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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
22 November 2018 (22.11.2018)

W!P O ¦ PCT

(10) International Publication Number
WO 2018/213726 Al

(51) International Patent Classification:
C12N 15/10 (2006.01) *C12N 15/62* (2006.01)

(21) International Application Number:
PCT/US20 18/033427

(22) International Filing Date:
18 May 2018 (18.05.2018)

(25) Filing Language: English

(26) Publication Langiiage: English

(30) Priority Data:
62/508,300 18 May 2017 (18.05.2017) 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) Inventor: **ZHANG, Feng;** 100 Pacific Street, Apt. 11, Cambridge, Massachusetts 02139 (US).

(74) Agent: **NIX, F. Brent;** Johnson, Marcou & Isaacs, LLC, P.O. Box 691, Hoschton, GA 30548 (US).

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

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

Declarations under Rule 4.17:

— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.1 7(H))*

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

WO 2018/213726 A1

(54) Title: SYSTEMS, METHODS, AND COMPOSITIONS FOR TARGETED NUCLEIC ACID EDITING

(57) Abstract: The invention provides for systems, methods, and compositions for targeting and editing nucleic acids. In particular, the invention provides non-naturally occurring or engineered DNA-targeting systems comprising a DNA-targeting Cpfl protein, at least one guide molecule, and at least one cytidine deaminase protein or catalytic domain thereof.

SYSTEMS, METHODS, AND COMPOSITIONS FOR TARGETED NUCLEIC ACID EDITING

RELATED APPLICATIONS AND INCORPORATION BY REFERENCE

[0001] This application claims priority to U.S. Provisional Application No. 62/508,300 filed May 18, 2017, which is incorporated herein by reference in its entirety

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under grant numbers MH100706 and MHl 10049 awarded by the National Institutes of Health. The government has certain rights in the invention.

FIELD OF THE INVENTION

[0003] The present invention generally relates to systems, methods, and compositions for targeting and editing nucleic acids, in particular for programmable deamination of cytosine at a target locus of interest.

BACKGROUND

[0004] Recent advances in genome sequencing techniques and analysis methods have significantly accelerated the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. Precise genome targeting technologies are needed to enable systematic reverse engineering of causal genetic variations by allowing selective perturbation of individual genetic elements, as well as to advance synthetic biology, biotechnological, and medical applications. Although genome-editing techniques such as designer zinc fingers, transcription activator-like effectors (TALEs), or homing meganucleases are available for producing targeted genome perturbations, there remains a need for new genome 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.

[0005] Point mutations A→G and T→C represent about 12% of known pathogenic SNPs. Programmable deamination of cytosine has been reported and may be used for correction of A→G and T→C point mutations. For example, Komor et al., Nature (2016)

533:420-424 reports targeted deamination of cytosine by APOBEC1 cytidine deaminase in a non-targeted DNA stranded displaced by the binding of a Cas9-guide RNA complex to a targeted DNA strand, which results in conversion of cytosine to uracil. See also Kim et al., Nature Biotechnology (2017) 35:371-376; Shimatani et al., Nature Biotechnology (2017) doi:10.1038/nbt.3833; Zong et al., Nature Biotechnology (2017) doi:10.1038/nbt.3811; Yang Nature Communication (2016) doi:10.1038/ncommsl3330.

[0006] Novel systems and methods which allow specific correction of these point mutations and pathogenic SNPs are of interest.

SUMMARY OF THE INVENTION

[0007] At least a first aspect of the invention relates to a method of modifying an Cytosine in a target locus of interest, the method comprising delivering to the locus: (a) a Cpfl nickase protein; (b) a guide molecule which comprises a guide sequence linked to a direct repeat sequence; and (c) a cytidine deaminase protein or catalytic domain thereof; wherein the cytidine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to the Cpfl nickase protein or the guide molecule or is adapted to link thereto after delivery; wherein guide molecule forms a complex with the Cpfl nickase and directs the complex to bind a first DNA strand at the target locus of interest, wherein the guide sequence is capable of hybridizing with a target sequence comprising said Cytosine within said first DNA strand to form a heteroduplex that comprises a non-pairing Adenine or Uracil opposite to said Cytosine; wherein the Cpfl nickase protein nicks a second DNA strand at the target locus of interest displaced by the formation of the heteroduplex; wherein the cytidine deaminase protein or catalytic domain thereof deaminates the Cytosine in the heteroduplex opposite to the non-pairing Adenine or Uracil.

[0008] In some embodiments, the cytidine deaminase protein or catalytic domain thereof is fused to N- or C-terminus of the Cpfl nickase protein. In some embodiments, the cytidine protein or catalytic domain thereof is fused to the Cpfl nickase protein by a linker. In some embodiments, the linker is $(GGGGS)_{3}$-n (SEQ ID NOS: 1-9), GSG_{5} (SEQ ID NO: 10) or LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO: 11).

[0009] In some embodiments, the cytidine deaminase protein or catalytic domain thereof is linked to an adaptor protein and the guide molecule or the Cpfl nickase protein comprises an aptamer sequence capable of binding to the adaptor protein. In some embodiments, the adaptor sequence is selected from MS2, PP7, QP, F2, GA, fr, JP501,

M12, R17, BZ13, JP34, JP500, KU1, Mil, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕC¾5, ϕC¾8r, ϕC¾12r, (^Cb23r, 7s and PRR1.

[0010] In some embodiments, the cytidine deaminase protein or catalytic domain thereof is inserted into an internal loop of the Cpfl nickase protein.

[0011] In some embodiments, said Cpfl nickase comprises a mutation in the Nuc domain. In some embodiments, the Cpfl nickase comprises a mutation corresponding to R1226A in AsCpfl.

[0012] In some embodiments, said Cpfl nickase has part or all of the Nuc domain removed.

[0013] In some embodiments, the guide sequence of the guide molecule comprises at least one further mismatches or non-pairing nucleotide with the target sequence, adjacent to said non-pairing Adenine or Uracil. In some embodiments, the guide sequence comprises a stretch of three to five consecutive mismatches with the target sequence. In some embodiments the guide sequence comprises less than 6 consecutive mismatches with the target sequence.

[0014] In some embodiments, the guide molecule binds to Cpfl and is capable of forming a heteroduplex of about 24 nt with the target sequence. In some embodiments, the guide molecule binds to Cpfl and is capable of forming a heteroduplex of more than 24 nt with the target sequence.

[0015] In some embodiments, the guide sequence comprises at least one further non-pairing nucleotide with said target sequence, adjacent to said non-pairing Adenine or Uracil. In some embodiments, the guide sequence comprises a stretch of three to five consecutive non-pairing nucleotides with said target sequence.

[0016] In some embodiments, the cytidine deaminase is a human, rat or lamprey cytidine deaminase. In some embodiments, the cytidine deaminase is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase, an activation-induced deaminase (AID), or a cytidine deaminase 1 (CDA1).

[0017] In some embodiments, the cytidine deaminase is an APOBECl deaminase comprising one or more mutations corresponding to W90A, W90Y, R 118A, H121R, H122R, R126A, R126E, or R132E in rat APOBECl, or an APOBEC3G deaminase comprising one or more mutations corresponding to W285A, W285Y, R313A, D316R, D317R, R320A, R320E, or R326E in human APOBEC3G.

[0018] In some embodiments, the Cpfl nickase protein is derived from a bacterial species selected from the group consisting of Francisella tularensis, Prevotella albensis, Lachnospiraceae bacterium, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium, Parcubacteria bacterium, Smithella sp., Acidaminococcus sp., Lachnospiraceae bacterium, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi, Leptospira inadai, Porphyromonas crevioricanis, Prevotella disiens and Porphyromonas macacae, Succinivibrio dextrinosolvens, Prevotella disiens, Flavobacterium branchiophilum, Helcococcus kunzii, Eubacterium sp., Microgenomates (Roizmanbacteria) bacterium, Flavobacterium sp., Prevotella brevis, Moraxella caprae, Bacteroidetes oral, Porphyromonas cansulci, Synergistes jonesii, Prevotella bryantii, Anaerovibrio sp., Butyrivibrio fibrisolvens, Candidatus Methanomethylophilus, Butyrivibrio sp., Oribacterium sp., Pseudobutyrivibrio ruminis and Proteocatella sphenisci.

[0019] In some embodiments, the natural PAM sequence is TTN, where N is A/C/G or T and the CRISPR-Cas protein is FnCpfl or wherein the PAM sequence is TTTV, where V is A/C or G and the CRISPR-Cas protein is PaCpflp, LbCpfl or AsCpfl. In some embodiments the Cpfl nickase protein has been modified to alter PAM-recognition.

[0020] In some embodiments, the target locus of interest is within a cell. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the cell is a non-human animal cell. In some embodiments, the cell is a human cell. In some embodiments, the cell is a plant cell.

[0021] In some embodiments, the target locus of interest is within an animal. In some embodiments, the target locus of interest is within a plant. In some embodiments, the target locus of interest is comprised in a DNA molecule *in vitro*.

[0022] In some embodiments, the components (a), (b) and (c) are delivered to the cell as a ribonucleoprotein complex.

[0023] In some embodiments, the components (a), (b) and (c) are delivered to the cell as one or more polynucleotide molecules. In some embodiments, the one or more polynucleotide molecules comprise one or more mRNA molecules encoding components (a) and/or (c).

[0024] In some embodiments, the one or more polynucleotide molecules are comprised within one or more vectors. In some embodiments, the one or more polynucleotide molecules comprise one or more regulatory elements operably configured to express the Cpfl nickase protein, the guide molecule, and the cytidine deaminase protein or catalytic

domain thereof, optionally wherein the one or more regulatory elements comprise inducible promoters.

[0025] In some embodiments, the Cpfl nickase protein and optionally the cytidine deaminase protein or catalytic domain thereof comprise one or more heterologous nuclear localization signal(s) (NLS(s)).

[0026] In some embodiments, the cytidine deaminase is delivered together with a uracil glycosylase inhibitor (UGI), where the UGI is covalently linked to said cytidine deaminase and/or said Cpfl nickase.

[0027] In some embodiments, the one or more polynucleotide molecules or the ribonucleoprotein complex are delivered via particles, vesicles, or one or more viral vectors.

[0028] In some embodiments, the particles comprise a lipid, a sugar, a metal or a protein. In some embodiments, the particles comprise lipid nanoparticles.

[0029] In some embodiments, the vesicles comprise exosomes or liposomes. In some embodiments, the one or more viral vectors comprise one or more of adenovirus, one or more lentivirus or one or more adeno-associated virus.

[0030] In some embodiments, the method modifies a cell, a cell line or an organism by manipulation of one or more target sequences at genomic loci of interest.

[0031] At least a second aspect of the invention relates to a method for treating or preventing a disease using the method described herein, wherein the deamination of the Cytosine at the target locus of interest remedies a disease caused by a T→C or A→G point mutation or pathogenic SNP.

[0032] At least a third aspect of the invention relates to a method for knock-out or knock-down an undesirable activity of a gene or regulatory element thereof, wherein the deamination of the Cytosine at the target locus of interest inactivates a target gene or a target regulatory element at the target locus.

[0033] At least a fourth aspect of the invention relates to a modified cell obtained from the method described above, or progeny thereof, wherein the cell comprises an Uracil or a Thymine in replace of the Cytosine in the target locus of interest compared to a corresponding cell not subjected to the method.

[0034] In some embodiments, the modified cell is a eukaryotic cell. In some embodiments, the modified cell is an animal cell. In some embodiments, the modified cell is a human cell. In some embodiments, the modified cell is a plant cell.

[0035] In some embodiments, the modified cell is a therapeutic T cell. In some embodiments, the modified cell is an antibody-producing B cell.

[0036] At least a fifth aspect of the invention relates to a non-human animal or a plant comprising the modified cell described herein.

[0037] At least a sixth aspect of the invention relates to a method for cell therapy, comprising administering to a patient in need thereof the modified cell described herein, wherein the presence of the modified cell remedies a disease in the patient.

[0038] At least a seventh aspect of the invention relates to an engineered, non-naturally occurring system suitable for modifying an Cytosine in a target locus of interest, comprising: a guide molecule which comprises a guide sequence linked to a direct repeat, or a nucleotide sequence encoding the guide molecule; a Cpfl nickase protein, or one or more nucleotide sequences encoding the Cpfl nickase protein; a cytidine deaminase protein or catalytic domain thereof, or one or more nucleotide sequences encoding; wherein the cytidine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to the Cpfl nickase protein or the guide molecule or is adapted to link thereto after delivery; wherein the guide sequence is capable of hybridizing with a target sequence on a first DNA strand comprising a Cytosine on a first strand at the target locus, but comprises an Adenine or Uracil at the position corresponding to said Cytosine; and wherein the Cpfl nickase protein is capable of nicking a non-target sequence in a second DNA strand, complementary to said first DNA strand. Accordingly, the application provides kits comprising or consisting of the components of the CD-functionalized CRISPR system described herein.

[0039] At least an eighth aspect of the invention relates to an engineered, non-naturally occurring vector system suitable for modifying an Cytosine in a target locus of interest, comprising one or more vectors comprising: a first regulatory element operably linked to one or more nucleotide sequences encoding a guide molecule which comprises a guide sequence linked to a direct repeat; a second regulatory element operably linked to a nucleotide sequence encoding a Cpfl nickase protein; and optionally a nucleotide sequence encoding a cytidine deaminase protein protein or catalytic domain thereof which is under control of the first or second regulatory element or operably linked to a third regulatory element; wherein, if the nucleotide sequence encoding a cytidine deaminase protein or catalytic domain thereof is operably linked to a third regulatory element, the cytidine deaminase protein or catalytic domain thereof is adapted to link to the guide molecule or the

Cpfl nickase protein after expression; wherein said guide sequence is capable of hybridizing with a target sequence comprising a Cytosine within said target locus, but comprises an Adenine or Uracil at the position corresponding to said Cytosine; wherein components (a), (b) and (c) are located on the same or different vectors of the system; and wherein the Cpfl nickase protein is capable of nicking a non-target sequence in a second DNA strand complementary to said first DNA strand. Accordingly, the application provides kits comprising or consisting of vectors encoding of the components of the CD-functionalized CRISPR system described herein.

[0040] At least an ninth aspect of the invention relates to *in vitro, ex vivo* or *in vivo* host cell or cell line or progeny thereof comprising the engineered, non-naturally occurring system or vector system described herein.

[0041] In some embodiments, the host cell is a eukaryotic cell. In some embodiments, the host cell is an animal cell. In some embodiments, the host cell is a human cell. In some embodiments, the host cell is a plant cell.

DETAILED DESCRIPTION

[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).

[0043] Methods for Targeted Deamination of Cytosine

[0044] In one aspect the present invention provides methods for targeted deamination of cytosine in a DNA, more particularly in a locus of interest. According to the methods of the invention, the cytidine deaminase (CD) protein is recruited specifically to the relevant Cytosine in the locus of interest by a CRISPR-Cas complex which can specifically bind to a target sequence. In order to achieve this, the cytidine deaminase protein can either be covalently linked to the CRISPR-Cas enzyme or be provided as a separate protein, but adapted so as to ensure recruitment thereof to the CRISPR-Cas complex.

[0045] In particular embodiments, of the methods of the present invention, recruitment of the cytidine deaminase to the target locus is ensured by fusing the cytidine deaminase or catalytic domain thereof to the CRISPR-Cas protein, which is a Cpfl protein. Methods of generating a fusion protein from two separate proteins are known in the art and typically

involve the use of spacers or linkers. The Cpfl protein can be fused to the cytidine deaminase protein or catalytic domain thereof on either the N- or C-terminal end thereof. In particular embodiments, the CRISPR-Cas protein is a Cpfl protein and is linked to the N-terminus of the deaminase protein or its catalytic domain.

[0046] The term "linker" as used in reference to a fusion protein refers to a molecule which joins the proteins to form a fusion protein. Generally, such molecules have no specific biological activity other than to join or to preserve some minimum distance or other spatial relationship between the proteins. However, in certain embodiments, the linker may be selected to influence some property of the linker and/or the fusion protein such as the folding, net charge, or hydrophobicity of the linker.

[0047] Suitable linkers for use in the methods of the present invention are well known to those of skill in the art and include, but are not limited to, straight or branched-chain carbon linkers, heterocyclic carbon linkers, or peptide linkers. However, as used herein the linker may also be a covalent bond (carbon-carbon bond or carbon-heteroatom bond). In particular embodiments, the linker is used to separate the CRISPR-Cas protein and the cytidine deaminase by a distance sufficient to ensure that each protein retains its required functional property. Preferred peptide linker sequences adopt a flexible extended conformation and do not exhibit a propensity for developing an ordered secondary structure. In certain embodiments, the linker can be a chemical moiety which can be monomeric, dimeric, multimeric or polymeric. Preferably, the linker comprises amino acids. Typical amino acids in flexible linkers include Gly, Asn and Ser. Accordingly, in particular embodiments, the linker comprises a combination of one or more of Gly, Asn and Ser amino acids. Other near neutral amino acids, such as Thr and Ala, also may be used in the linker sequence. Exemplary linkers are disclosed in Maratea et al. (1985), Gene 40: 39-46; Murphy et al. (1986) Proc. Nat'l. Acad. Sci. USA 83: 8258-62; U.S. Pat. No. 4,935,233; and U.S. Pat. No. 4,751,180. For example, GlySer linkers GGS, GGGS or GSG can be used. GGS, GSG, GGGS or GGGGS linkers can be used in repeats of 3 (such as $(GGS)_3$ (SEQ ID NO: 12), (GGGGS)3) or 5, 6, 7, 9 or even 12 or more, to provide suitable lengths. In particular embodiments, linkers such as $(GGGGS)_3$ (SEQ ID NO: 13) are preferably used herein. $(GGGGS)_6$ (SEQ ID NO: 14), $(GGGGS)_9$ (SEQ ID NO: 15) or (GGGGS)i_2 (SEQ ID NO: 16) may preferably be used as alternatives. Other preferred alternatives are (GGGGS) i (SEQ ID NO: 17), $(GGGGS)_2$ (SEQ ID NO: 18), $(GGGGS)_4$ (SEQ ID NO: 19), $(GGGGS)_5$ (SEQ ID NO: 20), $(GGGGS)_7$ (SEQ ID NO: 21), $(GGGGS)_8$ (SEQ ID NO: 22),

(GGGGS)io (SEQ ID NO: 23), or (GGGGS)n (SEQ ID NO: 24). In yet a further embodiment, LEPGEKP YKCPEC GK SF SQSGALTRHQRTHTR (SEQ ID NO: 25) is used as a linker. In yet an additional embodiment, the linker is XTEN linker. In particular embodiments, the CRISPR-cas protein is a Cpfl protein and is linked to the deaminase protein or its catalytic domain by means of an LEP GEKP YK CPEC GK SF SQSGAL TRHQRTHTR (SEQ ID NO:26) linker. In further particular embodiments, the Cpfl protein is linked C-terminally to the N-terminus of a deaminase protein or its catalytic domain by means of an LEPGEKP YKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO:27) linker. In addition, N- and C-terminal NLSs can also function as linker (e.g., PKKKRKVE ASSPKKRKVE AS (SEQ ID **NO:28)).**

[0048] In particular embodiments of the methods of the present invention, the cytidine 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 Cpfl 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, Qp, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, Mil, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, (^Cb8r, **(|>Cbl2r, (|>Cb23r,** 7s and PRRl. 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.

[0049] In particular embodiments of the methods and systems of the present invention, the guide molecule is provided with one or more distinct RNA loop(s) or disctinct sequence(s) that can recruit an adaptor protein. A guide molecule may be extended, without colliding with the Cpfl protein by the insertion of distinct RNA loop(s) or disctinct 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 CRISPR-Cas 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 stemloop and/or in a tetraloop. In these embodiments, the cytidine deaminase protein is fused to MS2. The cytidine deaminase

protein is then co-delivered together with the CRISPR-Cas protein and corresponding guide RNA.

[0050] The term "CD-functionalized CRISPR system" as used here refers to a nucleic acid targeting and editing system comprising (a) a CRISPR-Cas protein, more particularly a Cpfl protein which is catalytically inactive or a nickase; (b) a guide molecule which comprises a guide sequence; and (c) a cytidine deaminase protein or catalytic domain thereof; wherein the cytidine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to the CRISPR-Cas protein or the guide molecule or is adapted to link thereto after delivery; wherein the guide sequence is partially or substantially complementary to the target sequence but comprises a non-pairing Adenine or Uracil corresponding to the C being targeted for deamination, resulting in a C-A/U mismatch in a heteroduplex formed by the guide sequence and the target sequence. In particular embodiments, the guide sequence further comprises at least one additional mismatch or non-pairing nucleotide with the target sequence, adjacent to said non-pairing Adenine or Uracil. This results in a region of single strand DNA at the location of the Cytosine of interest in the target strand, allowing the cytidine deaminase to act specifically thereon. In some embodiments, the guide sequence of the guide molecule comprises at least one further mismatch or non-pairing nucleotide with the target sequence, adjacent to said non-pairing Adenine or Uracil. In some embodiments, the guide sequence comprises a stretch of three to five consecutive mismatches with the target sequence. In some embodiments the guide sequence comprises less than 8 consecutive mismatches with the target sequence. In some embodiments, the guide sequence comprises one or more mismatches 3' and one or more mismatches 5' of the non-pairing Adenine or Uracil. In particular embodiments, the guide sequence comprises one, two three or four mismatches with the target sequence on each side of the non-pairing Adenine of Uracil. Details on the aspect of the guide of the CD-functionalized CRISPR-Cas system are provided herein below.

[0051] 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.

[0052] 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 CRISPR-Cas protein, the cytidine deaminase protein, and

optionally the adaptor protein. The RNA molecules can be delivered via one or more lipid nanoparticles.

[0053] 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 CRISPR-Cas protein, the guide molecule, and the cytidine deaminase protein or catalytic domain thereof, optionally wherein the one or more regulatory elements comprise inducible promoters. For application in eukaryotic cells, the Cpfl protein and/or the cytidine deaminase are preferably NLS-tagged.

[0054] In some embodiments, the CRISPR-Cas protein is a Cpfl nickase. In some embodiments, the Cpfl nickase is capable of nicking a non-targeted DNA strand at the target locus of interest displaced by the formation of the heteroduplex between the targeted DNA strand and the guide molecule. In some embodiments, the Cpfl nickase comprises a mutation in the Nuc domain. Details on the aspect of the CRISPR-Cas protein in the CD-functionalized CRISPR-Cas system are provided herein elsewhere.

[0055] In some embodiments, the Cpfl nickase comprises a mutation corresponding to R1226A in AsCpfl.

[0056] In some embodiments, the CRISPR-Cas protein is a dead Cpfl. In some embodiments, the dead Cpfl comprises a mutation in the RuvC domain. In some embodiments, the dead Cpfl comprises a mutation corresponding to D908A or E993A in AsCpfl.

[0057] In some embodiments of the guide molecule is capable of hybridizing with a target sequence comprising the C to be deaminated within a first DNA strand at the target locus to form a heteroduplex which comprises a non-pairing A or U opposite to said C, optionally within a stretch of up to 5, 6, 7 or 8 mismatching nucleotides. Upon heteroduplex formation, the guide molecule forms a complex with the Cpfl protein and directs the complex to bind said first DNA strand at the target locus of interest, allowing the targeted C to be deaminased by the CD. Details on the aspect of the guide of the CD-functionalized CRISPR-Cas system are provided herein below.

[0058] In some embodiments, a Cpfl guide RNA having a canonical length (e.g., about 24 nt for AsCpfl) is used to form a heteroduplex with the target DNA. In some embodiments, a Cpfl guide molecule longer than the canonical length (e.g., >24 nt for

AsCpfl) is used to form a heteroduplex with the target DNA including outside of the Cpfl-guide RNA-target DNA complex. In certain example embodiments, the guide sequence has a length of about 20-53 nt, or about 25-53 nt, or about 29-53 nt capable of forming a DNA-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 duplex 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. In particular embodiments, the guide sequence comprises more than one mismatch corresponding to different adenosine sites in the target DNA sequence or wherein two guide molecules are used, each comprising a mismatch corresponding to a different adenosine sites in the target RNA sequence.

[0059] In at least a first design, the CD-functionalized CRISPR system comprises (a) a cytidine deaminase fused or linked to a CRISPR-Cas protein, wherein the CRISPR-Cas protein is catalytically inactive or a nickase, and (b) a guide molecule comprising a guide sequence designed to introduce a C-A/U mismatch in a heteroduplex formed between the guide sequence and the target sequence. In some embodiments, the CRISPR-Cas protein and/or the cytidine deaminase are NLS-tagged, on either the N- or C-terminus or both.

[0060] In at least a second design, the CD-functionalized CRISPR system comprises (a) a CRISPR-Cas protein that is catalytically inactive or a nickase, (b) a guide molecule comprising a guide sequence designed to introduce a C-A/U mismatch in a heteroduplex 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) a cytidine deaminase fused or linked to an adaptor protein, wherein the binding of the aptamer and the adaptor protein recruits the cytidine deaminase to the heteroduplex formed between the guide sequence and the target sequence for targeted deamination at the C of the C-A/U mismatch. In some embodiments, the adaptor protein and/or the cytidine deaminase are NLS-tagged, on either the N- or C-terminus or both. The CRISPR-Cas protein can also be NLS-tagged.

[0061] The use of different aptamers and corresponding adaptor proteins also allows orthogonal gene editing to be implemented. In one example in which cytidine deaminase are used in combination with adenosine 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-cytidine deaminase, while another sgRNA targeting locus B can be modified with PP7 loops, recruiting PP7-adenosine deaminase. In the same cell, orthogonal, locus-specific modifications are thus realized. This principle can be extended to incorporate other orthogonal RNA-binding proteins.

[0062] In at least a third design, the CD-functionalized CRISPR system comprises (a) a cytidine deaminase inserted into an internal loop or unstructured region of a CRISPR-Cas protein, wherein the CRISPR-Cas protein is catalytically inactive or a nickase, and (b) a guide molecule comprising a guide sequence designed to introduce a C-A/U mismatch in a heteroduplex formed between the guide sequence and the target sequence.

[0063] CRISPR-Cas protein split sites that are suitable for inseration of cytidine deaminase can be identified with the help of a crystal structure. One can use the crystal structure of an ortholog if a relatively high degree of homology exists between the ortholog and the intended CRISPR-Cas protein.

[0064] 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. 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.

[0065] For Cpfl, several small stretches of unstructured regions have been predicted within the Cpfl primary structure (see WO20 1620571 1, the content of which is hereby incorporated by reference). Unstructured regions, which are exposed to the solvent and not conserved within different Cpfl orthologs, are preferred sides for splits.

[0066] The following table presents non-limiting potential split regions within AsCpfl and LbCpf 1. A split site within such a region may be opportune.

Split region	AsCpf1	LbCpf1
1	575-588	566-571
2	631-645	754-757
3	653-664	-
4	818-844	-

[0067] For Fn, As and Lb Cpf1 mutants, it should be readily apparent what the corresponding position for a potential split site is, for example, based on a sequence alignment. For non-Fn, As and Lb enzymes one can use the crystal structure of an ortholog if a relatively high degree of homology exists between the ortholog and the intended Cpfl, or one can use computational prediction.

[0068] The CD-functionalized CRISPR system described herein can be used to target a specific Cytosine within a DNA sequence for deamination. For example, the guide molecule can form a complex with the CRISPR-Cas 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 A or U with the Cytosine of interest, which in particular embodiments is extended to a stretch of non-pairing nucleotides, the heteroduplex formed between the guide sequence and the target sequence comprises a region of single strand DNA, which directs the cytidine deaminase to contact and deaminate the C opposite to the non-pairing A or U, converting it to an Uracil (U). Since Uracil (U) base pairs with A and functions like T in cellular process, the targeted deamination of C described herein are useful for correction of undesirable T-C and A-G mutations, as well as for obtaining desirable G-A and C-T mutations. Alternatively, the guide sequence can be designed to be upstream or downstreat of the Cytosine of interest, allowing the DNA-RNA duplex to form upstream or downstream of the Cytosine of interesting, such that the cytidine deaminase can contact and deaminate the C in a region of single strand DNA outside the DNA-RNA duplex.

[0069] The use of a Cpfl nickase that nicks the non-targeted DNA strand is particularly advantageous. Subsequent to the deamination of the targeted C, the U:G mismatch obtained can be processed by mismatch repair mechanisms, which preferentially repair the nicked strand of a mismatch. The nick (single-strand DNA break) made by the Cpfl nickase on the non-targeted DNA strand containing the mismatched G facilitates removal of the

mismatched G via mismatch repair mechanisms, leading to conversion of the U:G mismatch to the desired U:A or T:A outcome (i.e., a G to A change on the non-targeted DNA strand).

[0070] In addition to facilitating the aforementioned G to A change on the non-targeted DNA strand, the Cpfl nickase can also facilitate repair of unintened deamination of C on the non-targeted DNA strand. Without being bound by limitation, in some embodiments, the CD-functionalized CRISPR system described herein may cause unintended deamination of at least one C on the non-targeted DNA strand, when the non-targeted DNA strand is displaced by the binding of the Cpfl -guide complex with the targeted DNA strand, and it is believed that nicking of the non-targeted DNA strand by the Cpfl nickase will facilitate repair of such unintended deamination by mismatch repair mechanisms, which preferentially repair the nicked strand of a mismatch.

[0071] In some embodiments, the CD-functionalized CRISPR system is used for targeted deamination in a DNA molecule *in vitro.* In some embodiments, the CD-functionalized CRISPR system is used for targeted deamination in a DNA molecule within a cell. The cell can be a eukaryotic cell, such as a animal cell, a mammalian cell, a human, or a plant cell.

[0072] The invention also relates to a method for treating or preventing a disease by the targeted deamination using the CD-functionalized CRISPR system, wherein the deamination of the C restores a healthy genotype at the target locus of interest, which remedies a disease caused by a T→C or A→G point mutation or a pathogenic SNP.

[0073] The invention also relates to a method for knocking-out or knocking-down an undesirable activity of a gene or regulatory element thereof, wherein the deamination of the Cytosine at the target locus of interest inactivates a target gene or a target regulatory element at the target locus. For example, in one embodiment, the targeted deamination by the CD-functionalized CRISPR system can cause a nonsense mutation resulting in a premature stop codon in an endogenous gene. This may alter the expression of the endogenous gene and can lead to a desirable trait in the edited cell. In another embodiment, the targeted deamination by the CD-functionalized CRISPR system can cause a nonconservative missense mutation resulting in a code for a different amino acid residue in an endogenous gene. This may alter the function of the endogenous gene expressed and can also lead to a desirable trait in the edited cell.

[0074] The invention also relates to a modified cell obtained by the targeted deamination using the CD-functionalized CRISPR system, or progeny thereof, wherein the

modified cell comprises an U or T in replace of the C in the target locus of interest compared to a corresponding cell before the targeted deamination. The modified cell can be a eukaryotic cell, such as an animal cell, a plant cell, a mammalian cell, or a human cell.

[0075] In some embodiments, the modified cell is a therapeutic T cell, such as a T cell sutiable for CAR-T therapies. The modification may result in one or more desirable traits in the therapeutic T cell, including but not limited to, reduced expression of an immune checkpoint receptor (e.g., PDA, CTLA4), reduced expression of HLA proteins (e.g., B2M, HLA-A), and reduced expression of an endogenous TCR.

[0076] In some embodiments, the modified cell is an antibody-producing B cell. The modification may results in one or more desirable traits in the B cell, including but not limited to, enhanced antibody production.

[0077] The invention also relates to a modified non-human animal or a modified plant. The modified non-human animal can be a farm animal. The modified plant can be an agricultural crop.

[0078] The invention further relates to a method for cell therapy, comprising administering to a patient in need thereof the modified cell described herein, wherein the presence of the modified cell remedies a disease in the patient. In one embodiment, the modified cell for cell therapy is a CAR-T cell capable of recognizing and/or attacking a tumor cell. In another embodiment, the modified cell for cell therapy is a stem cell, such as a neural stem cell, a mesenchymal stem cell, a hematopoietic stem cell, or an iPSC cell.

[0079] The invention additionally relates to an engineered, non-naturally occurring system suitable for modifying an Cytosine in a target locus of interest, comprising: a guide molecule which comprises a guide sequence, or a nucleotide sequence encoding the guide molecule; a CRISPR-Cas protein, or one or more nucleotide sequences encoding the CRISPR-Cas protein; a cytidine deaminase protein or catalytic domain thereof, or one or more nucleotide sequences encoding; wherein the cytidine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to the CRISPR-Cas protein or the guide molecule or is adapted to link thereto after delivery; wherein the guide sequence is capable of hybridizing with a target sequence comprising an Cytosine within the target locus, but comprises a Adenine or Uracil at the position corresponding to the Cytosine.

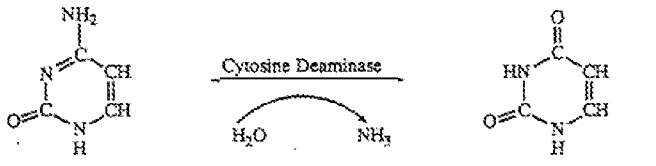
[0080] The invention additionally relates to an engineered, non-naturally occurring vector system suitable for modifying an Cytosine in a target locus of interest, comprising one or more vectors comprising: a first regulatory element operably linked to one or more

nucleotide sequences encoding a guide molecule which comprises a guide sequence; a second regulatory element operably linked to a nucleotide sequence encoding a CRISPR-Cas protein; and optionally a nucleotide sequence encoding a cytidine deaminase protein protein or catalytic domain thereof which is under control of the first or second regulatory element or operably linked to a third regulatory element; wherein, if the nucleotide sequence encoding a cytidine deaminase protein or catalytic domain thereof is operably linked to a third regulatory element, the cytidine deaminase protein or catalytic domain thereof is adapted to link to the guide molecule or the Crispr-Cas protein after expression; wherein the guide sequence is capable of hybridizing with a target sequence comprising an Cytosine within the target locus, but comprises an Adenine or a Uracil at the position corresponding to the Cytosine; wherein components (a), (b) and (c) are located on the same or different vectors of the system.

[0081] The invention additionally relates to in vitro, ex vivo or in vivo host cell or cell line or progeny thereof comprising the engineered, non-naturally occurring system or vector system described herein. The host cell can be a eukaryotic cell, such as an animal cell, a plant cell, a mammalian cell, or a human cell.

[0082] **Cytidine deaminase**

[0083] 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 an cytosine (or an 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)

[0084] 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 APOBECl 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.

[0085] 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 cytodine deaminase may edit on a single strand present outside of the binding component e.g. bound Casl3. 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 are of activity such as those disclosed in Kim *et al.,* Nature Biotechnology (2017) 35(4):371-377 (doi:10.1038/nbt.3803).

[0086] 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.

[0087] In some embodiments, the cytidine deaminase is a human APOBEC, including hAPOBECl or hAPOBEC3. In some embodiments, the cytidine deaminase is a human AID.

[0088] In some embodiments, the cytidine deaminase protein recognizes and converts one or more target cytosine residue(s) in a single-stranded bubble of a DNA-RNA heteroduplex into uracil residues (s). In some embodiments, the cytidine deaminase protein recognizes a binding window on the single-stranded bubble of a DNA-RNA heteroduplex. 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.

[0089] 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 a DNA-RNA heteroduplex 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.

[0090] In some embodiments, the cytidine deaminase comprises human APOBEC1 full protein (hAPOBECl) or the deaminase domain thereof (hAPOBECl-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 hAPOBECl, 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.

[0091] 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.

[0092] Certain mutations of APOBEC 1 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.

[0093] In some embodiments, the cytidine deaminase is an APOBEC 1 deaminase comprising one or more mutations at amino acid positions corresponding to W90, R 118, 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.

[0094] In some embodiments, the cytidine deaminase comprises a mutation at tryptophane ^{90} of the rat APOBECl amino acid sequence, or a corresponding position in a homologous APOBEC protein, such as tryptophane ^{285} of APOBEC3G. In some embodiments, the tryptophane residue at position 90 is replaced by an tyrosine or phenylalanine residue (W90Y or W90F).

[0095] In some embodiments, the cytidine deaminase comprises a mutation at Arginine ^{118} of the rat APOBECl 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 (Rl 18A).

[0096] In some embodiments, the cytidine deaminase comprises a mutation at Histidine ^{121} of the rat APOBECl 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).

[0097] In some embodiments, the cytidine deaminase comprises a mutation at Histidine ^{122} of the rat APOBECl 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).

[0098] In some embodiments, the cytidine deaminase comprises a mutation at Arginine ^{126} of the rat APOBECl amino acid sequence, or a corresponding position in a homologous APOBEC protein, such as Arginine ^{320} of APOBEC3G. In some embodiments, the arginine residue at position 126 is replaced by an alanine residue (R126A) or by a glutamic acid (R126E).

[0099] In some embodiments, the cytidine deaminase comprises a mutation at arginine ^{132} of the APOBECl 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).

[00100] 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 APOBECl, and mutations in a homologous APOBEC protein corresponding to the above.

[00101] In some embodiments, to reduce editing efficiency, the cytidine deaminase may comprise one or more of the mutations: W90A, R 118A, R132E, based on amino acid sequence positions of rat APOBECl, 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 efficicay to reduce off-target effects.

[00102] In some embodiments, the cytidine deaminase is wild-type rat APOBEC1 (rAPOBECl, SEQ ID NO:29) or a catalytic domain thereof. In some embodiments, the cytidine deaminase comprises one or more mutations in the rAPOBECl sequence, such that the editing efficiency, and/or substrate editing preference of rAPOBECl is changed according to specific needs.

[00103] rAPOBECl:

MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSIWRHTSQN
TNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYPHVTLFIYI
ARLYHHADPRNRQGLRDLISSGVTIQEVITEQESGYCWRNFVNYSPSNEAHWPRYPH
LWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHILWATGLK
(SEQ ID NO: 29)

[00104] In some embodiments, the cytidine deaminase is wild-type human APOBEC1 (hAPOBECl, SEQ ID NO:30) or a catalytic domain thereof. In some embodiments, the cytidine deaminase comprises one or more mutations in the hAPOBECl sequence, such that the editing efficiency, and/or substrate editing preference of hAPOBECl is changed according to specific needs.

[00105] hAPOBECl:

MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKIWRSSGK
NTTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRHPGVTLVI
YVARLFWHMDQQNRQGLRDLVNSGVTIQFMRASEYYHCWRNFVNYPPGDEAHW
PQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQTIPPHILLA
TGLIHPSVAWR (SEQ ID NO: 30)

[00106] In some embodiments, the cytidine deaminase is wild-type human APOBEC3G (hAPOBEC3G, SEQ ID NO:31) 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.

[00107] hAPOBEC3G:

MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMATFLAEDPK
VTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKFMNYDEFQHCWSKFVYSQ
RELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHETYLCYEVE

RMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDLDQDYRV TCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISI MTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 31)

[00108] In some embodiments, the cytidine deaminase is wild-type Petromyzon marinus CDAl (pmCDAl, SEQ ID NO:32) 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.

[00109] pmCDAl:

MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACFWGYAVN KPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILEWYNQEL RGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCCRKIFIQSS HNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV (SEQ ID NO: 32)

[00110] In some embodiments, the cytidine deaminase is wild-type human AID (hAID, SEQ ID NO:33) 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.

[00111] hAID:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNG CHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPYLSLRIFT ARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGL IIENSVRLSRQLRRILLPLYEVDDLRDAFRTLGLLD (SEQ ID NO: 33)

[00112] In some embodiments, the cytidine deaminase is truncated version of hAID (hAID-DC, SEQ ID NO:34) 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.

[00113] hAID-DC:

MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFGYLRNKNG CHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNPNLSLRIFT

ARLYFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTFKAWEGL HENSVRLSRQLRRILL (SEQ ID NO: 34)

[00114] Additional embodiments of the cytidine deaminase are disclosed in WO WO20 17/070632, titled "Nucleobase Editor and Uses Thereof," which is incorporated herein by reference in its entirety.

[00115] 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.

[00116] 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 2 1 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.

[00117] 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 Cpfl nickase 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 Cpfl nickase is fused to or linked to an APOBEC3G mutant that comprises a W285Y or W285F mutation.

[00118] 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 Cpfl nickase 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 Cpfl nickase 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 Cpfl nickase is fused to or linked to an APOBEC1 mutant that comprises a R132E mutation.

[00119] 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 APOBECl domain comprises double mutations of W90Y and R126E. In some embodiments, the APOBECl domain comprises double mutations of W90Y and R132E. In some embodiments, the APOBECl domain comprises double mutations of R126E and R132E. In some embodiments, the APOBECl domain comprises three mutations of W90Y, R126E and R132E.

[00120] 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.

[00121] 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.

[00122] According to the present invention, the substrate of the cytidine deaminase is an DNA single strand bubble of a DNA-RNA heteroduplex 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.

[00123] CRISPR-Cas Protein and Guide

[00124] In the methods and systems of the present invention use is made of a CRISPR-Cas protein and corresponding guide molecule. More particularly, the CRISPR-Cas protein is a class 2 CRISPR-Cas protein. In certain embodiments, said CRISPR-Cas protein Cpfl. The CRISPR-Cas system does not require the generation of customized proteins to target specific sequences but rather a single Cas protein can be programmed by guide molecule to recognize a specific nucleic acid target, in other words the Cas enzyme protein can be recruited to a specific nucleic acid target locus of interest using said guide molecule.

[00125] *Guide molecule*

[00126] The guide molecule or guide RNA of a Class 2 type V CRISPR-Cas protein comprises a tracr-mate sequence (encompassing a "direct repeat") and a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system). Indeed, in contrast to the type II CRISPR-Cas proteins, the CRISPR-Cas Cpfl protein does not rely on the presence of a tracr sequence. In embodiments, the CRISPR-Cas system or complex as described herein does not comprise and/or does not rely on the presence of a tracr sequence

(e.g. if the Cas protein is Cpfl). In certain embodiments, the guide molecule may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence.

[00127] In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence. 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 DNA sequence and a guide sequence promotes the formation of a CRISPR complex.

[00128] The terms "guide molecule" and "guide RNA" are used interchangeably herein to refer to RNA-based molecules that are capable of forming a complex with a CRISPR-Cas protein and comprises a guide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of the complex to the target nucleic acid sequence. The guide molecule or guide RNA specifically encompasses RNA-based molecules having one or more chemically modifications (e.g., by chemical linking two ribonucleotides or by replacement of one or more ribonucleotides with one or more deoxyribonucleotides), as described herein.

[00129] As used herein, the term "guide sequence" 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. In the context of the present invention the target nucleic acid sequence or target sequence is the sequence comprising the target cytosine to be deaminated also referred to herein as the "target cytosine". In some embodiments, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. However, as detailed herein, preferably, given that the presence of at least one mismatch with the target sequence is provided. Indeed, in the context of the present invention the guide molecule comprises a guide sequence that is designed to have at least one mismatch with the target sequence, such that a heteroduplex formed between the guide sequence and the target sequence comprises a non-pairing A or U in the guide sequence opposite to the target C for deamination on 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.

[00130] 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.

[00131] 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 comprising the cytosine to be deaminated. This is described more in detail below. Selection can encompass further steps which increase efficacy and specificity of deamination.

[00132] In some embodiments, the C-U/A mismatch corresponding to the target site of interest is located close to the center of the target sequence and thus the center of the guide sequence, thereby restricting the cytidine deaminase to a narrow editing window (e.g., about 4 bp wide). In some embodiments, the target sequence may comprise more than one target cytosine to be deaminated. In further embodiments the target sequence may further comprise one or more cytosines to be deaminated 3' to the target cytosine site. In these embodiments, further C-U/A mismatches can be provided in the guide sequence. Depending on their location within the target sequence, these may be located within the same or a separate stretch of mismatching or non-pairing nucleotides in the guide sequence. In some embodiments, to avoid off-target editing, the target sequence does not have further mismatched cytosines within the target sequence.

[00133] In some embodiments, a Cpfl guide sequence having a canonical length (e.g., about 24 nt for AsCpfl) is used to hybridize with the target DNA. In some embodiments, a Cpfl guide molecule longer than the canonical length (e.g., >24 nt for AsCpfl) is used to

hybridize with the target DNA, such that a region of the guide sequence hybridizes with a region of the DNA strand outside of the Cpfl -guide RNA-target DNA complex. This can be of interest where deamination of more than one cytosine 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 C-U/A mismatch outside of the canonical length of Cpfl guide, distant from the PAM sequence, which may decrease steric hindrance by Cpfl and increase the frequency of contact between the cytidine deaminase and the C-U/A mismatch.

[00134] 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).

[00135] In some embodiments, it is of interest to reduce the susceptibility of the guide molecule to RNA cleavage, such as to cleavage by Cpfl. Accordingly, in particular embodiments, the guide molecule is adjusted to avoide cleavage by Cpfl or other RNA-cleaving enzymes.

[00136] 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'-0-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'-0-methyl (M), 2'-0-methyl 3'phosphorothioate (MS), ^-constrained ethyl(cEt), or 2'-0-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,* E71 10-E71 11; Allerson et al., *J. Med. Chem.* 2005, 48:901-904; Bramsen et al., *Front. Genet.,* 2012, 3:154; Deng et al., *PNAS,* 2015, 112:1 1870-1 1875; Sharma et *a\., 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 DNA and one or more deoxyribonucletides and/or nucleotide analogs in a region that binds to Cpfl. 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. For Cpfl guide, in certain embodiments, the modification is not in the 5'-handle of the stem-loop regions. Chemical modification in the 5'-handle of the stem-loop region of a guide may abolish its function (see Li, et al., *Nature Biomedical Engineering,* 2017, 1:0066). In certain embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides of a guide is chemically modified. In some embodiments, 3-5 nucleotides at either the 3' or the 5' end of a guide is chemically modified. In some embodiments, only minor modifications are introduced in the seed region, such as 2'-F modifications. In some embodiments, 2'-F modification is introduced at the 3' end of a guide. In certain embodiments, three to five

nucleotides at the 5' and/or the 3' end of the guide are chemicially modified with **2'-0-**methyl (M), 2'**-0** -methyl 3' phosphorothioate (MS), S-constrained ethyl(cEt), or **2'-0-**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 chemicially modified with 2'**-0** -Me, 2'-F or ^-constrained ethyl(cEt). Such chemically modified guide can mediate enhanced levels of gene disruption (see Ragdarm et al., 0215, *PNAS,* E71 10-E71 11). In an embodiment of the invention, a guide is modified to comprise a chemical moiety at its 3' and/or 5' end. Such moieties include, but are not limited to amine, azide, alkyne, thio, dibenzocyclooctyne (DBCO), or Rhodamine. In certain embodiment, the chemical moiety is conjugated to the guide by a linker, such as an alkyl chain. In certain embodiments, the chemical moiety of the modified guide can be used to attach the guide to another molecule, such as DNA, RNA, protein, or nanoparticles. Such chemically modified guide can be used to identify or enrich cells generically edited by a CRISPR system (see Lee et al., *eLife,* 2017, 6:e25312, DOI: 10.7554).

[00137] In some embodiments, the guide comprises a modified Cpfl crRNA, having a 5'-handle and a guide segment further comprising a seed region and a 3'-terminus. In some embodiments, the modified guide can be used with a Cpfl of any one of Acidaminococcus sp. BV3L6 Cpfl (AsCpfl); Francisella tularensis subsp. Novicida U112 Cpfl (FnCpfl); L. bacterium MC2017 Cpfl (Lb3Cpfl); Butyrivibrio proteoclasticus Cpfl (BpCpfl); Parcubacteria bacterium GWC2011_GWC2_44_17 Cpfl (PbCpfl); Peregrinibacteria bacterium GW2011_GWA_33_10 Cpfl (PeCpfl); Leptospira inadai Cpfl (LiCpfl); Smithella sp. SC_K08D17 Cpfl (SsCpfl); L. bacterium MA2020 Cpfl (Lb2Cpfl); Porphyromonas crevioricanis Cpfl (PeCpfl); Porphyromonas macacae Cpfl (PmCpfl); Candidatus Methanoplasma termitum Cpfl (CMtCpfl); Eubacterium eligens Cpfl (EeCpfl); Moraxella bovoculi 237 Cpfl (MbCpfl); Prevotella disiens Cpfl (PdCpfl); or L. bacterium ND2006 Cpfl (LbCpfl).

[00138] In some embodiments, the modification to the guide is a chemical modification, an insertion, a deletion or a split. In some embodiments, the chemical modification includes, but is not limited to, incorporation of 2'**-0** -methyl (M) analogs, 2'-deoxy analogs, 2-thiouridine analogs, N6-methyladenosine analogs, 2'-fluoro analogs, 2-aminopurine, 5-

bromo-uridine, pseudouridine (Ψ), Nl-methylpseudouridine (ηeIΨ), 5-methoxyuridine(5moU), inosine, 7-methylguanosine, 2'-0-methyl 3'phosphorothioate (MS), S-constrained ethyl(cEt), phosphorothioate (PS), or 2'-0-methyl 3'thioPACE (MSP). In some embodiments, the guide comprises one or more of phosphorothioate modifications. In certain embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 1i, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 25 nucleotides of the guide are chemically modified. In certain embodiments, one or more nucleotides in the seed region are chemically modified. In certain embodiments, one or more nucleotides in the 3'-terminus are chemically modified. In certain embodiments, none of the nucleotides in the 5'-handle is chemically modified. In some embodiments, the chemical modification in the seed region is a minor modification, such as incorporation of a 2'-fluoro analog. In a specific embodiment, one nucleotide of the seed region is replaced with a 2'-fluoro analog. In some embodiments, 5 to 10 nucleotides in the 3'-terminus are chemically modified. Such chemical modifications at the 3'-terminus of the Cpfl CrRNA may improve Cpfl activity (see Li, et al., Nature Biomedical Engineering, 2017, 1:0066). In a specific embodiment, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides in the 3'-terminus are replaced with 2'-fluoro analogues. In a specific embodiment, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides in the 3'-terminus are replaced with 2'- O-methyl (M) analogs.

[00139] In some embodiments, the loop of the 5'-handle of the guide is modified. In some embodiments, the loop of the 5'-handle of the guide is modified to have a deletion, an insertion, a split, or chemical modifications. In certain embodiments, the modified loop comprises 3, 4, or 5 nucleotides. In certain embodiments, the loop comprises the sequence of UCUU, UUUU, UAUU, or UGUU.

[00140] In some embodiments, the guide molecule forms a stemloop 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.

[00141] 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: 11820-1 1821; Scaringe, Methods Enzymol. (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., J. Am. Chem. Soc. (201 1) 133: 11540-1 1546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989).

[00142] In certain embodiments, the guide molecule (capable of guiding Cpfl to a target locus) 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') 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 Cpfl guide sequence is approximately within the first 10 nucleotides of the guide sequence. In particular embodiments, the Cpfl is FnCpfl and the seed sequence is approximately within the first 5 nt on the 5' end of the guide sequence.

[00143] 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, preferrably 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 Cpfl guide molecule comprises (in 3' to 5' 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 architercture 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 Cpfl protein and/or target, for example the stemloop of the direct repeat sequence.

[00144] In particular embodiments, the stem comprises at least about 4bp 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 stemloop 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 stemloop at that position.

[00145] In particular embodiments the natural hairpin or stemloop structure of the guide molecule is extended or replaced by an extended stemloop. It has been demonstrated that extension of the stem can enhance the assembly of the guide molecule with the CRISPR-Cas proten (Chen et al. Cell. (2013); 155(7): 1479-1491). In particular embodiments the stem of the stemloop 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 stemloop.

[00146] 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 stemloop of the guide molecule, it is preferably ensured by a basepair flip.

[00147] 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.

[00148] In some embodiments, the guide molecule forms a duplex with a target DNA strand comprising at least one target cytosine residue to be edited. Upon hybridization of the guide RNA molecule to the target DNA strand, the cytidine deaminase binds to the single DNA strand 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.

[00149] A guide sequence, and hence a nucleic acid-targeting guide RNA may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be genomic DNA. The target sequence may be mitochondrial DNA.

[00150] 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 Cpfl protein, the compelementary sequence of the target sequence is downstream or 3' of the PAM. The precise sequence and length requirements for the PAM differ depending on the Cpfl 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 Cpfl orthologues are provided herein below and the skilled person will be able to identify further PAM sequences for use with a given Cpfl protein.

[00151] 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/naturel4592. As further detailed herein, the skilled person will understand that Cpfl proteins may be modified analogously.

[00152] In particular embodiments, the guide sequence is selected in order to ensure optimal efficiency of the deaminase on the cytosine to be deaminated. The position of the cytosine in the target strand relative to the cleavage site of the Cpfl nickase may be taken into account, In particular embodiments it is of interest to ensure that the nickase will act in the vicinity of the cytosine to be deaminated, on the non-target strand. For instance, in particular embodiments, the Cpfl nickase cuts the non-targeting strand 17 nucleotides downstream of the PAM (e.g. AsCpfl, LbCpfl) or 18 nucleotides downstream of the PAM (e.g. FnCpfl), and it can be of interest to design the guide that the cytosine which is to correspond to the cytosine to be deaminated is located in the guide sequence within 10 bp upstream or downstream of the nickase cleavage site in the sequence of the corresponding non-target strand.

[00153] In particular embodiment, the guide is an escorted guide. By "escorted" is meant that the Cpfl CRISPR-Cas system or complex or guide is delivered to a selected time or place within a cell, so that activity of the Cpfl CRISPR-Cas system or complex or guide is spatially or temporally controlled. For example, the activity and destination of the Cpfl 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.

[00154] The escorted Cpfl 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.

[00155] 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 flourescent protein (Paige, Jeremy S., Karen Y. Wu, and Sarnie R. Jaffrey. "RNA mimics of green fluorescent protein." Science 333.6042 (201 1): 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).

[00156] 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 an guide molecule that responds to normal or pathological physiological conditions, including without limitation pH, hypoxia, $\mathbf{0}_2$ concentration, temperature, protein concentration, enzymatic concentration, lipid structure, light exposure, mechanical disruption (e.g. ultrasound waves), magnetic fields, electric fields, or electromagnetic radiation.

[00157] 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.

[00158] 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.

[00159] 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 Cpfl 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 Cpfl CRISPR-Cas system or complex function; and optionally further determining that the expression of the genomic locus is altered.

[00160] There are several different designs of this chemical inducible system: 1. ABI-PYL based system inducible by Abscisic Acid (ABA) (see, e.g., http://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., http://www.nature.com/nmeth/journal/v2/n6/full/nmeth763.html), 3. GID1-GAI based system inducible by Gibberellin (GA) (see, e.g., http://www.nature.com/nchembio/journal/v8/n5/full/nchembio.922.html).

[00161] A chemical inducible system can be an estrogen receptor (ER) based system inducible by 4-hydroxytamoxifen (40HT) (see, e.g., http://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, estrogren receptor, estrogen-related receptor, glucocorticoid receptor, progesterone receptor, androgen receptor may be used in inducible systems analogous to the ER based inducible system.

[00162] 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., http://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 Cpfl 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 Cpfl CRISPR-Cas complex will be active and modulating target gene expression in cells.

[00163] 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.

[00164] 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.

[00165] 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).

[00166] 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.

[00167] 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).

[00168] 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).

[00169] 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 .mu.s duration. Such a pulse may be generated, for example, in known applications of the Electro Square Porator T820.

[00170] 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.

[00171] 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.

[00172] 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.

[00173] A preferred embodiment employs direct current at low voltage. Thus, Applicants disclose the use of an electric field which is applied to the cell, tissue or tissue mass at a field strength of between lV/cm and 20V/cm, for a period of 100 milliseconds or more, preferably 15 minutes or more.

[00174] Ultrasound is advantageously administered at a power level of from about 0.05 W/cm^2 to about 100 W/cm^2. Diagnostic or therapeutic ultrasound may be used, or combinations thereof.

[00175] 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]).

[00176] 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/cm^2 (FDA recommendation), although energy densities of up to 750 mW/cm^2 have been used. In physiotherapy, ultrasound is typically used as an energy source in a range up to about 3 to 4 W/cm^2 (WHO recommendation). In other therapeutic applications, higher intensities of ultrasound may be employed, for example, HIFU at 100 W/cm up to 1 kW/cm^2 (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.

[00177] 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. 1103-1 106.

[00178] 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.

[00179] 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}.

[00180] 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.

[00181] 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.

[00182] Advantageously, the target tissue is exposed to an ultrasound energy source at an acoustic power density of from about 0.05 Wcm^{-2} to about 10 Wcm^{-2} with a frequency ranging from about 0.015 to about 10 MHz (see WO 98/52609). However, alternatives are also possible, for example, exposure to an ultrasound energy source at an acoustic power density of above 100 Wcm^{-2}, but for reduced periods of time, for example, 1000 Wcm^{-2} for periods in the millisecond range or less.

[00183] 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.

[00184] 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.

[00185] 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.

[00186] 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.

[00187] 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 DNA binding to the mismatched basepairs at the 3' end. In particular embodiments of the invention, additional sequences comprising an extented 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 moleucle, the guide molecule is considered protected and results in improved specific binding of the CRISPR-Cas complex, while maintaining specific activity.

[00188] 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 DNA. 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.

[00189] *Crispr-Cas Enzyme*

[00190] In its unmodified form, a CRISPR-Cas protein is a catalytically active protein. This implies that upon formation of a nucleic acid-targeting complex (comprising a guide RNA hybridized to a target sequence one or both DNA strands in or near (e.g., within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence is modified (e.g. cleaved). As used herein the term "sequence(s) associated with a target locus of interest" refers to sequences near the vicinity of the target sequence (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from the target sequence, wherein the target sequence is comprised within a target locus of interest). The unmodified catalytically active Cpfl protein generates a staggered cut, whereby the cut sites are typically within the target sequence. More particularly, the staggered cut is typically 13-23 nucleotides distal to the PAM. In particular embodiments, the cut on the non-target strand is 17 nucleotides downstream of the PAM (i.e. between nucleotide 17 and 18 downstream of the PAM), while the cut on the target strand (i.e. strand hybridizing with the guide sequence) occurs a further 4 nucleotides further from the sequence complementary to the PAM (this is 21 nucleotides upstream of the complement of the PAM on the 3' strand or between nucleotide 21 and 22 upstream of the complement of the PAM).

[00191] In the methods according to the present invention, the CRISPR-Cas protein is preferably mutated with respect to a corresponding wild-type enzyme such that the mutated CRISPR-Cas protein lacks the ability to cleave one or both DNA strands of a target locus containing a target sequence. In particular embodiments, one or more catalytic domains of the Cpfl protein are mutated to produce a mutated Cas protein which cleaves only one DNA strand of a target sequence.

[00192] In particular embodiments, the CRISPR-Cas protein may be mutated with respect to a corresponding wild-type enzyme such that the mutated CRISPR-Cas protein lacks substantially all DNA cleavage activity. In some embodiments, a CRISPR-Cas protein may be considered to substantially lack all DNA and/or RNA cleavage activity when the cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the nucleic acid cleavage activity of the non-mutated form of the enzyme; an example can be when the nucleic acid cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form.

[00193] In certain embodiments of the methods provided herein the CRISPR-Cas protein is a mutated CRISPR-Cas protein which cleaves only one DNA strand, i.e. a nickase. More particularly, in the context of the present invention, the nickase ensures cleavage within the non-target sequence, i.e. the sequence which is on the opposite DNA strand of the target sequence and which is 3' of the PAM sequence. By means of further guidance, and without limitation, an arginine-to-alanine substitution (R1226A) in the Nuc domain of Cpfl from *Acidaminococcus sp.* converts Cpfl from a nuclease that cleaves both strands to a nickase (cleaves a single strand). It will be understood by the skilled person that where the enzyme is not AsCpfl, a mutation may be made at a residue in a corresponding position. In particular embodiments, the Cpfl is FnCpfl and the mutation is at the arginine at position R1218. In particular embodiments, the Cpfl is LbCpfl and the mutation is at the arginine at position R1138. In particular embodiments, the Cpfl is MbCpfl and the mutation is at the arginine at position R1293.

[00194] In certain embodiments of the methods provided herein the CRISPR-Cas protein has reduced or no catalytic activity. Where the CRISPR-Cas protein is a Cpfl protein, the mutations may include but are not limited to one or more mutations in the catalytic RuvC-like domain, such as D908A or E993A with reference to the positions in AsCpfl.

[00195] In some embodiments, a CRISPR-Cas protein is considered to substantially lack all DNA cleavage activity when the DNA cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the DNA cleavage activity of the non-mutated form of the enzyme; an example can be when the DNA cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form. In these embodiments, the CRISPR-Cas protein is used as a generic DNA binding protein. The mutations may be artificially introduced mutations or gain- or loss-of-function mutations.

[00196] In addition to the mutations described above, the CRISPR-Cas protein may be additionally modified. As used herein, the term "modified" with regard to a CRISPR-Cas protein generally refers to a CRISPR-Cas protein having one or more modifications or mutations (including point mutations, truncations, insertions, deletions, chimeras, fusion proteins, etc.) compared to the wild type Cas protein from which it is derived. By derived is meant that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as known in the art or as described herein.

[00197] The additional modifications of the CRISPR-Cas protein may or may not cause an altered functionality. By means of example, and in particular with reference to CRISPR-Cas protein, modifications which do not result in an altered functionality include for instance codon optimization for expression into a particular host, or providing the nuclease with a particular marker (e.g. for visualization). Modifications with may result in altered functionality may also include mutations, including point mutations, insertions, deletions, truncations (including split nucleases), etc.. Fusion proteins may without limitation include for instance fusions with heterologous domains or functional domains (e.g. localization signals, catalytic domains, etc.). In certain embodiments, various different modifications may be combined (e.g. a mutated nuclease which is catalytically inactive and which further is fused to a functional domain, such as for instance to induce DNA methylation or another nucleic acid modification, such as including without limitation a break (e.g. by a different nuclease (domain)), a mutation, a deletion, an insertion, a replacement, a ligation, a digestion, a break or a recombination). As used herein, "altered functionality" includes without limitation an altered specificity (e.g. altered target recognition, increased (e.g. "enhanced" Cas proteins) or decreased specificity, or altered PAM recognition), altered activity (e.g. increased or decreased catalytic activity, including catalytically inactive nucleases or nickases), and/or altered stability (e.g. fusions with destalilization domains). Suitable heterologous domains include without limitation a nuclease, a ligase, a repair protein, a methyltransferase, (viral) integrase, a recombinase, a transposase, an argonaute, a cytidine deaminase, a retron, a group II intron, a phosphatase, a phosphorylase, a sulpfurylase, a kinase, a polymerase, an exonuclease, etc.. Examples of all these modifications are known in the art. It will be understood that a "modified" nuclease as referred to herein, and in particular a "modified" Cas or "modified" CRISPR-Cas system or complex preferably still has the capacity to interact with or bind to the polynucleic acid (e.g.

in complex with theguide molecule). Such modified Cas protein can be combined with the deaminase protein or active domain thereof as described herein.

[00198] In certain embodiments, CRISPR-Cas protein may comprise one or more modifications resulting in enhanced activity and/or specificity, such as including mutating residues that stabilize the targeted or non-targeted strand (e.g. eCas9; "Rationally engineered Cas9 nucleases with improved specificity", Slaymaker et al. (2016), Science, 351(6268):84-88, incorporated herewith in its entirety by reference). In certain embodiments, the altered or modified activity of the engineered CRISPR protein comprises increased targeting efficiency or decreased off-target binding. In certain embodiments, the altered activity of the engineered CRISPR protein comprises modified cleavage activity. In certain embodiments, the altered activity comprises increased cleavage activity as to the target polynucleotide loci. In certain embodiments, the altered activity comprises decreased cleavage activity as to the target polynucleotide loci. In certain embodiments, the altered activity comprises decreased cleavage activity as to off-target polynucleotide loci. In certain embodiments, the altered or modified activity of the modified nuclease comprises altered helicase kinetics. In certain embodiments, the modified nuclease comprises a modification that alters association of the protein with the nucleic acid molecule comprising RNA (in the case of a Cas protein), or a strand of the target polynucleotide loci, or a strand of off-target polynucleotide loci. In an aspect of the invention, the engineered CRISPR protein comprises a modification that alters formation of the CRISPR complex. In certain embodiments, the altered activity comprises increased cleavage activity as to off-target polynucleotide loci. Accordingly, in certain embodiments, there is increased specificity for target polynucleotide loci as compared to off-target polynucleotide loci. In other embodiments, there is reduced specificity for target polynucleotide loci as compared to off-target polynucleotide loci. In certain embodiments, the mutations result in decreased off-target effects (e.g. cleavage or binding properties, activity, or kinetics), such as in case for Cas proteins for instance resulting in a lower tolerance for mismatches between target and guide RNA. Other mutations may lead to increased off-target effects (e.g. cleavage or binding properties, activity, or kinetics). Other mutations may lead to increased or decreased on-target effects (e.g. cleavage or binding properties, activity, or kinetics). In certain embodiments, the mutations result in altered (e.g. increased or decreased) helicase activity, association or formation of the functional nuclease complex (e.g. CRISPR-Cas complex). In certain embodiments, as described above, the mutations result in an altered PAM recognition, i.e. a

different PAM may be (in addition or in the alternative) be recognized, compared to the unmodified Cas protein. Particularly preferred mutations include positively charged residues and/or (evolutionary) conserved residues, such as conserved positively charged residues, in order to enhance specificity. In certain embodiments, such residues may be mutated to uncharged residues, such as alanine.

[00199] Base Excision Repair Inhibitor

[00200] In some embodiments, the CD-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 a U:G pairing in DNA may be responsible for a decrease in nucleobase editing efficiency in cells. Uracil DNA glycosylase catalyzes removal of uracil from DNA in cells, which may initiate base excision repair, such that the U:G pair is reversed to C:G. In some embodiments, the BER inhibitor is an uracyl glycosylase inhibitor or an active domain thereof.

[00201] In some embodiments, the BER inhibitor is an inhibitor of uracil DNA glycosylase (UDG). In some embodiments, the BER inhibitor is an inhibitor of human UDG. In some embodiments, the BER inhibitor is a polypeptide inhibitor. In some embodiments, the BER inhibitor is a protein that binds single-stranded DNA. For example, the BER inhibitor may be a Erwinia tasmaniensis single-stranded binding protein. In some embodiments, the BER inhibitor is a protein that binds uracil. In some embodiments, the BER inhibitor is a protein that binds uracil in DNA. In some embodiments, the BER inhibitor is a catalytically inactive UDG or binding domain thereof. In some embodiments, the BER inhibitor is a catalytically inactive UDG or binding domain thereof that does not excise uracil from the DNA. Other proteins that are capable of inhibiting (e.g., sterically blocking) UDG 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.

[00202] 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 uracil DNA glycosylase, inhibit base excision repair, protect the edited base, and/or promote fixing of the non-targeted strand. Accordingly, the use of the BER inhibitor described herein can increase the editing efficiency of a cytidine deaminase that is capable of catalyzing a C to U change.

[00203] In particular embodiments, the uracil glycosylase inhibitor (UGI) is the uracil DNA glycosylase inhibitor of *Bacillus subtilis* bacteriophage PBS1 or an active fragment thereof, such as an 83 residue protein of *Bacillus subtilis* bacteriophage PBS1.

[00204] Suitable UGI protein and nucleotide sequences are provided herein and additional suitable UGI sequences are known to those in the art, and include, for example, those published in Wang et al., Uracil-DNA glycosylase inhibitor gene of bacteriophage PBS2 encodes a binding protein specific for uracil-DNA glycosylase. J. Biol. Chem. 264: 1 163-1 171(1989); Lundquist et al., Site-directed mutagenesis and characterization of uracil-DNA glycosylase inhibitor protein. Role of specific carboxylic amino acids in complex formation with Escherichia coli uracil-DNA glycosylase. J. Biol. Chem. 272:21408-21419(1997); Ravishankar et al., X-ray analysis of a complex of Escherichia coli uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG. Nucleic Acids Res. 26:4880-4887(1998); and Putnam et al., Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with Escherichia coli uracil-DNA glycosylase. J. Mol. Biol. 287:331-346(1999), the entire contents of each are incorporated herein by reference.

[00205] In some embodiments, the UGI comprises the following amino acid sequence:

[00206] >sp|P14739|UNGI_BPPB2 Uracil-DNA glycosylase inhibitor

[00207] MTNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDEST DENVMLLTSDAPEYKPWALVIQDSNGENKIKML (SEQ ID NO: 35)

[00208] In some embodiments, the UGI domain comprises a wild-type UGI or a UGI as set forth in SEQ ID NO: 35. In some embodiments, the UGI proteins provided herein include fragments of UGI and proteins homologous to a UGI or a UGI fragment. For example, in some embodiments, a UGI domain comprises a fragment of the amino acid sequence set forth in SEQ ID NO: 35. In some embodiments, a UGI fragment comprises an amino acid sequence that comprises at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or at least 99.5% of the amino acid sequence as set forth in SEQ ID NO: 35. In some embodiments, a UGI comprises an amino acid sequence homologous to the amino acid sequence set forth in SEQ ID NO: 35 or an amino acid sequence homologous to a fragment of the amino acid sequence set forth in SEQ ID NO: 35. In some embodiments, proteins comprising UGI or fragments of UGI or homologs of UGI or UGI fragments are referred to as "UGI variants." A UGI variant shares homology to UGI, or a fragment

thereof. For example a UGI variant is at least 70% identical, at least 75% identical, at least 80%, identical, at least 85%> identical, at least 90% identical, at least 95% identical, at least 96%, identical, at least 97% identical, at least 98% identical, at least 99% identical, at least *99.5%* identical, or at least 99.9% identical to a wild type UGI or a UGI as set forth in SEQ ID NO: 35. In some embodiments, the UGI variant comprises a fragment of UGI, such that the fragment is at least 70% identical, at least 80% identical, at least 90% identical, at least 95%, identical, at least 96%) identical, at least 97% identical, at least 98% identical, at least 99%, identical, at least 99.5%) identical, or at least 99.9% to the corresponding fragment of wild-type UGI or a UGI as set forth in SEQ ID NO: 35.

[00209] It should be appreciated that additional proteins may be uracil glycosylase inhibitors. For example, other proteins that are capable of inhibiting (e.g., sterically blocking) a uracil-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. In some embodiments, a protein that binds DNA is used. In another embodiment, a substitute for UGI is used. In some embodiments, a uracil glycosylase inhibitor is a protein that binds single-stranded DNA. For example, a uracil glycosylase inhibitor may be a Erwinia tasmaniensis single-stranded binding protein. In some embodiments, the single-stranded binding protein comprises the amino acid sequence (SEQ ID NO: 78). In some embodiments, a uracil glycosylase inhibitor is a protein that binds uracil. In some embodiments, a uracil glycosylase inhibitor is a protein that binds uracil in DNA. In some embodiments, a uracil glycosylase inhibitor is a catalytically inactive uracil DNA-glycosylase protein. In some embodiments, a uracil glycosylase inhibitor is a catalytically inactive uracil DNA-glycosylase protein that does not excise uracil from the DNA. For example, a uracil glycosylase inhibitor is a UdgX. In some embodiments, the UdgX comprises the amino acid sequence (SEQ ID NO: 79). As another example, a uracil glycosylase inhibitor is a catalytically inactive UDG. In some embodiments, a catalytically inactive UDG comprises the amino acid sequence (SEQ ID NO: 80). It should be appreciated that other uracil glycosylase inhibitors would be apparent to the skilled artisan and are within the scope of this disclosure. In some embodiments, a uracil glycosylase inhibitor is a protein that is homologous to any one of SEQ ID NOs: 78-80. In some embodiments, a uracil glycosylase inhibitor is a protein that is at least 50% identical, at least 55% identical at least 60% identical, at least 65% identical, at least 70% identical, at least 75% identical, at least 80% identical at least 85% identical, at least 90%

identical, at least 95% identical, at least 96% identical, at least 98%> identical, at least 99% identical, or at least 99.5% identical to any one of SEQ ID NOs: 78-80.

[00210] Erwinia tasmaniensis SSB (themostable single-stranded DNA binding protein): MASRGVNKVILVGNLGQDPEVRYMPNGGAVANITLATSESWRDKQTGETK EKTEWHRVVLFGKLAEVAGEYLRKGSQVYIEGALQTRKWTDQAGVEKYTT EVVVNVGGTMQMLGGRSQGGGASAGGQNGGSNNGWGQPQQPQGGNQFSGG AQQQARPQQQPQQNNAPANNEPPIDFDDDIP (SEQ ID NO: 78)

[00211] UdgX (binds to Uracil in DNA but does not excise): MAGAQDFVPHTADLAELAAAAGECRGCGLYRDATQAVFGAGGRSARIMMI GEQPGDKEDLAGLPFVGPAGRLLDRALEAADIDRDALYVTNAVKHFKFTR AAGGKRRIUKTP SRTEVVACRPWLIAEMT SVEPDVVVLLGATAAKALLGN DFRVTQHRGEVLHVDDVPGDPALVATVHPSSLLRGPKEERESAFAGLVDD LRVAADVRP (SEQ ID NO: 79)

[00212] UDG (catalytically inactive human UDG, binds to Uracil in DNA but does not excise): MIGQKTLYSFFSPSPARKRHAPSPEPAVQGTGVAGVPEESGDAAAIPAK KAP AGQEEPGTPP SSPL SAEQLDRIQRNK AAALLRL AARNVPVGFGESW KKHLSGEFGKPYFIKLMGFVAEERKHYTVYPPPHQVFTWTQMCDIKDVK VVILGQEP YHGPNQAHGLCF SVQRPVPPPP SLENIYKEL STDIEDFVHP GHGDLSGWAKQGVLLLNAVLTVRAHQANSHKERGWEQFTDAVVSWLNQN SNGLVFLLWGSYAQKKGSAIDRKRHHVLQTAHPSPLSVYRGFFGCRHFS KTNELLQKSGKKPIDWKEL (SEQ ID NO: 80)

[00213] Accordingly, in the first design of the CD-functionalized CRISPR system discussed above, the CRISPR-Cas protein or the cytidine deaminase can be fused to or linked to a BER inhibitor (e.g., an inhibitor of uracyl DNA glycosylase). In some embodiments, the BER inhibitor can be comprised in one of the following structures (nCpfl=Cpfl nickase; dCpfl=dead Cpfl):

[CD]-[optional Hnker]-[nCpfl/dCpfl]-[optional linker]-[BER inhibitor];
[CD]-[optional linker]-[BER inhibitor]-[optional linker]-[nCpfl/dCpfl];
[BER inhibitor]-[optional linker]-[CD]-[optional linker]-[nCpfl/dCpfl];
[BER inhibitor]-[optional Hnker]-[nCpfl/dCpfl]-[optional linker]-[CD];
[nCpfl/dCpfl]-[optional linker]-[CD]-[optional linker]-[BER inhibitor];
[nCpfl/dCpfl]-[optional linker]-[BER inhibitor]-[optional linker]-[CD].

[00214] Similarly, in the second design of the CD-functionalized CRISPR system discussed above, the CRISPR-Cas protein, the cytidine deaminase, or the adaptor protein can be fused to or linked to a BER inhibitor (e.g., an inhibitor of uracil DNA glycosylase). In some embodiments, the BER inhibitor can be comprised in one of the following structures (nCpfl=Cpfl nickase; dCpfl=dead Cpfl):

[nCpfl/dCpfl]-[optional linker]-[BER inhibitor];

[BER inhibitor]-[optional Hnker]-[nCpfl/dCpfl];

[CD]-[optional linker]-[Adaptor]-[optional linker]-[BER inhibitor];

[CD]-[optional linker]-[BER inhibitor]-[optional linker]-[Adaptor];

[BER inhibitor]-[optional linker]-[CD]-[optional linker]-[Adaptor];

[BER inhibitor]-[optional linker]-[Adaptor]-[optional linker]-[CD];

[Adaptor]-[optional linker]-[CD]-[optional linker]-[BER inhibitor];

[Adaptor]-[optional linker]-[BER inhibitor]-[optional linker]-[CD].

[00215] In the third design of the CD-functionalized CRISPR system discussed above, the BER inhibitor can be inserted into an internal loop or unstructured region of a CRISPR-Cas protein.

[00216] Targeting to the Nucleus

[00217] In some embodiments, the methods of the present invention relate to modifying an Cytosine in a target locus of interest, whereby the target locus is within a cell. In order to improve targeting of the CRISPR-Cas protein and/or the cytidine deaminase protein or catalytic domain thereof used in the methods of the present invention to the nucleus, it may be advantageous to provide one or both of these components with one or more nuclear localization sequences (NLSs).

[00218] In preferred embodiments, the NLSs used in the context of the present invention are heterologous to the proteins. Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV (SEQ ID NO: 37) or PKKKRKVEAS (SEQ ID NO: 38); the NLS from nucleoplasmin (e.g., the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK (SEQ ID NO: 39)); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID NO: 40) or RQRRNELKRSP (SEQ ID NO: 41); the hRNPAl M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO: 42); the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID

NO: 43) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID NO: 44) and PPKKARED (SEQ ID NO: 45) of the myoma T protein; the sequence PQPKKKPL (SEQ ID NO: 46) of human p53; the sequence SALIKKKKKMAP (SEQ ID NO: 47) of mouse c-abl IV; the sequences DRLRR (SEQ ID NO: 48) and PKQKKRK (SEQ ID NO: 49) of the influenza virus NS1; the sequence RKLKKKIKKL (SEQ ID NO: 50) of the Hepatitis virus delta antigen; the sequence REKKKFLKRR (SEQ ID NO: 51) of the mouse Mxl protein; the sequence KRKGDEVDGVDEVAKKKSKK (SEQ ID NO: 52) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK (SEQ ID NO: 53) of the steroid hormone receptors (human) glucocorticoid. In general, the one or more NLSs are of sufficient strength to drive accumulation of the DNA-targeting Cas protein in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the CRISPR-Cas protein, the particular NLS(s) used, or a combination of these factors. Detection of accumulation in the nucleus may be performed by any suitable technique. For example, a detectable marker may be fused to the nucleic acid-targeting protein, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g., a stain specific for the nucleus such as DAPI). Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay for the effect of nucleic acid-targeting complex formation (e.g., assay for deaminase activity) at the target sequence, or assay for altered gene expression activity affected by DNA-targeting complex formation and/or DNA-targeting), as compared to a control not exposed to the CRISPR-Cas protein and deaminase protein, or exposed to a CRISPR-Cas and/or deaminase protein lacking the one or more NLSs.

[00219] The CRISPR-Cas and/or cytidine deaminase proteins may be provided with 1 or more, such as with, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more heterologous NLSs. In some embodiments, the proteins comprises about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the amino-terminus, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs at or near the carboxy-terminus, or a combination of these (e.g., zero or at least one or more NLS at the amino-terminus and zero or at one or more NLS at the carboxy terminus). When more than one NLS is present, each may be selected independently of the others, such that a single NLS may be present in more than one copy

and/or in combination with one or more other NLSs present in one or more copies. In some embodiments, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus. In preferred embodiments of the CRISPR-Cas proteins, an NLS attached to the C-terminal of the protein.

[00220] In certain embodiments of the methods provided herein, the CRISPR-Cas protein and the deaminase protein are delivered to the cell or expressed within the cell as separate proteins. In these embodiments, each of the CRISPR-Cas and deaminase protein can be provided with one or more NLSs as described herein. In certain embodiments, the CRISPR-Cas and deaminase proteins are delivered to the cell or expressed with the cell as a fusion protein. In these embodiments one or both of the CRISPR-Cas and deaminase protein is provided with one or more NLSs. Where the cytidine deaminase is fused to an adaptor protein (such as MS2) as described above, the one or more NLS can be provided on the adaptor protein, provided that this does not interfere with aptamer binding. In particular embodiments, the one or more NLS sequences may also function as linker sequences between the cytidine deaminase and the CRISPR-Cas protein.

[00221] In certain embodiments, guides of the invention comprise specific binding sites (e.g. aptamers) for adapter proteins, which may be linked to or fused to a cytidine deaminase or catalytic domain thereof. When such a guides forms a CRISPR complex (i.e. CRISPR-Cas protein binding to guide and target) the adapter proteins bind and, the cytidine deaminase or catalytic domain thereof associated with the adapter protein is positioned in a spatial orientation which is advantageous for the attributed function to be effective.

[00222] The skilled person will understand that modifications to the guide which allow for binding of the adapter + cytidine deaminase, but not proper positioning of the adapter + cytidine deaminase (e.g. due to steric hindrance within the three dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified guide may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein, preferably at either the tetra loop or stem loop 2, and most preferably at both the tetra loop and stem loop 2.

[00223] Use of orthogonal catalytically inactive CRISPR-Cas proteins

[00224] In particular embodiments, the Cpfl nickase is used in combination with an orthogonal catalytically inactive CRISPR-Cas protein to increase efficiency of said Cpfl nickase (as described in Chen et al. 2017, Nature Communications 8:14958;

doi:10.1038/ncommsl4958). More particularly, the orthogonal catalytically inactive CRISPR-Cas protein is characterized by a different PAM recognition site than the Cpfl nickase used in the CD-functionalized CRISPR system and the corresponding guide sequence is selected to bind to a target sequence proximal to that of the Cpfl nickase of the CD-functionalized CRISPR system. The orthogonal catalytically inactive CRISPR-Cas protein as used in the context of the present invention does not form part of the CD-functionalized CRISPR system but merely functions to increase the efficiency of said Cpfl nickase and is used in combination with a standard guide molecule as described in the art for said CRISPR-Cas protein. In particular embodiments, said orthogonal catalytically inactive CRISPR-Cas protein is a dead CRISPR-Cas protein, i.e. comprising one or more mutations which abolishes the nuclease activity of said CRISPR-Cas protein. In particular embodiments, the catalytically inactive orthogonal CRISPR-Cas protein is provided with two or more guide molecules which are capable of hybridizing to target sequences which are proximal to the target sequence of the Cpfl nickase. In particular embodiments, at least two guide molecules are used to target said catalytically inactive CRISPR-Cas protein, of which at least one guide molecule is capable of hybridizing to a target sequence 5" of the target sequence of the Cpfl nickase and at least one guide molecule is capable of hybridizing to a target sequence 3' of the target sequence of the Cpfl nickase of the CD-functionalized CRISPR system, whereby said one or more target sequences may be on the same or the opposite DNA strand as the target sequence of the Cpfl nickase. In particular embodiments, the guide sequences for the one or more guide molecules of the orthogonal catalytically inactive CRISPR-Cas protein are selected such that the target sequences are proximal to that of the guide molecule for the targeting of the CD-functionalized CRISPR, i.e. for the targeting of the Cpfl nickase. In particular embodiments, the one or more target sequences of the orthogonal catalytically inactive CRISPR-Cas enzyme are each separated from the target sequence of the Cpfl nickase by more than 5 but less than 450 basepairs. Optimal distances between the target sequences of the guides for use with the orthogonal catalytically inactive CRISPR-Cas protein and the target sequence of the CD-functionalized CRISPR system can be determined by the skilled person. In particular embodiments, the orthogonal CRISPR-Cas protein is a Class II, type II CRISPR protein. In particular embodiments, the orthogonal CRISPR-Cas protein is a Class II, type V CRISPR protein. In particular embodiments, the catalytically inactive orthogonal CRISPR-Cas protein In particular embodiments, the catalytically inactive orthogonal CRISPR-Cas protein has been

modified to alter its PAM specificity as described elsewhere herein. In particular embodiments, the Cpfl protein nickase is a nickase which, by itself has limited activity in human cells, but which, in combination with an inactive orthogonal CRISPR-Cas protein and one or more corresponding proximal guides ensures the required nickase activity.

[00225] CRISPR Development and Use

[00226] The present invention may be further illustrated and extended based on aspects of CRISPR-Cas 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:

- Multiplex genome engineering using CRISPR-Cas systems. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, F. Science Feb 15;339(6121):819-23 (2013);
- RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Jiang W., Bikard D., Cox D., Zhang F, Marraffini LA. Nat Biotechnol Mar;31(3):233-9 (2013);
- One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR-Cas-Mediated Genome Engineering. Wang H., Yang H., Shivalila CS., Dawlaty MM., Cheng AW., Zhang F., Jaenisch R. Cell May 9;153(4):910-8 (2013);
- Optical control of mammalian endogenous transcription and epigenetic states. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Piatt RJ, Scott DA, Church GM, Zhang F. Nature. Aug 22;500(7463):472-6. doi: 10.1038/Naturel2466. 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(l l):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);
- Crystal structure of cas9 in complex with guide RNA and target DNA. Nishimasu, H., Ran, FA., Hsu, PD., Konermann, S., Shehata, SI., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. Cell Feb 27, 156(5):935-49 (2014);
- Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Wu X., Scott DA., Kriz AJ., Chiu AC, Hsu PD., Dadon DB., Cheng AW., Trevino AE., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp PA. Nat Biotechnol. Apr 20. doi: 10.1038/nbt.2889 (2014);
- CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Piatt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, Graham DB, 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(l): 102-6 (2015);
- Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F., Nature. Jan 29;517(7536):583-8 (2015).

- A split-Cas9 architecture for inducible genome editing and transcription modulation, Zetsche B, Volz SE, Zhang F., (published online 02 February 2015) Nat Biotechnol. Feb;33(2): 139-42 (2015);
- Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis, Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R, Lee H, Zhang F, Sharp PA. Cell 160, 1246-1260, March 12, 2015 (multiplex screen in mouse), and
- In *vivo* genome editing using *Staphylococcus aureus* Cas9, Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F., (published online 01 April 2015), Nature. Apr 9;520(7546): 186-91 (2015).
- Shalem et al., "High-throughput functional genomics using CRISPR-Cas9," Nature Reviews Genetics 16, 299-31 1 (May 2015).
- Xu et al., "Sequence determinants of improved CRISPR sgRNA design," Genome Research 25, 1147-1 157 (August 2015).
- Parnas et al., "A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks," Cell 162, 675-686 (July 30, 2015).
- Ramanan et al., CRISPR-Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus," Scientific Reports 5:10833. doi: 10.1038/srepl0833 (June 2, 2015)
- Nishimasu et al., Crystal Structure of Staphylococcus aureus Cas9," Cell 162, 1113-1126 (Aug. 27, 2015)
- BCL1 1A enhancer dissection by Cas9-mediated in situ saturating mutagenesis, Canver et al., Nature 527(7577): 192-7 (Nov. 12, 2015) doi: 10.1038/naturel5521. Epub 2015 Sep 16.
- *Cpfl 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: 10.1 126/science.aad5227. Epub 2015 Dec 1.

- Gao *et al,* "Engineered Cpfl Enzymes with Altered PAM Specificities," bioRxiv 09161 1; doi: http://dx.doi.org/10.1 101/09161 1 (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:

- Cong *et al.* engineered type II CRISPR-Cas systems for use in eukaryotic cells based on both *Streptococcus thermophilics* Cas9 and also *Streptococcus pyogenes* Cas9 and demonstrated that Cas9 nucleases can be directed by short RNAs to induce precise cleavage of DNA in human and mouse cells. Their study further showed that Cas9 as converted into a nicking enzyme can be used to facilitate homology-directed repair in eukaryotic cells with minimal mutagenic activity. Additionally, their study demonstrated that multiple guide sequences can be encoded into a single CRISPR array to enable simultaneous editing of several at endogenous genomic loci sites within the mammalian genome, demonstrating easy programmability and wide applicability of the RNA-guided nuclease technology. This ability to use RNA to program sequence specific DNA cleavage in cells defined a new class of genome engineering tools. These studies further showed that other CRISPR loci are likely to be transplantable into mammalian cells and can also mediate mammalian genome cleavage. Importantly, it can be envisaged that several aspects of the CRISPR-Cas system can be further improved to increase its efficiency and versatility.
- Jiang *et al.* used the clustered, regularly interspaced, short palindromic repeats (CRISPR)-associated Cas9 endonuclease complexed with dual-RNAs to introduce precise mutations in the genomes of *Streptococcus pneumoniae* and *Escherichia coli.* The approach relied on dual-RNA:Cas9-directed cleavage at the targeted genomic site to kill unmutated cells and circumvents the need for selectable markers or counter-selection systems. The study reported reprogramming dual-RNA:Cas9 specificity by changing the sequence of short CRISPR RNA (crRNA) to make single- and multinucleotide changes carried on editing templates. The study showed that simultaneous use of two crRNAs enabled multiplex mutagenesis. Furthermore, when the approach was used in combination with recombineering, in *S. pneumoniae,* nearly 100% of cells that were recovered using the described approach contained the desired mutation, and in *E. coli,* 65% that were recovered contained the mutation.
- Wang *et al.* (2013) used the CRISPR-Cas system for the one-step generation of mice carrying mutations in multiple genes which were traditionally generated in multiple

steps by sequential recombination in embryonic stem cells and/or time-consuming intercrossing of mice with a single mutation. The CRISPR-Cas system will greatly accelerate the *in vivo* study of functionally redundant genes and of epistatic gene interactions.

- Konermann *et al.* (2013) addressed the need in the art for versatile and robust technologies that enable optical and chemical modulation of DNA-binding domains based CRISPR Cas9 enzyme and also Transcriptional Activator Like Effectors
- Ran *et al.* (2013-A) described an approach that combined a Cas9 nickase mutant with paired guide RNAs to introduce targeted double-strand breaks. This addresses the issue of the Cas9 nuclease from the microbial CRISPR-Cas system being targeted to specific genomic loci by a guide sequence, which can tolerate certain mismatches to the DNA target and thereby promote undesired off-target mutagenesis. Because individual nicks in the genome are repaired with high fidelity, simultaneous nicking *via* appropriately offset guide RNAs is required for double-stranded breaks and extends the number of specifically recognized bases for target cleavage. The authors demonstrated that using paired nicking can reduce off-target activity by 50- to 1,500-fold in cell lines and to facilitate gene knockout in mouse zygotes without sacrificing on-target cleavage efficiency. This versatile strategy enables a wide variety of genome editing applications that require high specificity.
- Hsu *et al.* (2013) characterized SpCas9 targeting specificity in human cells to inform the selection of target sites and avoid off-target effects. The study evaluated >700 guide RNA variants and SpCas9-induced indel mutation levels at >100 predicted genomic off-target loci in 293T and 293FT cells. The authors that SpCas9 tolerates mismatches between guide RNA and target DNA at different positions in a sequence-dependent manner, sensitive to the number, position and distribution of mismatches. The authors further showed that SpCas9-mediated cleavage is unaffected by DNA methylation and that the dosage of SpCas9 and guide RNA can be titrated to minimize off-target modification. Additionally, to facilitate mammalian genome engineering applications, the authors reported providing a web-based software tool to guide the selection and validation of target sequences as well as off-target analyses.
- Ran *et al.* (2013-B) described a set of tools for Cas9-mediated genome editing *via* non-homologous end joining (NHEJ) or homology-directed repair (HDR) in

mammalian cells, as well as generation of modified cell lines for downstream functional studies. To minimize off-target cleavage, the authors further described a double-nicking strategy using the Cas9 nickase mutant with paired guide RNAs. The protocol provided by the authors experimentally derived guidelines for the selection of target sites, evaluation of cleavage efficiency and analysis of off-target activity. The studies showed that beginning with target design, gene modifications can be achieved within as little as 1-2 weeks, and modified clonal cell lines can be derived within 2-3 weeks.

- Shalem *et al.* described a new way to interrogate gene function on a genome-wide scale. Their studies showed that delivery of a genome-scale CRISPR-Cas9 knockout (GeCKO) library targeted 18,080 genes with 64,751 unique guide sequences enabled both negative and positive selection screening in human cells. First, the authors showed use of the GeCKO library to identify genes essential for cell viability in cancer and pluripotent stem cells. Next, in a melanoma model, the authors screened for genes whose loss is involved in resistance to vemurafenib, a therapeutic that inhibits mutant protein kinase BRAF. Their studies showed that the highest-ranking candidates included previously validated genes NF1 and MED12 as well as novel hits NF2, CUL3, TADA2B, and TADAl. The authors observed a high level of consistency between independent guide RNAs targeting the same gene and a high rate of hit confirmation, and thus demonstrated the promise of genome-scale screening with Cas9.
- Nishimasu *et al.* reported the crystal structure of *Streptococcus pyogenes* Cas9 in complex with sgRNA and its target DNA at 2.5 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.

- 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.
- Piatt *et al.* established a Cre-dependent Cas9 knockin mouse. The authors demonstrated *in vivo* as well as *ex vivo* genome editing using adeno-associated virus (AAV)-, lentivirus-, or particle-mediated delivery of guide RNA in neurons, immune cells, and endothelial cells.
- Hsu *et al.* (2014) is a review article that discusses generally CRISPR-Cas9 history from yogurt to genome editing, including genetic screening of cells.
- Wang *et al.* (2014) relates to a pooled, loss-of-function genetic screening approach suitable for both positive and negative selection that uses a genome-scale lentiviral single guide RNA (sgRNA) library.
- Doench *et al.* created a pool of sgRNAs, tiling across all possible target sites of a panel of six endogenous mouse and three endogenous human genes and quantitatively assessed their ability to produce null alleles of their target gene by antibody staining and flow cytometry. The authors showed that optimization of the PAM improved activity and also provided an on-line tool for designing sgRNAs.
- Swiech *et al.* demonstrate that AAV-mediated SpCas9 genome editing can enable reverse genetic studies of gene function in the brain.
- Konermann *et al.* (2015) discusses the ability to attach multiple effector domains, e.g., transcriptional activator, functional and epigenomic regulators at appropriate positions on the guide such as stem or tetraloop with and without linkers.

- Zetsche *et al.* demonstrates that the Cas9 enzyme can be split into two and hence the assembly of Cas9 for activation can be controlled.
- Chen *et al.* relates to multiplex screening by demonstrating that a genome-wide *in vivo* CRISPR-Cas9 screen in mice reveals genes regulating lung metastasis.
- Ran *et al.* (2015) relates to SaCas9 and its ability to edit genomes and demonstrates that one cannot extrapolate from biochemical assays.
- Shalem *et al.* (2015) described ways in which catalytically inactive Cas9 (dCas9) fusions are used to synthetically repress (CRISPRi) or activate (CRISPRa) expression, showing. advances using Cas9 for genome-scale screens, including arrayed and pooled screens, knockout approaches that inactivate genomic loci and strategies that modulate transcriptional activity.
- Xu *et al.* (2015) assessed the DNA sequence features that contribute to single guide RNA (sgRNA) efficiency in CRISPR-based screens. The authors explored efficiency of CRISPR-Cas9 knockout and nucleotide preference at the cleavage site. The authors also found that the sequence preference for CRISPRi/a is substantially different from that for CRISPR-Cas9 knockout.
- Parnas *et al.* (2015) introduced genome-wide pooled CRISPR-Cas9 libraries into dendritic cells (DCs) to identify genes that control the induction of tumor necrosis factor (Tnf) by bacterial lipopolysaccharide (LPS). Known regulators of Tlr4 signaling and previously unknown candidates were identified and classified into three functional modules with distinct effects on the canonical responses to LPS.
- Ramanan *et al* (2015) demonstrated cleavage of viral episomal DNA (cccDNA) in infected cells. The HBV genome exists in the nuclei of infected hepatocytes as a 3.2kb double-stranded episomal DNA species called covalently closed circular DNA (cccDNA), which is a key component in the HBV life cycle whose replication is not inhibited by current therapies. The authors showed that sgRNAs specifically targeting highly conserved regions of HBV robustly suppresses viral replication and depleted cccDNA.
- Nishimasu *et al.* (2015) reported the crystal structures of SaCas9 in complex with a single guide RNA (sgRNA) and its double-stranded DNA targets, containing the 5'-TTGAAT-3' PAM and the 5'-TTGGGT-3' PAM. A structural comparison of SaCas9 with SpCas9 highlighted both structural conservation and divergence, explaining their distinct PAM specificities and orthologous sgRNA recognition.

- Canver *et al.* (2015) demonstrated a CRISPR-Cas9-based functional investigation of non-coding genomic elements. The authors we developed pooled CRISPR-Cas9 guide RNA libraries to perform *in situ* saturating mutagenesis of the human and mouse BCL1 1A enhancers which revealed critical features of the enhancers.
- Zetsche et al. (2015) reported characterization of Cpfl, a class 2 CRISPR nuclease from Francisella novicida U l 12 having features distinct from Cas9. Cpfl 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.
- Shmakov et al. (2015) reported three distinct Class 2 CRISPR-Cas systems. Two system CRISPR enzymes (C2cl and C2c3) contain RuvC-like endonuclease domains distantly related to Cpfl. Unlike Cpfl, C2cl depends on both crRNA and tracrRNA for DNA cleavage. The third enzyme (C2c2) contains two predicted HEPN RNase domains and is tracrRNA independent.
- 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.

[00227] The methods and tools provided herein are exemplified for Cpfl, a type II nuclease that does not make use of tracrRNA. Orthologs of Cpfl have been identified in different bacterial species as described herein. Further type II 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 Casl. In further embodiments, the CRISPR array is used as a seed to identify new effector proteins.

[00228] The effectiveness of the present invention has been demonstrated. Preassembled recombinant CRISPR-Cpfl complexes comprising Cpfl 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 Cpfl ribonucleoproteins, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3596. Genome-wide analyses shows that Cpfl is highly specific. By one measure, *in vitro* cleavage sites determined for Cpfl in human HEK293T cells were significantly fewer that for SpCas9. Kim, D. et al., Genome-wide analysis reveals specificities of Cpfl endonucleases in human cells, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3609. An efficient multiplexed system employing Cpfl 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 Cpfl gRNAs. doi: http://dx.doi.org/10.1 101/046417.

[00229] Also, "Dimeric CRISPR RNA-guided Fokl 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 Fokl Nucleases that recognize extended sequences and can edit endogenous genes with high efficiencies in human cells.

[00230] With respect to general information on CRISPR/Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, and making and using thereof, including as to amounts and formulations, as well as CRISPR-Cas-expressing eukaryotic cells, CRISPR-Cas expressing eukaryotes, such as a mouse, reference is made to: US Patents Nos. 8,999,641, 8,993,233, 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, and 8,945,839; US Patent Publications US 2014-0310830 (US App. Ser. No. 14/105,031), US 2014-0287938 A1 (U.S. App. Ser. No. 14/213,991), US 2014-0273234 A1 (U.S. App. Ser. No. 14/293,674), US2014-0273232 A1 (U.S. App. Ser. No. 14/290,575), US 2014-0273231 (U.S. App. Ser. No. 14/259,420), US 2014-0256046 A1 (U.S. App. Ser. No. 14/226,274), US 2014-0248702 A1 (U.S. App. Ser. No. 14/258,458), US 2014-0242700 A1 (U.S. App. Ser. No. 14/222,930), US 2014-0242699 A1 (U.S. App. Ser. No. 14/183,512), US 2014-0242664 A1 (U.S. App. Ser. No. 14/104,990), US 2014-0234972 A1 (U.S. App. Ser. No. 14/183,471), US 2014-0227787 A1 (U.S. App. Ser. No. 14/256,912), US 2014-0189896 A1 (U.S. App. Ser. No. 14/105,035), US 2014-0186958 (U.S. App. Ser. No. 14/105,017), US 2014-0186919 A1 (U.S. App. Ser. No. 14/104,977), US 2014-0186843 A1 (U.S. App. Ser. No. 14/104,900), US 2014-0179770 A1 (U.S. App. Ser. No.

14/104,837) and US 2014-0179006 A1 (U.S. App. Ser. No. 14/183,486), US 2014-0170753 (US App Ser No 14/183,429); US 2015-0184139 (U.S. App. Ser. No. 14/324,960); 14/054,414 European Patent Applications EP 2 771 468 (EP13818570.7), EP 2 764 103 (EP13824232.6), and EP 2 784 162 (EP14170383.5); and PCT Patent Publications WO2014/093661 (PCT/US2013/074743), WO2014/093694 (PCT/US2013/074790), WO2014/093595 (PCT/US2013/074611), WO2014/093718 (PCT/US2013/074825), WO2014/093709 (PCT/US2013/074812), WO2014/093622 (PCT/US2013/074667), WO2014/093635 (PCT/US2013/074691), WO2014/093655 (PCT/US2013/074736), WO2014/093712 (PCT/US2013/074819), WO2014/093701 (PCT/US2013/074800), WO2014/018423 (PCT/US2013/051418), WO2014/204723 (PCT/US2014/041790), WO2014/204724 (PCT/US2014/041800), WO2014/204725 (PCT/US2014/041803), WO2014/204726 (PCT/US2014/041804), WO2014/204727 (PCT/US2014/041806), WO2014/204728 (PCT/US2014/041808), WO2014/204729 (PCT/US2014/041809), WO2015/089351 (PCT/US2014/069897), WO2015/089354 (PCT/US2014/069902), WO2015/089364 (PCT/US2014/069925), WO2015/089427 (PCT/US2014/070068), WO2015/089462 (PCT/US2014/070127), WO2015/089419 (PCT/US2014/070057), WO2015/089465 (PCT/US2014/070135), WO2015/089486 (PCT/US2014/070175), WO2015/058052 (PCT/US2014/061077), WO2015/070083 (PCT/US2014/064663), WO2015/089354 (PCT/US2014/069902), WO2015/089351 (PCT/US2014/069897), WO2015/089364 (PCT/US2014/069925), WO2015/089427 (PCT/US2014/070068), WO2015/089473 (PCT/US2014/070152), WO2015/089486 (PCT/US2014/070175), WO2016/049258 (PCT/US2015/051830), WO2016/094867 (PCT/US2015/065385), WO2016/094872 (PCT/US2015/065393), WO2016/094874 (PCT/US2015/065396), WO2016/106244 (PCT/US2015/067177).

[00231] Mention is also made of US application 62/180,709, 17-Jun-15, PROTECTED GUIDE RNAS (PGRNAS); US application 62/091,455, filed, 12-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US application 62/096,708, 24-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US applications 62/091,462, 12-Dec-14, 62/096,324, 23-Dec-14, 62/180,681, 17-Jun-2015, and 62/237,496, 5-Oct-2015, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US application 62/091,456, 12-Dec-14 and 62/180,692, 17-Jun-2015, ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS; US application 62/091,461, 12-Dec-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND

COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs); US application 62/094,903, 19-Dec-14, UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE SEQUENCING; US application 62/096,761, 24-Dec-14, ENGINEERING OF SYSTEMS, METHODS AND OPTFMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION; US application 62/098,059, 30-Dec-14, 62/181,641, 18-Jun-2015, and 62/181,667, 18-Jun-2015, RNA-TARGETING SYSTEM; US application 62/096,656, 24-Dec-14 and 62/181,151, 17-Jun-2015, CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS; US application 62/096,697, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH AAV; US application 62/098,158, 30-Dec-14, ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS; US application 62/151,052, 22-Apr-15, CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING; US application 62/054,490, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS; US application 61/939,154, 12-F EB-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,484, 25-Sep-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTFMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,537, 4-Dec-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/054,651, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US application 62/067,886, 23-Oct-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US applications 62/054,675, 24-Sep-14 and 62/181,002, 17-Jun-2015, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES; US application 62/054,528, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS

SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS; US application 62/055,454, 25-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP); US application 62/055,460, 25-Sep-14, MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; US application 62/087,475, 4-Dec-14 and 62/181,690, 18-Jun-2015, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,487, 25-Sep-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,546, 4-Dec-14 and 62/181,687, 18-Jun-2015, MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; and US application 62/098,285, 30-Dec-14, CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS.

[00232] Mention is made of US applications 62/181,659, 18-Jun-2015 and 62/207,318, 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 applications 62/181,663, 18-Jun-2015 and 62/245,264, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS, US applications 62/181,675, 18-Jun-2015, 62/285,349, 22-Oct-2015, 62/296,522, 17-Feb-2016, and 62/320,231, 8-Apr-2016, NOVEL CRISPR ENZYMES AND SYSTEMS, US application 62/232,067, 24-Sep-2015, US Application 14/975,085, 18-Dec-2015, European application No. 16150428.7, US application 62/205,733, 16-Aug-2015, US application 62/201,542, 5-Aug-2015, US application 62/193,507, 16-M-2015, and US application 62/181,739, 18-Jun-2015, each entitled NOVEL CRISPR ENZYMES AND SYSTEMS and of US application 62/245,270, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS. Mention is also made of US application 61/939,256, 12-Feb-2014, and WO 2015/089473 (PCT/US20 14/070 152), 12-Dec-2014, each entitled ENGINEERING OF SYSTEMS, METHODS AND OPTFMIZED GUIDE COMPOSITIONS WITH NEW ARCHITECTURES FOR SEQUENCE MANIPULATION. Mention is also made of PCT/US2015/045504, 15-Aug-2015, US application 62/180,699, 17-Jun-2015, and US application 62/038,358, 17-Aug-2014, each entitled GENOME EDITING USING CAS9 NICKASES.

[00233] 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.

[00234] **Type-V CRISPR-Cas Protein**

[00235] The application describes methods using Type-V CRISPR-Cas proteins. This is exemplified herein with Cpfl, whereby a number of orthologs or homologs have been identified. It will be apparent to the skilled person that further orthologs or homologs can be identified and that any of the functionalities described herein may be engineered into other orthologs, incuding chimeric enzymes comprising fragments from multiple orthologs.

[00236] Computational methods of identifying novel CRISPR-Cas loci are described in EP300951 1 or US2016208243 and may comprise the following steps: detecting all contigs encoding the Casl protein; identifying all predicted protein coding genes within 20kB of the casl gene; comparing the identified genes with Cas protein-specific profiles and predicting CRISPR arrays; selecting unclassified candidate CRISPR-Cas loci containing proteins larger than 500 amino acids (>500 aa); analyzing selected candidates using methods such as PSI-BLAST and HHPred to screen for known protein domains, thereby identifying novel Class 2 CRISPR-Cas loci (see also Schmakov et al. 2015, Mol Cell. 60(3):385-97). In addition to the above mentioned steps, additional analysis of the candidates may be conducted by searching metagenomics databases for additional homologs. Additionally or alternatively, to expand the search to non-autonomous CRISPR-Cas systems, the same procedure can be performed with the CRISPR array used as the seed.

[00237] In one aspect the detecting all contigs encoding the Casl protein is performed by GenemarkS which a gene prediction program as further described in "GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions." John Besemer, Alexandre Lomsadze and Mark Borodovsky, Nucleic Acids Research (2001) 29, pp 2607-2618, herein incorporated by reference.

[00238] In one aspect the identifying all predicted protein coding genes is carried out by comparing the identified genes with Cas protein-specific profiles and annotating them according to NCBI Conserved Domain Database (CDD) which is a protein annotation resource that consists of a collection of well-annotated multiple sequence alignment models for ancient domains and full-length proteins. These are available as position-specific score matrices (PSSMs) for fast identification of conserved domains in protein sequences via RPS-BLAST. CDD content includes NCBI-curated domains, which use 3D-structure information to explicitly define domain boundaries and provide insights into sequence/structure/function relationships, as well as domain models imported from a number of external source databases (Pfam, SMART, COG, PRK, TIGRFAM). In a further aspect, CRISPR arrays were predicted using a PILER-CR program which is a public domain software for finding CRISPR repeats as described in "PILER-CR: fast and accurate identification of CRISPR repeats", Edgar, R.C., BMC Bioinformatics, Jan 20;8: 18(2007), herein incorporated by reference.

[00239] In a further aspect, the case by case analysis is performed using PSI-BLAST (Position-Specific Iterative Basic Local Alignment Search Tool). PSI-BLAST derives a position-specific scoring matrix (PSSM) or profile from the multiple sequence alignment of sequences detected above a given score threshold using protein-protein BLAST. This PSSM is used to further search the database for new matches, and is updated for subsequent iterations with these newly detected sequences. Thus, PSI-BLAST provides a means of detecting distant relationships between proteins.

[00240] In another aspect, the case by case analysis is performed using HHpred, a method for sequence database searching and structure prediction that is as easy to use as BLAST or PSI-BLAST and that is at the same time much more sensitive in finding remote homologs. In fact, HHpred' s sensitivity is competitive with the most powerful servers for structure prediction currently available. HHpred is the first server that is based on the pairwise comparison of profile hidden Markov models (HMMs). Whereas most conventional sequence search methods search sequence databases such as UniProt or the NR, HHpred searches alignment databases, like Pfam or SMART. This greatly simplifies the list of hits to a number of sequence families instead of a clutter of single sequences. All major publicly available profile and alignment databases are available through HHpred. HHpred accepts a single query sequence or a multiple alignment as input. Within only a few minutes it returns the search results in an easy-to-read format similar to that of PSI-BLAST.

Search options include local or global alignment and scoring secondary structure similarity. HHpred can produce pairwise query-template sequence alignments, merged query-template multiple alignments (e.g. for transitive searches), as well as 3D structural models calculated by the MODELLER software from HHpred alignments.

[00241] Orthologs ofCpfl

[00242] 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. Homologs and orthologs may be identified by homology modelling (see, e.g., Greer, Science vol. 228 (1985) 1055, and Blundell et al. Eur J Biochem vol 172 (1988), 513) or "structural BLAST" (Dey F, Cliff Zhang Q, Petrey D, Honig B. Toward a "structural BLAST": using structural relationships to infer function. Protein Sci. 2013 Apr;22(4):3 59-66. doi: 10.1002/pro.2225.). See also Shmakov et al. (2015) for application in the field of CRISPR-Cas loci. Homologous proteins may but need not be structurally related, or are only partially structurally related.

[00243] The Cpfl gene is found in several diverse bacterial genomes, typically in the same locus with casl, cas2, and cas4 genes and a CRISPR cassette (for example, FNFX 1_1431-FNFX1 _1428 of Francisella cf . novicida Fxl). Thus, the layout of this putative novel CRISPR-Cas system appears to be similar to that of type II-B. Furthermore, similar to Cas9, the Cpfl protein contains a readily identifiable C-terminal region that is homologous to the transposon ORF-B and includes an active RuvC-like nuclease, an arginine-rich region, and a Zn finger (absent in Cas9). However, unlike Cas9, Cpfl is also present in several genomes without a CRISPR-Cas context and its relatively high similarity with ORF-B suggests that it might be a transposon component. It was suggested that if this was a genuine CRISPR-Cas system and Cpfl is a functional analog of Cas9 it would be a novel CRISPR-Cas type, namely type V (See Annotation and Classification of CRISPR-Cas Systems. Makarova KS, Koonin EV. Methods Mol Biol. 2015;131 1:47-75). However, as described herein, Cpfl is denoted to be in subtype V-A to distinguish it from C2clp which does not have an identical domain structure and is hence denoted to be in subtype V-B.

[00244] The present invention encompasses the use of a Cpfl effector protein, derived from a Cpfl locus denoted as subtype V-A. Herein such effector proteins are also referred to as "Cpflp", e.g., a Cpfl protein (and such effector protein or Cpfl protein or protein derived from a Cpfl locus is also called "CRISPR-Cas protein").

[00245] In particular embodiments, the effector protein is a Cpfl effector protein from an organism from a genus comprising *Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Methanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes, Helcococcus, Leptospira, Desulfovibrio, Desulfonatronum, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium, Butyvibrio, Perigrinibacterium, Pareubacterium, Moraxella, Thiomicrospira or Acidaminococcus.* In particular embodiments, the Cpfl effector protein is selected from an organism from a genus selected from *Eubacterium, Lachnospiraceae, Leptotrichia, Francisella, Methanomethyophilus, Porphyromonas, Prevotella, Leptospira, Butyvibrio, Perigrinibacterium, Pareubacterium, Moraxella, Thiomicrospira or Acidaminococcus*

[00246] In further particular embodiments, the Cpfl effector protein is from an organism selected from *S. mutans, S. agalactiae, S. equisimilis, S. sanguinis, S. pneumonia; C. jejuni, C coli; N. salsuginis, N. tergarcus; S. auricularis, S. carnosus; N. meningitides, N. gonorrhoeae; L. monocytogenes, L. ivanovii; C. botulinum, C. difficile, C tetani, C sordellii, L inadai, F. tularensis 1, P. albensis, L. bacterium, B. proteoclasticus, P. bacterium, P. crevioricanis, P. disiens* and *P. macacae* .

[00247] The effector protein may comprise a chimeric effector protein comprising a first fragment from a first effector protein (e.g., a Cpfl) ortholog and a second fragment from a second effector (e.g., a Cpfl) protein ortholog, and wherein the first and second effector protein orthologs are different. At least one of the first and second effector protein (e.g., a Cpfl) orthologs may comprise an effector protein (e.g., a Cpfl) from an organism comprising *Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Methanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes,*

Helcococcus, Letospira, Desulfovibrio, Desulfonatronum, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium, Butyvibrio, Perigrinibacterium, Pareubacterium, Moraxella, Thiomicrospira or Acidaminococcus; e.g., a chimeric effector protein comprising a first fragment and a second fragment wherein each of the first and second fragments is selected from a Cpfl of an organism comprising *Streptococcus, Campylobacter, Nitratifractor, Staphylococcus, Parvibaculum, Roseburia, Neisseria, Gluconacetobacter, Azospirillum, Sphaerochaeta, Lactobacillus, Eubacterium, Corynebacter, Carnobacterium, Rhodobacter, Listeria, Paludibacter, Clostridium, Lachnospiraceae, Clostridiaridium, Leptotrichia, Francisella, Legionella, Alicyclobacillus, Methanomethyophilus, Porphyromonas, Prevotella, Bacteroidetes, Helcococcus, Letospira, Desulfovibrio, Desulfonatronum, Opitutaceae, Tuberibacillus, Bacillus, Brevibacilus, Methylobacterium, Butyvibrio, Perigrinibacterium, Pareubacterium, Moraxella, Thiomicrospira or Acidaminococcus* wherein the first and second fragments are not from the same bacteria; for instance a chimeric effector protein comprising a first fragment and a second fragment wherein each of the first and second fragments is selected from a Cpfl of *S. mutans, S. agalactiae, S. equisimilis, S. sanguinis, S. pneumonia; C. jejuni, C coli; N. salsuginis, N. tergarcus; S. auricularis, S. carnosus; N. meningitides, N. gonorrhoeae; L. monocytogenes, L. ivanovii; C. botulinum, C. difficile, C tetani, C sordellii; Francisella tularensis 1, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens* and *Porphyromonas macacae,* wherein the first and second fragments are not from the same bacteria.

[00248] In a more preferred embodiment, the Cpflp is derived from a bacterial species selected from *Francisella tularensis 1, Prevotella albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Moraxella bovoculi AAX08_00205, Moraxella bovoculi AAX11_00205, Butyrivibrio sp. NC3005, Thiomicrospira sp. XS5, Leptospira inadai, Lachnospiraceae bacterium ND2006,*

Porphyromonas crevioricanis 3, Prevotella disiens and *Porphyromonas macacae.* In certain embodiments, the Cpflp is derived from a bacterial species selected from *Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2'020.* In certain embodiments, the effector protein is derived from a subspecies of *Francisella tularensis 1,* including but not limited to *Francisella tularensis subsp. Novicida.* In certain preferred embodiments, the Cpflp is derived from a bacterial species selected from *Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium ND2006, Lachnospiraceae bacterium MA2020, Moraxella bovoculi AAX08_00205, Moraxella bovoculi AAX11_00205, Butyrivibrio sp. NC3005,* or *Thiomicrospira sp. XS5.*

[00249] In particular embodiments, the homologue or orthologue of Cpfl as referred to herein has a sequence homology or identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with Cpfl. In further embodiments, the homologue or orthologue of Cpfl as referred to herein has a sequence identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the wild type Cpfl. Where the Cpfl has one or more mutations (mutated), the homologue or orthologue of said Cpfl as referred to herein has a sequence identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the mutated Cpfl .

[00250] In an ambodiment, the Cpfl protein may be an ortholog of an organism of a genus which includes, but is not limited to Acidaminococcus sp, *Lachnospiraceae bacterium or Moraxella bovoculi;* in particular embodiments, the type V Cas protein may be an ortholog of an organism of a species which includes, but is not limited to *Acidaminococcus sp. BV3L6; Lachnospiraceae bacterium ND2006* (LbCpfl) or *Moraxella bovoculi 237.* In particular embodiments, the homologue or orthologue of Cpfl as referred to herein has a sequence homology or identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with one or more of the Cpfl sequences disclosed herein. In further embodiments, the homologue or orthologue of Cpf as referred to herein has a sequence identity of at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the wild type FnCpfl, AsCpfl or LbCpfl.

[00251] In particular embodiments, the Cpfl protein of the invention has a sequence homology or identity of at least 60%, more particularly at least 70, such as at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at

least 95% with FnCpfl, AsCpfl or LbCpfl. In further embodiments, the Cpfl protein as referred to herein has a sequence identity of at least 60%, such as at least 70%, more particularly at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with the wild type AsCpfl or LbCpfl. In particular embodiments, the Cpfl protein of the present invention has less than 60% sequence identity with FnCpfl. The skilled person will understand that this includes truncated forms of the Cpfl protein whereby the sequence identity is determined over the length of the truncated form. In particular embodiments, the Cpfl enzyme is not FnCpfl .

[00252] In some embodiments, the CRISPR effector protein is a Cpfl protein derived from an organism from the genus of *Eubacterium.* In some embodiments, the CRISPR effector protein is a Cpfl protein derived from an organism from the bacterial species of *Eubacterium rectale.* In some embodiments, the amino acid sequence of the Cpfl effector protein corresponds to NCBI Reference Sequence WP_055225123.1, NCBI Reference Sequence WP_055237260.1, NCBI Reference Sequence WP_055272206.1, or GenBank ID OLA16049.1. In some embodiments, the Cpfl effector protein has a sequence homology or sequence identity of at least 60%, more particularly at least 70%, such as at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95%, with NCBI Reference Sequence WP_055225123.1, NCBI Reference Sequence WP_055237260.1, NCBI Reference Sequence WP_055272206.1, or GenBank ID OLA16049.1. The skilled person will understand that this includes truncated forms of the Cpfl protein whereby the sequence identity is determined over the length of the truncated form. In some embodiments, the Cpfl effector recognizes the PAM sequence of TTTN or CTTN.

[00253] In some embodiments, the CRISPR effector protein is a Cpfl protein derived from an organism from the genus of Eubacterium. In some embodiments, the CRISPR effector protein is a Cpfl protein derived from an organism from the bacterial species of Eubacterium rectale. In some embodiments, the amino acid sequence of the Cpfl effector protein corresponds to NCBI Reference Sequence WP_055225123.1, NCBI Reference Sequence WP_055237260.1, NCBI Reference Sequence WP_055272206.1, or GenBank ID OLA16049.1. In some embodiments, the Cpfl effector protein has a sequence homology or sequence identity of at least 60%, more particularly at least 70, such as at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95%, with NCBI Reference Sequence WP_055225123.1, NCBI Reference Sequence

WP_055237260.1, NCBI Reference Sequence WP_055272206.1, or GenBank ID OLA16049.1. The skilled person will understand that this includes truncated forms of the Cpfl protein whereby the sequence identity is determined over the length of the truncated form. In some embodiments, the Cpfl effector recognizes the PAM sequence of TTTN or CTTN.

[00254] Codon optimized Cpfl sequences

[00255] Where the effector protein is to be administered as a nucleic acid, the application envisages the use of codon-optimized Cpfl sequences. An example of a codon optimized sequence, is in this instance a sequence optimized for expression in a eukaryote, e.g., humans (i.e. being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed; see, e.g., SaCas9 human codon optimized sequence in WO 2014/093622 (PCT/US20 13/074667) as an example of a codon optimized sequence (from knowledge in the art and this disclosure, codon optimizing coding nucleic acid molecule(s), especially as to effector protein (e.g., Cpfl) is within the ambit of the skilled artisan). Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species other than human, or for codon optimization for specific organs is known. In some embodiments, an enzyme coding sequence encoding a DNA/RNA-targeting Cas protein is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a 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 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.

[00256] In certain of the following, Cpfl amino acids are followed by *nuclear localization signals (NLS)* (italics), a glycine-serine (GS) linker (underlined), and **3x HA tag** (bold). In some embodiments, Cpfl amino acid sequences correspond to sequences without the NLS, GS linker and 3x HA tag.

[00257] 1- Franscisella tularensis subsp. novicida U112 (FnCpfl)

[00258] MSIYQEFVNKYSLSKTLRFELIPQGKTLENIKARGLILDDEKRAKDYKKA KQIIDKYHQFFIEEILSSVCISEDLLQNYSDVYFKLKKSDDDNLQKDFKSAKDTIKKQ ISEYIKDSEKFKNLFNQNLIDAKKGQESDLILWLKQSKDNGIELFKANSDITDIDEAL EIIKSFKGWTTYFKGFHENRKNVYSSNDIPTSIIYRIVDDNLPKFLENKAKYESLKDK APEAINYEQIKKDL AEELTFDID YKT SEVNQRVF SLDEVFEIANFNN YLNQ SGITKFN TIIGGKFWGENTKRKGINEYINLYSQQINDKTLKKYKMSVLFKQILSDTESKSFVID KLEDD SDVVTTMQ SFYEQIAAFKT VEEKSIKETL SLLFDDLK AQKLDL SKIYFKNDK

SLTDLSQQVFDDYSVIGTAVLEYITQQIAPKNLDNPSKKEQELIAKKTEKAKYLSLE
TIKLALEEFNKHRDIDKQCRFEEILANFAAIPMIFDEIAQNKDNLAQISIKYQNQGKK
DLLQASAEDDVKAIKDLLDQTNNLLHKLKIFHISQSEDKANILDKDEHFYLVFEECY
FELANIVPLYNKIRNYITQKPYSDEKFKLNFENSTLANGWDKNKEPDNTAILFIKDD
KYYLGVMNKKNNKIFDDKAIKENKGEGYKKIVYKLLPGANKMLPKVFFSAKSIKF
YNPSEDILRIRNHSTHTKNGSPQKGYEKFEFNIEDCRKFIDFYKQSISKHPEWKDFGF
RFSDTQRYNSIDEFYREVENQGYKLTFENISESYIDSVVNQGKLYLFQIYNKDFSAY
SKGRPNLHTLYWKALFDERNLQDVVYKLNGEAELFYRKQSIPKKITHPAKEAIANK
NKDNPKKESVFEYDLIKDKRFTEDKFFFHCPITINFKSSGANKFNDEINLLLKEKAND
VHILSIDRGERHLAYYTLVDGKGNIIKQDTFNIIGNDRMKTNYHDKLAAIEKDRDSA
RKDWKKINNIKEMKEGYLSQVVHEIAKLVIEYNAIVVFEDLNFGFKRGRFKVEKQV
YQKLEKMLIEKLNYLVFKDNEFDKTGGVLRAYQLTAPFETFKKMGKQTGIIYYVPA
GFTSKICPVTGFVNQLYPKYESVSKSQEFFSKFDKICYNLDKGYFEFSFDYKNFGDK
AAKGKWTIASFGSRLINFRNSDKNHNWDTREVYPTKELEKLLKDYSIEYGHGECIK
AAICGESDKKFFAKLTSVLNTILQMRNSKTGTELDYLISPVADVNGNFFDSRQAPKN
MPQDADANGAYHIGLKGLMLLGRIKNNQEGKKLNLVIKNEEWEFVQNRNNiC *RPA*
*ATKKAGQAKX KK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA** (SEQ ID N0 :54)

[00259] 3- Lachnospiraceae bacterium MC2017 (Lb3Cpfl)

[00260] MDYGNGQFERRAPLTKTITLRLKPIGETRETIREQKLLEQDAAFRKLVET
VTPIVDDCIRKIADNALCHFGTEYDFSCLGNAISKNDSKAIKKETEKVEKLLAKVLT
ENLPDGLRKVNDINSAAFIQDTLTSFVQDDADKRVLIQELKGKTVLMQRFLTTRITA
LTVWLPDRVFENFNIFIENAEKMRILLDSPLNEKIMKFDPDAEQYASLEFYGQCLSQ
KDIDSYNLIISGIYADDEVKNPGINEIVKEYNQQIRGDKDESPLPKLKKLHKQILMPV
EKAFFVRVLSNDSDARSILEKILKDTEMLPSKIIEAMKEADAGDIAVYGSRLHELSH
VIYGDHGKLSQIIYDKESKRISELMETLSPKERKESKKRLEGLEEHIRKSTYTFDELN
RYAEKNVMAAYIAAVEESCAEIMRKEKDLRTLLSKEDVKIRGNRHNTLIVKNYFN
AWTWRNLIRILRRKSEAEIDSDFYDVLDDSVEVLSLTYKGENLCRSYITKKIGSDLK
PEIATYGSALRPNSRWWSPGEKFNVKFHTIVRRDGRLYYFILPKGAKPVELEDMDG
DIECLQMRKIPNPTIFLPKLVFKDPEAFFRDNPEADEFVFLSGMKAPVTITRETYEAY
RYKLYTVGKLRDGEVSEEEYKRALLQVLTAYKEFLENRMIYADLNFGFKDLEEYK
DSSEFIKQVETHNTFMCWAKVSSSQLDDLVKSGNGLLFEIWSERLESYYKYGNEKV
LRGYEGVLLSILKDENLVSMRTLLNSRPMLVYRPKESSKPMVVHRDGSRVVDRFD
KDGKYIPPEVHDELYRFFNMLIKEKLGEKARKILDNKKVKVKVLESERVKWSKFY

DEQFAVTFSVKKNADCLDTTKDLNAEVMEQYSESNRLILIRNTTDILYYLVLDKNG
KVLKQRSLNIINDGARDVDWKERFRQVTKDRNEGYNEWDYSRTSNDLKEVYLNY
ALKEIAEAVIEYNAILIIEKMSNAFKDKYSFLDDVTFKGFETKLLAKLSDLHFRGIKD
GEPCSFTNPLQLCQNDSNKILQDGVIFMVPNSMTRSLDPDTGFIFAINDHNIRTKKA
KLNFLSKFDQLKVSSEGCLFMKYSGDSLPTHNTDNRVWNCCCNHPITNYDRETKKV
EFIEEPVEELSRVLEENGIETDTELNKLNERENVPGKVVDAIYSLVLNYLRGTVSGV
AGQRAVYYSPVTGKKYDISFIQAMNLNRKCDYYRIGSKERGEWTDFVAQLINX ***RPA***
*A/KKA(;QAKKKK*GS**YPYDVPDYAYPYDVPDYAYPYDI)\'PI)Y.** (SEQ ID NO:55)

[00261] 4- Butyrivibrio proteoclasticus (BpCpfl)

[00262] MLLYENYTKRNQITKSLRLELRPQGKTLRNIKELNLLEQDKAIYALLERL
KPVIDEGIKDIARDTLKNCEL SFEKLYEHFL SGDKKAYAKE SERLKKEIVKTLIKNLP
EGIGKISEINSAKYLNGVLYDFIDKTHKDSEEKQNILSDILETKGYLALFSKFLTSRIT
TLEQ SMPKRVIENFEIYAANIPKMQD ALERGAVSF AIEYESIC SVDYYNQILSQEDID
SYNRLISGIMDEDGAKEKGINQTISEKNIKIKSEHLEEKPFRILKQLHKQILEEREKAF
TIDHIDSDEEVVQVTKEAFEQTKEQWENIKKINGFYAKDPGDITLFIVVGPNQTHVL
SQLIYGEHDRIRLLLEEYEKNTLEVLPRRTKSEKARYDKFVNAVPKKVAKESHTFD
GLQKMTGDDRLFILYRDELARNYMRIKEAYGTFERDILKSRRGIKGNRDVQESLVS
FYDELTKFRSALRIINSGNDEKADPIFYNTFDGIFEKANRTYKAENLCRNYVTKSPA
DDARIMASCLGTPARLRTHWWNGEENFAINDVAMIRRGDEYYYFVLTPDVKPVDL
KTKDETDAQIFVQRKGAKSFLGLPKALFKCILEPYFESPEFIKNDKNCVIEEYVSKPL
TIDRRAYDIFKNGTFKKTNIGIDGLTEEKFKDDCRYLIDVYKEFIAVYTRYSCFNMS
GLKRADEYNDIGEFFSDVDTRLCTMEWIPVSFERFNDMVDKKEGLLFLVRSMFLYN
RPRKPYERTFIQLF SD SNMEHTSMLLNSRAMIQYRAASLPRRVTFIKKGSILVALRD S
NGEHIPMHIREAIYKMKNNFDISSEDFIMAKAYLAEFiDVAIKKANEDIIRNRRYTED
KFFLSLSYTKNADISARTLDYFNDKVEEDTQDSRMAVIVTRNLKDLTYVAVVDEKN
NVLEEKSLNEIDGVNYRELLKERTKIKYFiDKTRLWQYDVSSKGLKEAYVELAVTQI
SKLATKYNAVVVVESMS STFKDKF SFLDEQIFKAFEARLC ARMSDLSFNTIKEGEA
GSISNPIQVSNNNGNSYQDGVIYFLNNAYTRTLCPDTGFVDVFDKTRLITMQSKRQF
FAKMKDIRIDDGEMLFTFNLEEYPTKRLLDRKEWTVKIAGDGSYFDKDKGEYVYV
NDIVREQIIPALLEDKAVFDGNMAEKFLDKTAISGKSVELIYKWFANALYGIITKKD
GEKIYRSPITGTEIDVSKNTTYNFGKKFMFKQEYRGDGDFLDAFLNYMQAQDIAViC
RPAATKKAGQAKKKKGSYVYOWOYAYVYOWOYAYVYOWOYA (SEQ ID NO:56)

[00263] 5- Peregrinibacteria bacterium GW201 1_GWA_33_10 (PeCpfl)

[00264] MSNFFKNFTNLYELSKTLRFELKPVGDTLTNMKDHLEYDEKLQTFLKDQ NIDDAYQALKPQFDEIHEEFITDSLESKKAKEIDFSEYLDLFQEKKELNDSEKKLRNK IGETFNKAGEKWKKEKYPQYEWKKGSKIANGADILSCQDMLQFIKYKNPEDEKIK NYIDDTLKGFFTYFGGFNQNRANYYETKKEASTAVATRIVHENLPKFCDNVIQFKHI IKRKKDGTVEKTERKTEYLNAYQYLKNNNfCITQIKDAETEKMIESTPIAEKIFDVYY FSSCLSQKQIEEYNRIIGHYNLLINLYNQAKRSEGKHLSANEKKYKDLPKFKTLYKQ IGCGKKKDLFYTIKCDTEEEANKSRNEGKESHSVEEIINKAQEAINKYFKSNNDCENI NTVPDFINYILTKENYEGVYWSKAAMNTISDKYFANYHDLQDRLKEAKVFQKADK KSEDDIKIPEAIELSGLFGVLDSLADWQTTLFKSSILSNEDKLKIITDSQTPSEALLKMI FNDIEKNMESFLKETNDIITLKKYKGNKEGTEKIKQWFDYTLAINRMLKYFLVKEN KIKGNSLDTNISEALKTLIYSDDAEWFKWYDALRNYLTQKPQDEAKENKLKLNFD NPSLAGGWDVNKECSNFCVILKDKNEKKYLAIMKKGENTLFQKEWTEGRGKNLT KKSNPLFEINNCEILSKMEYDFWADVSKMIPKCSTQLKAVVNHFKQSDNEFIFPIGY KVTSGEKFREECKISKQDFELNNKVFNKNELSVTAMRYDLSSTQEKQYIKAFQKEY WELLFKQEKRDTKLTNNEIFNEWINFCNKKYSELLSWERKYKDALTNWINFCKYFL SKYPKTTLFNYSFKESENYNSLDEFYRDVDICSYKLNINTTINKSILDRLVEEGKLYL FEIKNQDSNDGKSIGHKNNLHTIYWNAIFENFDNRPKLNGEAEIFYRKAISKDKLGI VKGKKTKNGTEHKNYRFSKEKFILHVPITLNFCSNNEYVNDIVNTKFYNFSNLHFLG IDRGEKHLAYYSLVNfCNGEIVDQGTLNLPFTDKDGNQRSIKKEKYFYNKQEDKWE AKEVDCWNYNDLLDAMASNRDMARKNWQRIGTIKEAKNGYVSLVIRKIADLAVN NERPAFIVLEDLNTGFKRSRQKIDKSVYQKFELALAKKLNFLVDKNAKRDEIGSPTK ALQLTPPVNNYGDIENKKQAGFMLYTRANYTSQTDPATGWRKTIYLKAGPEETTY KKDGKIKNKSVKDQIIETFTDIGFDGKDYYFEYDKGEFVDEKTGEIKPKKWRLYSG ENGKSLDRFRGEREKDKYEWKIDKIDIVKILDDLFVNFDKNISLLKQLKEGVELTRN NEHGTGESLRFAINLIQQIRNTGNNERDNDFILSPVRDENGKHFDSREYWDKETKGE KISMPSSGDANGAFNIARKGIFMNAHILANSDSKDLSLFVSDEEWDLULNNKTEWK KQLNIFS*SRKAMAKRKKKRPAA TKKAGQAKKKKGS***YPYDVPDYAYPYDVPDYAYP YDVPDYA** (SEQ ID NO:57)

[00265] 6- Parcubacteria bacterium GWC201 1_GWC2_44_17 (PbCpfl)

[00266] MENIFDQFIGKYSLSKTLRFELKPVGKTEDFLKINKVFEKDQTIDDSYNQ AKFYFDSLHQKFIDAALASDKTSELSFQNFADVLEKQNKIILDKKREMGALRKRDK NAVGIDRLQKEINDAEDIIQKEKEKIYKDVRTLFDNEAESWKTYYQEREVDGKKITF

SKADLKQKGADFLTAAGILKVLKYEFPEEKEKEFQAKNQPSLFVEEKENPGQKRYI
FDSFDKFAGYLTKFQQTKKNLYAADGTSTAVATRIADNFIIFHQNTKVFRDKYKNN
HTDLGFDEENIFEIERYKNCLLQREIEHIKNENSYNKIIGRINKKIKEYRDQKAKDTK
LTKSDFPFFKNLDKQILGEVEKEKQLIEKTREKTEEDVLIERFKEFIENNEERFTAAK
KLMNAFCNGEFESEYEGIYLKNKAINTISRRWFVSDRDFELKLPQQKSKNKSEKNEP
KVKKFISIAEIKNAVEELDGDIFKAVFYDKKIIAQGGSKLEQFLVIWKYEFEYLFRDI
ERENGEKLLGYD SCLKIAKQLGIFPQEKEAREKATAVIKNYADAGLGIFQMMKYF S
LDDKDRKNTPGQLSTNFYAEYDGYYKDFEFIKYYNEFRNFITKKPFDEDKIKLNFEN
GALLKGWDENKEYDFMGVILKKEGRLYLGIMHKNHRKLFQSMGNAKGDNANRY
QKMIYKQIADASKDVPRLLLTSKKAMEKFKPSQEILRIKKEKTFKRESKNFSLRDLH
ALIEYYRNCIPQYSNWSFYDFQFQDTGKYQNIKEFTDDVQKYGYKISFRDIDDEYIN
QALNEGKMYLFEVVmDIYNTKNGSKNLHTLYFEHILSAENLNDPVFKLSGMAEIF
QRQPSVNEREKITTQKNQCILDKGDRAYKYRRYTEKKIMFHMSLVLNTGKGEIKQV
QFNKIINQRIS SSDNEMRVNVIGIDRGEKNLLYYS VVKQNGEIIEQASLNEINGVNYR
DKLIEREKERLKNRQSWKPVVKIKDLKKGYISHVIHKICQLIEKYSAIVVLEDLNMR
FKQIRGGIERSVYQQFEKALIDKLGYLVFKDNRDLRAPGGVLNGYQLSAPFVSFEK
MRKQTGILFYTQAEYTSKTDPITGFRKNVYISNSASLDKIKEAVKKFDAIGWDGKE
QSYFFKYNPYNLADEKYKNSTVSKEWAIFASAPRIRRQKGEDGYWKYDRVKVNEE
FEKLLKVWNFVNPKATDIKQEIIKKEKAGDLQGEKELDGRLRNFWHSFIYLFNLVL
ELRNSFSLQIKIKAGEVIAVDEGVDFIASPVKPFFTTPNPYIPSNLCWLAVENADANG
AYNIARKGVMILKKIREHAKKDPEFKKLPNLFISNAEWDEAARDWGKYAGTTALN
L*DH**KRPAA TKKAGQAKKKK*GS**YPYDVPDY**A**YPYDVPDY**A**YPYDVPDY**A (SEQ ID NO:58)

[00267] 7- Smithella sp. SC_K08D17 (SsCpfl)

[00268] MQTLFENFTNQYPVSKTLRFELIPQGKTKDFIEQKGLLKKDEDRAEKYK
KVKNIIDEYHKDFIEKSLNGLKLDGLEKYKTLYLKQEKDDKDKKAFDKEKENLRK
QIANAFRNNEKFKTLFAKELIKNDLMSFACEEDKKNVKEFEAFTTYFTGFHQNRAN
MYVADEKRTAIASRLIHENLPKFIDNIKIFEKMKKEAPELLSPFNQTLKDMKDVIKG
TTLEEIFSLDYFNKTLTQSGIDIYNSVIGGRTPEEGKTKIKGLNEYINTDFNQKQTDK
KKRQPKFKQLYKQILSDRQSLSFIAEAFKNDTEILEAIEKFYVNELLHFSNEGKSTNV
LDAIKNAVSNLESFNLTKMYFRSGASLTDVSRKVFGEWSIINRALDNYYATTYPIKP
REKSEKYEERKEKWLKQDFNVSLIQTAIDEYDNETVKGKNSGKVIADYFAKFCDD
KETDLIQKVmGYIAVKDLLNTPCPEl^KLGSNKDQVKQIKAFMDSFMDFMHFVRP

LSLKDTDKEKDETFYSLFTPLYDHLTQTIALYNKVRNYLTQKPYSTEKIKLNFENST
LLGGWDLNKETDNTAnLRKDNLYYLGIMDKRHNRIFRNVPKADKKDFCYEKMVY
KLLPGANKMLPKVFFSQSRIQEFTPSAKLLENYANETHKKGDNFNLNHCHKLIDFF
KDSINKHEDWKNFDFRFSATSTYADLSGFYHEVEHQGYKISFQSVADSFIDDLVNE
GKLYLFQIYNKDFSPFSKGKPNLHTLYWKMLFDENNLKDVVYKLNGEAEVFYRKK
SIAEKNTTIHKANESIINKNPDNPKATSTFNYDIVKDKRYTIDKFQFHIPITMNFKAEG
IFNMNQRVNQFLKANPDINIIGIDRGERHLLYYALINQKGKILKQDTLNVIANEKQK
VDYHNLLDKKEGDRATARQEWGVIETIKELKEGYLSQVIHKLTDLMIENNAIIVME
DLNFGFKRGRQKVEKQVYQKFEKMLIDKLNYLVDKNKKANELGGLLNAFQLANK
FESFQKMGKQNGFIFYVPAWNTSKTDPATGFIDFLKPRYENLNQAKDFFEKFDSIRL
NSKADYFEFAFDFKNFTEKADGGRTKWTVCTTNEDRYAWNRALNNNRGSQEKYD
ITAELKSLFDGKVDYKSGKDLKQQIASQESADFFKALMKNLSITLSLRHNNGEKGD
NEQDYILSPVADSKGRFFDSRKADDDMPKNADANGAYHIALKGLWCLEQISKTDD
LKKVKLAISNKEWLEFVQTLKG*Ki?PAATK¾* G*QAK***X** *KK***GS**
SYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID N0 :59)

[00269] 8- Acidaminococcus sp. BV3L6 (AsCpfl)

[00270] MTQFEGFTNLYQVSKTLRFELIPQGKTLKHIQEQGFIEEDKARNDHYKEL
KPIIDRIYKTYADQCLQLVQLDWENLSAAIDSYRKEKTEETRNALIEEQATYRNAIH
DYFIGRTDNLTDAINKRHAEIYKGLFKAELFNGKVLKQLGTVTTTEHENALLRSFD
KFTTYFSGFYENRKNVFSAEDISTAIPHRIVQDNFPKFKENCHIFTRLITAVPSLREHF
ENVKKAIGIFVSTSIEEVFSFPFYNQLLTQTQIDLYNQLLGGISREAGTEKIKGLNEVL
NLAIQKNDETAHIIASLPHRFIPLFKQILSDRNTLSFILEEFKSDEEVIQSFCKYKTLLR
NENVLETAEALFNELNSIDLTHIFISHKKLETISSALCDHWDTLRNALYERRISELTG
KITKSAKEKVQRSLKHEDINLQEIISAAGKELSEAFKQKTSEILSHAHAALDQPLPTT
LKKQEEKEILKSQLDSLLGLYHLLDWFAVDESNEVDPEFSARLTGIKLEMEPSLSFY
NKARNYATKKPYSVEKFKLNFQMPTLASGWDVNKEKNNGAILFVKNGLYYLGFM
PKQKGRYKALSFEPTEKTSEGFDKMYYDYFPDAAKMIPKCSTQLKAVTAHFQTHT
TPILLSNNFIEPLEITKEIYDLNNPEKEPKKFQTAYAKKTGDQKGYREALCKWIDFTR
DFLSKYTKTTSIDLSSLRPSSQYKDLGEYYAELNPLLYHISFQRIAEKEFMDAVETGK
LYLFQIYNKDFAKGHHGKPNLHTLYWTGLFSPENLAKTSIKLNGQAELFYRPKSRM
KRMAHRLGEKMLNKKLKDQKTPIPDTLYQELYDYVNHRLSHDLSDEARALLPNVI
TKEVSHEIIKDRRFTSDKFFFHVPITLNYQAANSPSKFNQRVNAYLKEHPETPIIGIDR
GERNLIYITVIDSTGKILEQRSLNTIQQFDYQKKLDNREKERVAARQAWSVVGTIKD

LKQGYL SQVIHEIVDLMIHYQAVVVLENLNFGFKSKRTGIAEKAVYQQFEKMLIDK LNCLVLKDYPAEKVGGVLNPYQLTDQFTSFAKMGTQSGFLFYVPAPYTSKIDPLTG FVDPFWKTIKNHESRKHFLEGFDFLHYDVKTGDFILHFKMNRNLSFQRGLPGFMP AWDIVFEKNETQFDAKGTPFIAGKRIVPVIENHRFTGRYRDLYPANELIALLEEKGIV FRDGSNILPKLLENDDSHAIDTMVALIRSVLQMRNSNAATGEDYINSPVRDLNGVC FD SRF QNPEWPMDADANGAYHIALKGQLLLNHLKE SKDLKLQNGI SNQDWLAYIQ ELRN*KRPAA TKKAGQAKKKK*G SYPYDVPDYAYPYD VPDYAYPYDVPDYA (SEQ ID NO: 60)

[00271] 9- Lachnospiraceae bacterium MA2020 (Lb2Cpfl)

[00272] MYYESLTKQYPVSKTIRNELIPIGKTLDNIRQNNILESDVKRKQNYEHVK GILDEYHKQLINEALDNCTLPSLKIAAEIYLKNQKEVSDREDFNKTQDLLRKEVVEK LKAHENF TKIGKKDILDLLEKLP SI SEDDYN ALE SFRNF YTYF TSYNKVRENL YSDK EKSSTVAYRLINENFPKFLDNVKSYRFVKTAGILADGLGEEEQDSLFIVETFNKTLT QDGIDTYNSQVGKINSSINLYNQKNQKANGFRKIPKMKMLYKQILSDREESFIDEFQ SDEVLIDNVESYGSVLIESLKSSKVSAFFDALRESKGKNVYVKNDLAKTAMSNIVFE NWRTFDDLLNQEYDLANENKKKDDKYFEKRQKELKKNKSYSLEHLCNLSEDSCN LIENYIHQISDDIENIIINNETFLRIVINEHDRSRKLAKNRKAVKAIKDFLDSIKVLERE LKLrNS SGQELEKDLIVYS AHEELLVELKQVD SL YNMTRNYLTKKPF STEKVKLNF NRSTLLNGWDRNKETDNLGVLLLKDGKYYLGFMNTSANKAFVNPPVAKTEKVFK KVDYKLLPVPNQMLPKVFFAKSNIDFYNPSSEIYSNYKKGTHKKGNMFSLEDCHNL IDFFKESISKHEDWSKFGFKFSDTASYNDISEFYREVEKQGYKLTYTDIDETYrNDLIE RNELYLFQIYNfCDFSMYSKGKLNLHTLYFMMLFDQRNIDDVVYKLNGEAEVFYRP ASISEDELnHKAGEEIKNKNPNRARTKETSTFSYDIVKDKRYSKDKFTLHIPITMNFG VDEVKRFNDAVNSAIRIDENVNVIGIDRGERNLLYVVVIDSKGNILEQISLNSIINKE YDIETDYH ALLDEREGGRDK ARKDWNTVENIRDLK AGYL SQVVNVVAKL VLKYN AIICLEDLNF GFKRGRQKVEKQVYQKFEKMLIDKLNYL VIDK SREQTSPKELGGAL NALQLTSKFKSFKELGKQSGVIYYVPAYLTSKIDPTTGFANLFYMKCENVEKSKRFF DGFDFIRFNALENVFEFGFDYRSFTQRACGrNSKWTVCTNGERIIKYRNPDKNNMFD EKVVVVTDEMKNLFEQYKIPYEDGRNVKDMIISNEEAEFYRRLYRLLQQTLQMRN STSDGTRDYIISPVKNKREAYFNSELSDGSVPKDADANGAYNIARKGLWVLEQIRQ KSEGEKrNLAMTNAEWLEYAQTHLLX *RPAATKK¾* G*QAK*X *KK*G SYPYDVPDYAYPY DVPDYAYPYDVPDYA (SEQ ID N0 :61)

[00273] 10- Candidatus Methanoplasma termitum (CMtCpfl)

[00274] MNNYDEFTKLYPIQKTIRFELKPQGRTMEHLETFNFFEEDRDRAEKYKIL KEAIDEYHKKFIDEHLTNMSLDWNSLKQISEKYYKSREEKDKKVFLSEQKRMRQEI VSEFKKDDRFKDLF SKKLF SELLKEEIYKKGNHQEID ALK SFDKF SGYFIGLHENRK NMYSDGDEITAISNPJV1^OTPKFLDM.QKYQEARKKYPEWIIKAESALVAHNIKM DEVF SLEYFNK VLNQEGIQR YNL ALGGY VTK SGEKMMGLND AL NL AHQ SEK SSKG RIHMTPLFKQILSEKESFSYIPDVFTEDSQLLPSIGGFFAQIENDKDGNIFDRALELISS YAEYDTERIYIRQADINRVSNVIFGEWGTLGGLMREYKADSINDINLERTCKKVDK WLDSKEFALSDVLEAIKRTGNNDAFNEYISKMRTAREKIDAARKEMKFISEKISGDE ESIHIIKTLLDSVQQFLHFFNLFKARQDIPLDGAFYAEFDEVHSKLFAIVPLYNKVRN YLTKNNLNTKKIKLNFKNPTLANGWDQNKVYDYASLIFLRDGNYYLGIINPKRKK NIKFEQGSGNGPFYRKMVYKQIPGPNKNLPRVFLTSTKGKKEYKPSKEIIEGYEADK HIRGDKFDLDFCHKLIDFFKESIEKHKDWSKFNFYFSPTESYGDISEFYLDVEKQGY RMHFENISAETIDEYVEKGDLFLFQIYNKDFVKAATGKKDMHTIYWNAAFSPENLQ DVVVKLNGEAELFYRDKSDIKEIVHREGEILVNRTYNGRTPVPDKIHKKLTDYHNG RTKDLGEAKEYLDKVRYFKAHYDITKDRRYLNDKIYFHVPLTLNFKANGKKNLNK MVIEKFLSDEKAHIIGIDRGERNLLYYSIIDRSGKIIDQQSLNVIDGFDYREKLNQREI EMKD ARQ SWN AIGKIKDLKEG YL SK AVF1EITKMAIQ YN AIVVMEELN YGFKRGRF KVEKQIYQKFENMLIDKMNYLVFKDAPDESPGGVLNAYQLTNPLESFAKLGKQTGI LFYVPAAYTSKIDPTTGFVNLFNTSSKTNAQERKEFLQKFESISYSAKDGGIFAFAFD YRKF GT SKTDF1KNVWT A YTNGERMR YIKEKKRNELFDP SKEIKEALT SSGIKYDGG QNILPDILRSNNNGLIYTMYSSFIAAIQMRVYDGKEDYIISPIKNSKGEFFRTDPKRRE LPIDADANGAYNIALRGELTMRAIAEKFDPDSEKMAKLELKHKDWFEFMQTRGDi: ***RPAATKKAGQAKKKKGSYVYOWOYAYVYOWOYAYVYOWOYA*** (SEQ ID NO: 62)

[00275] 11- Eubacterium eligens (EeCpfl)

[00276] MNGNRSIVYREFVGVIPVAKTLRNELRPVGHTQEHIIQNGLIQEDELRQE K STELKNIMDD YYRE YIDK SL SGVTDLDF TLLFELMNL VQ SSP SKDNKK ALEKEQ S KMREQICTHLQ SD SNYKNIFNAKLLKEILPDFIKNYNQ YD VKDK AGKLETL ALFNG F STYF TDFFEKRKNVF TKEAV ST SI AYRI VHEN SL IFL ANMT SYKKI SEKALDEIE VIE KNNQDKMGDWELNQIFNPDFYNMVLIQSGIDFYNEICGVVNAHMNLYCQQTKNN YNLFKMRKLHKQILAYTSTSFEVPKMFEDDMSVYNAVNAFIDETEKGNIIGKLKDI VNK YDELDEKRI YI SKDF YETL SCFM SGNWNLIT GCVENF YDENIH AK GK SKEEKV KKAVKEDKYKSINDVNDLVEKYIDEKERNEFKNSNAKQYIREISNIITDTETAHLEY

DDHISLIESEEKADEMKKRLDMYMNMYHWAKAFIVDEVLDRDEMFYSDIDDIYNI
LENIVPLYNRVRNYVTQKPYNSKKIKLNFQSPTLANGWSQSKEFDNNAIILIRDNKY
YLAIFNAKNKPDKKIIQGNSDKKNDNDYKKMVYNLLPGANKMLPKVFLSKKGIET
FKPSDYIISGYNAHKHIKTSENFDISFCRDLIDYFKNSIEKHAEWRKYEFKFSATDSY
SDISEFYREVEMQGYRIDWTYISEADINKLDEEGKIYLFQIYNKDFAENSTGKENLH
TMWKNIFSEEM ،KDniKLNGQAELFYRRASVKNPVKHKKDSVLVNfCTYKNQLDN
GDVVRIPIPDDIYNEIYKMYNGYIKE SDL SEAAKEYLDKVE VRTAQKDIVKDYRYT
VDKWIHTPITINYKVTARNNVNDMVVKYIAQNDDIHVIGIDRGERNLIYISVIDSHG
NIVKQKSYNILNNYDYKKKLVEKEKTREYARKNWKSIGNIKELKEGYISGVVHEIA
MLIVEYNAIIAMEDLNYGFKRGRFKVERQVYQKFESMLINKLNYFASKEKSVDEPG
GLLKGYQLTYVPDNIKNLGKQCGVIFYVPAAFTSKIDPSTGFISAFNFKSISTNASRK
QFFMQFDEIRYCAEKDMFSFGFDYNNFDTYNITMGKTQWTVYTNGERLQSEFNNA
RRTGKTKSINLTETIKLLLEDNEINYADGHDIRIDMEKMDEDKKSEFFAQLLSLYKL
TVQMRNSYTEAEEQENGISYDKIISPVINDEGEFFDSDNYKESDDKECKMPKDADA
NGAYCIALKGLYEVLKIKSEWTEDGFDRNCLKLPHAEWLDFIQNKRYEXKP^rXK

^Gg^XXXKGSYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID N0 :63)

[00277] 12- Moraxella bovoculi 237 (MbCpfl)

[00278] MLFQDFTHLYPL SKTVRFELKPIDRTLEHIHAKNFL SQDETMADMHQKV
KVILDDYHRDFIADMMGEVKLTKLAEFYDVYLKFRKNPKDDELQKQLKDLQAVL
RKEIVKPIGNGGKYKAGYDRLFGAKLFKDGKELGDLAKFVIAQEGESSPKLAHLAH
FEKFSTYFTGFHDNRKNMYSDEDKHTAIAYRLIHENLPRFIDNLQILTTIKQKHSALY
DQIINELTASGLDVSLASHLDGYHKLLTQEGITAYNTLLGGISGEAGSPKIQGINELIN
SHHNQHCHKSERIAKLRPLHKQILSDGMSVSFLPSKFADDSEMCQAVNEFYRHYAD
VFAKVQSLFDGFDDHQKDGIYVEHKNLNELSKQAFGDFALLGRVLDGYYVDVVN
PEFNERFAKAKTDNAKAKLTKEKDKFIKGVHSLASLEQAIEHYTARHDDESVQAG
KLGQYFKHGLAGVDNPIQKIHNNHSTIKGFLERERPAGERALPKIKSGKNPEMTQLR
QLKELLDNALNVAHFAKLLTTKTTLDNQDGNFYGEFGVLYDELAKIPTLYNKVRD
YLSQKPFSTEKYKLNFGNPTLLNGWDLNKEKDNFGVILQKDGCYYLALLDKAHKK
VFDNAPNTGKSIYQKMIYKYLEVRKQFPKVFFSKEAIAINYHPSKELVEIKDKGRQR
SDDERLKLYRFILECLKIHPKYDKKFEGAIGDIQLFKKDKKGREVPISEKDLFDKING
IF SSKPKLEMEDFFIGEFKRYNP SQDLVDQYNIYKKID SND NRKKENF YNMIPKFKK
DLVRYYYESMCKHEEWEESFEFSKKLQDIGCYVDVNELFTEIETRRLNYKISFCNIN
ADYIDELVEQGQL YLFQIYNKDF SPKAHGKPNLHTL YFKALF SEDNL ADPIYKLNG

EAQIFYRKASLDMNETTIHRAGEVLENKNPDNPKKRQFVYDIIKDKRYTQDKFMLH
WITMNFGVQGMTIKEFNKKVNQSIQQYDEVNVIGIDRGERHLLYLTVINSKGEILE
QCSLNDITTASANGTQMTTPYHKILDKREIERLNARVGWGEIETIKELKSGYLSHVV
HQISQLMLKYNAIVVLEDLNFGFKRGRFKVEKQIYQNFENALIKKLNHLVLKDKAD
DEIGSYKNALQLTNNFTDLKSIGKQTGFLFYVPAWNTSKIDPETGFVDLLKPRYENI
AQSQAFFGKFDKICYNADKDYFEFHIDYAKFTDKAKNSRQIWTICSHGDKRYVYD
KTANQNKGAAKGINVIWELKSLFARHHINEKQPNLVMDICQNNDKEFHKSLMYLL
KTLLALRYSNASSDEDFILSPVANDEGVFFNSALADDTQPQNADANGAYHIALKGL
WLLNELKNSDDLNKVKLAIDNQTWLNFAQNR*KRPAATKKA****GQAKKKKG***S**YPYDVP**
DYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:64)

[00279] 13- Leptospira inadai (LiCpf1)

[00280] MEDYSGFVNIYSIQKTLRFELKPVGKTLEHIEKKGFLKKDKIRAEDYKAV
KKIIDKYURAYIEEVFDSVLHQKKKKDKTRFSTQFIKEIKEFSELYYKTEKNIPDKER
LEALSEKLRKMLVGAFKGEFSEEVAEKYKNLFSKELIRNEIEKFCETDEERKQVSNF
KSFTTYFTGFHSNRQNIYSDEKKSTAIGYRJIHQNLPKFLDNLKIIESIQRRFKDFPWS
DLKKNLKKIDKNIKLTEYFSIDGFVNVLNQKGIDAYNTILGGKSEESGEKIQGLNEYI
NLYRQKNNIDRKNLPNVKILFKQILGDRETKSFIPEAFPDDQSVLNSITEFAKYLKLD
KKKKSIIAELKKFLSSFNRYELDGIYLANDNSLASISTFLFDDWSFIKKSVSFKYDES
VGDPKKKIKSPLKYEKEKEKWLKQKYYTISFLNDAIESYSKSQDEKRVKIRLEAYFA
EFKSKDDAKKQFDLLERIEEAYAIVEPLLGAEYPRDRNLKADKKEVGKIKDFLDSIK
SLQFFLKPLLSAEIFDEKDLGFYNQLEGYYEEIDSIGIILYNKVRNYLTGKIYSKEKFK
LNFENSTLLKGWDENREVANLCVIFREDQKYYLGVMDKENNTILSDIPKVKPNELF
YEKMVYKLIPTPHMQLPRIIFSSDNLSIYNPSKSILKIREAKSFKEGKNFKLKDCHKFI
DFYKESISKNEDWSRFDFKFSKTSSYENISEFYREVERQGYNLDFKKVSKFYIDSLVE
DGKLYLFQIYNKDFSIFSKGKPNLHTIYFRSLFSKENLKDVCLKLNGEAEMFFRKKSI
NYDEKKKREGHHPELFEKLKYPILKDKRYSEDKFQFHLPISLNFKSKERLNFNLKVN
EFLKRNKDINIIGIDRGERNLLYLVMINQKGEILKQTLLDSMQSGKGRPEINYKEKL
QEKEIERDKARKSWGTVENIKELKEGYLSIVIHQISKLMVENNAIVVLEDLNIGFKR
GRQKVERQVYQKFEKMLIDKLNFLVFKENKPTEPGGVLKAYQLTDEFQSFEKLSK
QTGFLFYVPSWNTSKIDPRTGFIDFLHPAYENIEKAKQWINKFDSIRFNSKMDWFEF
TADTRKFSENLMLGKNRVWVICTTNVERYFTSKTANSSIQYNSIQITEKLKELFVDIP
FSNGQDLKPEILRKNDAVFFKSLLFYIKTTLSLRQNNGKKGEEEKDFILSPVVDSKG
RFFNSLEASDDEPKDADANGAYHIALKGLMNLLVLNETKEENLSRPKWKIKNKDW

LEFVWERNR*KRPAATKKAGQAKX KK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA** (SEQ ID NO:65)

[00281] 14- Lachnospiraceae bacterium ND2006 (LbCpfl)

[00282] MSKLEKFTNCYSLSKTLRFKAIPVGKTQENIDNKRLLVEDEKRAEDYKG VKKLLDRYYLSFINDVLHSIKLKNLNNYISLFRKKTRTEKENKELEM.EINLRKEIAK AFKGNEGYKSLFKKDIIETILPEFLDDKDEIALVNSFNGFTTAFTGFFDNRENMFSEE AKSTSIAFRCINENLTRYISNMDIFEKVDAIFDKHEVQEIKEKILNSDYDVEDFFEGEF FNFVLTQEGIDVYNAIIGGFVTESGEKIKGLNEYINLYNQKTKQKLPKFKPLYKQVL SDRESL SFYGEGYT SDEEVLEVFRNTLNKNSEIF SSIKKLEKLFKNFDE Y SSAGIFVK NGP AI STI SKDIF GEWN VIRDKWN AE YDDIF1LKKKAVVTEK YEDDRRK SFKKIGSF S LEQLQEYADADLSVVEKLKEIIIQKVDEIYKVYGSSEKLFDADFVLEKSLKKNDAV VAIMKDLLDSVKSFENYIKAFFGEGKETNRDESFYGDFVLAYDILLKVDHIYDAIRN YVTQKPYSKDKFKLYFQNPQFMGGWDKDKETDYRATILRYGSKYYLAIMDKKYA KCLQKIDKDD VNGNYEKINYKLLPGPNKMLPKVFF SKKWMAYYNP SEDIQKIYKN GTFKKGDMFNLNDCHKLIDFFKDSISRYPKWSNAYDFNFSETEKYKDIAGFYREVE EQGYKVSFESASKKEVDKLVEEGKLYMFQIYNfCDFSDKSHGTPNLHTMYFKLLFD ENNHGQIRLSGGAELFMRRASLKKEELVVHPANSPIANKNPDNPKKTTTLSYDVYK DKRFSEDQYELHIPIAINKCPKNIFKINTEVRVLLKHDDNPYVIGIDRGERNLLYIVV VDGKGNIVEQYSLNEIINNFNGIRIKTDYHSLLDKKEKERFEARQNWTSIENIKELKA GYISQVVHKICELVEKYDAVIALEDLNSGFKNSRVKVEKQVYQKFEKMLIDKLNY MVDKKSNPCATGGALKGYQITNKFESFKSMSTQNGFIFYIPAWLTSKIDPSTGFVNL LKTK YTSIAD SKKFIS SFDRIMYVPEEDLFEF ALDYKNF SRTDADYIKKWKL Y SYGN RIRIFRNPKKNNWDWEE VCLT SAYKELFNKYGINYQQGDIRALLCEQ SDKAF Y SSF MALMSLMLQMRNSITGRTDVDFLISPVKNSDGIFYDSRNYEAQENAILPKNADANG AYNIARKVLWAIGQFKK AEDEKLDK VKIAI SNKEWLEYAQ*TSVKRKRPAA TKKAGQ AKKKK*GS**YPYI) VPI)** Y **AYPYI) VPI)** Y **AYPYI) VPI)** Y Λ (SEQ ID NO: 66)

[00283] 15- Porphyromonas crevioricanis (PcCpfl)

[00284] MDSLKDFTNLYPVSKTLRFELKPVGKTLENIEKAGILKEDEHRAESYRRV KKIIDTYHKVFIDSSLENMAKMGIENEIKAMLQSFCELYKKDHRTEGEDKALDKIRA VLRGLIVGAFTGVCGRRENTVQNEKYESLFKEKLIKEILPDFVLSTEAESLPFSVEEA TRSLKEFDSFTSYFAGFYENRKNIYSTKPQSTAIAYRLIHENLPKFIDNILVFQKIKEPI AKELEHIRADFSAGGYIKKDERLEDIFSLNYYIHVLSQAGIEKYNALIGKIVTEGDGE MKGLNEHINL YNQQRGREDRLPLFRPL YKQIL SDREQL SYLPESFEKDEELLRALKE

FYDHIAEDILGRTQQLMTSISEYDLSRIYVRNDSQLTDISKKMLGDWNAIYMARER
AYDHEQAPKRITAKYERDRIKALKGEESISLANLNSCIAFLDNVRDCRVDTYLSTLG
QKEGPHGLSNLVENVFASYHEAEQLLSFPYPEENNLIQDKDNVVLIKNLLDNISDLQ
RFLKPLWGMGDEPDKDERFYGEYNYIRGALDQVIPLYNKVRNYLTRKPYSTRKVK
LNFGNSQLLSGWDRNKEKDNSCVILRKGQNFYLAIMNNRHKRSFENKMLPEYKEG
EPYFEKMDYKFLPDPNKMLPKVFLSKKGIEIYKPSPKLLEQYGHGTHKKGDTFSMD
DLHELIDFFKHSIEAHEDWKQFGFKFSDTATYENVSSFYREVEDQGYKLSFRKVSES
YVYSLIDQGKLYLFQIYNKDFSPCSKGTPNLHTLYWRMLFDERNLADVIYKLDGKA
EIFFREK SLKNDHPTF1PAGKPIKKK SRQKKGEESLFE YDL VKDRR YTMDKF QFHVPI
TMNFKCSAGSKVNDMVNAHIREAKDMHVIGIDRGERNLLYICVIDSRGTILDQISLN
TINDIDYHDLLESRDKDRQQEHRNWQTIEGIKELKQGYLSQAVHRIAELMVAYKAV
VALEDLNMGFKRGRQKVESSVYQQFEKQLIDKLNYLVDKKKRPEDIGGLLRAYQF
TAPFKSFKEMGKQNGFLFYIPAWNTSNIDPTTGFVNLFHVQYENVDKAKSFFQKFD
SISYNPKKDWFEFAFDYKNFTKKAEGSRSMWILCTHGSRIKNFRNSQKNGQWDSEE
FALTEAFKSLFVRYEIDYTADLKTAIVDEKQKDFFVDLLKLFKLTVQMRNSWKEKD
LDYLISPVAGADGRFFDTREGNKSLPKDADANGAYNIALKGLWALRQIRQTSEGGK
LKLAISNKEWLQFVQERSYEKDX *RPAATKK¾GQAK*X *KK***GS**SYPYDVPDYAYPYDVP
DYAYPYDVPDYA (SEQ ID NO:67)

[00285] 16- Prevotella disiens (PdCpfl)

[00286] MENYQEFTNLFQLNKTLRFELKPIGKTCELLEEGKIFASGSFLEKDKVRA
DNVS YVKKEIDKKHKIFIEETLS SF SISNDLLKQYFDCYNELKAFKKDCKSDEEEVK
KTALRNKCT SIQRAMRE AI SQAFLK SPQKKLL AIKNLIENVFK ADEN VQHF SEF TSY
FSGFETNRENFYSDEEKSTSIAYRLVHDNLPIFIKNIYIFEKLKEQFDAKTLSEIFENY
KLYVAGS SLDEVF SLEYFNNTLTQKGIDN YNAVIGKIVKEDKQEIQGLNEHINL YNQ
KHKDRRLPFFISLKKQILSDREALSWLPDMFKNDSEVIKALKGFYIEDGFENNVLTP
LATLLS SLDKYNLNGIFIRNNEALS SLSQNVYRNF SIDEAIDANAELQTFNNYELIAN
ALRAKIKKETKQGRK SFEKYEE YIDKKVK AID SL SIQEINELVENYVSEFNSNSGNM
PRKVEDYFSLMRKGDFGSNDLIENIKTKLSAAEKLLGTKYQETAKDIFKKDENSKLI
KELLDATKQFQHFIKPLLGTGEEADRDLVFYGDFLPLYEKFEELTLLYNKVRNRLT
QKPYSKDKIRLCFNKPKLMTGWVDSKTEKSDNGTQYGGYLFRKKNEIGEYDYFLG
ISSKAQLFRKNEAVIGDYERLDYYQPKANTIYGSAYEGENSYKEDKKRLNKVIIAYI
EQIKQTNIKKSIIESISKYPNISDDDKVTPSSLLEKIKKVSIDSYNGILSFKSFQSVNKE
VIDNLLKTISPLKNKAEFLDLINKDYQIFTEVQAVIDEICKQKTFIYFPISNVELEKEM

GDKDKPLCLFQISNKDLSFAKTFSANLRKKRGAENLHTMLFKALMEGNQDNLDLG
SGAIFYRAKSLDGNKPTHPANEAIKCRNVANKDKVSLFTYDIYKNRRYMENKFLFH
L SIVQNYKAAND SAQLNS SATEYIRK ADDLHIIGIDRGERNLL YYS VIDMKGNIVEQ
D SLNIIRNNDLETDYFIDLLDKREKERKANRQNWE AVEGIKDLKKGYL SQAVHQI A
QLMLKYNAIIALEDLGQMFVTRGQKIEKAVYQQFEKSLVDKLSYLVDKKRPYNEL
GGILKAYQLASSITKNNSDKQNGFLFYVPAWNTSKIDPVTGFTDLLRPKAMTIKEA
QDFF GAFDNI SYNDKGYFEFETN YDKFKIRMK SAQTRWTICTF GNRIKRKKDKN Y
WNYEEVELTEEFKKLFKDSNIDYENCNLKEEIQNKDNRKFFDDLIKLLQLTLQMRN
SDDKGNDYIISPVANAEGQFFDSRNGDKKLPLDADANGAYNIARKGLWNIRQIKQT
KNDKKLNL SI SSTEWLDF VREKPYLK*KRPAATKKAGQAKKXK*G S**YPYDVPDYAYPY**
DVPDYAYPYDVPDYA (SEQ ID NO:68)

[00287] 17- Porphyromonas macacae (PmCpfl)

[00288] MKTQHFFEDFTSLYSLSKTIRFELKPIGKTLENIKKNGLIRRDEQRLDDYE
KLKK VIDEYHEDFI ANIL SSF SF SEEILQ SYIQNL SESEARAKIEKTMRDTL AKAF SED
ERYK SIFKKEL VKKDIP VWCP AYK SLCKKFDNF TT SL VPFHENRKNL YTSNEITASIP
YRIVHVNLPKFIQNIEALCELQKKMGADLYLEMMENLRNVWPSFVKTPDDLCNLK
TYNHLMVQSSISEYNRFVGGYSTEDGTKHQGrNEWrNIYRQRNKEMRLPGLVFLHK
QILAKVDSSSFISDTLENDDQVFCVLRQFRKLFWNTVSSKEDDAASLKDLFCGLSGY
DPEAIYVSDAHLATISKNIFDRWNYISDAIRRKTEVLMPRKKESVERYAEKISKQIKK
RQSYSLAELDDLLAHYSEESLPAGFSLLSYFTSLGGQKYLVSDGEVILYEEGSNIWD
EVLIAFRDLQVILDKDFTEKKLGKDEEAVSVIKKALDSALRLRKFFDLLSGTGAEIR
RDSSFYALYTDRMDKLKGLLKMYDKVRNYLTKKPYSIEKFKLHFDNPSLLSGWDK
NKELNNLSVIFRQNGYYYLGFMTPKGKNLFKTLPKLGAEEMFYEKMEYKQIAEPM
LMLPKVFFPKKTKPAFAPDQSVVDIYNKKTFKTGQKGFNKKDLYRLIDFYKEALTV
HEWKLFNF SF SPTEQYRNIGEFFDEVREQ AYKVSMVNVPAS YIDEAVENGKLYLFQ
IYNKDFSPYSKGIPNLHTLYWKALFSEQNQSRVYKLCGGGELFYRKASLHMQDTTV
HPKGISIHKKNLNKKGETSLFNYDLVKDKRFTEDKFFFHVPISrNYKNKKITNVNQM
VRDYIAQNDDLQIIGIDRGERNLLYISRIDTRGNLLEQFSLNVIESDKGDLRTDYQKI
LGDREQERLRRRQEWK SIE SnCDLKDGYM SQVVFIKICNMVVEFIKAIVVLENLNL SF
MKGRKKVEKSVYEKFERMLVDKLNYLVVDKKNLSNEPGGLYAAYQLTNPLFSFE
ELHRYPQSGILFFVDPWNTSLTDPSTGFVNLLGRINYTNVGDARKFFDRFNAIRYDG
KGNILFDLDL SRFDVRVETQRKL WTLTTF GSRIAK SKK SGKWM VERIENL SLCFLEL
FEQFNIGYRVEKDLKK AIL SQDRKEFYVRLIYLFNLMMQIRNSD GEEDYIL SP ALNE

KNLQFDSRLIEAKDLPVDADANGAYNVARKGLMVVQRIKRGDHESIHRIGRAQWL
R*Y* VQ*EGI**VEKRPAA TKKAGQAKKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA**
(SEQ ID NO: 69)

[00289] 18- Thiomicrospira sp. XS5 (TsCpfl)

[00290] MTKTFDSEFFNLYSLQKTVRFELKPVGETASFVEDFKNEGLKRVVSEDE
RRAVDYQKVKEIIDDYHRDFIEESLNYFPEQVSKDALEQAFHLYQKLKAAKVEERE
KALKEWEALQKKLREKVVKCFSDSNKARFSRIDKKELIKEDLINWLVAQNREDDIP
TVETFNNFTTYFTGFHENRKNIYSKDDHATAISFRLIHENLPKFFDNVISFNKLKEGF
PELKFDKVKEDLEVDYDLKHAFEIEYFVNFVTQAGIDQYNYLLGGKTLEDGTKKQ
GMNEQINLFKQQQTRDKARQIPKLIPLFKQILSERTESQSFIPKQFESDQELFDSLQKL
HNNCQDKFTVLQQAILGLAEADLKKVFIKTSDLNALSNTIFGNYSVFSDALNLYKES
LKTKKAQEAFEKLPAHSIHDLIQYLEQFNSSLDAEKQQSTDTVLNYFIKTDELYSRFI
KSTSEAFTQVQPLFELEALSSKRRPPESEDEGAKGQEGFEQIKRIKAYLDTLMEAVH
FAKPLYLVKGRKMIEGLDKDQSFYEAFEMAYQELESLIIPIYNfCARSYLSRKPFKAD
KFKINFDNNTLLSGWDANKETANASILFKKDGLYYLGIMPKGKTFLFDYFVSSEDS
EKLKQRRQKTAEEALAQDGESYFEKIRYKLLPGASKMLPKVFFSNKNIGFYNPSDDI
LRIRNTASHTKNGTPQKGHSKVEFNLNDCHKMIDFFKSSIQKHPEWGSFGFTFSDTS
DFEDMSAFYREVENQGYVISFDKIKETYIQSQVEQGNLYLFQIYNKDFSPYSKGKPN
LHTLYWKALFEEANLNNVVAKLNGEAEIFFRRHSIKASDKVVHPANQAIDNKNPHT
EKTQSTFEYDLVKDKRYTQDKFFFHVPISLNFKAQGVSKFNDKVNGFLKGNPDVNI
IGIDRGERHLLYFTVVNQKGEILVQESLNTLMSDKGHVNDYQQKLDKKEQERDAA
RKSWTTVENIKELKEGYLSHVVHKLAHLIIKYNAIVCLEDLNFGFKRGRFKVEKQV
YQKFEKALIDKLNYLVFKEKELGEVGHYLTAYQLTAPFESFKKLGKQSGILFYVPA
DYTSKIDPTTGFVNFLDLRYQSVEKAKQLLSDFNAIRFNSVQNYFEFEIDYKKLTPK
RKVGTQSKWVICTYGDVRYQNRRNQKGHWETEEVNVTEKLKALFASDSKTTTVID
YANDDNLIDVILEQDKASFFKELLWLLKLTMTLRHSKIKSEDDFILSPVKNEQGEFY
DSRKAGEVWPKDADANGAYHIALKGLWNLQQINQWEKGKTLNLAIKNQDWFSFI
Q*EKPY Q**EKRPAA TKKAGQAKKKK*GS**YPYDVPDYAYPYDVPDYAYPYDVPDYA**
(SEQ ID NO: 70)

[00291] 19- Moraxella bovoculi AAX08_00205 (Mb2Cpfl)

[00292] MLFQDFTHLYPLSKTVRFELKPIGRTLEHIHAKNFLSQDETMADMYQKV
KVILDDYHRDFIADMMGEVKLTKLAEFYDVYLKFRKNPKDDGLQKQLKDLQAVL
RKESVKPIGSGGKYKTGYDRLFGAKLFKDGKELGDLAKFVIAQEGESSPKLAHLAH

FEKFSTYFTGFHDNRKNMYSDEDKHTAIAYRLIHENLPRFIDNLQILTTIKQKHSALY
DQIINELTASGLDVSLASHLDGYHKLLTQEGITAYNRIIGEVNGYTNKHNQICHKSE
RIAKLRPLFDCQILSDGMGVSFLPSKFADDSEMCQAVNEFYRHYTDVFAKVQSLFDG
FDDHQKDGIYVEHKNLNELSKQAFGDFALLGRVLDGYYVDVVNPEFNERFAKAKT
DNAKAKLTKEKDKFIKGVHSLASLEQAIEHHTARF1DDESVQAGKLGQYFKHGLAG
VDNPIQKIHNNHSTIKGFLERERPAGERALPKIKSGKNPEMTQLRQLKELLDNALNV
AHFAKLLTTKTTLDNQDGNFYGEFGVLYDELAKIPTLYNKVRDYLSQKPFSTEKYK
LNFGNPTLLNGWDLNKEKDNFGVILQKDGCYYLALLDKAHKKVFDNAPNTGKNV
YQKMVYKLLPGPNKMLPKVFFAKSNLDYYNPSAELLDKYAKGTFIKKGDNFNLKD
CHALIDFFKAGFNKFlPEWQHFGFKFSPTSSYRDLSDFYREVEPQGYQVKFVDrNAD
YIDELVEQGKLYLFQIYNKDFSPKAHGKPNLHTLYFKALFSEDNLADPIYKLNGEA
QIFYRKASLDMNETTIHRAGEVLENKNPDNPKKRQFVYDIIKDKRYTQDKFMLHVP
ITMNFGVQGMTIKEFNKKVNQSIQQYDEVNVIGIDRGERHLLYLTVINSKGEILEQR
SLNDITTASANGTQVTTPYFIKILDKREIERLNARVGWGEIETIKELKSGYLSHVVHQI
NQLMLKYNAIVVLEDLNFGFKRGRFKVEKQIYQNFENALIKKLNHLVLKDKADDEI
GSYKNALQLTNNFTDLKSIGKQTGFLFYVPAWNTSKIDPETGFVDLLKPRYENIAQS
QAFFGKFDKICYNTDKGYFEFHIDYAKFTDKAKNSRQKWAICSHGDKRYVYDKTA
NQNKGAAKGFNVNDELKSLFARYHFNDKQPNLVMDICQNNDKEFFIKSLMCLLKTL
LALRYSNASSDEDFILSPVANDEGVFFNSALADDTQPQNADANGAYHIALKGLWLL
NELKNSDDLNKVKLAIDNQTWLNFAQNR*¾ PAATKKAGQAKKKK*GS**YPYDVPDYA**
YPYDVPDYAYPYDVPDYA (SEQ ID N0 :71)

[00293] 20- Moraxella bovoculi AAX1 1_00205 (Mb3Cpfl)

[00294] MLFQDFTULYPLSKTVRFELKPIGKTLEHIHAKNFLNQDETM ADMYQKV
KAILDDYHRDFIADMMGEVKLTKLAEFYDVYLKFRKNPKDDGLQKQLKDLQAVL
RKEIVKPIGNGGKYKAGYDRLFGAKLFKDGKELGDLAKFVIAQEGESSPKLAHLAH
FEKFSTYFTGFFIDNRKNMYSDEDKHTAIAYRLIHENLPRFIDNLQILATIKQKHSAL
YDQirNELTASGLDVSLASHLDGYFKLLTQEGITAYNTLLGGISGEAGSRKIQGrNEL
nSTSHFlNQHCFiKSERIAKLRPLFiKQILSDGMGVSFLPSKFADDSEVCQAVNEFYRHY
ADVFAKVQSLFDGFDDYQKDGIYVEYKNLNELSKQAFGDFALLGRVLDGYYVDV
VNPEFNERFAKAKTDNAKAKLTKEKDKFIKGVHSLASLEQAIEHYTARFIDDESVQ
AGKLGQYFKHGLAGVDNPIQKIHNNHSTIKGFLERERPAGERALPKIKSDKSPEIRQ
LKELLDNALNVAHFAKLLTTKTTLHNQDGNFYGEFGALYDELAKIATLYNKVRDY
LSQKPFSTEKYKLNFGNPTLLNGWDLNKEKDNFGVILQKDGCYYLALLDKAFIKKV

FDNAPNTGKSVYQKMIYKLLPGPNKMLPKVFFAKSNLDYYNPSAELLDKYAQGTH
KKGDNFNLKDCHALIDFFKAGINKHPEWQHFGFKFSPTSSYQDLSDFYREVEPQGY
QVKFVDINADYINELVEQGQLYLFQIYNKDFSPKAHGKPNLHTLYFKALFSEDNLV
NPIYKLNGEAEIFYRKASLDMNETTIHRAGEVLENKNPDNPKKRQFVYDIIKDKRYT
QDKFMLHVPITMNFGVQGMTIKEFNKKVNQSIQQYDEVNVIGIDRGERHLLYLTVI
NSKGEILEQRSLNDITTASANGTQMTTPYHKILDKREIERLNARVGWGEIETIKELKS
GYLSHVVHQISQLMLKYNAIVVLEDLNFGFKRGRFKVEKQIYQNFENALIKKLNHL
VLKDKADDEIGSYKNALQLTNNFTDLKSIGKQTGFLFYVPAWNTSKIDPETGFVDL
LKPRYENIAQSQAFFGKFDKICYNADRGYFEFHIDYAKFNDKAKNSRQIWKICSHG
DKRYVYDKTANQNKGATIGVNVNDELKSLFTRYHINDKQPNLVMDICQNNDKEFH
KSLMYLLKTLLALRYSNASSDEDFILSPVANDEGVFFNSALADDTQPQNADANGAY
HIALKGLWLLNELKNSDDLNKVKLAIDNQTWLNFAQNR*KRPAATKKAGQAKKKKG*
SYPYDVPDYAYPYDVPDYAYPYDVPDYA (SEQ ID NO:72)

[00295] 21- Butyrivibrio sp. NC3005 (BsCpfl)

[00296] MYYQNLTKKYPVSKTIRNELIPIGKTLENIRKNNILESDVKRKQDYEHVK
GIMDEYHKQLINEALDNYMLPSLNQAAEIYLKKHVDVEDREEFKKTQDLLRREVT
GRLKEHENYTKIGKKDILDLLEKLPSISEEDYNALESFRNFYTYFTSYNKVRENLYS
DEEKSSTVAYRLINENLPKFLDNIKSYAFVKAAGVLADCIEEEEQDALFMVETFNM
TLTQEGIDMYNYQIGKVNSAINLYNQKNHKVEEFKKIPKMKVLYKQILSDREEVFI
GEFKDDETLLSSIGAYGNVLMTYLKSEKINIFFDALRESEGKNVYVKNDLSKTTMS
NIVFGSWSAFDELLNQEYDLANENKKKDDKYFEKRQKELKKNKSYTLEQMSNLSK
EDISPIENYIERISEDIEKICIYNGEFEKIVVNEHDSSRKLSKNIKAVKVIKDYLDSIKEL
EHDIKLINGSGQELEKNLVVYVGQEEALEQLRPVDSLYNLTRNYLTKKPFSTEKVK
LNFNKSTLLNGWDKNKETDNLGILFFKDGKYYLGFMNTTANKAFVNPPAAKTENV
FKKVDYKLLPGSNKMLPKVFFAKSNIGYYNPSTELYSNYKKGTHKKGPSFSIDDCH
NLIDFFKESIKKHEDWSKFGFEFSDTADYRDISEFYREVEKQGYKLTFTDIDESYIND
LIEKNELYLFQIYNKDFSEYSKGKLNLHTLYFMMLFDQRNLDNVVYKLNGEAEVF
YRPASIAENELVIHKAGEGIKNKNPNRAKVKETSTFSYDIVKDKRYSKYKFTLHIPIT
MNFGVDEVRRFNDVINNALRTDDNVNVIGIDRGERNLLYVVVINSEGKILEQISLNS
IINKEYDIETNYHALLDEREDDRNKARKDWNTIENIKELKTGYLSQVVNVVAKLVL
KYNAIICLEDLNFGFKRGRQKVEKQVYQKFEKMLIEKLNYLVIDKSREQVSPEKMG
GALNALQLTSKFKSFAELGKQSGIIYYVPAYLTSKIDPTTGFVNLFYIKYENIEKAKQ
FFDGFDFIRFNKKDDMFEFSFDYKSFTQKACGIRSKWIVYTNGERIIKYPNPEKNNLF

DEKVINVTDEIKGLFKQYRIPYENGEDIKEIIISKAEADFYKRLFRLLHQTLQMRNST SDGTRDYIISPVKNDRGEFFCSEFSEGTMPKDADANGAYNIARKGLWVLEQIRQKD EGEKVNLSMTNAEWLKYAQLHLL*KRPAATKKAGQAK*X *KK*GS**YPYDVPDYAYPYD VPDYAYPYDVPDYA** (SEQ ID N0:73)

[00297] Further Cpfl orthologs include:

[00298] NCBI WP_055225123.1

[00299] MNNGTNNFQNFIGIS SLQKTLRNALIPTETTQQFIVKNGIIKEDELRGENR QILKDIMDDYYRGFISETLSSIDDIDWTSLFEKMEIQLKNGDNKDTLIKEQTEYRKAI HKKFANDDRFKNMFSAKLISDILPEFVIHNNNYSASEKEEKTQVIKLFSRFATSFKD YFKNRANCFSADDISSSCHRIVNDNAEIFFSNALVYRRIVKSLSNDDINKISGDMKD SLKEMSLEEIYSYEKYGEFITQEGISFYNDICGKVNSFMNLYCQKNKENKNLYKLQ KLHKQILCIADTSYEVPYKFESDEEVYQSVNGFLDNISSKHIVERLRKIGDNYNGYN LDKIYIVSKFYESVSQKTYRDWETINTALEIHYNNILPGNGKSKADKVKKAVKNDL QKSITEINELVSNYKLCSDDNIKAETYIHEISHILNNFEAQELKYNPEIHLVESELKAS ELKNVLDVIMNAFHWCSVFMTEELVDKDNNFYAELEEIYDEIYPVISLYNLVRNYV TQKPYSTKKIKLNFGIPTLADGWSKSKEYSNNAIILMRDNLYYLGIFNAKNKPDKKII EGNTSENKGDYKKMIYNLLPGPNKMIPKVFLSSKTGVETYKPSAYILEGYKQNKHI KSSKDFDITFCHDLID YFKNCIAIHPEWKNFGFDF SDTSTYEDISGF YREVELQGYKI DWTYISEKDIDLLQEKGQLYLFQIYNKDF SKKSTGNDNLHTMYLKNLF SEENLKDI VLKLNGEAEIFFRKSSIKNPIIHKKGSILVNRTYEAEEKDQFGNQIVRKNIPENIYQEL YKYFNDKSDKELSDEAAKLKNVVGHHEAATNIVKDYRYTYDKYFLHMPITINFKA NKTGFINDRILQYIAKEKDLHVIGIDRGERNLIYVSVIDTCGNIVEQKSFNIVNGYDY QIKLKQQEGARQIARKEWKEIGKIKEIKEGYL SLVIHEISKM VIKYNAII AMEDL SYG FKKGRFKVERQVYQKFETMLINKLNYLVFKDISITENGGLLKGYQLTYIPDKLKNV GHQCGCIFYVPAAYTSKIDPTTGFVNIFKFKDLTVDAKREFIKKFDSIRYDSEKNLFC FTFDYNNFITQNT VM SK SSW SVYTYGVRIKRRF VNGRF SNE SD TIDITKDMEKTLE MTDINWRD GHDLRQDIID YEI VQHIFEIFRLT VQMRNSL SELEDRD YDRLI SP VLNEN NIFYDSAKAGDALPKDADANGAYCIALKGLYEIKQITENWKEDGKFSRDKLKISNK DWFDFIQNKRYL (SEQ ID NO:74)

[00300] NCBI WP_055237260.1

[00301] MNNGTNNFQNFIGISSLQKTLRNALIPTETTQQFIVKNGIIKEDELRGENR QILKDIMDDYYRGFISETLSSIDDIDWTSLFEKMEIQLKNGDNKDTLIKEQAEKRKAI YKKFADDDRFKNMFSAKLISDILPEFVIHNNNYSASEKEEKTQVIKLFSRFATSFKD

YFKNRANCFSADDISSSCHRIVNDNAEIFFSNALVYRRIVKNLSNDDINKISGDMK
DSLKEMSLDEIYSYEKYGEFITQEGISFYNDICGKVNSFMNLYCQKNKENKNLYKL
RKLHKQILCIADTSYEVPYKFESDEEVYQSVNGFLDNISSKHIVERLRKIGDNYNGY
NLDKIYIVSRFYESVSQKTYRDWETrNTALEIHYNNILPGNGKSKADKVKKAVKND
LQKSITErNELVSNYKLCPDDNIKAETYIHEISHILNNFEAQELKYNPEIHLVESELKA
SELKNVLDVFMNAFHWCSVFMTEELVDKDNNFYAELEEIYDEIYPVISLYNLVRNY
VTQKPYSTKKIKLNFGIPTLADGWSKSKEYSNNAIILMRDNLYYLGIFNAKNKPDK
KIIEGNTSENKGDYKKMIYNLLPGPNKMIPKVFLSSKTGVETYKPSAYILEGYKQNK
HLKSSKDFDITFCRDLIDYFKNCIAIHPEWKNFGFDFSDTSTYEDISGFYREVELQGY
KIDWTYISEKDIDLLQEKGQLYLFQIYNKDFSKKSTGNDNLHTMYLKNLFSEENLK
DIVLKLNGEAEIFFRKSSIKNPnHKKGSILVNRTYEAEEKDQFGNIQIVRKTIPENIYQ
ELYKYFNDKSDKELSDEAAKLKNVVGHHEAATNIVKDYRYTYDKYFLHMPITrNF
KANKTSFrNDRILQYIAKENDLHVIGIDRGERNLIYVSVIDTCGNIVEQKSFNIVNGY
DYQIKLKQQEGARQIARKEWKEIGKIKEIKEGYLSLVIHEISKMVIKYNAIIAMEDLS
YGFKKGRFKVERQVYQKFETMLINKLNYLVFKDISITENGGLLKGYQLTYIPEKLK
NVGHQCGCIFYVPAAYTSKIDPTTGFANIFKFKDLTVDAKREFIKKFDSIRYDSEKNL
FCFTFDYNNFITQNTVMSKSSWSVYTYGVRIKRRFVNGRFSNESDTIDITKDMEKTL
EMTDINWRDGHDLRQDIIDYEIVQHIFEIFKLTVQMRNSLSELEDRDYDRLISPVLNE
NNIFYDSAKAGDALPKDADANGAYCIALKGLYEIKQITENWKEDGKFSRDKLKISN
KDWFDFIQNKRYL (SEQ ID NO: 75)

[00302] NCBI WP_055272206.1

[00303] MNNGTNNFQNFIGISSLQKTLRNALTPTETTQQFIVKNGIIKEDELRGENR
QILKDIMDDYYRGFISETLSSIDDIDWTSLFEKMEIQLKNGDNKDTLIKEQAEKRKAI
YKKFADDDRFKNMFSAKLISDILPEFVIHNNNYSASEKEEKTQVIKLFSRFATSFKD
YFKNRANCFSADDISSSCHRIVNDNAEIFFSNALVYRRIVKNLSNDDrNKISGDMK
DSLKKMSLEKIYSYEKYGEFITQEGISFYNDICGKVNSFMNLYCQKNKENKNLYKL
RKLHKQILCIADTSYEVPYKFESDEEVYQSVNGFLDNISSKHIVERLRKIGDNYNGY
NLDKIYIVSKFYESVSQKTYRDWETrNTALEIHYNNILPGNGKSKADKVKKAVKND
LQKSITErNELVSNYKLCPDDNIKAETYIHEISHILNNFEAQELKYNPEIHLVESELKA
SELKNVLDVFMNAFHWCSVFMTEELVDKDNNFYAELEEIYDEIYPVISLYNLVRNY
VTQKPYSTKKIKLNFGIPTLADGWSKSKEYSNNAIILMRDNLYYLGIFNAKNKPEKK
IIEGNTSENKGDYKKMIYNLLPGPNKMIPKVFLSSKTGVETYKPSAYILEGYKQNKH
LKSSKDFDITFCRDLIDYFKNCIAIHPEWKNFGFDFSDTSTYEDISGFYREVELQGYKI

DWTYISEKDIDLLQEKGQL YLFQIYNKDF SKK STGNDNLHTMYL KNL F SEENLKDV VLKLNGEAEIFFRKSSIKNPnHKKGSILVNRTYEAEEKDQFGNIQIVRKTIPENIYQEL YKYFNDKSDKELSDEAAKLKNAVGHHEAATNIVKDYRYTYDKYFLHMPITINFKA NKTSFINDRILQYIAKEKDLHVIGIDRGERNLIYVSVIDTCGNIVEQKSFNIVNGYDY QIKLKQQEGARQIARKEWKEIGKIKEIKEGYL SL VIHEISKM VIKYNAII AMEDL SYG FKKGRFKVERQVYQKFETMLINKLNYL VFKDISITENGGLLKGYQLTYIPEKLKNV GHQCGCIFYVPAAYTSKIDPTTGFVNIFKFKDLTVDAKREFIKKFDSIRYDSDKNLFC F TFD YNNFITQNT VM SK SSW SVY T YGVRIKRRF VNGRF SNE SD TIDITKDMEKTLE MTDINWRDGHDLRQDIIDYEIVQHIFEIFKLTVQMRNSLSELEDRNYDRLISPVLNEN NIFYDSAKAGDALPKDADANGAYCIALKGLYEIKQITENWKEDGKFSRDKLKISNK DWFDFIQNKRYL (SEQ ID NO: 76)

[00304] GenBank OLA16049. 1

[00305] MNNGTNNFQNFIGIS SLQKTLRNALIPTETTQQFIVKNGIIKEDELRGKNR QILKDIMDDYYRGFISETLSSIDDIDWTSLFEKMEIQLKNGDNKDTLIKEQAEKRKAI YKKF ADDDRFKNMF SAKLISDILPEF VIHNNNYS ASEKKEKTQ VIKLF SRF ATSFKD YFKNRANCF SADDIS SSSCURI VNDN AEIFF SN AL VYRRI VKNL SNDDINKI SGDMK DSLKEMSLEEIYSYEKYGEFITQEGISFYNDICGKVNSFMNLYCQKNKENKNLYKLR KLHKQILCIADTSYEVPYKFESDEEVYQSVNGFLDNISSKHIVERLRKIGDNYNDYN LDKIYIVSKFYESVSQKTYRDWETINTALEIHYNNILPGNGKSKADKVKKAVKNDL QKSITEINELVSNYKLCSDDNIKAETYIHEISHILNNFEAHELKYNPEIHLVESELKAS ELKNVLDnMNAFHWCSVFMTEELVDKDNNFYAELEEIYDEIYPVISLYNLVRNYVT QKPYSTKKIKLNFGIPTLADGWSKSKEYSNNAIILMRDNLYYLGIFNAKNKPDKKIIE GNT SENKGDYKKMI YNLLPGPNKMIPKVFL SSKTGVETYKP SAYILEGYKQNKULK SSKDFDITFCHDLID YFKNCIAIHPEWKNFGFDF SDTS AYEDISGF YREVELQGYKID WTYISEKDIDLLQEKGQLYLFQIYNKDFSKKSTGNDNLHTMYLKNLFSEENLKDIV LKLNGEAEIFFRKSSIKNPIIHKKGSILVNRTYEAEEKDQFGNIQIVRKTIPENIYQELY KYFNDKSDKELSDEAAKLKNVVGHHEAATNIVKDYRYTYDKYFLHMPITINFKAN KTSFINDRILQYIAKEKDLHVIGIDRGERNLIYVSVIDTCGNIVEQKSFNIVNGYDYQI KLKQQEGARQI ARKEWKEIGKIKEIKEGYL SL VIHEISKM VIKYNAII AMEDL SYGFK KGRFKVERQVYQKFETMLINKLNYL VFKDISITENGGLLKGYQLTYIPDKLKNVGH QCGCIFYVPAAYTSKIDPTTGFVNIFKFKDLTVDAKREFIKKFDSIRYDSEKNLFCFT FDYNNFITQNT VMSKS SWSVYTYGVRIKRRF VNGRF SNESDTIDITKDMEKTLEMT DINWRDGHDLRQDIIDYEIVQHIFEIFKLTVQMRNSLSELEDRDYDRLISPVLNENNI

FYDSAKAGYALPKDADANGAYCIALKGLYEIKQITENWKEDGKFSRDKLKISNKD WFDFIQNKRYL (SEQ ID NO: 77)

[00306] Modified Cpfl enzymes

[00307] In particular embodiments, it is of interest to make use of an engineered Cpfl protein as defined herein, such as Cpfl, wherein the protein complexes with a nucleic acid molecule comprising RNA to form a CRISPR complex, wherein when in the CRISPR complex, the nucleic acid molecule targets one or more target polynucleotide loci, the protein comprises at least one modification compared to unmodified Cpfl protein, and wherein the CRISPR complex comprising the modified protein has altered activity as compared to the complex comprising the unmodified Cpfl protein. It is to be understood that when referring herein to CRISPR "protein", the Cpfl protein preferably is a modified CRISPR-Cas protein (e.g. having increased or decreased (or no) enzymatic activity, such as without limitation including Cpfl. The term "CRISPR protein" may be used interchangeably with "CRISPR-Cas protein", irrespective of whether the CRISPR protein has altered, such as increased or decreased (or no) enzymatic activity, compared to the wild type CRISPR protein.

[00308] Computational analysis of the primary structure of Cpfl nucleases reveals three distinct regions. First a C-terminal RuvC like domain, which is the only functional characterized domain. Second a N-terminal alpha-helical region and thirst a mixed alpha and beta region, located between the RuvC like domain and the alpha-helical region.

[00309] Several small stretches of unstructured regions are predicted within the Cpfl primary structure. Unstructured regions, which are exposed to the solvent and not conserved within different Cpfl orthologs, are preferred sides for splits and insertions of small protein sequences . In addition, these sides can be used to generate chimeric proteins between Cpfl orthologs.

[00310] Based on the above information, mutants can be generated which lead to inactivation of the enzyme or which modify the double strand nuclease to nickase activity. In alternative embodiments, this information is used to develop enzymes with reduced off-target effects (described elsewhere herein)

[00311] In certain of the above-described Cpfl enzymes, the enzyme is modified by mutation of one or more residues (in the RuvC domain) including but not limited to positions R909, R912, R930, R947, K949, R951, R955, K965, K968, K1000, K1002, R1003, K1009, K1017, K1022, K1029, K1035, K1054, K1072, K1086, R1094, K1095,

K 1109, K 1118, K 1142, K 1150, K 1158, K 1159, R1220, R1226, R1242, and/or R 1252 with reference to amino acid position numbering of AsCpfl (Acidaminococcus sp. BV3L6). In certain embodiments, the Cpfl enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[00312] In certain of the above-described non-naturally-occurring CRISPR-Cas proteins, the enzyme is modified by mutation of one or more residues (in the RAD50) domain including but not limited positions K324, K335, K337, R331, K369, K370, R386, R392, R393, K400, K404, K406, K408, K414, K429, K436, K438, K459, K460, K464, R670, K675, R681, K686, K689, R699, K705, R725, K729, K739, K748, and/or K752 with reference to amino acid position numbering of AsCpfl (Acidaminococcus sp. BV3L6). In certain embodiments, the Cpfl enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[00313] In certain of the Cpfl enzymes, the enzyme is modified by mutation of one or more residues including but not limited positions R912, T923, R947, K949, R951, R955, K965, K968, K1000, R1003, K1009, K1017, K1022, K1029, K1072, K1086, F 1103, R1226, and/or R1252 with reference to amino acid position numbering of AsCpfl (Acidaminococcus sp. BV3L6). In certain embodiments, the Cpfl enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[00314] In certain embodiments, the Cpfl enzyme is modified by mutation of one or more residues including but not limited positions R833, R836, K847, K879, K881, R883, R887, K897, K900, K932, R935, K940, K948, K953, K960, K984, K1003, K1017, R1033, R 1138, R 1165, and/or R1252 with reference to amino acid position numbering of LbCpfl (Lachnospiraceae bacterium ND2006). In certain embodiments, the Cpfl enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[00315] In certain embodiments, the Cpfl enzyme is modified by mutation of one or more residues including but not limited positions K15, R18, K26, Q34, R43, K48, K51, R56, R84, K85, K87, N93, R103, N104, T 118, K123, K134, R176, K177, R192, K200, K226, K273, K275, T291, R301, K307, K369, S404, V409, K414, K436, K438, K468, D482, K516, R518, K524, K530, K532, K548, K559, K570, R574, K592, D596, K603, K607, K613, C647, R681, K686, H720, K739, K748, K757, T766, K780, R790, P791, K796, K809, K815, T816, K860, R862, R863, K868, K897, R909, R912, T923, R947, K949, R951, R955, K965, K968, K1000, R1003, K1009, K1017, K1022, K1029, A1053,

K1072, K1086, F1103, S1209, R1226, R1252, K1273, K1282, and/or K1288 with reference to amino acid position numbering of AsCpfl (Acidaminococcus sp. BV3L6). In certain embodiments, the Cpfl enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[00316] In certain embodiments, the enzyme is modified by mutation of one or more residues including but not limited positions K15, R18, K26, R34, R43, K48, K51, K56, K87, K88, D90, K96, K106, K107, K120, Q125, K143, R186, K187, R202, K210, K235, K296, K298, K314, K320, K326, K397, K444, K449, E454, A483, E491, K527, K541, K581, R583, K589, K595, K597, K613, K624, K635, K639, K656, K660, K667, K671, K677, K719, K725, K730, K763, K782, K791, R800, K809, K823, R833, K834, K839, K852, K858, K859, K869, K871, R872, K877, K905, R918, R921, K932, I960, K962, R964, R968, K978, K981, K1013, R1016, K1021, K1029, K1034, K1041, K1065, K1084, and/or K1098 with reference to amino acid position numbering of FnCpfl (Francisella novicida U112). In certain embodiments, the Cpfl enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[00317] In certain embodiments, the enzyme is modified by mutation of one or more residues including but not limited positions K15, R18, K26, K34, R43, K48, K51, R56, K83, K84, R86, K92, R102, K103, K116, K121, R158, E159, R174, R182, K206, K251, K253, K269, K271, K278, P342, K380, R385, K390, K415, K421, K457, K471, A506, R508, K514, K520, K522, K538, Y548, K560, K564, K580, K584, K591, K595, K601, K634, K640, R645, K679, K689, K707, T716, K725, R737, R747, R748, K753, K768, K774, K775, K785, K787, R788, Q793, K821, R833, R836, K847, K879, K881, R883, R887, K897, K900, K932, R935, K940, K948, K953, K960, K984, K1003, K1017, R1033, K1121, R1138, R1165, K1190, K1199, and/or K1208 with reference to amino acid position numbering of LbCpfl (Lachnospiraceae bacterium ND2006). In certain embodiments, the Cpfl enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[00318] In certain embodiments, the enzyme is modified by mutation of one or more residues including but not limited positions K14, R17, R25, K33, M42, Q47, K50, D55, K85, N86, K88, K94, R104, K105, K118, K123, K131, R174, K175, R190, R198, 1221, K267, Q269, K285, K291, K297, K357, K403, K409, K414, K448, K460, K501, K515, K550, R552, K558, K564, K566, K582, K593, K604, K608, K623, K627, K633, K637, E643, K780, Y787, K792, K830, Q846, K858, K867, K876, K890, R900, K901, M906,

K921, K927, K928, K937, K939, R940, K945, Q975, R987, R990, K1001, R1034, 11036, R1038, R1042, K1052, K1055, K1087, R1090, K1095, N 1103, K 1108, K 1115, K 1139, K 1158, R 1172, K 1188, K1276, R1293, A1319, K1340, K1349, and/or K1356 with reference to amino acid position numbering of MbCpfl (Moraxella bovoculi 237). In certain embodiments, the Cpfl enzymes comprising said one or more mutations have modified, more preferably increased specificity for the target.

[00319] In certain of the above-described Cpfl enzymes, the enzyme is modified by mutation of one or more residues including but not limited to positions D917, E1006, E1028, D1227, D1255A, N1257, according to FnCpfl protein or any corresponding ortholog. In an aspect the invention provides a herein-discussed composition wherein the Cpfl enzyme is an inactivated enzyme which comprises one or more mutations selected from the group consisting of D917A, E1006A, E1028A, D1227A, D1255A, N1257A, D917A, E1006A, E1028A, D1227A, D1255A and N1257A according to FnCpfl protein or corresponding positions in a Cpfl ortholog. In an aspect the invention provides a herein-discussed composition, wherein the CRISPR enzyme comprises D917, or E1006 and D917, or D917 and D1255, according to FnCpfl protein or a corresponding position in a Cpfl ortholog.

[00320] In one embodiment, the Cpfl protein is modified with a mutation at S1228 (e.g., S1228A) with reference to amino acid position numbering of AsCpfl. See Yamano *et al*, *Cell* 165:949-962 (2016), which is incorporated herein by reference in its entirety.

[00321] In certain embodiments, the Cpfl protein has been modified to recognize a non-natural PAM, such as recognizing a PAM having a sequence or comprising a sequence YCN, YCV, AYV, TYV, RYN, RCN, TGYV, NTTN, TTN, TRTN, TYTV, TYCT, TYCN, TRTN, NTTN, TACT, TYCC, TRTC, TATV, NTTV, TTV, TSTG, TVTS, TYYS, TCYS, TBYS, TCYS, TNYS, TYYS, TNTN, TSTG, TTCC, TCCC, TATC, TGTG, TCTG, TYCV, or TCTC. In particular embodiments, said mutated Cpfl comprises one or more mutated amino acid residue at position 11, 12, 13, 14, 15, 16, 17, 34, 36, 39, 40, 43, 46, 47, 50, 54, 57, 58, 111, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 626, 627, 628, 629, 630, 631,

632, 633, 634, 635, 636, 637, 638, 642, 643, 644, 645, 646, 647, 648, 649, 651, 652, 653, 654, 655, 656, 676, 679, 680, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 707, 711, 714, 715, 716, 717, 718, 719, 720, 721, 722, 739, 765, 768, 769, 773, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, or 1048 of AsCpfl or a position corresponding thereto in a Cpfl ortholog; preferably, one or more mutated amino acid residue at position 130, 131, 132, 133, 134, 135, 136, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 570, 571, 572, 573, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 630, 631, 632, 646, 647, 648, 649, 650, 651, 652, 653, 683, 684, 685, 686, 687, 688, 689, or 690;

[00322] In certain embodiments, the Cpfl protein is modified to have increased activity, i.e. wider PAM specificity. In particular embodiments, the Cpfl protein is modified by mutation of one or more residues including but not limited positions 539, 542, 547, 548, 550, 551, 552, 167, 604, and/or 607 of AsCpfl, or the corresponding position of an AsCpfl orthologue, homologue, or variant, preferably mutated amino acid residues at positions 542 or 542 and 607, wherein said mutations preferably are 542R and 607R, such as S542R and K607R; or preferably mutated amino acid residues at positions 542 and 548 (and optionally 552), wherein said mutations preferably are 542R and 548V (and optionally 552R), such as S542R and K548V (and optionally N552R); or at position 532, 538, 542, and/or 595 of LbCpfl, or the corresponding position of an AsCpfl orthologue, homologue, or variant, preferably mutated amino acid residues at positions 532 or 532 and 595, wherein said mutations preferably are 532R and 595R, such as G532R and K595R; or preferably mutated amino acid residues at positions 532 and 538 (and optionally 542), wherein said mutations preferably are 532R and 538V (and optionally 542R), such as G532R and K538V (and optionally Y542R), most preferably wherein said mutations are S542R and K607R, S542R and K548V, or S542R, K548V and N552R of AsCpfl.

[00323] Deactivated / inactivated Cpfl protein

[00324] Where the Cpfl protein has nuclease activity, the Cpfl protein may be modified to have diminished nuclease activity e.g., nuclease inactivation of at least 70%, at least 80%, at least 90%, at least 95%, at least 97%, or 100% as compared with the wild type enzyme; or to put in another way, a Cpfl enzyme having advantageously about 0% of the nuclease activity of the non-mutated or wild type Cpfl enzyme or CRISPR-Cas protein, or

no more than about 3% or about 5% or about 10% of the nuclease activity of the non-mutated or wild type Cpfl enzyme, e.g. of the non-mutated or wild type *Francisella novicida U112* (FnCpfl), *Acidaminococcus sp. BV3L6* (AsCpfl), *Lachnospiraceae bacterium ND2006* (LbCpfl) or *Moraxella bovoculi 237* (MbCpfl Cpfl enzyme or CRISPR-Cas protein. This is possible by introducing mutations into the nuclease domains of the Cpfl and orthologs thereof.

[00325] In preferred embodiments of the present invention at least one Cpfl protein is used which is a Cpfl nickase. More particularly, a Cpfl nickase is used which does not cleave the target strand but is capable of cleaving only the strand which is complementary to the target strand, i.e. the non-target DNA strand also referred to herein as the strand which is not complementary to the guide sequence. More particularly the Cpfl nickase is a Cpfl protein which comprises a mutation in the arginine at position 1226A in the Nuc domain of Cpfl from *Acidaminococcus sp.*, or a corresponding position in a Cpfl orthoiog. In further particular embodiments, the enzyme comprises an arginine-to-alanine substitution or an R1226A mutation. It will be understood by the skilled person that where the enzyme is not AsCpfl, a mutation may be made at a residue in a corresponding position. In particular embodiments, the Cpfl is FnCpfl and the mutation is at the arginine at position R1218. In particular embodiments, the Cpfl is LbCpfl and the mutation is at the arginine at position R 1138. In particular embodiments, the Cpfl is MbCpfl and the mutation is at the arginine at position R1293.

[00326] In certain embodiments, use is made additionally or alternatively of a CRISPR-Cas protein which is is engineered and can comprise one or more mutations that reduce or eliminate a nuclease activity. The amino acid positions in the FnCpflp RuvC domain include but are not limited to D917A, E1006A, E1028A, D1227A, D1255A, N1257A, D917A, E1006A, E1028A, D1227A, D1255A and N1257A. Applicants have also identified a putative second nuclease domain which is most similar to PD-(D/E)XK nuclease superfamily and Hindi endonuclease like. The point mutations to be generated in this putative nuclease domain to substantially reduce nuclease activity include but are not limited to N580A, N584A, T587A, W609A, D610A, K613A, E614A, D616A, K624A, D625A, K627A and Y629A. In a preferred embodiment, the mutation in the FnCpflp RuvC domain is D917A or E1006A, wherein the D917A or E1006A mutation completely inactivates the DNA cleavage activity of the FnCpfl effector protein. In another

embodiment, the mutation in the FnCpflp RuvC domain is D1255A, wherein the mutated FnCpfl effector protein has significantly reduced nucleolytic activity.

[00327] More particularly, the inactivated Cpfl enzymes include enzymes mutated in amino acid positions As908, As993, As1263 of AsCpfl or corresponding positions in Cpfl orthologs. Additionally, the inactivated Cpfl enzymes include enzymes mutated in amino acid position Lb832, 925, 947 or 1180 of LbCpfl or corresponding positions in Cpfl orthologs. More particularly, the inactivated Cpfl enzymes include enzymes comprising one or more of mutations AsD908A, AsE993A, AsD1263A of AsCpfl or corresponding mutations in Cpfl orthologs. Additionally, the inactivated Cpfl enzymes include enzymes comprising one or more of mutations LbD832A, E925A, D947A or D1180A of LbCpfl or corresponding mutations in Cpfl orthologs.

[00328] Mutations can also be made at neighboring residues, e.g., at amino acids near those indicated above that participate in the nuclease activity. In some embodiments, only the RuvC domain is inactivated, and in other embodiments, another putative nuclease domain is inactivated, wherein the effector protein complex functions as a nickase and cleaves only one DNA strand. In a preferred embodiment, the other putative nuclease domain is a HincII-like endonuclease domain.

[00329] The inactivated Cpfl or Cpfl nickase may have associated (e.g., via fusion protein) one or more functional domains, including for example, an cytidine deaminase or catalytic domain thereof. In some cases it is advantageous that additionally at least one heterologous NLS is provided. In some instances, it is advantageous to position the NLS at the N terminus. In general, the positioning of the one or more functional domain on the inactivated Cpfl or Cpfl nickase is one which allows for correct spatial orientation for the functional domain to affect the target with the attributed functional effect. For example, when the functional domain is an cytidine deaminase catalytic domain thereof, the cytidine deaminase catalytic domain is placed in a spatial orientation which allows it to contact and deaminate a target adenine. This may include positions other than the N- / C- terminus of Cpfl. In some embodiments, the cytidine deaminase protein or catalytic domain thereof is inserted into an internal loop of Cpfl .

[00330] Determination of PAM

[00331] Determination of PAM can be ensured as follows. This experiment closely parallels similar work in *E. coli* for the heterologous expression of StCas9 (Sapranauskas, R. et al. Nucleic Acids Res 39, 9275-9282 (2011)). Applicants introduce a plasmid

containing both a PAM and a resistance gene into the heterologous *E. coli,* and then plate on the corresponding antibiotic. If there is DNA cleavage of the plasmid, Applicants observe no viable colonies.

[00332] In further detail, the assay is as follows for a DNA target. Two *E.coli* strains are used in this assay. One carries a plasmid that encodes the endogenous effector protein locus from the bacterial strain. The other strain carries an empty plasmid (e.g.pACYC184, control strain). All possible 7 or 8 bp PAM sequences are presented on an antibiotic resistance plasmid (pUC19 with ampicillin resistance gene). The PAM is located next to the sequence of proto-spacer 1 (the DNA target to the first spacer in the endogenous effector protein locus). Two PAM libraries were cloned. One has a 8 random bp 5' of the proto-spacer (e.g. total of 65536 different PAM sequences = complexity). The other library has 7 random bp 3' of the proto-spacer (e.g. total complexity is 16384 different PAMs). Both libraries were cloned to have in average 500 plasmids per possible PAM. Test strain and control strain were transformed with 5'PAM and 3'PAM library in separate transformations and transformed cells were plated separately on ampicillin plates. Recognition and subsequent cutting/interference with the plasmid renders a cell vulnerable to ampicillin and prevents growth. Approximately 12h after transformation, all colonies formed by the test and control strains where harvested and plasmid DNA was isolated. Plasmid DNA was used as template for PCR amplification and subsequent deep sequencing. Representation of all PAMs in the untransfomed libraries showed the expected representation of PAMs in transformed cells. Representation of all PAMs found in control strains showed the actual representation. Representation of all PAMs in test strain showed which PAMs are not recognized by the enzyme and comparison to the control strain allows extracting the sequence of the depleted PAM.

[00333] The following PAMs have been identified for certain wild-type Cpfl orthologues: the Acidaminococcus sp. BV3L6 Cpfl (AsCpfl), Lachnospiraceae bacterium ND2006 Cpfl (LbCpfl) and Prevotella albensis (PaCpfl) can cleave target sites preceded by a TTTV PAM, where V is A/C or G, FnCpflp, can cleave sites preceded by TTN, where N is A/C/G or T. The Moraxella bovoculi AAX08_00205, Moraxella bovoculi AAX1 1_00205, Butyrivibrio sp. NC3005, Thiomicrospira sp. XS5, or Lachnospiraceae bacterium MA2020 PAM is 5' TTN, where N is A/C/G or T. The natural PAM sequence is TTTV or BTTV, wherein B is T/C or G and V is A/C or G and the effector protein is Moraxella lacunata Cpfl .

[00334] Delivery

[00335] In some embodiments, the components of the CD-functionalized CRISPR-Cas system may be delivered in various form, such as combinations of DNA/RNA or RNA/RNA or protein RNA. For example, the Cpfl protein 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.

[00336] 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.

[00337] Vectors

[00338] In general, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. It is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elementsVectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Vectors for and that result in expression in a eukaryotic cell can be referred to

herein as "eukaryotic expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[00339] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[00340] With regards to recombination and cloning methods, mention is made of U.S. patent application 10/815,730, published September 2, 2004 as US 2004-0171 156 Al, the contents of which are herein incorporated by reference in their entirety.

[00341] The term "regulatory element" is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g., transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g., liver, pancreas), or particular cell types (e.g., lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector comprises one or more pol III promoter (e.g., 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g., 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g., 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H I 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 EFla promoter. Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β-globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc. A vector can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., clustered regularly interspersed short palindromic repeats (CRISPR) transcripts, proteins, enzymes, mutant forms thereof, fusion proteins thereof, etc.). With regards to regulatory sequences, mention is made of U.S. patent application 10/491,026, the contents of which are incorporated by reference herein in their entirety. With regards to promoters, mention is made of PCT publication WO 201 1/028929 and U.S. application 12/51 1,940, the contents of which are incorporated by reference herein in their entirety.

[00342] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[00343] In particular embodiments, use is made of bicistronic vectors for the guide RNA and (optionally modified or mutated) the CRISPR-Cas protein fused to cytidine deaminase. Bicistronic expression vectors for guide RNA and (optionally modified or mutated) CRISPR-Cas protein fused to cytidine deaminase are preferred. In general and particularly in this embodiment, (optionally modified or mutated) CRISPR-Cas protein fused to cytidine deaminase is preferably driven by the CBh promoter. The RNA may preferably be driven by a Pol III promoter, such as a U6 promoter. Ideally the two are combined.

[00344] Vectors can be designed for expression of CRISPR transcripts (e.g. nucleic acid transcripts, proteins, or enzymes) in prokaryotic or eukaryotic cells. For example, CRISPR transcripts can be expressed in bacterial cells such as Escherichia coli, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN

ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

[00345] Vectors may be introduced and propagated in a prokaryote or prokaryotic cell. In some embodiments, a prokaryote is used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a eukaryotic cell (e.g. amplifying a plasmid as part of a viral vector packaging system). In some embodiments, a prokaryote is used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism. Expression of proteins in prokaryotes is most often carried out in Escherichia coli with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, such as to the amino terminus of the recombinant protein. Such fusion vectors may serve one or more purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Example fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion E. coli expression vectors include pTrc (Amrann et al., (1988) Gene 69:301-315) and pET lid (Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89). In some embodiments, a vector is a yeast expression vector. Examples of vectors for expression in yeast Saccharomyces cerivisae include pYepSecl (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). In some embodiments, a vector drives protein expression in insect cells using baculovirus

expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, et al., 1983. Mol. Cell. Biol. 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. Virology 170: 31-39).

[00346] In some embodiments, a vector is capable of driving expression of one or more sequences in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. Nature 329: 840) and pMT2PC (Kaufman, et al., 1987. EMBO J. 6: 187-195). When used in mammalian cells, the expression vector's control functions are typically provided by one or more regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, et al., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[00347] 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 a-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546). With regards to these prokaryotic and eukaryotic vectors, mention is made of U.S. Patent 6,750,059, the contents of which are incorporated by reference herein in their entirety. Other embodiments of the invention may relate to the use of viral vectors, with regards to which mention is made of U.S. Patent application 13/092,085, the contents of which are

incorporated by reference herein in their entirety. Tissue-specific regulatory elements are known in the art and in this regard, mention is made of U.S. Patent 7,776,321, the contents of which are incorporated by reference herein in their entirety. In some embodiments, a regulatory element is operably linked to one or more elements of a CRISPR system so as to drive expression of the one or more elements of the CRISPR system.

[00348] In some embodiments, one or more vectors driving expression of one or more elements of a nucleic acid-targeting system are introduced into a host cell such that expression of the elements of the nucleic acid-targeting system direct formation of a nucleic acid-targeting complex at one or more target sites. For example, a nucleic acid-targeting effector enzyme and a nucleic acid-targeting guide RNA could each be operably linked to separate regulatory elements on separate vectors. RNA(s) of the nucleic acid-targeting system can be delivered to a transgenic nucleic acid-targeting effector protein animal or mammal, e.g., an animal or mammal that constitutively or inducibly or conditionally expresses nucleic acid-targeting effector protein; or an animal or mammal that is otherwise expressing nucleic acid-targeting effector proteins or has cells containing nucleic acid-targeting effector proteins, such as by way of prior administration thereto of a vector or vectors that code for and express in vivo nucleic acid-targeting effector proteins. Alternatively, two or more of the elements expressed from the same or different regulatory elements, may be combined in a single vector, with one or more additional vectors providing any components of the nucleic acid-targeting system not included in the first vector. nucleic acid-targeting system elements that are combined in a single vector may be arranged in any suitable orientation, such as one element located 5' with respect to ("upstream" of) or 3' with respect to ("downstream" of) a second element. The coding sequence of one element may be located on the same or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In some embodiments, a single promoter drives expression of a transcript encoding a nucleic acid-targeting effector protein and the nucleic acid-targeting guide RNA, embedded within one or more intron sequences (e.g., each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the nucleic acid-targeting effector protein and the nucleic acid-targeting guide RNA may be operably linked to and expressed from the same promoter. Delivery vehicles, vectors, particles, nanoparticles, formulations and components thereof for expression of one or more elements of a nucleic acid-targeting system are as used in the foregoing documents, such as WO 2014/093622

(PCT/US20 13/074667). In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a "cloning site"). In some embodiments, one or more insertion sites (e.g., about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors. When multiple different guide sequences are used, a single expression construct may be used to target nucleic acid-targeting activity to multiple different, corresponding target sequences within a cell. For example, a single vector may comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide sequences. In some embodiments, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-sequence-containing vectors may be provided, and optionally delivered to a cell. In some embodiments, a vector comprises a regulatory element operably linked to an enzyme-coding sequence encoding a a nucleic acid-targeting effector protein. Nucleic acid-targeting effector protein or nucleic acid-targeting guide RNA or RNA(s) can be delivered separately; and advantageously at least one of these is delivered via a particle complex. nucleic acid-targeting effector protein mRNA can be delivered prior to the nucleic acid-targeting guide RNA to give time for nucleic acid-targeting effector protein to be expressed. Nucleic acid-targeting effector protein mRNA might be administered 1-12 hours (preferably around 2-6 hours) prior to the administration of nucleic acid-targeting guide RNA. Alternatively, nucleic acid-targeting effector protein mRNA and nucleic acid-targeting guide RNA can be administered together. Advantageously, a second booster dose of guide RNA can be administered 1-12 hours (preferably around 2-6 hours) after the initial administration of nucleic acid-targeting effector protein mRNA + guide RNA. Additional administrations of nucleic acid-targeting effector protein mRNA and/or guide RNA might be useful to achieve the most efficient levels of genome modification.

[00349] Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding components of a nucleic acid-targeting system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, Science

256:808-813 (1992); Nabel & Feigner, TIBTECH 11:21 1-217 (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): 1149-1 154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51(1):3 1-44 (1995); Haddada et al., in Current Topics in Microbiology and Immunology, Doerfler and Bohm (eds) (1995); and Yu et al., Gene Therapy 1:13-26 (1994).

[00350] Methods of non-viral delivery of nucleic acids include lipofection, nucleofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Feigner, WO 91/17424; WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration).

[00351] Plasmid delivery involves the cloning of a guide RNA into a CRISPR-Cas 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-Cas 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-Cas protein can be toxic to the cells. Finally, plasmids always hold 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).

[00352] 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). This is discussed more in detail below.

[00353] The use of RNA or DNA viral based systems for the delivery of nucleic acids takes advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro, and the modified cells may optionally be administered to patients (ex vivo). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[00354] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); 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).

[00355] In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system. Adeno-associated virus ("AAV") vectors may also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., Virology 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). Construction of

recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989).

[00356] The invention provides AAV that contains or consists essentially of an exogenous nucleic acid molecule encoding a CRISPR 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., Cpfl and a terminator, and one or more, advantageously up to the packaging size limit of the vector, e.g., in total (including the first cassette) five, cassettes comprising or consisting essentially of a promoter, nucleic acid molecule encoding guide RNA (gRNA) and a terminator (e.g., each cassette schematically represented as Promoter-gRNAl -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., Cas (Cpfl) and a terminator, and a second rAAV containing one or more cassettes each 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-gRNAl-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). Alternatively, because Cpfl can process its own crRNA/gRNA, a single crRNA/gRNA array can be used for multiplex gene editing. Hence, instead of including multiple cassettes to deliver the gRNAs, the rAAV may contain a single cassette comprising or consisting essentially of a promoter, a plurality of crRNA/gRNA, and a terminator (e.g., schematically represented as Promoter-gRNAl-gRNA2 …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). See Zetsche *et al Nature Biotechnology* 35, 31-34 (2017), which is incorporated herein by reference in its entirety. 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). 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.

[00357] 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.

[00358] 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, Huhl, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panel, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calul, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7,

NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/ 3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr -/-, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML Tl, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepalclc7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYOl, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-IOA, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN / OPCT cell lines, Peer, PNT-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.)).

[00359] In particular embodiments, transient expression and/or presence of one or more of the components of the CD-functionalized CRISPR system can be of interest, such as to reduce off-target effects. 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 CD-functionalized 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.

[00360] In some embodiments it is envisaged to introduce the RNA and/or protein directly to the host cell. For instance, the CRISPR-Cas protein can be delivered as encoding mRNA together with an in vitro transcribed guide RNA. Such methods can reduce the time to ensure effect of the CRISPR-Cas protein and further prevents long-term expression of the CRISPR system components.

[00361] 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, 539:1 11-1 14; 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, 1 1: 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.

[00362] Indeed, RNA delivery is a useful method of in vivo delivery. It is possible to deliver Cpfl, cytidine deaminase, and guide RNA into cells using liposomes or nanoparticles. Thus delivery of the CRISPR-Cas protein, such as a Cpfl, the delivery of the cytidine deaminase (which may be fused to the CRISPR-Cas protein or an adaptor protein), and/or delivery of the RNAs of the invention may be in RNA form and via microvesicles, liposomes or particle or particles. For example, Cpfl mRNA, cytidine deaminase mRNA, and guide RNA can be packaged into liposomal particles for delivery in vivo. Liposomal transfection reagents such as lipofectamine from Life Technologies and other reagents on the market can effectively deliver RNA molecules into the liver.

[00363] Means of delivery of RNA also preferred include delivery of RNA 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: 3 112-31 18, 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):21 12-26. 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 purify 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 (a-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 22:71 1-719 (June 201 1)) 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 Toe-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 a-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 μmoï of CRISPR Cas targeted to the brain may be contemplated. Zou et al. ((HUMAN GENE THERAPY 22:465-475 (April 201 1)) describes a method of lentiviral-mediated delivery of short-hairpin RNAs targeting PKCy for in vivo gene silencing in the spinal cord of rats. Zou et al. administered about 10 μï of a recombinant lentivirus having a titer of 1×10^9 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 1×10^9 transducing units (TU)/ml may be contemplated.

[00364] Dosage of vectors

[00365] 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.

[00366] 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.

[00367] In an embodiment herein the delivery is via an adenovirus, which may be at a single booster dose containing at least 1×10^5 particles (also referred to as particle units, pu) of adenoviral vector. In an embodiment herein, the dose preferably is at least about 1×10^6 particles (for example, about 1×10^6-1×10^{12} particles), more preferably at least about 1×10^7 particles, more preferably at least about 1×10^8 particles (e.g., about 1×10^8-1×10^{11} particles or about 1×10^8-1×10^{12} particles), and most preferably at least about $1 \times 10^{\circ}$ particles (e.g., about 1×10^9-1×10^{10} particles or about 1×10^9-1×10^{12} particles), or even at least about 1×10^{10} particles (e.g., about 1×10^{10}-1×10^{12} particles) of the adenoviral vector. Alternatively, the dose comprises no more than about 1×10^{14} particles, preferably no more than about 1×10^{13} particles, even more preferably no more than about 1×10^{12} particles, even more preferably no more than about 1×10^{11} particles, and most preferably

no more than about 1×10^{10} particles (e.g., no more than about 1×10^{9} articles). Thus, the dose may contain a single dose of adenoviral vector with, for example, about 1×10^{6} particle units (pu), about 2×10^{6} pu, about 4×10^{6} pu, about 1×10^{7} pu, about 2×10^{7} pu, about 4×10^{1} pu, about 1×10^{8} pu, about 2×10^{8} pu, about 4×10^{8} pu, about 1×10^{9} pu, about 2×10^{9} pu, about 4×10^{9} pu, about 1×10^{10} pu, about 2×10^{10} pu, about 4×10^{10} pu, about 1×10^{11} pu, about 2×10^{11} pu, about 4×10^{11} pu, about 1×10^{12} pu, about 2×10^{12} pu, or about 4×10^{12} pu of adenoviral vector. See, for example, the adenoviral vectors in U.S. Patent No. 8,454,972 B2 to Nabel, et. al., granted on June 4, 2013; incorporated by reference herein, and the dosages at col 29, lines 36-58 thereof. In an embodiment herein, the adenovirus is delivered via multiple doses.

[00368] In an embodiment herein, the delivery is via an AAV. A therapeutically effective dosage for in vivo delivery of the AAV to a human is believed to be in the range of from about 20 to about 50 ml of saline solution containing from about 1×10^{10} to about 1×10^{10} functional AAV/ml solution. The dosage may be adjusted to balance the therapeutic benefit against any side effects. In an embodiment herein, the AAV dose is generally in the range of concentrations of from about 1×10^{5} to 1×10^{50} genomes AAV, from about 1×10^{8} to 1×10^{20} genomes AAV, from about 1×10^{10} to about 1×10^{16} genomes, or about 1×10^{11} to about 1×10^{16} genomes AAV. A human dosage may be about 1×10^{13} genomes AAV. Such concentrations may be delivered in from about 0.001 ml to about 100 ml, about 0.05 to about 50 ml, or about 10 to about 25 ml of a carrier solution. Other effective dosages can be readily established by one of ordinary skill in the art through routine trials establishing dose response curves. See, for example, U.S. Patent No. 8,404,658 B2 to Hajjar, et al., granted on March 26, 2013, at col. 27, lines 45-60.

[00369] 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-Cas protein, 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.

[00370] 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.

[00371] 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.

[00372] RNA delivery

[00373] In particular embodiments, RNA based delivery is used. In these embodiments, mRNA of the CRISPR-Cas protein, mRNA of the cytidine deaminase (which may be fused to a CRISPR-Cas protein or an adaptor), are 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). In some embodiments, the mRNA(s) encoding Cpfl and/or cytidine deaminase can be chemically modified, which may lead to improved activity compared to plasmid-encoded Cpfl and/or cytidine deaminase. For example, uridines in the mRNA(s) can be partially or fully substituted with pseudouridine (Ψ), N^1-methylpseudouridine ($me^1Ψ$), 5-methoxyuridine(5moU). See Li *et al, Nature Biomedical Engineering* 1, 0066 DOI: 10.1038/s41551-017-0066 (2017), which is incorporated herein by reference in its entirety.

[00374] RNP

[00375] In particular embodiments, pre-complexed guide RNA, CRISPR-Cas protein, and cytidine deaminase (which may be fused to a CRISPR-Cas protein or an adaptor) are delived 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. An important advantage is that both 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(l):47-54), Chu et al. (2016, BMC Biotechnol. 16:4), and Wang et al. (2013, Cell. 9;153(4):910-8).

[00376] 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

[00377] Particles

[00378] 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 l,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.

[00379] 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.

[00380] As used herein, a particle delivery system/formulation is defined as any biological delivery system/formulation which includes a particle in accordance with the present invention. A particle in accordance with the present invention is any entity having a greatest dimension (e.g. diameter) of less than 100 microns (μm). In some embodiments, inventive particles have a greatest dimension of less than 10 μm. In some embodiments, inventive particles have a greatest dimension of less than 2000 nanometers (nm). In some embodiments, inventive particles have a greatest dimension of less than 1000 nanometers

(nm). In some embodiments, inventive particles have a greatest dimension of less than 900 nm, 800 nm, 700 nm, 600 nm, 500 nm, 400 nm, 300 nm, 200 nm, or 100 nm. Typically, inventive particles have a greatest dimension (e.g., diameter) of 500 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 250 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 200 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 150 nm or less. In some embodiments, inventive particles have a greatest dimension (e.g., diameter) of 100 nm or less. Smaller particles, e.g., having a greatest dimension of 50 nm or less are used in some embodiments of the invention. In some embodiments, inventive particles have a greatest dimension ranging between 25 nm and 200 nm.

[00381] In terms of this invention, it is preferred to have one or more components of CRISPR complex, e.g., CRISPR-Cas protein or mRNA, or cytidine deaminase (which may be fused to a CRISPR-Cas protein or an adaptor) 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.

[00382] 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.

[00383] 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.

[00384] 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.

[00385] 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.

[00386] 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-Cas protein or mRNA, cytidine deaminase (which may be fused to a CRISPR-Cas protein or an adaptor) 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.

[00387] 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.

[00388] CRISPR-Cas protein mRNA, cytidine deaminase (which may be fused to a CRISPR-Cas protein or an adaptor) or mRNA, and guide RNA may be delivered simultaneously using particles or lipid envelopes; for instance, CRISPR-Cas protein 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 l,2-dioleoyl-3-trimethylammonium-propane (DOTAP) or 1,2-ditetradecanoyl-s«-glycero-3-phosphocholine (DMPC) and/or wherein the hydrophilic polymer comprises ethylene glycol or polyethylene glycol (PEG); and/or wherein the particle further comprises cholesterol (e.g., particle from formulation 1 = DOTAP 100, DMPC 0, PEG 0, Cholesterol 0; formulation number 2 = DOTAP 90, DMPC 0, PEG 10, Cholesterol 0; formulation number 3 = DOTAP 90, DMPC 0, PEG 5, Cholesterol 5), wherein particles are formed using an efficient, multistep process wherein first, effector protein and RNA are mixed together, e.g., at a 1:1 molar ratio, e.g., at room temperature, e.g., for 30 minutes, e.g., in sterile, nuclease free IX PBS; and separately, DOTAP, DMPC, PEG, and cholesterol as applicable for the formulation are dissolved in alcohol, e.g., 100% ethanol; and, the two solutions are mixed together to form particles containing the complexes).

[00389] Nucleic acid-targeting effector proteins (e.g., a Type V protein such as Cpfl) 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 U S 9,301,923.

[00390] 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. 201 1 Jun 6;8(3):774-87. doi: 10.1021/mpl00390w. Epub 201 1 Apr 1) describes biodegradable core-shell structured nanoparticles with a poly (P-amino ester) (PBAE) core enveloped by a phospholipid bilayer shell. These were developed for in vivo mRNA delivery. The pH-responsive PBAE component was chosen to promote endosome

disruption, while the lipid surface layer was selected to minimize toxicity of the polycation core. Such are, therefore, preferred for delivering RNA of the present invention.

[00391] 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.

[00392] 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 CD-functionalized 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., NatNanotechnol. 2012 Jun 3;7(6):389-93.

[00393] US patent application 201 10293703 relates to lipidoid compounds are also particularly useful in the administration of polynucleotides, which may be applied to deliver the CD-functionalized 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.

[00394] US Patent Publication No. 201 10293703 also provides methods of preparing the aminoalcohol lipidoid compounds. One or more equivalents of an amine are allowed to react with one or more equivalents of an epoxide-terminated compound under suitable conditions to form an aminoalcohol lipidoid compound of the present invention. In certain embodiments, all the amino groups of the amine are fully reacted with the epoxide-terminated compound to form tertiary amines. In other embodiments, all the amino groups of the amine are not fully reacted with the epoxide-terminated compound to form tertiary amines thereby resulting in primary or secondary amines in the aminoalcohol lipidoid compound. These primary or secondary amines are left as is or may be reacted with another electrophile such as a different epoxide-terminated compound. As will be appreciated by one skilled in the art, reacting an amine with less than excess of epoxide-terminated compound will result in a plurality of different aminoalcohol lipidoid compounds with various numbers of tails. Certain amines may be fully functionalized with two epoxide-derived compound tails while other molecules will not be completely functionalized with epoxide-derived compound tails. For example, a diamine or polyamine may include one, two, three, or four epoxide-derived compound tails off the various amino moieties of the molecule resulting in primary, secondary, and tertiary amines. In certain embodiments, all the amino groups are not fully functionalized. In certain embodiments, two of the same types of epoxide-terminated compounds are used. In other embodiments, two or more different epoxide-terminated compounds are used. The synthesis of the aminoalcohol lipidoid compounds is performed with or without solvent, and the synthesis may be performed at higher temperatures ranging from 30-100 °C, preferably at approximately 50-90 °C. The prepared aminoalcohol lipidoid compounds may be optionally purified. For example, the mixture of aminoalcohol lipidoid compounds may be purified to yield an aminoalcohol lipidoid compound with a particular number of epoxide-derived compound tails. Or the mixture may be purified to yield a particular stereo- or regioisomer. The aminoalcohol lipidoid compounds may also be alkylated using an alkyl halide (e.g., methyl iodide) or other alkylating agent, and/or they may be acylated.

[00395] US Patent Publication No. 201 10293703 also provides libraries of aminoalcohol lipidoid compounds prepared by the inventive methods. These aminoalcohol lipidoid

compounds may be prepared and/or screened using high-throughput techniques involving liquid handlers, robots, microtiter plates, computers, etc. In certain embodiments, the aminoalcohol lipidoid compounds are screened for their ability to transfect polynucleotides or other agents (e.g., proteins, peptides, small molecules) into the cell.

[00396] 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 us to identify polymer coatings that inhibit macrophage activation in vitro. Furthermore, these coatings reduce the recruitment of inflammatory cells, and reduce fibrosis, following the subcutaneous implantation of carboxylated polystyrene microparticles. These polymers may be used to form polyelectrolyte complex capsules for cell encapsulation. The invention may also have many other biological applications such as antimicrobial coatings, DNA or siRNA delivery, and stem cell tissue engineering. The teachings of US Patent Publication No. 20130302401 may be applied to the CD-functionalized CRISPR-Cas system of the present invention.

[00397] Preassembled recombinant CRISPR-Cas complexes comprising Cpfl, cytidine deaminase (which may be fused to Cpfl or an adaptor protein), and guide RNA 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 Cpfl ribonucleoproteins, Nat Biotechnol. 2016 Jun 6. doi: 10.1038/nbt.3596.

[00398] 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.

[00399] In some embodiments, sugar-based particles may be used, for example GalNAc, as described herein and with reference to WO20141 18272 (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; 0stergaard et al., Bioconjugate Chem., 2015, 26 (8), pp 1451—1455). Similarly, poly(acrylate) polymers have been described for in vivo nucleic acid delivery (see WO2013 158141 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).

[00400] Nanoclews

[00401] Further, the CD-functionalized 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.

[00402] LNP

[00403] In some embodiments, delivery is by encapsulation of the Cpfl 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, acetampinophen, 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.

[00404] 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.

[00405] 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. 201 1). 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 l,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), l,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), l,2-dilinoleyloxy-keto-N,N-dimethyl-3-aminopropane (DLinKDMA), and l,2-dilinoleyl-4-(2-dimethylaminoethyl)-[l,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. 201 1). 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.

[00406] 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. 201 1). The cationic lipids l,2-dilineoyl-3-dimethylammonium -propane (DLinDAP), 1,2-

dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), l,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), l,2-dilinoleyl-4-(2-dimethylaminoethyl)-[l,3]-dioxolane (DLinKC2-DMA), (3-o-[2"-(methoxypolyethyleneglycol 2000) succinoyl]-l,2-dimyristoyl-sn-glycol (PEG-S-DMG), and R-3-[(ro-methoxy-poly(ethylene glycol)2000) carbamoyl]-l,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/1. This ethanol solution of lipid may be added drop-wise to 50 mmol/1 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/1 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.

[00407] 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/1, 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.

[00408] 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.

[00409] Literature that may be employed in conjunction with herein teachings include: Cutler et al., J. Am. Chem. Soc. 201 1 133:9254-9257, Hao et al., Small. 201 1 7:3158-3162, Zhang et al., ACS Nano. 201 1 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 109:1 1975-80, 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, 209ral52 (2013) and Mirkin, et al., Small, 10:186-192.

[00410] 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.

[00411] 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: l,4,7,10-tetraazacyclododecane-l,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.

[00412] 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 leukaemia has been administered siRNA by liposomal delivery, Davis et al.'s clinical trial is the initial human trial to systemically deliver siRNA with a targeted delivery system and to treat patients with solid cancer. To ascertain whether the targeted delivery system can provide effective delivery of functional siRNA to human 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).

[00413] 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 comprising polymer conjugated to a surfactant, hydrophilic polymer or lipid. 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 conatin a

biologically active material. 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 μιη and 30 μ m, incorporating a surfactant on the surface thereof for drug delivery to the pulmonary system. 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. 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. 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 can 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.

[00414] 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.

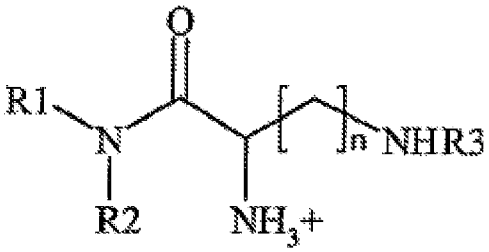
[00415] An epoxide-modified lipid-polymer may be utilized to deliver the CRISPR-Cas system of the present invention to pulmonary, cardiovascular or renal cells, however, one of skill in the art may adapt the system to deliver to other target organs. Dosage ranging from about 0.05 to about 0.6 mg/kg are envisioned. Dosages over several days or weeks are also envisioned, with a total dosage of about 2 mg/kg.

[00416] In some embodiments, the LNP for deliverting the RNA molecules is prepared by methods known in the art, such as those described in, for example, WO 2005/105152 (PCT/EP2005/004920), WO 2006/069782 (PCT/EP2005/0 14074), WO 2007/121947 (PCT/EP2007/003496), and WO 2015/082080 (PCT/EP20 14/003274), which are herein incorporated 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 *etal, 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.

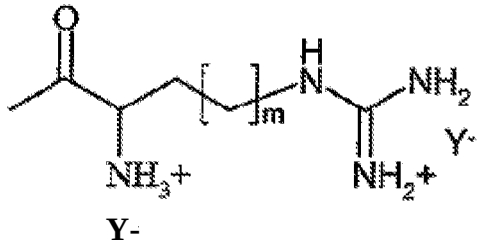
[00417] In some embodiments, the LNP includes any LNP disclosed in WO 2005/105152 (PCT/EP2005/004920), WO 2006/069782 (PCT/EP2005/014074), WO 2007/121947 (PCT/EP2007/003496), and WO 2015/082080 (PCT/EP2014/003274).

[00418] In some embodiments, the LNP includes at least one lipid having Formula I:



(Formula I),

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:



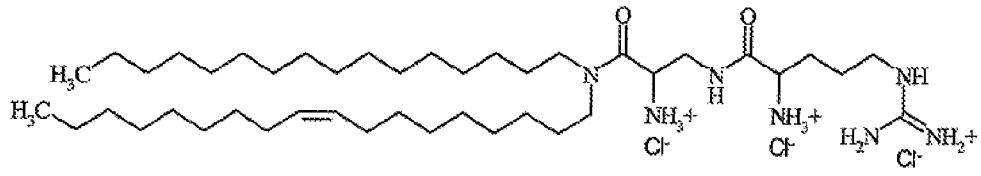
(Formula II),

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.

[00419] 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.

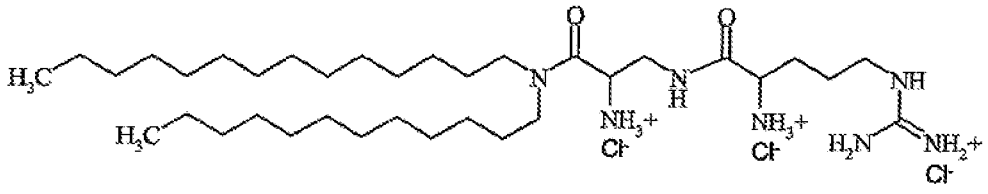
[00420] In some embodiments, the LNP comprises one or more lipids select from:

□ -arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrochloride (Formula III):



(Formula III);

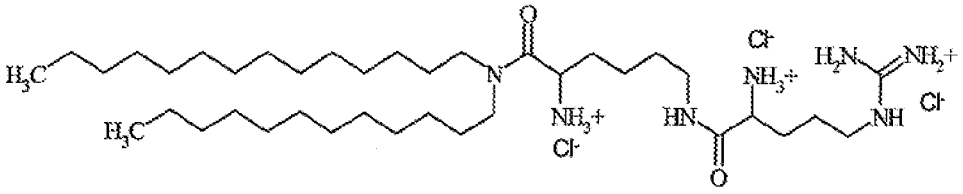
□ -arginyl-2,3-diamino propionic acid-N-lauryl-N-myristyl-amide trihydrochloride (Formula IV):



(Formula IV);

and

□ -arginyl-lysine-N-lauryl-N-myristyl-amide trihydrochloride (Formula V):



(Formula V).

[00421] 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 guide RNA and/or mRNA.

[00422] In some embodiments, the constituent of the LNP comprises an mRNA encoding a CRIPSR-Cas protein. In some embodiments, the constituent of the LNP comprises an mRNA encoding a Type-II or Type-V CRIPSR-Cas protein. In some embodiments, the constituent of the LNP comprises an mRNA encoding a cytidine deaminase (which may be fused to a CRISPR-Cas protein or an adaptor protein).

[00423] 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.

[00424] 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 l,2-dioleylsn-glycero-3-phosphoethanolamine (DOPE).

[00425] 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.

[00426] 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.

[00427] In some embodiments, the LNP includes any of □-3-arginyl-2,3-diaminopropionic acid-N-palmityl-N-oleyl-amide trihydrochloride, □-arginyl-2,3-diaminopropionic acid-N-lauryl-N-myristyl-amide trihydrochloride or □-arginyl-lysine-N-lauryl-N-myri sty1-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 □-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 ▫-arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrochloride (lipid), 1,2-diphytanoyl-sn-glycero-3-phosphoethanolamine (first helper lipid), and 1,2-disteroyl-sn-glycero-3-phosphoethanolamine-PEG2000 (second helper lipid).

[00428] 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 □-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)-l,2-distearoyl-sn-glycero3-phosphoethanolamine, or a sodium salt thereof.

[00429] 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.

[00430] 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.

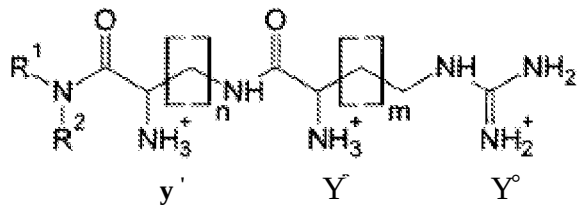
[00431] In some embodiments, the LNP includes at least one lipid, a first helper lipid, and a shielding compound that is removable from the lipid composition under *in vivo* conditions. In some embodiments, the LNP also includes a second helper lipid. In some embodiments, the first helper lipid is ceramide. 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.

[00432] 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.

[00433] 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.

[00434] In another embodiment, the LNP includes at least one cationic lipid having Formula VI:

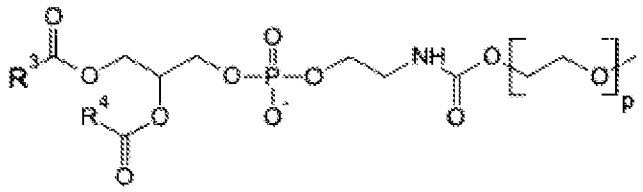


(Formula VI),

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:

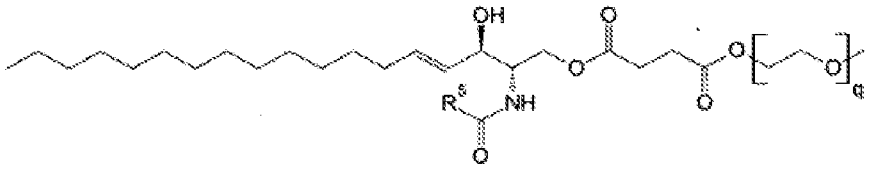
a PEGylated phosphoethanolamine of Formula VII:



(Formula VII),

wherein R^3 and R^4 are individually and independently linear C13-C17 alkyl, and p is any integer between 15 to 130;

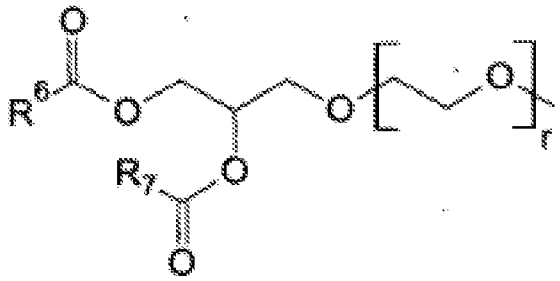
a PEGylated ceramide of Formula VIII:



(Formula VIII);

wherein R^5 is linear C7-C15 alkyl, and q is any number between 15 to 130; and

a PEGylated diacylglycerol of Formula IX:



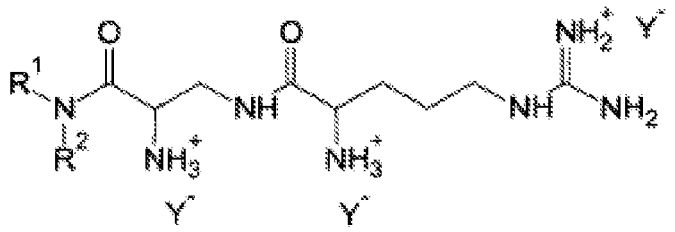
(Formula IX),

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.

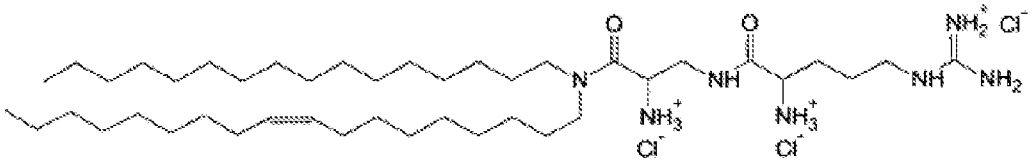
[00435] In some embodiments, R^1 and R^2 are different from each other. In some embodiments, R^1 is palmityl and R^2 is oleyl. In some embodiments, R^1 is lauryl and R^2 is myristyl. In some embodiments, R^1 and R^2 are the same. In some embodiments, each of R^1

and R^2 is individually and independently selected from the group consisting of C12 alkyl, C14 alkyl, C16 alkyl, C18 alkyl, C12 alkenyl, C14 alkenyl, C16 alkenyl and C18 alkenyl. In some embodiments, each of C12 alkenyl, C14 alkenyl, C16 alkenyl and CI 8 alkenyl comprises one or two double bonds. In some embodiments, C18 alkenyl is C18 alkenyl with one double bond between C9 and CIO. In some embodiments, CI8 alkenyl is cis-9-octadecyl.

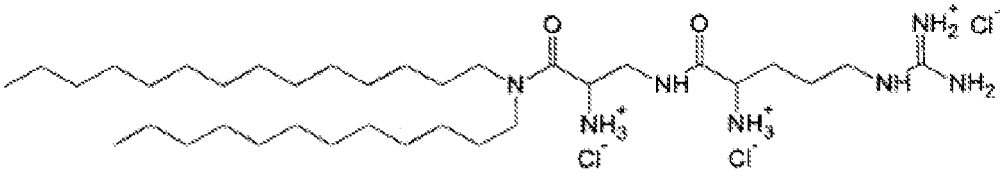
[00436] In some embodiments, the cationic lipid is a compound of Formula X:



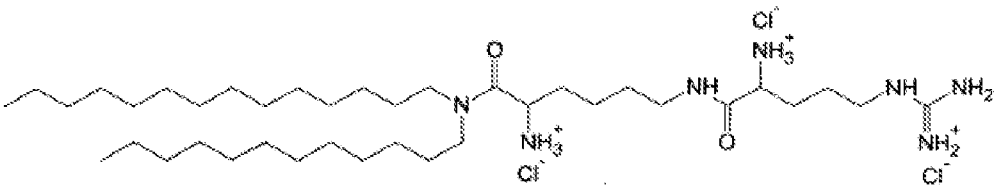
(Formula X). In some embodiments, Y^- is selected from halogenids, acetate and trifluoroacetate. In some embodiments, the cationic lipid is D-arginyl-2,3-diamino propionic acid-N-palmityl-N-oleyl-amide trihydrochloride of Formula III:



(Formula III). In some embodiments, the cationic lipid is D-arginyl-2,3-diamino propionic acid-N-lauryl-N-myri sty1-amide trihydrochloride of Formula IV:



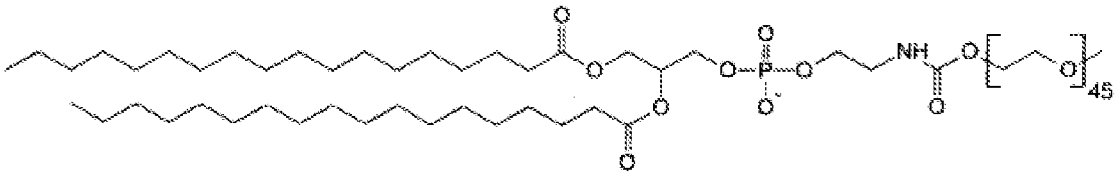
(Formula IV). In some embodiments, the cationic lipid is D-arginyl-lysine-N-lauryl-N-myristyl-amide trihydrochloride of Formula V:



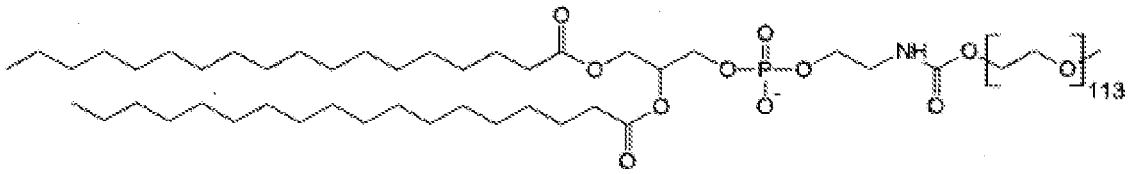
(Formula V).

[00437] In some embodiments, the sterol compound is cholesterol. In some embodiments, the sterol compound is stigmasterin.

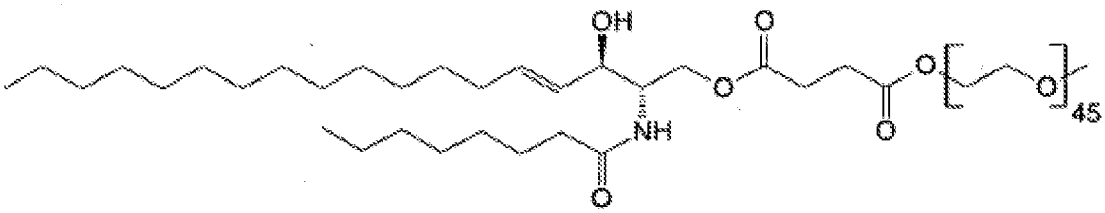
[00438] In some embodiments, the PEG moiety of the PEGylated lipid has a molecular weight from about 800 to 5,000 Da. In some embodiments, the molecular weight of the PEG moiety of the PEGylated lipid is about 800 Da. In some embodiments, the molecular weight of the PEG moiety of the PEGylated lipid 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, C15 alkyl and C17 alkyl. In some embodiments, the PEGylated phosphoethanolamine of Formula VII is l,2-distearoyl-s«-glycero-3-phosphoethanolamine-N-[methoxy(poly ethylene glycol)-2000] (ammonium salt):



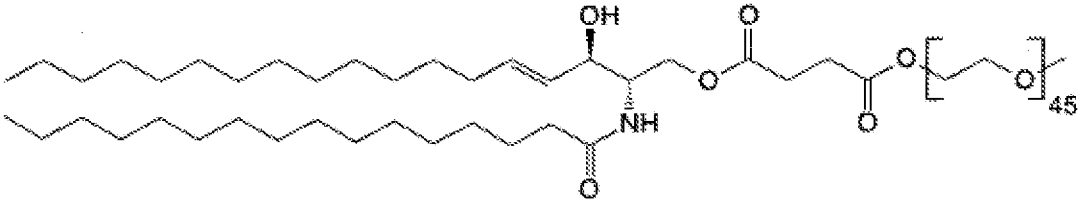
(Formula XI). In some embodiments, the PEGylated phosphoethanolamine of Formula VII is l,2-distearoyl-*s*«-glycero-3-phosphoethanolamine-N-[methoxy(polyethylene glycol)-5000] (ammonium salt):



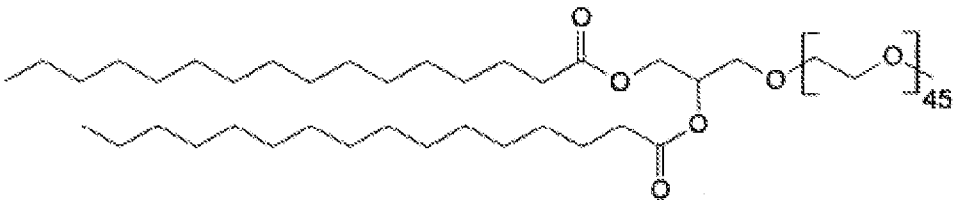
(Formula XII). In some embodiments, the PEGylated lipid is a PEGylated ceramide of Formula VIII, wherein R^5 is linear C7-C15 alkyl, and q is any integer from 18, 19 or 20, or from 44, 45 or 46 or from 113, 114 or 115. In some embodiments, R^5 is linear C7 alkyl. In some embodiments, R^5 is linear C15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



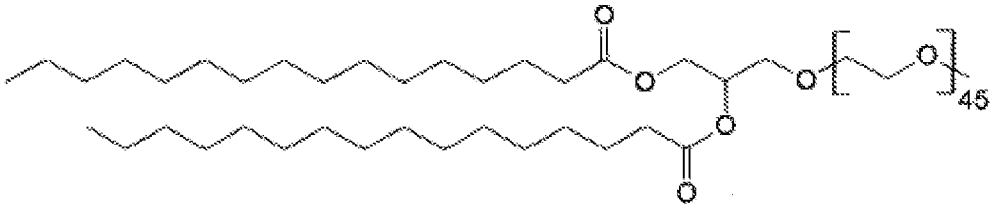
(Formula XIII). In some embodiments, the PEGylated ceramide of Formula VIII is N-palmitoyl-sphingosine- 1- {succinyl [methoxy(poly ethylene glycol)2000] }



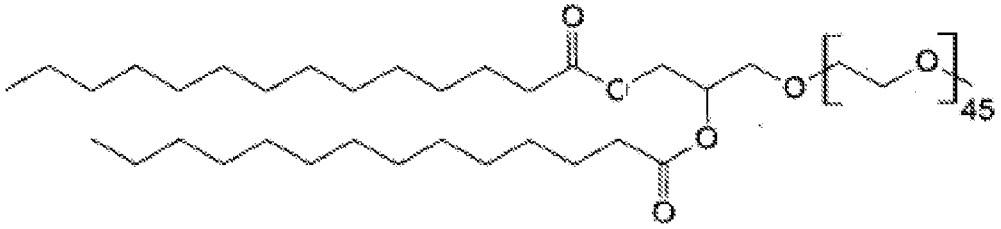
(Formula XIV). In some embodiments, the PEGylated lipid is a PEGylated diacylglycerol of Formula IX, wherein each of R^6 and R^7 is individually and independently linear C11-C17 alkyl, and r is any integer from 18, 19 or 20, or from 44, 45 or 46 or from 113, 114 or 115. In some embodiments, R^6 and R^7 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(poly ethylene glycol)2000]:



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



(Formula XVI). In some embodiments, the PEGylated diacylglycerol of Formula IX is:



(Formula XVII). 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.

[00439] 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.

[00440] Exosomes

[00441] 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. (201 1, 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.

[00442] 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.

[00443] Next, Alvarez-Erviti et al. investigated the possibility of loading modified exosomes with exogenous cargoes using electroporation protocols adapted for nanoscale applications. As electroporation for membrane particles at the nanometer scale is not well-characterized, nonspecific Cy5-labeled RNA was used for the empirical optimization of the electroporation protocol. The amount of encapsulated RNA was assayed after ultracentrifugation and lysis of exosomes. Electroporation at 400 V and 125 μT resulted in the greatest retention of RNA and was used for all subsequent experiments.

[00444] 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 6 1%, [+ or -] 13%> respectively, $P < 0.01$). Moreover, Applicants demonstrated a significant decrease (55%, $P < 0.05$) in the total [beta]-amyloid 1-42 levels, a main component of the amyloid plaques in Alzheimer's pathology, in the RVG-exosome-treated animals. The decrease observed was greater than the β-amyloid 1-40 decrease demonstrated in normal mice after intraventricular injection of BACE1 inhibitors. Alvarez-Erviti et al. carried out 5'-rapid amplification of cDNA ends (RACE) on BACE1 cleavage product, which provided evidence of RNAi-mediated knockdown by the siRNA.

[00445] Finally, Alvarez-Erviti et al. investigated whether RNA-RVG exosomes induced immune responses *in vivo* by assessing IL-6, IP-10, TNFa and IFN-a 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 CD-functionalized 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.

[00446] El-Andaloussi et al. (Nature Protocols 7,21 12-2126(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.

[00447] 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.

[00448] 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 OOOg 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.

[00449] Liposomes

[00450] Delivery or administration according to the invention can be performed with liposomes. Liposomes are spherical vesicle structures composed of a uni- or multilamellar

lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes have gained considerable attention as drug delivery carriers because they are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB) (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 201 1, Article ID 469679, 12 pages, 201 1. doi: 10.1 155/201 1/469679 for review).

[00451] Liposomes can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes as drug carriers. Although liposome formation is spontaneous when a lipid film is mixed with an aqueous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 201 1, Article ID 469679, 12 pages, 201 1. doi:10.1 155/201 1/469679 for review).

[00452] Several other additives may be added to liposomes in order to modify their structure and properties. For instance, either cholesterol or sphingomyelin may be added to the liposomal mixture in order to help stabilize the liposomal structure and to prevent the leakage of the liposomal inner cargo. Further, liposomes are prepared from hydrogenated egg phosphatidylcholine or egg phosphatidylcholine, cholesterol, and dicetyl phosphate, and their mean vesicle sizes were adjusted to about 50 and 100 nm. (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 201 1, Article ID 469679, 12 pages, 201 1. doi: 10.1 155/201 1/469679 for review).

[00453] A liposome formulation may be mainly comprised of natural phospholipids and lipids such as l,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 l,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) increases the stability (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 201 1, Article ID 469679, 12 pages, 201 1. doi: 10.1 155/201 1/469679 for review).

[00454] 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. Trojan Horse Liposomes may 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.

[00455] In another embodiment, the CD-functionalized 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), l,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).

[00456] 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-1 16 derived liver tumors (see, e.g., Li, Gene Therapy (2012) 19, 775-780). The SNALP liposomes may be prepared by formulating D-Lin-DMA and PEG-C-DMA with distearoylphosphatidylcholine (DSPC), Cholesterol and siRNA using a 25:1 lipid/siRNA ratio and a 48/40/10/2 molar ratio of Cholesterol/D-Lin-DMA/DSPC/PEG-C-DMA. The resulted SNALP liposomes are about 80-100 nm in size.

[00457] 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]-l,2-

dimyrestyloxypropylamine, and cationic l,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.

[00458] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich), l,2-distearoyl-sn-glycero-3-phosphocholine (DSPC; Avanti Polar Lipids Inc.), PEG-cDMA, and l,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.

[00459] 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.

[00460] 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.

[00461] Alnylam Pharmaceuticals has similarly advanced ALN-TTROl, 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-TTROl 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-lra 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-TTROl, was observed at 1 mg/kg.

[00462] 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_2HP0_4, 1 mM KH_2P0_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 CD-functionalized CRISPR Cas system of the present invention.

[00463] Other Lipids

[00464] Other cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[l,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- l-(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 0.1 1±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.

[00465] 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 (201 1)) describes the use of lipid envelopes to deliver RNA. Use of lipid envelopes is also preferred in the present invention.

[00466] In another embodiment, lipids may be formulated with the CD-functionalized 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.201 1.3) 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.

[00467] 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.

[00468] The CD-functionalized 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.

[00469] Nanomerics' technology addresses bioavailability challenges for a broad range of therapeutics, including low molecular weight hydrophobic drugs, peptides, and nucleic acid based therapeutics (plasmid, siRNA, miRNA). Specific administration routes for which the technology has demonstrated clear advantages include the oral route, transport across the blood-brain-barrier, delivery to solid tumours, as well as to the eye. See, e.g., Mazza et al., 2013, ACS Nano. 2013 Feb 26;7(2): 1016-26; Uchegbu and Siew, 2013, J Pharm Sci. 102(2):305-10 and Lalatsa et al., 2012, J Control Release. 2012 Jul 20; 161(2):523-36.

[00470] 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 synthesised 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(C"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.

[00471] US Patent Publication No. 20050019923 is based upon the observation that, contrary to earlier reports, cationic dendrimers, such as polypropylenimine dendrimers, display suitable properties, such as specific targeting and low toxicity, for use in the targeted delivery of bioactive molecules, such as genetic material. In addition, derivatives of the cationic dendrimer also display suitable properties for the targeted delivery of bioactive molecules. See also, Bioactive Polymers, US published application 20080267903, which discloses "Various polymers, including cationic polyamine polymers and dendrimeric polymers, are shown to possess anti-proliferative activity, and may therefore be useful for treatment of disorders characterised by undesirable cellular proliferation such as neoplasms and tumours, inflammatory disorders (including autoimmune disorders), psoriasis and atherosclerosis. The polymers may be used alone as active agents, or as delivery vehicles for other therapeutic agents, such as drug molecules or nucleic acids for gene therapy. In such cases, the polymers' own intrinsic anti-tumour activity may complement the activity of

the agent to be delivered." The disclosures of these patent publications may be employed in conjunction with herein teachings for delivery of CD-functionalized CRISPR Cas system(s) or component(s) thereof or nucleic acid molecule(s) coding therefor.

[00472] Supercharged proteins

[00473] 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 CD-functionalized 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. The creation and characterization of supercharged proteins has been reported in 2007 (Lawrence et al., 2007, Journal of the American Chemical Society 129, 101 10-101 12).

[00474] 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, 6111-61 16) (However, pilot experiments varying the dose of protein and RNA should be performed to optimize the procedure for specific cell lines): (1) One day before treatment, plate 1×10^5 cells per well in a 48-well plate. (2) On the day of treatment, dilute purified +36 GFP protein in serumfree media to a final concentration 200nM. Add RNA to a final concentration of 50nM. Vortex to mix and incubate at room temperature for lOmin. (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.

[00475] It has been further found +36 GFP to be an effective plasmid delivery reagent in a range of cells. As plasmid DNA is a larger cargo than siRNA, proportionately more +36 GFP protein is required to effectively complex plasmids. For effective plasmid delivery Applicants have developed a variant of +36 GFP bearing a C-terminal HA2 peptide tag, a known endosome-disrupting peptide derived from the influenza virus hemagglutinin protein. The following protocol has been effective in a variety of cells, but as above it is advised that plasmid DNA and supercharged protein doses be optimized for specific cell lines and delivery applications: (1) One day before treatment, plate 1×10^5 per well in a 48-well plate. (2) On the day of treatment, dilute purified p36 GFP protein in serumfree media to a final concentration 2 mM. Add lmg of plasmid DNA. Vortex to mix and incubate at room temperature for lOmin. (3) During incubation, aspirate media from cells and wash once with PBS. (4) Following incubation of p36 GFP and plasmid DNA, gently add the protein-DNA complexes to cells. (5) Incubate cells with complexes at 37 C for 4h. (6) Following incubation, aspirate the media and wash with PBS. Incubate cells in serum-containing media and incubate for a further 24-48h. (7) Analyze plasmid delivery (e.g., by plasmid-driven gene expression) as appropriate.

[00476] 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 CD-functionalized CRISPR Cas system of the present invention. These systems in conjunction with herein teaching can be employed in the delivery of CD-functionalized CRISPR Cas system(s) or component(s) thereof or nucleic acid molecule(s) coding therefor

[00477] Cell Penetrating Peptides (CPPs)

[00478] In yet another embodiment, cell penetrating peptides (CPPs) are contemplated for the delivery of the CD-functionalized 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 CD-

functionalized CRISPR Cas system or the entire CD-functionalized 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.

[00479] 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).

[00480] 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 CD-functionalized CRISPR-Cas system or components thereof. That CPPs can be employed to deliver the CD-functionalized 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, 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.

[00481] Aerosol delivery

[00482] 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.

[00483] Packaging and Promoters

[00484] The promoter used to drive CRISPR-Cas protein and cytidine deaminase coding nucleic acid molecule expression can include AAV ITR, which can serve as a promoter. This is advantageous for eliminating the need for an additional promoter element (which can take up space in the vector). The additional space freed up can be used to drive the expression of additional elements (gRNA, etc.). Also, ITR activity is relatively weaker, so can be used to reduce potential toxicity due to over expression of Cpf1.

[00485] For ubiquitous expression, promoters that can be used include: CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains, etc. For brain or other CNS expression, Synapsinl can be used for all neurons, CaMKIIalpha can be used for excitatory neurons, GAD67 or GAD65 or VGAT can be used for GABAergic neurons. For liver expression, Albumin promoter can be used. For lung expression, SP-B can be used. For endothelial cells, ICAM can be used. For hematopoietic cells, IFNbeta or CD45 can be used. For Osteoblasts, the OG-2 can be used.

[00486] The promoter used to drive guide RNA can include Pol III promoters such as U6 or HI, as well as use of Pol II promoter and intronic cassettes to express guide RNA.

[00487] Adeno associated virus (AAV)

[00488] The CRISPR-Cas protein, cytidine deaminase, 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 Cpfl and cytidine deaminase 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.

[00489] In terms of in vivo delivery, AAV is advantageous over other viral vectors for a couple of reasons: low toxicity (this may be due to the purification method not requiring

ultra centrifugation of cell particles that can activate the immune response); and low probability of causing insertional mutagenesis because it doesn't integrate into the host genome.

[00490] AAV has a packaging limit of 4.5 or 4.75 Kb. This means that Cpfl as well as a promoter and transcription terminator have to be all fit into the same viral vector. Constructs larger than 4.5 or 4.75 Kb will lead to significantly reduced virus production. SpCas9 is quite large, the gene itself is over 4.1 Kb, which makes it difficult for packing into AAV. Therefore embodiments of the invention include utilizing homologs of Cpfl that are shorter.

[00491] As to AAV, the AAV can be AAV1, AAV2, AAV5 or any combination thereof. One can select the AAV of the AAV with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. The herein promoters and vectors are preferred individually. A tabulation of certain AAV serotypes as to these cells (see Grimm, D. et al, J. Virol. 82: 5887-591 1 (2008)) is as follows:

Cell Line	AAV-1	AAV-2	AAV-3	AAV-4	AAV-5	AAV-6	AAV-8	AAV-9
Huh-7	13	100	2.5	0.0	0.1	10	0.7	0.0
HEK293	25	100	2.5	0.1	0.1	5	0.7	0.1
HeLa	3	100	2.0	0.1	6.7	1	0.2	0.1
HepG2	3	100	16.7	0.3	1.7	5	0.3	ND
HeplA	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
HT1 180	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

[00492] Lentiviruses

[00493] Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells. The most commonly known lentivirus is

the human immunodeficiency virus (HIV), which uses the envelope glycoproteins of other viruses to target a broad range of cell types.

[00494] Lentiviruses may be prepared as follows. After cloning pCasESlO (which contains a lentiviral transfer plasmid backbone), HEK293FT at low passage (p=5) were seeded in a T-75 flask to 50% confluence the day before transfection in DMEM with 10% fetal bovine serum and without antibiotics. After 20 hours, media was changed to OptiMEM (serum-free) media and transfection was done 4 hours later. Cells were transfected with 10 μg of lentiviral transfer plasmid (pCasESlO) and the following packaging plasmids: 5 μg of pMD2.G (VSV-g pseudotype), and 7.5ug of psPAX2 (gag/pol/rev/tat). Transfection was done in 4mL OptiMEM with a cationic lipid delivery agent (50uL Lipofectamine 2000 and lOOul Plus reagent). After 6 hours, the media was changed to antibiotic-free DMEM with 10%, fetal bovine serum. These methods use serum during cell culture, but serum-free methods are preferred.

[00495] Lentivirus may be purified as follows. Viral supernatants were harvested after 48 hours. Supernatants were first cleared of debris and filtered through a 0.45um low protein binding (PVDF) filter. They were then spun in a ultracentrifuge for 2 hours at 24,000 rpm. Viral pellets were resuspended in 50ul of DMEM overnight at 4C. They were then aliquotted and immediately frozen at -80°C.

[00496] In another embodiment, minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV) are also contemplated, especially for ocular gene therapy (see, e.g., Balagaan, J Gene Med 2006; 8: 275 - 285). In another embodiment, RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostatin and angiostatin that is delivered via a subretinal injection for the treatment of the web form of age-related macular degeneration is also contemplated (see, e.g., Binley et al., HUMAN GENE THERAPY 23:980-991 (September 2012)) and this vector may be modified for the CD-functionalized CRISPR-Cas system of the present invention.

[00497] 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 CD-functionalized 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 μηοΙ/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.).

[00498] Lentiviral vectors have been disclosed as in the treatment for Parkinson's Disease, see, e.g., US Patent Publication No. 20120295960 and US Patent Nos. 7303910 and 7351585. Lentiviral vectors have also been disclosed for the treatment of ocular diseases, see e.g., US Patent Publication Nos. 20060281 180, 20090007284, US201 101 17189; US200900 17543; US20070054961, US201003 17109. Lentiviral vectors have also been disclosed for delivery to the brain, see, e.g., US Patent Publication Nos. US201 10293571; US201 10293571, US20040013648, US20070025970, US200901 11106 and US Patent No. US7259015.

[00499] Application in Non-Animal Organisms

[00500] The CD-functionalized CRISPR system(s) (e.g., single or multiplexed) can be used in conjunction with recent advances in crop genomics. The systems described herein can be used to perform efficient and cost effective plant gene or genome interrogation or editing or manipulation—for instance, for rapid investigation and/or selection and/or interrogations and/or comparison and/or manipulations and/or transformation of plant genes or genomes; e.g., to create, identify, develop, optimize, or confer trait(s) or characteristic(s) to plant(s) or to transform a plant genome. There can accordingly be improved production of plants, new plants with new combinations of traits or characteristics or new plants with enhanced traits. The CD-functionalized CRISPR system 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 Cpfl effector protein system 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). Emodiments 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. 1186/1746-481 1-9-39); 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.2013.1 14; 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/sstl 19. Epub 2013 Aug 17; Xu, "Gene targeting using the Agrobacterium turnefaciens-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. 201 1 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 applciations, including those mentioned herein.

[00501] Application of CD-functionalized CRISPR system to plants and yeast

[00502] 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.

[00503] The methods for genome editing using the CD-functionalized 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 preferred embodiments, target plants and plant cells for engineering include, but are not limited to, those monocotyledonous and dicotyledonous plants, such as crops including grain crops (e.g., wheat, maize, rice, millet, barley), fruit crops (e.g., tomato, apple, pear, strawberry, orange), forage crops (e.g., alfalfa), root vegetable crops (e.g., carrot, potato, sugar beets, yam), leafy vegetable crops (e.g., lettuce, spinach); flowering plants (e.g., petunia, rose, chrysanthemum), conifers and pine trees (e.g., pine fir, spruce); plants used in phytoremediation (e.g., heavy metal accumulating plants); oil crops (e.g., sunflower, rape seed) and plants used for experimental purposes (e.g., Arabidopsis). Thus, the methods and systems 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, Violates, 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.

[00504] The CD-functionalized CRISPR systems and methods of use described herein can be used over a broad range of plant species, included in the non-limitative list of dicot, monocot or gymnosperm genera hereunder: *Atropa, Alseodaphne, Anacardium, Arachis, Beilschmiedia, Brassica, 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.*

[00505] The CD-functionalized CRISPR systems and methods of use 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.

[00506] A part of a plant, i.e., a "plant tissue" may be treated according to the methods of the present invention to produce an improved plant. Plant tissue also encompasses plant cells. The term "plant cell" as used herein refers to individual units of a living plant, either in an intact whole plant or in an isolated form grown in in vitro tissue cultures, on media or agar, in suspension in a growth media or buffer or as a part of higher organized unites, such as, for example, plant tissue, a plant organ, or a whole plant.

[00507] 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.

[00508] 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 descendents of any of these, such as cuttings or seed.

[00509] 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.

[00510] 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.

[00511] 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.

[00512] 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.

[00513] 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).

[00514] 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.

[00515] In some embodiments, the fungal cell is a polyploid cell. As used herein, a "polyploid" cell may refer to any cell whose genome is present in more than one copy. A polyploid cell may refer to a type of cell that is naturally found in a polyploid state, or it may refer to a cell that has been induced to exist in a polyploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A polyploid cell may refer to a cell whose entire genome is polyploid, or it may refer to a cell that is polyploid in a particular genomic locus of interest. Without wishing to be bound to theory, it is thought that the abundance of guideRNA may more often be a rate-limiting component in genome engineering of polyploid cells than in haploid cells, and thus the methods using the CD-functionalized CRISPR system described herein may take advantage of using a certain fungal cell type.

[00516] 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.

[00517] As used herein, a "yeast expression vector" refers to a nucleic acid that contains one or more sequences encoding an RNA and/or polypeptide and may further contain any desired elements that control the expression of the nucleic acid(s), as well as any elements that enable the replication and maintenance of the expression vector inside the yeast cell. Many suitable yeast expression vectors and features thereof are known in the art; for example, various vectors and techniques are illustrated in in Yeast Protocols, 2nd edition, Xiao, W., ed. (Humana Press, New York, 2007) and Buckholz, R.G. and Gleeson, M.A. (1991) Biotechnology (NY) 9(1 1): 1067-72. Yeast vectors may contain, without limitation, a centromeric (CEN) sequence, an autonomous replication sequence (ARS), a promoter, such as an RNA Polymerase III promoter, operably linked to a sequence or gene of interest, a terminator such as an RNA polymerase III terminator, an origin of replication, and a marker gene (e.g., auxotrophic, antibiotic, or other selectable markers). Examples of expression vectors for use in yeast may include plasmids, yeast artificial chromosomes, 2μ plasmids, yeast integrative plasmids, yeast replicative plasmids, shuttle vectors, and episomal plasmids.

[00518] Stable integration of CD-functionalized CRISPR system components in the genome of plants and plant cells

[00519] In particular embodiments, it is envisaged that the polynucleotides encoding the components of the CD-functionalized CRISPR system are introduced for stable integration into the genome of a plant cell. In these embodiments, the design of the transformation vector or the expression system can be adjusted depending on for when, where and under what conditions the guide RNA and/or fusion protein of cytidine deaminase and Cpfl are expressed.

[00520] In particular embodiments, it is envisaged to introduce the components of the CD-functionalized CRISPR system stably into the genomic DNA of a plant cell. Additionally or alternatively, it is envisaged to introduce the components of the CD-functionalized CRISPR system for stable integration into the DNA of a plant organelle such as, but not limited to a plastid, e mitochondrion or a chloroplast.

[00521] The expression system for stable integration into the genome of a plant cell may contain one or more of the following elements: a promoter element that can be used to express the RNA and/or fusion protein of cytidine deaminase and Cpfl in a plant cell; a 5' untranslated region to enhance expression ; an intron element to further enhance expression in certain cells, such as monocot cells; a multiple-cloning site to provide convenient

restriction sites for inserting the guide RNA and/or the fusion protein of cytidine deaminase and Cpfl encoding sequences and other desired elements; and a 3' untranslated region to provide for efficient termination of the expressed transcript.

[00522] 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.

[00523] In a particular embodiment, a CD-functionalized CRISPR expression system comprises at least: 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 a nucleotide sequence encoding a fusion protein of cytidine deaminase and Cpfl, wherein components (a) or (b) are located on the same or on different constructs, and whereby the different nucleotide sequences can be under control of the same or a different regulatory element operable in a plant cell.

[00524] DNA construct(s) containing the components of the CD-functionalized CRISPR system, and, where applicable, template sequence may be introduced into the genome of a plant, plant part, or plant cell by a variety of conventional techniques. The process generally comprises the steps of selecting a suitable host cell or host tissue, introducing the construct(s) into the host cell or host tissue, and regenerating plant cells or plants therefrom.

[00525] 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(l):l l-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).).

[00526] In particular embodiments, the DNA constructs containing components of the CD-functionalized CRISPR system may be introduced into the plant by *Agrobacterium*-mediated transformation. The DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The foreign DNA can be incorporated into the genome of plants by infecting the plants or by incubating plant protoplasts with Agrobacterium bacteria, containing one or more Ti

(tumor-inducing) plasmids. (see e.g. Fraley et al., (1985), Rogers et al., (1987) and U.S. Pat. No. 5,563,055).

[00527] Plant promoters

[00528] In order to ensure appropriate expression in a plant cell, the components of the CD-functionalized CRISPR system described herein are typically placed under control of a plant promoter, i.e. a promoter operable in plant cells. The use of different types of promoters is envisaged.

[00529] 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 CD-functionalized CRISPR components are expressed under the control of a constitutive promoter, such as the cauliflower mosaic virus 35S promoter issue-preferred promoters can be utilized to target enhanced expression in certain cell types within a particular plant tissue, for instance vascular cells in leaves or roots or in specific cells of the seed. Examples of particular promoters for use in the CD-functionalized CRISPR system are found in Kawamata et al., (1997) Plant Cell Physiol 38:792-803; Yamamoto et al., (1997) Plant J 12:255-65; Hire et al, (1992) Plant Mol Biol 20:207-18,Kuster et al, (1995) Plant Mol Biol 29:759-72, and Capana et al., (1994) Plant Mol Biol 25:681 -91.

[00530] Inducible promoters can be of interest to express one or more of the components of the CD-functionalized CRISPR system under limited circumstances to avoid non-specific activity of the deaminase. In particular embodiments, one or more elements of the CD-functionalized CRISPR system are expressed under control of an inducible promoter. Examples of promoters that are inducible and that allow for spatiotemporal control of gene editing or gene expression may use a form of energy. The form of energy may include but is not limited to sound energy, electromagnetic radiation, chemical energy and/or thermal energy. Examples of inducible systems include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc), or

light inducible systems (Phytochrome, LOV domains, or cryptochrome)., such as a Light Inducible Transcriptional Effector (LITE) that direct changes in transcriptional activity in a sequence-specific manner. The components of a light inducible system may include a fusion protein of cytidine deaminase and Cpfl, a light-responsive cytochrome heterodimer (e.g. from Arabidopsis thaliana). 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.

[00531] 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-11-27, WO93/01294), activated by hydrophobic electrophilic compounds used as pre-emergent herbicides, and the tobacco PR-1 a promoter (Ono et al., (2004) Biosci Biotechnol Biochem 68:803-7) activated by salicylic acid. Promoters which are regulated by antibiotics, such as tetracycline-inducible and tetracycline-repressible promoters (Gatz et al., (1991) Mol Gen Genet 227:229-37; U.S. Patent Nos. 5,814,618 and 5,789,156) can also be used herein.

[00532] Translocation to and/or expression in specific plant organelles

[00533] The expression system may comprise elements for translocation to and/or expression in a specific plant organelle.

[00534] Chloroplast targeting

[00535] In particular embodiments, it is envisaged that the CD-functionalized CRISPR system is used to specifically modify chloroplast genes or to ensure expression in the chloroplast. For this purpose use is made of chloroplast transformation methods or compartimentalization of the CD-functionalized CRISPR components to the chloroplast. For instance, the introduction of genetic modifications in the plastid genome can reduce biosafety issues such as gene flow through pollen.

[00536] 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 pastid can be used as described in WO20 1006 1186.

[00537] Alternatively, it is envisaged to target one or more of the CD-functionalized CRISPR components to the plant chloroplast. This is achieved by incorporating in the expression construct a sequence encoding a chloroplast transit peptide (CTP) or plastid transit peptide, operably linked to the 5' region of the sequence encoding the fusion protein of cytidine deaminase and Cpfl. The CTP is removed in a processing step during translocation into the chloroplast. Chloroplast targeting of expressed proteins is well known to the skilled artisan (see for instance Protein Transport into Chloroplasts, 2010, Annual Review of Plant Biology, Vol. 61: 157-180) . In such embodiments it is also desired to target the 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 CD-functionalized CRISPR system components.

[00538] Introduction of polynucleotides encoding the CD-functionalized CRISPR system in Algal cells.

[00539] Transgenic algae (or other plants such as rape) may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol) or other products. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[00540] US 8945839 describes a method for engineering Micro-Algae *{Chlamydomonas reinhardtii* cells) species) using Cas9. Using similar tools, the methods of the CD-functionalized CRISPR system described herein can be applied on *Chlamydomonas* species and other algae. In particular embodiments, a CRISPR-Cas protein (e.g., Cpfl), cytidine deaminase (which may be fused to the CRISPR-Cas protein or an aptamer-binding adaptor protein), and guide RNA are introduced in algae expressed using a vector that expresses the fusion protein of cytidine deaminase and Cpfl under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2 -tubulin. Guide RNA is optionally delivered using a vector containing T7 promoter. Alternatively, Cpfl mRNA and in vitro transcribed guide RNA can be delivered to algal cells. Electroporation protocols are available to the skilled person such as the standard recommended protocol from the GeneArt Chlamydomonas Engineering kit.

[00541] Introduction of CD-functionalized CRISPR system components in yeast cells

[00542] In particular embodiments, the invention relates to the use of the CD-functionalized CRISPR system for genome editing of yeast cells. Methods for transforming yeast cells which can be used to introduce polynucleotides encoding the CD-functionalized CRISPR system components are described in 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.

[00543] Transient expression of CD-functionalized CRISPR system components in plants and plant cell

[00544] In particular embodiments, it is envisaged that the guide RNA and/or CRISPR-Cas gene are transiently expressed in the plant cell. In these embodiments, the CD-functionalized CRISPR system can ensure modification of a target gene only when both the guide RNA, the CRISPR-Cas protein (e.g., Cpfl), and cytidine deaminase (which may be fused to the CRISPR-Cas protein or an aptamer-binding adaptor protein), are present in a cell, such that genomic modification can further be controlled. As the expression of the CRISPR-Cas protein is transient, plants regenerated from such plant cells typically contain no foreign DNA. In particular embodiments the CRISPR-Cas protein is stably expressed by the plant cell and the guide sequence is transiently expressed.

[00545] In particular embodiments, the CD-functionalized CRISPR system components can be introduced in the plant cells using a plant viral vector (Scholthof et al. 1996, Annu Rev Phytopathol. 1996;34:299-323). In further particular embodiments, said viral vector is a vector from a DNA virus. For example, geminivirus (e.g., cabbage leaf curl virus, bean yellow dwarf virus, wheat dwarf virus, tomato leaf curl virus, maize streak virus, tobacco leaf curl virus, or tomato golden mosaic virus) or nanovirus (e.g., Faba bean necrotic yellow virus). In other particular embodiments, said viral vector is a vector from an RNA virus. For example, 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.

[00546] In particular embodiments, the vector used for transient expression of CD-functionalized CRISPR system 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 guide

RNAs in stable transgenic plants expressing a CRISPR enzyme (Scientific Reports 5, Article number: 14926 (2015), doi: 10.1038/srep14926).

[00547] In particular embodiments, double-stranded DNA fragments encoding the guide RNA and/or the CRISPR-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.)

[00548] In other embodiments, an RNA polynucleotide encoding the CRISPR-Cas protein (e.g., Cpfl) and/or cytidine deaminase (which may be fused to the CRISPR-Cas protein or an aptamer-binding adaptor 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).

[00549] Combinations of the different methods described above are also envisaged.

[00550] Delivery of CD-functionalized CRISPR system components to the plant cell

[00551] In particular embodiments, it is of interest to deliver one or more components of the CD-functionalized CRISPR system directly to the plant cell. This is of interest, inter alia, for the generation of non-transgenic plants (see below). In particular embodiments, one or more of the CD-functionalized CRISPR system components is prepared outside the plant or plant cell and delivered to the cell. For instance in particular embodiments, the CRISPR-Cas protein is prepared in vitro prior to introduction to the plant cell. The CRISPR-Cas protein can be prepared by various methods known by one of skill in the art and include recombinant production. After expression, the CRISPR-Cas protein is isolated, refolded if needed, purified and optionally treated to remove any purification tags, such as a His-tag. Once crude, partially purified, or more completely purified CRISPR-Cas protein is obtained, the protein may be introduced to the plant cell.

[00552] In particular embodiments, the CRISPR-Cas protein is mixed with guide RNA targeting the gene of interest to form a pre-assembled ribonucleoprotein.

[00553] The individual components or pre-assembled ribonucleoprotein can be introduced into the plant cell via electroporation, by bombardment with CRISPR-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).

[00554] In particular embodiments, the CD-functionalized CRISPR system components are introduced into the plant cells using nanoparticles. The components, either as protein or nucleic acid or in a combination thereof, can be uploaded onto or packaged in nanoparticles and applied to the plants (such as for instance described in WO 2008042156 and US 20130185823). In particular, embodiments of the invention comprise nanoparticles uploaded with or packed with DNA molecule(s) encoding the CRISPR-Cas protein (e.g., Cpfl), DNA molecule(s) encoding cytidine deaminase (which may be fused to the CRISPR-Cas protein or an aptamer-binding adaptor protein), and DNA molecules encoding the guide RNA and/or isolated guide RNA as described in WO20 150894 19.

[00555] Further means of introducing one or more components of the CD-functionalized CRISPR system to the plant cell is by using cell penetrating peptides (CPP). Accordingly, in particular, embodiments the invention comprises compositions comprising a cell penetrating peptide linked to the CRISPR-Cas protein. In particular embodiments of the present invention, the CRISPR-Cas protein and/or guide RNA is coupled to one or more CPPs to effectively transport them inside plant protoplasts. Ramakrishna (Genome Res. 2014 Jun;24(6): 1020-7 for Cas9 in human cells). In other embodiments, the CRISPR-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 are then regenerated to plant cells and further to plants. CPPs are generally described as short peptides of fewer than 35 amino acids either derived from proteins or from chimeric sequences which are capable of transporting biomolecules across cell membrane in a receptor independent manner. CPP can be cationic peptides, peptides having hydrophobic sequences, amphipatic peptides, peptides having proline-rich and anti-microbial sequence, and chimeric or bipartite peptides (Pooga and Langel 2005). CPPs are able to penetrate biological membranes and as such trigger the movement of various biomolecules across cell membranes into the cytoplasm and to improve their intracellular routing, and hence facilitate interaction of the biolomolecule with the target. Examples of CPP include amongst others: Tat, a nuclear transcriptional activator protein required for viral replication by HIV

typel, 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.

[00556] Use of the CD-functionalized CRISPR system to make genetically modified non-transgenic plants

[00557] 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 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.

[00558] In particular embodiments, this is ensured by transient expression of the CD-functionalized CRISPR system components . In particular embodiments one or more of the components are expressed on one or more viral vectors which produce sufficient CRISPR-Cas protein, cytidine deaminase, and guide RNA to consistently steadily ensure modification of a gene of interest according to a method described herein.

[00559] In particular embodiments, transient expression of CD-functionalized CRISPR system constructs is ensured in plant protoplasts and thus not integrated into the genome. The limited window of expression can be sufficient to allow the CD-functionalized CRISPR system to ensure modification of a target gene as described herein.

[00560] In particular embodiments, the different components of the CD-functionalized CRISPR system are introduced in the plant cell, protoplast or plant tissue either separately or in mixture, with the aid of pariculate delivering molecules such as nanoparticles or CPP molecules as described herein above.

[00561] The expression of the CD-functionalized CRISPR system components can induce targeted modification of the genome, by deaminase activity of the cytidine deaminase. The different strategies described herein above allow CRISPR-mediated targeted genome editing without requiring the introduction of the CD-functionalized CRISPR systemt components into the plant genome. Components which are transiently introduced into the plant cell are typically removed upon crossing.

[00562] Plant cultures and regeneration

[00563] 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.).

[00564] 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 CD-functionalized CRISPR system 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).

[00565] Generation of plants with enhanced agronomic traits

[00566] The CD-functionalized CRISPR systems provided herein can be used to introduce targeted A-G and T-C mutations. 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.

[00567] In particular embodiments, the CD-functionalized CRISPR system as described herein is used to introduce targeted A-G and T-C mutations. Such mutation can be a nonsense mutation (e.g., premature stop codon) or a missense mutation (e.g., encoding different amino acid residue). This is of interest where the A-G and T-C mutations in certain endogenous genes can confer or contribute to a desired trait.

[00568] 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.

[00569] 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 CRISPR-Cas protein, the cytidine deaminase, and the guide RNA, where the delivering is via Agrobacterium. The polynucleotide sequence encoding the components of the CD-functionalized 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.

[00570] 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.

[00571] Use of CD-functionalized CRISPR system to modify polyploid plants

[00572] 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 CD-functionalized CRISPR system 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 defences 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).

[00573] Examplary genes conferring agronomic traits

[00574] In particular embodiments, the invention encompasses methods which involve targeted A-G and T-C mutations in endogenous genes and their regulatory elements, such as listed below:

[00575] 1. Genes that confer resistance to pests or diseases:

[00576] Plant 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/10.1 101/064824 Epub. July 23, 2016 (tomato plants with deletions in the S1DMR6-1 which is normally upregulated during pathogen infection).

[00577] Genes conferring resistance to a pest, such as soybean cyst nematode. See e.g., PCT Application WO 96/30517; PCT Application WO 93/19181.

[00578] Bacillus thuringiensis proteins see, e.g., Geiser et al., Gene 48: 109 (1986).

[00579] Lectins, see, for example, Van Damme et al., Plant Molec. Biol. 24:25 (1994.

[00580] Vitamin-binding protein, such as avidin, see PCT application US93/06487, teaching the use of avidin and avidin homologues as larvicides against insect pests.

[00581] 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. 2 1:985 (1993)), Sumitani et al., Biosci. Biotech. Biochem. 57:1243 (1993) and U.S. Pat. No. 5,494,813.

[00582] 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).

[00583] 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.

[00584] 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).

[00585] Enzymes responsible for a hyperaccumulation of a monoterpene, a sesquiterpene, a steroid, hydroxamic acid, a phenylpropanoid derivative or another nonprotein molecule with insecticidal activity.

[00586] 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. 2 1 :673 (1993).

[00587] 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).

[00588] Viral-invasive proteins or a complex toxin derived therefrom. See Beachy et al., Ann. rev. Phytopathol. 28:451 (1990).

[00589] 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).

[00590] A developmental-arrestive protein produced in nature by a plant. For example, Logemann et al., Bio/Technology 10:305 (1992).

[00591] In plants, pathogens are often host-specific. For example, some Fusarium species will causes tomato wilt but attacks 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 CD-functionalized CRISPR 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 CD-functionalized 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.

[00592] 2. Genes involved in plant diseases, such as those listed in WO 2013046247:

[00593] 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;

[00594] 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, Coiletotrichum acutatum, Phytophtora cactorum;

[00595] Pear diseases: Venturia nashicola, V. pirina, Alternaria alternata Japanese pear pathotype, Gymnosporangium haraeanum, Phytophtora cactorum;

[00596] Peach diseases: Monilinia fructicola, Cladosporium carpophilum, Phomopsis sp.;

[00597] Grape diseases: Elsinoe ampelina, Glomerella cingulata, Uninula necator, Phakopsora ampelopsidis, Guignardia bidwellii, Plasmopara viticola;

[00598] Persimmon diseases: Gloesporium kaki, Cercospora kaki, Mycosphaerela nawae;

[00599] Gourd diseases: Colletotrichum lagenarium, Sphaerotheca fuliginea, Mycosphaerella melonis, Fusarium oxysporum, Pseudoperonospora cubensis, Phytophthora sp., Pythium sp.;

[00600] Tomato diseases: Alternaria solani, Cladosporium fulvum, Phytophthora infestans; Pseudomonas syringae pv. Tomato; Phytophthora capsici; Xanthomonas

[00601] Eggplant diseases: Phomopsis vexans, Erysiphe cichoracearum; Brassicaceous vegetable diseases: Alternaria japonica, Cercosporella brassicae, Plasmodiophora brassicae, Peronospora parasitica;

[00602] Welsh onion diseases: Puccinia allii, Peronospora destructor;

[00603] Soybean diseases: Cercospora kikuchii, Elsinoe glycines, Diaporthe phaseolorum var. sojae, Septoria glycines, Cercospora sojina, Phakopsora pachyrhizi, Phytophthora sojae, Rhizoctonia solani, Coiynespora casiicola, Sclerotinia sclerotiorum;

[00604] Kidney bean diseases: Colletrichum lindemthianum;

[00605] Peanut diseases: Cercospora personata, Cercospora arachidicola, Sclerotium rolfsii;

[00606] Pea diseases pea: Erysiphe pisi;

[00607] Potato diseases: Alternaria solani, Phytophthora infestans, Phytophthora erythroseptica, Spongospora subterranean, f. sp. Subterranean;

[00608] Strawberry diseases: Sphaerotheca humuli, Glomerella cingulata;

[00609] Tea diseases: Exobasidium reticulatum, Elsinoe leucospila, Pestalotiopsis sp., Colletotrichum theae-sinensis;

[00610] Tobacco diseases: Alternaria longipes, Erysiphe cichoracearum, Colletotrichum tabacum, Peronospora tabacina, Phytophthora nicotianae;

[00611] Rapeseed diseases: Sclerotinia sclerotiorum, Rhizoctonia solani;

[00612] Cotton diseases: Rhizoctonia solani;

[00613] Beet diseases: Cercospora beticola, Thanatephorus cucumeris, Thanatephorus cucumeris, Aphanomyces cochlioides;

[00614] Rose diseases: Diplocarpon rosae, Sphaerotheca pannosa, Peronospora sparsa;

[00615] Diseases of chrysanthemum and asteraceae: Bremia lactuca, Septoria chrysanthemi-indici, Puccinia horiana;

[00616] Diseases of various plants: Pythium aphanidermatum, Pythium debarianum, Pythium graminicola, Pythium irregulare, Pythium ultimum, Botrytis cinerea, Sclerotinia sclerotiorum;

[00617] Radish diseases: Alternaria brassicicola;

[00618] Zoysia diseases: Sclerotinia homeocarpa, Rhizoctonia solani;

[00619] Banana diseases: Mycosphaerella fijiensis, Mycosphaerella musicola;

[00620] Sunflower diseases: Plasmopara halstedii;

[00621] 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;

[00622] Virus diseases of various plants mediated by Polymixa spp., Olpidium spp., or the like.

[00623] 3. Examples of genes that confer resistance to herbicides:

[00624] 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.

[00625] 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.

[00626] 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).

[00627] 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 7,1 12,665.

[00628] Hydroxyphenylpyruvatedioxygenases (HPPD) inhibitors, naturally occuring 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.

[00629] 4. Examples of genes involved in Abiotic stress tolerance:

[00630] 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 WO 00/04173 or, WO/2006/045633.

[00631] 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.

[00632] 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.

[00633] 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 97/1 1188, WO 97/26362, WO 97/32985, WO 97/42328, WO 97/44472, WO 97/45545,

WO 98/27212, WO 98/40503, W099/58688, 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 00/1 1192, 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 94/1 1520, 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.

[00634] 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 2009/0144850, US 2007/0266453, and WO 2002/08391 1. 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).

[00635] 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.

[00636] In addition to targeted mutation of single genes, CD-functionalized CRISPR system 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.

[00637] Use of CD-functionalized CRISPR system to create male sterile plants

[00638] 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 and systems 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 CD-functionalized CRISPR system 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.

[00639] Increasing the fertility stage in plants

[00640] In particular embodiments, the methods and systems provided herein are 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)

[00641] Use of CD-functionalized CRISPR system to generate genetic variation in a crop of interest

[00642] 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 this application of the CD-functionalized CRISPR system a library of guide RNAs targeting different locations in the plant genome is provided and is introduced into plant cells together with the CRISPR-Cas protein and cytidine deaminase. In this way 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 the method is a method for the generation of stress-tolerant crop varieties

[00643] Use of CD-functionalized CRISPR to affect fruit-ripening

[00644] 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-l-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.

[00645] Thus in particular embodiments, the methods comprise the use of the CD-functionalized CRISPR system 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.

[00646] Increasing storage life of plants

[00647] In particular embodiments, the 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 a gene that prevents the accumulation of reducing sugars in potato tubers. Upon high-temperature processing, these reducing sugars react with free amino acids, resulting in brown, bitter-tasting products and elevated levels of acrylamide, which is a potential carcinogen. In particular embodiments, the methods provided herein are used to reduce or inhibit expression of the vacuolar invertase

gene (VInv), which encodes a protein that breaks down sucrose to glucose and fructose (Clasen et al. DOI: 10. 1111/pbi. 12370).

[00648] The use of the CD-functionalized CRISPR system to ensure a value added trait

[00649] In particular embodiments the CD-functionalized CRISPR system 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.

[00650] Examples of nutritionally improved crops include (Newell-McGloughlin, Plant Physiology, July 2008, Vol. 147, pp. 939-953):

[00651] 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).

[00652] 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).

[00653] 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 1140-1 145 [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 2 1 205S-21 1S [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 2 1 988-992), 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)

[00654] Carbohydrates, such as Fructans described for Chicory (Smeekens (1997) Trends Plant Sci 2 286-287, Sprenger et al. (1997) FEBS Lett 400 355-358, Sevenier 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),

[00655] 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 2 1 191S-198S, Cahoon et al. (2003) Nat Biotechnol 2 1 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 2 1 177-181), Tomato (Rosati et al. (2000) Plant J 24 413-419, Fraser et al. (2001) J Sci Food Agric 8 1 822-827, Mehta et al. (2002) Nat Biotechnol 20 613-618, Diaz 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.

[00656] 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 femlic acids, resveratrol; United Press International (2002)); and

[00657] 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 wheate (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).

[00658] 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:

[00659] Carotenoids, such as a-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

[00660] Lutein present in green vegetables which contributes to maintenance of healthy vision

[00661] Lycopene present in tomato and tomato products, which is believed to reduce the risk of prostate cancer

[00662] Zeaxanthin, present in citrus and maize, which contributes to mainteance of healthy vision

[00663] 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)

[00664] 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

[00665] 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

[00666] Glucosinolates, indoles, isothiocyanates, such as Sulforaphane, present in Cruciferous vegetables (broccoli, kale), horseradish, which neutralize free radicals, may reduce risk of cancer

[00667] 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

[00668] Plant stand s/sterols present in maize, soy, wheat and wooden oils which May reduce risk of coronary heart disease by lowering blood cholesterol levels

[00669] Fructans, inulins, fructo-oligosaccharides present in Jerusalem artichoke, shallot, onion powder which may improve gastrointestinal health

[00670] Saponins present in soybean, which may lower LDL cholesterol

[00671] Soybean protein present in soybean which may reduce risk of heart disease

[00672] 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.

[00673] Sulfides and thiols such as diallyl sulphide present in onion, garlic, olive, leek and scallon and Allyl methyl trisulfide, dithiolthiones present in cruciferous vegetables which may lower LDL cholesterol, helps to maintain healthy immune system

[00674] Tannins, such as proanthocyanidins, present in cranberry, cocoa, which may improve urinary tract health, may reduce risk of CVD and high blood pressure.

[00675] 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.

[00676] 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 CD-functionalized 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 CD-functionalized 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 CD-functionalized CRISPR system are described herein above.

[00677] 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).

[00678] 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 Dofl 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 AtDof 1.1 (OBP2) up-regulated all steps in the glucosinolate biosynthetic pathway in Arabidopsis (Skirycz et al., 2006 Plant J 47: 10-24).

[00679] Reducing allergen in plants

[00680] In particular embodiments the 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: 11676-1 1680).

[00681] Peanut allergies and allergies to legumes generally are a real and serious health concern. The CD-functionalized CRISPR system of the present invention can be used to identify and then mutate genes encoding allergenic proteins of such legumes. Without limitation as to such genes and proteins, Nicolaou et al. identifies allergenic proteins in peanuts, soybeans, lentils, peas, lupin, green beans, and mung beans. See, Nicolaou et al., Current Opinion in Allergy and Clinical Immunology 201 1;1 1(3):222).

[00682] Screening methods for endogenous genes of interest

[00683] 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 CD-functionalized 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.

[00684] Further applications of the CD-functionalized CRISPR system in plants and yeasts

[00685] Use of CD-functionalized CRISPR system in biofuel production

[00686] 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.

[00687] Enhancing plant properties for biofuel production

[00688] In particular embodiments, the methods using the CD-functionalized CRISPR system as described herein are used to alter the properties of the cell wall in order to facilitate access by key hydrolysing agents for a more efficient release of sugars for fermentation. In particular embodiments, the biosynthesis of cellulose and/or lignin are modified. Cellulose is the major component of the cell wall. The biosynthesis of cellulose and lignin are co-regulated. By reducing the proportion of lignin in a plant the proportion of cellulose can be increased. In particular embodiments, the methods described herein are used to downregulate lignin biosynthesis in the plant so as to increase fermentable carbohydrates. More particularly, the methods described herein are used to downregulate at least a first lignin biosynthesis gene selected from the group consisting of 4-coumarate 3-hydroxylase (C3H), phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), hydroxycinnamoyl transferase (HCT), caffeic acid O-methyltransferase (COMT), caffeoyl CoA 3-O-methyltransferase (CCoAOMT), ferulate 5- hydroxylase (F5H), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl CoA-reductase (CCR), 4- coumarate-CoA ligase (4CL), monolignol-lignin-specific glycosyltransferase, and aldehyde dehydrogenase (ALDH) as disclosed in WO 2008064289 A2.

[00689] 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.

[00690] Modifying yeast for Biofuel production

[00691] In particular embodiments, the CD-functionalized CRISPR system provided herein is used for bioethanol production by recombinant micro-organisms. For instance, the CD-functionalized CRISPR system 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. In some embodiments, the CD-functionalized CRISPR system is used to modify endogenous metabolic pathways which compete with the biofuel production pathway.

[00692] Accordingly, in more particular embodiments, the methods described herein are used to modify a micro-organism as follows: 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.

[00693] Modifying Algae and plants for production of vegetable oils or biofuels

[00694] 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.

[00695] According to particular embodiments of the invention, the CD-functionalized CRISPR system is used to generate lipid-rich diatoms which are useful in biofuel production.

[00696] 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 CD-functionalized CRISPR system and methods described herein can be used to specifically activate such genes in diatoms as to increase their lipid content.

[00697] 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 guide RNA 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 guide RNA expression. Hlavova et al. (Biotechnol. Adv. 2015) discusses development of species or strains of microalgae using techniques such as CRISPR to target nuclear and chloroplast genes for insertional mutagenesis and screening.

[00698] US 8945839 describes a method for engineering Micro-Algae (Chlamydomonas reinhardtii cells) species) using Cas9 . Using similar tools, the methods of the CD-functionalized CRISPR system described herein can be applied on Chlamydomonas species and other algae. In particular embodiments, a CRISPR-Cas protein (e.g., Cpfl), cytidine deaminase (which may be fused to the CRISPR-Cas protein or an aptamer-binding adaptor protein), and guide RNA are introduced in algae expressed using a vector that expresses the CRISPR-Cas protein and optionally the cytidine deaminase 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, 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.

[00699] The use of CD-functionalized CRISPR system in the generation of micro¬ organisms capable of fatty acid production

[00700] In particular embodiments, the 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"),

[00701] 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 encloding 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 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.

[00702] Additionally or alternatively, said micro-organism 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.

[00703] In particular embodiments, the micro-organism 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.*

[00704] The use of CD-functionalized CRISPR system in the generation of micro¬ organisms capable of organic acid production

[00705] The methods 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 (1-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).

[00706] 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.

[00707] The use of CD-functionalized CRISPR system in the generation of improved xylose or cellobiose utilizing yeasts strains

[00708] In particular embodiments, the CD-functionalized CRISPR 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 CD-functionalized CRISPR system into a yeast strain (for instance S288C) and strains can be selected with enhanced xylose or cellobiose utilization capacity, as described in WO2015138855.

[00709] The use of CD-functionalized CRISPR system in the generation of improved yeasts strains for use in isoprenoid biosynthesis

[00710] Tadas Jakociunas 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 CD-functionalized 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.

[00711] Improved plants and yeast cells

[00712] The present invention also provides plants and yeast cells obtainable and obtained by the methods provided herein. The improved plants obtained by the methods described herein may be useful in food or feed production through expression of genes which, for instance ensure tolerance to plant pests, herbicides, drought, low or high temperatures, excessive water, etc.

[00713] 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.

[00714] 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.

[00715] The invention also provides for improved parts of a plant. Plant parts include, but are not limited to, leaves, stems, roots, tubers, seeds, endosperm, ovule, and pollen. Plant parts as envisaged herein may be viable, nonviable, regeneratable, and/or non-regeneratable.

[00716] It is also encompassed herein to provide plant cells and plants generated according to the methods of the invention. Gametes, seeds, embryos, either zygotic or somatic, progeny or hybrids of plants comprising the genetic modification, which are produced by traditional breeding methods, are also included within the scope of the present invention. Such plants may contain a heterologous or foreign DNA sequence inserted at or instead of a target sequence. Alternatively, such plants may contain only an alteration (mutation, deletion, insertion, substitution) in one or more nucleotides. As such, such plants

will only be different from their progenitor plants by the presence of the particular modification.

[00717] 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.

[00718] The methods for genome editing using the CD-functionalized CRISPR system as 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.

[00719] 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.

[00720] The methods described herein generally 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, 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 CD-functionalized CRISPR system for plant, algae, fungi, yeast, etc. genome editing include, but are not limited to: editing of endogenous genes to confer an agricultural trait of interest. Examplary 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); Sscreening methods for endogenous genes of interest; biofuel, fatty acid, organic acid, etc production.

[00721] CD-functionalized CRISPR System Can Be Used In Non-Human Organisms

[00722] 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;l l l(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 CD-functionalized CRISPR system may be applied to a similar system.

[00723] The methods of Lee et al., (Proc Natl Acad Sci U S A. 2014 May 20;l l l(20):7260-5) may be applied to the present invention analogously as follows. Mutated pigs are produced by targeted modification of 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.

[00724] The present invention is also applicable to modifying SNPs of other animals, such as cows. 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 guide RNA sequences were cloned into the Church lab guide RNA 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 Xbal-Agel fragment from the hCas9 plasmid (encompassing the hCas9 cDNA) into the RCIScript plasmid.

[00725] Heo et al. (Stem Cells Dev. 2015 Feb l;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 GSK3P and MEK inhibitor (2i) treatment. Heo et al. observed that these bovine iPSCs are highly similar to naive 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.

[00726] 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 (http://www.animalgenome.org/cattle/maps/db.html). Thus, the present invention maybe 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.

[00727] Qingjian Zou et al. (Journal of Molecular Cell Biology Advance Access published October 12, 2015) demonstrated increased muscle mass in dogs by targeting 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 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 CD-functionalized CRISPR systems provided herein.

[00728] Livestock - Pigs

[00729] Viral targets in livestock may include, in some embodiments, porcine CD163, for example on porcine macrophages. CD163 is associated with infection (thought to be through viral cell entry) by PRRSv (Porcine Reproductive and Respiratory Syndrome virus, an arterivirus). Infection by PRRSv, especially of porcine alveolar macrophages (found in the lung), results in a previously incurable porcine syndrome ("Mystery swine disease" or "blue ear disease") that causes suffering, including reproductive failure, weight loss and high mortality rates in domestic pigs. Opportunistic infections, such as enzootic pneumonia, meningitis and ear oedema, are often seen due to immune deficiency through

loss of macrophage activity. It also has significant economic and environmental repercussions due to increased antibiotic use and financial loss (an estimated $660m per year).

[00730] As reported by Kristin M Whitworth and Dr Randall Prather et al. (Nature Biotech 3434 published online 07 December 2015) at the University of Missouri and in collaboration with Genus Pic, CD163 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.

[00731] Accordingly, in some embodiments, porcine alveolar macrophages may be targeted by the CRISPR protein. In some embodiments, porcine CD163 may be targeted by the CRISPR protein. In some embodiments, porcine CD163 may be knocked out through induction of a DSB or through insertions or deletions, for example targeting deletion or modification of exon 7, including one or more of those described above, or in other regions of the gene, for example deletion or modification of exon 5.

[00732] An edited pig and its progeny are also envisaged, for example a CD163 knock out pig. This may be for livestock, breeding or modelling purposes (i.e. a porcine model). Semen comprising the gene knock out is also provided.

[00733] CD163 is a member of the scavenger receptor cysteine-rich (SRCR) superfamily. Based on in vitro studies SRCR domain 5 of the protein is the domain responsible for unpackaging and release of the viral genome. As such, other members of the SRCR superfamily may also be targeted in order to assess resistance to other viruses. PRRSV is also a member of the mammalian arterivirus group, which also includes murine lactate dehydrogenase-elevating virus, simian hemorrhagic fever virus and equine arteritis virus. The arteriviruses share important pathogenesis properties, including macrophage

tropism and the capacity to cause both severe disease and persistent infection. Accordingly, arteriviruses, and in particular murine lactate dehydrogenase-elevating virus, simian hemorrhagic fever virus and equine arteritis virus, may be targeted, for example through porcine CD 163 or homologues thereof in other species, and murine, simian and equine models and knockout also provided.

[00734] Indeed, this approach may be extended to viruses or bacteria that cause other livestock diseases that may be transmitted to humans, such as Swine Influenza Virus (SIV) strains which include influenza C and the subtypes of influenza A known as H1N1, H1N2, H2N1, H3N1, H3N2, and H2N3, as well as pneumonia, meningitis and oedema mentioned above.

[00735] In some embodiments, the CD-functionalized CRISPR system described herein can be used to genetically modify a pig genome to inactivate one or more porcine endogenous retrovirus (PERVs) loci to facilitate clinical application of porcine-to-human xenotransplantation. See Yang *et al., Science* 350(6264): 1101-1 104 (2015), which is incorporated herein by reference in its entirey. In some embodiments, the CD-functionalized CRISPR system described herein can be used to produce a genetically modified pig that does not comprise any active porcine endogenous retrovirus (PERVs) locus.

[00736] Therapeutic Targeting with CD-functionalized CRISPR System

[00737] As will be apparent, it is envisaged that CD-functionalized CRISPR system can be used to target any polynucleotide sequence of interest. The invention provides a non-naturally occurring or engineered composition, or one or more polynucleotides encoding components of said composition, or vector or delivery systems comprising one or more polynucleotides encoding components of said composition for use in a modifying a target cell *in vivo, ex vivo* or *in vitro* and, may be conducted in a manner alters the cell such that once modified the progeny or cell line of the CRISPR modified cell retains the altered phenotype. The modified cells and progeny may be part of a multi-cellular organism such as a plant or animal with *ex vivo* or *in vivo* application of CRISPR system to desired cell types. The CRISPR invention may be a therapeutic method of treatment. The therapeutic method of treatment may comprise gene or genome editing, or gene therapy.

[00738] Adoptive Cell Therapies

[00739] The present invention also contemplates use of the CD-functionalized CRISPR system described herein 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 a and β chains with selected peptide specificity (see U.S. Patent No. 8,697,854; PCT Patent Publications: WO2003020763, WO2004033685, WO2004044004, WO20051 14215, WO2006000830, WO2008038002, WO2008039818, WO2004074322, WO20051 13595, WO2006125962, WO2013 166321, WO2013039889, WO2014018863, WO2014083173; U.S. Patent No. 8,088,379).

[00740] 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; 6,004,81 1; 6,284,240; 6,392,013; 6,410,014; 6,753,162; 8,21 1,422; and, PCT Publication W092 15322). 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 CD8a hinge domain and a CD8a transmembrane domain, to the transmembrane and intracellular signaling domains of either CD3C or FcRy (scFv-CD3C or scFv-FcRy; 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-CD3Q see U.S. Patent Nos. 8,91 1,993; 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 CD3C-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 aPTCR, 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.

[00741] 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.

[00742] 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.

[00743] 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 immunoreponsive 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.

[00744] 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.

[00745] 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.

[00746] The administration of the cells or population of cells can consist of the administration of 10^4- 10^9 cells per kg body weight, preferably 10^5 to 10^6 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 10^6 to 10^9 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.

[00747] 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.

[00748] 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 WO201 1146862; PCT Patent Publication WO201401 1987; PCT Patent Publication WO20 13040371; Zhou et al. BLOOD, 2014, 123/25:3895 - 3905; Di Stasi et al., The New England Journal of Medicine 201 1; 365:1673-1683; Sadelain M, The New England Journal of Medicine 201 1; 365:1735-173; Ramos et al., Stem Cells 28(6): 1107-15 (2010)).

[00749] In a further refinement of adoptive therapies, genome editing with a CD-functionalized 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 ULA type II and/or type I molecules, or to knockout selected genes that may inhibit the desired immune response, such as the PD1 gene.

[00750] Cells may be edited using a CD-functionalized CRISPR system as described herein. CD-functionalized 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: WO2013 176915, WO2014059173, WO2014172606, WO2014184744, and WO2014191 128). Editing may result in inactivation of a gene.

[00751] 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, a 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 a 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 a 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 TCRa or TCRP 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.

[00752] 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 l;1 12(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 a-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.

[00753] 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.

[00754] 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).

[00755] WO20 14 172606 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.

[00756] 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, SMALM, 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.

[00757] In other embodiments, at least two genes are edited. Pairs of genes may include, but are not limited to PDl and TCRa, PDl and TCRp, CTLA-4 and TCRa, CTLA-4 and TCRp, LAG3 and TCRa, LAG3 and TCRp, Tim3 and TCRa, Tim3 and TCRp, BTLA and TCRa, BTLA and TCRp, BY55 and TCRa, BY55 and TCRp, TIGIT and TCRa, TIGIT and TCRp, B7H5 and TCRa, B7H5 and TCRp, LAIR1 and TCRa, LAIR1 and TCRp, SIGLEC10 and TCRa, SIGLEC10 and TCRp, 2B4 and TCRa, 2B4 and TCRp.

[00758] 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.*

[00759] The practice of the present invention employs techniques known in the field 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.).

[00760] Correction of Disease-Associated Mutations and Pathogenic SNPs

[00761] In one aspect, the invention described herein provides methods for modifying an cytidine residue at a target locus with the aim of remedying and/or preventing a diseased condition that is or is likely to be caused by a T-to-C or A-to-G point mutation or a pathogenic single nucleotide polymorphism (SNP).

[00762] Pathogenic T-to-C or A-to-G mutations/SNPs associated with various diseases are reported in the ClinVar database, including but not limited to genetic diseases, cancer, metabolic diseases, or lysosomal storage diseases. Accordingly, an aspect of the invention relates to a method for correcting one or more pathogenic T-to-C or A-to-G mutations/SNPs associated with any of these diseases, as discussed below.

[00763] In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic T-to-C or A-to-G mutations/SNPs reported in the ClinVar database. In some embodiments, the methods, systems, and compositions described herein are used to correct one or more pathogenic T-to-C or A-to-G mutations/SNPs associated with any of the diseases or disorders disclosed in WO20 17/070632, titled "Nucleobase Editor and Uses Thereof," which is incorporated herein by reference in its entirety. Exemplary diseases or disorders that may be treated include, without limitation, 3-Methylglutaconic aciduria type 2, 46,XY gonadal dysgenesis, 4-Alpha-hydroxyphenylpyruvate hydroxylase deficiency, 6- pyruvoyl-tetrahydropterin synthase deficiency, achromatopsia, Acid-labile subunit deficiency, Acrodysostosis, acroerythrokeratoderma, ACTH resistance, ACTH-independent macronodular adrenal hyperplasia, Activated PBK-delta syndrome, Acute intermittent porphyria, Acute myeloid leukemia, Adams-Oliver syndrome 1/5/6, Adenylosuccinate lyase deficiency, Adrenoleukodystrophy, Adult neuronal ceroid lipofuscinosis, Adult onset ataxia with oculomotor apraxia, Advanced sleep phase syndrome, Age-related macular degeneration, Alagille syndrome, Alexander disease, Allan-Herndon-Dudley syndrome, Alport syndrome, X-linked recessive, Alternating hemiplegia of childhood, Alveolar capillary dysplasia with misalignment of pulmonary veins, Amelogenesis imperfecta, Amyloidogenic transthyretin

amyloidosis, Amyotrophic lateral sclerosis, Anemia (nonspherocytic hemolytic, due to G6PD deficiency), Anemia (sideroblastic, pyridoxine-refractory, autosomal recessive), Anonychia, Antithrombin III deficiency, Aortic aneurysm, Aplastic anemia, Apolipoprotein C2 deficiency, Apparent mineralocorticoid excess, Aromatase deficiency, Arrhythmogenic right ventricular cardiomyopathy, Familial hypertrophic cardiomyopathy, Hypertrophic cardiomyopathy, Arthrogryposis multiplex congenital, Aspartylglycosaminuria, Asphyxiating thoracic dystrophy, Ataxia with vitamin E deficiency, Ataxia (spastic), Atrial fibrillation, Atrial septal defect, atypical hemolytic-uremic syndrome, autosomal dominant CD1 1C+/CD1C+ dendritic cell deficiency, Autosomal dominant progressive external ophthalmoplegia with mitochondrial DNA deletions, Baraitser-Winter syndrome, Bartter syndrome, Basa ganglia calcification, Beckwith-Wiedemann syndrome, Benign familial neonatal seizures, Benign scapuloperoneal muscular dystrophy, Bernard Soulier syndrome, Beta thalassemia intermedia, Beta-D-mannosidosis, Bietti crystalline corneoretinal dystrophy, Bile acid malabsorption, Biotinidase deficiency, Borjeson-Forssman- Lehmann syndrome, Boucher Neuhauser syndrome, Bowen-Conradi syndrome, Brachydactyly, Brown-Vialetto-Van laere syndrome, Brugada syndrome, Cardiac arrhythmia, Cardiofaciocutaneous syndrome, Cardiomyopathy, Carnevale syndrome, Carnitine palmitoyltransferase II deficiency, Carpenter syndrome, Cataract, Catecholaminergic polymorphic ventricular tachycardia, Central core disease, Centromeric instability of chromosomes 1,9 and 16 and immunodeficiency, Cerebral autosomal dominant arteriopathy, Cerebro-oculo-facio-skeletal syndrome, Ceroid lipofuscinosis, Charcot-Marie-Tooth disease, Cholestanol storage disease, Chondrocalcinosis, Chondrodysplasia, Chronic progressive multiple sclerosis, Coenzyme Q10 deficiency, Cohen syndrome, Combined deficiency of factor V and factor VIII, Combined immunodeficiency, Combined oxidative phosphorylation deficiency, Combined partial 1 7-alpha- hydroxylase/1 7,20-lyase deficiency, Complement factor d deficiency, Complete combined 17- alpha-hydroxylase/1 7,20-lyase deficiency, Cone-rod dystrophy, Congenital contractural arachnodactyly, Congenital disorder of glycosylation, Congenital lipomatous overgrowth, Neoplasm of ovary, PIK3CA Related Overgrowth Spectrum, Congenital long QT syndrome, Congenital muscular dystrophy, Congenital muscular hypertrophy-cerebral syndrome, Congenital myasthenic syndrome, Congenital myopathy with fiber type di sproportion, Eichsfeld type congenital muscular dystrophy, Congenital stationary night blindness, Corneal dystrophy, Cornelia de Lange syndrome, Craniometaphyseal dysplasia,

Crigler Najjar syndrome, Crouzon syndrome, Cutis laxa with osteodystrophy, Cyanosis, Cystic fibrosis, Cystinosis, Cytochrome-c oxidase deficiency, Mitochondrial complex I deficiency, D-2-hydroxyglutaric aciduria, Danon disease, Deafness with labyrinthine aplasia microtia and microdontia (LAMM), Deafness, Deficiency of acetyl-CoA acetyltransferase, Deficiency of ferroxidase, Deficiency of UDPglucose-hexose-1-phosphate uridylyltransferase, Dejerine-Sottas disease, Desbuquois syndrome, DFNA, Diabetes mellitus type 2, Diabetes-deafness syndrome, Diamond-Blackfan anemia, Diastrophic dysplasia, Dihydropteridine reductase deficiency, Dihydropyrimidinase deficiency, Dilated cardiomyopathy, Disseminated atypical mycobacterial infection, Distal arthrogryposis, Distal hereditary motor neuronopathy, Donnai Barrow syndrome, Duchenne muscular dystrophy, Becker muscular dystrophy, Dyschromatosis universalis hereditaria, Dyskeratosis congenital, Dystonia, Early infantile epileptic encephalopathy, Ehlers-Danlos syndrome, Eichsfeld type congenital muscular dystrophy, Emery-Dreifuss muscular dystrophy, Enamel-renal syndrome, Epidermolysis bullosa dystrophica inversa, Epidermolysis bullosa herpetiformis, Epilepsy, Episodic ataxia, Erythrokeratodermia variabilis, Erythropoietic protoporphyria, Exercise intolerance, Exudative vitreoretinopathy, Fabry disease, Factor V deficiency, Factor VII deficiency, Factor xiii deficiency, Familial adenomatous polyposis, breast cancer, ovarian cancer, cold urticaria!, chronic infantile neurological, cutaneous and articular syndrome, hemiplegic migraine, hypercholesterolemia, hypertrophic cardiomyopathy, hypoalphalipoproteinemia, hypokalemia-hypomagnesemia, juvenile gout, hyperlipoproteinemia, visceral amyloidosis, hypophosphatemic vitamin D refractory rickets, FG syndrome, Fibrosis of extraocular muscles, Finnish congenital nephrotic syndrome, focal epilepsy, Focal segmental glomerulosclerosis, Frontonasal dysplasia, Frontotemporal dementia, Fructose-biphosphatase deficiency, Gamstorp-Wohlfart syndrome, Ganglioside sialidase deficiency, GATA-I -related thrombocytopenia, Gaucher disease, Giant axonal neuropathy, Glanzmann thrombasthenia, Glomerulocystic kidney disease, Glomerulopathy, Glucocorticoid resistance, Glucose-6- phosphate transport defect, Glutaric aciduria, Glycogen storage disease, Gorlin syndrome, Holoprosencephaly, GRACILE syndrome, Haemorrhagic telangiectasia, Hemochromatosis, Hemoglobin H disease, Hemolytic anemia, Hemophagocytic lymphohistiocytosis, Carcinoma of colon, Myhre syndrome, leukoencephalopathy, Hereditary factor IX deficiency disease, Hereditary factor VIII deficiency disease, Hereditary factor XI deficiency disease, Hereditary fructosuria,

Hereditary Nonpolyposis Colorectal Neoplasm, Hereditary pancreatitis, Hereditary pyropoikilocytosis, Elliptocytosis, Heterotaxy, Heterotopia, Histiocytic medullary reticulosis, Histiocytosis-lymphadenopathy plus syndrome, HNSHA due to aldolase A deficiency, Holocarboxylase synthetase deficiency, Homocysteinemia, Rowel-Evans syndrome, Hydatidiform mole, Hypercalciuric hypercalcemia, Hyperimmunoglobulin D, Mevalonic aciduria, Hyperinsulinemic hypoglycemia, Hyperkalemic Periodic Paralysis, Paramyotonia congenita of von Eulenburg, Hyperlipoproteinemia, Hypermanganesemia, Hypermethioninemia, Hyperphosphatasemia, Hypertension, hypomagnesemia, Hypobetalipoproteinemia, Hypocalcemia, Hypogonadotropic hypogonadism, Hypogonadotropic hypogonadism, Hypohidrotic ectodermal dysplasia, Hyper-IgM immunodeficiency, Hypohidrotic X-linked ectodermal dysplasia, Hypomagnesemia, Hypoparathyroidism, Idiopathic fibrosing alveolitis, Immunodeficiency, Immunoglobulin A deficiency, Infantile hypophosphatasia, Infantile Parkinsonism-dystonia, Insulin-dependent diabetes mellitus, Intermediate maple syrup urine disease, Ischiopatellar dysplasia, Islet cell hyperplasia, Isolated growth hormone deficiency, Isolated lutropin deficiency, Isovaleric acidemia, Joubert syndrome, Juvenile polyposis syndrome, Juvenile retinoschisis, Kallmann syndrome, Kartagener syndrome, Kugelberg-W elander disease, Lattice corneal dystrophy, Leber congenital amaurosis, Leber optic atrophy, Left ventricular noncompaction, Leigh disease, Mitochondrial complex I deficiency, Leprechaunism syndrome, Arthrogryposis, Anterior horn cell disease, Leukocyte adhesion deficiency, Leukodystrophy, Leukoencephalopathy, Ovarioleukodystrophy, L-ferritin deficiency, Li-Fraumeni syndrome, Limb-girdle muscular dystrophy- dystroglycanopathy, Loeys-Dietz syndrome, Long QT syndrome, Macrocephaly/autism syndrome, Macular corneal dystrophy, Macular dystrophy, Malignant hyperthermia susceptibility, Malignant tumor of prostate, Maple syrup urine disease, Marden Walker like syndrome, Marfan syndrome, Marie Unna hereditary hypotrichosis, Mast cell disease, Meconium ileus, Medium-chain acyl-coenzyme A dehydrogenase deficiency, Melnick- Fraser syndrome, Mental retardation, Merosin deficient congenital muscular dystrophy, Mesothelioma, Metachromatic leukodystrophy, Metaphyseal chondrodysplasia, Methemoglobinemia, methylmalonic aciduria, homocystinuria, Microcephaly, chorioretinopathy, lymphedema, Microphthalmia, Mild non-PKU hyperphenylalanemia, Mitchell-Riley syndrome, mitochondrial 3-hydroxy-3-methylglutaryl-CoA synthase deficiency, Mitochondrial complex I deficiency, Mitochondrial complex III deficiency, Mitochondrial myopathy, Mucolipidosis III,

Mucopolysaccharidosis, Multiple sulfatase deficiency, Myasthenic syndrome, Mycobacterium tuberculosis, Myeloperoxidase deficiency, Myhre syndrome, Myoclonic epilepsy, Myofibrillar myopathy, Myoglobinuria, Myopathy, Myopia, Myotonia congenital, Navajo neurohepatopathy, Nemaline myopathy, Neoplasm of stomach, Nephrogenic diabetes insipidus, Nephronophthisis, Nephrotic syndrome, Neurofibromatosis, Neutral lipid storage disease, Niemann-Pick disease, Non-ketotic hyperglycinemia, Noonan syndrome, Noonan syndrome-like disorder, Norum disease, Macular degeneration, N-terminal acetyltransferase deficiency, Oculocutaneous albinism, Oculodentodigital dysplasia, Ohdo syndrome, Optic nerve aplasia, Ornithine carbamoyltransferase deficiency, Orofaciodigital syndrome, Osteogenesis imperfecta, Osteopetrosis, Ovarian dysgenesis, Pachyonychia, Palmoplantar keratoderma, nonepidermolytic, Papillon-Lef\xc3\xa8vre syndrome, Haim-Munk syndrome, Periodontitis, Peeling skin syndrome, Pendred syndrome, Peroxisomal fatty acyl-coa reductase I disorder, Peroxisome biogenesis disorder, Pfeiffer syndrome, Phenylketonuria, Phenylketonuria, Hyperphenylalaninemia, non- PKU, Pituitary hormone deficiency, Pityriasis rubra pilaris, Polyarteritis nodosa, Polycystic kidney disease, Polycystic lipomembranous osteodysplasia, Polymicrogyria, Pontocerebellar hypoplasia, Porokeratosis, Posterior column ataxia, Primary erythromelalgia, hyperoxaluria, Progressive familial intrahepatic cholestasis, Progressive pseudorheumatoid dysplasia, Propionic acidemia, Pseudohermaphroditism, Pseudohypoaldosteronism, Pseudoxanthoma elasticum-like disorder, Purine-nucleoside phosphorylase deficiency, Pyridoxal 5-phosphate-dependent epilepsy, Renal dysplasia, retinal pigmentary dystrophy, cerebellar ataxia, skeletal dysplasia, Reticular dysgenesis, Retinitis pigmentosa, Usher syndrome, Retinoblastoma, Retinopathy, RRM2B-related mitochondrial disease, Rubinstein-Taybi syndrome, Schnyder crystalline corneal dystrophy, Sebaceous tumor, Severe congenital neutropenia, Severe myoclonic epilepsy in infancy, Severe X-linked myotubular myopathy, onychodysplasia, facial dysmorphism, hypotrichosis, Short-rib thoracic dysplasia, Sialic acid storage disease, Sialidosis, Sideroblastic anemia, Small fiber neuropathy, Smith-Magenis syndrome, Sorsby fundus dystrophy, Spastic ataxia, Spastic paraplegia, Spermatogenic failure, Spherocytosis, Sphingomyelin/cholesterol lipidosis, Spinocerebellar ataxia, Split-hand/foot malformation, Spondyloepimetaphyseal dysplasia, Platyspondylic lethal skeletal dysplasia, Squamous cell carcinoma of the head and neck, Stargardt disease, Sucrase-isomaltase deficiency, Sudden infant death syndrome, Supravalvar aortic stenosis, Surfactant metabolism dysfunction,

Tangier disease, Tatton-Brown-rahman syndrome, Thoracic aortic aneurysms and aortic dissections, Thrombophilia, Thyroid hormone resistance, TNF receptor-associated periodic fever syndrome (TRAPS), Tooth agenesis, Torsades de pointes, Transposition of great arteries, Treacher Collins syndrome, Tuberous sclerosis syndrome, Tyrosinase-negative oculocutaneous albinism, Tyrosinase-positive oculocutaneous albinism, Tyrosinemia, UDPglucose-4-epimerase deficiency, Ullrich congenital muscular dystrophy, Bethlem myopathy Usher syndrome, UV-sensitive syndrome, Van der Woude syndrome, popliteal pterygium syndrome, Very long chain acyl-CoA dehydrogenase deficiency, Vesicoureteral reflux, Vitreoretinochoroidopathy, Von Rippel-Lindau syndrome, von Willebrand disease, Waardenburg syndrome, Warsaw breakage syndrome, WFSI -Related Disorders, Wilson disease, Xeroderma pigmentosum, X-linked agammaglobulinemia, X-linked hereditary motor and sensory neuropathy, X-linked severe combined immunodeficiency, and Zellweger syndrome.

[00764] In certain embodiments, , the methods, systems, and compositions described herein are used to correct one or more pathogenic T-to-C or A-to-G mutations/SNPs as provided in the Table below.

Candidate	**Gene**	**Disease**
NM_007262.4(PARK7):c.497T>C (p.Leu166Pro)	PARK7	Parkinson disease 7
NM_174936.3(PCSK9):c.646T>C (p.Phe216Leu)	PCSK9	Hypercholesterolemia, autosomal dominant, 3
NM_000642.2(AGL):c.3083+2T>C	AGL	Glycogen storage disease type III
NM_213653.3(HFE2):c.842T>C (p.Ile281Thr)	HFE2	Hemochromatosis type 2A
NM_170707.3(LMNA):c.799T>C (p.Tyr267His)	LMNA	Primary dilated cardiomyopathy|not provided
NM_000488.3(SERPINC1):c.1141T>C (p.Ser381Pro)	SERPINC1	Antithrombin III deficiency
NM_000465.3(BARD1):c.1159T>C (p.Phe387Leu)	BARD1	Familial cancer of breast|not specified|Hereditary cancer-predisposing syndrome
NM_000030.2(AGXT):c.613T>C (p.Ser205Pro)	AGXT	Primary hyperoxaluria, type I|not provided
NM_001302946.1(TRNT1):c.668T>C (p.Ile223Thr)	TRNT1	Sideroblastic anemia with B-cell immunodeficiency, periodic fevers, and developmental delay
NM_138694.3(PKHD1):c.8068T>C	PKHD1	Autosomal recessive polycystic

(p.Trp2690Arg)		kidney disease
NM_000162.3(GCK):c.l l69T>C (p.Ile390Thr)	GCK	Maturity-onset diabetes of the young, type 2
NM_017890.4(VPS13B):c.7504+2T>C	VPS13B	Cohen syndrome
NM_000155.3(GALT):c.512T>C (p.Phel71Ser)	GALT	Deficiency of UDPglucose-hexose-1-phosphate uridylyltransferase
NM_000277. l(PAH):c.691T>C (p.Ser231Pro)	PAH	Phenylketonuria|not provided
NM_000138.4(FBNl):c.4531T>C (p.Cysl51 1Arg)	FBN1	Marfan syndrome
NM_000527.4(LDLR):c.l745T>C (p.Leu582Pro)	LDLR	Familial hypercholesterolemia

[00765] ADDITIONAL EMBODIMENTS

[00766] Embodiment 1. A method of modifying a Cytosine in a target locus of interest, comprising delivering to said locus: (a) a Cpfl nickase protein; (b) a guide molecule which comprises a guide sequence linked to a direct repeat; and (c) a cytidine deaminase protein or catalytic domain thereof; wherein said cytidine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to said Cpfl nickase protein or said guide molecule or is adapted to link thereto after delivery; wherein guide molecule forms a complex with said Cpfl nickase protein and directs said complex to bind a first DNA strand at said target locus of interest, wherein said guide sequence is capable of hybridizing with a target sequence comprising said Cytosine within said first DNA strand to form a heteroduplex, wherein said guide sequence comprises a non-pairing Adenine or Uracil at a position corresponding to said Cytosine resulting in a C-A or C-U mismatch in said heteroduplex formed; wherein said Cpfl nickase protein nicks a second DNA strand at said target locus of interest displaced by formation of said heteroduplex; and wherein said cytidine deaminase protein or catalytic domain thereof deaminates said Cytosine in said heteroduplex.

[00767] Embodiment 2. The method of Embodiment 1, wherein said cytidine deaminase protein or catalytic domain thereof is fused to N- or C-terminus of said Cpfl nickase protein.

[00768] Embodiment 3. The method of Embodiment 2, wherein said cytidine deaminase protein or catalytic domain thereof is fused to said Cpfl nickase protein by a linker.

[00769] Embodiment 4. The method of Embodiment 3, wherein said linker is $(GGGGS)_{3-11}$ (SEQ ID NOS: 1-9), GSG_5 (SEQ ID NO: 10) or LEP GEKP YKCPEC GK SF SQ SGAL TRHQRTHTR (SEQ ID NO: 11).

[00770] Embodiment 5. The method of Embodiment 1, wherein said cytidine deaminase protein or catalytic domain thereof is linked to an adaptor protein, and said guide molecule or said Cpfl nickase protein comprises an aptamer sequence capable of binding to said adaptor protein.

[00771] Embodiment 6. The method of Embodiment 5, wherein said adaptor sequence is selected from MS2, PP7, Q[illegible]3, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M i l,

MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1.

[00772] Embodiment 7. The method of Embodiment 1, wherein said cytidine deaminase protein or catalytic domain thereof is inserted into an internal loop of said Cpf1 nickase protein.

[00773] Embodiment 8. The method of any of Embodiments 1-7, wherein said Cpf1 nickase protein comprises a mutation in the Nuc domain.

[00774] Embodiment 9. The method of Embodiment 8, wherein said Cpf1 nickase protein comprises a mutation corresponding to R1226A in AsCpf1.

[00775] Embodiment 10. The method of any of Embodiments 1-7, wherein said Cpf1 nickase protein has at least part of the Nuc domain removed.

[00776] Embodiment 11. The method of any of Embodiments 1-10, wherein said guide molecule binds to said Cpf1 nickase protein and is capable of forming said heteroduplex of about 24 nt with said target sequence.

[00777] Embodiment 12. The method of any of Embodiments 1-10, wherein said guide molecule binds to said Cpf1 nickase protein and is capable of forming said heteroduplex of more than 24 nt with said target sequence.

[00778] Embodiment 13. The method of any of Embodiments 1-12, wherein said guide sequence comprises at least one further non-pairing nucleotide with said target sequence, adjacent to said non-pairing Adenine or Uracil.

[00779] Embodiment 14. The method of Embodiment 13, wherein said guide sequence comprises a stretch of three to five consecutive non-pairing nucleotides with said target sequence.

[00780] Embodiment 15. The method of any of the preceding Embodiments, wherein said cytidine deaminase protein or catalytic domain thereof is a human, rat or lamprey cytidine deaminase protein or catalytic domain thereof.

[00781] Embodiment 16. The method of Embodiment 15, wherein said cytidine deaminase protein or catalytic domain thereof is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase, an activation-induced deaminase (AID), or a cytidine deaminase 1 (CDA1).

[00782] Embodiment 17. The method of Embodiment 16, wherein said cytidine deaminase protein or catalytic domain thereof is an APOBEC1 deaminase comprising one or more mutations corresponding to W90A, W90Y, R 118A, 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.

[00783] Embodiment 18. The method of any of the preceding Embodiments, wherein said Cpfl nickase protein and optionally said cytidine deaminase protein or catalytic domain thereof comprise one or more heterologous nuclear localization signal(s) (NLS(s)).

[00784] Embodiment 19. The method of any of the preceding Embodiments, wherein said cytidine deaminase protein or catalytic domain thereof is delivered together with a uracil glycosylase inhibitor (UGI), where said UGI is covalently linked to said cytidine deaminase protein or catalytic domain thereof and/or said Cpfl nickase protein.

[00785] Embodiment 20. The method of any of the preceding Embodiments, wherein said method comprises, determining said target sequence of interest and selecting a cytidine deaminase protein or catalytic domain thereof which most efficiently deaminates said Cytosine present in said target sequence.

[00786] Embodiment 21. The method of any of the preceding Embodiments, wherein said Cpfl nickase protein is obtained from a Cpfl nuclease derived from a bacterial species selected from the group consisting of Francisella tularensis, Prevotella albensis, Lachnospiraceae bacterium, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium, Parcubacteria bacterium, Smithella sp., Acidaminococcus sp., Lachnospiraceae bacterium, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi, Leptospira inadai, Porphyromonas crevioricanis, Prevotella disiens and Porphyromonas macacae, Succinivibrio dextrinosolvens, Prevotella disiens, Flavobacterium branchiophilum, Helcococcus kunzii, Eubacterium sp., Microgenomates (Roizmanbacteria) bacterium, Flavobacterium sp., Prevotella brevis, Moraxella caprae, Bacteroidetes oral, Porphyromonas cansulci, Synergistes jonesii, Prevotella bryantii, Anaerovibrio sp., Butyrivibrio fibrisolvens, Candidatus Methanomethylophilus, Butyrivibrio sp., Oribacterium sp., Pseudobutyrivibrio ruminis and Proteocatella sphenisci.

[00787] Embodiment 22. The method of Embodiment 21, wherein said Cpfl nickase protein is a FnCpfl nickase and recognizes a PAM sequence of TTN, wherein N is A/C/G

or T, or said Cpfl nickase protein is a PaCpflp, LbCpfl or AsCpfl nickase and recognizes a PAM sequence of TTTV, wherein V is A/C or G.

[00788] Embodiment 23. The method of Embodiment 21, wherein said Cpfl nickase protein has been modified to and recognizes an altered PAM sequence.

[00789] Embodiment 24. The method of any of the preceding Embodiments, wherein said target locus of interest is within a cell.

[00790] Embodiment 25. The method of Embodiment 24, wherein said cell is a eukaryotic cell.

[00791] Embodiment 26. The method of Embodiment 24, wherein said cell is a non-human animal cell.

[00792] Embodiment 27. The method of Embodiment 24, wherein said cell is a human cell.

[00793] Embodiment 28. The method of Embodiment 24, wherein said cell is a plant cell.

[00794] Embodiment 29. The method of any of the preceding Embodiments, wherein said target locus of interest is within an animal.

[00795] Embodiment 30. The method of any of the preceding Embodiments, wherein said target locus of interest is within a plant.

[00796] Embodiment 31. The method of any of the preceding Embodiments, wherein said target locus of interest is comprised in a DNA molecule in vitro.

[00797] Embodiment 32. The method of any of the preceding Embodiments, wherein said components (a), (b) and (c) are delivered to the cell as a ribonucleoprotein complex.

[00798] Embodiment 33. The method of any of the preceding Embodiments, wherein said components (a), (b) and (c) are delivered to the cell as one or more polynucleotide molecules.

[00799] Embodiment 34. The method of Embodiment 33, wherein said one or more polynucleotide molecules comprise one or more mRNA molecules encoding components (a) and/or (c).

[00800] Embodiment 35. The method of Embodiment 33, wherein said one or more polynucleotide molecules are comprised within one or more vectors.

[00801] Embodiment 36. The method of Embodiment 35, wherein said one or more polynucleotide molecules comprise one or more regulatory elements operably configured to express said Cpfl nickase protein, said guide molecule, and said cytidine deaminase protein or catalytic domain thereof, optionally wherein said one or more regulatory elements comprise inducible promoters.

[00802] Embodiment 37. The method of any of Embodiments 32-36, wherein said one or more polynucleotide molecules or said ribonucleoprotein complex are delivered via particles, vesicles, or one or more viral vectors.

[00803] Embodiment 38. The method of Embodiment 37, wherein said particles comprise a lipid, a sugar, a metal or a protein.

[00804] Embodiment 39. The method of Embodiment 38, wherein said particles comprise lipid nanoparticles.

[00805] Embodiment 40. The method of Embodiment 37, wherein said vesicles comprise exosomes or liposomes.

[00806] Embodiment 41. The method of Embodiment 37, wherein said one or more viral vectors comprise one or more of adenovirus, one or more lentivirus or one or more adeno-associated virus.

[00807] Embodiment 42. The method of any of the preceding Embodiments, which is a method of modifying a cell, a cell line or an organism by manipulation of one or more target sequences at genomic loci of interest.

[00808] Embodiment 43. The method of Embodiment 42, wherein deamination of said Cytosine at said target locus of interest remedies a disease caused by a T→C or A→G point mutation or pathogenic SNP.

[00809] Embodiment 44. The method of Embodiment 42, wherein deamination of said Cytosine at said target locus of interest inactivates a target gene at said target locus.

[00810] Embodiment 45. A modified cell obtained from the method of any of the preceding Embodiments, or progeny thereof, wherein said cell comprises a Uracil or a Thymine in replace of said Cytosine in said target locus of interest compared to a corresponding cell not subjected to said method.

[00811] Embodiment 46. The modified cell or progeny thereof of Embodiment 45, wherein said cell is a eukaryotic cell.

[00812] Embodiment 47. The modified cell or progeny thereof of Embodiment 45, wherein said cell is an animal cell.

[00813] Embodiment 48. The modified cell or progeny thereof of Embodiment 45, wherein said cell is a human cell.

[00814] Embodiment 49. The modified cell or progeny thereof of Embodiment 45, wherein said cell is a therapeutic T cell.

[00815] Embodiment 50. The modified cell or progeny thereof of Embodiment 45, wherein said cell is an antibody-producing B cell.

[00816] Embodiment 51. The modified cell or progeny thereof of Embodiment 45, wherein said cell is a plant cell.

[00817] Embodiment 52. A non-human animal comprising said modified cell of Embodiment 47.

[00818] Embodiment 53. A plant comprising said modified cell of Embodiment 51.

[00819] Embodiment 54. A method for cell therapy, comprising administering to a patient in need thereof said modified cell of any of Embodiments 45-50, wherein presence of said modified cell remedies a disease in said patient.

[00820] Embodiment 55. An engineered, non-naturally occurring system suitable for modifying a Cytosine in a target locus of interest, comprising: a guide molecule which comprises a guide sequence, or a nucleotide sequence encoding said guide molecule; a Cpf1 nickase protein, or a nucleotide sequence encoding said Cpf1 nickase protein; a cytidine deaminase protein or catalytic domain thereof, or a nucleotide sequence encoding said cytidine deaminase protein or catalytic domain thereof; wherein said cytidine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to said Cpf1 nickase protein or said guide molecule or is adapted to link thereto after delivery; wherein said guide sequence is capable of hybridizing with a target sequence comprising a Cytosine on a first DNA strand at said target locus to form a heteroduplex, wherein said guide sequence comprises a non-pairing Adenine or Uracil at a position corresponding to said Cytosine resulting in a C-A or C-U mismatch in said heteroduplex formed; and wherein said Cpf1 nickase protein is capable of nicking a second DNA strand at said target locus.

[00821] Embodiment 56. An engineered, non-naturally occurring vector system suitable for modifying an Cytosine in a target locus of interest, comprising said nucleotide sequences of a), b) and c) of Embodiment 55.

[00822] Embodiment 57. The engineered, non-naturally occurring vector system of Embodiment 56, comprising one or more vectors comprising: a first regulatory element operably linked to a nucleotide sequence encoding said guide molecule which comprises a guide sequence, a second regulatory element operably linked to a nucleotide sequence encoding said Cpfl nickase protein; and a nucleotide sequence encoding said cytidine deaminase protein 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 cytidine deaminase protein or catalytic domain thereof is operably linked to a third regulatory element, said cytidine protein or catalytic domain thereof is adapted to link to said guide molecule or said Cpfl nickase protein after expression; wherein components (a), (b) and (c) are located on the same or different vectors of said system.

[00823] Embodiment 58. An in vitro or ex vivo host cell or progeny thereof or cell line or progeny thereof comprising a system of any of Embodiments 55-57.

[00824] Embodiment 59. The host cell or progeny thereof or cell line or progeny thereof of Embodiment 58, wherein said cell is a eukaryotic cell.

[00825] Embodiment 60. The host cell or progeny thereof or cell line or progeny thereof of Embodiment 58, wherein said cell is an animal cell.

[00826] Embodiment 61. The host cell or progeny thereof or cell line or progeny thereof of Embodiment 58, wherein said cell is a human cell.

[00827] Embodiment 62. The host cell or progeny thereof or cell line or progeny thereof of Embodiment 58, wherein said cell is a plant cell.

WORKING EXAMPLES

[00828] EXAMPLE 1

[00829] Cytosine deaminases (CDs) typically deaminates cytosines at specific sites in single stranded DNA. Previous efforts have attempted to fuse CD to Cas9 to achieve RNA-guided cytosine deamination on genomic DNA, with the CD editing a non-targeted DNA strand displaced by the binding of a Cas9-guide RNA complex to a targeted DNA strand.

[00830] The fact that CD can be directed to effect cytosine deamination on single stranded DNA of interest presents a unique opportunity to develop an RNA guided CD by taking advantage of the the bubble formed between the guide RNA and its complementary DNA target in the presence of mismatching nucleotides. By using Cpfl nickase to recruit a CD, the CD enzyme will then act on the cytosine in the accessible single stranded DNA in the mismatching bubble, whereas the Cpfl nickase will nick the non-targeted DNA strand displaced by the binding of a Cpfl -guide RNA complex to the targeted DNA strand, which would facilitate subsequent conversion of the corresponding guanine on the non-targeted DNA.

[00831] In one embodiment, nickase Cpfl can be used to nick the strand of DNA that is not complementary to the guide RNA. For AsCpfl the mutation would be R1226A.

[00832] Designs for the recruitment of CD to a specific locus:

[00833] 1. NLS-tagged nickase Cpfl is fused to CD on either the N- or C-terminal end. A variety of linkers are used including flexible linkers such as GSG_5 or less flexible linkers such as LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR.

[00834] 2. The guide RNA scaffold is modified with aptamers such as MS2 binding sites (e.g. Konermann et al., Nature 2015). NLS-tagged CD-MS2 binding protein fusions is co-introduced into target cells along with (NLS-tagged nickase Cpfl) and corresponding guide RNA.

[00835] 3. CD is inserted into an internal loop of NLS-tagged nickase Cpfl.

[00836] Designs for the RNA guide:

[00837] 1. Normal length of RNA guide (24nt for AsCpfl) is designed to target the genomic locus of interest.

[00838] 2. RNA guide with longer than canonical length is used to form heteroduplexes outside of the protein-guide RNA-target DNA complex.

[00839] For each of these RNA guide designs, the base on the RNA that is opposite of the cytosine on the DNA strand would be specified as an A or U as opposed to G. The guide can be designed to comprise a stretch of two or more consecutive mismatches or non-pairing nucleotides, including up to about 6 consecutive mismatches or non-pairing nucleotides, preferably 3-5 consecutive mismatches or non-pairing nucleotides, and more preferably 3-4 consecutive mismatches or non-pairing nucleotides. The guide can be designed to comprise one or two mismatches or non-pairing nucleotides on each side of and adjacent to the aforementioned A or U opposite to the targeted C.

[00840] Choice and Designs of CDs:

[00841] A number of CDs are used, and each will have varying levels of activity. These CDs include:

[00842] Rat APOBEC 1 (rAPOBEC 1):

[00843] MSSETGPVAVDPTLRRRIEPHEFEVFFDPRELRKETCLLYEINWGGRHSI WRHTSQNTNKHVEVNFIEKFTTERYFCPNTRCSITWFLSWSPCGECSRAITEFLSRYP HVTLFIYIARLYHHADPRNRQGLRDLISSGVTIQIMTEQESGYCWRNFVNYSPSNEA HWPRYPHLWVRLYVLELYCIILGLPPCLNILRRKQPQLTFFTIALQSCHYQRLPPHIL WATGLK (SEQ ID NO: 29)

[00844] Human APOBEC 1 (hAPOBEC 1)

[00845] MTSEKGPSTGDPTLRRRIEPWEFDVFYDPRELRKEACLLYEIKWGMSRKI WRSSGKNTTNHVEVNFIKKFTSERDFHPSMSCSITWFLSWSPCWECSQAIREFLSRH PGVTLVIYVARLFWHMDQQNRQGLRDLVNSGVTIQFMRASEYYHCWRNFVNYPP GDEAHWPQYPPLWMMLYALELHCIILSLPPCLKISRRWQNHLTFFRLHLQNCHYQT IPPHILLATGLIHPSVAWR (SEQ ID NO: 30)

[00846] Human APOBEC3G (hAPOBEC3G):

[00847] MELKYHPEMRFFHWFSKWRKLHRDQEYEVTWYISWSPCTKCTRDMAT FLAEDPKVTLTIFVARLYYFWDPDYQEALRSLCQKRDGPRATMKFMNYDEFQHCW SKFVYSQRELFEPWNNLPKYYILLHIMLGEILRHSMDPPTFTFNFNNEPWVRGRHET YLCYEVERMHNDTWVLLNQRRGFLCNQAPHKHGFLEGRHAELCFLDVIPFWKLDL DQDYRVTCFTSWSPCFSCAQEMAKFISKNKHVSLCIFTARIYDDQGRCQEGLRTLA EAGAKISIMTYSEFKHCWDTFVDHQGCPFQPWDGLDEHSQDLSGRLRAILQNQEN (SEQ ID NO: 31)

[00848] Petromyzon marinus (Lamprey) CDA1 (pmCDAl):

[00849] MTDAEYVRIHEKLDIYTFKKQFFNNKKSVSHRCYVLFELKRRGERRACF WGYAVNKPQSGTERGIHAEIFSIRKVEEYLRDNPGQFTINWYSSWSPCADCAEKILE WYNQELRGNGHTLKIWACKLYYEKNARNQIGLWNLRDNGVGLNVMVSEHYQCC RKIFIQSSHNQLNENRWLEKTLKRAEKRRSELSIMIQVKILHTTKSPAV (SEQ ID NO: 32)

[00850] Human AID (hAJD):

[00851] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFG YLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNP

YLSLRIFTAEI.YFCEDRKAEPEGLRRLHRAGVQIAIMTFKDYFYCWNTFVENHERTF KAWEGLHENSVRLSRQLRRILLPLYEVDDLRDAFRTLGLLD (SEQ ID NO: 33)

[00852] Human AID-DC (hAID-DC, truncated version of hAID):

[00853] MDSLLMNRRKFLYQFKNVRWAKGRRETYLCYVVKRRDSATSFSLDFG YLRNKNGCHVELLFLRYISDWDLDPGRCYRVTWFTSWSPCYDCARHVADFLRGNP NLSLRIFTARLWCEDRKAEPEGLRRLHRAGVQIAFMTFKDYFYCWNTFVENHERTF KAWEGLHENS VRL SRQLRRILL (SEQ ID NO: 34)

[00854] Mutations can also be used to increase the specificity of base editing. For example, rAPOBECl mutants having one or more mutations of W90Y, R126E, and R132E can be used to increase the specificity of base editing.

[00855] In addition, a uracil glycosylase inhibitor (UGI) can be linked to the Cpfl and/or the CD to inhibit uracil-DNA glycosylase base-excision repair enzyme and increase the efficiency of base editing. For example, rAPOBECl mutants having one or more mutations of W90Y, R126E, and R132E can be used to increase the specificity of deamination.

[00856] Uracil glycosylase inhibitor (UGI):

[00857] TNLSDIIEKETGKQLVIQESILMLPEEVEEVIGNKPESDILVHTAYDESTDE NVMLLTSDAPEYKPWALVIQDSNGENKIKML (SEQ ID NO: 35)

[00858] The embodiments, illustratively described herein may suitably be practiced in the absence of any element or elements, limitation or limitations, not specifically disclosed herein. Thus, for example, the terms "comprising," "including," "containing," etc. shall be read expansively and without limitation. Additionally, the terms and expressions employed herein have been used as terms of description and not of limitation, and there is no intention in the use of such terms and expressions of excluding any equivalents of the features shown and described or portions thereof, but it is recognized that various modifications are possible within the scope of the claimed technology. Additionally, the phrase "consisting essentially of will be understood to include those elements specifically recited and those additional elements that do not materially affect the basic and novel characteristics of the claimed technology. The phrase "consisting of excludes any element not specified.

[00859] The present disclosure is not to be limited in terms of the particular embodiments described in this application. Many modifications and variations can be made without departing from its spirit and scope, as will be apparent to those skilled in the art. Functionally equivalent methods and compositions within the scope of the disclosure, in addition to those enumerated herein, will be apparent to those skilled in the art from the

foregoing descriptions. Such modifications and variations are intended to fall within the scope of the appended claims. The present disclosure is to be limited only by the terms of the appended claims, along with the full scope of equivalents to which such claims are entitled. It is to be understood that this disclosure is not limited to particular methods, reagents, compounds compositions or biological systems, which can of course vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to be limiting.

[00860] In addition, where features or aspects of the disclosure are described in terms of Markush groups, those skilled in the art will recognize that the disclosure is also thereby described in terms of any individual member or subgroup of members of the Markush group.

[00861] As will be understood by one skilled in the art, for any and all purposes, particularly in terms of providing a written description, all ranges disclosed herein also encompass any and all possible subranges and combinations of subranges thereof. Any listed range can be easily recognized as sufficiently describing and enabling the same range being broken down into at least equal halves, thirds, quarters, fifths, tenths, etc. As a non-limiting example, each range discussed herein can be readily broken down into a lower third, middle third and upper third, etc. As will also be understood by one skilled in the art all language such as "up to," "at least," "greater than," "less than," and the like, include the number recited and refer to ranges which can be subsequently broken down into subranges as discussed above. Finally, as will be understood by one skilled in the art, a range includes each individual member.

[00862] All publications, patent applications, issued patents, and other documents referred to in this specification are herein incorporated by reference as if each individual publication, patent application, issued patent, or other document was specifically and individually indicated to be incorporated by reference in its entirety. Definitions that are contained in text incorporated by reference are excluded to the extent that they contradict definitions in this disclosure.

[00863] Other embodiments are set forth in the following claims.

WHAT IS CLAIMED IS:

1. A method of modifying a Cytosine in a target locus of interest, comprising delivering to said locus:

(a) a Cpfl nickase protein;

(b) a guide molecule which comprises a guide sequence linked to a direct repeat; and

(c) a cytidine deaminase protein or catalytic domain thereof;

wherein said cytidine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to said Cpfl nickase protein or said guide molecule or is adapted to link thereto after delivery;

wherein guide molecule forms a complex with said Cpfl nickase protein and directs said complex to bind a first DNA strand at said target locus of interest, wherein said guide sequence is capable of hybridizing with a target sequence comprising said Cytosine within said first DNA strand to form a heteroduplex, wherein said guide sequence comprises a non-pairing Adenine or Uracil at a position corresponding to said Cytosine resulting in a C-A or C-U mismatch in said heteroduplex formed;

wherein said Cpfl nickase protein nicks a second DNA strand at said target locus of interest displaced by formation of said heteroduplex; and

wherein said cytidine deaminase protein or catalytic domain thereof deaminates said Cytosine in said heteroduplex.

2. The method of claim 1, wherein said cytidine deaminase protein or catalytic domain thereof is fused to N- or C-terminus of said Cpfl nickase protein.

3. The method of claim 2, wherein said cytidine deaminase protein or catalytic domain thereof is fused to said Cpfl nickase protein by a linker.

4. The method of claim 3, wherein said linker is $(GGGGS)_3$-n (SEQ ID NOS: 1-9), GSG_5 (SEQ ID NO: 10) or LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO: 11).

5. The method of claim 1, wherein said cytidine deaminase protein or catalytic domain thereof is linked to an adaptor protein, and said guide molecule or said Cpfl nickase protein comprises an aptamer sequence capable of binding to said adaptor protein.

6. The method of claim 5, wherein said adaptor sequence is selected from MS2, PP7, Qp, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KUl, Mil, MXl, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, (^Cb8r, (^Cbl2r, (^Cb23r, 7s and PRR1.

7. The method of claim 1, wherein said cytidine deaminase protein or catalytic domain thereof is inserted into an internal loop of said Cpfl nickase protein.

8. The method of any of claims 1-7, wherein said Cpfl nickase protein comprises a mutation in the Nuc domain.

9. The method of claim 8, wherein said Cpfl nickase protein comprises a mutation corresponding to R1226A in AsCpfl.

10. The method of any of claims 1-7, wherein said Cpfl nickase protein has at least part of the Nuc domain removed.

11. The method of any of claims 1-10, wherein said guide molecule binds to said Cpfl nickase protein and is capable of forming said heteroduplex of about 24 nt with said target sequence.

12. The method of any of claims 1-10, wherein said guide molecule binds to said Cpfl nickase protein and is capable of forming said heteroduplex of more than 24 nt with said target sequence.

13. The method of any of claims 1-12, wherein said guide sequence comprises at least one further non-pairing nucleotide with said target sequence, adjacent to said non-pairing Adenine or Uracil.

14. The method of claim 13, wherein said guide sequence comprises a stretch of three to five consecutive non-pairing nucleotides or six to eight consecutive non-pairing nucleotides with said target sequence.

15. The method of any of the preceding claims, wherein said cytidine deaminase protein or catalytic domain thereof is a human, rat or lamprey cytidine deaminase protein or catalytic domain thereof.

16. The method of claim 15, wherein said cytidine deaminase protein or catalytic domain thereof is an apolipoprotein B mRNA-editing complex (APOBEC) family deaminase, an activation-induced deaminase (AID), or a cytidine deaminase 1 (CDA1).

17. The method of claim 16, wherein said cytidine deaminase protein or catalytic domain thereof is an APOBEC1 deaminase comprising one or more mutations corresponding to W90A, W90Y, R 118A, H121R, H122R, R126A, R126E, or R132E in rat APOBECl, or an APOBEC3G deaminase comprising one or more mutations corresponding to W285A, W285Y, R313A, D316R, D317R, R320A, R320E, or R326E in human APOBEC3G.

18. The method of any of the preceding claims, wherein said Cpfl nickase protein and optionally said cytidine deaminase protein or catalytic domain thereof comprise one or more heterologous nuclear localization signal(s) (NLS(s)).

19. The method of any of the preceding claims, wherein said cytidine deaminase protein or catalytic domain thereof is delivered together with a uracil glycosylase inhibitor (UGI), where said UGI is covalently linked to said cytidine deaminase protein or catalytic domain thereof and/or said Cpfl nickase protein.

20. The method of any of the preceding claims, wherein said method comprises, determining said target sequence of interest and selecting a cytidine deaminase protein or catalytic domain thereof which most efficiently deaminates said Cytosine present in said target sequence.

21. The method of any of the preceding claims, wherein said Cpfl nickase protein is obtained from a Cpfl nuclease derived from a bacterial species selected from the group consisting of Francisella tularensis, Prevotella albensis, Lachnospiraceae bacterium, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium, Parcubacteria bacterium, Smithella sp., Acidaminococcus sp., Lachnospiraceae bacterium, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi, Leptospira inadai, Porphyromonas crevioricanis, Prevotella disiens and Porphyromonas macacae, Succinivibrio dextrinosolvens, Prevotella disiens, Flavobacterium branchiophilum, Helcococcus kunzii, Eubacterium sp., Microgenomates (Roizmanbacteria) bacterium, Flavobacterium sp., Prevotella brevis, Moraxella caprae, Bacteroidetes oral, Porphyromonas cansulci, Synergistes jonesii, Prevotella bryantii, Anaerovibrio sp., Butyrivibrio fibrisolvens, Candidatus Methanomethylophilus, Butyrivibrio sp., Oribacterium sp., Pseudobutyrivibrio ruminis and Proteocatella sphenisci.

22. The method of claim 21, wherein said Cpfl nickase protein is a FnCpfl nickase and recognizes a PAM sequence of TTN, wherein N is A/C/G or T, or said Cpfl

nickase protein is a PaCpflp, LbCpfl or AsCpfl nickase and recognizes a PAM sequence of TTTV, wherein V is A/C or G.

23. The method of claim 21, wherein said Cpfl nickase protein has been modified to and recognizes an altered PAM sequence.

24. The method of any of the preceding claims, wherein said target locus of interest is within a cell.

25. The method of claim 24, wherein said cell is a eukaryotic cell.

26. The method of claim 24, wherein said cell is a non-human animal cell, a human cell, or a plant cell.

27. The method of any of the preceding claims, wherein said target locus of interest is within an animal, within a plant, or comprised in a DNA molecule *in vitro.*

28. The method of any of the preceding claims, wherein said components (a), (b) and (c) are delivered to the cell as a ribonucleoprotein complex.

29. The method of any of the preceding claims, wherein said components (a), (b) and (c) are delivered to the cell as one or more polynucleotide molecules.

30. The method of claim 29, wherein said one or more polynucleotide molecules comprise one or more mRNA molecules encoding components (a) and/or (c).

31. The method of claim 29, wherein said one or more polynucleotide molecules are comprised within one or more vectors.

32. The method of claim 31, wherein said one or more polynucleotide molecules comprise one or more regulatory elements operably configured to express said Cpfl nickase protein, said guide molecule, and said cytidine deaminase protein or catalytic domain thereof, optionally wherein said one or more regulatory elements comprise inducible promoters.

33. The method of any of claims 28-32, wherein said one or more polynucleotide molecules or said ribonucleoprotein complex are delivered via particles, vesicles, or one or more viral vectors.

34. The method of claim 33, wherein said particles comprise a lipid, a sugar, a metal or a protein.

35. The method of claim 34, wherein said particles comprise lipid nanoparticles.

36. The method of claim 33, wherein said vesicles comprise exosomes or liposomes.

37. The method of claim 33, wherein said one or more viral vectors comprise one or more of adenovirus, one or more lentivirus or one or more adeno-associated virus.

38. The method of any of the preceding claims, which is a method of modifying a cell, a cell line or an organism by manipulation of one or more target sequences at genomic loci of interest.

39. The method of claim 38, wherein deamination of said Cytosine at said target locus of interest remedies a disease caused by a T→C or A→G point mutation or pathogenic SNP.

40. The method of claim 38, wherein deamination of said Cytosine at said target locus of interest inactivates a target gene at said target locus.

41. A modified cell obtained from the method of any of the preceding claims, or progeny thereof, wherein said cell comprises a Uracil or a Thymine in replace of said Cytosine in said target locus of interest compared to a corresponding cell not subjected to said method.

42. The modified cell or progeny thereof of claim 41, wherein said cell is a eukaryotic cell.

43. The modified cell or progeny thereof of claim 41, wherein said cell is a non-human animal cell, a human cell, or a plant cell.

44. The modified cell or progeny thereof of claim 41, wherein said cell is a therapeutic T cell or an antibody-producing B cell.

45. A non-human animal or a plant comprising said modified non-human animal cell of claim 43.

46. A method for cell therapy, comprising administering to a patient in need thereof said modified cell of any of claims 41-44, wherein presence of said modified cell remedies a disease in said patient.

47. An engineered, non-naturally occurring system suitable for modifying a Cytosine in a target locus of interest, comprising

a) a guide molecule which comprises a guide sequence, or a nucleotide sequence encoding said guide molecule;

b) a Cpfl nickase protein, or a nucleotide sequence encoding said Cpfl nickase protein ;

c) a cytidine deaminase protein or catalytic domain thereof, or a nucleotide sequence encoding said cytidine deaminase protein or catalytic domain thereof;

wherein said cytidine deaminase protein or catalytic domain thereof is covalently or non-covalently linked to said Cpfl nickase protein or said guide molecule or is adapted to link thereto after delivery;

wherein said guide sequence is capable of hybridizing with a target sequence comprising a Cytosine on a first DNA strand at said target locus to form a heteroduplex, wherein said guide sequence comprises a non-pairing Adenine or Uracil at a position corresponding to said Cytosine resulting in a C-A or C-U mismatch in said heteroduplex formed; and

wherein said Cpfl nickase protein is capable of nicking a second DNA strand at said target locus.

48. An engineered, non-naturally occurring vector system suitable for modifying an Cytosine in a target locus of interest, comprising said nucleotide sequences of a), b) and c) of claim 47.

49. The engineered, non-naturally occurring vector system of claim 48, comprising one or more vectors comprising

a) a first regulatory element operably linked to a nucleotide sequence encoding said guide molecule which comprises a guide sequence,

b) a second regulatory element operably linked to a nucleotide sequence encoding said Cpfl nickase protein; and

c) a nucleotide sequence encoding said cytidine deaminase protein 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 cytidine deaminase protein or catalytic domain thereof is operably linked to a third regulatory element, said cytidine protein or catalytic domain thereof is adapted to link to said guide molecule or said Cpfl nickase protein after expression;

wherein components (a), (b) and (c) are located on the same or different vectors of said system.

50. An *in vitro* or *ex vivo* host cell or progeny thereof or cell line or progeny thereof comprising a system of any of claims 47-49, wherein optionally said cell is a eukaryotic cell.

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/033427

A. CLASSIFICATION OF SUBJECT MATTER
INV. C12N15/10 C12N15/62
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 , 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	WO 2015/089406 Al (HARVARD COLLEGE [US]) 18 June 2015 (2015-06-18)	41-43 ,45
Y	abstract paragraphs [0009] , [0028] , [0040] , [0043] , [0114] claims 4,7, 17 , 20 claims 27-30,39,43,47 ,51	1-40,44, 46-50
Y	MALZAHN AIMEE ET AL: "Plant genome editing with TALEN and CRISPR" , CELL & BIOSCIENCE, BI0MED CENTRAL LTD, LONDON, UK, vol . 7, 24 April 2017 (2017-04-24) , pages 1-18, XP002785201 , ISSN : 2045-3701, D0I : 10. 1186/S13578-017-0148-4 abstract page 12	1-40,44, 46-50

-/ -

[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) orwhich 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
24 October 2018	06/11/2018

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	Barz, Wolfgang

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2018/033427

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	BERND ZETSCHE ET AL: "Cpfl Is a Singl e RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System" , CELL, vol . 163 , no. 3 , 1 October 2015 (2015-10-01) , pages 759-771, XP055267511 , AMSTERDAM, NL ISSN : 0092-8674, D0I : 10. 1016/j .eel 1.2015. 09. 038 the whole document	1-50
A	Y BI LL KIM ET AL: "Increasing the genome-targeti ng scope and preci si on of base edi t i ng wi t h engi neered Cas9-cyti dine deami nase fusi ons" , NATURE BIOTECHNOLOGY, vol . 35, no. 4 , 13 February 2017 (2017-02-13) , pages 371-376, XP055484491 , ISSN : 1087-0156, D0I : 10. 1038/nbt.3803 abstract	1-50
A	ALEXIS C. K0M0R ET AL: "Programmable edi t i ng of a target base i n genomi c DNA wi thout doubl e-stranded DNA cl eavage" , NATURE, vol . 533 , no. 7603, 20 Apri l 2016 (2016-04-20) , pages 420-424, XP055483559, London ISSN : 0028-0836, D0I : 10. 1038/naturel7946 abstract	1-50

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2018/033427

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2015089406 Al	18-06-2015	AU 2014362208 Al	30-06-2016
		CA 2933625 Al	18-06-2015
		CN 105934516 A	07-09-2016
		EP 3080265 Al	19-10-2016
		JP 2017500035 A	05-01-2017
		US 2015165054 Al	18-06-2015
		US 2015166980 Al	18-06-2015
		US 2015166981 Al	18-06-2015
		US 2015166982 Al	18-06-2015
		US 2015166983 Al	18-06-2015
		US 2015166984 Al	18-06-2015
		US 2015166985 Al	18-06-2015
		US 2016304846 Al	20-10-2016
		WO 2015089406 Al	18-06-2015