, as described in WO2015138855.

[01354] The use of Type VII Cas or system thereof in the generation of improved yeasts strains for use in isoprenoid biosynthesis

[01355] Tadas Jakočiūnas et al. described the successful application of a multiplex CRISPR/Cas9 system for genome engineering of up to 5 different genomic loci in one transformation step in baker's yeast Saccharomyces cerevisiae (Metabolic Engineering Volume 28, March 2015, Pages 213–222) resulting in strains with high mevalonate production, a key intermediate for the industrially important isoprenoid biosynthesis pathway. In particular embodiments, the Cas CRISPR system may be applied in a multiplex genome engineering method as described herein for identifying additional high producing yeast strains for use in isoprenoid synthesis.

Use of Type VII CRISPR-Cas to generate lactic acid producing yeasts strains

[01356] In another embodiment, successful application of a multiplex Type VII Cas CRISPR system is encompassed. In analogy with Vratislav Stovicek et al. (Metabolic Engineering Communications, Volume 2, December 2015, Pages 13–22), improved lactic acid-producing strains can be designed and obtained in a single transformation event. In a particular embodiment, the Type VII CRISPR-Cas system is configured and used for simultaneously inserting the heterologous lactate dehydrogenase gene and disruption of two endogenous genes PDC1 and PDC5 genes.

Further applications of the Type VII Cas CRISPR system in plants

[01357] In particular embodiments, the CRISPR system, and preferably the Cas CRISPR system described herein, can be used for visualization of genetic element dynamics. For example, CRISPR imaging can visualize either repetitive or non-repetitive genomic sequences,

report telomere length change and telomere movements and monitor the dynamics of gene loci throughout the cell cycle (Chen et al., Cell, 2013). These methods may also be applied to plants.

[01358] Other applications of the CRISPR system, and preferably the Cas CRISPR system described herein, is the targeted gene disruption positive-selection screening *in vitro* and *in vivo* (Malina et al., Genes and Development, 2013). These methods may also be applied to plants.

[01359] In particular embodiments, fusion of inactive Cas endonucleases with histone-modifying enzymes can introduce custom changes in the complex epigenome (Rusk et al., Nature Methods, 2014). These methods may also be applied to plants.

[01360] In particular embodiments, the CRISPR system, and preferably the Type VII CRISPR-Cas system described herein, can be used to purify a specific portion of the chromatin and identify the associated proteins, thus elucidating their regulatory roles in transcription (Waldrip et al., Epigenetics, 2014). These methods may also be applied to plants.

[01361] In particular embodiments, present invention can be used as a therapy for virus removal in plant systems as it is able to cleave both viral DNA and RNA. Previous studies in human systems have demonstrated the success of utilizing CRISPR in targeting the single strand RNA virus, hepatitis C (A. Price, et al., Proc. Natl. Acad. Sci, 2015) as well as the double stranded DNA virus, hepatitis B (V. Ramanan, et al., Sci. Rep, 2015). These methods may also be adapted for using the Cas CRISPR system in plants.

[01362] In particular embodiments, present invention could be used to alter genome complexity. In further particular embodiment, the CRISPR system, and preferably the Cas CRISPR system described herein, can be used to disrupt or alter chromosome number and generate haploid plants, which only contain chromosomes from one parent. Such plants can be induced to undergo chromosome duplication and converted into diploid plants containing only homozygous alleles (Karimi-Ashtiyani et al., PNAS, 2015; Anton et al., Nucleus, 2014). These methods may also be applied to plants.

[01363] In particular embodiments, the Cas CRISPR system described herein, can be used for self-cleavage. In these embodiments, the promotor of the Cas enzyme and gRNA can be a constitutive promotor and a second gRNA is introduced in the same transformation cassette, but controlled by an inducible promoter. This second gRNA can be designated to induce site-specific cleavage in the Cas gene in order to create a non-functional Cas. In a further particular embodiment, the second gRNA induces cleavage on both ends of the transformation cassette,

resulting in the removal of the cassette from the host genome. This system offers a controlled duration of cellular exposure to the Cas enzyme and further minimizes off-target editing. Furthermore, cleavage of both ends of a CRISPR/Cas cassette can be used to generate transgene-free T0 plants with bi-allelic mutations (as described for Cas9 e.g. Moore et al., Nucleic Acids Research, 2014; Schaeffer et al., Plant Science, 2015). The methods of Moore et al. may be applied to the Cas CRISPR systems described herein.

[01364] Sugano et al. (Plant Cell Physiol. 2014 Mar;55(3):475-81. doi: 10.1093/pcp/pcu014. Epub 2014 Jan 18) reports the application of CRISPR-Cas9 to targeted mutagenesis in the liverwort Marchantia polymorpha L., which has emerged as a model species for studying land plant evolution. The U6 promoter of M. polymorpha was identified and cloned to express the gRNA. The target sequence of the gRNA was designed to disrupt the gene encoding auxin response factor 1 (ARF1) in M. polymorpha. Using Agrobacterium-mediated transformation, Sugano et al. isolated stable mutants in the gametophyte generation of M. polymorpha. CRISPR-Cas9-based site-directed mutagenesis *in vivo* was achieved using either the Cauliflower mosaic virus 35S or M. polymorpha EF1α promoter to express Cas9. Isolated mutant individuals showing an auxin-resistant phenotype were not chimeric. Moreover, stable mutants were produced by asexual reproduction of T1 plants. Multiple arf1 alleles were easily established using CRIPSR-Cas9-based targeted mutagenesis. The methods of Sugano et al. may be applied to the Cas effector protein system of the present invention.

[01365] Kabadi et al. (Nucleic Acids Res. 2014 Oct 29;42(19):e147. doi: 10.1093/nar/gku749. Epub 2014 Aug 13) developed a single lentiviral system to express a Cas9 variant, a reporter gene and up to four sgRNAs from independent RNA polymerase III promoters that are incorporated into the vector by a convenient Golden Gate cloning method. Each sgRNA was efficiently expressed and can mediate multiplex gene editing and sustained transcriptional activation in immortalized and primary human cells. The methods of Kabadi et al. may be applied to the Cas effector protein system of the present invention.

[01366] Ling et al. (BMC Plant Biology 2014, 14:327) developed a CRISPR-Cas9 binary vector set based on the pGreen or pCAMBIA backbone, as well as a gRNA This toolkit requires no restriction enzymes besides BsaI to generate final constructs harboring maize-codon optimized Cas9 and one or more gRNAs with high efficiency in as little as one cloning step. The toolkit was validated using maize protoplasts, transgenic maize lines, and transgenic Arabidopsis lines and was shown to exhibit high efficiency and specificity. More importantly,

using this toolkit, targeted mutations of three Arabidopsis genes were detected in transgenic seedlings of the T1 generation. Moreover, the multiple-gene mutations could be inherited by the next generation. (guide RNA)module vector set, as a toolkit for multiplex genome editing in plants. The toolbox of Lin et al. may be applied to the Cas effector protein system of the present invention.

[01367] Protocols for targeted plant genome editing via CRISPR-Cas are also available based on those disclosed for the CRISPR-Cas9 system in volume 1284 of the series Methods in Molecular Biology pp 239-255 10 February 2015. A detailed procedure to design, construct, and evaluate dual gRNAs for plant codon optimized Cas9 (pcoCas9) mediated genome editing using Arabidopsis thaliana and Nicotiana benthamiana protoplasts s model cellular systems are described. Strategies to apply the CRISPR-Cas9 system to generating targeted genome modifications in whole plants are also discussed. The protocols described in the chapter may be applied to the Cas effector protein system of the present invention.

[01368] Ma et al. (Mol Plant. 2015 Aug 3;8(8):1274-84. doi: 10.1016/j.molp.2015.04.007) reports robust CRISPR-Cas9 vector system, utilizing a plant codon optimized Cas9 gene, for convenient and high-efficiency multiplex genome editing in monocot and dicot plants. Ma et al. designed PCR-based procedures to rapidly generate multiple sgRNA expression cassettes, which can be assembled into the binary CRISPR-Cas9 vectors in one round of cloning by Golden Gate ligation or Gibson Assembly. With this system, Ma et al. edited 46 target sites in rice with an average 85.4% rate of mutation, mostly in biallelic and homozygous status. Ma et al. provide examples of loss-of-function gene mutations in T0 rice and T1Arabidopsis plants by simultaneous targeting of multiple (up to eight) members of a gene family, multiple genes in a biosynthetic pathway, or multiple sites in a single gene. The methods of Ma et al. may be applied to the Cas effector protein system of the present invention.

[01369] Lowder et al. (Plant Physiol. 2015 Aug 21. pii: pp.00636.2015) also developed a CRISPR-Cas9 toolbox enables multiplex genome editing and transcriptional regulation of expressed, silenced or non-coding genes in plants. This toolbox provides researchers with a protocol and reagents to quickly and efficiently assemble functional CRISPR-Cas9 T-DNA constructs for monocots and dicots using Golden Gate and Gateway cloning methods. It comes with a full suite of capabilities, including multiplexed gene editing and transcriptional activation or repression of plant endogenous genes. T-DNA based transformation technology is fundamental to modern plant biotechnology, genetics, molecular biology and physiology. A

method for the assembly of Cas (WT, nickase or dCas) was developed and gRNA(s) into a T-DNA destination-vector of interest. The assembly method is based on both Golden Gate assembly and MultiSite Gateway recombination. Three modules are required for assembly. The first module is a Cas entry vector, which contains promoterless Cas or its derivative genes flanked by attL1 and attR5 sites. The second module is a gRNA entry vector which contains entry gRNA expression cassettes flanked by attL5 and attL2 sites. The third module includes attR1-attR2-containing destination T-DNA vectors that provide promoters of choice for Cas expression. The toolbox of Lowder et al. may be applied to the Cas effector protein system of the present invention.

[01370] Wang et al. (bioRxiv 051342; doi: https://doi.org/10.1101/051342; Epub. May 12, 2016) demonstrate editing of homoeologous copies of four genes affecting important agronomic traits in hexaploid wheat using a multiplexed gene editing construct with several gRNA-tRNA units under the control of a single promoter.

[01371] In an advantageous embodiment, the plant may be a tree. The present invention may also utilize the herein disclosed CRISPR Cas system for herbaceous systems (see, e.g., Belhaj et al., Plant Methods 9: 39 and Harrison et al., Genes & Development 28: 1859–1872). In a particularly advantageous embodiment, the CRISPR Cas system of the present invention may target single nucleotide polymorphisms (SNPs) in trees (see, e.g., Zhou et al., New Phytologist, Volume 208, Issue 2, pages 298–301, October 2015). In the Zhou et al. study, the authors applied a CRISPR Cas system in the woody perennial Populus using the 4-coumarate:CoA ligase (4CL) gene family as a case study and achieved 100% mutational efficiency for two 4CL genes targeted, with every transformant examined carrying biallelic modifications. In the Zhou et al., study, the CRISPR-Cas9 system was highly sensitive to single nucleotide polymorphisms (SNPs), as cleavage for a third 4CL gene was abolished due to SNPs in the target sequence. These methods may be applied to the Cas effector protein system of the present invention.

[01372] The methods of Zhou et al. (New Phytologist, Volume 208, Issue 2, pages 298–301, October 2015) may be applied to the present invention as follows. Two 4CL genes, 4CL1 and 4CL2, associated with lignin and flavonoid biosynthesis, respectively are targeted for CRISPR-Cas9 editing. The Populus tremula × alba clone 717-1B4 routinely used for transformation is divergent from the genome-sequenced Populus trichocarpa. Therefore, the 4CL1 and 4CL2 gRNAs designed from the reference genome are interrogated with in-house 717 RNA-Seq data to ensure the absence of SNPs which could limit Cas efficiency. A third gRNA designed for

4CL5, a genome duplicate of 4CL1, is also included. The corresponding 717 sequence harbors one SNP in each allele near/within the PAM, both of which are expected to abolish targeting by the 4CL5-gRNA. All three gRNA target sites are located within the first exon. For 717 transformation, the gRNA is expressed from the Medicago U6.6 promoter, along with a human codon-optimized Cas under control of the CaMV 35S promoter in a binary vector. Transformation with the Cas-only vector can serve as a control. Randomly selected 4CL1 and 4CL2 lines are subjected to amplicon-sequencing. The data is then processed and biallelic mutations are confirmed in all cases. These methods may be applied to the Cas effector protein system of the present invention.

[01373] In plants, pathogens are often host-specific. For example, Fusarium oxysporum f. sp. lycopersici causes tomato wilt but attacks only tomato, and F. oxysporum f. dianthii Puccinia graminis f. sp. tritici attacks only wheat. Plants have existing and induced defenses to resist most pathogens. Mutations and recombination events across plant generations lead to genetic variability that gives rise to susceptibility, especially as pathogens reproduce with more frequency than plants. In plants there can be non-host resistance, e.g., the host and pathogen are incompatible. There can also be Horizontal Resistance, e.g., partial resistance against all races of a pathogen, typically controlled by many genes and Vertical Resistance, e.g., complete resistance to some races of a pathogen but not to other races, typically controlled by a few genes. In a Gene-for-Gene level, plants and pathogens evolve together, and the genetic changes in one balance changes in other. Accordingly, using Natural Variability, breeders combine most useful genes for Yield, Quality, Uniformity, Hardiness, Resistance. The sources of resistance genes include native or foreign Varieties, Heirloom Varieties, Wild Plant Relatives, and Induced Mutations, e.g., treating plant material with mutagenic agents. Using the present invention, plant breeders are provided with a new tool to induce mutations. Accordingly, one skilled in the art can analyze the genome of sources of resistance genes, and in Varieties having desired characteristics or traits employ the present invention to induce the rise of resistance genes, with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs.

[01374] **Table 13** below provides additional references and related fields for which the CRISPR-Cas complexes, modified effector proteins, systems, and methods of optimization may be used to improve bioproduction.

Table 13

2-17-2014	PCT/US15/63434 (WO2016/099887)	Compositions and methods for efficient gene editing in E. coli using guide RNA/Cas endonuclease systems in combination with circular polynucleotide modification templates.
08-13-2014	PCT/US15/41256 (WO2016/025131)	Genetic targeting in non-conventional yeast using an RNA-guided endonuclease.
11-06-2014	PCT/US15/58760 (WO2016/073433)	Peptide-mediated delivery of RNA-guided endonuclease into cells.
10-12-2015	PCT/US16/56404 (WO2017/066175)	Protected DNA templates for gene modification and increased homologous recombination in cells and methods of use.
12-11-2015	PCT/US16/65070 (WO2017/100158)	Methods and compositions for enhanced nuclease-mediated genome modification and reduced off-target site effects.
12-18-2015	PCT/US16/65537 (WO 2017/105991)	Methods and compositions for T-RNA based guide RNA expression.
12-18-2015	PCT/US16/66772 (WO2017/106414)	Methods and compositions for polymerase II (Pol-II) based guide RNA expression.
12-16-2014	PCT/US15/65693 (WO2016/100272)	Fungal genome modification systems and methods of use.
12-16-2014	PCT/US15/66195 (WO2016/100571)	Fungal genome modification systems and methods of use
12-16-2014	PCT/US15/66192 (WO 2016/100568)	Fungal genome modification systems and methods of use.
12-16-2014	PCT/US15/66178 (WO 2016/100562)	Use of a helper strain with silenced NHEJ to improve homologous integration of targeted DNA cassettes in Trichoderma reesei.
07-28-2015	PCT/US16/44489 (WO 2017/019867)	Genome editing systems and methods of use.

Use of Type VII CRISPR-Cas Systems in Non-Human Animals

[01375] Modified non-human animals and cells are described in greater detail elsewhere herein. See e.g. "Modified Cells and Organisms" herein. Modifying the non-human animal and/or cell by a Type VII CRISPR-Cas system described herein can, without limitation, treat or prevent a disease or a symptom thereof in the non-human animal or cell thereof, be used to screen the efficacy of one or more agents delivered to the modified non-human animal or cell thereof, confer a desired trait to the non-human animal, generate a model of a condition or disease, or a combination thereof.

[01376] In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example, a mammal, amphibian, reptile, avian, or a fish. In some embodiments, the organism may be an arthropod such as an insect. The present invention may also be extended to other agricultural applications such as, for example, farm and production animals. In some embodiments, the modified animal or cell thereof is a non-human primate. In some embodiments, the modified animal is a non-primate animal.

[01377] In some embodiments, the modified non-human animal is a model system for a disease or condition. In some embodiment, such a modified animal is a modified pig or other livestock animal. For example, pigs have many features that make them attractive as biomedical models, especially in regenerative medicine. In particular, pigs with severe combined immunodeficiency (SCID) may provide useful models for regenerative medicine, xenotransplantation (discussed also elsewhere herein), and tumor development and will aid in developing therapies for human SCID patients. Lee et al., (Proc Natl Acad Sci U S A. 2014 May 20;111(20):7260-5) utilized a reporter-guided transcription activator-like effector nuclease (TALEN) system to generated targeted modifications of recombination activating gene (RAG) 2 in somatic cells at high efficiency, including some that affected both alleles. The Type VII effector protein may be applied to a similar system.

[01378] In some embodiments, the desired trait conferred is one that is commercially relevant, such as disease resistance, meat quality, speed, marbling, muscling, heat tolerance, cold tolerance, improved reproductive characteristics, coat color, height, stamina, improved feed efficiency, increased phytase utilization, among others.

[01379] Cell-based therapies previously described can be used in non-human organisms as well in the same manner. Autologous therapies can be particularly advantageous, for example, in companion animals (e.g. dogs, cats, and horses) to improve quality of life and/or performance as well as treat disease, injury, and disorders.

[01380] 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

[01381] Type VII CRISPR effectors were identified that are related to HNH nucleases. (Figs. 1-3) The novel effector resemble HNH nucleases with comprises a split RuvC and bridge helix with a architecture reminiscent of Cas9. The effectors comprise RuvC-I, Bridge Helix, RuvC-II, HNH, and RuvC-III regions, though recognizable PI and REC1/REC2 domains have not been identified. While split RuvC HNH nucleases are not associated with CRISPR arrays, members of this new family of Type VII CRISPR effectors have a CRISPR array present.

[01382] **Table 1** shows sequences of four Type VII CRISPR effectors and Figs. 4 and 5 show sequence comparisons of VII-cA1 with VII-cA2 and VII-cB1 with VII-cB2. Type VII effectors have been found as small as 578 amino acids in length, comprise RuvCI, RuvCII, RuvCIII, Bridge Helix, and HNH domains, and comprise catalytic amino acid residues corresponding to D10, 893H and 896D of SpCas9, as well as conserved Asp and His catalytic residues in their HNH domains. The exemplified Type VII effectors have 66% sequence similarity (52% identity) to each other, and their respective direct repeats (DRs) have 78% identity. In this respect, the Type VII proteins are distinct from Cas9s, demonstrating 5-6% identity (13% similarity) to SpCas9 and 9% identity (16% similarity) to CjCas9.

[01383] The exemplified Type VII effectors also differ from split RuvC HNH nucleases, having associated CRISPR arrays. Certain Type VII effectors comprise an insertion (Fig. 4) which is proposed to interact with guide nucleic acids.

[01384] The following DRs were identified for the Type VII cA1, cA2, cA3, and cA4 proteins:

VII cA1: GTTACAACGCACGGTGCAGCAACCCTGTGACTGAAAC (**SEQ ID NO: 30**);

VII cA2: GTTGCAATGTATAGTGCAGCGACCCTGGGAATGAAAC (**SEQ ID NO: 31**);

VII cB1: GTTTTGATCTCGCTCCCCCCGGAGCAATTTGGTATT (**SEQ ID NO: 32**);

VII cB2: GTTTTGATCTCGCCCCCCACGGGGCAATTTGGTAAG (**SEQ ID NO: 33**).

Example 2

[01385] The Type VII CRISPR effectors shown in **Table 1** are encoded by the following nucleotide sequences in **Table 14** below. Maps of the VII_cA1 and VII_cA2 regions are shown in **FIG. 7** and **FIG. 8** respectively. The VII_cA1, VII_cA2, VII_cB1 and VII_cB2 nucleotide

sequences are human codon optimized in their respective regions encoding Type VII effector proteins.

Table 14

VII_cA1 **SEQ ID NO: 34**	TGTCAAGTGTGAGGCTTGTGGTTACAGCCAGTGTTAGCCGTTTATTCGGCTAACCAGTTT ATTGAATTGCAGCTTAATCATTTTGAGCGTTATTATTTTCATTTTGCAATATCAAAAAGG GGTGGCATTGGTTGGATAAATTTTATGTCTTGAAAAATGGGATAGGTTATCTGATCAAAA GGGTAGGTCGTAGTTATGCCACGGGGTGTTATTTTTCAGCAACGGCGATAGATCCGAGGG TGATGAGATTTGACACGTTGGATTTAGCGCAGGAATTTATAGAGAGGCACAATTTGGGTG AATTTATTATTGTGGAGGTGTATGTGGATGTTGCTGTTAATTTAGCAAGAAACATTTTTG GCAGTTAGTGTGTTGTTTCAGATTTCAGTCGTAGGGTTGCTGTATCGCTGTAACTCGTAT TTTGTGATAGGTGATAGTGAGTTTCATCAGTAGGGGTTATTGCACCGTGCGTTGTAACTC TGGTGTTACGCAGGTAGTGATAGTTATTGAGGTAGTTTCAGTCACAGGGTTGCTGCACCG TGCGTTGTAACTTTTTTTGCAGCATGTCTCGGCACAGTGATAGTTTGTGTTTCAGTCACA GGGTTGCTGCACCGTGCGTTGTAACGTTTAGTTTTCTGCGTGAAAGCTTATCTTTTGCTT CAGTTTCAGTCACAGGGTTGCTGCACCGTGCGTTGTAACTGAGTTGAGAAACCCCGGATC TGTGGGGTTTTTTGATATTTTCGCACGGTATTTTTCCGAAGATTGTTTTTTGGAAAAAGA AAGCCTGGCTGTCATTATTGATAGGAATTGTTGTATTTATTTACAGCGGTTCGACAAGAG ACCACACGACCCGAGTGAGGGTAAAGGCTACAAATAAGGTTAGGTTTATTATCCGACCCT GTTGTGTTGCGCCAGCTTCAGGCCCTCGAATCCGTTGTTGATAGAGATTGCTATGCATGC ATGCAGCGCTTGTATTTGCGATTGAAGGGATGACGGGAATAGCCGGTTTGTAGCTCGTGG AGGCGTGGGTCACGTGTATATACATGTGATCGTGATACCTGCTTAGGCAGAGAAGATCGA TCTTTTGCTAAAGAGATCGATAAGGGGGTAACCTCAATTTTAACGTATGTTTTAAATGCT GATAATTCTCCGTTGATGCCTTGTCATCCTGCTCGCGCAAGAGAGCTCGTGCGGAAAGGC AGAGCCTACTGGGTCAAGAGGGACACCGTGAAGCTGAAAACCTTCCTGCAGAACGCCGTG TGCCAGCACACAGTGCTGGGAATTGATGCCGGCGCTATCCACATTGGACTGGCCGTGGTG GCCAAGAGAGAGAAGAGAAAGCCCAAGGTGCTGTTCAGAGCCCAGGTGGACGCCAGACCT ATGAGCGAGATCAAGAGCGTGATGGACAAGCGGCGGAGCCTGCGGAGAAACAGAAGAAGC AGAAAGGGCTACAGACAGCCCTGGTACAGCCCCACAAGAAGAGTGGGACTGACCGCCAAG CTGTTCAAGAAGTCCGGCGAGTTCCTGTGCAAGATCCCTCTGGCCGAGGGCAAGAGACTG CTGAAGCCTGACAAAGGCGGCAACAGACAGGCCAAAAGACTCGGCGGAAACAAGTACCAG CTGAGCACCATGAAGGAAATCAAGGCCCCTAAGCGCGAGAAAGAGCCCGGCAAGAACGAC ATCGTGGTGTACAGATTCGACGGCAGACCCCTGACCTGGTTCAGCCCTAAGAGCCGGAAG TTCGTGAAGCAGCTGTGCAAGTGCCTGAGAGGCTACGGCAAAGTGCGGAGACTGCCTGTG GCCGACGTGAAGAAAGAGGGCGACAAAGTCAAGAGCATCTGGCTGAAGTCCAGACGGGCC GAGAGACTGATCCCCAAGAACATCCTGCCTGAGGTGGTGCCTCCATTCCGGAAGATGGGC CAGAAGAGGAAGCCTTCTGGCTGGCTGGCCCCTTCTGCCAACGCTCTGGTCAACGCCTAC TTCAACGGCATCAAGCGCGTGGGCAGCTACGTGCCAATCGCCAGCGTCAGATTTGAGGAC GTGGCCTTCGACACCCGGAAGCTGGAAGATTGCCTGGACGGAAAGGGCGACTGCAGCACC GTGAATTACGCCCGGCAGAACCGGAAGGATTTCGTGCTGGCTAGAGATGGCTGGAAGTGC GTGTACTGCGGCAGAAAGAGCAACTTCAGCAAGAAGGGCCTGAGACTGACCCAGGATCAC GTGGTGCCCAAGTCCAAGAACGGCAGCAACGCCATCGGAAACCTGGTGGCCTGTTGCGAG AAGTGCAACAGCGATAAGGGCAGCCTGGACCTGGACGACTACCTGAAGAAGAGAGAGCAG CAGGGCGATAAGCCCAGCGACACCAAGAGAATCAAGAGATACGTCGTGGGCCTGAGCCAG AAGTCCGAGAGCTTTAGAAGCAAGGCCCATGTGGGCCAAGTCAAGACAAAGGTGCTGCGG CTGCTGCGCGAGCGATTCGAGGATGTGTCTATCACCAAGGGCTACGTGACCAAGAAAGAC AGAGACTGGATGGGCCTGACCAAGAGCCACGACTGTGATGCCGTGGCCATTGCCGCCTAT GGCGAGCGAGTTGGCTTTATGCCTCCTCTGAGAACCTACCACGTGCGGCGGTATACCTGG GAGGGCTCTAGAAGAAGGCAGTACAGCGCCAATCCTATGAAGAAATACTCCTCTAAGTGC GTGCTGATCGCCAAGGGCACCCCTAGAGAAAGATGGGTGCAGCCCAACAGCTACAACCCC TACGCCGATACCAAAGTGGGCAGAGTGTACAGCAGAGATGTGGTGCTGGCCGTCGGAGGC AACAAGGGAACAGAACCTGTGCTGGGCTACATCCGGGGCATCTGTAAAGATGGCGCCCTG TTCATCAGCAGACTGCCTACAGGCGGCAAGAAAAAAGAGCGGAGCAAGCGGCAGGCCAAA TACGTGAAAAAGATCATCAGCCGGTGCAGGGGCCTGATGGAAGTGCTTTAAAATTGCCGG GATAAACCCGGCAATCAGTTTGTTTTATTCCAACCTTTATTTTTAAGGTGATTCGGTATC CTTGTTAAATAATAAGGGGGTGTAATAAATGAATCTGGCTTGTGTGATAAGCCAAGGGTT TAAAGTAAGGATCGTGATGGATGAGTTTGATTATGATCCTAGGGGTGATGATAATCTGGG CACGATGGTTTGCTCTCATCGAAGATATCGTTTGGGTGATGAGCAGGCCTGGAATGTCGA

	AGGGTATTCCAGCTGGGATCAGTGGTTGGATGGAGAAATCCTGAAAACCTATGGTGAGGG CAATGTCATATATTTACCTTTGTATCTATATGATCACAGTGGGCTTCGGATGAATACAGT AGGTTTTAGTTGTCCGTGGGATTCTGGACAGGTGGGGTGGATATATGTAACCAGGGAGAG GGCAGCAAAAGAGACTGGTTGTGCAAAAAGGGCCAGAGAAATACTTCAGGATGAGGTAGC AACGTATGATCGCTATCTTTCTGGATGTAGTAGTTATGGCTACATTGTTGAGGATATGGA TAACAATGTAGTTGAGTCTTGCGGCGGTTTTATGTCCGTAGATGAAATAAAAGATCACCA TAGTAAGTATAGTGATCTTTTGAAACATTGTATAAGTATTTGCGATTCATAGCAAGAGTA TGGTTTTGTCCATGCTTTGTATTTAATTAAATCGGCGTAAAGCCGGTTATTTTTTATGAA TTAAATAAAAAGGGGATGAAGCAATGTTGTAC
VII_cA2 **SEQ ID NO: 35**	ACTTCGATGAAGAGGCCTTCCGCTATATAGAATCCTATTCCGCCGTAATCATACCTGCTG CTAAGGCAGCTTATGATGATTTCAATACATGGCTGAACTCTCCCAGCAAGGATGGACCGT TGCTAACCACCCTCAAAGAAAATGCTGCGGTCATCAAAGAGATTAACCAACGGTACTGGA CAGAAGAATTTCCTGATGACAAAGATATAGAAAAATGGCGGGTCCAGCGTAACCTTAACG GGCGCGAGTGGGTCATAAAAGGCCGAGAACTGGCAGACGCGCTTTATTTTGTTGCGGCGA ACACCGATACTTTTGCTTACCAGTTGGAGCTGATCGTTGAAAATCAAGGTCCACTCTCGG ACGAACAACTTGAATCGCTAATAGCATGCTCACAAGATGTGAAATCCCAAGTTGGCAGGC TGAACAGATTGCTTTTCACCGGTTACGTGCGGTCGGTAATTGACCAGTGAAGACGTTATG TTCTATTTGTTTTTAATTTAAAGCCATTGTGCCAAGGTTTATCGGGTGATGACGACCAGT ACTCCCAAACCCTGTTAAGAGGTTTTCTGGTATATTAGGATAAATGTATAATATATTCAA GGATTGGCAAGTTGTTATATTTACCCTTTGAAAAAGAGGTTGCTATTTCGCAATACTTCT GTAACAATATCCTTACTATGTTGGATCTTGTTCTTAGTGGTACAGAAAAACTTAAGGGAA AGAGGTGATGATGTTTAGATATTAATTTTCCTTAGATTAATTGAAACCGAAACAAAGGTG GCGTTTCATTCCCAGGGTCGCTGCACTATACATTGCAACGCAGCCATCATGGGCGTTACG CCGCAATTCTTACGCATGTTTCATTCCCAGGGTCGCTGCACTATACATTGCAACCGGCAG AATCGTTCGAGGAAAGAACGGGAATTAAGCCGTTTCATTCCCAGGGTCGCTGCACTATAC ATTGCAACCCAAAGAAAGAAACTTAAAAATTAATTATTTATATTTCTAATATAAAGTGTT ATTTTCGAAGATTTGATTTCTGCTCAAAAACCTTGGGCATCTTCGGAAAATATGGCCCTG GGAAGAGAAACCACACGACCTAGTAACTAACTAGGTAAAAGCTGCAAATAAGTGCTTAGG GGATTATCTCCCCTATCCGAACTTGGGGTGTTGCGCCAGCTCCAAGCTCTCGAATCCGTT GTTGACAGAGGCTGCTCTATGAGGAATGCAGCAACCAACATTAGGGCTGTCAAAGGGATG ACGGGAATAGCCGGTTTGCAGCTCGTGGAGGCGCGCTGTTGCCCCTTACCAATAATAGTA AGGGAGAAAGCTTCATTGAAGCTGTAAAGGAGGTAACTCCTTATTATTACATATGTGATG AACCAAGATAATAAGCCACTTATGCCTTGTCACCCTGCCAGAGCCAGAGAACTCGTCAGA AGAGGCAGAGCCTACTGGGTCCGACGGGACACAATCAAGCTGAGCGAGTACAAAGAAAAC GCCGTGCTGCAGCGGACCTTCCTGGGAATTGATGCCGGCGCTGAGCACATTGGACTGGCC GTGGTTGCCGAGAGAGAGAAGCACAAGCCCAGAGTGATCTTCAGAGCACAGGTGGAAGCC CGGCCTATGCGCGAGATCAAAGAGTGCCTGGACCAGCGGCGGAGCTACAGAAGAGGAAGA AGATCCCGGAAGGCCTACAGACGGCCCTGGTTTAGCCCTACCAGACGGGTTTCACTGAGC GCCAGACTGTACAACAAGCAGGGCGAGTTCCTGTGCAGAATCCCTCTGGCCTACGGCAAG TACCTGACCAAGCCTGACAAGAGCGGCAGACCCCAGGCCAAGAGACTGTACGGCAACAAC TACCAGCTGCTGACCAGCAAGAACCACGAGAGCCCCATCAGAGAGAAGAAGCCTGGCAAG CAGGACATCCTGGTGTACCGGAACGATGGCAGACTGCTGACTTGGTACGACCCCAAGAGA GGCGACCTGGAAGGCATCGAGAAGAGACTGCTGGGCACCTCCAGACGGCACAACATCCCT ATTGCCAGCCTGGAAAAAGAGGGCTCCAAGAAGATCATCCGGCTGCTGTCTGTGTCCGCC GAGGACGCCATTCCTAAAGCCGCTCTGCCCAATCCTATGCCTCGGTTCAGAAAGATGGGC CAGAAGCGGAAGCCTCCTGGCTGGCTTTCTCCAAGCGCCAACTCTCTGCTGAACGCCTAC TACCGGGGCATCAAAGAAATCGGCAGCTACGTGCCCATCAGCCTCGTCAGACTGGAAGTG GCCGCCTTCGACACCCAGAAGATCGAGGACCCTACAATCCACGGCGTGGGCTACCAGAAG CCTGCTCTGTACGCCAAGCAGAACCGGAAGGACTTCGTGATCGCCAGAGATGGCTGGCGC TGTGTGTACTGCGGCGTGAACAATGTGCCCCTGACACAGGACCACATCCAGCCTCAGTCT AAAGGCGGCCCTGACAGAATCGGCAATCTGGTGGCCTGCTGCTTCAAGTGCCAGAACGAG AAGTCCAATCTGACCCTGGACGAGTTCCTCAAGACCAAGTCTCCCCAGCAGGCTAAGCGG ATCAAAGAATACGTGTCCAGCATCACGCAGAAGAACGAGAGCCTGCAGAAGAATGAGTCC CTGAGAAGCGCCGCTCACGTGGGCCAGATCAAGTCTAAACTGCTGCAGGGCCTGATCGAC CTGTTCGGCGACAGCAACGTGAAGAGAACCTGGGGCTACATCACCAAGCGGGATAGACTG GCCCTGCATCTGCCCAAGAGCCACGACTACGACGCCATTGCCATTGCTAGCCTGGGAAGA AGAATCGACACCGAGAACATCCCTCCACTGCGGCTGTACTACATGCGGAGATACACCAAG GACAGCCCCAGGCGGCAGAAGTACAAGGCCAATCCTCTGAGCAACGCCAAGAAGGATAAC AAGAATCTGGCCCCTCTGTTCATCGGCCAGGGCAGCAGATCTAGATGGGTGCAGCAGGTC

	CAAGTGAACAGCTACGCCGAGACACCTGTGGGAATTGTGCGGCTGAGGGACCTGATCAAG ACCATCGACGGCAGGATCGGCTACGTGCGGAAGATCAACAGCGGCAGAAAGCCCAAGCCT GGCCAGAAACCTGTGCGCATCCTGACACTGAGCCACAAGCCTCAGGGCAAAGAGGGCCAG TTTAGCACCAGCGCCAACAAGGTGGAAAAGATCATCAGGCGGAGATGCGGCCTGCTGGAA ATGCTGTACGTGTGAATAATGTTTTAGGAGTTCATTAAATGTCTCAGTATTTTGAAACAG TTTTAGTTAGTAGTAATCTCTTACTTTCCGTTTAAATTCGACAAGTTGTTTCTTAAAGTA AAGGCAAACCGTTAAAGAATCAGTATTCTCTACATCTTCAGCTTTATAACGTAACTGCTT AGCTCCAATGAGCACTTCAAGTGGAGTTGATCCGGTTCAGTGGGTACAACGAGGCTTGGT TTCCACGCTGTTTCTTTTAATTTCTCTTGTTTGAACCACCTCTCTTCGAATTCAGCTGGG GATAGATATCCTATCCTAGGGCGGAATGTCGCCGCCTGCGGTTGTAGAAACCCTCGATAT ATTCAAAGATCGCAGACCGGAGACATTGACGAGTAGGGGTGTCCATCAAACCGGGTCAAG CCCATTATTTATTTTTATCCTTAAGTTTCTTAAGCAAATCCTCTATCTTCTCCCTCTGCT TTTTCTCCCCTTCCGGGTCTAGCTCAAAGGTAACTTTAAGCCAGCTACCTTCACGAGATC CCTTGGGAAGCAACTTGCGAGGGATGTTTAATTTTATTTCTTCGTCTCCAAAAAGAACTA CAGCAAAATCACCTTCAAAACGGTCTATAACAGCTTTCATTATTGTCTCTCCCTTCTTAA TTACTTAGGTGGATTACAAACTTTACAAGGAACGTATCCTGCACTTTTTGCTTCCTCGAC AGAATTAAACCATA
VII_cB1 **SEQ ID NO: 36**	CGCACTTTTTCCGGATCCATGATGGCCTGCGACATAATTATTTCCTGAAATGCCTGTGTC GCGGCCAAGGTGCCTCATCCGCCGGCATAAATCCACCCGCGTTTTTTCTTTTCTTCACCT CATTCTGAAATACAATGCGGCTTCGGTGGCCGATGTGGCGGAATTGGCAGACGCGCCAGA CTCAAAATTTGATTTTTACCCATTTTTTCGGGTTGCAATCCATCGCTTGAAATAGCGTCA ACAACATTGATTTTATTGGTCAAAACTTGTTTCCAGCCTTTCCGCTCAATGACCCGTTAG GTGTCTCGAAAAAGGTGCAACTCTAGCACAAAAACTAGCACAAAATCAAATTAGACCAGA TGGCCCTCTTTTTGTGTTAGCTTGTGCATGTCCAACTGCTATAAATAGTGACGAAGCAAG CGCACCCACGGATTGTTTGACCGTCGGAAAAATTCCCCGTTCGCGTTCCCGTGCTTGTCA TCTGTGTTGCTGATTTGAAGGCCAAGCCACATGCCAGTCACGAGAACACTCACCCCATTC AAGACTGGTAGTGTCAAATTTGCAACCAGCACCGCGCCCGCGTAACCCTTCGCGGAACTG GTTTCAAACTCGGCACCACCAGGACAAGTCTCGCGCGAACGGACGGCTTCCGTTGGCTTG AATTGTAGCCGCGAATGGAATATCCCGGTTCGGCCAACTGGTGCTGCGGGAGGAATCATC CTGCCCCGAATGGGATAACATGACCAGAATGGCCCAGCTGGCCAGACAGCTGAGAGAAGA GGCTGCTAGATGGCTGGGAGTGTTCGGCGACGATGCTAACGTGCGGCAGAAGATCGGCAC CCCTACAGGCTACCAGACCAGCGTGTGTAGAGAGAGCTGGGGCGAGAACCTGCCTAAGAA GGACAGAAGCAACCTGCGGCACCACCTGTGGGATGCCGCCGTGATTTCTCACATCCCTCC AGGCAAGGGCATGAACCTGACCAATTGCGCCGGCATCTTTCTGCACACCGGCACCAATCA GCACGGCAACATCACCATGAGAGCTATCCCTGGCCTGGGACCTGACCTGAACAGCTTCGA GAACGGCACCAAGGACCAGTGCCTGGTGGACGAAAGAAGGGCCACCAAGAGCAAGAAGCA GAGATTCGACCAGACCATCTACTCCCCACCTGACACCGAAGGCGTGATGTGGGCTAGAGA TCCCCTGGATAAGCTGGCCACAAAGAAGAACCAGAGCGCCGAGGGACTGCTGCAGCTGCT TAGAGATGCCGGAATGGACGAGAACCAGCTGCCTAGCTCCAGATTCGAGGAATGGTGGAA TGGCGTGCAGGCCCAGTGGTTCACCAGAAACGATGCCCAGGCTCTGTTTCAGACCCTGGA AATCCCATCTCCTAAGATCATCAGCGAGGACGCCCTGAACAAGTGGTGGGGCAACTACGA GAAGAAGAAGATCAACTACAAGAGCCTGCGGCTGCTGCTGAAAGAGTCCGGCATCACACC TGTGGAAGTGTCCGACCTGAGACTGCAGAATGCCCTGAGCGGACACAGCGAGCCTGGAAG CCTGACAAGAAAGGACGGCACCCCAATTAGAGGCGTGCGGGGAACAGCCAAGCAGATGAC ACCTATGGCCGTGATTCCCCACCGGAACCACAAGGGCGAGACAATCGGCTTCAAGTTCGC CAAAGAATCCTTCGTGCGGGCCGAGATCTGGGTCACCGAGAAAACCACCAAGGATGGCAA GGTGGAACTGGACTACCACTGCCGGCTGATCCCTCATCCTAGAGGCCTGGCTAACCTGCA GCTGAGGGTCATGAAGTGTACCGGCCAGCCTCTGACCTGGGAGAGAAGGCTGACAGACGA CGAGCTGGTGGAACTCGGCCTGGCCAAAAAACTGGCCACTCTGAAGAACGAGAGAGAGCG GCTGAAGAACAGCCACGCCAAGGCCGTGGAAAAGTGGCAGAAGTCCCAGCAGAAGAAAAA GGACAGCGACACCCTGGGACTTGCCGCCGATGCTGCTACAAGCAGCCCCAAAGAAATCCC CGAGCTGTGGCTGAGCCTGAGAAAGATCTACACCGGCCTGCCTCCACACGCCAAGAGACT GACAAATGCCAACGGCGCCGACATCAGCAGAGTGACCAAAGGCGATCTGCTGTTCGTGCC CCTGAGAAAAGTTCCCGGCGCTGGCAAAGGCCACGGCAAGTTCTGTAAAAGGGGCGAAGT GGCCACAGGCGGCAGCTGCTGGTTTAGAGTGTCTGCCATTAAGGCCGCTGGCGAGATCGA ACTGCACCTGGCCGAGTTTAAGCTGCCCAAGATCAAGGATGACAAGAAGGCCACAGAGCA GCAAGAGTGGCTGACCACCATTTGGGAGCAGAGGCCTAGCAGCGAGGATGACATTGCCTG GCTGCTGGAACAGAGCAGAGCCCACGATCAGAGCCCTCATCCAGTGAAGTGAGGAGCAAG ACCTGCACGTCACGGATGGCAAGTAGGCAAAACCCGCTCGTGTAGAGCGGGTTTTGCGGC

	ACGAATACCAAATTGCTCCGGGGGGAGCGAGATCAAAACGAAAACACCGACGGGCAGGCC AAGTTCCAAAAAATACCAAATTGCTCCGGGGGGAGCGAGATCAAAACAATGGCGAACAAC GCACGCGCTGACGCCAGAAAATACCAAATTGCTCCGGGGGGAGCGAGATCAAAACCGGCG GCGGGCAAGCCGCTGCGGGTTTTTGAATTTTTCACGGTTCAGACAATTTCCGTCCGCCAT TTTTTCGGCAGCGGCTTGCCCGCCGCCGGTTTTAATTTCTCCCGACCCCATGGAACTGAA CTTTTAAGAAAAAAAAATCCGATTATTCCCGCCACAGATTCACGGTTTTGCAGCCTGGTA AATGTTTTTTATTCCGAGTGATAACGGCCAGACCAATCAACTTGCCACCATCT
VII_cB2 **SEQ ID NO: 37**	CAGAGGCCGTCGTCCTTTGAAACTGCTGATGACTCATGGGCAGAGGCTTCAGAGCCAGCA TCAAGAGCAAGAACCACCACAACCTGAGCTACGGCGAGACAAACGGACTGGCCAGACAGC TGAGAGAGGAAGCCGCTCACTGGCTGGGCATCTTCGGAGATGATGCCAAAGTGCGGCAGC GGATCGGCACACCTACAGGCTATCAGACCAGCGTGTGCAGAGGCAGCTGGTCCATCAATC TGCCCGCCGACTTCTACCCCAAGAAGAACAGAAACAACCTGCGGCACCACCTGTGGGATG CCGCCATCCTGGCTAATATCCCACCTGGCAAGGGCCTGAACCTGACCAATTGCGGCGGCA TCTTCTTTCACGCCGGACGGAACAAGCACGGCAATATCGCTATGCAGGCCCTGCCTGGAC TGGGCCCTGATCTGAATGCCTTTGAGAAGGCCACCAGCAACCAGTGCCTGGTGGAAAAGA AGCAGACCACCAACAGCAAGAAGTCCCGGTTCCAGCAGTGGATCTACGGCTTCCCCGACA AGAACGGCGAGATGTGGGCCAGAGATGCCATCGAGACACTGGCCGACAAGCCCAACCTGC TGAGCAGACTGAGAGATGCCGGAATCGACGAAGCCCAGCTGCCAACCAGACACTTTGAAG TGTGGCAAGAGCGGCGGCAGGCCCAGTACTTTACAAGAGAAGAGGCCTGGGTCGCCGTGC AGGCTCTTGCACTTCCTGCCGAACACGCCATTCCAGTGTCCACCTTCGACGAGTGGTGGA ACGACAGACTGAAGTCCGACAAAAAGAAAGTGACCGACAAGTCCCTGCGGGCCCTGCTGG CTAAAGCTAGAATCCCTAAAGCCGCCGTGACCGACGTGAAGCTGACAGGCGTGCTGGTCA ACAGAGGCGATCCTGGACCTCTGACCAGAAAGGACGGCACCCCTATCTATGGCGTGTCCG GAACAGCCAAGCAGATGACCCCTATGGCCGTGGTGCCCCACAGAAACCACAAGGGCGAAA CCATCGGCTTCAAGCTGGCCACCGAGAGCTTCATCAGAGCCGAGATCTGGACCACAGAGA AGCGGGACAAAAAGGGCGACATCGTGAAGGACGACACCGGCAAGCCTGTGCTGGAATACC ACAGACGGCTGATCCCTCATCCTCGGGGCCTGAAGAACCTGAGACTGCGGACCATGAGAT GCACAGGCAAGAGACTGAGCTGGGAGAGAAGGCTGACCGACGAGGAAATCGTGGAACTGG GACTGAAAGAACACGCCGAAGTGAAGCGGCTGCGGAAGGACTATGACAAGGCCGTGTCTC AGCACGTGAAGCACACCAGAGCACAGGCCCCTGAATGTGAACTGGCCCTGTGCCAGAGCA AAGTGACCATGCCTGGACAGCCCAAGCCTCCTGTGATCAGCCTGCGGAAGATCTACACAG GCCTGCCTCCAAACGCCAAGCGGCTGACAAATGCTCAGGGCGCCGACATCAGCTGCTTCG CCAAAGGCGATCTGCTGCTGGTGCCTGTGTCTCAGAGCGGCGCCATCTGCGAGTCTGATA AGGCCCCTTACAGAGAGCTGTGGTACAGAATCACAGCCCTGAAGGCCAACGGCCAGGTGG AACTGAAACTGGCCGAGTTCAAAGAGATCAAGCGGCCCAGCGAGAAAGAGCTGAAAGAGG GCAAGTCTCTGACCCCTGAGCAAGAGTGGTTCGCCGATGCCTTTCTGCAGCAGCCCGGCA ACGATGAGCTGATCGCTGGACTGCTCCAGCGGACCAGAGTGAACGATCAGAGCAGCCACA GCGTGAAGTGAGGAACAACATCTTCACGTAAACGATGGCAAGTGAGCCGAACCCGCTCTA TTCAAGCGGGTTCGGCTGCGCCCTTACCAAATTGCCCCGTGGGGGGCGAGATCAAAACGC GACAGGCGCGCAGATCCTGAAGGCAACGGACTTACCAAATTGCCCCGTGGGGGGCGAGAT CAAAACCTATCGTCGTGAGTGGCCGCGAGTCGGGAAATCTTACCAAATTGCCCCGTGGGG GGCGAGATCAAAACGACGTTGCGACTGCGCCAGGGCGCACTGCGAGTTTTGTAGGGGGTG TGGTAGTCTGTTGGCAAAGGTGTGACCGACGAGAATTGCGCGGGGCTGACTGTGGGGACA GTCTGCCACAAGGAACTAACGATTCCCTACCACCGCGAA

***

[01386] Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the invention as claimed should not be unduly limited to such specific

embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention. This application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure come within known customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

[01387] Further attributes, features, and embodiments of the present invention can be understood by reference to the following numbered aspects of the disclosed invention. Reference to disclosure in any of the preceding aspects is applicable to any preceding numbered aspect and to any combination of any number of preceding aspects, as recognized by appropriate antecedent disclosure in any combination of preceding aspects that can be made. The following numbered aspects are provided:

1. An engineered system comprising

i) a Type VII CRISPR effector protein,

wherein the Type VII CRISPR effector protein is capable of complexing with a crRNA to form a CRISPR-Cas complex, wherein the crRNA comprises

a) a direct repeat (DR) sequence; and

b) a guide sequence capable of hybridizing to a target nucleic acid sequence.

2. The system of aspect 1, wherein the Type VII CRISPR effector does not have a REC1 domain, a REC2 domain, or both.

3. The system of any one of aspects 1-2, wherein the Type VII CRISPR effector comprises a bridge helix domain, a RuvC domain or RuvC-like domain, and an insertion between the bridge helix and RuvC or RuvC-like domain.

4. The system of aspect 3, wherein the insertion is capable of binding RNA.

5. The system of any one of aspects 3-4, wherein the Type VII CRISPR effector comprises a RuvC I domain, a RuvC II domain, or both.

6. The system of any one of aspects 1-5, wherein the Type VII CRISPR effector comprises an HNH domain.

7. The system of any of aspects 1-6, further comprising the crRNA.

8. The system of aspects 1-7, which further comprises a tracr RNA and the CRISPR complex comprises the Type VII CRISPR effector protein complexed with the crRNA and the tracr RNA.

9. The system of any one of aspects 1 to 8, which comprises two or more crRNAs.

10. The composition of any one of aspects 1 to 9, wherein the guide sequence is designed to hybridize to a target nucleic acid in a prokaryotic cell.

11. The system of any one of aspects 1 to 10, wherein the guide sequence is designed to hybridize to a target nucleic acid in a eukaryotic cell.

12. The system of any one of aspects 1 to 11, wherein the Type VII CRISPR effector protein comprises one or more nuclear localization signals (NLSs).

13. The system of any of aspects 1 to 12, wherein the Type VII CRISPR effector protein is a nickase.

14. The system of any of aspects 1 to 13, wherein the Type VII CRISPR effector protein is catalytically inactive.

15. The system of any one of aspects 1 to 14, wherein the Type VII CRISPR effector protein is coupled to or associated with one or more functional domains.

16. The system of aspect 15, wherein the functional domain has methylase activity, demethylase activity, translation activation activity, translation initiation activity, translation repression activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity, single-strand DNA cleavage activity, double-strand DNA cleavage activity, molecular switch activity, chemical inducibility, light inducibility, nucleic acid binding activity, deaminase activity, or a combination thereof.

17. The system of aspect 15, wherein the functional domain cleaves the target nucleic acid.

18. The system of aspect 15, wherein the functional domain modifies transcription or translation of the target nucleic acid.

19. The system of aspect 15, wherein the functional domain comprises an adenosine deaminase, or catalytic domain thereof, or cytidine deaminase, or catalytic domain thereof.

20. The system of aspect 19, wherein the adenosine deaminase or catalytic domain thereof, comprises one or more mutations that increase activity or specificity of the adenosine deaminase relative to wild type.

21. The system of aspect 20, wherein the mutation confers an ability of the adenosine deaminase to deaminate cytidine.

22. A vector system comprising:

one or more polynucleotides each encoding one or more components of the system as set forth in any one of aspects 1-21, wherein the one or more polynucleotides are each optionally included in one or more vectors.

23. The vector system of aspect 22, further comprising one or more regulatory elements, wherein each regulatory element is operably linked to one or more of the one or more polynucleotides of the vector system.

24. The vector system of aspect 23, wherein a first regulatory element is operably linked to a polynucleotide encoding a Type VII CRISPR effector protein, and optionally comprises a second regulatory element operably linked to a polynucleotide encoding the crRNA, when optionally present.

25. The vector system of aspect 23, wherein a first regulatory element is operably linked to a polynucleotide encoding a Type VII CRISPR effector protein, a second regulatory element operably linked to a polynucleotide encoding the crRNA when optionally present, and a third regulatory element operably linked to a polynucleotide encoding the tracr RNA when optionally present, or a second regulatory element operably linked to a polynucleotide encoding the crRNA and the tracr RNA when both are optionally present.

26. The vector system of any of aspects 22-25, wherein the polynucleotide encoding the Type VII CRISPR effector protein is codon optimized for expression in a eukaryotic cell.

27. The vector system of any of aspects 22-26, wherein the polynucleotides encoding the Type VII CRISPR effector protein, the optional crRNA, and optional tracrRNA are comprised in a single vector.

28. The vector system of any of aspects 22-27, wherein the one or more vectors comprise viral vectors.

29. The vector system of any of claims 22-28, wherein the one or more vectors comprise one or more retroviral, lentiviral, adenoviral, adeno-associated or herpes simplex viral vectors.

30. A delivery system configured to deliver a one or more components of the engineered system of any of claims 1-21, comprising

i) a type VII CRISPR effector protein, and optionally

ii) a crRNA comprising a direct repeat (DR) sequence and a guide sequence that is designed to hybridize to a target nucleic acid sequence, and

whereby there is formed a CRISPR complex comprising the Type VII CRISPR effector protein complexed with the crRNA.

31. The delivery system of claim 30, which comprises one or more vectors or one or more polynucleotide molecules, the one or more vectors or polynucleotide molecules comprising one or more polynucleotide molecules encoding the Type VII CRISPR effector protein and one or more optional nucleic acid components of the non-naturally occurring or engineered Type VII CRISPR system.

32. The delivery system of any one of aspects 30-31, which comprises a delivery vehicle comprising liposome(s), particle(s), exosome(s), microvesicle(s), a gene-gun or one or more viral vector(s).

33. A method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with an engineered system as in any one of aspects 1-21 or a vector system as in any of claims 22-29,

wherein the guide sequence directs sequence-specific binding to the target nucleic acid sequence, whereby the target nucleic acid sequence, the expression of the target nucleic acid, or both is/are modified.

34. The method of aspect 33, wherein modifying occurs *in vitro*, *ex vivo*, or *in vivo*.

35. The method of any one of aspects 334, wherein modifying the target nucleic acid comprises cleaving the target nucleic acid.

36. The method of any one of aspects 33-35, wherein modifying expression of the target nucleic acid comprises increasing or decreasing transcription or translation of the target nucleic acid.

37. The method of any one of aspects 33-36, wherein the target nucleic acid is in a prokaryotic cell.

38. The method of any one of aspects 33-36, wherein the target nucleic acid is in a eukaryotic cell.

39. A cell comprising a modified target nucleic acid of interest, wherein the target nucleic acid of interest has been modified according to the method of any of claims 33-36.

40. The cell according to aspect 37-39, wherein the modification of the target nucleic acid of interest results in:

the cell comprising altered expression of at least one gene product;

the cell comprising altered expression of at least one gene product, wherein the expression of the at least one gene product is increased;

the cell comprising altered expression of at least one gene product, wherein the expression of the at least one gene product is decreased; or

the cell comprising altered expression of at least one gene product, wherein the expressed gene produce is altered.

41. A cell comprising: an engineered system as in any one of aspects 1-21 or a vector system as in any one of aspects 23-29.

42. The cell of any of aspect 41, wherein the cell is a prokaryotic cell.

43. The cell of any of aspect 41, wherein the cell is a eukaryotic cell.

44. The eukaryotic cell according to any one of aspects or 38 or 43, wherein the cell is a mammalian cell or a human cell.

45. A cell line of or comprising the cell according to any one of aspects 37-44 or progeny thereof.

46. A multicellular organism comprising one or more cells according to any one of aspects 37-44.

47. A plant or animal model comprising one or more cells according to any one of aspects 37-44.

48. A gene product from a cell of any one of aspects 37-44 or the cell line of aspect 45, the organism of aspect 46 or the plant or animal model of aspect 47.

49. The gene product of aspect 48, wherein the amount of gene product expressed is greater than or less than the amount of gene product from a cell that does not have altered expression.

CLAIMS

What is claimed is:

1. An engineered system, comprising:

a Type VII CRISPR effector protein,

wherein the Type VII CRISPR effector protein is capable of complexing with a crRNA to form a CRISPR-Cas complex, wherein the crRNA comprises a direct repeat (DR) sequence and a guide sequence capable of hybridizing to a target nucleic acid sequence.

2. The system of claim 1, wherein the Type VII CRISPR effector does not have a REC1 domain, a REC2 domain, or both.

3. The system of claim 1, wherein the Type VII CRISPR effector comprises a bridge helix domain, a RuvC domain or RuvC-like domain, and an insertion between the bridge helix and RuvC or RuvC-like domain.

4. The system of claim 3, wherein the insertion is capable of binding RNA.

5. The system of claim 3, wherein the Type VII CRISPR effector comprises a RuvC I domain, a RuvC II domain, or both.

6. The system of claim 1, wherein the Type VII CRISPR effector comprises an HNH domain.

7. The system of claim 1, further comprising the crRNA.

8. The system of claim 1, further comprising a tracr RNA and the CRISPR complex comprises the Type VII CRISPR effector protein complexed with the crRNA and the tracr RNA.

9. The system of claim 1, further comprising two or more crRNAs.

10. The composition of claim 1, wherein the guide sequence is capable of hybridizing to a target nucleic acid in a prokaryotic cell.

11. The system of claim 1, wherein the guide sequence is capable of hybridizing to a target nucleic acid in a eukaryotic cell.

12. The system of claim 1, wherein the Type VII CRISPR effector protein comprises one or more nuclear localization signals (NLSs).

13. The system of claim 1, wherein the Type VII CRISPR effector protein is a nickase.

14. The system of claim 1, wherein the Type VII CRISPR effector protein is catalytically inactive.

15. The system of claim 1, wherein the Type VII CRISPR effector protein is coupled to or associated with one or more functional domains.

16. The system of claim 15, wherein the functional domain has methylase activity, demethylase activity, translation activation activity, translation initiation activity, translation repression activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, nuclease activity, single-strand RNA cleavage activity, double-strand RNA cleavage activity, single-strand DNA cleavage activity, double-strand DNA cleavage activity, molecular switch activity, chemical inducibility, light inducibility, nucleic acid binding activity, deaminase activity, or a combination thereof.

17. The system of claim 15, wherein the functional domain cleaves the target nucleic acid.

18. The system of claim 15, wherein the functional domain modifies transcription or translation of the target nucleic acid.

19. The system of claim 15, wherein the functional domain comprises an adenosine deaminase, or catalytic domain thereof, or cytidine deaminase, or catalytic domain thereof.

20. The system of claim 19, wherein the adenosine deaminase or catalytic domain thereof, comprises one or more mutations that increase activity or specificity of the adenosine deaminase relative to wild type.

21. The system of claim 20, wherein the mutation confers an ability of the adenosine deaminase to deaminate cytidine.

22. A vector system comprising:

one or more polynucleotides each encoding one or more components of the system as set forth in any one of claims 1-21, wherein the one or more polynucleotides are each optionally included in one or more vectors.

23. The vector system of claim 22, further comprising one or more regulatory elements, wherein each regulatory element is operably linked to one or more of the one or more polynucleotides of the vector system.

24. The vector system of claim 23, wherein a first regulatory element is operably linked to a polynucleotide encoding a Type VII CRISPR effector protein, and optionally comprises a second regulatory element operably linked to a polynucleotide encoding the crRNA, when optionally present.

25. The vector system of claim 23, wherein a first regulatory element is operably linked to a polynucleotide encoding a Type VII CRISPR effector protein, a second regulatory element operably linked to a polynucleotide encoding the crRNA when optionally present, and a third regulatory element operably linked to a polynucleotide encoding the tracr RNA when optionally present, or a second regulatory element operably linked to a polynucleotide encoding the crRNA and the tracr RNA when both are optionally present.

26. The vector system of claim 22, wherein the polynucleotide encoding the Type VII CRISPR effector protein is a codon optimized for expression in a eukaryotic cell.

27. The vector system of claim 22, wherein the polynucleotides encoding the Type VII CRISPR effector protein, the optional crRNA, and optional tracrRNA are comprised in a single vector.

28. The vector system of claim 23, wherein the one or more vectors comprise viral vectors.

29. The vector system of claim 23, wherein the one or more vectors comprise one or more retroviral, lentiviral, adenoviral, adeno-associated or herpes simplex viral vectors.

30. A delivery system configured to deliver a one or more components of the engineered system of any of claims 1-21 or the vector system thereof, comprising:

i) a type VII CRISPR effector protein, and optionally

ii) a crRNA comprising a direct repeat (DR) sequence and a guide sequence that is designed to hybridize to a target nucleic acid sequence, and

whereby there is formed a CRISPR complex comprising the Type VII CRISPR effector protein complexed with the crRNA.

31. The delivery system of claim 30, further comprising one or more vectors or one or more polynucleotide molecules, the one or more vectors or polynucleotide molecules comprising one or more polynucleotide molecules encoding the Type VII CRISPR effector protein and one or more optional nucleic acid components of the non-naturally occurring or engineered Type VII CRISPR system.

32. The delivery system of claim 31, further comprising a delivery vehicle comprising liposome(s), particle(s), exosome(s), microvesicle(s), a gene-gun or one or more viral vector(s).

33. A method of modifying a target nucleic acid, the method comprising contacting the target nucleic acid with an engineered system as in any one of claims 1-21,

wherein the guide sequence directs sequence-specific binding to the target nucleic acid sequence, whereby the target nucleic acid sequence, the expression of the target nucleic acid, or both is/are modified.

34. The method of claim 33, wherein modifying occurs *in vitro*, *ex vivo*, or *in vivo*.

35. The method of claim 33, wherein modifying the target nucleic acid comprises cleaving the target nucleic acid.

36. The method of claim 33, wherein modifying expression of the target nucleic acid comprises increasing or decreasing transcription or translation of the target nucleic acid.

37. The method of claim 33, wherein the target nucleic acid is in a prokaryotic cell.

38. The method of claim 33, wherein the target nucleic acid is in a eukaryotic cell.

39. A cell comprising a modified target nucleic acid of interest, wherein the target nucleic acid of interest has been modified according to the method of claim 33.

40. The cell according to claim 39, wherein the modification of the target nucleic acid of interest results in:

the cell comprising altered expression of at least one gene product;

the cell comprising altered expression of at least one gene product, wherein the expression of the at least one gene product is increased;

the cell comprising altered expression of at least one gene product, wherein the expression of the at least one gene product is decreased; or

the cell comprising altered expression of at least one gene product, wherein the expressed gene produce is altered.

41. A cell comprising: an engineered system as in any one of claims 1-21 or a vector system thereof.

42. The cell of claim 41, wherein the cell is a prokaryotic cell.

43. The cell of claim 41, wherein the cell is a eukaryotic cell.

44. The eukaryotic cell of claim 38 or 43, wherein the cell is a mammalian cell or a human cell.

45. A cell line of or comprising the cell according to claim 41or progeny thereof.

46. The cell line of claim 45, wherein the cell line comprises one or more polynucleotide modifications.

47. A multicellular organism comprising one or more cells according to claim 41.

48. A plant or animal model comprising one or more cells according to claim 41.

49. A gene product from a cell of claim 41, the cell line of claim 45, the organism of claim 47 or the plant or animal model of claim 48.

50. The gene product of aspect 49, wherein the amount of gene product expressed is greater than or less than the amount of gene product from a cell that does not have altered expression.

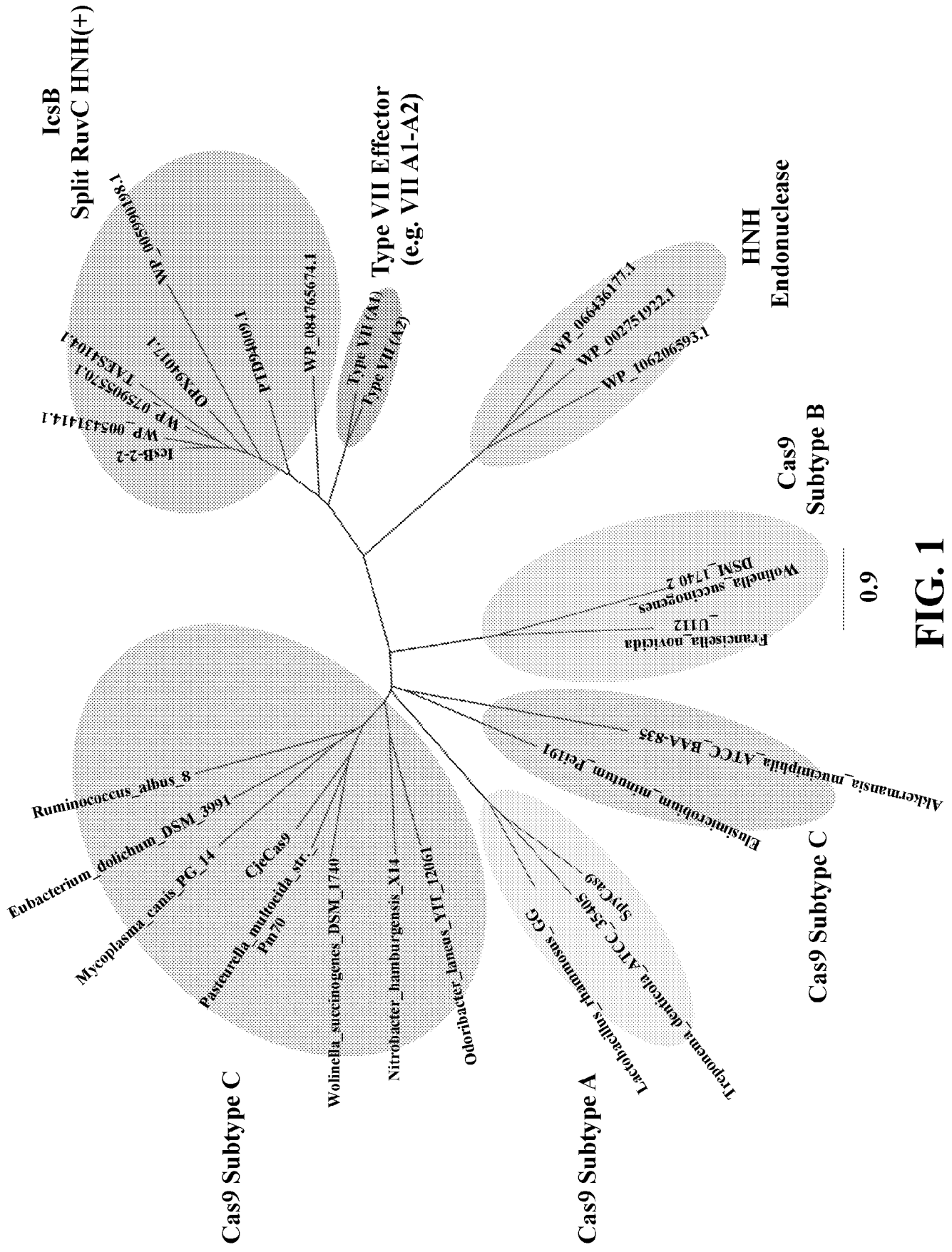

IcsB
Split RuvC HNH(+)
Type VII Effector
(e.g. VII A1-A2)
HNH
Endonuclease
Cas9
Subtype B
Cas9 Subtype C
Cas9 Subtype A
Cas9 Subtype C
IcsB-2-2
WP_005313141.1
WP_075905570.1
OPX94017.1
WP_084765674.1
Type VII (A1)
Type VII (A2)
WP_066436177.1
WP_002751922.1
WP_106206593.1
Francisella_noviciida_U112
Wolinella_succinogenes_DSM_1740 2
0.9
Akkermansia_muciniphila_ATCC_BAA-835
Elusimicrobium_minutum_Pei191
Treponema_denticola_ATCC_35405
Lactobacillus_thamnosus_GG
Odoribacter_laneus_YIT_12061
Nitrobacter_hamburgensis_X14
Wolinella_succinogenes_DSM_1740
Pasteurella_multocida_str._Pm70
Mycoplasma_canis_PG_14
Eubacterium_dolichum_DSM_3991
Ruminococcus_albus_8
CjeCas9
SpyCas9

FIG. 1

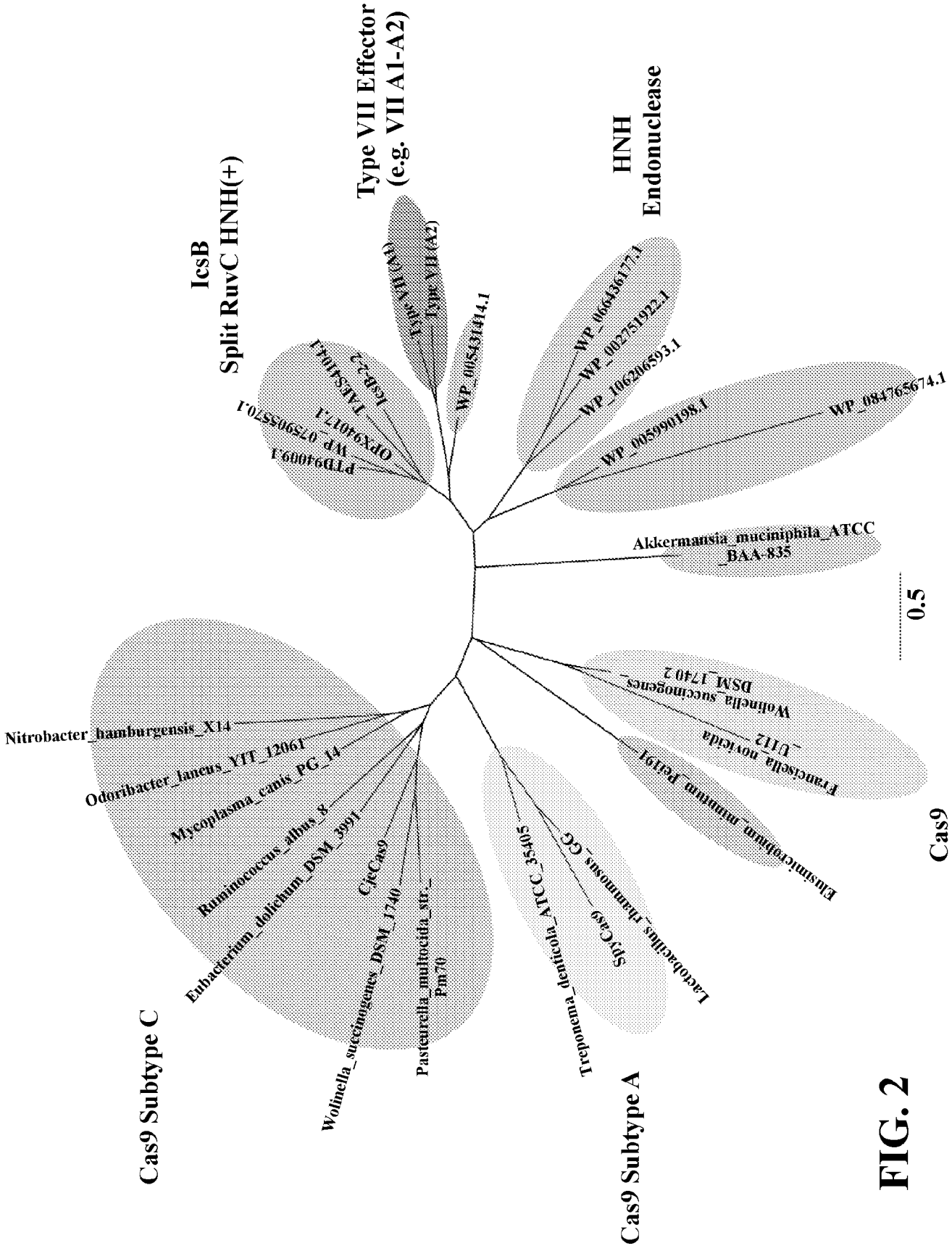

Type VII Effector
(e.g. VII A1-A2)
HNH
Endonuclease
IcsB
Split RuvC HNH(+)
Type VII(A2)
Type VII(A1)
WP_005431414.1
WP_066436177.1
WP_002751922.1
WP_106206593.1
WP_005990198.1
WP_084765674.1
PTD94009.1
WP_075905870.1
OPX94017.1
TADS41061
IcsB-2.2
Akkermansia_muciniphila_ATCC_BAA-835
0.5
Nitrobacter_hamburgensis_X14
Odoribacter_laneus_YIT_12061
Mycoplasma_canis_PG_14
Ruminococcus_albus_8
Eubacterium_dolichum_DSM_3991
CjeCas9
Wolinella_succinogenes_DSM_1740
Pasteurella_multocida_str._Pm70
Treponema_denticola_ATCC_35405
SpyCas9
Lactobacillus_thamnosus_GG
Elusimicrobium_minutum_Pei191
Francisella_novicida_U112
Wolinella_succinogenes_DSM_17402
Cas9 Subtype C
Cas9 Subtype A
Cas9
Subtype B

FIG. 2

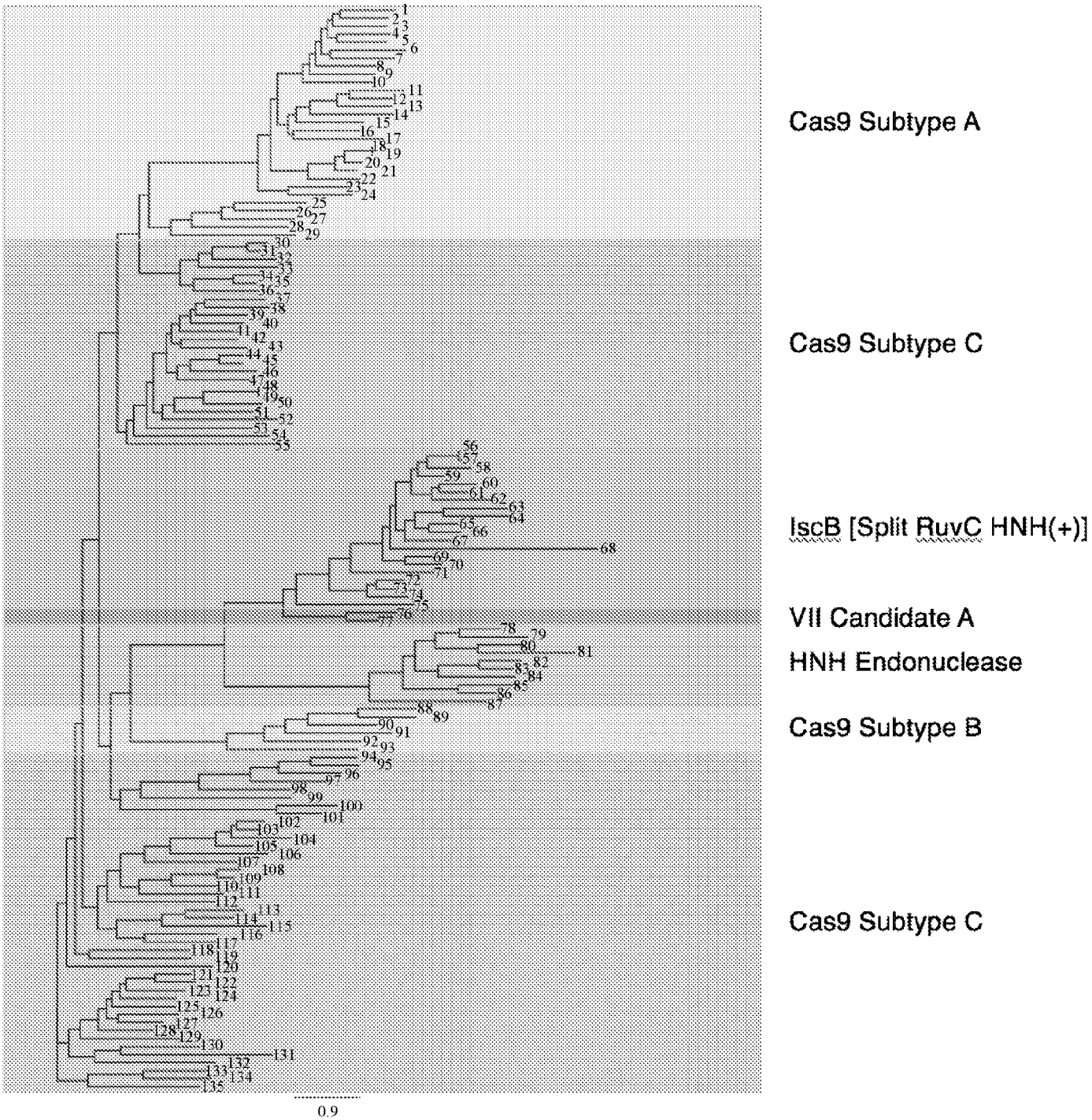

Cas9 Subtype A
Cas9 Subtype C
IscB [Split RuvC HNH(+)]
VII Candidate A
HNH Endonuclease
Cas9 Subtype B
Cas9 Subtype C
0.9

FIG. 3

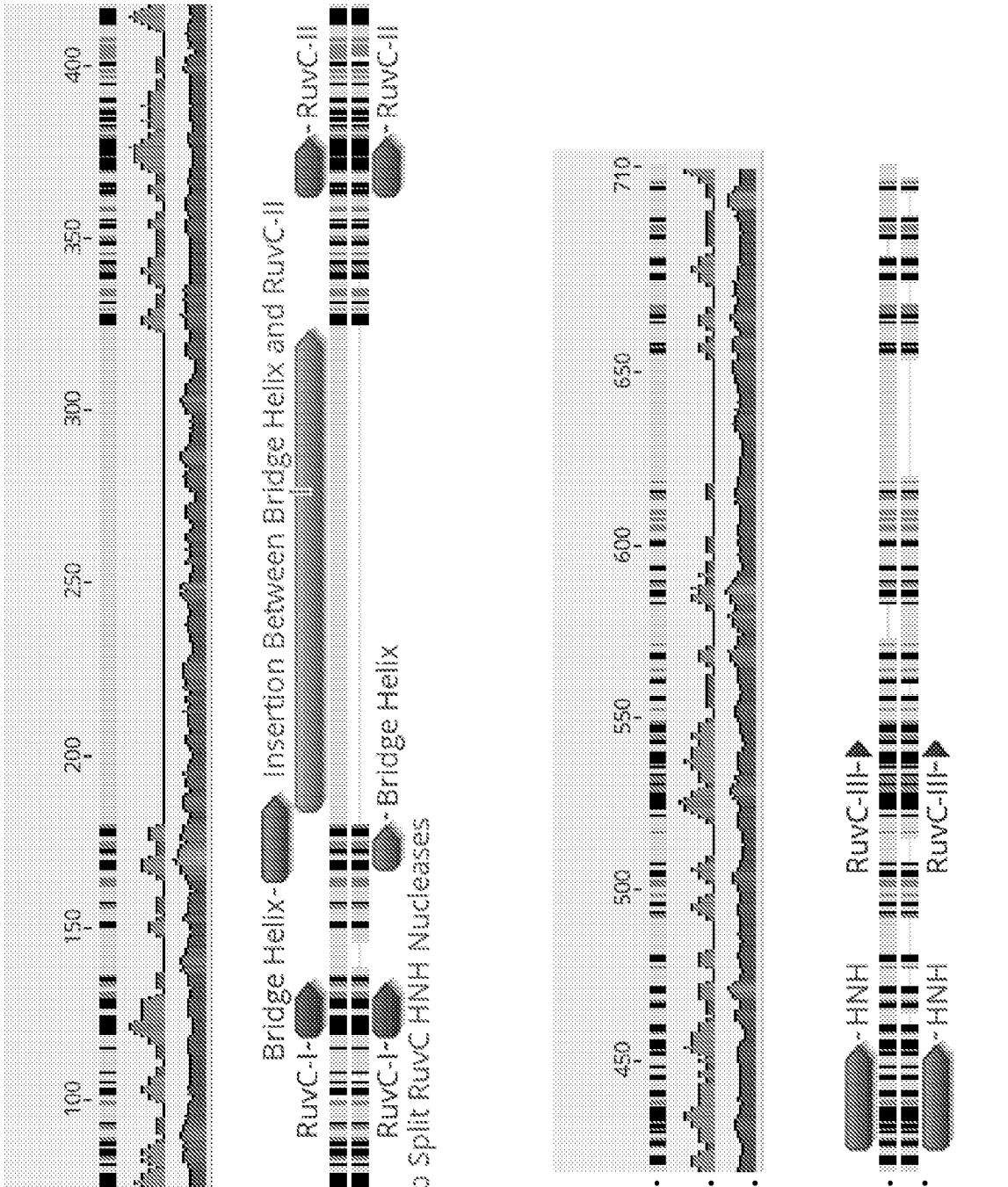

1
50
100
150
200
250
300
350
400
Bridge Helix
Insertion Between Bridge Helix and RuvC-II
RuvC-II
RuvC-I
RuvC-II
Bridge Helix
RuvC-I
RuvC-II
Deletion Specific To Split RuvC HNH Nucleases
1.
2.
3.
4.
5.
450
500
550
600
650
710
HNH
RuvC-III
HNH
RuvC-III
1. Consensus
2. Identity
3. Mean PI
4. VII Candidate A1
5. Split RuvC HNH Nuclease

FIG. 4

```
VII_cA1      --------MPCHPARARELVRKGRAYWVKRDTVKLKTFLQNAVCQHTVLGIDAGAIHIGL    52
VII_cA2      MNQDNKPLMPCHPARARELVRRGRAYWVRRDTIKLSEYKENAVLQRTFLGIDAGAEHIGL    60
                      **************:******:***:**. : :*** *:*.******* ****

VII_cA1      AVVAKREKRKPKVLFRAQVDARPMSEIKSVMDKRRSLRRNRRSRKGYRQPWYSPTRRVGL    112
VII_cA2      AVVAEREKHKPRVIFRAQVEARPMREIKECLDQRRSYRRGRRSRKAYRRPWFSPTRRVSL    120
             ****:***:**:*:*****:**** ***. :*:*** **.*****.**:**:******.*

VII_cA1      TAKLFKKSGEFLCKIPLAEGKRLLKPDKGGNRQAKRLGGNKYQLSTMKEIKAPKREKEPG    172
VII_cA2      SARLYNKQGEFLCRIPLAYGKYLTKPDKSGRPQAKRLYGNNYQLLTSKNHESPIREKKPG    180
             :*:*::*.*****:**** ** * ****.*. ***** **:*** * *: ::* ***:**

VII_cA1      KNDIVVYRFDGRPLTWFSPKSRKFVKQLCKCLRGYGKVRRLPVADVKKEGDKVKSIWLKS    232
VII_cA2      KQDILVYRNDGRLLTWYDPKRGDLE-GIEKRLLGTSRRHNIPIASLEKEGSKK-IIRLLS    238
             *:**:*** *** ***:.**  .:   : * * * .: :.:*:*.::***.*   * * *

VII_cA1      RRAERLIPKNILPEVVPPFRKMGQKRKPSGWLAPSANALVNAYFNGIKRVGSYVPIASVR    292
VII_cA2      VSAEDAIPKAALPNPMPRFRKMGQKRKPPGWLSPSANSLLNAYYRGIKEIGSYVPISLVR    298
               **  ***  **: :* ********** ***:****:*:***:.***.:******: **

VII_cA1      FEDVAFDTRKLED-CLDGKGDCSTVNYARQNRKDFVLARDGWKCVYCGRKSNFSKKGLRL    351
VII_cA2      LEVAAFDTQKIEDPTIHGVGYQKPALYAKQNRKDFVIARDGWRCVYCGVN------NVPL    352
             :* .****:*:**  :.* *  . . **:*******:*****:***** :      .: *

VII_cA1      TQDHVVPKSKNGSNAIGNLVACCEKCNSDKGSLDLDDYLKKREQQGDKPSDTKRIKRYVV    411
VII_cA2      TQDHIQPQSKGGPDRIGNLVACCFKCQNEKSNLTLDEFLKTKS-----PQQAKRIKEYVS    407
             ****: *:**.* : ******** **:.:*..* **::**.:.     *.::****.**

VII_cA1      GLSQK------SESFRSKAHVGQVKTKVLRLLRERFEDVS--ITKGYVTKKDRDWMGLTK    463
VII_cA2      SITQKNESLQKNESLRSAAHVGQIKSKLLQGLIDLFGDSNVKRTWGYITKRDRLALHLPK    467
             .::**      .**:** *****:*:*:*: * : * * .   * **:**:**  : * *

VII_cA1      SHDCDAVAIAAYGERVG--FMPPLRTYHVRRYTWEGSRRRQYSANPMKKYSSK----CVL    517
VII_cA2      SHDYDAIAIASLGRRIDTENIPPLRLYYMRRYTKDSPRRQKYKANPLSNAKKDNKNLAPL    527
             *** **:***: *.*:.   :**** *::**** :. **::*.***:.: ...    . *

VII_cA1      IAKGTPRERWVQPNSYNPYADTKVGRVYSRDVVLAVGGNKGTEPVLGYIRGICKDG----    573
VII_cA2      FIGQGSRSRWVQQVQVNSYAETPVGIVRLRDLIKTIDGRIG------YVRKINSGRKPKP    581
             :     *.****  . * **:* ** *  **:: ::.*. *      *:* * ..

VII_cA1      ------ALFISRLPTGGKKKERSKRQAKYVKKIISRCRGLMEVL--                  611
VII_cA2      GQKPVRILTLSHKPQGKE--GQFSTSANKVEKIIRRRCGLLEMLYV                  625
```

VII cA1: SEQ ID NO: 1
VII cA2: SEQ ID NO: 2

FIG. 5

```
VII-cB1      --------------------MAQLARQLREEAARWLGVFGDDANVRQKIGTPTGYQTSV    39
VII-cB2      MGRGFRASIKSKNHHNLSYGETNGLARQLREEAAHWLGIFGDDAKVRQRIGTPTGYQTSV    60
                                   **********:***:*****:***:***********

VII-cB1      CRESWGENL-----PKKDRSNLRHHLWDAAVISHIPPGKGMNLTNCAGIFLHTGTNQHGN    94
VII-cB2      CRGSWSINLPADFYPKKNRNNLRHHLWDAAILANIPPGKGLNLTNCGGIFFHAGRNKHGN    120
             ** **. **     ***:*.**********::::******:*****.***:*:* *:***

VII-cB1      ITMRAIPGLGPDLNSFENGTKDQCLVDERRATKSKKQRFDQTIYSPPDTEGVMWARDPLD    154
VII-cB2      IAMQALPGLGPDLNAFEKATSNQCLVEKKQTTNSKKSRFQQWIYGFPDKNGEMWARDAIE    180
             *:*:*:********:**:.*.:****:::::*:***.**:* **. **.:* ***** ::

VII-cB1      KLATKKNQSAEGLLQLLRDAGMDENQLPSSRFEEWWNGVQAQWFTRNDAQALFQTLEIPS    214
VII-cB2      TLADKP-----NLLSRLRDAGIDEAQLPTRHFEVWQERRQAQYFTREEAWVAVQALALPA    235
             .** *      .**. *****:** ***: :** * :  ***:***::* . .*:* :*:

VII-cB1      PKIISEDALNKWWGNY---EKKKINYKSLRLLLKESGITPVEVSDLRLQNALSGHSEPGS    271
VII-cB2      EHAIPVSTFDEWWNDRLKSDKKKVTDKSLRALLAKARIPKAAVTDVKLTGVLVNRGDPGP    295
              : *  .::::**.:    :***:. **** ** :: *  . *:*::* ..* .:.:**

VII-cB1      LTRKDGTPIRGVRGTAKQMTPMAVIPHRNHKGETIGFKFAKESFVRAEIWVTEKTTKDG-    330
VII-cB2      LTRKDGTPIYGVSGTAKQMTPMAVVPHRNHKGETIGFKLATESFIRAEIWTTEKRDKKGD    355
             ********* ** **********:*************:*.***:*****.***  *.*

VII-cB1      -------KVELDYHCRLIPHPRGLANLQLRVMKCTGQPLTWERRLTDDELVELGLAKKLA    383
VII-cB2      IVKDDTGKPVLEYHRRLIPHPRGLKNLRLRTMRCTGKRLSWERRLTDEEIVELGLKEHAE    415
                    *  *:** ********* **:**.*:***: *:*******:*:***** ::

VII-cB1      TLKNERERLKNSHAKAVEKWQKSQQKKKDSDTLGLAADAATSSPKEIPELWLSLRKIYTG    443
VII-cB2      VK-----RLRKDYDKAVSQHVKHTRAQAPECELALCQ-SKVTMPGQPKPPVISLRKIYTG    469
             .      **::.: ***.:  *  : :  .  *.*.  : .: * :     :********

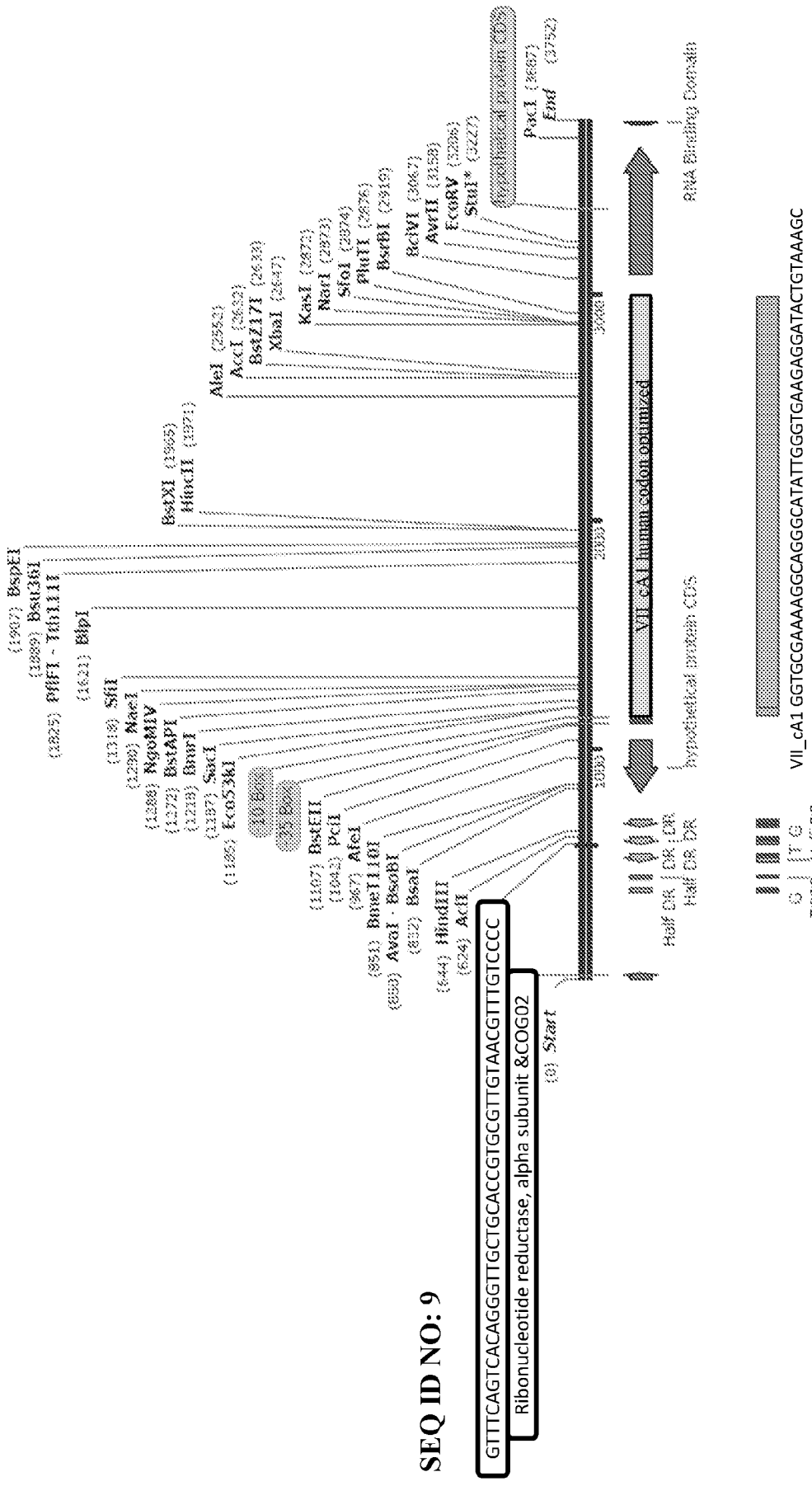
VII-cB1      LPPHAKRLTNANGADISRVTKGDLLFVPLRKVPGAGKGHGKFCKRGEVATGGSCWFRVSA    503
VII-cB2      LPPNAKRLTNAQGADISCFAKGDLLLVPVSQS-------GAICESD-KAPYRELWYRITA    521
             ***:*******:***** .:*****:**: :        * :*: .  *   . *:*::*

VII-cB1      IKAAGEIELHLAEFKLPK------IKDDKKATEQQEWLTTIWEQRPSSEDDIAWLLEQSR    557
VII-cB2      LKANGQVELKLAEFKEIKRPSEKELKEGKSLTPEQEWFADAFLQQPGNDELIAGLLQRTR    581
             :** *::**:*****  *      :*:.*. * :***::  : *:*..:: ** **:::*

VII-cB1      AHDQSPHPVK                                                      567
VII-cB2      VNDQSSHSVK                                                      591
```

VII cB1: SEQ ID NO: 3
VII cB2: SEQ ID NO: 4

FIG. 6



SEQ ID NO: 9

GTTTCAGTCACAGGGTTGCTGCACCGTGCGTTGTAACGTTTGTCCCC

Ribonucleotide reductase, alpha subunit &COG02

SEQ ID NO: 10

VII_cA1 GGTGCGAAAAGGCAGGGCATATTGGGTGAAGAGGATACTGTAAAGC

FIG. 7

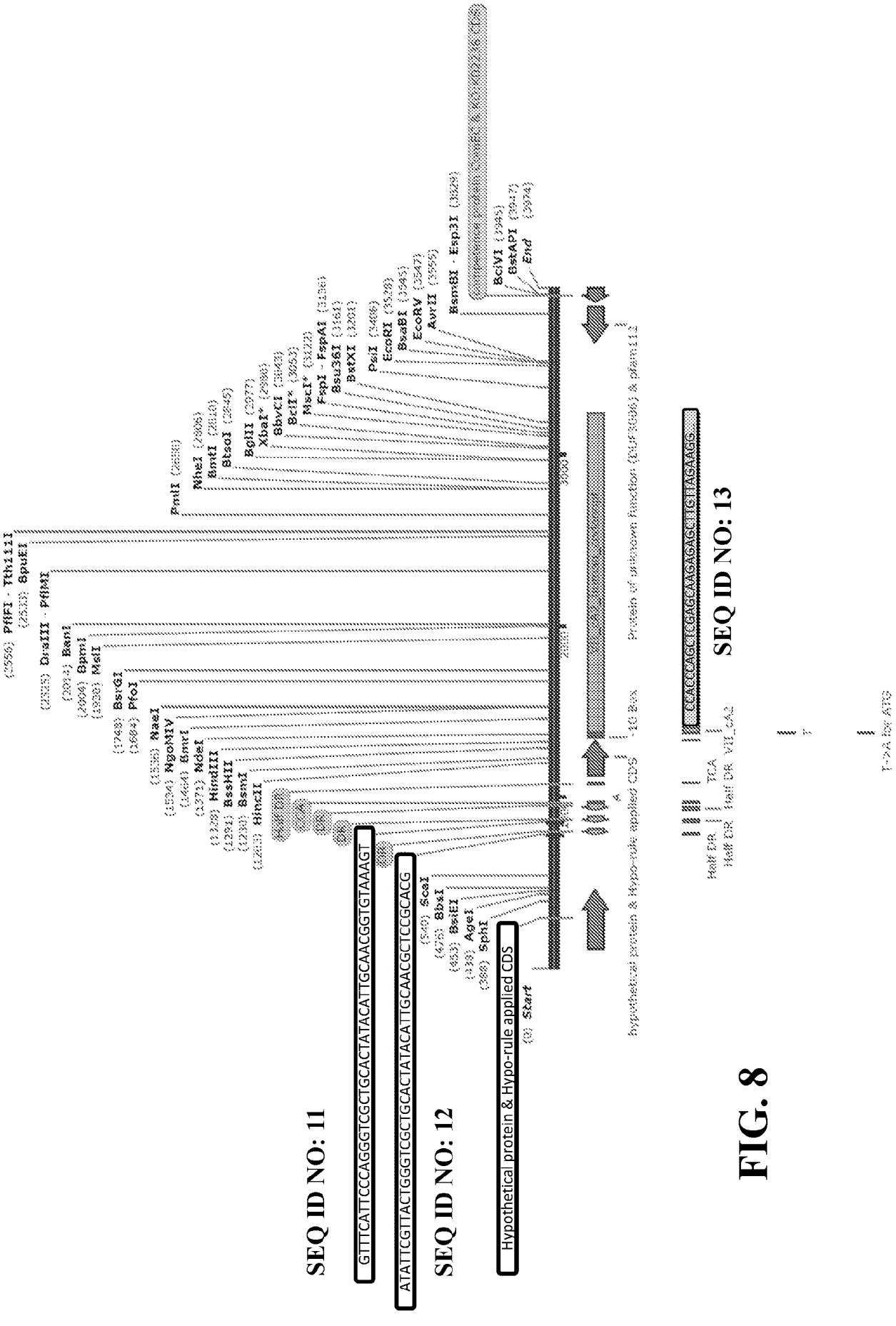

SEQ ID NO: 11
GTTTCATTCCCAGGGTCGCTGCACTATACATTGCAACGGTGTAAAGT
SEQ ID NO: 12
ATATTCGTTACTGGGTCGCTGCACTATACATTGCAACGCTCCGCACG
Hypothetical protein & Hypo-rule applied CDS
(0) Start
ScaI
BbsI
BsiEI
AgeI
SphI
hypothetical protein & Hypo-rule applied CDS
10 Bps
Protein of unknown function (DUF3036) & pfam10
Half DR
TCA
Half DR
VIC_CAS
CCACCCAGCTCGAGCAAGAAGAGCTTGTTAGAAGG
SEQ ID NO: 13
T-DNA for AMS

FIG. 8

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2020/023441

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

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

B. FIELDS SEARCHED

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

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

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Y ESTHER TAK ET AL: "Inducible and multiplex gene regulation using CRISPR-Cpf1-based transcription factors", NATURE METHODS, vol. 14, no. 12, 30 October 2017 (2017-10-30), pages 1163-1166, XP055577293, New York ISSN: 1548-7091, DOI: 10.1038/nmeth.4483 Abstract; page 1163, column 2, paragraph 1 ----- -/--	49,50

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

[X] See patent family annex.

* Special categories of cited documents :

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
3 June 2020	12/06/2020

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

1

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

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	YUYI TANG ET AL: "Class 2 CRISPR/Cas: an expanding biotechnology toolbox for and beyond genome editing", CELL & BIOSCIENCE, vol. 8, no. 1, 12 November 2018 (2018-11-12), XP055700477, DOI: 10.1186/s13578-018-0255-x the whole document	1-50
A	YOSHIZUMI ISHINO ET AL: "History of CRISPR-Cas from encouter with a mysterious repeated sequence to genome editing technology", JOURNAL OF BACTERIOLOGY, vol. 200, no. 7, 22 January 2018 (2018-01-22), XP055700490, ISSN: 0021-9193, DOI: 10.1128/JB.00580-17 the whole document	1-50
X,P	WO 2020/051460 A1 (SANFORD BURNHAM PREBYS MEDICAL DISCOVERY INST [US]) 12 March 2020 (2020-03-12) paragraphs [0086], [0087], [0099]; claims 1-39	1,7,8, 11,15, 16,18, 22,30, 33-35, 38-41, 43,44, 46,47, 49,50

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2020/023441

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2020051460 A1	12-03-2020	NONE	