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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 August 2024 (15.08.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/168253 A1

(51) International Patent Classification:
C12N 15/86 (2006.01) *C12N 15/00* (2006.01)
A61K 9/127 (2006.01) *C12N 15/10* (2006.01)
A61K 9/51 (2006.01) *C12N 15/11* (2006.01)
C12N 9/22 (2006.01)

(21) International Application Number:
PCT/US2024/015176

(22) International Filing Date:
09 February 2024 (09.02.2024)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/484,404 10 February 2023 (10.02.2023) US

(71) Applicants: **POSSIBLE MEDICINES LLC** [US/US]; 2850 Quarry Lake Drive, Baltimore, MD 21208 (US). **THE BOARD OF TRUSTEES OF THE LELAND STANFORD JUNIOR UNIVERSITY** [US/US]; Building 170, Third Floor, Main Quad, P.O. Box 20386, Stanford, CA 94305 (US).

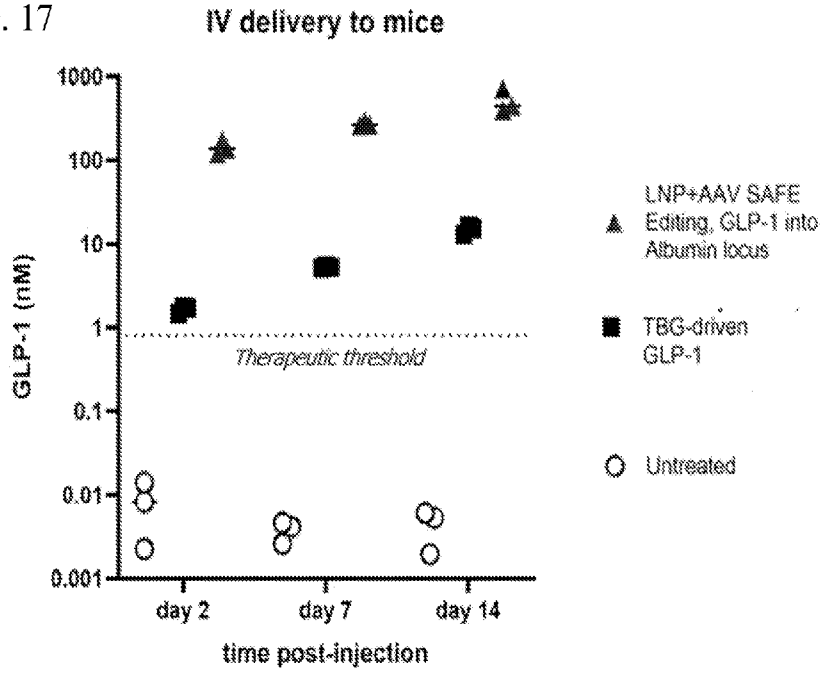
(72) Inventors: **HINDERER, Christian**; c/o Possible Medicines LLC, 2850 Quarry Lake Drive, Baltimore, MD 21208 (US). **HE, Zhenhua**; c/o Possible Medicines LLC, 2850 Quarry Lake Drive, Baltimore, MD 21208 (US). **KARNAL, Alex**; c/o Possible Medicines LLC, 2850 Quarry Lake Drive, Baltimore, MD 21208 (US). **CONG, Le**; c/o The Board Of Trustees Of The Leland Stanford Junior University, Building 170, Third Floor, Main Quad, P.O. Box 20386, Stanford, CA 94305 (US). **WANG, Chengkun**; c/o The Board Of Trustees Of The Leland Stanford Junior University, Building 170, Third Floor, Main Quad, P.O. Box 20386, Stanford, CA 94305 (US).

(74) Agent: **KOWALSKI, Thomas, J.** et al.; Duane Morris LLP, 1540 Broadway, New York, NY 10036 (US).

(81) Designated States *(unless otherwise indicated, for every kind of national protection available)*: AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CV, CZ, DE, DJ, DK, DM,

(54) Title: DELIVERY OF AN RNA GUIDED RECOMBINATION SYSTEM

FIG. 17



(57) Abstract: The present invention relates to genetic modification systems utilizing RNA-guided recombination-editing systems using phage recombination enzymes. Manipulation of a target sequence involves delivering an RNA guided-recombination editing system with a viral vector system and a lipid-based delivery system having a nucleic acid molecule with a guide RNA sequence that is complementary to a target DNA sequence and a recombination protein with a nuclease, a single stranded DNA annealing protein (SSAP) or a combination thereof for expression in a cell.

[Continued on next page]

WO 2024/168253 A1

DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

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

Published:

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

— *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))*

DELIVERY OF AN RNA GUIDED RECOMBINATION SYSTEM

RELATED APPLICATIONS AND INCORPORATION BY REFERENCE

[0001] This application claims priority to US provisional application Serial No. 63/484,404, filed February 10, 2023, which is incorporated by reference herein in its entirety.

[0002] The foregoing applications, and all documents cited therein or during their prosecution ("appln cited documents") and all documents cited or referenced in the appln cited documents, and all documents cited or referenced herein ("herein cited documents"), and all documents cited or referenced in herein cited documents, together with any manufacturer's instructions, descriptions, product specifications, and product sheets for any products mentioned herein or in any document incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention. More specifically, all referenced documents are incorporated by reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

FIELD OF THE INVENTION

[0003] The present invention relates to genetic modification systems utilizing RNA-guided recombination-editing systems using phage recombination enzymes as well as methods, vectors, nucleic acid compositions, and kids thereof.

BACKGROUND OF THE INVENTION

[0004] The Clustered Regularly Interspersed Short Palindromic Repeats (CRISPR) system, originally found in bacteria and archaea as part of the immune system to defend against invading viruses, forms the basis for genome editing technologies that can be programmed to target specific stretches of a genome or other DNA for editing precise locations. While various CRISPR-based tools are available, the majority are geared towards editing short sequences. Long- sequence editing is highly sought after in the engineering of model systems, therapeutic cell production and gene therapy. Prior studies have developed technologies to improve Cas9-mediated homology directed repair (HDR) (K. S. Pawelczak, et al., *ACS Chem. Biol.* 13, 389-396 (2018)), and tools leveraging nucleic acid modification enzymes with Cas9, e.g. prime-editing (E. V. Anzalone, et al., *Nature*. 576, 149-157 (2019)) that demonstrated editing up to 80-base-pairs (bp) in length. Despite these progresses, there are continues demands for large-scale mammalian genome engineering with high efficiency and fidelity.

[0005] Citation or identification of any document in this application is not an admission that such document is available as prior art to the present invention.

SUMMARY OF THE INVENTION

[0006] The invention relates to methods and compositions for modifying an organism or a non-human organism by manipulation of a target sequence, which comprises delivering in a cell, an RNA guided-recombination system or composition comprising: (i) a nucleic acid molecule comprising a guide RNA sequence that is complementary to a target DNA sequence; wherein the target DNA sequence comprises a genomic DNA sequence in the cell, and (ii) a recombination protein, wherein the recombination protein comprises a nuclease, a single stranded DNA annealing protein (SSAP) or a combination thereof; or, (iii) nucleic acid molecule(s) encoding or delivering the nucleic acid molecule, and/or the recombination protein for expression *in vivo* in a cell; or, (iv) vector(s) containing the nucleic acid molecule(s) for expression *in vivo* in a cell; wherein the delivering comprises a delivery system comprising a viral vector system and/or a lipid-based delivery system.

[0007] In certain embodiments, the method also includes a Cas protein. In certain embodiments, the method includes a donor nucleic acid. In certain embodiments, the target DNA sequence comprises a genomic sequence of albumin (ALB), AAVS1, HSP90AA1, DYNLT1, ACTB, BCAP31, HIST1H2BK, CLTA, ROSA26, PCSK, or RAB11A.

[0008] In certain embodiments, the method also includes a recruitment system comprising at least one aptamer sequence; and an aptamer binding protein functionally linked to the recombination protein as part of a fusion protein. In certain embodiments, the at least one aptamer sequence is an RNA aptamer sequence or a peptide aptamer sequence. In certain embodiments, the nucleic acid molecule or nucleic acid molecules additionally comprise the at least one RNA aptamer sequence. In certain embodiments, the at least one RNA aptamer sequences comprise the same sequence. In certain embodiments, the aptamer binding protein comprises an MS2 coat protein, or a functional derivative or variant thereof. In certain embodiments, the at least one peptide aptamer sequence is conjugated to the Cas protein. In certain embodiments, the at least one peptide aptamer sequence comprises between 1 and 24 peptide aptamer sequences. In certain embodiments, the aptamer sequences comprise the same sequence.

[0009] In certain embodiments, the recombination protein N-terminus is linked to the aptamer binding protein C-terminus. In certain embodiments, the recombination protein comprises a microbial recombination protein or active portion thereof.

[0010] In certain embodiments, the Cas protein is catalytically inactive (less than 5% nuclease activity as compared with a wild type or non-mutated Cas protein) or catalytically dead. In certain embodiments, the Cas protein comprises a Cas9, Cas12a, or Cas12f.

[0011] In certain embodiments, the single stranded DNA annealing protein and the Cas protein are functionally linked to each other and comprise a fusion protein.

[0012] In certain embodiments, the RNA guided recombination system is delivered via a viral vector system and a lipid-based system. In certain advantageous embodiments, the viral vector system is an AAV viral vector system. In certain particularly advantageous embodiments, the viral vector system is an AAV8 viral vector system. In ceratin advantageous embodiments, the lipid-based system is a lipid nanoparticle (LNP).

[0013] In an aspect, the invention provides a system or composition for delivering into an organism or a non-human organism: (i) a nucleic acid molecule comprising a guide RNA sequence that is complementary to a target DNA sequence; wherein the target DNA sequence comprises a genomic DNA sequence in the cell, and (ii) a recombination protein, wherein the recombination protein comprises a nuclease, a single stranded DNA annealing protein (SSAP) or a combination thereof; or, (iii) nucleic acid molecule(s) encoding or delivering (i), and/or (ii) for expression *in vivo* in a cell; or, (iv) vector(s) containing the nucleic acid molecule(s) of (iii) for expression *in vivo* in a cell; wherein the system or composition comprises a delivery system comprising a viral vector system and/or a lipid-based delivery system.

[0014] In certain embodiments, the system or composition comprises a donor nucleic acid. In certain embodiments, the system or composition comprises a donor nucleic acid, wherein the donor nucleic acid comprises a sequence that encodes GLP-1 or a GLP-1 analog.

[0015] In certain embodiments, the system or composition targets a DNA sequence which comprises a genomic sequence of albumin (ALB), AAVS1, HSP90AA1, DYNLT1, ACTB, BCAP31, HIST1H2BK, CLTA, ROSA26, PCSK, or RAB11A.

[0016] In certain embodiments, the system or composition comprises a recruitment system comprising at least one aptamer sequence; and an aptamer binding protein functionally linked to the recombination protein as part of a fusion protein. In certain embodiments the at least one

aptamer sequence is an RNA aptamer sequence or a peptide aptamer sequence. In certain embodiments, the nucleic acid molecule or nucleic acid molecules comprise the at least one RNA aptamer sequence. In certain embodiments, the at least one RNA aptamer sequences comprise the same sequence.

[0017] In certain embodiments the aptamer binding protein comprises an MS2 coat protein, or a functional derivative or variant thereof. In certain embodiments, the at least one peptide aptamer sequence is conjugated to the Cas protein. In certain embodiments the at least one peptide aptamer sequence comprises between 1 and 24 peptide aptamer sequences. In certain embodiments the aptamer sequences comprise the same sequence.

[0018] In certain embodiments the recombination protein N-terminus is linked to the aptamer binding protein C-terminus. In certain embodiments recombination protein comprises a microbial recombination protein or active portion thereof.

[0019] In certain embodiments the Cas protein comprises a Cas9, Cas12a, or Cas12f. In certain embodiments the Cas protein is catalytically inactive (less than 5% nuclease activity as compared with a wild type or non-mutated Cas protein) or catalytically dead.

[0020] In certain embodiments the single stranded DNA annealing protein and the Cas protein are functionally linked to each other and comprise a fusion protein.

[0021] In certain embodiments, the RNA guided recombination system is delivered via a viral vector system and a lipid-based system.

[0022] In certain embodiments, the viral vector system comprises an AAV viral vector system. In certain embodiments, viral vector system comprises an AAV8 viral vector system. In certain embodiments the lipid nanoparticle (LNP).

[0023] In an aspect, the invention provides a kit comprising components of the delivery system or composition.

[0024] Accordingly, it is an object of the invention not to encompass within the invention any previously known product, process of making the product, or method of using the product such that Applicants reserve the right and hereby disclose a disclaimer of any previously known product, process, or method. It is further noted that the invention does not intend to encompass within the scope of the invention any product, process, or making of the product or method of using the product, which does not meet the written description and enablement requirements of the USPTO (35 U.S.C. §112, first paragraph) or the EPO (Article 83 of the EPC), such that Applicants reserve

the right and hereby disclose a disclaimer of any previously described product, process of making the product, or method of using the product. It may be advantageous in the practice of the invention to be in compliance with Art. 53(c) EPC and Rule 28(b) and (c) EPC. All rights to explicitly disclaim any embodiments that are the subject of any granted patent(s) of applicant in the lineage of this application or in any other lineage or in any prior filed application of any third party is explicitly reserved. Nothing herein is to be construed as a promise.

[0025] It is noted that in this disclosure and particularly in the claims and/or paragraphs, terms such as "comprises", "comprised", "comprising" and the like can have the meaning attributed to it in U.S. Patent law; e.g., they can mean "includes", "included", "including", and the like; and that terms such as "consisting essentially of" and "consists essentially of" have the meaning ascribed to them in U.S. Patent law, e.g., they allow for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention.

[0026] These and other embodiments are disclosed or are obvious from and encompassed by, the following Detailed Description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0027] The patent or application file contains at least one drawing executed in color. Copies of this patent or patent application publication with color drawing(s) will be provided by the Office upon request and payment of the necessary fee.

[0028] The following detailed description, given by way of example, but not intended to limit the invention solely to the specific embodiments described, may best be understood in conjunction with the accompanying drawings.

[0029] FIG 1. Experimental design for Milestone 1. Milestone 1 compared the blood serum levels of untreated mice and mice treated with AAV delivery alone, LNP delivery alone, and AAV + LNP delivery.

[0030] FIG. 2. Experimental design for Milestone 2. Milestone 2 compared the blood serum levels of mice treated with LNP delivery alone and AAV + LNP delivery.

[0031] FIG. 3. Test articles for Milestone 1.

[0032] FIG. 4. Test articles for Milestone 2.

[0033] FIG. 5. Reagents used for Milestones 1 and 2.

[0034] FIG. 6. Instruments used for Milestones 1 and 2.

[0035] FIG. 7. Drug administration information for Milestone 1.

[0036] FIG. 8. Drug administration information for Milestone 2.

[0037] FIG. 9. Blood GLP-1 measurement dilution factors.

[0038] FIG. 10. LNP preparation for Milestone 1.

[0039] FIG. 11. LNP preparation for Milestone 2.

[0040] FIG. 12. Final state of mice in Milestone 1.

[0041] FIG. 13. Final state of mice in Milestone 2.

[0042] FIG. 14. Serum concentrations of GLP-1 over time.

[0043] FIG. 15. Detailed description of serum concentrations of GLP-1.

[0044] FIG. 16. Schematic of serum concentrations of GLP-1 over time.

[0045] FIG. 17. Robust *in vivo* dulaglutide expression from the albumin locus using SAFE editing. Safe editing, which comprises AAV and LNP delivered dCas9-SSAP + sgRNA + hGLP1 donor template, achieved over 500nM serum dulaglutide 14 days after treatment. Expression exceeded AAV gene transfer with TBG promoter and resulted in a 650-fold therapeutic serum concentration of dulaglutide.

[0046] FIG. 18. Study design schematic. AAV viral delivery systems carrying GLP-1 donor template is injected into mice with LNP delivery systems carrying dCas9 +SSAP mRNA and Alb MS2 gRNA and injected into mice, which were studied over a 14 day period (see FIG. 17).

DETAILED DESCRIPTION OF THE INVENTION

[0047] The present invention is directed to a system and the components for DNA editing. In particular, the disclosed system is based on CRISPR targeting and homology directed repair by phage recombination enzymes. The disclosed system is an RNA guided recombination system. The system results in superior recombination efficiency and accuracy at a kilobase scale. The present invention is also directed to nucleic acids encoding an RNA guided recombination system and vectors encoding an RNA guided recombination system.

[0048] With respect to general information on CRISPR-Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, AAV, and making and using thereof, including as to amounts and formulations, all useful in the practice of the instant invention, reference is made to: U.S. Pat. Nos. 8,697,359, 8,771,945, 8,795,965, 8,865,406, 8,871,445, 8,889,356, 8,889,418 and 8,895,308; US Patent Publications US 2014-0310830 (U.S. application Ser. No. 14/105,031), US 2014-0287938 A1 (U.S. application Ser. No. 14/213,991), US 2014-0273234 A1 (U.S. application Ser. No. 14/293,674), US2014-

0273232 A1 (U.S. application Ser. No. 14/290,575), US 2014-0273231 (U.S. application Ser. No. 14/259,420), US 2014-0256046 A1 (U.S. application Ser. No. 14/226,274), US 2014-0248702 A1 (U.S. application Ser. No. 14/258,458), US 2014-0242700 A1 (U.S. application Ser. No. 14/222,930), US 2014-0242699 A1 (U.S. application Ser. No. 14/183,512), US 2014-0242664 A1 (U.S. application Ser. No. 14/104,990), US 2014-0234972 A1 (U.S. application Ser. No. 14/183,471), US 2014-0227787 A1 (U.S. application Ser. No. 14/256,912), US 2014-0189896 A1 (U.S. application Ser. No. 14/105,035), US 2014-0186958 (U.S. application Ser. No. 14/105,017), US 2014-0186919 A1 (U.S. application Ser. No. 14/104,977), US 2014-0186843 A1 (U.S. application Ser. No. 14/104,900), US 2014-0179770 A1 (U.S. application Ser. No. 14/104,837) and US 2014-0179006 A1 (U.S. application Ser. No. 14/183,486), US 2014-0170753 (U.S. application Ser. No. 14/183,429); 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 WO 2014/093661 (PCT/US2013/074743), WO 2014/093694 (PCT/US2013/074790), WO 2014/093595 (PCT/US2013/074611), WO 2014/093718 (PCT/US2013/074825), WO 2014/093709 (PCT/US2013/074812), WO 2014/093622 (PCT/US2013/074667), WO 2014/093635 (PCT/US2013/074691), WO 2014/093655 (PCT/US2013/074736), WO 2014/093712 (PCT/US2013/074819), WO2014/093701 (PCT/US2013/074800), and WO2014/018423 (PCT/US2013/051418). Reference is also made to U.S. provisional patent applications 61/758,468; 61/802,174; 61/806,375; 61/814,263; 61/819,803 and 61/828,130, filed on Jan. 30, 2013; Mar. 15, 2013; Mar. 28, 2013; Apr. 20, 2013; May 6, 2013 and May 28, 2013 respectively. Reference is also made to U.S. provisional patent application 61/836,123, filed on Jun. 17, 2013. Reference is additionally made to U.S. provisional patent applications 61/835,931, 61/835,936, 61/836,127, 61/836,101, 61/836,080 and 61/835,973, each filed Jun. 17, 2013. Further reference is made to U.S. provisional patent applications 61/862,468 and 61/862,355 filed on Aug. 5, 2013; 61/871,301 filed on Aug. 28, 2013; 61/960,777 filed on Sep. 25, 2013 and 61/961,980 filed on Oct. 28, 2013. Reference is yet further made to: PCT Patent applications Nos: PCT/US2014/041803, PCT/US2014/041800, PCT/US2014/041809, PCT/US2014/041804 and PCT/US2014/041806, each filed Jun. 10, 2014 6/10/14; PCT/US2014/041808 filed Jun. 11, 2014; and PCT/US2014/62558 filed Oct. 28, 2014, and U.S. Provisional Patent Applications Ser. Nos. 61/915,150, 61/915,301, 61/915,267 and 61/915,260, each filed Dec. 12, 2013; 61/757,972 and 61/768,959, filed on Jan. 29, 2013 and Feb. 25, 2013; 61/835,936, 61/836,127, 61/836,101,

61/836,080, 61/835,973, and 61/835,931, filed Jun. 17, 2013; 62/010,888 and 62/010,879, both filed Jun. 11, 2014; 62/010,329 and 62/010,441, each filed Jun. 10, 2014; 61/939,228 and 61/939,242, each filed Feb. 12, 2014; 61/980,012, filed Apr. 15, 2014; 62/038,358, filed Aug. 17, 2014; 62/054,490, 62/055,484, 62/055,460 and 62/055,487, each filed Sep. 25, 2014; and 62/069,243, filed Oct. 27, 2014. Reference is also made to U.S. provisional patent applications Nos. 62/055,484, 62/055,460, and 62/055,487, filed Sep. 25, 2014; U.S. provisional patent application 61/980,012, filed Apr. 15, 2014; and U.S. provisional patent application 61/939,242 filed Feb. 12, 2014. Reference is made to PCT application designating, inter alia, the United States, application No. PCT/US14/41806, filed Jun. 10, 2014. Reference is made to U.S. provisional patent application 61/930,214 filed on Jan. 22, 2014. Reference is made to U.S. provisional patent applications 61/915,251; 61/915,260 and 61/915,267, each filed on Dec. 12, 2013. Reference is made to U.S. provisional patent application U.S. Ser. No. 61/980,012 filed Apr. 15, 2014. Reference is made to PCT application designating, inter alia, the United States, application No. PCT/US14/41806, filed Jun. 10, 2014. Reference is made to U.S. provisional patent application 61/930,214 filed on Jan. 22, 2014. Reference is made to U.S. provisional patent applications 61/915,251; 61/915,260 and 61/915,267, each filed on Dec. 12, 2013. 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.

[0049] Also with respect to general information on CRISPR-Cas Systems, mention is made of the following (also hereby incorporated herein by reference):

[0050] *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* February 15; 339(6121):819-23 (2013);

[0051] *RNA-guided editing of bacterial genomes using CRISPR-Cas systems.* Jiang W., Bikard D., Cox D., Zhang F, Marraffini L A. Nat Biotechnol March; 31(3):233-9 (2013);

[0052] *One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Genome Engineering*. Wang H., Yang H., Shivalila C S., Dawlaty M M., Cheng A W., Zhang F., Jaenisch R. Cell May 9; 153(4):910-8 (2013);

[0053] *Optical control of mammalian endogenous transcription and epigenetic states*. Konermann S, Brigham M D, Trevino A E, Hsu P D, Heidenreich M, Cong L, Platt R J, Scott D A, Church G M, Zhang F. Nature. 2013 Aug. 22; 500(7463):472-6. doi: 10.1038 Nature 12466. Epub 2013 Aug. 23;

[0054] *Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity*. Ran, F A., Hsu, P D., Lin, C Y., Gootenberg, J S., Konermann, S., Trevino, A E., Scott, D A., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. Cell August 28. pii: S0092-8674(13)01015-5. (2013);

[0055] *DNA targeting specificity of RNA-guided Cas9 nucleases*. Hsu, P., Scott, D., Weinstein, J., Ran, F A., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, T J., Marraffini, L A., Bao, G., & Zhang, F. *Nat Biotechnol* doi:10.1038/nbt.2647 (2013);

[0056] *Genome engineering using the CRISPR-Cas9 system*. Ran, F A., Hsu, P D., Wright, J., Agarwala, V., Scott, D A., Zhang, F. Nature Protocols November; 8(11):2281-308. (2013);

[0057] *Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells*. Shalem, O., Sanjana, N E., Hartenian, E., Shi, X., Scott, D A., Mikkelson, T., Heckl, D., Ebert, B L., Root, D E., Doench, J G., Zhang, F. *Science* December 12. (2013). [Epub ahead of print];

[0058] *Crystal structure of cas9 in complex with guide RNA and target DNA*. Nishimasu, H., Ran, F A., Hsu, P D., Konermann, S., Shehata, S I., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. *Cell* February 27. (2014). 156(5):935-49;

[0059] *Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells*. Wu X., Scott D A., Kriz A J., Chiu A C., Hsu P D., Dadon D B., Cheng A W., Trevino A E., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp P A. *Nat Biotechnol*. (2014) April 20. doi: 10.1038/nbt.2889,

[0060] *CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling*, Platt et al., Cell 159(2): 440-455 (2014) DOI: 10.1016/j.cell.2014.09.014,

[0061] Development and Applications of CRISPR-Cas9 for Genome Engineering, Hsu et al, Cell 157, 1262-1278 (Jun. 5, 2014) (Hsu 2014),

[0062] *Genetic screens in human cells using the CRISPR/Cas9 system*, Wang et al., Science. 2014 Jan. 3; 343(6166): 80-84. doi:10.1126/science.1246981,

[0063] *Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation*, Doench et al., Nature Biotechnology published online 3 Sep. 2014; doi:10.1038/nbt.3026, and

[0064] In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9, Swiech et al, Nature Biotechnology; published online 19 Oct. 2014; doi:10.1038/nbt.3055.

[0065] Each of which is incorporated herein by reference, and discussed briefly below:

[0066] Cong et al. engineered type II CRISPR/Cas systems for use in eukaryotic cells based on both *Streptococcus thermophilus* Cas9 and also *Streptoccocus 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.

[0067] 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 donors. The study showed that simultaneous use of two crRNAs enabled multiplex mutagenesis. Furthermore, when the approach was used in combination with recombination, 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.

[0068] Konermann et al. 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

[0069] Cas9 nuclease from the microbial CRISPR-Cas system is targeted to specific genomic loci by a 20 nt guide sequence, which can tolerate certain mismatches to the DNA target and thereby promote undesired off-target mutagenesis. To address this, Ran et al. described an approach that combined a Cas9 nickase mutant with paired guide RNAs to introduce targeted double-strand breaks. 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.

[0070] Hsu et al. 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 reported that SpCas9 tolerates mismatches between guide RNA and target DNA at different positions in a sequence-dependent manner, sensitive to the number, position and distribution of mismatches. The authors further showed that SpCas9-mediated cleavage is unaffected by DNA methylation and that the dosage of SpCas9 and sgRNA can be titrated to minimize off-target modification. Additionally, to facilitate mammalian genome engineering applications, the authors reported providing a web-based software tool to guide the selection and validation of target sequences as well as off-target analyses.

[0071] Ran et al. 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.

[0072] Shalem et al. described a new way to interrogate gene function on a genome-wide scale. Their studies showed that delivery of a genome-scale CRISPR-Cas9 knockout (GeCKO) library targeted 18,080 genes with 64,751 unique guide sequences enabled both negative and positive selection screening in human cells. First, the authors showed use of the GeCKO library to identify genes essential for cell viability in cancer and pluripotent stem cells. Next, in a melanoma model, the authors screened for genes whose loss is involved in resistance to vemurafenib, a therapeutic that inhibits mutant protein kinase BRAF. Their studies showed that the highest-ranking candidates included previously validated genes NF1 and MED12 as well as novel hits NF2, CUL3, TADA2B, and TADA1. The authors observed a high level of consistency between independent guide RNAs targeting the same gene and a high rate of hit confirmation, and thus demonstrated the promise of genome-scale screening with Cas9.

[0073] Nishimasu et al. reported the crystal structure of *Streptococcus pyogenes* Cas9 in complex with sgRNA and its target DNA at 2.5 Å resolution. The structure revealed a bilobed architecture composed of target recognition and nuclease lobes, accommodating the sgRNA:DNA heteroduplex in a positively charged groove at their interface. Whereas the recognition lobe is essential for binding sgRNA and DNA, the nuclease lobe contains the HNH and RuvC nuclease domains, which are properly positioned for cleavage of the complementary and non-complementary strands of the target DNA, respectively. The nuclease lobe also contains a carboxyl-terminal domain responsible for the interaction with the protospacer adjacent motif (PAM). This high-resolution structure and accompanying functional analyses have revealed the molecular mechanism of RNA-guided DNA targeting by Cas9, thus paving the way for the rational design of new, versatile genome-editing technologies.

[0074] 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.

[0075] Hsu 2014 is a review article that discusses generally CRISPR-Cas9 history from yogurt to genome editing, including genetic screening of cells, that is in the information, data and findings of the applications in the lineage of this specification filed prior to Jun. 5, 2014. The general teachings of Hsu 2014 do not involve the specific models, animals of the instant specification.

[0076] The terms "comprise(s)," "include(s)," "having," "has," "can," "contain(s)," and variants thereof, as used herein, are intended to be open-ended transitional phrases, terms, or words that do not preclude the possibility of additional acts or structures. The singular forms "a," "and" and "the" include plural references unless the context clearly dictates otherwise. The present invention also contemplates other embodiments "comprising," "consisting of' and "consisting essentially of," the embodiments or elements presented herein, whether explicitly set forth or not.

[0077] For the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 6-9, the numbers 7 and 8 are contemplated in addition to 6 and 9, and for the range 6.0-7.0, the number 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated.

[0078] A cell has been "genetically modified," "transformed," or "transfected" by exogenous DNA, e.g., a recombinant expression vector, when such DNA has been introduced inside the cell. The presence of the exogenous DNA results in permanent or transient genetic change. The transforming DNA may or may not be integrated (covalently linked) into the genome of the cell. In prokaryotes, yeast, and mammalian cells for example, the transforming DNA may be maintained on an episomal element such as a plasmid. With respect to eukaryotic cells, a stably transformed cell is one in which the transforming DNA has become integrated into a chromosome so that it is inherited by daughter cells through chromosome replication. This stability is demonstrated by the ability of the eukaryotic cell to establish cell lines or clones that comprise a population of daughter cells containing the transforming DNA A "clone" is a population of cells derived from a single cell or common ancestor by mitosis. A "cell line" is a clone of a primary cell that is capable of stable growth in vitro for many generations.

[0079] A "vector" or "expression vector" is a replicon, such as plasmid, phage, virus, or cosmid, to which another DNA segment, e.g., an "insert," may be attached or incorporated so as to bring about the replication of the attached segment in a cell.

[0080] In general, the CRISPR-Cas or CRISPR system refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas9, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell, and may include nucleic acids in or from mitochondrial, organelles, vesicles, liposomes or particles present within the cell. In some embodiments, especially for non-nuclear uses, NLSs are not preferred. In some embodiments, direct repeats may be identified in silico by searching for repetitive motifs that fulfill any or all of the following criteria: 1. found in a 2 Kb window of genomic sequence flanking the type II CRISPR locus; 2. span from 20 to 50 bp; and 3. interspaced by 20 to 50 bp. In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[0081] In embodiments of the invention the terms guide sequence and guide RNA are used interchangeably. In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding

target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g. the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, Calif.), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence is about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length. In some embodiments, a guide sequence is less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length. In some embodiments the guide sequence is 10-30 nucleotides long. The ability of a guide sequence to direct sequence-specific binding of a CRISPR complex to a target sequence may be assessed by any suitable assay. For example, the components of a CRISPR system sufficient to form a CRISPR complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the CRISPR sequence, followed by an assessment of preferential cleavage within the target sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target polynucleotide sequence may be evaluated in a test tube by providing the target sequence, components of a CRISPR complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[0082] In a classic CRISPR-Cas system, the degree of complementarity between a guide sequence and its corresponding target sequence can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or 100%; a guide or RNA or sgRNA can be about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length; or guide or RNA or sgRNA can be less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length; and advantageously tracr RNA is 30 or 50 nucleotides in length. However, an aspect of the invention is to reduce off-target interactions, e.g., reduce the guide interacting with a target sequence having low complementarity. Indeed, in the examples, it is shown that the invention involves mutations that result in the

CRISPR-Cas system being able to distinguish between target and off-target sequences that have greater than 80% to about 95% complementarity, e.g., 83%-84% or 88-89% or 94-95% complementarity (for instance, distinguishing between a target having 18 nucleotides from an off-target of 18 nucleotides having 1, 2 or 3 mismatches). Accordingly, in the context of the present invention the degree of complementarity between a guide sequence and its corresponding target sequence is greater than 94.5% or 95% or 95.5% or 96% or 96.5% or 97% or 97.5% or 98% or 98.5% or 99% or 99.5% or 99.9%, or 100%. Off target is less than 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% or 94% or 93% or 92% or 91% or 90% or 89% or 88% or 87% or 86% or 85% or 84% or 83% or 82% or 81% or 80% complementarity between the sequence and the guide, with it advantageous that off target is 100% or 99.9% or 99.5% or 99% or 99% or 98.5% or 98% or 97.5% or 97% or 96.5% or 96% or 95.5% or 95% or 94.5% complementarity between the sequence and the guide.

[0083] In some embodiments according to the invention, the escorted guide RNA (capable of guiding Cas to a target locus) may comprise (1) a guide sequence capable of hybridizing to a genomic target locus in the eukaryotic cell; (2) a tracr sequence; and (3) a tracr mate sequence. All (1) to (3) may reside in a single RNA, i.e. an sgRNA (arranged in a 5′ to 3′ orientation), or the tracr RNA may be a different RNA than the RNA containing the guide and tracr sequence. The tracr hybridizes to the tracr mate sequence and directs the CRISPR/Cas complex to the target sequence.

[0084] The methods according to the invention as described herein comprehend inducing one or more mutations in a eukaryotic cell (in vitro, i.e. in an isolated eukaryotic cell) as herein discussed comprising delivering to cell a vector as herein discussed. The mutation(s) can include the introduction, deletion, or substitution of one or more nucleotides at each target sequence of cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1-75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations include the introduction, deletion, or substitution of 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24,

25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 40, 45, 50, 75, 100, 200, 300, 400 or 500 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s).

[0085] For minimization of toxicity and off-target effect, it will be important to control the concentration of Cas mRNA and guide RNA delivered. Optimal concentrations of Cas mRNA and guide RNA can be determined by testing different concentrations in a cellular or non-human eukaryote animal model and using deep sequencing the analyze the extent of modification at potential off-target genomic loci. Alternatively, to minimize the level of toxicity and off-target effect, Cas nickase mRNA (for example *S. pyogenes* Cas9 with the D10A mutation), or enzymatically inactive Cas mRNA (dCas) can be delivered with a pair of guide RNAs targeting a site of interest. Guide sequences and strategies to minimize toxicity and off-target effects can be as in WO 2014/093622 (PCT/US2013/074667); or, via mutation as herein.

[0086] A Cas protein can comprise a modified form of a wild type Cas protein. The modified form of the wild type Cas protein can comprise an amino acid change (e.g., deletion, insertion, or substitution) that reduces the nucleic acid-cleaving activity of the Cas protein. For example, the modified form of the Cas protein can have less than 90%, less than 80%, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nucleic acid-cleaving activity of the wild-type Cas protein (e.g., Cas9 from *S. pyogenes*). The modified form of Cas protein can have no substantial nucleic acid-cleaving activity. When a Cas protein is a modified form that has no substantial nucleic acid-cleaving activity, it can be referred to as enzymatically inactive and/or "dead" (abbreviated by "d"). A dead Cas protein (e.g., dCas12f, dCas9) can bind to a target polynucleotide but may not cleave the target polynucleotide. In some embodiments, a dead Cas protein is a dead Cas9 protein.

[0087] A dCas9 polypeptide can associate with a single guide RNA (sgRNA) to activate or repress transcription of target DNA. sgRNAs can be introduced into cells expressing the engineered chimeric receptor polypeptide. In some cases, such cells contain one or more different sgRNAs that target the same nucleic acid. In other cases, the sgRNAs target different nucleic acids in the cell. The nucleic acids targeted by the guide RNA can be any that are expressed in a cell

such as an immune cell. The nucleic acids targeted may be a gene involved in immune cell regulation. In some embodiments, the nucleic acid is associated with cancer. The nucleic acid associated with cancer can be a cell cycle gene, cell response gene, apoptosis gene, or phagocytosis gene. The recombinant guide RNA can be recognized by a CRISPR protein, a nuclease-null CRISPR protein, variants thereof, or derivatives thereof.

[0088] Enzymatically inactive can refer to a polypeptide that can bind to a nucleic acid sequence in a polynucleotide in a sequence-specific manner, but may not cleave a target polynucleotide. An enzymatically inactive site-directed polypeptide can comprise an enzymatically inactive domain (e.g. nuclease domain). Enzymatically inactive can refer to no activity. Enzymatically inactive can refer to substantially no activity. Enzymatically inactive can refer to essentially no activity. Enzymatically inactive can refer to an activity less than 1%, less than 2%, less than 3%, less than 4%, less than 5%, less than 6%, less than 7%, less than 8%, less than 9%, or less than 10% activity compared to a wild-type exemplary activity (e.g., nucleic acid cleaving activity, wild-type Cas9 activity).

[0089] One or a plurality of the nuclease domains (e.g., RuvC, HNH) of a Cas protein can be deleted or mutated so that they are no longer functional or comprise reduced nuclease activity. For example, in a Cas protein comprising at least two nuclease domains (e.g., Cas9), if one of the nuclease domains is deleted or mutated, the resulting Cas protein, known as a nickase, can generate a single-strand break at a CRISPR RNA (crRNA) recognition sequence within a double-stranded DNA but not a double-strand break. Such a nickase can cleave the complementary strand or the non-complementary strand, but may not cleave both. If all of the nuclease domains of a Cas protein (e.g., both RuvC and HNH nuclease domains in a Cas9 protein; RuvC nuclease domain in a Cpf1 protein) are deleted or mutated, the resulting Cas protein can have a reduced or no ability to cleave both strands of a double-stranded DNA. An example of a mutation that can convert a Cas9 protein into a nickase is a D10A (aspartate to alanine at position 10 of Cas9) mutation in the RuvC domain of Cas9 from *S. pyogenes*. H939A (histidine to alanine at amino acid position 839) or H840A (histidine to alanine at amino acid position 840) in the HNH domain of Cas9 from *S. pyogenes* can convert the Cas9 into a nickase. An example of a mutation that can convert a Cas9 protein into a dead Cas9 is a D10A (aspartate to alanine at position 10 of Cas9) mutation in the RuvC domain and H939A (histidine to alanine at amino acid position 839) or H840A (histidine to alanine at amino acid position 840) in the HNH domain of Cas9 from *S. pyogenes*.

[0090] A dead Cas protein can comprise one or more mutations relative to a wild-type version of the protein. The mutation can result in less than 90%, less than 80%, less than 70%, less than 60%, less than 50%, less than 40%, less than 30%, less than 20%, less than 10%, less than 5%, or less than 1% of the nucleic acid-cleaving activity in one or more of the plurality of nucleic acid-cleaving domains of the wild-type Cas protein. The mutation can result in one or more of the plurality of nucleic acid-cleaving domains retaining the ability to cleave the complementary strand of the target nucleic acid but reducing its ability to cleave the non-complementary strand of the target nucleic acid. The mutation can result in one or more of the plurality of nucleic acid-cleaving domains retaining the ability to cleave the non-complementary strand of the target nucleic acid but reducing its ability to cleave the complementary strand of the target nucleic acid. The mutation can result in one or more of the plurality of nucleic acid-cleaving domains lacking the ability to cleave the complementary strand and the non-complementary strand of the target nucleic acid. The residues to be mutated in a nuclease domain can correspond to one or more catalytic residues of the nuclease. For example, residues in the wild type exemplary *S. pyogenes* Cas9 polypeptide such as Asp10, His840, Asn854 and Asn856 can be mutated to inactivate one or more of the plurality of nucleic acid-cleaving domains (e.g., nuclease domains). The residues to be mutated in a nuclease domain of a Cas protein can correspond to residues Asp10, His840, Asn854 and Asn856 in the wild type *S. pyogenes* Cas9 polypeptide, for example, as determined by sequence and/or structural alignment.

[0091] As non-limiting examples, residues D10, G12, G17, E762, H840, N854, N863, H982, H983, A984, D986, and/or A987 (or the corresponding mutations of any of the Cas proteins) can be mutated. For example, e.g., D10A, G12A, G17A, E762A, H840A, N854A, N863A, H982A, H983A, A984A, and/or D986A. Mutations other than alanine substitutions can be suitable.

[0092] A D10A mutation can be combined with one or more of H840A, N854A, or N856A mutations to produce a Cas9 protein substantially lacking DNA cleavage activity (e.g., a dead Cas9 protein). A H840A mutation can be combined with one or more of D10A, N854A, or N856A mutations to produce a site-directed polypeptide substantially lacking DNA cleavage activity. A N854A mutation can be combined with one or more of H840A, D10A, or N856A mutations to produce a site-directed polypeptide substantially lacking DNA cleavage activity. A N856A mutation can be combined with one or more of H840A, N854A, or D10A mutations to produce a site-directed polypeptide substantially lacking DNA cleavage activity.

[0093] In some embodiments, for instance, when the system includes a Cas9 nickase or a catalytically dead Cas9, two nucleic acid molecules comprising a guide RNA sequence may be utilized. The two nucleic acid molecules may have the same or different guide RNA sequences, thus complementary to the same or different target DNA sequence. In some embodiments, the guide RNA sequences of the two nucleic acid molecules are complementary to a target DNA sequences at opposite ends (e.g., 3' or 5') and/or on opposite strands of the insert location.

[0094] In some embodiments, the Cas protein are Cas12f, which can also be referred to as Cms1, CRISPR nucleases are a class of CRISPR nucleases that have certain desirable properties relative to other CRISPR nucleases such as Cas9 nucleases.

[0095] The Cas12f polypeptides interact with specific guide RNAs (gRNAs), which direct the Cas12f endonuclease to a specific target site, at which site the Cas12f endonuclease introduces a double-stranded break that can be repaired by a DNA repair process such that the DNA sequence is modified. Since the specificity is provided by the guide RNA, the Cas12f polypeptide is universal and can be used with different guide RNAs to target different genomic sequences. Cas12f endonucleases have certain advantages over the Cas nucleases (e.g., Cas9) traditionally used with CRISPR arrays. For example, Cas12f-associated CRISPR arrays are processed into mature crRNAs without the requirement of an additional trans-activating crRNA (tracrRNA). Also, Cas12f-crRNA complexes can cleave target DNA preceded by a short protospacer-adjacent motif (PAM) that is often T-rich, in contrast to the G-rich PAM following the target DNA for many Cas9 systems. Further, Cas12f nucleases can introduce a staggered DNA double-stranded break. The methods disclosed herein can be used to target and modify specific chromosomal sequences and/or introduce exogenous sequences at targeted locations in the genome of eukaryotic and prokaryotic cells. The methods can further be used to introduce sequences or modify regions within organelles (e.g., chloroplasts and/or mitochondria). Furthermore, the targeting is specific with limited off target effects.

[0096] In general, Cas12f polypeptides comprise at least one RNA recognition and/or RNA binding domain. RNA recognition and/or RNA binding domains interact with guide RNAs. Typically the guide RNA comprises a region with a stem-loop structure that interacts with the Cas12f polypeptide. Cas12f polypeptides can also comprise nuclease domains (i.e., DNase or RNase domains), DNA binding domains, helicase domains, RNAse domains, protein-protein interaction domains, dimerization domains, as well as other domains. In specific embodiments, a

Cas12f polypeptide, or a polynucleotide encoding a Cas12f polypeptide, comprises: an RNA-binding portion that interacts with the DNA-targeting RNA, and an activity portion that exhibits site-directed enzymatic activity, such as a RuvC endonuclease domain.

[0097] Cas12f polypeptides can be wild type Cas12f polypeptides, modified Cas12f polypeptides, or a fragment of a wild type or modified Cas12f polypeptide. The Cas12f polypeptide can be modified to increase nucleic acid binding affinity and/or specificity, alter an enzymatic activity, and/or change another property of the protein. For example, nuclease (i.e., DNase, RNase) domains of the Cas12f polypeptide can be modified, deleted, or inactivated. Alternatively, the Cas12f polypeptide can be truncated to remove domains that are not essential for the function of the protein.

[0098] In some embodiments, the Cas12f polypeptide can be derived from a wild type Cas12f polypeptide or fragment thereof. In other embodiments, the Cas12f polypeptide can be derived from a modified Cas12f polypeptide. For example, the amino acid sequence of the Cas12f polypeptide can be modified to alter one or more properties (e.g., nuclease activity, affinity, stability, etc.) of the protein. Alternatively, domains of the Cas12f polypeptide not involved in RNA-guided cleavage can be eliminated from the protein such that the modified Cas12f polypeptide is smaller than the wild type Cas12f polypeptide.

[0099] The nuclease domain can be modified using well-known methods, such as site-directed mutagenesis, PCR-mediated mutagenesis, and total gene synthesis, as well as other methods known in the art. Cas12f proteins with inactivated nuclease domains (dCas12f proteins) can be used to modulate gene expression without modifying DNA sequences. Non-limiting examples of genetic modifications to Cas12f that result in inactivated nuclease domains include D225A, D401A, and/or E324A. In certain embodiments, a dCas12f protein may be targeted to particular regions of a genome such as promoters for a gene or genes of interest through the use of appropriate gRNAs. The dCas12f protein can bind to the desired region of DNA and may interfere with RNA polymerase binding to this region of DNA and/or with the binding of transcription factors to this region of DNA. This technique may be used to up- or down-regulate the expression of one or more genes of interest. In certain other embodiments, the dCas12f protein may be fused to a repressor domain to further downregulate the expression of a gene or genes whose expression is regulated by interactions of RNA polymerase, transcription factors, or other transcriptional regulators with the region of chromosomal DNA targeted by the gRNA. In certain other embodiments, the dCas12f

protein may be fused to an activation domain to effect an upregulation of a gene or genes whose expression is regulated by interactions of RNA polymerase, transcription factors, or other transcriptional regulators with the region of chromosomal DNA targeted by the gRNA.

[00100] The dCas12f polypeptides disclosed herein can further comprise at least one nuclear localization signal (NLS). In general, an NLS comprises a stretch of basic amino acids. Nuclear localization signals are known in the art (see, e.g., Lange et al., J. Biol. Chem. (2007) 282:5101-5105). The NLS can be located at the N-terminus, the C-terminus, or in an internal location of the dCas12f polypeptide. In some embodiments, the dCas12f polypeptide can further comprise at least one cell-penetrating domain. The cell-penetrating domain can be located at the N-terminus, the C-terminus, or in an internal location of the protein.

[00101] The dCas12f polypeptide disclosed herein can further comprise at least one plastid targeting signal peptide, at least one mitochondrial targeting signal peptide, or a signal peptide targeting the dCas12f polypeptide to both plastids and mitochondria. Plastid, mitochondrial, and dual-targeting signal peptide localization signals are known in the art (see, e.g., Nassoury and Morse (2005) *Biochim Biophys Acta* 1743:5-19; Kunze and Berger (2015) *Front Physiol* 6:259, Herrmann and Neupert (2003) *IUBMB Life* 55:219-225; Soll (2002) *Curr Opin Plant Biol* 5:529-535; Carrie and Small (2013) *Biochim Biophys Acta* 1833:253-259; Carrie et al. (2009) *FEBS J* 276:1187-1195; Silva-Filho (2003) *Curr Opin Plant Biol* 6:589-595; Peeters and Small (2001) *Biochim Biophys Acta* 1541:54-63; Murcha et al. (2014) *J Exp Bot* 65:6301-6335; Mackenzie (2005) *Trends Cell Biol* 15:548-554; Glaser et al. (1998) *Plant Mol Biol* 38:311-338). The plastid, mitochondrial, or dual-targeting signal peptide can be located at the N-terminus, the C-terminus, or in an internal location of the Cas12f polypeptide.

[00102] In still other embodiments, the Cas12f polypeptide can also comprise at least one marker domain. Non-limiting examples of marker domains include fluorescent proteins, purification tags, and epitope tags. In certain embodiments, the marker domain can be a fluorescent protein. Non limiting examples of suitable fluorescent proteins include green fluorescent proteins (e.g., GFP, GFP-2, tagGFP, turboGFP, EGFP, Emerald, Azami Green, Monomeric Azami Green, CopGFP, AceGFP, ZsGreen1), yellow fluorescent proteins (e.g. YFP, EYFP, Citrine, Venus, YPet, PhiYFP, ZsYellow1), blue fluorescent proteins (e.g. EBFP, EBFP2, Azurite, mKalama1, GFPuv, Sapphire, T-sapphire), cyan fluorescent proteins (e.g. ECFP, Cerulean, CyPet, AmCyan1, Midoriishi-Cyan), red fluorescent proteins (mKate, mKate2, mPlum, DsRed monomer, mCherry,

mRFP1, DsRed-Express, DsRed2, DsRed-Monomer, HcRed-Tandem, HcRed1, AsRed2, eqFP611, mRasberry, mStrawberry, Jred), and orange fluorescent proteins (mOrange, mKO, Kusabira-Orange, Monomeric Kusabira-Orange, mTangerine, tdTomato) or any other suitable fluorescent protein. In other embodiments, the marker domain can be a purification tag and/or an epitope tag. Exemplary tags include, but are not limited to, glutathione-S-transferase (GST), chitin binding protein (CBP), maltose binding protein, thioredoxin (TRX), poly(NANP), tandem affinity purification (TAP) tag, myc, AcV5, AU1, AU5, E, ECS, E2, FLAG, HA, nus, Softag 1, Softag 3, Strep, SBP, Glu-Glu, HSV, KT3, S, S1, T7, V5, VSV-G, 6×His, biotin carboxyl carrier protein (BCCP), and calmodulin.

[00103] In certain embodiments, the Cas12f polypeptide may be part of a protein-RNA complex comprising a guide RNA. The guide RNA interacts with the Cas12f polypeptide to direct the Cas12f polypeptide to a specific target site, wherein the 5′ end of the guide RNA can base pair with a specific protospacer sequence of the nucleotide sequence of interest in the plant genome, whether part of the nuclear, plastid, and/or mitochondrial genome. As used herein, the term "DNA-targeting RNA" refers to a guide RNA that interacts with the Cas12f polypeptide and the target site of the nucleotide sequence of interest in the genome of a plant cell. A DNA-targeting RNA, or a DNA polynucleotide encoding a DNA-targeting RNA, can comprise: a first segment comprising a nucleotide sequence that is complementary to a sequence in the target DNA, and a second segment that interacts with a Cas12f polypeptide.

[00104] The polynucleotides encoding Cas12f polypeptides disclosed herein can be used to isolate corresponding sequences from other prokaryotic or eukaryotic organisms, or from metagenomically-derived sequences whose native host organism is unclear or unknown. In this manner, methods such as PCR, hybridization, and the like can be used to identify such sequences based on their sequence homology or identity to the sequences set forth herein. Sequences isolated based on their sequence identity to the entire Cas12f sequences set forth herein or to variants and fragments thereof are encompassed by the present invention. Such sequences include sequences that are orthologs of the disclosed Cas12f sequences. "Orthologs" is intended to mean genes derived from a common ancestral gene and which are found in different species as a result of speciation. Genes found in different species are considered orthologs when their nucleotide sequences and/or their encoded protein sequences share at least about 75%, about 80%, about 85%, about 90%, about 91%, about 92%, about 93%, about 94%, about 95%, about 96%, about 97%,

about 98%, about 99%, or greater sequence identity. Functions of orthologs are often highly conserved among species. Thus, isolated polynucleotides that encode polypeptides having Cas12f endonuclease activity and which share at least about 75% or more sequence identity to the sequences disclosed herein, are encompassed by the present invention. As used herein, Cas12f endonuclease activity refers to CRISPR endonuclease activity wherein, a guide RNA (gRNA) associated with a Cas12f polypeptide causes the Cas12f -gRNA complex to bind to a pre-determined nucleotide sequence that is complementary to the gRNA; and wherein Cas12f activity can introduce a double-stranded break at or near the site targeted by the gRNA. In certain embodiments, this double-stranded break may be a staggered DNA double-stranded break. As used herein a "staggered DNA double-stranded break" can result in a double strand break with about 1, about 2, about 3, about 4, about 5, about 6, about 7, about 8, about 9, or about 10 nucleotides of overhang on either the 3′ or 5′ ends following cleavage. In specific embodiments, the Cas12f polypeptide introduces a staggered DNA double-stranded break with a 5′ overhang. The double strand break can occur at or near the sequence to which the DNA-targeting RNA (e.g., guide RNA) sequence is targeted.

[00105] Fragments and variants of the Cas12f polynucleotides and Cas12f amino acid sequences encoded thereby that retain Cas12f nuclease activity are encompassed herein. By "Cas12f nuclease activity" is intended the binding of a pre-determined DNA sequence as mediated by a guide RNA. In embodiments wherein the Cas12f nuclease retains a functional RuvC domain, Cas12f nuclease activity can further comprise double-strand break induction. By "fragment" is intended a portion of the polynucleotide or a portion of the amino acid sequence. "Variants" is intended to mean substantially similar sequences. For polynucleotides, a variant comprises a polynucleotide having deletions (i.e., truncations) at the 5′ and/or 3′ end; deletion and/or addition of one or more nucleotides at one or more internal sites in the native polynucleotide; and/or substitution of one or more nucleotides at one or more sites in the native polynucleotide. As used herein, a "native" polynucleotide or polypeptide comprises a naturally occurring nucleotide sequence or amino acid sequence, respectively. Generally, variants of a particular polynucleotide of the invention will have at least about 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more sequence identity to that particular polynucleotide as determined by sequence alignment programs and parameters as described elsewhere herein.

[00106] Fusion proteins are provided herein comprising a Fusion proteins are provided herein comprising a Cas12f polypeptide, or a fragment or variant thereof, and an effector domain. The Cas12f polypeptide can be directed to a target site by a guide RNA, at which site the effector domain can modify or effect the targeted nucleic acid sequence. The effector domain can be a cleavage domain, an epigenetic modification domain, a transcriptional activation domain, or a transcriptional repressor domain. The fusion protein can further comprise at least one additional domain chosen from a nuclear localization signal, plastid signal peptide, mitochondrial signal peptide, signal peptide capable of protein trafficking to multiple subcellular locations, a cell-penetrating domain, or a marker domain, any of which can be located at the N-terminus, C-terminus, or an internal location of the fusion protein. The Cas12f polypeptide can be located at the N-terminus, the C-terminus, or in an internal location of the fusion protein. The Cas12f polypeptide can be directly fused to the effector domain, or can be fused with a linker. In specific embodiments, the linker sequence fusing the Cas12f polypeptide with the effector domain can be at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 30, 40, or 50 amino acids in length. For example, the linker can range from 1-5, 1-10, 1-20, 1-50, 2-3, 3-10, 3-20, 5-20, or 10-50 amino acids in length.

[00107] In some embodiments, the Cas12f polypeptide of the fusion protein can be derived from a wild type Cas12f protein. The Cas12f -derived protein can be a modified variant or a fragment. In some embodiments, the Cas12f polypeptide can be modified to contain a nuclease domain (e.g. a RuvC or RuvC-like domain) with reduced or eliminated nuclease activity. For example, the Cas12f -derived polypeptide can be modified such that the nuclease domain is deleted or mutated such that it is no longer functional (i.e., the nuclease activity is absent).

[00108] As used herein, "fusion" can refer to a protein and/or nucleic acid comprising one or more non-native sequences (e.g., moieties). A fusion can comprise one or more of the same non-native sequences. A fusion can comprise one or more of different non-native sequences. A fusion can be a chimera. A fusion can comprise a nucleic acid affinity tag. A fusion can comprise a barcode. A fusion can comprise a peptide affinity tag. A fusion can provide for subcellular localization of the site-directed polypeptide (e.g., a nuclear localization signal (NLS) for targeting to the nucleus, a mitochondrial localization signal for targeting to the mitochondria, a chloroplast localization signal for targeting to a chloroplast, an endoplasmic reticulum (ER) retention signal, and the like). A fusion can provide a non-native sequence (e.g., affinity tag) that can be used to

track or purify. A fusion can be a small molecule such as biotin or a dye such as alexa fluor dyes, Cyanine3 dye, Cyanine5 dye.

[00109] A fusion can refer to any protein with a functional effect. For example, a fusion protein can comprise methyltransferase activity, demethylase activity, dismutase activity, alkylation activity, depurination activity, oxidation activity, pyrimidine dimer forming activity, integrase activity, transposase activity, recombinase activity, polymerase activity, ligase activity, helicase activity, photolyase activity or glycosylase activity, acetyltransferase activity, deacetylase activity, kinase activity, phosphatase activity, ubiquitin ligase activity, deubiquitinating activity, adenylation activity, deadenylation activity, SUMOylating activity, deSUMOylating activity, ribosylation activity, deribosylation activity, myristoylation activity, remodelling activity, protease activity, oxidoreductase activity, transferase activity, hydrolase activity, lyase activity, isomerase activity, synthase activity, synthetase activity, or demyristoylation activity. An effector protein can modify a genomic locus. A fusion protein can be a fusion in a Cas protein. An fusion protein can be a non-native sequence in a Cas protein.

[00110] In some embodiments, the system further comprises a recruitment system comprising at least one aptamer sequence and an aptamer binding protein functionally linked to the recombination protein as part of a fusion protein.

[00111] In some embodiments, the aptamer sequence is an RNA aptamer sequence. In some embodiments, the nucleic acid molecule comprising the guide RNA also comprises one or more RNA aptamers, or distinct RNA secondary structures or sequences that can recruit and bind another molecular species, an adaptor molecule, such as a nucleic acid or protein. Several CRISPR systems are compatible with guide RNA insertions and extensions, including but not limited to SpCas9, SaCas9, and LbCasl2a (aka Cpfl). The RNA 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 molecular species. In some embodiments, the nucleic acid comprises two or more aptamer sequences. The aptamer sequences may be the same or different and may target the same or different adaptor proteins. In select embodiments, the nucleic acid comprises two aptamer sequences.

[00112] Any RNA aptamer/ aptamer binding protein pair known may be selected and used in connection with the present invention (see, e.g., Jayasena, S.D., Clinical Chemistry, 1999. 45(9):

p. 1628-1650; Gelinas, et al., Current Opinion in Structural Biology, 2016. 36: p. 122-132; and Hasegawa, H., Molecules, 2016; 21(4): p. 421, incorporated herein by reference).

[00113] A number of RNA aptamer binding, or adaptor, proteins exist, including a diverse array of bacteriophage coat proteins. Examples of such coat proteins include but are not limited to: MS2, Q~, F2, GA, fr, JP501, Ml2, RI 7, BZ13, JP34, JP500, KUI, Ml 1, MXI, TW18, VK, SP, Fl, ID2, NL95, TW19, AP205, ~Cb5, ~Cb8r, ~Cbl2r, ~Cb23r, 7s and PRRI. In some embodiments, the RNA aptamer binds MS2 bacteriophage coat protein or a functional derivative, fragment or variant thereof. MS2 binding RNA aptamers commonly have a simple stem-loop structure, classically defined by a 19 nucleotide RNA molecule with a single bulged adenine on the 5' leg of the stem (Witherall G.W., et al., (1991) Prog. Nucleic Acid Res. Mol. Biol., 40, 185-220, incorporated herein by reference). However, a number of vastly different primary sequences were found to be able to bind the MS2 coat protein (MCP) (Parrott AM, et al., Nucleic Acids Res. 2000;28(2):489-497, Buenrostro JD, et al. Natura Biotechnology 2014; 32, 562-568, and incorporated herein by reference). Any of the RNA aptamer sequence known to bind the MS2 bacteriophage coat protein (MCP) may be utilized in connection with the present invention to bind to fusion proteins comprising MS2. In select embodiments, the MS2 RNA aptamer sequence comprises: AACAUGAGGAUCACCCAUGUCUGCAG, AGCAUGAGGAUCACCCAUGUCUGCAG, or AGCGUGAGGAUCACCCAUGCCUGCAG.

[00114] N-proteins (Nut-utilization site proteins) of bacteriophages contain arginine-rich conserved RNA recognition motifs of ~20 amino acids, referred to as N peptides. The RNA aptamer may bind a phage N peptide or a functional derivative, fragment or variant thereof. In some embodiments, the phage N peptide is the lambda or P22 phage N peptide or a functional derivative, fragment or variant thereof.

[00115] In select embodiments, the N peptide is lambda phage N22 peptide, or a functional derivative, fragment or variant thereof. In some embodiments, the N22 peptide comprises an amino acid sequence with at least 70% similarity to the ammo acid sequence GNARTRRRERRAEKQAQWKAAN. N22 peptide, the 22 amino acid RNA binding domain of the 11. bacteriophage antiterminator protein N (11.N-(1-22) or 11.N peptide), is capable of specifically binding to specific stem-loop structures, including but not limited to the BoxB stem-loop. See, for example Cilley and Williamson, RNA 1997; 3(1):57-67, incorporated herein by reference. A number of different BoxB stem-loop primary sequences are known to bind the N22

peptide and any of those may be utilized in connection with the present invention. In some embodiments, the N22 peptide RNA aptamer sequence comprises a nucleotide sequence with at least 70% similarity to an RNA sequence selected from the group consisting of GCCCUGAAAAAGGGC, GCCCUGAAGAAGGGC, GCGCUGAAAAAGCGC, GCCCUGACAAAGGGC, and GCGCUGACAAAGCGC.

[00116] In select embodiments, the N peptide is the P22 phage N peptide, or a functional derivative, fragment or variant thereof. A number of different BoxB stem-loop primary sequences are known to bind the P22 phage N peptide and variants thereof and any of those may be utilized in connection with the present invention. See, for example Cocozaki, Ghattas, and Smith, Journal of Bacteriology 2008; 190(23):7699-7708, incorporated herein by reference. In some embodiments, the P22 phage N peptide comprises an amino acid sequence with at least 70% similarity to the amino acid sequence GNAKTRRHERRRKLAIERDTI. In some embodiments, the P22 phage N peptide RNA aptamer sequence comprises a sequence with at least 70% similarity to an RNA sequence selected from the group consisting of GCGCUGACAAAGCG and CCGCCGACAACGCGG.

[00117] In certain embodiments, different aptamer/aptamer binding protein pairs can be selected to bring together a combination of recombination proteins and functions. In some embodiments, the aptamer sequence is a peptide aptamer sequence. The peptide aptamers can be naturally occurring or synthetic peptides that are specifically recognized by an affinity agent. Such aptamers include, but are not limited to, a c-Myc affinity tag, an HA affinity tag, a His affinity tag, an S affinity tag, a methionine-His affinity tag, an RGD-His affinity tag, a 7x His tag, a FLAG octapeptide, a strep tag or strep tag II, a VS tag, or a VSV-G epitope. Corresponding aptamer binding proteins are well-known in the art and include, for example, primary antibodies, biotin, affimers, single domain antibodies, and antibody mimetics.

[00118] An exemplary peptide aptamer includes a GCN4 peptide (Tanenbaum et al., Cell 2014; 159(3):635-646, incorporated herein by reference). Antibodies, or GCN4 binding protein can be used as the aptamer binding proteins.

[00119] In some embodiments, the peptide aptamer sequence is conjugated to the Cas protein. The peptide aptamer sequence may be fused to the Cas in any orientation (e.g., N-terminus to Cterminus, C-terminus to N-terminus, N-terminus to N-terminus). In select embodiments, the peptide aptamer is fused to the C-terminus of the Cas protein.

[00120] In some embodiments, between 1 and 24 peptide aptamer sequences may be conjugated to the Cas protein. The aptamer sequences may be the same or different and may target the same or different aptamer binding proteins. In select embodiments, 1 to 24 tandem repeats of the same peptide aptamer sequence are conjugated to the Cas protein. In some embodiments between 4 and 18 tandem repeats are conjugated to the Cas protein. The individual aptamers may be separated by a linker region. Suitable linker regions are known in the art. The linker may be flexible or configured to allow the binding of affinity agents to adjacent aptamers without or with decreased steric hindrance. The linker sequences may provide an unstructured or linear region of the polypeptide, for example, with the inclusion of one or more glycine and/or serine residues. The linker sequences can be at least about 2, 3, 4, 5, 6, 7, 8, 9, 10 or more amino acids in length.

[00121] In some embodiments, the fusion protein comprises a recombination protein functionally linked to an aptamer binding protein. In some embodiments, the recombination protein comprises a microbial recombination protein. In some embodiments, the recombination protein comprises a recombinase. In some embodiments, the recombination protein comprises a nuclease. In some embodiments, the recombination protein comprises an endonuclease. In certain embodiments, the recombination protein comprises 5' -3' exonuclease activity . In certain embodiments, the recombination protein comprises 3 '-5' exonuclease activity . In certain embodiments, the recombination protein comprises ssDNA binding activity . In certain embodiments, the recombination protein comprises ssDNA annealing activity .

[00122] The bacteriophage A-encoded genetic recombination machinery, named the 11, red system, comprises the exo and bet genes, assisted by the gam gene, together designated 11, red genes. Exo is a 5'-3' exonuclease which targets dsDNA and Bet is a ssDNA-binding protein. Bet functions include protecting ssDNA from degradation and promoting annealing of complementary ssDNA strands. Another bacteriophage system found in E. coli is the Rae prophage system, comprising recE and recT genes which are functionally similar to exo and bet . In some embodiments, the microbial recombination protein may be RecE, RecT, lambda exonuclease (Exo), Bet protein (betA, redB), exonuclease gp6, single-stranded DNA-binding protein gp2.5, or a derivative or variant thereof.

[00123] Recombination proteins and functional fragments thereof useful in the invention include nucleases, ssDNA-binding proteins (SSBs), and ssDNA annealing proteins (SSAPs) . Among microbial proteins, these include, without limitation, E. coli proteins such as ExoI (xonA;

sbcB), ExoIII (xthA), ExoIV (orn), ExoVII (xseA, xseB), ExoIX (ygdG), ExoX (exoX), DNA poll 5' Exo (ExoVI) (polA), DNA Pol I 3' Exo (ExoII) (polA), DNA Pol II 3' Exo (polB), DNA Pol III 3' Exo (dnaQ, mutD), RecBCD (recB, recC, recD), and RecJ (reel) and their functional fragments.

[00124] While double-stranded DNA contains genetic information, use of the information involves single-stranded intermediates. Whereas the single-stranded intermediates form secondary structures and are sensitive to chemical and nucleolytic degradation, cells encode ssDNA binding proteins (SSBs) that bind to and stabilize ssDNA. Useful SSBs include, without limitation, SSBs of prokaryotes, bacteriophage, eukaryotes, mammals, mitochondria, and viruses . While SSBs are found in every organism, the proteins themselves share surprisingly little sequence similarity, and may differ in subunit composition and oligomerization states. SSB proteins may comprise certain structural features . One is use of oligonucleotide/oligosaccharide-binding (OB) domains to bind ssDNA through a combination of electrostatic and base-stacking interactions with the phosphodiester backbone and nucleotide bases . Another feature is oligomerization that brings together DNA-binding OB folds . Eukaryotic SSBs are regulated by phosphorylation on serine and threonine residues . Tyrosine phosphorylation of microbial SSBs is observed in taxonomically distant bacteria and substantially increases affinity for ssDNA. The human mitochondrial ssDNAbinding protein is structurally similar to SSB from Escherichia coli (EcoSSB), but lacks the Cterminal disordered domain. Eukaryotic replication protein A (RP A) shares function, but not sequence homology with bacterial SSB. The herpes simplex virus (HSV-1) SSB, ICP8, is a nuclear protein that, along other replication proteins is required for viral DNA replication.

[00125] Without being bound by theory, it is thought that exonuclease activities and ssDNA binding activities of the recombination proteins of the invention uncover and protect single stranded regions of template and target DNAs, thereby facilitating recombination. Also, targeting can be cooperative, involving target directed CRISPR-mediated nicking, cleaving, or binding of chromosomal DNA coordinated with recombination directed by homology arms designed into donor DNAs . In certain embodiments of the invention, off-target effects are minimized . For example, whereas targeted recombination involves coordinated CRISPR and recombination functions, at off-target sites, homology with the HR donor DNA is absent and nick repair may be favored.

[00126] Single stranded DNA annealing proteins (SSAPs) also are ubiquitous among organisms with diverse sequences and have been classified into families and superfamilies by bioinformatics

and experimental analysis. Moreover, phage encoded SSAPs are recognized to encode their own SSAP recombinases which substitute for classic RecA proteins while functioning with host proteins to control DNA metabolism . Steczkiewiz classified SSAPs into seven families (RecA, Gp2.5, RecT/Red~, Erf, Rad52/22, Sak3, and Sak4) organized into three superfamilies including prokaryotes, eukaryotes, and phage (Steczkiewicz et al., 2021, Front. Microbial 12:644622).

[00127] In certain embodiments, a microbial recombination protein is RecE or RecT, or a derivative or variant thereof. Derivatives or variants of RecE and RecT are functionally equivalent proteins or polypeptides which possess substantially similar function to wild type RecE and RecT. RecE and RecT derivatives or variants include biologically active amino acid sequences similar to the wild-type sequences but differing due to amino acid substitutions, additions, deletions, truncations, post-translational modifications, or other modifications. In some embodiments, the derivatives may improve translation, purification, biological half-life, activity, or eliminate or lessen any undesirable side effects or reactions. The derivatives or variants may be naturally occurring polypeptides, synthetic or chemically synthesized polypeptides or genetically engineered peptide polypeptides. RecE and RecT bioactivities are known to, and easily assayed by, those of ordinary skill in the art, and include, for example exonuclease and single-stranded nucleic acid binding, respectively.

[00128] A Cas protein as used herein can be a wildtype or a modified form of a Cas protein. A Cas protein can be an active variant, inactive variant, or fragment of a wild type or modified Cas protein. A Cas protein can comprise an amino acid change such as a deletion, insertion, substitution, variant, mutation, fusion, chimera, or any combination thereof relative to a wild-type version of the Cas protein. A Cas protein can be a polypeptide with at least about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity or sequence similarity to a wild type exemplary Cas protein. A Cas protein can be a polypeptide with at most about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100% sequence identity and/or sequence similarity to a wild type exemplary Cas protein. Variants or fragments can comprise at least about 5%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity or sequence similarity to a wild type or modified Cas protein or a portion thereof. Variants or fragments can be targeted to a nucleic acid locus in complex with a guide nucleic acid while lacking nucleic acid cleavage activity.

[00129] A Cas protein can be a fusion protein. For example, a Cas protein can be fused to a cleavage domain, an epigenetic modification domain, a transcriptional activation domain, or a transcriptional repressor domain. A Cas protein can also be fused to a heterologous polypeptide providing increased or decreased stability. The fused domain or heterologous polypeptide can be located at the N-terminus, the C-terminus, or internally within the Cas protein.

[00130] In a fusion protein, a microbial recombination protein may be linked to either terminus of an aptamer binding protein in any orientation (e.g., N-terminus to C-terminus, C-terminus to N-terminus, N-terminus to Nterminus). In select embodiments, a microbial recombination protein N-terminus is linked to the aptamer binding protein C-terminus. Thus, the overall fusion protein from N- to C-terminus comprises the aptamer binding protein (N- to C-terminus) linked to the microbial recombination protein (N- to C-terminus).

[00131] Typically, in the context of an endogenous CRISPR system, formation of a CRISPR complex (comprising a guide sequence hybridized to a target sequence and complexed with one or more Cas proteins) results in cleavage of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. Without wishing to be bound by theory, the tracr sequence, which may comprise or consist of all or a portion of a wild-type tracr sequence (e.g. about or more than about 20, 26, 32, 45, 48, 54, 63, 67, 85, or more nucleotides of a wild-type tracr sequence), may also form part of a CRISPR complex, such as by hybridization along at least a portion of the tracr sequence to all or a portion of a tracr mate sequence that is operably linked to the guide sequence.

[00132] The nucleic acid molecule encoding a Cas is advantageously codon optimized Cas. An example of a codon optimized sequence, is in this instance a sequence optimized for expression in a eukaryote, e.g., humans (i.e. being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed; see, e.g., SaCas9 human codon optimized sequence in WO 2014/093622 (PCT/US2013/074667). Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species other than human, or for codon optimization for specific organs is known.

[00133] In some embodiments, an enzyme coding sequence encoding a Cas is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock,

or non-human mammal or primate. In some embodiments, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded.

[00134] 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 Cas correspond to the most frequently used codon for a particular amino acid.

[00135] In some embodiments, the Cas sequence is fused to one or more nuclear localization sequences (NLSs), such as about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more NLSs. In some embodiments, the Cas 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.

[00136] In an embodiment of the invention, the Cas comprises at most 6 NLSs. 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. Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV; the NLS from nucleoplasmin (e.g. the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK); the c-myc NLS having the amino acid sequence PAAKRVKLD or RQRRNELKRSP; the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY; the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV of the IBB domain from importin-alpha; the sequences VSRKRPRP and PPKKARED of the myoma T protein; the sequence PQPKKKPL of human p53; the sequence SALIKKKKKMAP of mouse c-abl IV; the sequences DRLRR and PKQKKRK of the influenza virus NS1; the sequence RKLKKKIKKL of the Hepatitis virus delta antigen; the sequence REKKKFLKRR of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK of the steroid hormone receptors (human) glucocorticoid.

[00137] In general, the one or more NLSs are of sufficient strength to drive accumulation of the Cas 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 Cas, 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 Cas, such that location within a cell may be visualized, such as in combination with a means for detecting the location of the nucleus (e.g. a stain specific for the nucleus such as DAPI). Cell nuclei may also be isolated from cells, the contents of which may then be analyzed by any suitable process for detecting protein, such as immunohistochemistry, Western blot, or enzyme activity assay. Accumulation in the nucleus may also be determined indirectly, such as by an assay for the effect of CRISPR complex formation (e.g. assay for DNA cleavage or mutation at the target sequence, or assay for altered gene expression activity affected by CRISPR complex formation and/or Cas

enzyme activity), as compared to a control no exposed to the Cas or complex, or exposed to a Cas lacking the one or more NLSs. In other embodiments, no NLS is required.

[00138] As used herein, the term "crRNA" or "guide RNA" or "single guide RNA" or "sgRNA" or "one or more nucleic acid components" of a Type II CRISPR-Cas locus effector protein comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. 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.

[00139] 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, Calif.), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid-targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art.

[00140] A guide sequence, and hence a nucleic acid-targeting guide RNA may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence

may be any RNA sequence. In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmatic RNA (scRNA). In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[00141] In some embodiments, a nucleic acid-targeting guide RNA is selected to reduce the degree secondary structure within the RNA-targeting guide RNA. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the nucleic acid-targeting guide RNA participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g., A. R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[00142] In certain embodiments, a guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat (DR) sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence. In certain embodiments, the direct repeat sequence may be located upstream (i.e., 5′) from the guide sequence or spacer sequence. In other embodiments, the direct repeat sequence may be located downstream (i.e., 3′) from the guide sequence or spacer sequence.

[00143] In certain embodiments, the crRNA comprises a stem loop. In certain embodiments, the direct repeat sequence forms a stem loop.

[00144] The "tracrRNA" sequence or analogous terms includes any polynucleotide sequence that has sufficient complementarity with a crRNA sequence to hybridize. In general, degree of complementarity is with reference to the optimal alignment of the tracr mate sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the tracr sequence or tracr mate sequence. In some embodiments, the degree of complementarity between the tracr sequence and the tracr mate sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher.

[00145] A guide sequence may be selected to target any target sequence. In some embodiments, the target sequence is a sequence within a genome of a cell. Exemplary target sequences include those that are unique in the target genome. For example, for the *S. pyogenes* Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXGG where NNNNNNNNNNNNXGG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. pyogenes* Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXGG where NNNNNNNNNNNNXGG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. For the *S. thermophiles* CRISPR1 Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNNXXAGAAW where NNNNNNNNNNNNXXAGAAW (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. thermophiles* CRISPR1 Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXXAGAAW where NNNNNNNNNNNNXXAGAAW (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. For the *S. pyogenes* Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXGGXG where NNNNNNNNNNNNXGGXG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. A unique target sequence in a genome may include an *S. pyogenes* Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXGGXG where NNNNNNNNNNNXGGXG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. In each of these sequences "M" may be A, G, T, or C, and need not be considered in identifying a sequence as

unique. In some embodiments, a guide sequence is selected to reduce the degree secondary structure within the guide sequence. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the guide sequence 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).

[00146] In general, a tracr mate sequence includes any sequence that has sufficient complementarity with a tracr sequence to promote one or more of: (1) excision of a guide sequence flanked by tracr mate sequences in a cell containing the corresponding tracr sequence; and (2) formation of a CRISPR complex at a target sequence, wherein the CRISPR complex comprises the tracr mate sequence hybridized to the tracr sequence. In general, degree of complementarity is with reference to the optimal alignment of the tracr mate sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the tracr sequence or tracr mate sequence. In some embodiments, the degree of complementarity between the tracr sequence and tracr mate sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and tracr mate sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. In an embodiment of the invention, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In some embodiments, the transcript has two, three, four or five hairpins. In a further embodiment of the invention, the transcript has at most five hairpins.

[00147] In a hairpin structure the portion of the sequence 5′ of the final "N" and upstream of the loop corresponds to the tracr mate sequence, and the portion of the sequence 3′ of the loop corresponds to the tracr sequence Further non-limiting examples of single polynucleotides comprising a guide sequence, a tracr mate sequence, and a tracr sequence are as follows (listed 5′ to 3′), where "N" represents a base of a guide sequence, the first block of lower case letters represent the tracr mate sequence, and the second block of lower case letters represent the tracr sequence, and the final poly-T sequence represents the transcription terminator: (1) NNNNNNNNNNNNNNNNNNNgtttttgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagataa ggcttcatgccgaaatcaacaccctgtcatttatggcagggtgttttcgttatttaaTTTTTT; (2) NNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccg aaatcaacaccctgtcatttatggcagggtgttttcgttatttaaTTTTTT; (3) NNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccg aaatcaacaccctgtcatttatggcagggtgtTTTTTT; (4) NNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaactt gaaaaagtggcaccgagtcggtgcTTTTTT; (5) NNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagttaaaataaggctagtccgttatcaac ttgaaaaagtgTTTTTTT; and (6) NNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaagttaaaataaggctagtccgttatcaTT TTTTTT.

[00148] In some embodiments, sequences (1) to (3) are used in combination with Cas9 from *S. thermophilus* CRISPR1. In some embodiments, sequences (4) to (6) are used in combination with Cas9 from *S. pyogenes*. In some embodiments, the tracr sequence is a separate transcript from a transcript comprising the tracr mate sequence. In some embodiments it may be preferred in a CRISPR complex that the tracr sequence has one or more hairpins and is 30 or more nucleotides in length, 40 or more nucleotides in length, or 50 or more nucleotides in length; the guide sequence is between 10 to 30 nucleotides in length, the CRISPR/Cas enzyme is a Type II Cas9 enzyme.

[00149] In some embodiments, candidate tracrRNA may be subsequently predicted by sequences that fulfill any or all of the following criteria: 1. sequence homology to direct repeats (motif search in Geneious with up to 18-bp mismatches); 2. presence of a predicted Rho-independent transcriptional terminator in direction of transcription; and 3. stable hairpin secondary

structure between tracrRNA and direct repeat. In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[00150] In some embodiments, chimeric synthetic guide RNAs (sgRNAs) designs may incorporate at least 12 bp of duplex structure between the direct repeat and tracrRNA.

[00151] Donor nucleic acids can be single-strand or double-stranded DNA and comprise (1) various lengths of homology arms (HA) to match a genomic target region, and (2) a transgene, i.e. knock-in sequence or replacement sequence etc. There is no limit to the sized of the transgene .

[00152] In another embodiment, the target genomic DNA sequence can comprise a gene, the mutation of which contributes to a particular disease in combination with mutations in other genes. Diseases caused by the contribution of multiple genes which lack simple (e.g., Mendelian) inheritance patterns are referred to in the art as a "multifactorial" or "polygenic" disease. Examples of multifactorial or polygenic diseases include, but are not limited to, asthma, diabetes, epilepsy, hypertension, bipolar disorder, and schizophrenia. Certain developmental abnormalities also can be inherited in a multifactorial or polygenic pattern and include, for example, cleft lip/palate, congenital heart defects, and neural tube defects.

[00153] In another embodiment, the method of altering a target genomic DNA sequence can be used to delete nucleic acids from a target sequence in a cell by cleaving the target sequence and allowing the cell to repair the cleaved sequence in the absence of an exogenously provided donor nucleic acid molecule. Deletion of a nucleic acid sequence in this manner can be used in a variety of applications, such as, for example, to remove disease-causing trinucleotide repeat sequences in neurons, to create gene knock-outs or knock-downs, and to generate mutations for disease models in research.

[00154] The term "donor nucleic acid molecule" refers to a nucleotide sequence that is inserted into the target DNA (e.g., genomic DNA). As described above the donor DNA may include, for example, a gene or part of a gene, a sequence encoding a tag or localization sequence, or a regulating element. The donor nucleic acid molecule may be of any length. In some embodiments, the donor nucleic acid molecule is between 10 and 10,000 nucleotides in length. For example, between about 100 and 5,000 nucleotides in length, between about 200 and 2,000 nucleotides in length, between about 500 and 1,000 nucleotides in length, between about 500 and 5,000 nucleotides in length, between about 1,000 and 5,000 nucleotides in length, or between about 1,000 and 10,000 nucleotides in length.

[00155] The CRISPR system is derived advantageously from a type II CRISPR system. In some embodiments, one or more elements of a CRISPR system is derived from a particular organism comprising an endogenous CRISPR system, such as *Streptococcus pyogenes*. In some embodiments of the invention, the CRISPR system is a type II CRISPR system and the Cas enzyme is Cas9, which catalyzes DNA cleavage. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Cas12, Cas12a, Cas12f (also known as Csm1), Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologues thereof, or modified versions thereof.

[00156] In an embodiment, the Cas9 protein may be an ortholog of an organism of a genus which includes but is not limited to *Corynebacter, Sutterella, Legionella, Treponema, Filifactor, Eubacterium, Streptococcus, Lactobacillus, Mycoplasma, Bacteroides, Flaviivola, Flavobacterium, Sphaerochaeta, Azospirillum, Gluconacetobacter, Neisseria, Roseburia, Parvibaculum, Staphylococcus, Nitratifractor, Mycoplasma* and *Campylobacter*. Species of an organism of such a genus can be as otherwise herein discussed.

[00157] It will be appreciated that any of the functionalities described herein may be engineered into CRISPR enzymes from other orthologs, including chimeric enzymes comprising fragments from multiple orthologs. Examples of such orthologs are described elsewhere herein. Thus, chimeric enzymes may comprise fragments of CRISPR enzyme orthologs of an organism which includes but is not limited to *Corynebacter, Sutterella, Legionella, Treponema, Filifactor, Eubacterium, Streptococcus, Lactobacillus, Mycoplasma, Bacteroides, Flaviivola, Flavobacterium, Sphaerochaeta, Azospirillum, Gluconacetobacter, Neisseria, Roseburia, Parvibaculum, Staphylococcus, Nitratifractor, Mycoplasma* and *Campylobacter*. A chimeric enzyme can comprise a first fragment and a second fragment, and the fragments can be of CRISPR enzyme orthologs of organisms of genuses herein mentioned or of species herein mentioned; advantageously the fragments are from CRISPR enzyme orthologs of different species

[00158] In some embodiments, the unmodified CRISPR enzyme has DNA cleavage activity, such as Cas9. In some embodiments, the CRISPR enzyme directs cleavage of one or both strands at the location of a target sequence, such as within the target sequence and/or within the complement of the target sequence. In some embodiments, the CRISPR enzyme directs cleavage

of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence. In some embodiments, a vector encodes a CRISPR enzyme that is mutated to with respect to a corresponding wild-type enzyme such that the mutated CRISPR enzyme lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence. For example, an aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of Cas9 from *S. pyogenes* converts Cas9 from a nuclease that cleaves both strands to a nickase (cleaves a single strand). Other examples of mutations that render Cas9 a nickase include, without limitation, H840A, N854A, and N863A. As a further example, two or more catalytic domains of Cas9 (RuvC I, RuvC II, and RuvC III or the HNH domain) may be mutated to produce a mutated Cas9 substantially lacking all DNA cleavage activity.

[00159] In some embodiments, a D10A mutation is combined with one or more of H840A, N854A, or N863A mutations to produce a Cas9 enzyme substantially lacking all DNA cleavage activity. In some embodiments, a CRISPR enzyme 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. Where the enzyme is not SpCas9, mutations may be made at any or all residues corresponding to positions 10, 762, 840, 854, 863 and/or 986 of SpCas9 (which may be ascertained for instance by standard sequence comparison tools). In particular, any or all of the following mutations are envisioned in SpCas9: D10A, E762A, H840A, N854A, N863A and/or D986A; as well as conservative substitution for any of the replacement amino acids is also envisaged. The same (or conservative substitutions of these mutations) at corresponding positions in other Cas9s are also envisioned. In other Cas9s, residues corresponding to SpCas9 D10 and H840 are also envisioned. Orthologs of SpCas9 can be used in the practice of the invention.

[00160] A Cas enzyme may be identified Cas9 as this can refer to the general class of enzymes that share homology to the biggest nuclease with multiple nuclease domains from the type II CRISPR system. The Cas9 enzyme is from, or is derived from, spCas9 (*S. pyogenes* Cas9) or saCas9 (*S. aureus* Cas9). StCas9" refers to wild type Cas9 from *S. thermophilus*, the protein sequence of which is given in the SwissProt database under accession number G3ECR1. Similarly,

S pyogenes Cas9 or spCas9 is included in SwissProt under accession number Q99ZW2. By derived, Applicants mean that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as described herein. It will be appreciated that the terms Cas and CRISPR enzyme are generally used herein interchangeably, unless otherwise apparent. As mentioned above, many of the residue numberings used herein refer to the Cas9 enzyme from the type II CRISPR locus in *Streptococcus pyogenes*. However, it will be appreciated that this invention includes many more Cas9s from other species of microbes, such as SpCas9, SaCa9, St1Cas9 and so forth. Enzymatic action by Cas9 derived from *Streptococcus pyogenes* or any closely related Cas9 generates double stranded breaks at target site sequences which hybridize to 20 nucleotides of the guide sequence and that have a protospacer-adjacent motif (PAM) sequence (examples include NGG/NRG or a PAM that can be determined as described herein) following the 20 nucleotides of the target sequence. CRISPR activity through Cas9 for site-specific DNA recognition and cleavage is defined by the guide sequence, the tracr sequence that hybridizes in part to the guide sequence and the PAM sequence.

[00161] More aspects of the CRISPR system are described in Karginov and Hannon, The CRISPR system: small RNA-guided defence in bacteria and archaea, Mol Cell 2010, Jan. 15; 37(1): 7. The type II CRISPR locus from *Streptococcus pyogenes* SF370, which contains a cluster of four genes Cas9, Cas1, Cas2, and Csn1, as well as two non-coding RNA elements, tracrRNA and a characteristic array of repetitive sequences (direct repeats) interspaced by short stretches of non-repetitive sequences (spacers, about 30 bp each). In this system, targeted DNA double-strand break (DSB) is generated in four sequential steps. First, two non-coding RNAs, the pre-crRNA array and tracrRNA, are transcribed from the CRISPR locus. Second, tracrRNA hybridizes to the direct repeats of pre-crRNA, which is then processed into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas9 to the DNA target consisting of the protospacer and the corresponding PAM via heteroduplex formation between the spacer region of the crRNA and the protospacer DNA. Finally, Cas9 mediates cleavage of target DNA upstream of PAM to create a DSB within the protospacer. A pre-crRNA array consisting of a single spacer flanked by two direct repeats (DRs) is also encompassed by the term "tracr-mate sequences"). In certain embodiments, Cas9 may be constitutively present or inducibly present or conditionally present or administered or delivered. Cas9 optimization may be used to enhance

function or to develop new functions, one can generate chimeric Cas9 proteins. And Cas9 may be used as a generic DNA binding protein.

[00162] Aspects of the invention relate to the expression of the gene product being decreased or a donor polynucleotide being further introduced into the DNA molecule encoding the gene product or an intervening sequence being excised precisely by allowing the two 5′ overhangs to reanneal and ligate or the activity or function of the gene product being altered or the expression of the gene product being increased. In an embodiment of the invention, the gene product is a protein. Only sgRNA pairs creating 5′ overhangs with less than 8 bp overlap between the guide sequences (offset greater than −8 bp) were able to mediate detectable indel formation. Importantly, each guide used in these assays is able to efficiently induce indels when paired with wildtype Cas9, indicating that the relative positions of the guide pairs are the most important parameters in predicting double nicking activity.

[00163] Since Cas9n and Cas9H840A nick opposite strands of DNA, substitution of Cas9n with Cas9H840A with a given sgRNA pair should have resulted in the inversion of the overhang type; but no indel formation is observed as with Cas9H840A indicating that Cas9H840A is a CRISPR enzyme substantially lacking all DNA cleavage activity (which is 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; whereby an example can be when the DNA cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form, e.g., when no indel formation is observed as with Cas9H840A in the eukaryotic system in contrast to the biochemical or prokaryotic systems). Nonetheless, a pair of sgRNAs that will generate a 5′ overhang with Cas9n should in principle generate the corresponding 3′ overhang instead, and double nicking. Therefore, sgRNA pairs that lead to the generation of a 3′ overhang with Cas9n can be used with another mutated Cas9 to generate a 5′ overhang, and double nicking. Accordingly, in some embodiments, a recombination donor, or template, is also provided. A recombination donor may be a component of another vector as described herein, contained in a separate vector, or provided as a separate polynucleotide. In some embodiments, a recombination donor is designed to serve as a template in homologous recombination, such as within or near a target sequence nicked or cleaved by a CRISPR enzyme as a part of a CRISPR complex. A donor polynucleotide may be of any suitable length, such as about or more than about 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, or more nucleotides in length. In some

embodiments, the donor polynucleotide is complementary to a portion of a polynucleotide comprising the target sequence. When optimally aligned, a donor polynucleotide might overlap with one or more nucleotides of a target sequences (e.g. about or more than about 1, 5, 10, 15, 20, or more nucleotides). In some embodiments, when a donor sequence and a polynucleotide comprising a target sequence are optimally aligned, the nearest nucleotide of the donor polynucleotide is within about 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, or more nucleotides from the target sequence.

[00164] In prokaryotic hosts such as E. coli, homologous gene replacements can be effected with bacteriophage l Red homologous recombination systems which comprise a bacteriophage l exonuclease, a bacteriophage l Beta protein, a single-stranded DNA annealing protein (SSAP) which facilitates annealing of complementary DNA strands, and a DNA donor (Murphy, 2016). Bacteriophage l Red homologous recombination systems have been combined with CRISPR-Cas9 systems in prokaryotes to effect recombination at target sequences in bacterial genomes (Jiang et al., 2013; Wang et al., 2016).

[00165] In some embodiments, these aspects employ a single stranded DNA annealing protein (SSAP) system. Non-limiting examples of SSAPs include RecT, Redβ, ERF and Rad52.

[00166] In certain embodiments or the invention, there is provided a system or composition for RNA-guided recombination that does not rely primarily on CRISPR protens. In such embodiments, the system or composition comprises: a nucleic acid molecule comprising a guide RNA sequence that is complementary to a target DNA sequence and a recombination protein . In certain embodiments, the system or composition is capable of promoting R-loop formation . In certain embodiments, the system or composition is capable of recombination. In certain embodiments, the system or composition is free of CRISPR proteins. In certain embodiments, the recombination protein comprises a microbial recombination protein. In certain embodiments, the recombination protein comprises a viral recombination protein. In certain embodiments, the recombination protein comprises a eukaryotic recombination protein. In certain embodiments, the recombination protein comprises a mitochondrial recombination protein . In various embodiments, the recombination protein comprises a single stranded DNA annealing protein (SSAP), a nuclease, or a combination of two or more thereof. In certain embodiments, the system or composition does not comprise a Cas9. In certain embodiments, the system or composition does not comprise a Cas. In certain embodiments, the system or composition does not comprise a CRISPR.

[00167] In some embodiments, the recombination protein may be functionally linked as a fusion protein or chimera or chimeric molecule to a nuclease. In some embodiments, the recombination protein may be functionally linked as a fusion protein or chimera or chimeric molecule to an endonuclease. In some embodiments, the recombination protein may be functionally linked as a fusion protein or chimera or chimeric molecule to an exonuclease. In some embodiments, the recombination protein may be functionally linked as a fusion protein or chimera or chimeric molecule to a nuclease and/or a Cas or dCas. In some embodiments, the recombination protein may be functionally linked as a fusion protein or chimera or chimeric molecule to a endonuclease and/or a Cas or dCas. In some embodiments, the recombination protein may be functionally linked as a fusion protein or chimera or chimeric molecule to a exonuclease and/or a Cas or dCas. In some embodiments, the recombination protein may be expressed independently from, not a fusion protein with a nuclease. In some embodiments, the recombination protein may be expressed independently from, not a fusion protein with an endonuclease. In some embodiments, the recombination protein may be expressed independently from, not a fusion protein with an exonuclease. In some embodiments, the recombination protein may be expressed independently from, not a fusion protein with a nuclease and/or a Cas or dCas. In some embodiments, the recombination protein may be functionally linked as a fusion protein or chimera or chimeric molecule to an aptamer and/or aptamer binding protein. In some embodiments, the recombination protein may be expressed independently, not as a fusion protein, with an aptamer and/or aptamer binding protein. In some embodiments, the recombination protein may be functionally linked as a fusion protein or chimera or chimeric molecule to a nuclease and/or Cas or dCas and/or to an aptamer and/or aptamer binding protein. In some embodiments, the recombination protein may be expressed independently from, not a fusion protein with a nuclease and/or a Cas or dCas and/or an aptamer and/or aptamer binding protein. In some embodiments, the aptamer and/or aptamer binding protein is an MCP protein. In some embodiments the recombination protein may be an SSAP.

[00168] The term "nuclease" as used herein, refers to an agent, such as a protein or small molecule, that is capable of cleaving phosphodiester bonds that join nucleotide residues in a nucleic acid molecule. In some embodiments, the nuclease is but woven, e.g., an enzyme that is capable of binding to a nucleic acid molecule and cleaving phosphodiester bonds linking nucleotide residues in the nucleic acid molecule. The nuclease may be an endonuclease, which

cleaves a phosphodiester bond in a polynucleotide strand, or an exonuclease, which cleaves a phosphodiester bond at the end of a polynucleotide strand. In some embodiments, the nuclease is a site-specific nuclease that binds to and/or cleaves a particular phosphodiester bond within a particular nucleotide sequence, which is also referred to herein as a "recognition sequence", "nuclease target site", or "target site". In some embodiments, the nuclease is an RNA-guided (i.e., RNA-programmable) nuclease that complexes (e.g., binds) to RNA having a sequence complementary to the target site, thereby providing sequence specificity of the nuclease. In some embodiments, the nuclease recognizes a single-stranded target site, while in other embodiments, the nuclease recognizes a double-stranded target site, e.g., a double-stranded DNA target site. Target sites for many naturally occurring nucleases, for example many naturally occurring DNA restriction nucleases, are well known to those skilled in the art. In many cases, DNA nucleases such as EcoRI, HindIII or BamHI recognize palindromic double-stranded DNA target sites that are 4 to 10 base pairs in length and cut each of the two DNA strands at specific positions within the target site. Some endonucleases symmetrically cleave a double-stranded nucleic acid target site, i.e., cleave both strands at the same position, such that the ends comprise base-paired nucleotides, also referred to herein as blunt ends. Other endonucleases cleave double-stranded nucleic acid target sites asymmetrically, i.e., each strand is cleaved at a different position such that the ends contain unpaired nucleotides. Unpaired nucleotides at the ends of a double-stranded DNA molecule are also referred to as "overhangs", e.g., "5 '-overhangs" or "3' -overhangs", depending on whether the unpaired nucleotide forms the 5 'or 3' end of the corresponding DNA strand. The ends of a double-stranded DNA molecule that terminate in unpaired nucleotides are also referred to as sticky ends, so they can "stick" to the ends of other double-stranded DNA molecules that contain complementary unpaired nucleotides. Nuclease proteins typically comprise a "binding domain" that mediates interaction of the protein with a nucleic acid substrate (in some cases also specifically binding to a target site) and a "cleavage domain" that catalyzes the cleavage of phosphodiester bonds within the nucleic acid backbone. In some embodiments, the nuclease protein is capable of binding and cleaving a nucleic acid molecule in a monomeric form, while in other embodiments, the nuclease protein must dimerize or otherwise cleave a target nucleic acid molecule. Binding and cleavage domains of naturally occurring nucleases, as well as mode binding and cleavage domains that can be fused to create nucleases, are well known to those of skill in the art. For example, a zinc finger or transcriptional activator-like element can be used as a

binding domain to specifically bind a desired target site and fused or conjugated to a cleavage domain, such as the cleavage domain of fokl, to create an engineered nuclease that cleaves the target site.

[00169] Non-limiting examples of an exonuclease include exonuclease I, exonuclease II, exonnuclease III, exonuclease IV, exonuclease V, exonuclease VII, exonuclease VIII, lambda exonuclease, Xrn1, mung bean nuclease, TREX2, exonuclease T, T7 exonuclease, strandase exonuclease, 3'-5' exophosphodiesterase, and Bal31 nuclease.

[00170] Without being bound by theory, the system can be thought of as comprising a guide nucleic acid that promotes R-loop formation by binding to target DNA and a recombination protein that promotes recombination between the target nucleic acids and donor nucleic acids.

[00171] In certain embodiments, the guide RNA and the recombination protein are effectively linked . In some embodiments, the linkage is covalent. In some embodiments, the linkage is noncovalent. In certain embodiments, the guide nucleic acid comprises and aptamer sequence and the recombination protein comprises or is joined to an aptamer binding domain.

[00172] In certain aspects the invention involves vectors, e.g. for delivering or introducing in a cell Cas and/or RNA capable of guiding Cas to a target locus (i.e. guide RNA), but also for propagating these components (e.g. in prokaryotic cells). A used herein, a "vector" is a tool that allows or facilitates the transfer of an entity from one environment to another. It is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements. In general, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art.

[00173] One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated

viruses (AAVs)). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[00174] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). With regards to recombination and cloning methods, mention is made of U.S. patent application Ser. No. 10/815,730, published Sep. 2, 2004 as US 2004-0171156 A1, the contents of which are herein incorporated by reference in their entirety.

[00175] The vector(s) can include the regulatory element(s), e.g., promoter(s). The vector(s) can comprise Cas encoding sequences, and/or a single, but possibly also can comprise at least 3 or 8 or 16 or 32 or 48 or 50 guide RNA(s) (e.g., sgRNAs) encoding sequences, such as 1-2, 1-3, 1-4 1-5, 3-6, 3-7, 3-8, 3-9, 3-10, 3-8, 3-16, 3-30, 3-32, 3-48, 3-50 RNA(s) (e.g., sgRNAs). In a single vector there can be a promoter for each RNA (e.g., sgRNA), advantageously when there are up to about 16 RNA(s) (e.g., sgRNAs); and, when a single vector provides for more than 16 RNA(s) (e.g., sgRNAs), one or more promoter(s) can drive expression of more than one of the RNA(s) (e.g., sgRNAs), e.g., when there are 32 RNA(s) (e.g., sgRNAs), each promoter can drive expression of two RNA(s) (e.g., sgRNAs), and when there are 48 RNA(s) (e.g., sgRNAs), each promoter can drive expression of three RNA(s) (e.g., sgRNAs). By simple arithmetic and well established cloning protocols and the teachings in this disclosure one skilled in the art can readily practice the invention as to the RNA(s) (e.g., sgRNA(s) for a suitable exemplary vector such as

AAV, and a suitable promoter such as the U6 promoter, e.g., U6-sgRNAs. For example, the packaging limit of AAV is ~4.7 kb. The length of a single U6-sgRNA (plus restriction sites for cloning) is 361 bp. Therefore, the skilled person can readily fit about 12-16, e.g., 13 U6-sgRNA cassettes in a single vector. This can be assembled by any suitable means, such as a golden gate strategy used for TALE assembly (www.genome-engineering.org/taleffectors).

[00176] The skilled person can also use a tandem guide strategy to increase the number of U6-sgRNAs by approximately 1.5 times, e.g., to increase from 12-16, e.g., 13 to approximately 18-24, e.g., about 19 U6-sgRNAs. Therefore, one skilled in the art can readily reach approximately 18-24, e.g., about 19 promoter-RNAs, e.g., U6-sgRNAs in a single vector, e.g., an AAV vector. A further means for increasing the number of promoters and RNAs, e.g., sgRNA(s) in a vector is to use a single promoter (e.g., U6) to express an array of RNAs, e.g., sgRNAs separated by cleavable sequences.

[00177] A further means for increasing the number of promoter-RNAs, e.g., sgRNAs in a vector, is to express an array of promoter-RNAs, e.g., sgRNAs separated by cleavable sequences in the intron of a coding sequence or gene; and, in this instance it is advantageous to use a polymerase II promoter, which can have increased expression and enable the transcription of long RNA in a tissue specific manner. (see, e.g., Chung et al., Polycistronic RNA polymerase II expression vectors for RNA interference based on BIC/miR-155. Nucleic Acids Res. 2006, 34(7):e53. doi: 10.1093/nar/gkl143. and nature.com/mt/journal/v16/n9/abs/mt2008144a.html). In an embodiment, AAV may package U6 tandem sgRNA targeting up to about 50 genes. Accordingly, from the knowledge in the art and the teachings in this disclosure the skilled person can readily make and use vector(s), e.g., a single vector, expressing multiple RNAs or guides or sgRNAs under the control or operatively or functionally linked to one or more promoters-especially as to the numbers of RNAs or guides or sgRNAs discussed herein, without any undue experimentation.

[00178] The guide RNA(s), e.g., sgRNA(s) encoding sequences and/or Cas encoding sequences, can be functionally or operatively linked to regulatory element(s) and hence the regulatory element(s) drive expression. The promoter(s) can be constitutive promoter(s) and/or conditional promoter(s) and/or inducible promoter(s) and/or tissue specific promoter(s). The promoter can be selected from the group consisting of RNA polymerases, pol I, pol II, pol III, T7, U6, H1, retroviral Rous sarcoma virus (RSV) LTR promoter, the cytomegalovirus (CMV)

promoter, the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter. An advantageous promoter is the promoter is U6.

[00179] In some embodiments, one or more vectors driving expression of one or more elements of a CRISPR system are introduced into a host cell such that expression of the elements of the CRISPR system direct formation of a CRISPR complex at one or more target sites. For example, a Cas enzyme, a guide sequence linked to a tracr-mate sequence, and a tracr sequence could each be operably linked to separate regulatory elements on separate vectors. Or, RNA(s) of the CRISPR System can be delivered to a transgenic Cas9 animal or mammal, e.g., an animal or mammal that constitutively or inducibly or conditionally expresses Cas9; or an animal or mammal that is otherwise expressing Cas9 or has cells containing Cas9, such as by way of prior administration thereto of a vector or vectors that code for and express in vivo Cas9. 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 CRISPR system not included in the first vector.

[00180] CRISPR 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 CRISPR enzyme and one or more of the guide sequence, tracr mate sequence (optionally operably linked to the guide sequence), and a tracr sequence embedded within one or more intron sequences (e.g. each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the CRISPR enzyme, guide sequence, tracr mate sequence, and tracr sequence are operably linked to and expressed from the same promoter.

[00181] Delivery vehicles, vectors, particles, nanoparticles, formulations and components thereof for expression of one or more elements of a CRISPR system are as used in the foregoing documents, such as WO 2014/093622 (PCT/US2013/074667). In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a "cloning site"). In some embodiments, one or more insertion sites (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or

downstream of one or more sequence elements of one or more vectors. In some embodiments, a vector comprises an insertion site upstream of a tracr mate sequence, and optionally downstream of a regulatory element operably linked to the tracr mate sequence, such that following insertion of a guide sequence into the insertion site and upon expression the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell.

[00182] In some embodiments, a vector comprises two or more insertion sites, each insertion site being located between two tracr mate sequences so as to allow insertion of a guide sequence at each site. In such an arrangement, the two or more guide sequences may comprise two or more copies of a single guide sequence, two or more different guide sequences, or combinations of these. When multiple different guide sequences are used, a single expression construct may be used to target CRISPR 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.

[00183] In some embodiments, a vector comprises a regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, such as a Cas protein. CRISPR enzyme or CRISPR enzyme mRNA or CRISPR guide RNA or RNA(s) can be delivered separately; and advantageously at least one of these is delivered via a nanoparticle complex. CRISPR enzyme mRNA can be delivered prior to the guide RNA to give time for CRISPR enzyme to be expressed. CRISPR enzyme mRNA might be administered 1-12 hours (preferably around 2-6 hours) prior to the administration of guide RNA. Alternatively, CRISPR enzyme mRNA and guide RNA can be administered together. Advantageously, a second booster dose of guide RNA can be administered 1-12 hours (preferably around 2-6 hours) after the initial administration of CRISPR enzyme mRNA+guide RNA. Additional administrations of CRISPR enzyme mRNA and/or guide RNA might be useful to achieve the most efficient levels of genome modification.

[00184] In one aspect, the invention provides methods for using one or more elements of a CRISPR system. The CRISPR complex of the invention provides an effective means for modifying a target polynucleotide. The CRISPR complex of the invention has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target polynucleotide in a multiplicity of cell types. As such the CRISPR complex of the invention has a

broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis.

[00185] An exemplary CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within the target polynucleotide. The guide sequence is linked to a tracr mate sequence, which in turn hybridizes to a tracr sequence. In one embodiment, this invention provides a method of cleaving a target polynucleotide. The method comprises modifying a target polynucleotide using a CRISPR complex that binds to the target polynucleotide and effect cleavage of said target polynucleotide. Typically, the CRISPR complex of the invention, when introduced into a cell, creates a break (e.g., a single or a double strand break) in the genome sequence. For example, the method can be used to cleave a disease gene in a cell. The break created by the CRISPR complex can be repaired by a repair processes such as the error prone non-homologous end joining (NHEJ) pathway or the high fidelity homology-directed repair (HDR). During these repair process, an exogenous polynucleotide donor can be introduced into the genome sequence. In some methods, the HDR process is used modify genome sequence. For example, an exogenous polynucleotide donor comprising a sequence to be integrated flanked by an upstream sequence and a downstream sequence is introduced into a cell. The upstream and downstream sequences share sequence similarity with either side of the site of integration in the chromosome. Where desired, a donor polynucleotide can be DNA, e.g., a DNA plasmid, a bacterial artificial chromosome (BAC), a yeast artificial chromosome (YAC), a viral vector, a linear piece of DNA, a PCR fragment, a naked nucleic acid, or a nucleic acid complexed with a delivery vehicle such as a liposome or poloxamer.

[00186] The exogenous polynucleotide donor comprises a sequence to be integrated (e.g., a mutated gene). The sequence for integration may be a sequence endogenous or exogenous to the cell. Examples of a sequence to be integrated include polynucleotides encoding a protein or a non-coding RNA (e.g., a microRNA). Thus, the sequence for integration may be operably linked to an appropriate control sequence or sequences. Alternatively, the sequence to be integrated may provide a regulatory function. The upstream and downstream sequences in the exogenous polynucleotide donor are selected to promote recombination between the chromosomal sequence of interest and the donor polynucleotide. The upstream sequence is a nucleic acid sequence that shares sequence similarity with the genome sequence upstream of the targeted site for integration. Similarly, the downstream sequence is a nucleic acid sequence that shares sequence similarity with

the chromosomal sequence downstream of the targeted site of integration. The upstream and downstream sequences in the exogenous polynucleotide donor can have 75%, 80%, 85%, 90%, 95%, or 100% sequence identity with the targeted genome sequence. Preferably, the upstream and downstream sequences in the exogenous polynucleotide donor have about 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with the targeted genome sequence.

[00187] In some methods, the upstream and downstream sequences in the exogenous polynucleotide donor have about 99% or 100% sequence identity with the targeted genome sequence. An upstream or downstream sequence may comprise from about 20 bp to about 2500 bp, for example, about 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1100, 1200, 1300, 1400, 1500, 1600, 1700, 1800, 1900, 2000, 2100, 2200, 2300, 2400, or 2500 bp. In some methods, the exemplary upstream or downstream sequence have about 200 bp to about 2000 bp, about 600 bp to about 1000 bp, or more particularly about 700 bp to about 1000 bp. In some methods, the exogenous polynucleotide donor may further comprise a marker. Such a marker may make it easy to screen for targeted integrations. Examples of suitable markers include restriction sites, fluorescent proteins, or selectable markers. The exogenous polynucleotide donor of the invention can be constructed using recombinant techniques (see, for example, Sambrook et al., 2001 and Ausubel et al., 1996).

[00188] In a method for modifying a target polynucleotide by integrating an exogenous polynucleotide donor, a double stranded break is introduced into the genome sequence by the CRISPR complex, the break is repaired via homologous recombination an exogenous polynucleotide donor such that the donor is integrated into the genome. The presence of a double-stranded break facilitates integration of the donor. In other embodiments, this invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. The method comprises increasing or decreasing expression of a target polynucleotide by using a CRISPR complex that binds to the polynucleotide. In some methods, a target polynucleotide can be inactivated to effect the modification of the expression in a cell. For example, upon the binding of a CRISPR complex to a target sequence in a cell, the target polynucleotide is inactivated such that the sequence is not transcribed, the coded protein is not produced, or the sequence does not function as the wild-type sequence does. For example, a protein or microRNA coding sequence may be inactivated such that the protein or microRNA or pre-microRNA transcript is not produced. In some methods, a control sequence can be inactivated such that it no longer functions as a control sequence. As used herein,

"control sequence" refers to any nucleic acid sequence that effects the transcription, translation, or accessibility of a nucleic acid sequence. Examples of a control sequence include, a promoter, a transcription terminator, and an enhancer are control sequences.

[00189] The target polynucleotide of a CRISPR complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA). Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide.

[00190] A "disease-associated" gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non-disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at a normal or abnormal level. The target polynucleotide of a CRISPR complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA).

[00191] The target polynucleotide of a CRISPR complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA). The target can be a control element or a regulatory element or a promoter or an enhancer or a silencer. The promoter may, in some embodiments, be in the region of +200 bp or even +1000 bp from the TTS. In some embodiments, the regulatory region may be an enhancer.

The enhancer is typically more than +1000 bp from the TTS. More in particular, expression of eukaryotic protein-coding genes generally is regulated through multiple cis-acting transcription-control regions. Some control elements are located close to the start site (promoter-proximal elements), whereas others lie more distant (enhancers and silencers) Promoters determine the site of transcription initiation and direct binding of RNA polymerase II.

[00192] Three types of promoter sequences have been identified in eukaryotic DNA. The TATA box, the most common, is prevalent in rapidly transcribed genes. Initiator promoters infrequently are found in some genes, and CpG islands are characteristic of transcribed genes. Promoter-proximal elements occur within ≈200 base pairs of the start site. Several such elements, containing up to ≈20 base pairs, may help regulate a particular gene. Enhancers, which are usually ≈100-200 base pairs in length, contain multiple 8- to 20-bp control elements. They may be located from 200 base pairs to tens of kilobases upstream or downstream from a promoter, within an intron, or downstream from the final exon of a gene. Promoter-proximal elements and enhancers may be cell-type specific, functioning only in specific differentiated cell types. However, any of these regions can be the target sequence and are encompassed by the concept that the target can be a control element or a regulatory element or a promoter or an enhancer or a silencer.

[00193] Without wishing to be bound by theory, it is believed that the target sequence should be associated with a PAM (protospacer adjacent motif); that is, a short sequence recognized by the CRISPR complex. The precise sequence and length requirements for the PAM differ depending on the CRISPR enzyme used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence). Examples of PAM sequences are given in the examples section below, and the skilled person will be able to identify further PAM sequences for use with a given CRISPR enzyme. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence.

[00194] In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such that said binding results in increased or

decreased expression of said polynucleotide; wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. Similar considerations and conditions apply as above for methods of modifying a target polynucleotide. In fact, these sampling, culturing and re-introduction options apply across the aspects of the present invention. In one aspect, the invention provides for methods of modifying a target polynucleotide in a eukaryotic cell, which may be in vivo, ex vivo or in vitro. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal, and modifying the cell or cells. Culturing may occur at any stage ex vivo. The cell or cells may even be re-introduced into the non-human animal or plant. For re-introduced cells it is particularly preferred that the cells are stem cells.

[00195] In any aspect of the invention, the CRISPR complex may comprise a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence, wherein said guide sequence may be linked to a tracr mate sequence which in turn may hybridize to a tracr sequence.

[00196] The invention relates to the engineering and optimization of systems, methods and compositions used for the control of gene expression involving sequence targeting, such as genome perturbation or gene-editing, that relate to the CRISPR-Cas system and components thereof. In some embodiments, the Cas enzyme is Cas9. An advantage of the present methods is that the CRISPR system minimizes or avoids off-target binding and its resulting side effects. This is achieved using systems arranged to have a high degree of sequence specificity for the target DNA.

[00197] In an embodiment, nucleic acid molecule(s) encoding CRISPR-Cas9 or an ortholog or homolog thereof, may be codon-optimized for expression in a eukaryotic cell. A eukaryote can be as herein discussed. Nucleic acid molecule(s) can be engineered or non-naturally occurring.

[00198] In an embodiment, the CRISPR-Cas9 effector protein may comprise one or more mutations. The mutations may be artificially introduced mutations and may include but are not limited to one or more mutations in a catalytic domain, to provide a nickase, for example. Examples of catalytic domains with reference to a Cas enzyme may include but are not limited to RuvC I, RuvC II, RuvC III, and HNH domains.

[00199] In an embodiment, the CRISPR-Cas9 effector protein may be used as a generic nucleic acid binding protein with fusion to or being operably linked to a functional domain. Exemplary functional domains may include but are not limited to translational initiator, translational activator,

translational repressor, nucleases, in particular ribonucleases, a spliceosome, beads, a light inducible/controllable domain or a chemically inducible/controllable domain.

[00200] In some embodiments, the CRISPR-Cas9 effector protein may have cleavage activity. In some embodiments, the Cas9 effector protein may direct cleavage of one or both nucleic acid strands at the location of or near a target sequence, such as within the target sequence and/or within the complement of the target sequence or at sequences associated with the target sequence. In some embodiments, the Cas9 effector protein may direct cleavage of one or both DNA or RNA strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence.

[00201] In some embodiments, the cleavage may be blunt, i.e., generating blunt ends. In some embodiments, the cleavage may be staggered, i.e., generating sticky ends. In some embodiments, the cleavage may be a staggered cut with a 5′ overhang, e.g., a 5′ overhang of 1 to 5 nucleotides. In some embodiments, the cleavage may be a staggered cut with a 3′ overhang, e.g., a 3′ overhang of 1 to 5 nucleotides. In some embodiments, a vector encodes a nucleic acid-targeting Cas protein that may be mutated with respect to a corresponding wild-type enzyme such that the mutated nucleic acid-targeting Cas protein lacks the ability to cleave one or both DNA or RNA strands of a target polynucleotide containing a target sequence. As a further example, two or more catalytic domains of Cas (RuvC I, RuvC II, and RuvC III or the HNH domain) may be mutated to produce a mutated Cas substantially lacking all RNA cleavage activity. As described herein, corresponding catalytic domains of a Cas9 effector protein may also be mutated to produce a mutated Cas9 lacking all DNA cleavage activity or having substantially reduced DNA cleavage activity.

[00202] In some embodiments, a nucleic acid-targeting effector protein may be considered to substantially lack all RNA cleavage activity when the RNA cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the nucleic acid cleavage activity of the non-mutated form of the enzyme; an example can be when the nucleic acid cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form. An effector protein may be identified with reference to the general class of enzymes that share homology to the biggest nuclease with multiple nuclease domains from the Type II CRISPR system. Most preferably, the effector protein is a Type II protein such as Cas9. By derived, Applicants mean that the derived enzyme is largely based, in the sense of having a high degree of

sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as known in the art or as described herein.

[00203] In certain embodiments, Cas9 may be constitutively present or inducibly present or conditionally present or administered or delivered. Cas9 optimization may be used to enhance function or to develop new functions, one can generate chimeric Cas proteins. And Cas may be used as a generic nucleic acid binding protein.

[00204] Typically, in the context of an endogenous nucleic acid-targeting system, formation of a nucleic acid-targeting complex (comprising a guide RNA hybridized to a target sequence and complexed with one or more nucleic acid-targeting effector proteins) results in cleavage of one or both DNA or RNA strands in or near (e.g., within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. As used herein the term "sequence(s) associated with a target locus of interest" refers to sequences near the vicinity of the target sequence (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from the target sequence, wherein the target sequence is comprised within a target locus of interest).

[00205] An example of a codon optimized sequence, is in this instance a sequence optimized for expression in a eukaryote, e.g., humans (i.e. being optimized for expression in humans), or for another eukaryote, animal or mammal as herein discussed; see, e.g., SaCas9 human codon optimized sequence in WO 2014/093622 (PCT/US2013/074667) 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., Cas9) 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-targeting Cas protein is codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In some embodiments, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded.

[00206] 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.

[00207] 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.

[00208] In one aspect, the invention provides methods for using one or more elements of a nucleic acid-targeting system. The nucleic acid-targeting complex of the invention provides an effective means for modifying a target DNA (double stranded, linear or super-coiled). The nucleic acid-targeting complex of the invention has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target DNA in a multiplicity of cell types. As such the nucleic acid-targeting complex of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis. An exemplary nucleic acid-targeting complex comprises a DNA targeting effector protein complexed with a guide RNA hybridized to a target sequence within the target locus of interest.

[00209] In some embodiments, the method may comprise allowing a nucleic acid-targeting complex to bind to the target DNA and thereby modifying the target DNA, wherein the nucleic

acid-targeting complex comprises a nucleic acid-targeting effector protein complexed with a guide RNA hybridized to a target sequence within said target DNA or RNA. In one aspect, the invention provides a method of modifying expression of DNA in a eukaryotic cell.

[00210] In some embodiments, the method comprises allowing a nucleic acid-targeting complex to bind to the DNA such that said binding results in increased or decreased expression of said DNA; wherein the nucleic acid-targeting complex comprises a nucleic acid-targeting effector protein complexed with a guide RNA. Similar considerations and conditions apply as above for methods of modifying a target DNA. In fact, these sampling, culturing and re-introduction options apply across the aspects of the present invention. In one aspect, the invention provides for methods of modifying a target DNA in a eukaryotic cell, which may be in vivo, ex vivo or in vitro. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal, and modifying the cell or cells. Culturing may occur at any stage ex vivo. The cell or cells may even be re-introduced into the non-human animal or plant. For re-introduced cells it is particularly preferred that the cells are stem cells.

[00211] The invention relates to the engineering and optimization of systems, methods and compositions used for the control of gene expression involving DNA or RNA sequence targeting, that relate to the nucleic acid-targeting system and components thereof. An advantage of the present methods is that the CRISPR system minimizes or avoids off-target binding and its resulting side effects. This is achieved using systems arranged to have a high degree of sequence specificity for the target DNA.

[00212] Through this disclosure and the knowledge in the art, TALEs, CRISPR-Cas system, specifically the novel CRISPR systems described herein, or components thereof or nucleic acid molecules thereof (including, for instance HDR donor) or nucleic acid molecules encoding or providing components thereof may be delivered by a delivery system herein described both generally and in detail.

[00213] Vector delivery, e.g., plasmid, viral delivery: The CRISPR enzyme, for instance a Cas9, and/or any of the present RNAs, for instance a guide RNA, can be delivered using any suitable vector, e.g., plasmid or viral vectors, such as adeno associated virus (AAV), lentivirus, adenovirus or other viral vector types, or combinations thereof. Cas9 and one or more guide RNAs can be packaged into one or more vectors, e.g., plasmid or viral vectors. In some embodiments, the vector, e.g., plasmid or viral vector is delivered to the tissue of interest by, for example, an intramuscular

injection, while other times the delivery is via intravenous, transdermal, intranasal, oral, mucosal, or other delivery methods. Such delivery may be either via a single dose, or multiple doses. One skilled in the art understands that the actual dosage to be delivered herein may vary greatly depending upon a variety of factors, such as the vector choice, the target cell, organism, or tissue, the general condition of the subject to be treated, the degree of transformation/modification sought, the administration route, the administration mode, the type of transformation/modification sought, etc.

[00214] 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.

[00215] 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 is at least about 1×10^6 particles (for example, about 1×10^6-1×10^{12} particles), at least about 1×10^7 particles, 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 at least about 1×10^0 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, no more than about 1×10^{13} particles, no more than about 1×10^{12} particles, no more than about 1×10^{11} particles, no more than about 1×10^{10} particles (e.g., no more than about 1×10^{9} articles). Thus, the dose may contain a single dose of adenoviral vector with, for example, about 1×10^{6} particle units (pu), about 2×10^{6} pu, about 4×10^{6} pu, about 1×10^{7} pu, about 2×10^{7} pu, about 4×10^{7} pu, about 1×10^{8} pu, about 2×10^{8} pu, about 4×10^{8} pu, about 1×10^{9} pu, about 2×10^{9} pu, about 4×10^{9} pu, about 1×10^{10} pu, about 2×10^{10} pu, about 4×10^{10} pu, about 1×10^{11} pu, about 2×10^{11} pu, about 4×10^{11} pu, about 1×10^{12} pu, about 2×10^{12} pu, or about 4×10^{12} pu of adenoviral vector. See, for example, the adenoviral vectors in U.S. Pat. No. 8,454,972 B2 to Nabel, et. al., granted on Jun. 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.

[00216] 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. Pat. No. 8,404,658 B2 to Hajjar, et al., granted on Mar. 26, 2013, at col. 27, lines 45-60.

[00217] In an embodiment herein the delivery is via a plasmid. In such plasmid compositions, the dosage should be a sufficient amount of plasmid to elicit a response. For instance, suitable quantities of plasmid DNA in plasmid compositions can be from about 0.1 to about 2 mg, or from about 1 μg to about 10 μg per 70 kg individual. Plasmids of the invention will generally comprise (i) a promoter; (ii) a sequence encoding a CRISPR enzyme, operably linked to said promoter; (iii) a selectable marker; (iv) an origin of replication; and (v) a transcription terminator downstream of

and operably linked to (ii). The plasmid can also encode the RNA components of a CRISPR complex, but one or more of these may instead be encoded on a different vector.

[00218] 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 20 g and from mice experiments one can scale up to a 70 kg individual.

[00219] In some embodiments the RNA molecules of the invention are delivered in liposome or lipofectin formulations and the like and can be prepared by methods well known to those skilled in the art. Such methods are described, for example, in U.S. Pat. Nos. 5,593,972, 5,589,466, and 5,580,859, which are herein incorporated by reference. Delivery systems aimed specifically at the enhanced and improved delivery of siRNA into mammalian cells have been developed, (see, for example, Shen et al FEBS Let. 2003, 539:111-114; Xia et al., Nat. Biotech. 2002, 20:1006-1010; Reich et al., Mol. Vision. 2003, 9: 210-216; Sorensen et al., J. Mol. Biol. 2003, 327: 761-766; Lewis et al., Nat. Gen. 2002, 32: 107-108 and Simeoni et al., NAR 2003, 31, 11: 2717-2724) and may be applied to the present invention. siRNA has recently been successfully used for inhibition of gene expression in primates (see for example. Tolentino et al., Retina 24(4):660 which may also be applied to the present invention.

[00220] Indeed, RNA delivery is a useful method of in vivo delivery. It is possible to deliver Cas9 and gRNA (and, for instance, HR repair donor) into cells using liposomes or nanoparticles. Thus delivery of the CRISPR enzyme, such as a Cas9 and/or delivery of the RNAs of the invention may be in RNA form and via microvesicles, liposomes or nanoparticles. For example, Cas9 mRNA and gRNA can be packaged into liposomal particles for delivery in vivo. Liposomal transfection reagents such as lipofectamine from Life Technologies and other reagents on the market can effectively deliver RNA molecules into the liver.

[00221] Means of delivery of RNA also envisioned include delivery of RNA via nanoparticles (Cho, S., Goldberg, M., Son, S., Xu, Q., Yang, F., Mei, Y., Bogatyrev, S., Langer, R. and Anderson, D., Lipid-like nanoparticles for small interfering RNA delivery to endothelial cells, Advanced Functional Materials, 19: 3112-3118, 2010) or exosomes (Schroeder, A., Levins, C., Cortez, C., Langer, R., and Anderson, D., Lipid-based nanotherapeutics for siRNA delivery, Journal of Internal Medicine, 267: 9-21, 2010, PMID: 20059641). Indeed, exosomes have been shown to be particularly useful in delivery siRNA, a system with some parallels to the CRISPR

system. For instance, El-Andaloussi S, et al. ("Exosome-mediated delivery of siRNA in vitro and in vivo." Nat Protoc. 2012 December; 7(12):2112-26. doi: 10.1038/nprot.2012.131. Epub 2012 Nov. 15.) describe how exosomes are promising tools for drug delivery across different biological barriers and can be harnessed for delivery of siRNA in vitro and in vivo. Their approach is to generate targeted exosomes through transfection of an expression vector, comprising an exosomal protein fused with a peptide ligand. The exosomes are then 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 (α-tocopherol) may be conjugated with CRISPR Cas and delivered to the brain along with high density lipoprotein (HDL), for example in a similar manner as was done by Uno et al. (HUMAN GENE THERAPY 22:711-719 (June 2011)) for delivering short-interfering RNA (siRNA) to the brain. Mice were infused via Osmotic minipumps (model 1007D; Alzet, Cupertino, Calif.) filled with phosphate-buffered saline (PBS) or free TocsiBACE or Toc-siBACE/HDL and connected with Brain Infusion Kit 3 (Alzet). A brain-infusion cannula was placed about 0.5 mm posterior to the bregma at midline for infusion into the dorsal third ventricle. Uno et al. found that as little as 3 nmol of Toc-siRNA with HDL could induce a target reduction in comparable degree by the same ICV infusion method. A similar dosage of CRISPR Cas conjugated to α-tocopherol and co-administered with HDL targeted to the brain may be contemplated for humans in the present invention, for example, about 3 nmol to about 3 μmol of CRISPR Cas targeted to the brain may be contemplated. Zou et al. ((HUMAN GENE THERAPY 22:465-475 (April 2011)) describes a method of lentiviral-mediated delivery of short-hairpin RNAs targeting PKCγ for in vivo gene silencing in the spinal cord of rats. Zou et al. administered about 10 μl of a recombinant lentivirus having a titer of 1×10^9 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.

[00222] Cas9 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, U.S. Pat. No. 8,454,972 (formulations, doses for adenovirus), U.S. Pat. No. 8,404,658 (formulations, doses for AAV) and U.S. Pat. No. 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 U.S. Pat. No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in U.S. Pat. No. 8,404,658 and as in clinical trials involving adenovirus.

[00223] For plasmid delivery, the route of administration, formulation and dose can be as in U.S. Pat. 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 Cas9 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.

[00224] 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). Low probability of causing insertional mutagenesis because it doesn't integrate into the host genome.

[00225] As to AAV, the AAV can be AAV1, AAV2, AAV5 or any combination thereof. One can select the AAV with regard to tropism toward 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. Non-limiting examples of the AAV types that can be used in combination or alone include: AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAV11, AAV12, AAV13, AAV-rh10, AAV-DJ, AAV-DJ/8, AAV-PHP.eB, AAV2-retro, AAV2-QuadYF, AAV2.7m8, AAV2/1, AAV2/5, AAV2/6, AAV2/7, AAV2/8, AAV2/9, AAV PHP.S, AAV2/6.2, AAV2/Rh8. AAV2/rh10, AAV2/rh32.33, AAV2/hu11, AAV2/7m8, ShH10, Anc80L65, AAV2/AAV.CAP-B10.

[00226] Some AAV tissue tropisms are summarized in the following table.

AAV Serotype	Tissue tropism							
	CNS	Retina	Lung	Liver	Pancreas	Kidney	Heart	Muscle
AAV1	x	x			x		x	x
AAV2		x		x		x		
AAV3		x	x	x			x	
AAV4	x	x					x	
AAV5	x	x	x		x			
AAV6	x		x	x			x	x
AAV7				x				x
AAV8		x		x	x			x
AAV9	x		x	x			x	x
AAV-DJ		x	x	x		x		
AAV-DJ/8		x		x				x
AAV-Rh10	x		x	x			x	x
AAV-retro	x	x						x
AAV-PHP.B	x						x	x
AAV8-PHP.eB	x							x
AAV-PHP.S	x						x	x

[00227] The RNAs may be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other viral vector types, or combinations thereof. The RNAs can be packaged into one or more viral vectors. In some embodiments, the viral vector is delivered to the tissue of interest by, for example, an intramuscular injection, while other times the viral 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 chose, 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.

[00228] 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. Such a dosage formulation is readily ascertainable by one skilled 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.

[00229] 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 is at least about 1×10^6 particles (for example, about 1×10^6-1×10^{12} particles), at least about 1×10^{10} particles, 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 at least about 1×10^0 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, no more than about 1×10^{13} particles, no more than about 1×10^{12} particles, no more than about 1×10^{11} particles, and no more than about 1×10^{10} particles (e.g., no more than about 1×10^9 articles). Thus, the dose may contain a single dose of adenoviral vector with, for example, about 1×10^6 particle units (pu), about 2×10^6 pu, about 4×10^6 pu, about 1×10^7 pu, about 2×10^7 pu, about 4×10^7 pu, about 1×10^8 pu, about 2×10^8 pu, about

4×10^8 pu, about 1×10^9 pu, about 2×10^9 pu, about 4×10^9 pu, about 1×10^{10} pu, about 2×10^{10} pu, about 4×10^{10} pu, about 1×10^{11} pu, about 2×10^{11} pu, about 4×10^{11} pu, about 1×10^{12} pu, about 2×10^{12} pu, or about 4×10^{12} pu of adenoviral vector. See, for example, the adenoviral vectors in U.S. Pat. No. 8,454,972 B2 to Nabel, et. al., granted on Jun. 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.

[00230] 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. Pat. No. 8,404,658 B2 to Hajjar, et al., granted on Mar. 26, 2013, at col. 27, lines 45-60.

[00231] 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. Mice used in experiments are about 20 g. From that which is administered to a 20 g mouse, one can extrapolate to a 70 kg individual.

[00232] 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.

[00233] 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.

[00234] 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 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.

[00235] Particle characterization (including e.g., characterizing morphology, dimension, etc.) is done using a variety of different techniques. Common techniques are electron microscopy (TEM, SEM), atomic force microscopy (AFM), dynamic light scattering (DLS), X-ray photoelectron spectroscopy (XPS), powder X-ray diffraction (XRD), Fourier transform infrared spectroscopy (FTIR), matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF), ultraviolet-visible spectroscopy, dual polarisation interferometry and nuclear magnetic resonance (NMR). Characterization (dimension measurements) may be made as to native particles (i.e., preloading) or after loading of the cargo (herein cargo refers to one or more RNAs and/or vectors encoding the same, and may include additional components, 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 embodiments, particle dimension (e.g., diameter) characterization is based on measurements using dynamic laser scattering (DLS).

[00236] 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.

[00237] In terms of this invention, it is envisioned to have one or more components of the system delivered using nanoparticles or lipid envelopes. CRISPR enzyme mRNA and guide RNA may be delivered simultaneously using nanoparticles or lipid envelopes. Other delivery systems or vectors may be used in conjunction with the nanoparticle aspects of the invention.

[00238] In general, a "nanoparticle" refers to any particle having a diameter of less than 1000 nm. In certain embodiments, nanoparticles of the invention have a greatest dimension (e.g., diameter) of 500 nm or less. In other embodiments, nanoparticles of the invention have a greatest dimension ranging between 25 nm and 200 nm. In other embodiments, nanoparticles of the invention have a greatest dimension of 100 nm or less. In other embodiments, nanoparticles of the invention have a greatest dimension ranging between 35 nm and 60 nm.

[00239] 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.

[00240] 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.

[00241] For example, Su X, Fricke J, Kavanagh D G, Irvine D J ("In vitro and in vivo mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles" Mol Pharm. 2011 Jun. 6;

8(3):774-87. doi: 10.1021/mp100390w. Epub 2011 Apr. 1) describes biodegradable core-shell structured nanoparticles with a poly(β-amino ester) (PBAE) core enveloped by a phospholipid bilayer shell. These were developed for in vivo mRNA delivery. The pH-responsive PBAE component was chosen to promote endosome disruption, while the lipid surface layer was selected to minimize toxicity of the polycation core. Such are, therefore, envisioned for delivering RNA of the present invention.

[00242] In one embodiment, 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 deliver 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.

[00243] In one embodiment, nanoparticles that can deliver RNA to a cancer cell to stop tumor growth developed by Dan Anderson's lab at MIT may be used/and or adapted to the CRISPR Cas system of the present invention. In particular, the Anderson lab developed fully automated, combinatorial systems for the synthesis, purification, characterization, and formulation of new biomaterials and nanoformulations. See, e.g., Alabi et al., Proc Natl Acad Sci USA. 2013 Aug. 6; 110(32):12881-6; Zhang et al., Adv Mater. 2013 Sep. 6; 25(33):4641-5; Jiang et al., Nano Lett. 2013 Mar. 13; 13(3):1059-64; Karagiannis et al., ACS Nano. 2012 Oct. 23; 6(10):8484-7; Whitehead et al., ACS Nano. 2012 Aug. 28; 6(8):6922-9 and Lee et al., Nat Nanotechnol. 2012 Jun. 3; 7(6):389-93.

[00244] US patent application 20110293703 relates to lipidoid compounds are also particularly useful in the administration of polynucleotides, which may be applied to deliver the CRISPR Cas system of the present invention. In one aspect, the aminoalcohol lipidoid compounds are combined

with an agent to be delivered to a cell or a subject to form microparticles, nanoparticles, liposomes, or micelles. The agent to be delivered by the particles, liposomes, or micelles may be in the form of a gas, liquid, or solid, and the agent may be a polynucleotide, protein, peptide, or small molecule. The minoalcohol 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.

[00245] US Patent Publication No. 0110293703 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.

[00246] 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.

[00247] US Patent Publication No. 0110293703 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.

[00248] 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 system of the present invention.

[00249] In another embodiment, lipid nanoparticles (LNPs) are contemplated. In particular, an antitransthyretin small interfering RNA encapsulated in lipid nanoparticles (see, e.g., Coelho et al., N Engl J Med 2013; 369:819-29) may be applied to the 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. Lipids include, but are not limited to, DLin-KC2-DMA4, C12-200 and colipids

disteroylphosphatidyl choline, cholesterol, and PEG-DMG may be formulated RNA instead of siRNA (see, e.g., Novobrantseva, Molecular Therapy—Nucleic Acids (2012) 1, e4; doi:10.1038/mtna.2011.3) using a spontaneous vesicle formation procedure. The component molar ratio may be about 50/10/38.5/1.5 (DLin-KC2-DMA or C12-200/disteroylphosphatidyl choline/cholesterol/PEG-DMG). The final lipid:siRNA weight ratio may be ~12:1 and 9:1 in the case of DLin-KC2-DMA and C12-200 lipid nanoparticles (LNPs), respectively. The formulations may have mean particle diameters of ~80 nm with >90% entrapment efficiency. A 3 mg/kg dose may be contemplated.

[00250] 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 CRISPR Cas to the liver. A dosage of about four doses of 6 mg/kg of the LNP (or RNA of the CRISPR-Cas) 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.

[00251] 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, December 2011). Negatively charged polymers such as siRNA oligonucleotides may be loaded into LNPs at low pH values (e.g., pH 4) where the ionizable lipids display a positive charge. However, at physiological pH values, the LNPs exhibit a low surface charge compatible with longer circulation times. Four species of ionizable cationic lipids have been focused upon, namely 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-

dilinoleyloxy-keto-N,N-dimethyl-3-aminopropane (DLinKDMA), and 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA). It has been shown that LNP siRNA systems containing these lipids exhibit remarkably different gene silencing properties in hepatocytes in vivo, with potencies varying according to the series DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP employing a Factor VII gene silencing model (see, e.g., Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, December 2011). A dosage of 1 μg/ml levels may be contemplated, especially for a formulation containing DLinKC2-DMA. 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, December 2011). The cationic lipids 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA), (3-o-[2″-(methoxypolyethyleneglycol 2000) succinoyl]-1,2-dimyristoyl-sn-glycol (PEG-S-DMG), and R-3-[(w-methoxy-poly(ethylene glycol)2000) carbamoyl]-1,2-dimyristyloxlpropyl-3-amine (PEG-C-DOMG) may be provided by Tekmira Pharmaceuticals (Vancouver, Canada) or synthesized. Cholesterol may be purchased from Sigma (St Louis, Mo.). The specific CRISPR Cas RNA may be encapsulated in LNPs containing DLinDAP, DLinDMA, DLinK-DMA, and DLinKC2-DMA (cationic lipid:DSPC:CHOL: PEGS-DMG or PEG-C-DOMG at 40:10:40:10 molar ratios). When required, 0.2% SP-DiOC18 (Invitrogen, Burlington, Canada) may be incorporated to assess cellular uptake, intracellular delivery, and biodistribution. Encapsulation may be performed by dissolving lipid mixtures comprised of cationic lipid:DSPC:cholesterol:PEG-c-DOMG (40:10:40:10 molar ratio) in ethanol to a final lipid concentration of 10 mmol/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).

[00252] 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.

[00253] 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, Calif.). The particle size for all three LNP systems may be ~70 nm in diameter. siRNA encapsulation efficiency may be determined by removal of free siRNA 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. siRNA to lipid ratio was determined by measurement of cholesterol content in vesicles using the Cholesterol E enzymatic assay from Wako Chemicals USA (Richmond, Va.). PEGylated liposomes (or LNPs) can also be used for delivery.

[00254] Preparation of large LNPs may be used/and or adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, December 2011. A lipid premix solution (20.4 mg/ml total lipid concentration) may be prepared in ethanol containing DLinKC2-DMA, DSPC, and cholesterol at 50:10:38.5 molar ratios. Sodium acetate may be added to the lipid premix at a molar ratio of 0.75:1 (sodium acetate:DLinKC2-DMA). The lipids may be subsequently hydrated by combining the mixture with 1.85 volumes of citrate buffer (10 mmol/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 siRNA 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.

[00255] Delivery or administration according to the invention can be performed with liposomes. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic

phospholipid bilayer. Liposomes have gained considerable attention as drug delivery carriers because they are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB) (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[00256] Liposomes can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes as drug carriers. Although liposome formation is spontaneous when a lipid film is mixed with an aqueous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[00257] Several other additives may be added to liposomes in order to modify their structure and properties. For instance, either cholesterol or sphingomyelin may be added to the liposomal mixture in order to help stabilize the liposomal structure and to prevent the leakage of the liposomal inner cargo. Further, liposomes are prepared from hydrogenated egg phosphatidylcholine or egg phosphatidylcholine, cholesterol, and dicetyl phosphate, and their mean vesicle sizes were adjusted to about 50 and 100 nm. (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[00258] Conventional liposome formulation is mainly comprised of natural phospholipids and lipids such as 1,2-distearoryl-sn-glycero-3-phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines and monosialoganglioside. Since this formulation is made up of phospholipids only, liposomal formulations have encountered many challenges, one of the ones being the instability in plasma. Several attempts to overcome these challenges have been made, specifically in the manipulation of the lipid membrane. One of these attempts focused on the manipulation of cholesterol. Addition of cholesterol to conventional formulations reduces rapid release of the encapsulated bioactive compound into the plasma or 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) increases the stability (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[00259] In an embodiment, Trojan Horse liposomes (also known as Molecular Trojan Horses) are desirable and protocols may be found at cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long. These particles allow delivery of a transgene to the entire brain after an intravascular injection. Without being bound by limitation, it is believed that neutral lipid particles with specific antibodies conjugated to surface allow crossing of the blood brain barrier via endocytosis. Applicant postulates utilizing Trojan Horse Liposomes to deliver the CRISPR family of nucleases to the brain via an intravascular injection, which would allow whole brain transgenic animals without the need for embryonic manipulation. About 1-5 g of nucleic acid molecule, e.g., DNA, RNA, may be contemplated for in vivo administration in liposomes.

[00260] In another embodiment, the system 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 abpit 1 or 2.5 mg/kg are also contemplated (see, e.g., Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006). The SNALP formulation may contain the lipids 3-N-[(wmethoxypoly(ethylene glycol) 2000) carbamoyl]-1,2-dimyristyloxy-propylamine (PEG-C-DMA), 1,2-dilinoleyloxy-N,N-dimethyl-3-aminopropane (DLinDMA), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC) and cholesterol, in a 2:40:10:48 molar percent ratio (see, e.g., Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006).

[00261] In another embodiment, stable nucleic-acid-lipid particles (SNALPs) have proven to be effective delivery molecules to highly vascularized HepG2-derived liver tumors but not in poorly vascularized HCT-116 derived liver tumors (see, e.g., Li, Gene Therapy (2012) 19, 775-780). The SNALP liposomes may be prepared by formulating D-Lin-DMA and PEG-C-DMA with distearoylphosphatidylcholine (DSPC), Cholesterol and siRNA using a 25:1 lipid/siRNA ratio and a 48/40/10/2 molar ratio of Cholesterol/D-Lin-DMA/DSPC/PEG-C-DMA. The resulted SNALP liposomes are about 80-100 nm in size.

[00262] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich, St Louis, Mo., USA), dipalmitoylphosphatidylcholine (Avanti Polar Lipids, Alabaster,

Ala., USA), 3-N-[(w-methoxy poly(ethylene glycol)2000)carbamoyl]-1,2-dimyrestyloxypropylamine, and cationic 1,2-dilinoleyloxy-3-N,Ndimethylaminopropane (see, e.g., Geisbert et al., Lancet 2010; 375: 1896-905). A dosage of about 2 mg/kg total CRISPR Cas per dose administered as, for example, a bolus intravenous infusion may be contemplated.

[00263] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC; Avanti Polar Lipids Inc.), PEG-cDMA, and 1,2-dilinoleyloxy-3-(N;N-dimethyl)aminopropane (DLinDMA) (see, e.g., Judge, J. Clin. Invest. 119:661-673 (2009)). Formulations used for in vivo studies may comprise a final lipid/RNA mass ratio of about 9:1.

[00264] Other cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA) may be utilized to encapsulate CRISPR Cas similar to SiRNA (see, e.g., Jayaraman, Angew. Chem. Int. Ed. 2012, 51, 8529-8533). A preformed vesicle with the following lipid composition may be contemplated: amino lipid, di stearoylphosphatidylcholine (DSPC), cholesterol and (R)-2,3-bis(octadecyloxy) propyl-1-(methoxy poly(ethylene glycol)2000)propylcarbamate (PEG-lipid) in the molar ratio 40/10/40/10, respectively, and a FVII siRNA/total lipid ratio of approximately 0.05 (w/w). To ensure a narrow particle size distribution in the range of 70-90 nm and a low polydispersity index of 0.11_0.04 (n=56), the particles may be extruded up to three times through 80 nm membranes prior to adding the CRISPR Cas 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.

[00265] In certain embodiments, delivery of editing systems or components comprises delivery of a ribonucleoprotein (RNP) complex. According to the invention, targeting nucleic acids, including but not limited to gRNAs, dgRNAs can be provided in complexes, such as without limitation, complexes comprising Cas9 or dCas9. In certain embodiments, an RNP complex comprises a guide nucleic acid and a Cas9 fusion protein, such as without limitation a complex comprising dCas9-SSAP . In certain embodiments, an RNP complex comprises a guide nucleic acid and a recombination protein, e.g., an SSAP or SSB, which may be adapted or modified to bind to the guide nucleic acid. In certain embodiments, the guide nucleic acid and the recombination protein or Cas9 fusion protein comprise binding elements that promote complex formation . In a non-limiting example, a recombination protein comprises an MCP domain and a

guide RNA comprises an MS2 aptamer, whereby binding of the MS2 aptamer to the MCP domain produces an RNP.

[00266] Upon introducing the systems described herein into a cell comprising a target genomic DNA sequence, the guide RNA sequence binds to the target genomic DNA sequence in the cell genome, the Cas protein associates with the guide RNA and may induce a double strand break or single strand nick in the target genomic DNA sequence and the aptamer recruits the microbial recombination proteins to the target genomic DNA sequence through the aptamer binding protein of the fusion protein, thereby altering the target genomic DNA sequence in the cell. When introducing the compositions, or vectors described herein into the cell, the nucleic acid molecule comprising a guide RNA sequence, the Cas9 protein, and the fusion protein are first expressed in the cell.

[00267] In some embodiments, the cell is in an organism or host, such that introducing the disclosed systems, compositions, vectors into the cell comprises administration to a subject. The method may comprise providing or administering to the subject, in vivo, or by transplantation of ex vivo treated cells, systems, compositions, vectors of the present system.

[00268] A "subject" may be human or non-human and may include, for example, animal strains or species used as "model systems" for research purposes, such a mouse model as described herein. Likewise, subject may include either adults or juveniles (e.g. , children). Moreover, subject may mean any living organism, preferably a mammal (e.g. , human or non-human) that may benefit from the administration of compositions contemplated herein. Examples of mammals include, but are not limited to, any member of the Mammalian class: humans, non-human primates such as chimpanzees, and other apes and monkey species; farm animals such as cattle, horses, sheep, goats, swine; domestic animals such as rabbits, dogs, and cats; laboratory animals including rodents, such as rats, mice and guinea pigs, and the like. Examples of non-mammals include, but are not limited to, birds, fish and the like. In some embodiments of the methods and compositions provided herein, the mammal is a human.

[00269] As used herein, the terms "providing", "administering," "introducing," are used interchangeably herein and refer to the placement of the systems of the invention into a subject by a method or route which results in at least partial localization of the system to a desired site. The systems can be administered by any appropriate route which results in delivery to a desired location in the subject.

[00270] In some embodiments, the systems and methods described herein may be used to correct one or more defects or mutations in a gene (referred to as "gene correction"). In such cases, the target genomic DNA sequence encodes a defective version of a gene, and the system further comprises a donor nucleic acid molecule which encodes a wild-type or corrected version of the gene. Thus, in other words, the target genomic DNA sequence is a "disease-associated" gene. The term "disease-associated gene," refers to any gene or polynucleotide whose gene products are expressed at an abnormal level or in an abnormal form in cells obtained from a disease-affected individual as compared with tissues or cells obtained from an individual not affected by the disease. A disease-associated gene may be expressed at an abnormally high level or at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene, the mutation or genetic variation of which is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease.

[00271] Examples of genes responsible for such "single gene" or "monogenic" diseases include, but are not limited to, adenosine deaminase, a-1 antitrypsin, cystic fibrosis transmembrane conductance regulator (CFTR), ~-hemoglobin (HBB), oculocutaneous albinism II (OCA2), Huntingtin (HTT), dystrophia myotonica-protein kinase (DMPK), low-density lipoprotein receptor (LDLR), apolipoprotein B (APOB), neurofibromin I (NFI), polycystic kidney disease I (PKD I), polycystic kidney disease 2 (PKD2), coagulation factor VIII (F8), dystrophin (DMD), phosphate-regulating endopeptidase homologue, X-linked (PHEX), methyl-CpG-binding protein 2 (MECP2), and ubiquitin-specific peptidase 9Y, Y-linked (USP9Y). Other single gene or monogenic diseases are known in the art and described in, e.g., Chial, H. Rare Genetic Disorders: Learning About Genetic Disease Through Gene Mapping, SNPs, and Microarray Data, Nature Education 1(1):192 (2008), incorporated herein by reference; Online Mendelian Inheritance in Man (OMIM); and the Human Gene Mutation Database (HGMD).

[00272] The invention provides knock-ins of large transgenes at therapeutically relevant loci in the human genome . In certain embodiments, the locus provides cell or tissue-specific expression. In certain embodiments, the invention comprises insertion of nucleic acids into the albumin (ALB) locus. The ALB locus provides for liver targeting in human hepatocytes, is highly expressed and in a liver-specific manner. In certain embodiments, the invention comprises insertion of nucleic acids into the AAVSI locus. The AAVSI locus is a safe-harbor locus for gene therapy that is well

expressed in certain tissue types and can be used in a wide variety of treatments, with low expression in liver. US Patent Publication 2018/0214490 Al describes gene therapy for lysosomal storage diseases, including targeting transgenes to safe harbo" loci such as the AAVSI, HPRT and CCR5 genes in human cells, and Rosa26 in murine cells. US Patent 9267154 describes integration of exogenous nucleic acid sequences into the PPPIR12C locus, which is widely expressed in most tissues . describes cell-specific expression by targeting transgenes (e.g., encoding chimeric antigen receptors (CARs)) to the T-cell receptor a constant (TRAC) locus. Exemplary and non-limiting loci that can be targeted according to the invention include ALB, HSP90AA1, DYN2TI, ACRB, BCAP31, HIST1H2BK, CLTA, and RABIIA.

[00273] Exemplary, non-limiting guides for targeting ALB include the following:

sgRNA sg3 guide sequence	GTAAATATCTACTAAGACAA
sgRNA sg4 guide sequence	GCATCTTCAGGGAGTAGCTT
sgRNA scaffold with MS2 aptamer	gtttAagagctaggccAACATGAGGATCACCCATGTCTGCAGggcctagcAAG TTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGA GTCGGTGCGCGCACATGAGGATCACCCATGTGCT
sgRNA sg3 combined with scaffold	GTAAATATCTACTAAGACAAgtttAagagctaggccAACATGAGGATCACCCATGTCTG CAGggcctagcAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC GAGTCGGTGCGCGCACATGAGGATCACCCATGTGCT
sgRNA sg4 combined with scaffold	GCATCTTCAGGGAGTAGCTTgtttAagagctaggccAACATGAGGATCACCCATGTCTG CAGggcctagcAAGTTTAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACC GAGTCGGTGCGCGCACATGAGGATCACCCATGTGCT

[00274] An exemplary, non-limiting fusion protein comprising an MCP aptamer binding sequence, linker, and SSAP protein and encoding nucleotide sequence (RNA) is shown in the following table. In the example, the SSAP comprises RecT.

HA_MCP_SSAP-RecT	
AA sequence	MPAAKRVKLDYPYDVPDYAASNFTQFVLVDNGGTGDVTVAPSNFANGV AEWISSNSRSQAYKVTCSVRQSSAQKRKYTIKVEVPKVATQTVGGVELPV AAWRSYLNMELTIPIFATNSDCELIVKAMQGLLKDGNPIPSAIAANSGIYS ASGGSSGGSSGSETPGTSESATPESSGGSSGGSGGSMTKQPPIAKADLQKT QGNRAPAAVKNSDVISFINQPSMKEQLAAALPRHMTAERMIRIATTEIRK VPALGNCDTMSFVSAIVQCSQLGLEPGSALGHAYLLPFGNKNEKSGKKN VQLIIGYRGMIDLARRSGQIASLSARVVREGDEFSFEFGLDEKLIHRPGENE DAPVTHVYAVARLKDGGTQFEVMTRKQIELVRSLSKAGNNGPWVTHWE EMAKKTAIRRLFKYLPVSIEIQRAVSMDEKEPLTIDPADSSVLTGEYSVIDN SEE*
DNA sequence (ATG and TAA in upper case)	ATGcctgcggccaaaagagtgaagcttgattacccatacgatgttccagattacgctgcttcaaactttactcagtt cgtgctcgtggacaatggtgggacaggggatgtgacagtggctccttctaatttcgctaatggggtggcagagtgg atcagctccaactcacggagccaggcctacaaggtgacatgcagcgtcaggcagtctagtgcccagaagagaaa gtataccatcaaggtggaggtccccaaagtggctacccagacagtgggcggagtcgaactgcctgtcgccgcttgg aggtcctacctgaacatggagctcactatcccaattttcgctaccaattctgactgtgaactcatcgtgaaggcaatg caggggctcctcaaagacggtaatcctatcccttccgccatcgccgctaactcaggtatctacagcgctTCCGGA GGATCTAGCGGAGGCTCCTCTGGCTCTGAGACACCTGGCACAAGCGAGAGCGCAACAC CTGAAAGCAGCGGGGGCAGCAGCGGGGGGTCAggaggatccATGACAAAACAACCTCCA ATAGCAAAAGCTGATCTGCAGAAAACCCAGGGCAATAGAGCCCCTGCTGCTGTGAAGA ATAGTGACGTGATTAGCTTCATCAACCAGCCTAGCATGAAGGAGCAGCTGGCCGCCGC CCTGCCTCGCCACATGACCGCAGAGCGGATGATTCGGATCGCTACAACCGAGATCAGA AAAGTGCCCGCTCTGGGGAACTGTGATACCATGTCTTTCGTGAGCGCCATTGTGCAGTG TAGCCAGCTGGGCCTGGAGCCCGGAAGTGCCCTGGGCCACGCCTACCTGCTGCCCTTC GGCAACAAGAACGAGAAATCCGGTAAAAAGAACGTGCAGCTGATCATCGGTTATAGA GGTATGATCGATCTGGCTCGGCGGTCCGGCCAAATCGCCAGCCTGTCTGCCAGGGTGG TGAGAGAGGGCGACGAGTTCAGCTTTGAGTTTGGCCTGGACGAAAAGTTGATCCATCG GCCCGGAGAGAACGAGGACGCCCCTGTCACCCACGTGTATGCCGTGGCTAGACTGAAG GACGGCGGTACACAGTTCGAAGTTATGACCCGGAAGCAGATCGAACTGGTGAGGTCCC TGTCCAAGGCCGGCAATAACGGGCCCTGGGTGACCCACTGGGAAGAAATGGCCAAGA AAACCGCCATTAGACGCCTGTTCAAGTACCTGCCCGTGTCCATTGAGATTCAGAGGGCC GTGTCTATGGATGAGAAAGAGCCCCTGACCATCGACCCTGCTGACAGCAGCGTGCTGA CCGGCGAGTACTCTGTGATTGACAATTCTGAGGAGTAA

[00275] In some embodiments, validation of the genome edit is performed using the screening method Tracking of Indels by Decomposition (TIDE), which uses the trace file generated in Sanger sequencing and a simple algorithm to quantify editing efficiency by insertion and deletion frequency.

[00276] In some embodiments, the disclosed invention is utilized to deliver GLP-1 or a GLP-1 analog to provide therapy for type II diabetes, among others. A GLP-1 analog is defined as a molecule having a modification including one or more amino acid substitutions, deletions,

inversions, or additions when compared with GLP-1. GLP-1 analogs known in the art include, e.g., GLP-1(7-34) and GLP-1(7-35), GLP-1(7-36), Val^{8}-GLP-1(7-37), Gln^{9}-GLP-1(7-37), D-Gln^{9}-GLP-1(7-37), Thr^{16}-Lys^{18}-GLP-1(7-37), and Lys^{18}-GLP-1(7-37), disclosed in U.S. Pat. Nos. 5,118,666, 5,545,618, and 6,583,111. These compounds are the biologically processed forms of GLP-1 having insulinotropic properties. In some embodiments, the invention is utilized to deliver a long-acting GLP-1 analog, such as, but not limited to dulaglutide (also referred to herein as GLP1-IgG4Fc or hGLP1-Protracted) or liraglutide.

[00277] Glucagon-like peptide-1 (GLP-1), which is involved in insulin resistance, regulates insulin secretion in the pancreas. It is known that in type 2 diabetes there is an increased expression of dipeptidyl peptidase-4 (DPP4), which is known to decompose GLP-1, which in turn leads to increased insulin resistance. In practice, sitagliptin, a known DPP4 inhibitor, and the like are used as medicines for treating type 2 diabetes. However, when such DPP4 inhibitors are used in a formulation, the effect is temporary. Thus, these medicines must be taken every day, and the use thereof with respect to type 2 diabetes accompanying kidney disorders is limited because of side effects. In addition, various other side effects including allergic reactions caused by the medicine have been reported (Lenski M et al., "Effects of DPP-4 inhibition on cardiac metabolism and function in mice." J. Mol. Cell Cardiol., 51(6), 906-918, 2011).

[00278] The disclosed systems and methods overcome challenges encountered during conventional gene editing, including low efficiency and off-target events, particularly with kilobase-scale nucleic acids. In some embodiments, the disclosed systems and methods improve the efficiency of gene editing. For example, the disclosed systems and methods can have an increase in efficiency over conventional CRISPR-Cas9 systems and methods, as shown in Example 1 and Figure 17. In some embodiments, the improvement in efficiency is accompanied by a reduction in off-target events. Another aspect of increasing the overall accuracy of a gene editing system is reducing the on-target insertion-deletions (indels), a byproduct of HDR editing.

[00279] The invention further provides kits containing one or more reagents or other components useful, necessary, or sufficient for practicing any of the methods described herein. For example, kits may include CRISPRreagents (Cas protein, guide RNA, vectors, compositions, etc.), recombination reagents (recombination protein-aptamer binding protein fusion protein, the aptamer sequence, vectors, compositions, etc.) transfection or administration reagents, negative and positive control samples (e.g., cells, donor DNA), cells, containers housing one or more

components (e.g., microcentrifuge tubes, boxes), detectable labels, detection and analysis instruments, software, instructions, and the like.

[00280] Although the present invention and its advantages have been described in detail, it should be understood that various changes, substitutions and alterations can be made herein without departing from the spirit and scope of the invention as defined in the appended claims.

[00281] The present invention will be further illustrated in the following Examples which are given for illustration purposes only and are not intended to limit the invention in any way.

Examples

Dose Ranging Study of AAV GLP-1 Delivery in Mouse Model

[00282] Applicants evaluated the efficiency of the mRNA/AAV GLP-1 delivery system in animal models.

[00283] Applicants grouped the experiments into two milestones. Milestone 1 consisted of 6 groups of mice, consisting of 3 mice per group. The experimental design for Milestone 1 is outlined in Figure 1. The test article information for Milestone 1 is outlined in Figure 3.

[00284] Milestone 2 consisted of 3 groups of mice, consisting of 3 mice per group. The experimental design for Milestone 2 is outlined in Figure 2. The test article information for Milestone 2 is outlined in Figure 4.

[00285] Reagents used during experimentation are outlined in Figure 5 and instruments used during experimentation are outlined in Figure 6.

[00286] Animal and housing conditions were as follows:

[00287] Animal:

[00288] Species: Mus Musculus

[00289] Strain: C57BL/6J

[00290] Gender: Male

[00291] Age: 6-8 weeks

[00292] Body weight: 18-22g

[00293] Number of animals: 18 mice

[00294] Animal supplier: Biocytogen Biotechnology Co., Ltd

[00295] Animal Certificate No.: 202271194

[00296] Housing Conditions:

[00297] All mice were kept in SPF room in individually ventilated cages at constant temperature and humidity with 5 animals in each cage.

[00298] Temperature: 20-26°C

[00299] Humidity: 40-70%

[00300] Light cycle: 12 hours light, 12 hours dark.

[00301] Cages: Made of polycarbonate. The size was 350 mm x 160 mm x 120 mm. The bedding material was corncob, which was changed twice per week.

[00302] Diet: Animals had free access to food and water. Food was sterilized by irradiation and water and bedding were sterilized by high-pressure sterilization during the entire study period.

[00303] Cage identification: The detailed corresponding labels for each cage contained the following information: number of animals, gender, strain, receiving date, study number, group number, study stage, and study responsible person, etc.

[00304] Administration for Milestone 1:

[00305] Applicants calculated the detailed drug administration information according to Figure 7.

[00306] The LNP was stored at -80°C and thawed on ice.

[00307] The LNP was diluted according to the calculated dosage based on the information provided in Figure 7.

[00308] The diluted LNP was transferred on ice to the animal facilities.

[00309] The diluted LNP was injected intravenously into mice grouped according to body weight. Before injection, the LNP was placed at room temperature for 5 minutes and mixed.

[00310] The AAV was injected 8 hours after LNP injection according to the same procedure.

[00311] Administration for Milestone 2:

[00312] Applicants calculated the detailed drug administration information according to Figure 8.

[00313] The AAV was stored at -80°C and thawed on ice.

[00314] The LNP was stored at 4°C.

[00315] The AAV and LNP were diluted according to the calculated dosage based on information provided in Figure 8.

[00316] The diluted AAV and LNP were transferred on ice to the animal facilities.

[00317] The diluted AAV was injected intravenously into mice grouped according to Figure 8.

[00318] The diluted AAV was injected first, followed by the diluted LNP, 4 hours later.

[00319] Prior to injection, the diluted LNP was placed at room temperature for 5 minutes and mixed.

[00320] Animal blood collection:

[00321] 50μL of whole blood was collected from the retro-orbital venous plexus of mice through a 0.3 mm capillary. After standing at 4°C for 30 minutes, the blood was centrifuges at 7500 rpm for 10 min at 4°C. The upper layer of serum was carefully drawn and stored at -80°C.

[00322] Milestone 1: The mouse serum was collected in the first and second groups at 2 days, 7 days, and 14 days after administration.

[00323] All mice from group 6 of milestone 1 died after administration.

[00324] Milestone 2: The mouse serum was collected in group 3 at 2 days, 7 days, and 14 days after administration.

[00325] Animal liver collection:

[00326] On day 14 after administration, the mice were euthanized and livers were harvested. The livers of each mouse was divided into 3 EP tubes for quick freezing with liquid nitrogen and stored at -80°C.

[00327] Milestone 1: Livers were collected from the mice in groups 1, 3, 4, and 5.

[00328] Milestone 2: Livers were collected from the mice in groups 1, 2, and 3.

[00329] Blood GLP-1 measurement (ELISA, EXGLPHS-35K, Millipore):

[00330] The dilution factor of the test was carried out according to Figure 9 to ensure all samples were within detection range.

[00331] Observations:

[00332] All of the procedures related to animal handling, care, and treatment in this study were performed according to the guidelines approved by the Institutional Animal Care and Use Committee (IACUC) and following the guidance of the Association for Assessment and Accreditation of Laboratory Animal Care (AAALAC). At the time of routine monitoring, the animals were monitored daily for tumor growth and behavior, such as mobility and food and water consumption, as well as eye/hair matting and any other abnormal effect as stated in the protocol. Body weights were measured twice weekly. Death and observed clinical signs were recorded.

[00333] Results:

[00334] Applicants prepared the LNP for Milestone 1 according to Figure 10 and for Milestone 2 according to Figure 11.

[00335] After the Milestone 1 LNP was prepared, it was stored at -80°C for about 3 days, then thawed and diluted for animal experiments.

[00336] After the Milestone 2 LNP was prepared, it was temporarily stored at 4°C for no more than 24 hours and then directly used for animal experiments.

[00337] Clinical observation results:

[00338] There were no obvious abnormalities in the mice after AAV injection. After LNP injection, the mice generally appeared in poor condition and showed a positive correlation with the concentration gradient of the death rate. The living mice gradually recovered within 6 hours, and then behaved normally. The final state of the mice and the general LNP information is shown in Figures 12 and 13.

[00339] Serum GLP-1 concentrations:

[00340] The serum concentrations of GLP-1 at various time points are shown in Figures 14-16.

* * *

[00341] Having thus described in detail preferred embodiments of the present invention, it is to be understood that the invention defined by the above paragraphs is not to be limited to particular details set forth in the above description as many apparent variations thereof are possible without departing from the spirit or scope of the present invention.

WHAT IS CLAIMED IS:

1. A method of modifying an organism or a non-human organism by manipulation of a target sequence, which comprises delivering in a cell, an RNA guided-recombination system or composition comprising:

(i) a nucleic acid molecule comprising a guide RNA sequence that is complementary to a target DNA sequence; wherein the target DNA sequence comprises a genomic DNA sequence in the cell, and

(ii) a recombination protein, wherein the recombination protein comprises a nuclease, a single stranded DNA annealing protein (SSAP) or a combination thereof;

or,

(iii) nucleic acid molecule(s) encoding or delivering (i), and/or (ii) for expression *in vivo* in a cell;

or,

(iv) vector(s) containing the nucleic acid molecule(s) of (iii) for expression *in vivo* in a cell;

wherein the delivering comprises a delivery system comprising a viral vector system and/or a lipid-based delivery system.

2. The method of claim 1 further comprising a Cas protein.

3. The method of claim 1 or 2 further comprising a donor nucleic acid.

4. The method of claim 1 or 2 further comprising a donor nucleic acid, wherein the donor nucleic acid comprises a sequence that encodes GLP-1 or a GLP-1 analog.

5. The method of any one of claims 1 to 4, wherein the target DNA sequence comprises a genomic sequence of albumin (ALB), AAVS1, HSP90AA1, DYNLT1, ACTB, BCAP31, HIST1H2BK, CLTA, ROSA26, PCSK, or RAB11A.

6. The method of any one of claims 1 to 5, further comprising a recruitment system comprising at least one aptamer sequence; and

an aptamer binding protein functionally linked to the recombination protein as part of a fusion protein.

7. The method of claim 6, wherein the at least one aptamer sequence is an RNA aptamer sequence or a peptide aptamer sequence.

8. The method of any one of claims 6 or 7, wherein the nucleic acid molecule or nucleic acid molecules additionally comprise the at least one RNA aptamer sequence.

9. The method of claim 8, wherein the at least one RNA aptamer sequences comprise the same sequence.

10. The method of any one claims 6-9, wherein the aptamer binding protein comprises an MS2 coat protein, or a functional derivative or variant thereof.

11. The method of any one of claims 6 or 7, wherein the at least one peptide aptamer sequence is conjugated to the Cas protein.

12. The method of claim 11, wherein the at least one peptide aptamer sequence comprises between 1 and 24 peptide aptamer sequences.

13. The method of any one of claims 11 or 12, wherein the aptamer sequences comprise the same sequence.

14. The method of any one of claims 6-13, wherein the recombination protein N-terminus is linked to the aptamer binding protein C-terminus.

15. The method of any one of claims 1-14, wherein the recombination protein comprises a microbial recombination protein or active portion thereof.

16. The method of any one of claims 2-15, wherein the Cas protein is catalytically inactive (less than 5% nuclease activity as compared with a wild type or non-mutated Cas protein) or catalytically dead.

17. The method of any one of claims 2-16, wherein the Cas protein comprises a Cas9, Cas12a, or Cas12f.

18. The method of any of claims 2-17, wherein the single stranded DNA annealing protein and the Cas protein are functionally linked to each other and comprise a fusion protein.

19. The method of any of claims 1-18, wherein the RNA guided recombination system is delivered via a viral vector system and a lipid-based system.

20. The method of claim 19, wherein the viral vector system comprises an AAV viral vector system.

21. The method of claim 20, wherein the viral vector system comprises an AAV8 viral vector system.

22. The method of claim 19, wherein the lipid-based system comprises a lipid nanoparticle (LNP).

23. A system or composition for delivering into an organism or a non-human organism:

(i) a nucleic acid molecule comprising a guide RNA sequence that is complementary to a target DNA sequence; wherein the target DNA sequence comprises a genomic DNA sequence in the cell, and

(ii) a recombination protein, wherein the recombination protein comprises a nuclease, a single stranded DNA annealing protein (SSAP) or a combination thereof;

or,

(iii) nucleic acid molecule(s) encoding or delivering (i), and/or (ii) for expression *in vivo* in a cell;

or,

(iv) vector(s) containing the nucleic acid molecule(s) of (iii) for expression *in vivo* in a cell;

wherein the system or composition comprises a delivery system comprising a viral vector system and/or a lipid-based delivery system.

24. The system or composition of claim 23, which comprises a Cas protein.

25. The system or composition of any one of claims 23 or 24, which further comprises a donor nucleic acid.

26. The system or composition of any one of claims 23 or 24, which further comprises a donor nucleic acid, wherein the donor nucleic acid comprises a sequence that encodes GLP-1 or a GLP-1 analog.

27. The system or composition of any one of claims 23 to 26, wherein the target DNA sequence comprises a genomic sequence of albumin (ALB), AAVS1, HSP90AA1, DYNLT1, ACTB, BCAP31, HIST1H2BK, CLTA, ROSA26, PCSK, or RAB11A.

28. The system or composition of any one of claims 23 to 26, further comprising a recruitment system comprising at least one aptamer sequence; and
an aptamer binding protein functionally linked to the recombination protein as part of a fusion protein.

29. The system or composition of claim 28, wherein the at least one aptamer sequence is an RNA aptamer sequence or a peptide aptamer sequence.

30. The system or composition of any one of claims 28 or 29, wherein the nucleic acid molecule or nucleic acid molecules additionally comprise the at least one RNA aptamer sequence.

31. The system or composition of claim 30, wherein the at least one RNA aptamer sequences comprise the same sequence.

32. The system or composition of any one of claims 28-31, wherein the aptamer binding protein comprises an MS2 coat protein, or a functional derivative or variant thereof.

33. The system or composition of any one of claims 28 or 29, wherein the at least one peptide aptamer sequence is conjugated to the Cas protein.

34. The system or composition of claim 33, wherein the at least one peptide aptamer sequence comprises between 1 and 24 peptide aptamer sequences.

35. The system or composition of any one of claims 33 or 34, wherein the aptamer sequences comprise the same sequence.

36. The system or composition of any one of claims 28-35, wherein the recombination protein N-terminus is linked to the aptamer binding protein C-terminus.

37. The system or composition of any one of claims 23-36, wherein the recombination protein comprises a microbial recombination protein or active portion thereof.

38. The system or composition of any one of claims 24-37, wherein the Cas protein is catalytically inactive (less than 5% nuclease activity as compared with a wild type or non-mutated Cas protein) or catalytically dead.

39. The system or composition of any one of claims 24-38, wherein the Cas protein comprises a Cas9, Cas12a, or Cas12f.

40. The system or composition of any one of claims 24-39, wherein the single stranded DNA annealing protein and the Cas protein are functionally linked to each other and comprise a fusion protein.

41. The system or composition of any one of claims 23-40, wherein the RNA guided recombination system is delivered via a viral vector system and a lipid-based system.

42. The method of claim 41, wherein the viral vector system comprises an AAV viral vector system.

43. The method of claim 42, wherein the viral vector system comprises an AAV8 viral vector system.

44. The method of claim 41, wherein the lipid-based system comprises a lipid nanoparticle (LNP).

45. A kit comprising the system of any one of claims 23 to 44.

FIG. 1

Experimental design for Milestone 1						
G	**N**	**Treatment-AAV/LNP**	**Mice Strain**	**AAV/LNP Dose**	**Dose Route**	**Schedule**
1	3	Vehicle	C57BL/6J	—	*i.v.*	Single
2	3	Albumin_sg3_TBG-hGLP1-Protracted		4*10E11		
3	3	Cas9+sgRNA-3		0.5 mg/kg		
4	3	Cas9+sgRNA-3		1.5 mg/kg		
5	3	Cas9+sgRNA-3		4.5 mg/kg		
6	3	AlbG3-dCAS9-SSAP+ sgRNA-3 Albumin_sg3_SA-hGLP1-Protracted		1.5 mg/kg 4*10E11		

FIG. 2

Experimental design for Milestone 2						
G	**N**	**Treatment-AAV/LNP**	**Mice Strain**	**AAV/LNP Dose**	**Dose Route**	**Schedule**
1	3	Cas9+sgRNA-3	C57BL/6J	1.5 mg/kg	*i.v.*	Single
2	3	Cas9+sgRNA-3		3 mg/kg		
3	3	AlbG3-dCAS9-SSAP+ sgRNA-3 Albumin_sg3_SA-hGLP1-Protracted		1.5 mg/kg 4*10E11		

FIG. 3

Test Articles for Milestone 1					
Category	**Sample ID**	**Lot**	**Suppler**	**Conc. /Titer**	**Storage**
LNP	Cas9+sgRNA-3	NA	Probio	LNP: 5.5 mg/mL Total RNA: 177.5 µg/mL	-80°C
	AlbG3-dCAS9-SSAP+ sgRNA-3	NA		LNP: 5 mg/mL Total RNA: 178.0 µg/mL	
AAV	Albumin_sg3_TBG-hGLP1-Protracted	NA		4.03*E13	-80°C
	Albumin_sg3_SA-hGLP1-Protracted	NA		2.83*E13	

FIG. 4

Test Articles for Milestone 2					
Category	**Sample ID**	**Lot**	**Suppler**	**Conc. /Titer**	**Storage**
LNP	Cas9+sgRNA-3	NA	Probio	LNP: 5.4 mg/mL Total RNA: 111 µg/mL	4°C
	AlbG3-dCAS9-SSAP+ sgRNA-3	NA		LNP: 6.0 mg/mL Total RNA: 128 µg/mL	
AAV	Albumin_sg3_SA-hGLP1-Protracted	NA		2.83*E13	-80°C

FIG. 5

Reagents			
Name	**Manufacturer**	**Cat**	**Lot**
Tween 20	Sigma Aldrich	P1379-100ML	WXBD0981V
PBS	SOLARBIO	P1010-100	20220728
BSA	Sigma Aldrich	V900933-100G	WXBD8440V
High Sensitivity GLP-1 Active Chemiluminescent ELISA Kit 96-Well Plate	EMD Millipore	EZGLPHS-35K	3900909 3909605

FIG. 6

Instruments		
Name	**Model**	**Manufacturer**
Orbital microtiter plate shaker	MX100-4A	ALLSHENG
Pipette	1-10 μL, 10-100 μL, 20-200 μL, 100-1000 μL, 30-300 μL	Eppendorf
Luminometer plate reader	PHERAstar FSX	BMG Labtech

FIG. 7

Drug Administration Information for Milestone 1	
Group-1：Vehicle, 10μL/g	
/	LNP dilution buffer solvent (μL)
Theoretic quantity/volume	1200

Group-2: AAV(Albumin_sg3_TBG-hGLP1-Protracted), 20*10E11/mL, 200 μL/mice		
/	AAV (403*10E11/mL) (μL)	AAV dilution buffer solvent (μL)
Theoretic quantity/volume	60	1140

Group-3: LNP (Cas9+sgRNA-3), 50 μg/mL, 10 μL/g		
/	LNP (177.5 μg/mL) (μL)	LNP dilution buffer solvent (μL)
Theoretic quantity/volume	338	862

Group-4: LNP (Cas9+sgRNA-3), 150 μg/mL, 10 μL/g		
/	LNP (177.5 μg/mL) (μL)	LNP dilution buffer solvent (μL)
Theoretic quantity/volume	1017	183

Group-5: LNP (Cas9+sgRNA-3), 450 μg/mL, 25.4 μL/g	
/	LNP (177.5 μg/mL) (μL)
Theoretic quantity/volume	2000

Group-6: LNP (AlbG3-dCAS9-SSAP+ sgRNA-3), 150 μg/mL, 10 μL/g		
/	LNP (178.0 μg/mL) (μL)	LNP dilution buffer solvent (μL)
Theoretic quantity/volume	1011	189

AAV(Albumin_sg3_SA-hGLP1-Protracted), 20*10E11/mL, 200 μL/mice		
/	AAV (283*10E11/mL) (μL)	AAV dilution buffer solvent (μL)
Theoretic quantity/volume	85	1115

FIG. 8

Drug Administration Information for Milestone 2	
Group-1: LNP (Cas9+sgRNA-3), 150 μg/mL, 13.5 μL/g	
/	LNP (111 μg/mL) (μL)
Theoretic quantity/volume	1000

Group-2: LNP (Cas9+sgRNA-3), 300 μg/mL, 27 μL/g	
/	LNP (111 μg/mL) (μL)
Theoretic quantity/volume	2000

Group-3: LNP (AlbG3-dCAS9-SSAP+ sgRNA-3), 150 μg/mL, 11.7 μL/g	
/	LNP (128 μg/mL) (μL)
Theoretic quantity/volume	900

AAV(Albumin_sg3_SA-hGLP1-Protracted), 20*10E11/mL, 200 μL/mice		
/	AAV (283*10E11/mL) (μL)	AAV dilution buffer solvent (μL)
Theoretic quantity/volume	85	1115

FIG. 9

Blood GLP-1 Measurement Dilution Factors			
	Vehicle	Albumin_sg3_TBG-hGLP1-Protracted	AlbG3-dCAS9-SSAP+ sgRNA-3 Albumin_sg3_SA-hGLP1-Protracted
2 d	4	200	2000
7 d	4	200	10000
14 d	4	200	10000

FIG. 10

LNP Preparation for Milestone 1					
LNP	Quantity	Encapsulation Efficiency	Mean Intensity	Mean PDI	Storage
Cas9+sgRNA-3	Total RNA=177.5 µg/mL LNP=5.5 mg/mL	88.1%	147.5 nm	0.1565	-80 °C
AlbG3-dCAS9-SSAP+ sgRNA-3	Total RNA=178 µg/mL LNP: 5.0 mg/mL	86.7%	182.3 nm	0.1622	-80 °C

FIG. 11

LNP Preparation for Milestone 2					
LNP	Quantity	Encapsulation Efficiency	Mean Intensity	Mean PDI	Storage
Cas9+sgRNA-3	Total RNA=111 µg/mL LNP=5.4 mg/mL	88.5%	109.6 nm	0.081	4 °C
AlbG3-dCAS9-SSAP+ sgRNA-3	Total RNA=128 µg/mL LNP: 6.0 mg/mL	83.7%	83.72 nm	0.148	4 °C

FIG. 12

The Final State of Mice in Milestone 1				
G	**N**	Treatment-AAV/LNP	AAV/LNP Dose	Latest animal status
1	3	Vehicle	—	Three mice were normal
2	3	Albumin_sg3_TBG-hGLP1-Protracted	4*10E11	Three mice were normal
3	3	Cas9+sgRNA-3	0.5 mg/kg	Three mice were normal
4	3	Cas9+sgRNA-3	1.5 mg/kg	One mouse was dead; two mice were normal
5	3	Cas9+sgRNA-3	4.5 mg/kg	Two mice were dead; one mouse was normal
6	3	AlbG3-dCAS9-SSAP+ sgRNA-3 Albumin_sg3_SA-hGLP1-Protracted	1.5 mg/kg 4*10E11	Three mice were dead;

FIG. 13

The Final State of Mice in Milestone 2				
G	**N**	Treatment-AAV/LNP	AAV/LNP Dose	Latest animal status
1	3	Cas9+sgRNA-3	1.5 mg/kg	Three mice were normal
2	3	Cas9+sgRNA-3	3. mg/kg	Two mice were dead, One mouse was normal
3	3	AlbG3-dCAS9-SSAP+ sgRNA-3 Albumin_sg3_SA-hGLP1-Protracted	1.5 mg/kg 4*10E11	Three mice were normal

FIG. 14

Serum concentrations of GLP-1 (Mean±SD; pM)

Group/Time	Vehicle	Albumin_sg3_TBG-hGLP1-Protracted	AlbG3-dCAS9-SSAP+ sgRNA-3 Albumin_sg3_SA-hGLP1-Protracted
2 d	8.2±5.91	1642.95±163.38	143608.27±28429.45
7 d	3.8±1.07	5287.75±176.47	272784.16±16654.43
14 d	4.46±2.2	14772.61±1623.04	516135.68±174997.12

FIG. 15

Serum concentrations of GLP-1 (Mean±SD; pM)

Group	Sample ID	well 1	well 2	Mean	SD	%CV	RE Conc	Mean±SD
Group 1 Vehicle (milestone 1)	171-2d	587	578	582.5	4.5	0.77%	2.25	8.2 ±5.91
	181-2d	1061	1129	1095	34	3.11%	8.25	
	190-2d	1655	1645	1650	5	0.30%	14.08	
	171-7d	780	687	733.5	46.5	6.34%	4.14	3.8 ±1.07
	181-7d	588	631	609.5	21.5	3.53%	2.60	
	190-7d	741	815	778	37	4.76%	4.67	
	171-14d	868	800	834	34	4.08%	5.32	4.46 ±2.2
	181-14d	571	549	560	11	1.96%	1.96	
	190-14d	903	901	902	1	0.11%	6.10	
Group 2 Albumin_sg3_TBG -hGLP1-Protracted (milestone 1)	175-2d	6269	6441	6355	86	1.35%	1768.26	1642.95 ±163.38
	183-2d	5138	4997	5067.5	70.5	1.39%	1458.17	
	187-2d	6057	6101	6079	22	0.36%	1702.41	
	175-7d	22021	22543	22282	261	1.17%	5483.20	5287.75 ±176.47
	183-7d	20811	20866	20838.5	27.5	0.13%	5140.13	
	187-7d	21131	21388	21259.5	128.5	0.60%	5239.93	
	175-14d	62567	60550	61558.5	1008.5	1.64%	16141.42	14772.61 ±1623.04
	183-14d	50825	51097	50961	136	0.27%	12979.59	
	187-14d	58507	58440	58473.5	33.5	0.06%	15196.83	
Group 3 LNP (AlbG3-dCAS9-SSAP+ sgRNA-3) AAV(Albumin_sg3 _SA-hGLP1-Protracted) (milestone 2)	191-2d	48442	46284	47363	1079	2.28%	119579.39	143608.27 ±28429.45
	192-2d	54270	52109	53189.5	1080.5	2.03%	136252.80	
	193-2d	67198	64563	65880.5	1317.5	2.00%	174992.62	
	191-7d	24241	23692	23966.5	274.5	1.15%	291991.60	272784.16 ±16654.43
	192-7d	21815	21156	21485.5	329.5	1.53%	262358.37	
	193-7d	21370	21878	21624	254	1.17%	264002.50	
	191-14d	32030	31862	31946	84	0.26%	390240.45	516135.68 ±174997.12
	192-14d	36365	35653	36009	356	0.99%	442200.92	
	193-14d	56888	54597	55742.5	1145.5	2.05%	715965.67	

FIG. 16

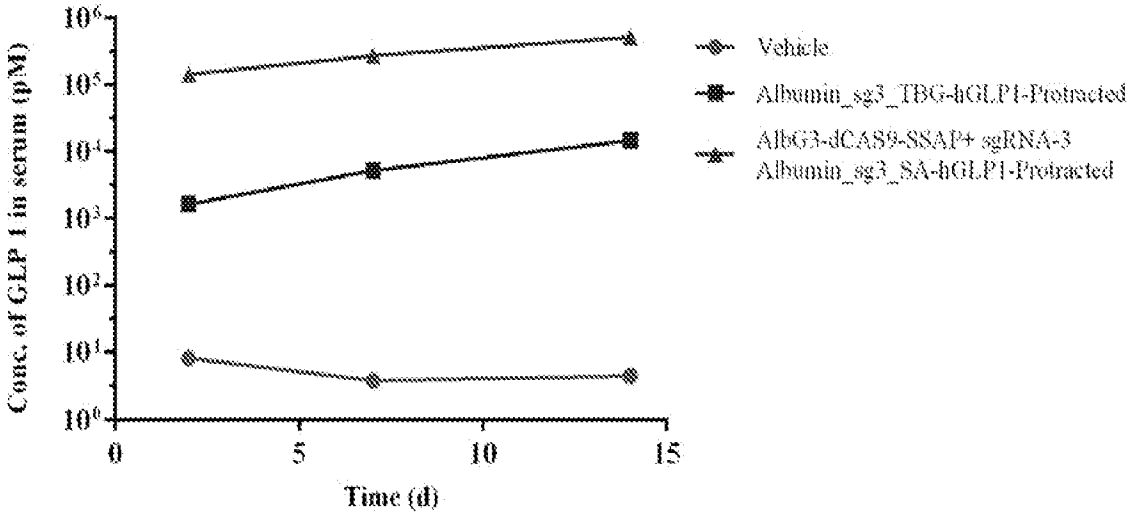

Conc. of GLP 1 in serum (pM)
10^6
10^5
10^4
10^3
10^2
10^1
10^0
0
5
10
15
Time (d)
Vehicle
Albumin_sg3_TBG-hGLP1-Protracted
AlbG3-dCAS9-SSAP+ sgRNA-3
Albumin_sg3_SA-hGLP1-Protracted

FIG. 17

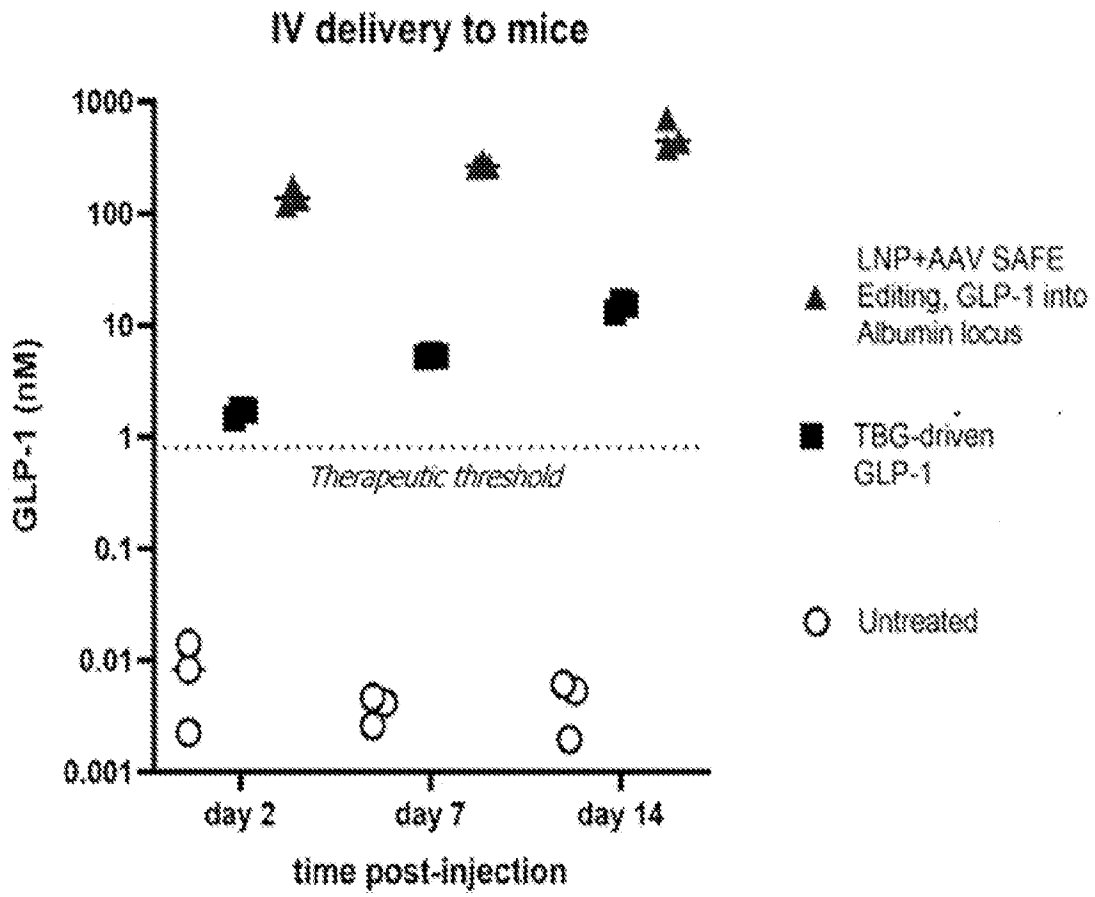
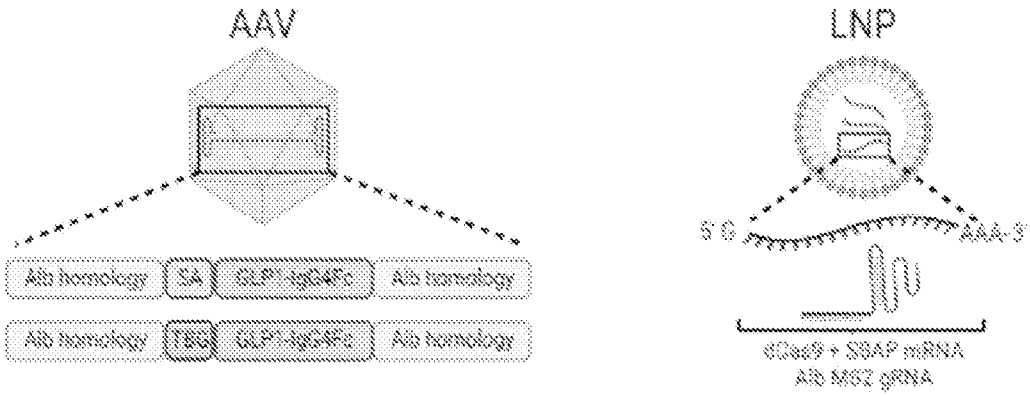
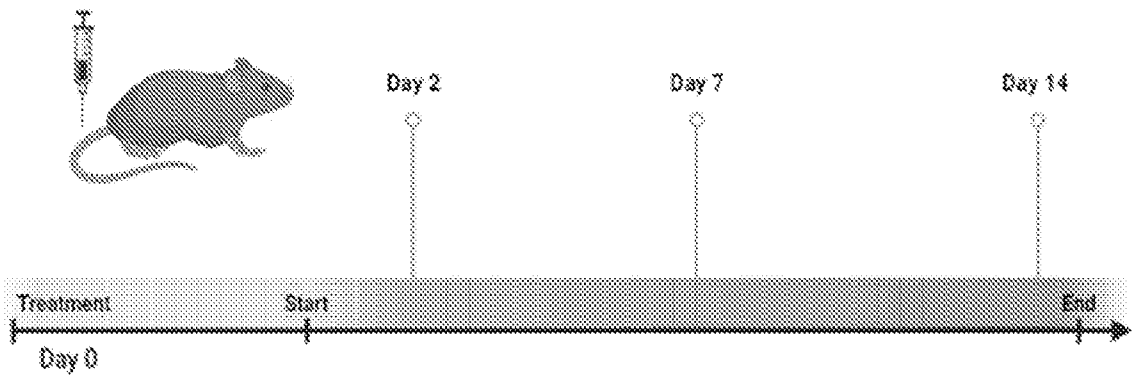

IV delivery to mice
GLP-1 (nM)
1000
100
10
1
0.1
0.01
0.001
Therapeutic threshold
day 2
day 7
day 14
time post-injection
LNP+AAV SAFE Editing, GLP-1 into Albumin locus
TBG-driven GLP-1
Untreated

FIG. 18



Group	AAV	LNP
Untreated (n = 3)	NA	NA
AAV-only control (n = 3)	AAV8 4 x 10^{11} GC TBG promoter-Dulaglutide	NA
SAFE editing (n = 3)	AAV8 4 x 10^{11} GC Splice acceptor-Dulaglutide	SSAP, dCas9, sgRNA 1.5 mg/kg



INTERNATIONAL SEARCH REPORT

International application No
PCT/US2024/015176

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N15/86 A61K9/127 A61K9/51 C12N9/22 C12N15/00
C12N15/10 C12N15/11
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 C40B C07K A61K

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

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WANG CHENGKUN ET AL: "CRISPR-Cas12a System With Synergistic Phage Recombination Proteins for Multiplex Precision Editing in Human Cells", FRONTIERS IN CELL AND DEVELOPMENTAL BIOLOGY, vol. 9, 14 June 2022 (2022-06-14), XP093172938, CH ISSN: 2296-634X, DOI: 10.3389/fcell.2021.719705 abstract, p.1 - 2, p. 5, column 2, paragraph 2-3, p. 2-3, figure 1,2, supplementary material, p. 2, right column, paragraph 2	1-7, 11-13, 15,17, 18, 23-29, 32-35, 37,39,40

-/--

[x] Further documents are listed in the continuation of Box C.

[x] See patent family annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance;; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance;; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
18 July 2024	29/07/2024

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	Empl, Laura

1

Form PCT/ISA/210 (second sheet) (April 2005)

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WANG CHENGKUN ET AL: "dCas9-based gene editing for cleavage-free genomic knock-in of long sequences", NATURE CELL BIOLOGY, NATURE PUBLISHING GROUP UK, LONDON, vol. 24, no. 2, 1 February 2022 (2022-02-01), pages 268-278, XP037691428, ISSN: 1465-7392, DOI: 10.1038/S41556-021-00836-1 [retrieved on 2022-02-10] see abstract, p.268, p. 277, figure 7, figure 1 and 3, p.269	1-7,9, 10,13, 15-18, 23-29, 31,32, 35,37-40
X	WO 2021/178432 A1 (UNIV LELAND STANFORD JUNIOR [US]) 10 September 2021 (2021-09-10) claims 1-14, 19-23,26-27, 31, 37-42, figure 33A, 35, paragraphs [00129-001335], example 7	1-10, 13-18, 23-32, 35-40
X	WO 2020/264016 A1 (INARI AGRICULTURE INC [US]) 30 December 2020 (2020-12-30) paragraphs [0179-0184], claims 157-158,paragraphs [0077-0082],claims 1-3, 13-14, 36-38, 49-52, figures 3,6, 10, [0137]	1-4,15, 17, 23-26, 37,39,45
X	HAO YIN ET AL: "Therapeutic genome editing by combined viral and non-viral delivery of CRISPR system components in vivo", NATURE BIOTECHNOLOGY, vol. 34, no. 3, 1 February 2016 (2016-02-01), pages 328-333, XP055540393, New York ISSN: 1087-0156, DOI: 10.1038/nbt.3471 abstract, figures 1-2, p.334, left column, supplementary figures	1,4, 19-23, 26,41-44

-/--

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2024/015176

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	JONATHAN D. FINN ET AL: "A Single Administration of CRISPR/Cas9 Lipid Nanoparticles Achieves Robust and Persistent In Vivo Genome Editing", CELL REPORTS, vol. 22, no. 9, 1 February 2018 (2018-02-01), pages 2227-2235, XP055527484, US ISSN: 2211-1247, DOI: 10.1016/j.celrep.2018.02.014 abstract, p.2228, left column, p.2231, left column, paragraph 2-p.2232, right column -----	1,23

1

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2024/015176

Box No. I Nucleotide and/or amino acid sequence(s) (Continuation of item 1.c of the first sheet)

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:

   a. ☐ forming part of the international application as filed.

   b. ☒ furnished subsequent to the international filing date for the purposes of international search (Rule 13*ter*.1(a)).

      ☒ accompanied by a statement to the effect that the sequence listing does not go beyond the disclosure in the international application as filed.

2. ☐ With regard to any nucleotide and/or amino acid sequence disclosed in the international application, this report has been established to the extent that a meaningful search could be carried out without a WIPO Standard ST.26 compliant sequence listing.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2024/015176

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2021178432 A1	10-09-2021	AU 2021231769 A1	29-09-2022
		BR 112022017196 A2	25-10-2022
		CA 3173526 A1	10-09-2021
		CN 115667283 A	31-01-2023
		EP 4114845 A1	11-01-2023
		IL 296057 A	01-10-2022
		JP 2023515670 A	13-04-2023
		KR 20220151175 A	14-11-2022
		US 2023091242 A1	23-03-2023
		WO 2021178432 A1	10-09-2021
WO 2020264016 A1	30-12-2020	AR 123323 A1	23-11-2022
		AU 2020301236 A1	25-11-2021
		BR 112021026220 A2	15-02-2022
		CA 3138663 A1	30-12-2020
		CN 114630910 A	14-06-2022
		EP 3990633 A1	04-05-2022
		IL 289106 A	01-02-2022
		JP 2022539338 A	08-09-2022
		KR 20220052905 A	28-04-2022
		SG 11202112092T A	29-11-2021
		US 2020407754 A1	31-12-2020
		US 2021277422 A1	09-09-2021
		WO 2020264016 A1	30-12-2020