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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
02 September 2021 (02.09.2021)

WIPO | PCT

(10) International Publication Number
WO 2021/173734 A1

(51) International Patent Classification:
C12N 15/90 (2006.01)

(21) International Application Number:
PCT/US2021/019494

(22) International Filing Date:
24 February 2021 (24.02.2021)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

62/980,904	24 February 2020 (24.02.2020)	US
62/980,922	24 February 2020 (24.02.2020)	US
63/000,224	26 March 2020 (26.03.2020)	US

(71) Applicants: **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, Massachusetts 02142 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US).

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

(74) Agent: **RUTLEDGE, Rachel D.** et al.; Johnson, Marcou, Isaacs & Nix, LLC, Customer No. 95280, P.O. Box 691, Hoschton, Georgia 30548 (US).

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

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

(54) Title: NOVEL TYPE IV AND TYPE I CRISPR-CAS SYSTEMS AND METHODS OF USE THEREOF

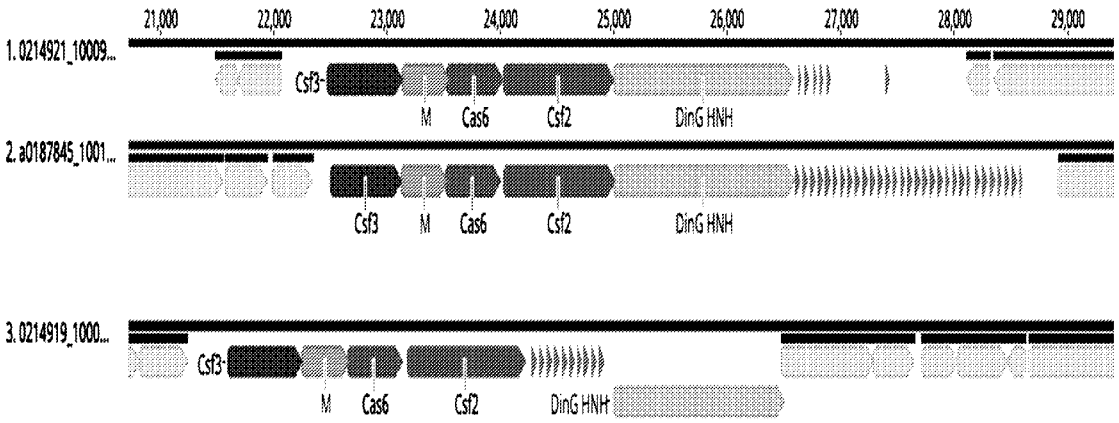


FIG. 3

(57) Abstract: The present disclosure provides for systems, methods, and compositions for targeting nucleic acids. In particular, the invention provides novel class 1, Type IV and novel class I, Type I Cas proteins and their use in modifying target sequences.

[Continued on next page]

WO 2021/173734 A1

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

Published:

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

NOVEL TYPE IV AND TYPE I CRISPR-CAS SYSTEMS AND METHODS OF USE THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/980,904, filed February 24, 2020; U.S. Provisional Application No. 63/000,224, filed March 26, 2020, and U.S. Provisional Application No. 62/980,922, filed February 24, 2020. The entire contents of the above-identified applications are hereby fully incorporated herein by reference.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under Grant Nos. MH110049 and HL141201 awarded by the National Institutes of Health. The government has certain rights in the invention.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0003] The contents of the electronic sequence listing ("BROD-5060WP_ST25.txt"; Size is 50,909,273 bytes and it was created on February 24, 2021) is herein incorporated by reference in its entirety.

TECHNICAL FIELD

[0004] The subject matter disclosed herein generally relates to systems, methods and compositions used for the modification and control of gene expression using Class 1 CRISPR-Cas systems and components thereof.

BACKGROUND

[0005] The CRISPR-Cas systems of bacterial and archaeal adaptive immunity are some such systems that show extreme diversity of protein composition and genomic loci architecture. There exists a pressing need for alternative and robust systems and techniques for targeting nucleic acids or polynucleotides.

SUMMARY

[0006] In one aspect, the present disclosure provides a non-naturally occurring, engineered composition comprising a Cas protein that comprises an HNH domain that is less than 600 amino acids in size, and at least one guide sequence capable of complexing with the Cas protein and directing binding of the guide-Cas protein complex to a target polynucleotide.

[0007] In some embodiments, the Cas protein is a class 1, Type IV Cas protein. In some embodiments, the class 1, Type IV Cas protein is a DinG protein. In some embodiments, the composition further comprises one or more Cas proteins. In some embodiments, the one or more Cas proteins comprise Cas7 or a Cas7-like Cas protein, Cas5 or a Cas5-like Cas protein, Cas6 or a Cas6-like protein, and a Cse3 family protein. In some embodiments, the one or more Cas proteins comprise Csf2, Csf3, Cas6, and Pfam08798. In some embodiments, the Cas protein is a protein with nucleotide sequence listed in Table 11 and SEQ ID NOs 1-405. In some embodiments, the target sequence comprises a protospacer adjacent motif (PAM) at 5' side of the target sequence. In some embodiments, the PAM sequence comprises CC. In some embodiments, the PAM sequence comprises (C/T)CN. In some embodiments, the target sequence comprises a PAM at 3' side of the target sequence. In some embodiments, the PAM sequence is GG. In some embodiments, the composition comprises cascade/helicase activity. In some embodiments, the composition does not comprise Cas3 DNA shredding activity. In some embodiments, the composition further comprises a donor polynucleotide. In some embodiments, the donor polynucleotide: a. introduces one or more mutations to the target polynucleotide, b. corrects a premature stop codon in the target polynucleotide, c. disrupts a splicing site, d. restores a splicing site, or e. a combination thereof. In some embodiments, the one or more mutations introduced by the donor polynucleotide comprises substitutions, deletions, insertions, or a combination thereof. In some embodiments, the one or more mutations causes a shift in an open reading frame on the target polynucleotide. In some embodiments, the composition further comprises a plurality of guide molecules capable of complexing with the Cas protein and directing binding of the guide-Cas protein complex to one or more target polynucleotides.

[0008] In another aspect, the present disclosure provides a composition comprising one or more polynucleotides encoding: one or more class 1, Type IV Cas proteins or functional fragments thereof, wherein the one or more class 1, Type IV Cas proteins comprises DinG protein with a length less than 600 amino acids; and one or more guide molecules capable of complexing with the class 1, Type IV Cas protein and directing binding of the guide-Cas protein complex to one or more target polynucleotides.

[0009] In some embodiments, the composition further comprises a donor polynucleotide. In some embodiments, the donor polynucleotide comprises a polynucleotide insert. In some embodiments, the one or more polynucleotides encode part of all of the components of the composition herein. In some embodiments, the one or more class 1, Type IV Cas proteins comprise Csf2 (Cas7 like), Csf3 (Cas5 like), Cas6, and Pfam08798 (Cse3 family).

[0010] In another aspect, the present disclosure provides a vector comprising the one or more polynucleotides herein.

[0011] In another aspect, the present disclosure provides an engineered cell comprising the composition herein.

[0012] In another aspect, the present disclosure provides a method of modifying a target polynucleotide sequence in a cell, comprising introducing to the cell: one or more class 1, Type IV Cas proteins or functional fragments thereof, wherein the one or more class 1, Type IV Cas proteins comprises DinG protein with a length less than 600 amino acids; and one or more guide molecules capable of complexing with the class 1, Type IV Cas protein and directing binding of the guide-Cas protein complex to one or more target polynucleotides.

[0013] In some embodiments, the method further comprises introducing a donor polynucleotide. In some embodiments, the donor polynucleotide: a. introduces one or more mutations to the target polynucleotide, b. corrects a premature stop codon in the target polynucleotide, c. disrupts a splicing site, d. restores a splicing site, or e. a combination thereof. In some embodiments, the one or more mutations introduced by the donor polynucleotide comprises substitutions, deletions, insertions, or a combination thereof. In some embodiments, the one or more mutations causes a shift in an open reading frame on the target polynucleotide. In some embodiments, the cell is a prokaryotic cell. In some embodiments, the cell is a eukaryotic cell. In some embodiments, the cell is a mammalian cell, a cell of a non-human primate, or a human cell. In some embodiments, the cell is a plan cell.

[0014] In another aspect, the present disclosure provides a non-naturally occurring, engineered composition comprising a Cas protein that comprises an HNH domain and a helicase domain, and at least one guide sequence capable of complexing with the Cas protein and directing binding of the guide-Cas protein complex to a target polynucleotide. In some embodiments, the helicase domain is a DinG domain. In some embodiments, the Cas protein is a class 1 Type IV Cas protein. In another aspect, the present disclosure provides a composition comprising one or more polynucleotides encoding one or more components of the composition herein. In another aspect, the present disclosure provides a method of modifying a target polynucleotide sequence in a cell, comprising introducing to the cell one or more components of composition herein.

[0015] The present disclosure provides systems and methods for nucleic acid modification. In one aspect, the present disclosure provides systems for one or more components of a class 1, Type I CRISPR-Cas system, such as a class 1, Type I Cas protein(s) and/or a guide RNA molecule(s) capable of formulating a complex with the class 1, Type I Cas protein(s). In some

embodiments, a class 1, Type I Cas protein comprises an HNH domain that is less than 400 amino acids. Nucleic acid sequences of exemplary Type I Cas proteins in the systems include those in Table 2, SEQ ID NOs. 495-1212 in the Sequence Listing herein. In an aspect, the class 1, Type I Cas system that comprises an HNH domain that is less than 400 amino acids in length, and at least one guide sequence capable of complexing with the Cas protein and directing binding of the guide-Cas protein complex to a target polynucleotide is provided. The HNH domain may be comprised in an McRA protein/subunit protein. The HNH domain may be comprised in a Cas5 protein. In an aspect, composition comprises cascade/helicase activity. The composition may comprise a system that does not comprise Cas3, in an aspect the system does not comprise Cas3 DNA shredding activity.

[0016] In another aspect, embodiments disclosed herein include systems and uses for such Cas proteins including diagnostics, base editing therapeutics and methods of detection. Fusion proteins comprising one or more class 1, Type I Cas proteins herein, and nucleotide deaminase may also be used for base editing. Delivery of the proteins and systems disclosed is also provided, including to a variety of cells and via a variety of particles, vesicles and vectors.

[0017] Methods of modifying a target polynucleotide sequence in the cell comprising use of the systems described herein.

[0018] These and other aspects, objects, features, and advantages of the example embodiments will become apparent to those having ordinary skill in the art upon consideration of the following detailed description of illustrated example embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0019] An understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention may be utilized, and the accompanying drawings of which:

[0020] **FIG. 1** shows determining of PAM sequences of exemplary class 1, Type IV Cas (SEQ ID NO:406-412).

[0021] **FIG. 2** shows the alignment of sequences of exemplary Type IV Cas proteins comprising HNH and DinG.

[0022] **FIG. 3** shows examples of loci comprising DinG and HNH sequences.

[0023] **FIG. 4** shows identification of PAM by spacer BLAST.

[0024] **FIG. 5** shows an exemplary Type IV Cas performed plasmid interference with 5' (C/T)CN PAM.

[0025] **FIG. 6** shows determination of PAM sequences of exemplary Type I Cas (SEQ ID NOs: 1122-1130).

[0026] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

DEFINITIONS

[0027] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.): Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.): Antibodies A Laboratory Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011).

[0028] As used herein, the singular forms "a", "an", and "the" include both singular and plural referents unless the context clearly dictates otherwise.

[0029] The term "optional" or "optionally" means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[0030] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0031] The terms "about" or "approximately" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" or "approximately" refers is itself also specifically, and preferably, disclosed.

[0032] As used herein, a "biological sample" may contain whole cells and/or live cells and/or cell debris. The biological sample may contain (or be derived from) a "bodily fluid". The present invention encompasses embodiments wherein the bodily fluid is selected from amniotic fluid, aqueous humor, vitreous humor, bile, blood serum, breast milk, cerebrospinal fluid, cerumen (earwax), chyle, chyme, endolymph, perilymph, exudates, feces, female ejaculate, gastric acid, gastric juice, lymph, mucus (including nasal drainage and phlegm), pericardial fluid, peritoneal fluid, pleural fluid, pus, rheum, saliva, sebum (skin oil), semen, sputum, synovial fluid, sweat, tears, urine, vaginal secretion, vomit and mixtures of one or more thereof. Biological samples include cell cultures, bodily fluids, cell cultures from bodily fluids. Bodily fluids may be obtained from a mammal organism, for example by puncture, or other collecting or sampling procedures.

[0033] The terms "subject," "individual," and "patient" are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. Tissues, cells and their progeny of a biological entity obtained in vivo or cultured in vitro are also encompassed.

[0034] Various embodiments are described hereinafter. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s). Reference throughout this specification to "one embodiment", "an embodiment," "an example embodiment," means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases "in one embodiment," "in an embodiment," or "an example embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person

skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

[0035] All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

OVERVIEW

[0036] The present disclosure provides systems and methods for nucleic acid targeting and modification. In one aspect, the present disclosure provides systems comprising one or more components of a Class 1, Type IV CRISPR-Cas system, such as a Type IV Cas protein(s) and/or a guide RNA molecule(s) capable of formulating a complex with the Type IV Cas protein(s). In one example embodiment, a Type IV Cas may comprises a helicase comprising a DinG domain and a HNH domain. In another example embodiment, the Type IV Cas helicase protein comprising a DinG domain and HNH domain may be less than 600 amino acids in size. Nucleic acid sequences of exemplary Type IV Cas proteins, which are alternately referred to herein as Type IV subunit proteins, in the systems include those in Table 1, and SEQ ID NOs: 1-405 in the Sequence Listing filed herewith.

[0037] In another aspect, embodiments disclosed herein include systems and uses for such Cas systems including research reagents, therapeutics, and diagnostics. Fusion proteins comprising one or more Type IV or Type I Cas proteins/subunits are detailed herein, and nucleotide deaminase may also be used for base editing. For Type IV Cas systems, fusion Delivery of the proteins and systems disclosed is also provided, including to a variety of cells and via a variety of particles, vesicles and vectors.

[0038] In another aspect, the present disclosure provides systems comprising one or more components of a class 1, Type I CRISPR-Cas system, such as class 1, Type I Cas protein(s), which are alternately referred to herein as Type I subunit proteins and/or a guide RNA molecule(s) capable of formulating a complex with the class 1, Type I Cas protein(s)/subunit(s). In some embodiments, a class 1, Type I Cas helicase protein comprises an HNH domain comprising nuclease activity that is less than 400 amino acids. Nucleic acid sequences of exemplary Type I Cas proteins in the systems include those in Table 12, SEQ ID

NOs. 495-1212 in the Sequence Listing incorporated herein. In an aspect, the Class 1, Type I Cas system comprises a Cas5 or Cas5-like subunit comprising an HNH domain. In one embodiment the Class 1, Type I system comprises helicase activity. In an aspect, the Class 1, Type I system does not comprise Cas3 DNA shredding activity.

[0039] In another aspect, embodiments disclosed herein include systems and uses for such Cas systems including. research reagents, therapeutics, and diagnostics..Fusion proteins comprising one or more class 1, Type I Cas proteins/subunits are detailed herein, and nucleotide deaminase may also be used for base editing. Delivery of the proteins and systems disclosed is also provided, including to a variety of cells and via a variety of particles, vesicles and vectors.

SYSTEMS AND COMPOSITIONS IN GENERAL

[0040] In one aspect, the present disclosure provides for systems and compositions for modification of nucleic acids. In general, the systems or composition may comprise one or more Class 1 Cas systems that comprise one or more HNH domains. The HNH domains may have less than 1000 (e.g., less than 600) amino acids. The systems and compositions may further comprise one or more guide sequences. The guide sequences may be capable of complexing with the Cas protein(s)/system and/or hybridizing to a target sequence. The guide sequence may be capable of directing the binding of the guide-Cas protein complex to the target polynucleotide.

[0041] In some example embodiments, the CRISPR-Cas system may be class 1, Type IV CRISPR-Cas system. In some embodiments, the class 1, Type IV system may comprise a helicase and one or more of Cas7 or Cas7-like proteins (e.g. Csf2), Cas5 or Cas5-like Cas proteins (e.g. Csf3) or Cas6-like Cas protein, a Cse3 family protein (e.g., Pfam08798). The helicase may comprise a DinG domain. In certain example embodiments, the helicase comprises a DinG and a HNH domain. In one example, the systems and compositions comprise a Type IV CRISPR-Cas system comprising a helicase comprising a DinG domain and a HNH domain, Csf2, Csf3, Cas6, and/or pfam08798. The systems and compositions may further comprise one or more guide sequences. The guide sequences may be capable of complexing with the Cas protein(s) and/or hybridizing to a target sequence. The guide sequence may be capable of directing the binding of the guide-Cas protein complex to the target polynucleotide.

[0042] In one aspect, the CRISPR-Cas system may be a class I, Type IV CRISPR-Cas system comprising a fusion of multimeric subunit components.

[0043] In one embodiment, the class I, Type IV CRISPR-Cas system may comprise a fusion at the N-terminus, the C-terminus, or any accessible part of the protein complex, e.g. on a subunit of the Cas systems.

[0044] In some example embodiments, the CRISPR-Cas system may be a class 1, Type I CRISPR-Cas system. In some embodiments, the class 1 Type I system may comprise an HNH domain that is less than 400 amino acids and one or more of Cas1, Cas2, Cas6, Cas7, a Cas8e-Cse fusion, CasE, Cse2, and McrA. In one example, the systems and compositions comprise a Cas system that comprises an HNH domain, and Cas1, Cas2, Cas6, Cas7, a Cas8e-Cse fusion, CasE, Cse2, and/or McrA subunit proteins. In some embodiments, the composition comprises cascade/helicase activity. In one embodiment, the Type I system comprises cascade/helicase activity without Cas3 shredding activity typically present in Type I systems. The systems and compositions may further comprise one or more guide sequences. The guide sequences may be capable of complexing with the Cas protein(s) and/or hybridizing to a target sequence. The guide sequence may be capable of directing the binding of the guide-Cas protein complex to the target polynucleotide.

[0045] In one aspect, the CRISPR-Cas system may be a class I, Type I CRISPR-Cas system comprising a fusion of multimeric subunit components.

[0046] In one embodiment, the class I, Type I CRISPR-Cas system comprises fusion at the N-terminus, the C-terminus, and any accessible part of the protein complex.

In one embodiment the class I, Type I CRISPR-Cas system comprises fusion of the Cas3 Cas5, Cas6, Cas7, Cas8 and Cas11 subunits, and in any combination thereof.

CAS PROTEINS

[0047] The methods, systems, and tools provided herein may be designed for use with Class 1 CRISPR proteins. In certain example embodiments, the Class 1 system may be Type I, Type III or Type IV Cas proteins as described in Makarova et al. "Evolutionary classification of CRISPR-Cas systems: a burst of class 2 and derived variants" Nature Reviews Microbiology, 18:67-81 (Feb 2020)., incorporated in its entirety herein by reference, and particularly as described in Figure 1, p. 326. The Class 1 systems typically use a multi-protein effector complex, which can, in some embodiments, include ancillary proteins, such as one or more proteins in a complex referred to as a CRISPR-associated complex for antiviral defense (Cascade), one or more adaptation proteins (e.g. Cas1, Cas2, RNA nuclease), and/or one or more accessory proteins (e.g. Cas 4, DNA nuclease), CRISPR associated Rossman fold (CARF) domain containing proteins, and/or RNA transcriptase. Although Class 1 systems have limited sequence similarity, Class 1 system proteins can be identified by their similar

architectures, including one or more Repeat Associated Mysterious Protein (RAMP) family subunits, e.g. Cas 5, Cas6, Cas7. RAMP proteins are characterized by having one or more RNA recognition motif domains. Large subunits (for example cas8 or cas10) and small subunits (for example, cas11) are also typical of Class 1 systems. See, e.g., Figures 1 and 2. Koonin EV, Makarova KS. 2019 Origins and evolution of CRISPR-Cas systems. Phil. Trans. R. Soc. B 374: 20180087, doi: 10.1098/rstb.2018.0087. In one aspect, Class 1 systems are characterized by the signature protein Cas3. In Class 1, Type I systems, the helicase-nuclease fusion enzyme Cas3 is specifically recruited to the R-loop-forming Cascade, nicks the non-target strand (NTS) DNA, and degrades, or shreds, its upstream region (PAM-proximal side), and degrades target DNA. See, Xiao et al., Science. 2018 Jul 6;361(6397):eaat0839. doi: 10.1126/science.aat0839. Epub 2018 Jun 7.

Type IV Systems

[0048] The Cas systems may be class 1 Type IV Cas comprising one or more Cas subunit proteins, which may also be referred to herein as a multimeric Cas protein or system. In some embodiments, the systems may comprise one or more class 1 Type IV Cas subunits or proteins. The Cas system may include Cas subunit proteins that have at least one HNH domain. In certain examples, the Cas subunit may comprise only one HNH domain.

[0049] In some embodiments, the HNH subunit protein of the system is less than 2000 amino acids in size. For example, the Cas protein may be less than 2000, less than 1900, less than 1800, less than 1700, less than 1600, less than 1500, less than 1400, less than 1300, less than 1200, less than 1100, less than 1000, less than 950, less than 900, less than 890, less than 880, less than 870, less than 860, less than 850, less than 840, less than 830, less than 820, less than 810, less than 800, less than 790, less than 780, less than 770, less than 760, less than 750, less than 700, less than 650, less than 600 amino acids in size. In some examples, the Cas protein is less than 600 amino acids in size.

[0050] The present Type IV systems may lack one or more hallmark components of CRISPR-Cas systems including an effector nuclease, even when a CRISPR array may be present. See, e.g. Pinilla Redondo, et al., Nucleic Acids Research, Volume 48, Issue 4, 28 February 2020, Pages 2000–2012, doi:10.1093/nar/gkz1197, for general Type IV systems. In an embodiment, novel type IV CRISPR-Cas proteins are detailed. In one embodiment, the type IV protein comprises an HNH domain, a well-characterized nuclease domain. See, e.g. Keeble A.H., Maté M.J., Kleanthous C. (2005) HNH Endonucleases. In: Belfort M., Wood D.W., Stoddard B.L., Derbyshire V. (eds) Homing Endonucleases and Inteins. Nucleic Acids and

Molecular Biology, vol 16. Springer, Berlin, Heidelberg. Doi: 10.1007/3-540-29474-0_4; Zuo et al., eLife 2019;8:e46500 DOI: 10.7554/eLife.46500.

[0051] In one aspect, the present disclosure provides for systems and compositions wherein the CRISPR-Cas system may be a class 1, Type IV CRISPR-Cas system comprising a helicase (e.g., DinG HNH), Csf3 (e.g., Cas5-like), Cas6, Csf2 (Cas7-like), Cse3 and pfam08798.

[0052] In one embodiment, the class 1, Type IV CRISPR-Cas system comprises DinG HNH Csf3 (Cas5-like), Cas6 and Csf2 (Cas7-like).

[0053] In one embodiment, the class 1, Type IV CRISPR-Cas system comprises DinG HNH, pfam08798, Csf2, and Csf3.

[0054] In one embodiment, the class 1, Type IV CRISPR-Cas system comprises DinG HNH pfam08798, Csf2 and Csf3.

[0055] In one embodiment, the class 1, Type IV CRISPR-Cas system comprises DinG HNH, Csf3, Cse3 and Csf2.

[0056] In one embodiment, the class 1, Type IV CRISPR-Cas system comprises DinG HNH, Csf3, pfam08798 and Csf2.

[0057] In one embodiment, the class I, Type IV CRISPR-Cas system comprises fusion of the DinG, Cas5, Cas6 (e.g., cas6b, cas6e, cas6e/f), Cas7, Cas8 and Cas8-like, Cas10-like, Cas11, Cas2 and Cas4 subunits in any combination thereof.

[0058] In an embodiment, the class 1, Type IV CRISPR-Cas system comprises onr or more components from Table 1.

Table 1.

Map Label	Feature Label	Sequence
DinG_HNH_1_017 2379_10000168_or ganized	DR(SEQ ID NO:455)	ggtcaagccaccctgcatctgggtggttactacaac
	Csf3 (Cas5 like) (SEQ ID NO:456	atggacaagatcaaagtaaccgcgcacctgctggccccagttgtcacaggggcggctacatgaccctcgacggtctgctggcggcactcctgttcg atgacctgcaagacgtggacgccgcacacgcggccatcccgattcggcagaccgacggcctgcaccacgccagcgccgccatcatggaggcaag tcgggagcgtatctcgttcattgccgcactgcggcccggccacagcatcgatcctgacctgatccagaagaacaagcacgggcagttgcaccgcaag ttcgacacgtccctgaccaacgtattgaacagctacgccctgctgactgcgcccacggtgacgtggtatgccgagggtgatgcggatcacatcgagc ggctgatctccccggttcagttcatcggcaagcgccgagcctctgggtttggccgcgtggcgcgctggcaagtcgagcctgatgatctggatggtatc gaaggcccgttcggcgaaccgttgcgacccgtgccggtagggatgttcaagggtgacacgtcgcatcccatcgtcgacgctgcatggcgaccggcc tactggaacccgattcaccgcacggcgtgttacgcgcccgtttaa
	Cas6) (SEQ ID NO:457)	atggactacgcaatgcaaaagccgccaagtgtgcgcggctacgccctgcaccggatcgccgctggcttggctgatggcgagaaggtgctgttcgcg gactgtggcgatcacctgctgctgcgtaccagcaagccgctggacgttgatcggcgcaaggttcgcaccatcacggcgggagacatcgtcggctttg aacttcgcgcctgcgtcggcaagaaggtcaagggcaagcaccgctacttccactcgtccgactggcgcagccggcatgactggctggaacgacag ggcgaagcgcacggctttgggatcatcacgctgcactgcacgacgcagttgcagaccgtcaaggcaggcgaccgggaattcacggtcgatcagac cgacttcaccggcgttctaaaggtcacggatgccgaacgcttcaacagcgctgtcgccaatggcgtcggcaacaaggccaaggctttcgggtttggc atgttggtcatctaa
	Csf2 (Cas7 like) (SEQ ID NO:)	gtgaaaactatcgtcatcaacgccacccctgactaccgtcggccccttgtccatcaccatgcccatcgctgaaggccagatgggcaacaagttcaacaat ttccccgtcgtcacgcgcggcatggatgatgacggcaacaagcaggagaccggctacctgccggccaccacgctgcgcggctttctccggcgtgcc atcgtcacggatcagatgacgcaggccgccgccgacggcaagccctacaccttgcagaaagcctatgccgacctgatcgggcaggatgcggcgtc cgagtcgaaggacgacatcgacctgctcaagatgcgccagacccgcgaggaaaatcccgtcctcgatctgttcggctccggcttgggcatcaagtcg cggcttctggtatcccacttcatgccgacccacaacgtgctgccggaagtcgctaccggcgtccgcaaagacttggaggacaccgatggtgtgctgg agtcgctgtcgtcgcccgaccgcgatagctatcttggacgctccggtgcgaacagcaagcgcgctgccgcccctgacggtggcaaaggggttggagg gcaagattcgcaagacgaagaaggccggcgaggacatcgccgaacttgaggcggcgctgaaggaggcgcagggtatcgtcgcgcgctatgaatc cgacatgggcaacatggtcaactcatcgcgcacgctgaccaactattttgcgcttccggcaggcatcgaactcaagggccgcttggtcatcgagaagg cgcgcgaccgcgatctgccgatggtggaacttgcgctgaacgaattgtccaagcgtccgatcttgggggcacaaagcgcccgtggctgcggcgaaa tcgaagggacattcgatgttctggtggatggcgttcttgccaagaaaatcagcatcggtggttggcaagcggcgaccgtgaccgatttttgtggcgacg aattagcgaaatcggctgcgtaa

	DinG_HNH (DinG Helicase) (SEQ ID NO:458)	atgtttcgttctaccgaacaatcaacctttgcggcaatagtcgccaagcatatggcgagtagtgatgcccctctgctgttggaaggggagtcggcatcg gcaagacccgcgcctatatgtcagcgctggttcaatccgggcaacgggtcgctgtcgtcctgccgacgcatcagttgatcgaccagttgctcgcatcgt ccgatctgggggcgacgcgcggcaccgctacgctggcggcgtttcggcctgcccgaatgttttgaacgccgggccgactacgaggccaacaagcgc gcggcgactgacgcacaactgatgttctgtaccgccgccagcgtcatcattgaccatcgcttgggcggcgagtacaacggttccaccacccgcgcct acatcgtgttcgatgaggctgaccaactgccggacatggcggcactgcaaagcgacttcaccatcacgcagacagaagtggctgacatgggtatccg cctgacgaccgtcgaggatgcgttgcaggcgattctggcgaaaccgccacgcatcgtagaaccggaaaccaaggccgcagcgcagatcatcctcg atgccatcgaggaaccggcgtggtatcagtccgcaggcatcagcgacgacggcgacatcatgctgacgcacaagctgcccggccgcctgctgaag aagatcagcaaccgcagcaacgtggcgttcgtctcggccaccctgacggtcggcggtcgcttcaaggattttcagaactcgatgggcatcgagaaga tcagccacctgtcgtccgcgattgaaccgcggcagcatggttcgctggatctccacatgatcccgatggcggtcgatacgccggaatggatcagcgc agtcgtacaagtggcgaacgatgcgccgaagcccgtgctggtggcaactccatcgcatgacctatccagcctgctcggcggacgtctgccgggggc tgtcgtcaggtctgtcgaggaaacggcggctcaggctggcgacagggtgagcgccgctggcatcctgatttctgccgggggcatgggcagggctgga tacaccgctgcgctggcgctcgattgtcgtgccgcgcgtgccgttcggtcagccggtcgtgattgacgggcaagccacgaccacctatttcgatgcgc gcaacaccgccgtgcgccgcttgcgtcaagtcatcgggcgcggactgaggacaccggatgcagtgtgcagcgtgtacatcttggatgccagagcgg agaaactaagcggcttcgtaccggagcgctttctaacgagtgggctactcggacatatatcgaaggggcgaaaaaggaggtcgtactgtccaaggct gaacgcgacccccgcagttcgtcgggcggcgctgaatcactacggacggaagtgcatggcctgcggcttcatgccgaaggtcgattcacagcttgac gtgcatcaccttgaccccatcgccgaggggcagcggcgtaccaagatcgaagacgtggtggtgctgtgcgcgaactgccaccgactggcgcacgc gaccaatccgccgacacggctggatgttacggagacagggcagccgcttccttgctga
DinG_HNH_2_0206130_10001073_organized	DR (SEQ ID NO:459)	atcaagccacccctctaaatgggtggttactagaac
	KOON_pfam08798 (SEQ ID NO:460)	atggatcacacaatcacaaaaccagcgcgcgtcgaaggctacgccgtccaccgtatggtcgccgctttggtagacgacgcacccgcgctgcatctcg atgccggcgacaccctaaccattcgcaccgatgcgcccctgcccagtgggcatattgtgggcttctcgttgcgtgcctgtgtctctgtcaagaacaagg ggcggcaccgctatttccccagtgctgactggcgctctcgccacagttggctcgagcgaaaaggcgtacagcatggcttcgaagtcatcaccgtgcat tgctctgccaaacacttggtcatcgaaaaacaatctggccaattcagtgttgatgacaccttgtttacgggtgtcctcaaagttgtcgattacgacaagtac aacgccgcgctgctatcgggcatcggcaacacggcacgcgccttcgggttcggcatgttagttatttaa
	KOON_mkCas0162 Csf2 (SEQ ID NO:461)	atgccaaagcaaaccatcgttatcgacaccgttatcaccacacaaggcccgctgtctatcgccatgcccgtcgcacagggcacacgggcaaacgact acaacaactttccattgatggcccgtgggcaggatgaactgggcgccgtattgcaaacagcctacctgcccgctagtacggtgcgtgggttccttcgcc gtgcggctggcattgccgcgatggaaaagcgcggcgtaggcaacaccacattgcagcaagcgtattccgatattttggggcagtctaaagacgccaa agaaacggtggacttggtccagctgtctgccatacgcgaggcagacccctactcgacctctttggttcgtggtctatcaagtcgcgcttactggtgagt aacttcttgcccacaaccaacgtgttgccacaggccatcacgggtgtcagaaaagacttagaagacaccgaaggcgtgttgaccctcttgagtgagac cgaccaacaggcctattacaaccgcagtgacatcaacagccagcgcgcagccgctgacgcagtcgtcaagggcctcactcgtgacatcagaaaagc caaaaaagacggcaacgcactcactgacctagaggcggcgcttgtgatcgccgacgcgaaggccgtggccttaaaaacggatatgggcgagatgg cgaactcatcacgcaccatcaccgagttttttgccttgcccgctaacatggccttgcacggacgcatcgtggttgagcaagccaaggagcgtgaccttg cattgatcgtggaggcgctagacaccctgtctattcgccccatgttgggcgcccaagtagcgcgcggctgtggcgaggtggcgggcagttcaccat caaggtggatggcatcgtggtcaagacagtcaccacgggcggctgggtgcctgcgcaggtcgaagacttctcgtctgatgcgcctgcaccagccgct gtctaa
	KOON_mkCas0093 Csf3 (SEQ ID NO:462)	gtggaccagttacgaatcaccgcccacctcagcggcccccttgtgacgggggcggctacctcacctttgatgcgttgctcgcctctgccacttacagc ttgacgcaagacattgagcaggcccacaaccagctgccacttaaacaagttgccggcctgtaccacgccagcgccgccattttatgaacccgttgccat cggggcattggctattgtgcaagccatgcgtcccgacgacctgtggcttgaccaccagtggttaaagaagaacaagcgcggcgctgtgcacaccaag ttcgacaacctatccgacaacatactgaccacctacaagaccatagcggcaccgaccatgacctggtactgcgaaggtgatgcagaccgcattcgtgt cttactagcgcatctaccctcattggtaaaaagcgggtttgctccgtgacccagtggaccgttgaggcgggcgagcttgacggtctgcatggctatgc cgatgagccactgcggcctgtacccgttgagctgtgggatggtgacaagagcagccccatcgtggacgcggcatggcgtcccgcttattgggatgcg actgggcgcaccgcttgttttgtgaactga
	-> DinG Helicase (1,444)[631.5] | DinG HNH (SEQ ID NO:463)	atgtttcgttcctcagagcagcagcatttctgcaatgtgcttattaagcatttgaaacgcgtcgatgcgcccatattgctggaaggcggggtgggtcttgg gaagacgcgtgcctacctggccgcactggcaactacgacgcaaacggttgccattgtgttacccacacaccaactgattgaccagttgctggctagcg ctgacttgtcagccgttgccttgccattggccgtgttcaggccagccagtcggtttgacacacgggaggcgtaccaagcgcagcgagaggcggcccg tgctgcgcgtgtcatgctctgcaccgctgcaagcgtgattattgaccaacgccttgcgggggcttacaacggcgcaacacagcgtgatgtggtgttattt gatgaagccgaccagttacccgatatggcggcattgcagagcgactttgtgattgaggcggcgctactgcagggccagccgctgcacgcggccttag agcgcataagcaccacccgttcacagtcggttgaggccgagcaaaaggctgctgcgcgggtcatgcttgagatattggccgaaccggtgggctacg catccgtcgggcttgatgatgatggcaatgcgcggctacaccacgcccgccaggtcgcttattaaagaaaattagcaatcgcggtagcacggtgtttct ttctgcaacgctgacccacaacgataagtttgatgacttcaagatcgccatgggcctagacgcggtgagtagcctgtcaacaagcattgagccagaac agcatggcacgctcgacttcaccgtacacgcgcacgaggtggacacgcctgactggtttgcggccaccgttcagtgcatacaaagcgcgcagcgcc ccacgttggtggtcacacctagccatgacctcacagagcggctacaaaccgccgtgggtgagcaagcagacttgcttatacgtgcgggtgcatgggc gggtcttgacacgccaacgcgctggcgtaccgtggttattccgcgcgtgccatttggcgcaccagtcattgttgacggggtggcggtcacgtcttacttg gccgccaaggtgacggccagtcggcggatgcgccaagcaattggccgtgggcttagaacacctgaggcggtgtgcacggttgtggttttggatgcg cgggcacacaacttaggtcgctttgtgccagagcgttttgcggcagattggtcgtctcggcaagtgtttgaggagggtgcccgccatgcggtggtgctg tccaaagcggagcgcgacccccgctttggccaaagccgcgctcagacactacggcgctcaatgccaacaagcgggttgttccgtgacagaggcgtgt cagctggacgtgcaccaccttgaccccgtaagcgaaggcgtgcgccgcacaacctgaaggatgtgaccgtgctgtgtgcaaaccaccaccgacag gcccaccacgacatgaaaatggggctgaacgtgtcgtag
DinG_HNH_3_209632_10012131_organized_PAM	DR (SEQ ID NO:464)	atgcagccacccctctcaaggggtggttactacaac
	KOON_mkCas0093 Csf3 (SEQ ID NO:465)	atggaaaaactaaccgtaacggcgcatatgaagacaccgttcataacaggtgggggctacatgaccttttgacgctttgctcgccggtatcttatttgacc agctgcaagatgttgaagccgcgcacgcctccgtcccctatcaaatgcactgacggcttgttccatgcaagcgcggccatgtacgaaccttggatcagt caagagttgcgttcgttgcaaacatgagagccgaccactctcttgatgcagatctgtttccacggaagaaggacggaacaatgcgcaagatgggccga ctaaggcggcgtgactttggcgctgtcatgaatagctacgattcaaaacaagttgattcacttgaatggcatgttgaaggtgattccgaggagatattaaa actaattgagccgatttcttttattggcaaacgcagagcaagtggctttggcgaagttcatcgttggagcattgaatcttctgaactggatggactcatggg ctacctgaacgagcctttgcgcccaatacctgttgagcgtttcaagggtaatacaggctcactcaaagtagatgctgcttggcggccggcctattggcat ccagccaacagagcgatttgctacgctccggagacacctcaatga
	KOON_pfam08798 Cse3 family (SEQ ID NO:466)	atgcttcgtgaatatgaaataaacaaaccggccaatttacgtggataccccttgcatagaatggtggctgaattaacaaagaatgagccaagcctcttca gcgatgaaggtgcgaccatgcttgtgcgaactgaagcaacaattgacttagaaccaaaggctttgaaacagataagccttgggcagctaacggcattt gagctccgcgcatgcgtttctaaaaaacgtaagggcaaacatatttacccgcctagcgacgactggagagtgagacacgaatggctaaagaataaag ctaatagcaacgggtttgagatattgaccctgcattcttcatcaaaaatggtgcggatacaggaccaagcacgttccttcaaagttgatcaaactgacttta cgggaattttaaaagtcacagatgaagcccaattcagaatggccctagctaaaggaattggctccactgccaagacttttggttttggcatgctagttattt aa
	KOON_mkCas0162	atgatcgttattaaagcaagtattaaaactgaagcacccctctcaattgccatgcctgtggcccaaggatcccgagaaaatgcgtacggcaactttccatt aatggtacgcggcatagatgaagagggagaaaagcaatacaccggatacctgcctgctacaactattcgaggctttctgagacgaggcatcgtgttgg

	Csf2 (SEQ ID NO:467)	aggaaatgaaagaagcggcggaaaacaatcaacattacaaactgaaccaagcctatgctgaactgattggtcaggatgccgcaagcgaacaagggg gcggggatatagacctcgtcgctttaaaggatacgcgtgaaggatcacccatattagatttgtttggctcgggactggggatcaaatcaaggctgcgag tcagtcatttttatacccctcgcacaatctactgcctgacccgattacaggtacaagaaaagatttagatgataatgaagaagcctttgacctcatagatgcag atgaaaaaagtgagttcttcgagcgatccgaagccaacagcaagcgcgcaagagccgcaagccagctggacggattacagcgacaacgcagggc cgctgaacgcaagggcgaagatacaaaggatttagccgaacaagtggctgctgcagaagcgctggtgaagaaatacgaggcggttatgggcgatat gaagaactcgtcacgaacaatcctaacgcattacgccttagccgctggtttagagctagagggtaaattggtcatagagaacccgaaggaacgcgaca tcggactgcttgaaacagggttaaatgccttgtcgttattccccattcttggcgcgcactcagcaagagggtgtggcgaaatatcaggtacatttgattttat taaggatggcgcattgtttaaacgaatcacaataggaggctttgaagctgcaagagttgatgacttttag
	-> DinG Helicase (4,478)[689.7] | DinG HNH(480,5 (SEQ ID NO:468)	atgactgttaacttccgatcaaccgaacaagctaatttttcaaagcttctacaggctcacctatctaacccagatgcccctcttcttttggaaggtgcgaccg gattgggcaagacgagggcctatctttatccgttgttggaagctgctaacaacggtaagaaaatcgcaatcattcttcccacacgtcaactgattgagca gttgctgaactcagaagacattgacgtatgcaataaagccggcattgagattcaagcgttcatcccgaagcgttactttgaagatgcagatcaatacggc gagcagcgtcaaaaagctttaaatgccccggtgatgctctgcaccagcgcctcagtaatcattgatcaaagactgggcggtgattacaacggcgtcac caagcgggattacttgttatttgacgaggctgatcagttacctgatgttgcagcgcttcagtctgacgtgacaattaactatcaaacactgggggatctcaa cataaagccggtatccgctcagcaagcagcagaagaacttctgaaaaagaaacacgttgagccagaagacagggccaaagctaagattattctagaa gcgttagctgagccatttggtttatgtccgccggccttgatgatgaaggcgggattgttctgtatcacaagctacccgggagactccttaaacggatcgc taatcagcgaaatactgcattcgtttctgcaaccctgactattggtaacaagtttgacgacttcacgcgagctatgggaatcagtgaaatcagccgtcattc atcagccgttgagccaaagctgcatggctatttaacctttcgttgcaccagtcggaggaagttaatagtcccgaatggattcaggcagtgattgatgcaa tcgacgaagctcaaaaacctgccttgttgttacctctagctacgagacagcgactaatcttgttgagctatatccaagctcaattgcgaggaaccaagaa gaatcagagacagctagccaagcagctatacgtttagccaatgatgattcaaaagatacccttaattgcgacagcggcttgggctggcttggacacgcc cgtggtatggaaatctattgttgttccaaagatcccgttctcaaaacctacgattattgatggtcaggttgaaagtcactatatcaatgcaaggaacgcagc catacgacggatgcgccaagttatcgggagaggtttacgaagcccggatgccgaatgcgacatcattatttgtgatcctaggtggcaaaaaatagagtc tttcgtccctgatagattcaaacaaaactggtccaacaaaaccttcagcgaaggcgatcggtatgaactgactctttcggcagttgaaagaaatccatctg ttaggaggaatgcaattaagcattatggcttgaagtgttatgcctgtgacctcacgcccaaaacatcaaatcaaatagacgttcaccatcttgacccgattg ctgaagggcaacgacagacaataatggaagatgtgatacccgctttgtgcgaattgtcatcgcttggcacactccgagcgaccgcctctaacaatagag gaaatcaagagcgtactataa
DinG_HNH_4_ 0206130_1001 2455_organized _PAM	DR (SEQ ID NO:469)	ctgaagctaccctcgttttgggtggttactgaaac
	KOON_mkCas0093 Csf3 (SEQ ID NO:470)	atgcagtgtgtagaaataaaagcaattctcaaaacgccaattattctaggcggaggttatctgacgctagatgccctttttagcatctcttatctttgagaaat gtggcgatgttgaggtagcccataccacggttccaattaagcaaaccaacggacttttttcacgctagcgctgcactgctagaaccaatttcaagaagttct gtctcatttgttgccagtctcagagcagagcatgacctcaaccctgatttgctaaaaaaaaagagcgggagctctaaaaaaaaataaagcggttacgacgg agtgattttggcaacgttttaaatacttatgggtctcttcaagccgctcatatatcatggttcgctgagggcgataaagatgaaattagccagcttcttcagg acgtgaagtttattggaaagcgtcgcgccagtggatttggcgaaattgaaacttgggttatcgaggacagcgacatagacggtgttgtcgggccgtttg atgaacctctgcggccagttccagttgaaatgttcattggtgatcagtcttctctgaaagtagatgccgcttggaaaccagcttattggaaccctgaaaaca gagcgatctgttatgcgccggagctgatttcatga
	KOON_pfam08798 (SEQ ID NO:471)	atgctgcttgatcacgtaatctctaaaccaaaaaatattggaggctacgggctccaccaactcgttcaagggctaacagatggggaaagccccttatttgt ggatatgggcgacaaatttatagttcgcactgtcaaaacaattactgaaactggcaccccccaaggtccgtagcagataaagacgttatcggttttgag ttacgagcttgcgtcagtaaaaaactgaaaggcaagcacatatactatcccacttcagattgggctagcagacataattggctaaaacgtcagggcaaa aagcacggttttgaaccactcacgatgaattgtcgctcaaatcttgcaaaaataactgatggaaaaggtcgcaagttcacggttgatcaaactgattttgtt ggggttctaaaagttttggatgccttggcttttcagaaagctgttggacaaggcgttggttctactgccaaagcctttggttttggaatgttaattatttaa
	KOON_mkCas0162 Csf2 (SEQ ID NO:472)	atgggtgacataattatataacggagcaattactctggtgggaccgttatcaatttgtgcgcacaacttagatgactacaacgggtttcccataatgtctcgg ggcactcctgaggatggcaaacttttcaaaaactggatacctgcccgcgaccacaatacgggggaaattgcgtcgcgcttatgttttttgaaaatatgaaa gaagctgctgctaaaaatgagcattacccacttagacgtgtttatgcagaactgattgggcaggacccagatagtgaaaaacaggctggtgaaatagac ttacttcaaatcaaaaaaacgcgggatgactcaccggtgcttgatttgtttggcagtggtctcagtgtggcaagccgtttgagagtggggcatttttttcca agtatgaatattcttcctgaggaaattggttcgataaggaaagatttggatgaaactgaagaagttgttgaagccctagctccagaagacgtgcaacgttt ctttgagcggaaagacctcaccagccgacgttctcaagcagttgcgctagtaacgtcattagaacgaaaaattagagctgttaaaaaaacaagtggccc gacaaatgaacttgaggtggagctcaaagaagctgagggtctagttgaaaaatacaatcaaggcatgggcgacatgcaggttggcgcaaaacagctt cctacctatttcgctctgccacccggattgatattgtcagggcgcattgtgatcgtaaatgcgaaagaccacgatctggccattattgaaaagggtttggat tgtatgagcctatcccctgtgcttggagcacaatccgcgcgcggctgtggcgaaatagagggcagttttgatgttcaaatagatggaaaaactgttaaaa aaatcacgataggaggctacaaatcaagcagaattgatgttttttaa
	-> DinG Helicase (1,449)[600.1] | DinG HNH(451,4 (SEQ ID NO:473)	atgtttcgctccaatgaacaagaatatttcaatacactgcttgagcgacatatgtctgttctggatggacctctgttgcttgagggaggcactggtctaggta aaacccgcgcttatcttgctgcaataaaaggttttcaaggcaaagtcgcaatcgttttgcccacgcaccagcttatagaacaacttatcatttcaagtgattt aaaagcagttggcattgaaattaaagccttccgtccaagaccgatgtttgacacacaggctgactatgatttgcaaaaagatgaggcgtcagcggctgc ggttatgatttgcacatctgcaagtgtaataatcgatcaaaggctttctggtggttacaacggagcgacacagcgagactatattcttttgatgaagctga cgctctaccagacgctgctgcattgcagaaagatttggaaatctcagcatttgaaatcaaagatgctggcgtaaaaacctctgtggtcgcaaatgctttgg acgacttcattgacaaaaccgccgatgtagagctgaaagcaaaagctcgtataattaaagaagcatcggaaaacccggcttggtaccatacgtgggg tctacagaggagggcggcattagcctctttcaccgaatgcctggcttgctactcacccgaattgccaatcagaaaaacgttgccttcatcagtgcgacctt gaaagtcgcaagttcatttgatgacttcaagcgcagcctagggattaaaaaggaaagccgcctttcggatagtatcgaccccaaaatacacggtgatat gcttgtcgagtatccacttttctgagaatattgtgactgtaatatctggcgcaccaaaaccttgtctagttgccacgacaagctttcagtcttccgaagagctc gctgcattaatcccaacggctgttttgcgcaccagagatgaaacgcccagtatggctgctttgcgggtctctgaagacggcgttttgatcgcaactagtg catgggccggtctggatactcccagacgttgggcatcggttgtcgttcccaaagtgccttttcggcaaaccggttgagcaagacggtcacatcgaaagc cggtatttggatagtcgtaatgtggcaattcggcgaatgcgtcaggtattaggccgagggcttcgaacacctgacgcgaaatgcacattctacatctgcg actccagatacaaatctcttggaaatttcttacctgatagggttttctaaaagctggtgggagggggtaaaaaaaccattgagctttcaaaatcagaacgct gtccggcaattagaaaggccgcattaaaacattatggatgttttgccaccgctgcgatcttaggccaaaaaatccgacaattatcgaggttcatcatttatt tccaatcgctgatggaacccgtaagacaaaaatagaggatgttatccccctgtgcagaaattgccatgagcttgctcatacagaaaatccaccgatacc catagatcgattgaaaacgactgatttgtag
DinG_HNH_5_ a0187845_1001 177_organized	DR (SEQ ID NO:474)	atgcaagctgcccttttgtcggggcagctactagaac
	KOON_mkCas0093 Csf3 (SEQ ID NO:475)	atgcaggctttcgtgctgcgcgcgcatctgcgcacgcccttaattcagatgggatacaccactctggatgcgctgttgatggcgacacttgagacaggt gatgtcagccacctgctgcgctgtgaaacgggcctttactacgccagtgccggtttcccggtggatggaattgccggccccaatgccgcattcatcgcc agcatgcgtccggagcacaccccgagtggttggatgtaattgctcccaatacgatgaagggcgatgtgcaaattggcctcgcccgtcagcgtgaag gcggcaacgtgctgaatagttatcccaccacgactgcgtcagccgtggagtggtatgccaccggggacatggcagcagtgttggaggtcgtcgcccc catcatgttcatcggcaagcgacgtgcttcgggctatggtgaggtagttcgctgggaagctgagcccggggtactggatggcattgttggtcacctggg tgagccgcttcgccccatccccgtcgagcggtggactgagggcggcgattgggtcccggtggaggctgcgtggaggcctccgtactgggaggtgc gcaaccgaaccaagtgctacgtccccgaggtggctctgtga

	KOON_pfam08798 (SEQ ID NO:476)	atggacatcctgatcgacaagcctccgcgattgaaaggttatgcgttgcaccgtttggtcgagcagtaccagcaagggcgccccgcactatgggctga cgaagggcaacaactacgcatccgcccgcgtgatgcgcaagctccactctttgaagttagcaagctgctgggctttgcgaccacggcctgtgtggtgc ccagtacccacgccaaacacaaatacctgccgctcaacgattggcgcggccggcgtgagtggcttgagaaacaagccggcaagaacggctttgag ctggttggggtgcacgtctcaccaggcatgcgccgtattgagacccacgatggccgcagcttcaccatcgatgccacggagttcacgggcattctcaa agtcacagagccagagaggtttgccgcgtgcctggtgaaaggcctgggccgtgtcggcaaagccttcggcttgaacctgcttgtagttcagtaa
	KOON_mkCas0162 Csf2 (SEQ ID NO:477)	atgactaaccgcatcatcggcaccgtgagcctagccaccgtgcagcccatcgccttcacccatcatggcattgagggccagccactcatgacgcgag gcgttgatgccgaagggcgccccttgcgtaccgtttttctaccggcctctcagctccggggccgcattcgacatgaggctgcaatggccgcactgcac agcaagggtgcgcctgtgaagcttgaagaggcctatctcgctgcgctagggcaggatctgcgccccgaagaagatcgggccgaagaactcgttcgtt tgcgtgagcagcaggtgctgcgtgaagctcagccccctgctcgatttgtttggaacctggaaggttgccagtcgcttgaaggtcggccatttgctgcccg atgtcaatgtcgctcccgatagtttcagcatcatccgccgcgacctcgactccaatgaagaactgatgaacttgctgccagacgaagagcaagaccgct ttttatgcccgccaggatagtcagtccatggcctccaaggtggacggcatgatcaaggtggccgagcgcgaactgcgtgccgccaagaaggccaaga acgaggctcttactgctgagttgacacagaagctggaagaactgaagaccgccaagcagtctcacaagggtgatgatgaatcctccaattccaagcac ttgctcagcgtcgaagccatccctgcaggggtcaatctgagtggacgctgggtcatcgagcgagcccgtccgcgcgatctggacgtgatcgttgctgc cttcgacgggcccagccggcgccctgtcatgggtgcagccacagcccgtggctgcggcgaggtttccggggcggttgtgttctctgactctgagggc gaagtcctgcttcgcatcaagtttggccagtacaccccggccatcgttgagtggacagtgtcgggcaaggccttcatgacccagcccaggccgcatg a
	-> DinG Helicase (2,461)[605.1] | DinG HNH(462,5 (SEQ ID NO:478)	atgatctatcgcagcaaggagcaggccgcttatgcgaagctgattgcaggcgcgctgcgccccggctcccctttgctggctggcggtgcggctggcc tgggcaaagctcatggctacagcatcccgcttctggcatcgggcctgcgtgttgccattgccctgtccacgcgccagctcattgatcagttcgtgagca gtgatgcccttgccgaggcgaatagggggccggcaaagagtggtcatggccctcaagacgcgggcccatttcgagaccgctgctgactatcaagccc accgagcccaggctttggcggcagatgtcctcgtcctgacgcatgcggctgccttgatcgacagctggatccccggttatgccgagctgcggcagcg cgacgtggtgctgtttgatgaagctgatctactggccgatgcggcagaccttcggggcagcttcaccatcccggaagatcaactgcgtggacgtgact gtgagcccagggaagccgccacctatgtgatgcagcacaccgatgacccggaaatcaaggctgcggcgtcagcgattttgtatgcccttgatcaccc cacgccttacaaggtcgtgggtcgagatcttgagggcaacctggtcctcaagcaccgcatgcccgggcgcatgctcaagcccctggtgcgcgatgtg gccagctgcatcttcaccagtggcacgttgatggtcaatggccgttttgactactttgtgcgggccttgggtctgtcgggcatcgacccagcctcgcgcc acattgatcccctgactcatggggacctgaacatcatctgtgccgagagcgagatgagcccggagcaaaaggcggcgcagatttcgcaggctgcgc gccccgcgcttgtgctcacttcaagccacgaggatacccaggcgctgggggtaacttgcccctggccacggtccgcatgccgggtgaggcactgc tggacgctctggagcgatgccctgacgatggcatcctgattgcggcaggtgcatggtcgggccttgacagtccacgccttcgctggtggtccgtcatc atcccaaggcgccatacggccccccagtcgaagtggatggacagcaacttacccattacctagacagtcgcgtggtccgccgtggcggcgcatctgcc aggggctgcatcgcggtttgcgcaccccggatgcccattgcacgctgatccttcaagatcctcgctgtcgtcggcccgagctgatggaagccatccctt caaggttcaagggagcgctgatcagtgccatggaaggcgaagttgaggcggccctgcgcaatgagcgtgagcgacatagaccccttcaggaccctg gcgctcaagcactatggatgccagtgcatgtggcatgagggctgcgatgtcaaagcgccttatttgctggatgtgcatcaccgcgatcccgtctcagaa ggggtgcgccgcaccgctctgcaagacctgatggttctttgtaaaaaccaccatgctgaagcccatcagcagatgcggctactcattttcaatggagga gattcaacaaaaattgatgatggatcaatgatttag
DinG_HNH_6_a0209199_1004312_organized	DR (SEQ ID NO:479)	gttgtggcccccattaatttgtggggcaactgaaac
	KOON_mkCas0093 Csf3 (SEQ ID NO:480)	atgtttgaacaagtgaaaattgaggcaacccttcgtactccgattattattcaggggtatatgacatttgatgcgctattaggtgcattactgtttgagcggct acaggatgtcgacaaaacgcactccagcattccaattcaatgtgatgatgggctgttttatgcttctgcagctcaggttgaagtcattgatcgtggtggtttg gcaattgctgctaatctgagggcccagcatgacttgaacccaaatcatgtaaagaaaagtcgtgatggctcaaaactgcacaaaagtttaggattgaag cgccgtagacagtttggagcagttctaaataaatacagaacaatcactgtaaattctatcagttgggatgtgactggcgatattaaccagattgaagacttg cttaatgatgctttttttatcggcaaaaaaagaacagcaggttatggtgaagttataagttgggacttcaagaacagtgaaacagacggtctcatagatgc gcaaggacaggcactaaggcctattccggtagaacgttttaatggtgataagtctttgccaataattgataccgcttggaagcctgcttattggaatcctga gaacaggtcagcctgttatgctcctgagcctgtctcatga
	KOON_pfam08798 (SEQ ID NO:481)	atgctgtttgattataaaatattaaaatctgatgcaaaaaatgcctatgacttgcatcagcttgtcgctgccacctttaagggtgagagtggcagaaatttgtt ctgtgattacggcgaccacatacgtgtgcggagcacaaagcacactctaaatgcgaaacaaggtgtcgccgttactgtgcctgccgtcggcgcaatag ctttcattgaattgcgagcatcctgttatgtgtctgggggaggaaaaaaatatttcctgatgcaaggtgattggagagccaggcatacctggcttgaaaaa aagagcgcccaaagtggcttttcggttttaaatgtaacctgtcaatcagaactgatgaaaatcccaaagccaggtgcacagttcacgcttgattgcaccg actttgctgcatgcatcaagatcaatgacatagaaaaattcgaaactgccttgcgcgatggaattggatcgaaaggcagagcgtttggttttggaatgctt gttttatga
	KOON_pfam03787 Csf2 (SEQ ID NO:482)	atggtaaatacgatctttaatatggttctaacatctgagcagcctttgagtatcacgctacccgtggccgagtacactcgggcaaataaatttgctaactctc cagttatgacccgagggctcgacgcggaaggtgcgaagcaagaaacatgttacattcctgctacgacttttcgcggaatgctcaggcgtcatgtggttg tcccgctgatggaggagcgagccagaaacggtgatcactatactttggatgaagcctaccgtgatttgattgggcaggatgctgccagcgaacaacag ccagatgaaattgatctgatggctttaaaaaatatgcgtgcacaaaatccaattgtcgatttgtttggtctcggtttgggtgttcaatcacgcctaaaagtca gtcacttttcttccagagcataatattttgccagaggtatttaccacccgctcgaaaggacttaggcgatgcagacgatggtgcaatactggacttggttcgc gatgaagagagagaaaaatatctcaaaagagccaacgcaaattcaaagagatctgctgcagaaagccttctggcgcggctggtgaagcagaagaaa aaggacggttcatcagcagacttggatgctcaaatagttgcgacagaaagtaaggttgaagagtaccgtgatgcgatggatggcatgaccaattcttct cggaccatacttacgcattacgccatgcctgcgggtgtacagtggacgggacgtttaattattacaaactccactgaaaaagatatcgatgctctgtgtgg agccttagatcgtttcagtagctatccgatgattggtgcccagcatgcccgtggatgcggtgagattagtggtattgctaatatagaaacagatagaagg ctggttaaaagaattagtttttggtggctttgcgccagcacaaatagacaacgtagcctag
	-> DinG Helicase (4,483)[573.6] | DinG HNH(485,5 (SEQ ID NO:483)	atgatcagaatttttagatcggaagagcaagatcagttttttaagctgatccaaaaacatatagaaacccctgatgcaccttactgctggaaggagccact ggactgggtaaaacgctgccgttccttgttgccgcagcggtaagtggtaaaaaggttgcgatagttttcccaacccatcagctgatcaatcaggctttgg atagctgggacttaaaagcaacacttgcggcctatcccgcagtgattgtgaaggcctttagacctaggtcatatttcgaagcctcacctgcagattacgc ggcacacaaacaagaagcgttagacgcagacataatgttctgcacctcggtatccgtgatcatagaccaaagactcaatggcgattataatggctcaac agacagagatgtgattatctttgatgaggcagatcagcttcctcaatttgctgctttgacctctgatatttcactcagcagaattgatctaagaagcctcaga attaatgaatcttcgccaaaagagaccattgatgcaatcttcaaaaagaagacattgccgtcagaaatcagggcaaaggcgcgcttgatagaagaagc gttaggccttggcgatgtttggtttattaaggtgggttttgcggatgatggagccttagaggtgcgctcacgcttgcctggaaggctgttaaaacgcgtatc caaccgagcatcgacaatttttataagcgcaaccctgtctgtaaatggtacatttaaagacttttgtcgtgcgatgggaatagaagagaagagtagactgt ctgcaatagttgagcccaagaaacatggggacctacaatttgattttataactgatgaagagcttggaagcgatgcttggctggacgcagtttgtgaagc aattgaatatgccgatcatcctgtactggtcgctacaccgtcgcattctttagctcaacagattgcaagtaatgctactcaggcgacctaagagccgcgg atgaaacaacgtcgcaagctgctgcgcgtatgggagatacaaaggttctcatagcagcgggtgcatgggcagggcttgatacgccggtacaatggaa aacggtttgtggtacctagaatacccctacactgggccaaacaatctcagagagttttggtatgatgatgaagaggacttttggcgtccagaagccagatcc gatgtcatctttttatagattcaaaaaatgctgcggctcgtaggctgaagcaggtctttggcagaggcttaagaacacccgatgccgaatgctcaataatac tttgcgacccgcgaattgatcgttttaataagattgcaccagccagatttcagaaaacgtattttgagggaatggccatttcagttaatgcaacaaaatctga acgttcgcctctgcttagaaaagcagcactcgctcattatggtgaaacctgcatggcctgtgacttttattcctgagtctttaaggcaattagaaatccaccat cttgagccccttgctgatcgtgggccaagccacacgaccctgaatgacgtcgtggtattgtgcagaaactgccacggtttggcacatgacactaatccg cccatacagttgccagagcttaaattacgatctcaaaagttgaaaaagaaaagacctggattgtccatgatttaa

DinG_HNH_7_a0335030_0007884_organized	DR (SEQ ID NO:484)	gatcaagccgcctgttagatgggcggttactggaac
	KOON_mkCas0093 Csf3 (SEQ ID NO:485)	atggttcctttcacattgagggcacggctggcagcgcccatcatccggcgtgggttcggcaccttggatgccctactgatggctgtgcttggtcgtggg gacgtttccgacctgctccagtgtgaacaggagttgtacttcgcaagtgccgggatcatggccgtcgggggcgcagcacaacccgcctcatttgtggc cagcatgcggcctgagcacacacctgagtggcttgacgtcgtggcgcccaacaccgtcaataccgatctgtccatcccggatggtgccaagggcgt gcccgttacaacgacctgaggatcggcctgtccaggcagcgcgagtgcggcaacgtcatcagcagttacaccgccgtttccggtggttccgtcgagt ggtacgccacgggtgatgcccaagctgtgctcgatgtcgtcaagcacgtgccattcataggcaaacgccgtgccgcaggctatggagaggtggaag gctgggaggttgaggagggcgtgcttgacggggtctgcgggttcctgggggaacctttacggccagttcctgtttcgcgctgggtccatggcggtgac tggataccagttgaagccgcctggaaggccccgtattgggacgtccgtaaccgagcgaagtgcttcgtgcccgagatcgcatga
	KOON_pfam08798 (SEQ ID NO:486)	gtggtggacctcctgatcaacaaacctgcaagcgttcgcggctacgccctgcaccgtctggttgcagagtacctccaaggccgccctgcactgtgggc ggatgaaggtggccaagtccgaatacgaccgcgagacgccacgcctcctgcgtacgtcgccggcacccctgctgggttcaccgtccgcgcctgcgt ttcacacagcaccaaaaaccgccaccactatctgcctgtgggcgactggcggcggccgccgcacatggctggaggccaaagctgcgccgaatgggtt cgaggtcgttgctgttcatgtcagtggtgatatggcactgatcgaggcacatgacggacgccgtttcacggtggacgcaaccgatttcactggcctgct gcgcgtgacagacccgcaagccttttctcgatgtcttgttacaggggtgggcaaggttggcaaggcgttcggcctgggcctgctcatcgtgcagtga
	KOON_mkCas0162 Csf2 (SEQ ID NO:487)	atgaatccggcattcatcctggccgatgtgctcatccatacggcccagccgctgacctacacccatcacagcgtcaaggatttgccgctgatgacccgg ggcgtcgatgcggaagggcgcccccaacggaccgtgtacctgcctgcaggccaattcagaggccgtttgcgccatgaagccgccatggcggcact gcgcgccaagcctgagaaggtcagacttgaagatgcatatcttttggctcttggccaggacctcgctcggcaagaagaagatgaacccgaggcggtg cgcctgggtgaccagatcaagttccggcaagacaacccgctgctggacctgtttggcacctggaaagttgccagccgcctgtttgtgtcccatctaatg ccacaggtcaatgtcatgcccgaccgggtcagccacatcaggcgcgatctcgacagcagcgaggacttgatgggcatgctggcgggtgaagagca ggaccgcatgtatgaccgacaggccaagcaggcactggcctccaaggctgggaccctgatcaagctggcggaggagaacgaacgggcggtacga aaggcaaaggacacggcgctcattgatcagcttgaatcgaagctgaaggaacttgaagaactgaagaagcagcacaaggggaagacgactccga caacaccaagcacctcgtggagttgcaggtaattccggcaggaattgatctgctgggcaagatcaccgtacaagcgccccgacgaacagacatcgg gttccttatccaggcacttcagggtgtgagcatgaagcccttgatgggtgcccagcgcgctcgtggctgcggcgaaattcgaggtcgcgcttctttcgt accccagatggcgaagtgctgacggcagttgagtttggcggctttcagcctgcgaaggtcgactggactgcggagggccaagccttcatggggcca gcggcagcctga
	-> DinG Helicase (2,464)[653.9] | DinG HNH(465,5 (SEQ ID NO:488)	gtgaacttccgcagcccagagcaacaggcctatgcccgcctgatcgaaggtgcgctgaagcctggaccacccctgctggccggggcagcaccgggg tctgggcaagacacacgggtacaccttgccccttggtggcaagtgggcgccgcattgcggtggccatgagtacccgtcaactcatccagcagtacctgt ccagtgatgccatgcgcgctgcacaagcgttgcgtcctgcgagcgtggttgcgcttcaaacccgccgagacttcagcaatgacgcctcccacagagc gcacagggctgcggcgatgcaggctgacgtactggtggtcacccatgcggctgcgctgatcgacagttggcacccaggttacgcagacctgcgttcc cgtgacgtggtggtgtttgatgaagcggacctgctggctgacgcagcagacctgcgcagcaccttcagcatcgatgcggaagttcaggaggaatgcg gcgctgacaaactggacctgctccaggctgtgcaggcggtaaagcagaaggccaaggcgacggaggatcgtgcagcggcgtcggccatcctctat gcactggagcatccggcttggtacaaggttgtgggccgtgaagaagacggcactttgacgctgaagcacaggatgccggggcgaatgctgcgtcgtt tggtcgaggaggcgccgcgttgcatcttcaccagcggaacgctgcaggtgggtggccgtttcgaccacttcgtgcaggcagtcggcttgaagtcgat ggcgcccgagagccggcacattgacccggtaaagcatggtgacctgagcgtggaactcgctgctgaagacatgagtccgcagcagatggcgctgc gcatccaggcagcagagcggcctgttctggtactgacgacgagtcatgccaccacacaaaccctaggcgagttgtgcacaggtgcagtctggcgtgc gcagggcgaaacgctcggtgatgcattgaaccgtgtgccagacgatggcatcttcatcgcagcaggtgcttggtcgggggttggatgcccgcgcctg cggtggcgtaccgtggtgattccgaagactccctaccagaggccagccatcctggatgggcagcaggtcacgcggtatctggactcacgggtggtg gctatccggcgtataactcaaggcctgcaccgaggtctgcgcacgcctgatgcccgctgcactctcctgctgcttgacccgcgcagcggtcgtgacga acttcgagaagccatacctgcgcgatttcgggtgaattggcagagttttgacgagggctcgccaagaatgcgcgcccatgtggccggtgaaaggcat ccctccctgcgtaacgctgccctgaaggcgctgggggcgcattgcagtcaccctggctgtgaagtcacagccgtgcacctcctggacgtccatcacct gaaaccagttgctgaagggcagcgcaaaaccagcttggctgacatcgccgtgctttgcaaaaaccatcacgccgaggcccatcaccagatgcggca ggccgaaattcttgaggtcaaggaggaagaatcctga
DinG_HNH_8_0105052_10007820_organized_PAM	DR (SEQ ID NO:489)	ggtgaagccaccccggaattgggtggctactgaaac
	KOON_mkCas0093 Csf3 (SEQ ID NO:490)	atgcgccgcttccgggtaactgcccacatgaattcccccatcatcttcgggggcggttacgctactcttgacgggctacttggagccctgcttttgatca actggacgatgtggatgcagcccatgcggcgatcccgctggccaacaccgacggccttttccatgcctcggcggcgttctacgaacccgtcgaggtc ggcaagacgacctttcgttgccagcttgcgggcgcaacacgatctggatgccgacctgctcaggaaaaacaagacgggcaccgcgacgcatcgggc catcagcacgaagcgcgagcgcgacttcggcaatgtgttcaatacttatcgccgggtcaccacgccgaagatcgactggtatgccgagggcgacca agccgccgtgctgcgcctactccaggacgcgcgcttcatcggcaagcgtcgggccagcggattttggcgaagtcgagaggtggaccatcgaagagg acgaactcgacggcattaccggccacctagatgagcccctgcggcccgtgccggtggagctattcaaaggcgaccagagcgcgatcaaggcggac gcggcctggcgacccgcgtactggaaccctgcacatcgggcgatctgtttcgtgccgggagacggccaatga
	KOON_pfam08798 (SEQ ID NO:491)	atgcttgaccacgtcatacgaagcccgcccgcgttgccggctacatgctgcaccaactggcggaagccatgaccgacggtgcacggcccctgttcg tggaccgaggcgaccatctcgtcgtgcgtaccgacaaggcgatcacgccaaatccgacccccacgcgcgcagcgcaaaatggagaagtcatcgcc ttcgaactgcgtgcctgcgtggccaagaagaacaaggggcacccccatctattttccccctcggggattggcgcgcgcgtcacgaatggctcaatcgcaa ggcaggacaaagcggcttcgaggtcctggcgcttcgctcaaccaccgcgatggcgaagatcgccaaggcaggccgcacttttaccatcgatcagac agacttcgtcggggtcctgcgcgtcaacgacgtggcgctgttcgagatcgcgatgaccaaggtgtcgggtcggtaggccgcgccttcggttttggca tgctccacatcaactag
	KOON_mkCas0162 Csf2 (SEQ ID NO:492)	atgacgacaacgatcatcaacgcgacgatccacaccgtcggtcccttttcgattgcgatgcccgtcgcggaaggcggcaccgccaatgcctacaaga acttcccggtgatgacacgggggtggacgaggatggcaacaagaagctgaccggctacctgccagcaaccacccttcgtggcttcctgcgccgtg ccgttgtcctgcgcgacatgaagaaggcggcggcggccagcctgccctacacgttacagcgcgcctataccgaactgatcgggcaggacgccgaa agcgagagggcggaagacgataggatcaacctcgttgaattgcgcaagacgagagaggcttctccggttctcgatcttttcggtagcggattgggcat caagtcgcgcctgatggtctcgcacttcctgccgatgtccaatgtcttgcctgatgcgttcacgggcgtacgcaaggacttggaagatacccgaggcgt tctggatatggtcaaggcggacgaagtcgaggcgtacctcggacggtcggaggcgaacagccggcgcgctgccgccgagggcctgctgtcgcaa atgaagcgtaagctgaacgcaaccaaaaggaagggcgagcctaccgccgatctggaagcccaggtttccgtgcttgagacgaaagtcgagggcct gaaggctgagatgggcgatatgcggaacagttcgcgtaggccggttggccacttttgcactgcctgccgggatcgatctcaagggacgattggtgatc aatagctcgaaggacatcgaccttgaaatgatcgagcacgccttgaacgaactctccttgaagccggtgttgggcgcgcaatcagcgcgcggttgtgg agagataagcggcacgttcgacgtcgtggtggaaggcgtcttgaagaagaggcttgcaatcggcggctacgaacctgcccggagtatcgatttttcct caacgccagaaggagaaaaggctgcttga
	DinG_Helicase (SEQ ID NO:493)	atgttccgatcatccgaacagcaaaacttcgccgcgatggtcgccgcccacatctccaaggccgacggcccgttgctgctcgaaggtggcaccggca tcgggaagacgcgcgcctatctgtcggcgctcatcggctccggcaaacgggtcgcgattgtgttgccgacccaccagctcatcgagcagctgcttgct tccaccgatctcgcagaggccaatgccagaggcatcaccgtcgccgctttcgaccggcgtcgaagttcgataccagggccgactacttggcgaacc gtgaagtcgcggccagtgcgcaactcatgttgtgcactgccgccagtgtcatcatcgaccgccgtctggacggcaaatacaacggcgcgaccgaac gggattatcttctcttcgatgaggctgaccagttgccggcgatggccgctcttcagtccgacatggaggttaccggcgaggccctcaagacgctgggg atcgaaggcgaaacgacccgtaatctggccgacgccgtcgtcgcgtccaagaaggccgagccggaacagcgcgccgccgccaagatgatcctcg

		aagtcctgggcgagcccgcgtggtttcatgtggcaggcaaggccgacgatggtggcatccgcctctggcatcggctaccggggcggctactgcgcg agatcgccaaccaaccaaacgtcgcctttgtcagcgccaccctgactgtcggaggtctgtttaacgacttcaagtgtgcgctcggcatccagagtgaaa gcgagctgtcgggtgtcgtcgaacctgagcgacacggcacggtcgagttccggaccgttgtgtcgcaagccgtggacgccgacggctggctcgaa aaggtggcggcgaccgtcaggcaggcagaggctcccgtgctggtcgtaacgccctcgttcgaccgtgctgagaagattggcgcgctggtggaagg cgcgacggtgcggcagcgcgaggaagccgttggcgacgccatggctcgaatgggcggctcccgaaccctgattggggcagccgcctgggccgg cctcgacacgcccatccagtggcggtcaattgtcgtgcctgttgtgcctttcgagcgcctgacggagatcgaagggcaagtcgagagccactacatcg atgcgcgcaacacggccgtgcgacgcctgcgccaggtgatcggtcggggactgcggtcgcccgatgcccgctgcaccgtctacttcctcgatcctc gtgtcttgaagctaccggccttctggcccagccggttttcctcggcctgggtggcccgccctcgcgatccgggtgtcctcgaaggggcgcgaggcga ggtggtcctgtccaaagccgagcgcgacacgtctcttagaaagcgcgcccttgcccactacggtctgcattgtcgggcgtgtggtctggtacccaagg tcgccgcccagatcgaggtacatcacctcgaccccatctcggaaggcatccggcgcacgacgttagcggacctgtgccccctttgtcggaactgtcac tcacttgcccactccacgagcccgcctatgccggtcgaggaggttgcgaagatcgtgggggcctag

Type I Systems

[0059] In an embodiment, the Class 1, Type I system comprises an HNH domain. In one embodiment, the Class 1, Type I system comprises a Cas5 subunit comprising an HNH domain. In one embodiment, the Class 1, Type I system comprises an McrA (5-methylcytosine-specific restriction endonuclease) or McrA-like subunit comprising an HNH domain. See, e.g,. Loenen et al., Nucleic Acids Research, Volume 42, Issue 1, 1 January 2014, Pages 56–69, doi:10.1093/nar/gkt747.

[0060] In one aspect, the present disclosure provides for systems and compositions wherein the CRISPR-Cas system may be a class 1, Type I CRISPR-Cas system comprising a Cas1, Cas2, Cas8e_Cse1, Cas5, Cas7, pfam08798, Cse2, McrA, CasE, and Cas6.

[0061] In one embodiment, the class 1, Type I CRISPR-Cas system comprises Cas1, Cas2. Cas8e_Cse1, Cas5, Cas7 and pfam08798.

[0062] In one embodiment, the class 1, Type I CRISPR-Cas system comprises Cas8e, Cse2. Cas7, McrA, CasE, Cas1 and Cas2

[0063] In one embodiment, the class 1, Type I CRISPR-Cas system comprises Cas8e, Cas7, McrA, Cas6, Cas1 and Cas2.

[0064] In one embodiment, the class 1, Type I CRISPR-Cas system comprises Cas8e, Cse2, Cas7, Cas1 and Cas2.

[0065] In one embodiment, the class 1, Type I CRISPR-Cas system comprises Cas8e, Cas7, Cas5, Cas6, Cas1,and Cas2.

[0066] In one embodiment, the class 1, Type I CRISPR-Cas system comprises Cas8e, Cse2, Cas7, Cas1and Cas2.

[0067] In an embodiment, the class 1, Type I CRISPR-Cas system comprises one or more components from Table 2.

[0068] Nucleic acids sequences of exemplary Type I Cas proteins are shown in SEQ ID NOs:495-500 in the Sequence Listing herein. All loci sequences of exemplary Type I Cas proteins are shown in SEQ ID NOs: 501-936 in the Sequence Listing herein. Complete loci

sequences of exemplary Type I Cas proteins are shown in SEQ ID NOs:937-1121 or may comprise 80%, .

Table 2

Map Label	Feature Label	Sequence
c5eH_1-0172376_10001016_organized	DR (SEQ ID NO:1173)	GTCTTCCCCACGCATGTGGGGGTGAACCG
	Cas2 (SEQ ID NO:1174)	tcatgaagcttctggttgcccgcctttagcccgaagaagcgtcaccaaggctaaaccctcgcgttcgatcagcatccgccttgccgaaccatattgccggaactggaa gccctggggcgtcgggcttgaccagatttgcgccacatagcccagaggcggcctttcggtcaccttcgcccaaagctcatccctcaccctccggctgcaatggccca ggaaaactcccggcttgatttccatcagccagcgagacagctttcccctcaagctcgtcgggatttttctccgagatcatcagcagcat
	Cas1 (SEQ ID NO:1175)	tcatccctcaccctccggctgcaatggcccaggaaaactcccggcttgatttccatcagccagcgagacagctttcccctcaagctcgtcgggatttttctccgagatcat cagcagcatcgagcacctccgcgatatccggcaatatgcgcctcatcaacttggcctggtgaaaagccacacggcaggccaatcgtgtttcccgctcaagcccggg cccgtccttggcggccatcgtgaatgccatcggtaccgtcaattccgacttgtacaagtccgcgatgtcgtaaacgaagctcagcattttccccacatggatgaacccg aggccggcggaatatccggctccgacgatcgcggcatggcacaggccgtaaaggcacgcgttcgccgccgaaaggcaacgattgacagcgttggccttgaacca gtcatcctggtcgtaggtccgcccgtcccatggaactcccgttttcctgctccattcctcgtagcaaacgcggacccggtatccttccatgcccgcaactgttccattgt cttggcggaaacatcctgtcccgggaatcgcttggcatacatccggcgggcgacctgcatccgtgtctcctcccgactcaccagcatcgcctggttgatcagccgcc gcgcgctgaatgtcgcgcccgtggaatgggcgtaaagcctcacgccctgctcgccaacccacgccacccatgcaggcgttgtcggccaggatcctcatcgcggcgt gggtgacggaggtgcccggacccagcttcaggagcgccaactggtcgatcggcaccggggtctcgccatcggcatcgcggaataccagcgccgaggcggtgca atccagcttgccatgttcgaggtaaagatgtgaccagcggtcggaaaaccgcggcagatcgtgcaggcggcgcat
	KOON_pfam09481 - Cas11 Fusion Cas8e_Cse1 (SEQ ID NO:1176)	atgggttccgtcgtattggataattccgacccggcacgatggaacacgctgtcatcaccctggttggaagcgctggatagccgcggcgtcaacatccggatctcgcc gctcgaggcgctcgccaattcggcggacttgtgccaacttgccttttccaaccctctcgaccttttcgcggcccacaggtttttgctgtcgctcctgtactggaaggcgg gagccgctaacggcacccgggcgctgcgcgagcggttggtttcgggaaccatgccccagggcgttatcgaatcgattcgggaggaagcatccgcattcaaccttttt gatcctgaaaggccattcctccaggatcccgcttgccgggacgacaaaaaggaatccccgccctcctatcttttcgccgaaatggcctcgggaacgaatgtcgccca cttccatcacggggaggacgattcatcccggctatgccttgctgtgcgactcgggtttattgcgactggcgccctggatgcaatcgggcggcaaggggttgaaacc cgccatccacggcgcgccgccgatcatggcgatcgccatgggggacacattgcgtcacaccctgggcctcaacctgattccgtttgacgggcatttgggaaatccg caatggaccggaccattcaagccgacgcgcgagggggccgtcaccccgctcctcgaggcgttcacctggaacacgcgacgcgtccatttgctcgatcctttctccgc tggactgctgttccaattgcggatcggaaagcggtttcggtgtcggaccgatcgttttcaaggcgaacgaggcttgcaaggatgattcggatgaggggcaagcgtatg ccgacaattggcgcgacccggcggcctttacaccgtgaaggatggcgatactgtcaacaagaccgtcaagtcgaccgacgagtcacaggcctccgtcgccgccg atctcacgcgccttgattcctcaaaagaggggcaaggcggaagtaccggcgccacggtgtctcctgacgcaggccaatccgggacacgcgaactggatccttgtcat tccctgcaccaatcccgccaataacaagacattcgatcaccgcttggtggcgatacggggagagctgcccggcgccatcgtaccgggccgtaacgatgtggatgcg attcctgttcttgccggggacccgcgcacgctgcgcaagccctggtcggcgaacccacgccggggcgcgagggtattcgtgaacgccgccagccgcctgcaaaa gggtgactgggcggtgttggcggggtccgccggcaagccccttgggacacgacgccgcggccttcgacatcttctcttccctttattggccaatcaggagccgggtttc ctcggcgccttcgcgggaagccgcctggatgacgcctcaaactgatggcggtttccaacctcggtggtaatgccgccatgggtgatattcgcccgtggaatcatttgaa aaacagtatgcggaaacagaaaagccggaacggcatcaacattgcctacccgaggcggattcccaagggcgttcaccttgaaaccgagcttgggggatcatcag gcgttttttccggcaagggactcgactgggcgggcctttgccaattcctcaacgaccatctttcctga
	KOON_pfam09344 Cas7 (SEQ ID NO:1177)	atgtttgtccagatccacatgcttcaaagtgttccgcccggcaatctcaaccgggatgattccgggttgcccaagaaatgcgtttttggcggcgtcacccgcgggcgca tcagcagccagtgtctcaagcgcaacatcaggctttcacctctgttcaaggaagctttcgcggaacacttggccgttcgctcgaaatacctttcccacgctggtcgccaa cgccttgagggaaaacagggccgaagccatccctgaaaaggaaatcgaggcggtcaccgcggccttggccgcgaggttcaagaaagaggaaaaagccagggat tccgaacaggacaccggggaggacgatgaggctgccgatgccccgaggctttggcggtcactccgcaactggtatttttccgccccgcgttcgccggcaagatcg cagaactcgcgcggaacttgaggagcaccgatccgctggcctacaagcttttcaccaagatggaaagaatctcgaaaaacgatgagagaagcgccaaggccggc gtcgaaaaattcatcaggtcggtgttcgaggcgcgggaatcgttgtcggtcgatgtcgcgcttttcggtcgcatgaccaccagcgacatcgtaatggatatcgaagcct cgtgccaggtggcgcatgcgatcgcgacccatgagaccttgattgaaagcgattacttcacggcgatggacgacctcaacgacagcttgggcaagggtgccggtgc ggcattcctggggtcgggcgacaccgagacattcttcaattccgccgtttattacaagtacctgaacatggacacgggcgcattttgcgcaacctccggggtgatgac acctccgtggcctccaaggccgccgcggttctccttgaagccgcggcactctcgacgcccacgggaaaacagaactccttcgccgcccattccgtaccggaattggt gctggtcgaggtctccccaaggaagcagccgatcagttacgccaacgcgttcctggggcccgtgcaaggggaggacctgatggcggcctcggcggacgccctgg ctgattattgcgggacagtcgcggcggcattttcaccgccgaaaactcaaagattttaccttgccgttggcagggcccggaacttcaggatccccatggaaggtgaac ggatcgacagcctggccaacttggccccttgaggttgggaaaggtgtcaaggctgccaacgcgggtgccaagtga
	KOON_pfam09704: Cas5 (SEQ ID NO:1178)	gtgagcgacaacaccctgttgctgcgcctggcagggcccatgcaggcgtggggcaccgattcccgattccagttgcgccggaccgaatccgttcccggcaagtcg ggcgtcatcgggctggttctttgcgccatgggaatcaggcgcgaagaggcctccgcgaaggctgccgaattgtcggccttgaaaatggccgtccgtgtcgaccggg ccggccacaccgactgggactaccacaccgtcggcgccggcatcggaatccgtcaagccgagggggggcgtcaagaaaacggcaagcacgggtgaaccggag accttctgtccaggcgccagtacctttgggacgcctctttcacggtggcgcttcagggtcccggcaacctggttgcggatgccgcccaggcgctgagtgaacccgtt tggcctccgttcctcgggcgaaaatgctgcgtgccaaccgaacccgtcttttggccgctggccgtttcgcaaacctcgcggaggcgctgggggcgatcgaatggac ggaccatcatggatcgtcgaggaacgggacgacggaactccaggcattcattgaacacgacgaaggcggcccggtcccaccaggtgccttcgtggtttacgacac accgcgaaccctgcgaaatcccagccatgggccaaggttggtggttcccggcacgatcaatgttcgggttgcgaccgttcaaccggccaacgcaatcggcaggcg aggagtgaactacgccgcggcccaatggcgagcggtgagggaaaggcggttggaactggaccgtttcctctgcgtcttctgcaagatgcccgcggaggatgtgca tcatgtgacctacgaaagggctggaaacgaatccctcgacgacctgaggtcgttgtgcaagatctgccatgacgcctgcacgcaactcgaatacggctccggcatga ctcgggaaaggattgatccggcagaacccgcggcgcgcaacgccatcatggagcaggttggcaaacttctttcacaaaggcgcctggcaaggcgtcaggccatac gcgatttggcccgcggcgacttttttcgacaatgttcccgatgcggggatatcctga
	KOON_pfam08798 (SEQ ID NO:1179)	atgttcctgtcccaattgatcgtcaatacggcgggaaatccggaccaacccgtcccggcagggaatgggtgcgatccacttacagggtgcatcagcgaatctggat ggcttttcccgacgaacacaagcgccatgccgatccgttttttccttggcgcctggtcaacgcctgatggcgtcaagcctcgaaggacggaggcgggattccttttccg cgtcgaaccggatttgccgacccgggttttggtccaatcggttttccgcccggattgggaatatgccttccagaacgccaggcatttgctggcggaaaacgccaaatgtg agggaattcaatccggacccagtcgccggccaaccatacaaattcaggctcgccatgctcatggtgaagaggaaaacacacaagggatccgaatcagccgggccc gtcgagcatccgatccgttgcctggttcccctcgcccgtggctggcgagccttttcggcccgatccgaaattcacggcttggcgggaaaagctatcctcggaagccttg cggcatgggtttgcgctgggcgagtatccctcaaggctttcagtttcgccgatcaggcacctcatgatgaaaccggacaaggacggcaaggccatgcccttcaacgc cgccatgttcgagggcgagttgacctgcaccgacccggaaaaactcaaggccgcgatggttggtggaatcggccgtggcaaggctttcggcatgggcctgctttcg ctcgcgatccttcccaagaaagcctccggctag
c5eH_2-0172382_100091	DR (SEQ ID NO:1180)	GTCTTCCCCACGCATGTGGGGGTGAACCG

64_organized	KOON_pfam09481: Cas8e (SEQ ID NO:1181)	atggaggagagctcaatgtataatttgcttgaacaagattgggtgagggtgctctaccgcagcggagaacctaagctaatctcgctatcgagactgcatcgggatgcg gataaattgcatcttgcctacagtaatcccatggaccggctggcggttttccgctttcttctcgctttgggctattggtgttttgccaatacgaaccaaacaccggcacgag ataaacccatacctgaagcatggctgtcctggctggaggagaacaagcagtatttgaaacttttatggtgacggtcccaggatgtatcaggacccgggcgttaaacgca cacgtcccatcaccgacttgctgcacgaaattcccacgggaaacaatctttcgcatttcaaacacaccttggactatgtctccggcttgtgcctgccctgctgcatacaa ggattgctgcgtttccccatttatgcaacctcgggattgccacaactcaaggccagcattaacggcatgccccccatttttatgccctgccatgggggaacactctggcag aaactctttacctgaattgggagccacgcgaattgatgggaaccccggcctgggttttcaacaccgttgaccccgaagacggcccctgtgccactcttggagggtcttac ttcacttgccaggcgctgtgtattacatgaccccatacccggagatggcgcctgttccgcatgtggcgaccgaagcggaagcttggtatattcctgtggttacgacacg gctggtgttgttgccaatcccatgtgggaagacccgcacgtatgctacgatgtcaaaaaaaccaaagcaaacaaggcgttgaatttaatctctgaaaaaccaatgatgc aggatagaaactggcaacgtgcattcactggttcttccaagttggagtccggcgatgaaagcatcaatctgctctgcattggagtatcagttaaccaagccgtcataaat gacgtgtgggagtccaccataagaattccgccccgggaagccggaaaagacccccaaaagcataccgaattggacttggattaaccgtcgacttgttggtggatatga aagaaacagattgtattgggccgataaatcagtttccggaaaacccagggctccgctcagtgaatcgctcatcagcgcagtttctccgcatattcttcattatgcgattca ggagacgggtggttttttgccggataacggtgaattgcccaaatcggtacagattgagtatagaaacatgctgaaagctgttgccaaatcccaggcaccggctgtcac aatagaggatgccatcgcgtccaagtttgtgcataatcagcgtccacagcttgagcaagccaaagcttccaagcaggataaaaaggaggaccaagttgattga
	KOON_pfam09485: Cse2 (SEQ ID NO:1182)	ttgattgaaacatccacagctcaattcatcaaaagaattgcagaactgaagtctggagacctgggtcggctgagacaattgagaggttgtgacctgcatgagaatctgg acgggttcgacctgttcaccggcctgtggtggacattgagaaaaaactcgccccgggcgccccgacgtgagattgcctggctcgttaccaaactgtatgcgcagttcc cattctctcaagaggacgggctaaccttgcctgtaatccttggacgtatttacaggactctggcaacgcaaaaggaaaaacgaatcttcgccatgaggacggatgccat tctgcaagcgcaggcaaaccagttggaacacccctaagctgggccctgcggatgcttgccctgtatggctacgcttcgctggattggattcaacttaccgacgacctt tcggtttggcatgaaatagtgcgacaccgatggacggaggcatttctaacacaatcatcaaaggacaaaaacaaggagaattaa
	KOON_pfam09344: Cas7 (SEQ ID NO:1183)	atgttaatcgaaatgcacatgattcaaaaccacagcccggcaaatctcaaccgggacgaccttggcgcgcccaagacctgttatttcggcggagtgctacgttcacgc atttccagccaatgcatcaaacgcagcatcagaaccagcgatgattttcagactctgcagggcgggttcgcacacgcaggctggtggaacttatccagctggaagc tggggacacagattgcaggaaaagagcacaggaaattctgacctattgcggatttaaacatgaaaaagacaacagcaaaatgctgattttcatgtccagagacaaggt gaaatatcctgccgagattatcttggatgaatccctggaaatagcagacgctgcccagaaggttgccgatgtgattacccaggccacgtgttccccgatattgcgcttt ttggaaggatgttggaacctgaaacaagcaagactaaatggagagatacaacagtggaggcagccatgcaggttgcacacgccctttccacacacgccgcccgtcc ggaaatcgattattatgtggcggcagatgacgttaagggagaagatgccggtgccggctttgttgacgaagcgatgttttgcatccgcctgcttctataaatatttctccatt gattgggaccaactggttaaaaatttgaaaggcaacgaagaactcgccgcacaaacagttggggcatttatcaagtctgctgcgctcaccaatcccagcggcaaaca aaacagtttcgcagcccataacccgcccgatggcattttggtggaattcaaggatacacccgtcagttatgccaatgctttcgttcgtcccgtaaagcttaccgccgaat cggacttggtggaactcagcataggccaactcgccaactatgttaaagatgtaagcgaagcatattatcccggccacatggaaggcaacgcattttggttctctcccaa cggacgccatccctgttaccacaagacggggacatacaagccgagcaagttggacaactggaagaactgataaagaaaacggttcataaatgcggttatgagtgg gaagccatcaagcaaaaaaaagtagaaggctga
	KOON_pfam09704 McrA (SEQ ID NO:1184)	atgagcgccaacaccctattcctaagactggagggaccacttcaatcctggggcgggcatgaggcgaaatttttccattcgccgcacttctgatgctcccaccaaatca ggagtctttggcttgctttgcgccgcactgggaattggcagagtcgaagcaggtgaaagctgcctgcccgaactggcgaacttacgcatgggagtgcgcatagaca ggctcggtgtccgctggtgggattaccacaccgtgaacgctgggattccgatgcttaaatctgggctagaggcgtctgaagacctcagcgatgtgggcaaagccag ggcagaccgactgcttaaaaaagaactgagtttcgacaaacgcgggttgctcagccgtcgcgaatatctggcggatgccagttttctggtggccttgcagggtgaacc ttcaacagttcaaaaactggcggaagctctgcaaaaacctgtctggacactatatctgggcagaaagacctgcacaccatcacgccctgtgggagagcatacccccg gtaaatatgcgagccttgaggaagcattgttggccataccgtgcaggtgagggataaacacgaaaccctgccccagtccctccaatgtgttctggactggattccgg cttatgatggtgaacccgctccggattgcgcggaaatccactacgacctacccatgtcttttgacctccccggcatttgtgattcgccgtgaactgccct taagtgtagatatcgaaacacaacaggttatcattggaagaccctggcgtccggcacgcagtagggccgattacacaaacactcaatataagagagtccgcgcagaa cgtctggtgatggatgtcgggctatgcatggtatgcaaggcgcctgccacaacagtgcaacacgtaaactatcgtcgcactggagggagggaagagcctgaagattt gcgcgccctgtgtcggctgtgccacgacgcatgcaccatgctggaatatggcagcggcatgaccatggaccgcatagacccgtgtgaccccgcctggcgtaaccg catcataagcaaacgcaatgaaatcatcgaatttccgctcgctggaacaacgcagacgaaagcttgacccccaaggaggaaaacggatga
	KOON_pfam08798 CasE (SEQ ID NO:1185)	atgtatctttctcaattactcatagataccggtggaaatccggacagactcagaccgggacgttcgtggctcaacaatatttacaacgtccaccgcaggctttccatggc ctttcccatgcggcaacgaaaagatggtgacccccggttcctgcagccttttaacccggctgactttgagcgtcaccgctttctgttcagggtggacaacaacatcgac ggagaggaaaccagagccatggttctggtccagtcatcaatactgccggattggaactacgccttccaaaacgccaccggctttttggcggctcctcctcaggtaaaa gaatacacccaaatttttccccccggacagcagctgcgcttccgcatcctcatcaacgccagcgtaaagcgcaaaatcgaagaaatgaaagacggggactccggaa gagttaagacggggaaagtgctgcataagcgtgttgccttcacttgggatgctgatgtttcccctccgatgcacttgccgaatggttcgccaaaaaggcgtccaaact ggggtttacgccaattaaaacagagctggttaagcttggctggatatatggctccaaacctgaaccgcgaggtgaaagaccacgaagagaaaatgaatacccgcccc ctgaaatatcgttctgcactgctggaaggatatcttaaagtggatgagccggagttattcaagaagactttggtcacgggaattggcagtgcaaaaagcatgggatttgg gctactttccgtaatttcccacacaagatgcgtaa
	Cas1 (SEQ ID NO:1186)	atgcgtaacaaagacctgcgctccatcccaaagatgagagaccgcatcacctacctgtatgtggaacatgcaaaggtggataaaaaacataaatccatagcctttcat gatgccatgggttccacacaggtgccatgtgctggactggctcttttgatgctcgggcccggcgtatccattactcatgccgccataatggtgttggccgaatgtggttgt cttgttgcctgggtgggtgaacaaggcgttcggttttatgcttcaggcgccggcaccacccgcagttctgcaaacctgatgcgtcaagctgccgcatgggctaatcag gataaacatatgaaaatagtgcgcacgatgtacgagatgaggttcatggaaaaactggacacctcatataccctgcccagattagggtaaggagggtgcgagagt gcgagacacatatttcaggttgtccaaagaatatggagtcaccatgaaggccagagaatacccaaagagacaattggattaaaacagacccgtgcaaccgtgcaatttc cgtggcaaactcgtgcctatacggaatatgtcatgctgcaatagtggctgcgggttattctactgccttgggctttatccatacgggaaaacaactctcgttcgtatacgat attgcagacctttataagatggaaacttcagtgcccgctgccttttatggcaacagctgaaggaagtgacagccttgaccggagaacgcgaatagcctgtcgtgaccaa ttccatcaatttggtttgctgaaaaaaatcgtcgccgacatccatgcccttttcgacatttttggaaactccggatgatgaacccgacctggtggatatcgatgccgcagca cccggtgggctgtggaacccggataaaagtataagtccaggcgggcaatcatggggtttggattcacaaggaactgaagatgatggttctgattcttga
	Cas2 (SEQ ID NO:1187)	atgatggttctgattcttgaaaatgttccaatctcactccgtggtgaaatcacccgttggttgcttgaaatataaagcaggcgtatttgtcggcaacgtgtcagctctggtca gagaaaagatatgggataaagtcaggtcagattgtgccggcggcacttgcctcatggtcttcagtaaacattgcgaacaaggctttgacataaagttttggggcgctgc aggaaaaattgtcacagaatgggaaggtttaaaacttatgactaaacatccgtaa
c5eH_3-0187892_10009586_organized	DR (SEQ ID NO:1188)	GTCTTCCCCACGCATGTGGGGGTGAACCG
	KOON_pfam09481 Cas8e (SEQ ID NO:1189)	atgactttcaatttgctcgttgaaccgtggattcccgtcctccgtacaaatggacgccctgcgcgagtaggcgtccgggaagcgctaacgactgcgggcgccatccga cagatcgccgcgtcgaacccaatggataacgtcgctctgctccggtttttgctggccgtgctcatgtggtgcaagccgcgtctaagtgacgaagaacgcacacagctc gacgtcgcgaagggcgtcccagaagagtggctggccgaacgactcggcggtgcaagtacgccgtgcgaggagttcaacctgctcggtgacggcaaccggttctt ccaggatcgatctacccccaaataacgcgcgccccatcggcgacctgctcgttgagtttccgacagaaaccaagatcgcgcgcttccgccacgtccgagacacgcag tacggcttctgtccggcctgcagcgcgctcggcgttgtgcggttctgcgctttcgccaacgcgtacggcggcggtcgctacacatccgcccttaacgggcctacgcc cgcatatgtgatctcgcaccgcccacgctagttggcactctgctgctacactggcctatgaacaatccggtcaaacgatcgccgccgtggcttttgtgctgatccgccg tccgaaagcgacctagacctcccacggcctcgcctggcgctcgcgccgcatttggctaaacgatcccgatgccgacgggattgcgcatactgtggtcaccgaag cacactcattcgacaactcgcgttcaccggggcgtggaagccgccatttgaaaccaaggggcaggagaagaggttctgggagaaggatccgcatctgatcgtcgtg gagaaggaccgcgaggcgccagcggaggagacggcagccgatgacagtgacgacgcgccccaccgtgacgccaagagcagcgagcgccgccgcacgaga acaacgctcggtttcccttccccggggtcgcgagtcgcggcacattcgaggttctggcgcagagcccttcgcgctaggatcgggcacgccagtggcaaaaacgcg gacgatgcaaccgctgtgatcgttgcgggaccccgctgccaacaaagggctttatcaggatgctagcgaagtccgcttgccgcgtctttctaacgatgacggcaccgc

		cgcgtcgattgatgctctggcgaaggcggttgatgctctgagcgccgtgttgaagtgctccacacgcaacccggagcgtcagcaccccgagcacaaggccgcgct cgattcgttggcaccttcgctagaactgggttttcgcgaggatactcaagcgttactgatcccgagcttctcaatcgatcacgcgacagagcgactcttacccgtagtga ataccgtccttcgcgccacgtccgcaggctcgccgctgcgccgggcgcgagggagaggcgcgcgcacgcgaggcgttgaatcgcgttctccactcggtcgctgcc gcgagcggttcgcccgcgaaatatcaaaatgagcccagcgcgggcgaacaaagggcccgtgaacacaagcgcacgcgcgcgacagagtctaggagaaccaac gcgtga
	KOON_pfam09344 Cas7 (SEQ ID NO:1190)	atgctgatcgaaatccacatgatccagaaccacagcccgtcaaacctcaatcgcgacgacttagggggcacccaagacctgcctcttcggcggcgttacccgcgcac gaatctcgagccagtgcctcaaacgcagcatccgacgcagcgcggagtttcagcgcgctttggagaaaggcggcggggtccggacgcggcgcttggtcgcaattc tgtcggagcgcgcgcacggctcgccctccgcacccccacgacttgatcgagctggtgactaaggcgttcgaaaagggtggggtgaagaaggatcccacgaaggac gacccgcacgcaacgaacattttgttgttcgtgccagaggggacggtagtcgcgatggcgaaagcggtcaaggctgagtacggcaagcccgcggctcaacggac ggccgatggcttggctgagacactgacggaactcttcggtcagaaggcgtcagtgccagacattgcgctttgcggacgcatgaccgaacttgactcagacggcctgt ttcgcaagatcgacttttcgcgttgaggcggcgcttgcagcggctcacgccatctcaacgcacgaggccatcaacgaggtcgattacttcacggctgtcgatgaccag gcggtcgggaccggtgccgggcatgtcaacgaggccatgttcgcatcggcctgcttctacaaacacttctgcatcgactggaacccaacttctcgacaacctcggcgg tgacgagccgctggcgagtacgaccgtgtgccagttcctgcgcgctgcggccatgaatgtgccgagcgggaagcaaaagagctttgcggccttcaatccgccgga cggcattctagttgaagtcaagcacgaacaccgcctgcccattagctacgcgaacgcgttcgctgatcccgtcccgctcaagactctgcgcggactgattggcgaaa gcatcgcgcgcctggggcagcatgttcacgacattacggccgggtacggagtcaaagcgaagcggttctggttttctccggccgggcgccaccggctgacgtggat tgaccgaacggctaaggagggtgaaaaggagaagctggccatagacggcgaagatggcacttttgattcgttcgacaaattcgtcagcgccgtaatgcaggaactg ggaaggaccgcctcgtga
	KOON_pfam09704 McrA (SEQ ID NO:1191)	gtgactggaagtaccgttttcatacggctggaagggccactacaatcgtggggcgacacgtcgaagttcgtgattcgccgaagcatggaggccccgacgaaatccg gcgtccttgggctgatctgttgcgccatggtgctttcgcgagaggccgcgcgcgaggtgctcacagacctgaatacctttggttatgggcgtccgtatcgatcgggcgg gcatgcgttggtgggactttcacaccgtcggtgccaaggtcggcatttctgaatgcaaagggccacgtgaagaagacggccagtacgggcgagtatgaaacactcat cacccggcgcgagtatctcgccgacgcgagctttctcgttgcactgcaaggcgacgatgagctgactaagatcgtgcacgctgcgctcgagtcgcccaagtggccg attttccttggccgcaagtgctgcgcgccaagcgtgccgctactcgctcctccgtcggacggcgaatcctggacgaacccgcaggtacacggcagcttgagcgctg ccctcgccgccgtatcctggcgccctcgatacgattccgaatcgccgcccatacctacgacgcgcgacggaagggcggagcttgattgcttgatggagtggcgggc cctaagcgacagcgacgtcgcgcccgatgcggccgaggtatggtacgatgtaccggtctgtttcgatccccccgtccacgagccgcggatcgtcacgcgaagcact acggtagtcccgatcgcgcgagcgctccaacaacacactccgccgccgccgcggccgcgcgctgattacaccagcagagagtacaagaagcaacgcgcggccc gtcttgaggctgatggcggtgtgtgcgtcttctgcaaatcgcccgccagcaccgtccagcacgtaagctaccgtcgcgcgggcggcgctgagacgcaggacgacc tccgggcgctttgccggctgtgtcacgatgccgtcacgatgatcgaatacggcctgggcatgggcctcgaccggatcgacccgtgccaacccgttggcgccaaga cattctgcgtactcgggccgaaatcgtgcgtttccgcgcgcttcaagtacggcgccgacgcctttagcccccgacgaggtcgaatga
	KOON_pfam08798 Cas6 (SEQ ID NO:1192)	atgtttctatcgtcgctcctgatcgacaccggcgaaaaccccgaccggccgcgtcccggacgactgtggctgcgcaacatgtaccgcgtacaccagcgactttgcat ggcgtttccgactgcagcgcgcagcgaaaacgatccgcagttcgtacaaccctaccagcgggacgacttcattcacgtgcacggatgtcgcacgaacgatcaggct tttctctttcgtgttgattctctcgccggcggcagagtagttgttctgatccagtcggcgatcaagccggattgggattacggattccataatgcgcagcaccttctcgccg cgccacccgccatcaagttgtttgatccgcagctcgcggccgggcaacgcgctcgatttcggctgcaggccaatcctaccaaacgcctgcgccgcgaatcgccaaa cgctagcggtgatccgacaccgcaacagtgggtgggcaagcgcgtcgccgtgcccgctgacaagctccaggactggctcgcgcgccgcgcagagagctacgga ttccacttgcgcgaattgacggagttgtgtccgggatacgtgtacgttagtcgcactccagagtccgccggcggtctacggctgcgatctgtcaggtacgagggtctg ctggaagtcactgactccgagtcgctggagcgtgccatcatccgtggcatcggccccgccaaagccttcggcttcggcttgctctcgctggcgccggtccggtaa
	Cas1 (SEQ ID NO:1193)	atggacgatctgcgctcgctgcccgcttcgacgatagcctcagctatctctatctcgagcatgcccgcatcgagcaggacgacaaggccatcgccttctacgacaa ggccggcctcacgcccgtgcccatcgcggcgctggcgctgctgatgctgggcccccggcaccaccatcacgcacgcggccatcaaggctctggccgacaacaact gcaccgtcgcctgggtcggcgaggacggcttgcgcttctacgccgccggcaccggcggcacgcgcggcagtgccgcctcctgcgccaggcccggctgatgag ccatccgcagttgcgcctggcggtcgtgatgcgcatgtaccgaaagcgcttccccggcggactcgaccgcgaacctgacgttgcaacagatccgcggcatggaggg catccgcgtccgcgaggcttacgccgccgccagcaaggctaccggcgtgccctggaccgggcggaactacaagcgtggcgactgggccgctgccgaccccgtg aaccgcgcactgtcgaccggcaacgcttgccttttacggcctggttcaggccgccctcctgtcggccggctactcgcccgccctcggcttcatccacaccggcaagca gctctcgttcgtctacgacatcgccgacctgtacaaggccgacctggtcgtgccggccgccttcgaaaccgccggctggcaatgcccgtccgtcgaacgccagatg cgcatcgcctgccgcgccaagttcaaggacgctcgcctcatgcagcgcattctgcccgacatcgccgagctcctcgcgctcgagggcttgcccagcgtcgccgac gacccggcctgggacgttgaatcggccgcgcccggcgcgctctgggacccggacggccgcctcgaaggggcctcaactatggtactgatgttgcttga
	Cas2 (SEQ ID NO:1194)	atggtactgatgttgcttgaacgcgtcgcccctctgtgcgcggcgaactcacccgctggctcatccagcctcgcaccggcgtcttcgtcgggaccgtctccgccctc gtccgcgacaagctttgggacaaggtcgtacactccctacaggctgccaaacccgaaaaagcggaaagcgcgccggcgcgattatggtctataccaccaatacc gagcagggcttcgacatccgcgtcttcggccaaaccgatcgatccatcgaagacttcgagggcctcatccttgcaaaaaccatcgcccaaccctaa
c5eH_4-a0073932_1000118_organized	DR (SEQ ID NO:1195)	GTCTTCCCCACGCATGTGGGGGTGAACCG
	KOON_pfam09481 Cas8e (SEQ ID NO:1196)	atgaacgactttaacctgatagaggagccgtggattcctgttatttaccacaatggcaactattgccgtgtcggcatcaagcaggctctttccgatgcccatctcatccgc gacatcgccacctccaaccccatggaccgcgcgggccttctacgctttctcctctctatagtttattggtgccatggaatgccagcccctcaagaaacgtctcggctagg tgatcctttaccggattgcacaccacatcttggtgcaaagacaaatttttttctgccttttcgggcagaaacaccgtttttaccaatcaccaacggcaaaccgagaagtaaa aattggtgaactgtttcaagaaattccatccggaaacaatttctggcattttcgacatgtcaccgataaagaacatggatatttgcccagcctgttgtgctctaggcctgctg cgccttcctcttttttcagtcagtggattgccagacttgaaatcggggattaatggaactccacccatctatgtctttcattatggcaaatctttagctcaaattctccgaaaaa actggttgccaatgaagcaaaacttgctgggcacgccatgttggatggaccccaccttggttccaccaatcggagacccctatcccgctgttaatcggattaaccgtccc agcacgcaaagtttggctcggacaacccattcaccatggccgttgctctttgtgcgggaacccatcaaaacacctggttttcgaggctcacattcaatccgccggtacg ttggaaaattcaaattggcaagacccacatgttgtttatttcctaaaccaacaacaacgaaagtcatgcaaagctcccaatctcgtcgtaggcggtgaatttcaaatggac cgcccttatgccgatcttatggcatgtctcttggaacgggtccaatttcccgcaactgcgacaagccaacttctaattgtcggttttgccacggacaaagctaaaaatata gatgtttgggagagaactgtaacttttggtccttaacgaaatgcagacaaccgcagccacagcttttccacctggtggcgaagtgtaagtgcattgtccaacaatgtttac actattattaaaacaaccccctacaatcggcctcgtgggtccaaaccctcgtatgttcactgccatttttcccacatcagaccccatgtggaacatataatatcatctaacat gcatattctagtaagtgaaccaaacgaaggttggcatgaagcacgccgtgagttcagcgcttttttgccgatgttggctcaatccctctctcccggcgtgacctcccaag ccgttcaaaagcgccgtgccatcgctcaggccttgccacctaatccctgcccgaaaccggaaccatccagttcacccctccactcaaaaccccaacttgtaa
	KOON_pfam09485 Cse2 (SEQ ID NO:1197)	atgagcgccacccttgaatacattacaaaactgtcccatataaaggccggctcccgcacttttggttcgcactcatgccaatctctcattgaaccattccgtggatgccttg atttgttcaccggtctttggtggcccttgcgccaacaatcacaatttgcaccccgtcgccgcgtggcctggctggtcgccaaacttcacgcagtcagtcctctgccacac caaccaggccagacccctccccatccagctcgcgaagatttggcataggcagccagacaccaatacaaaaaacaggttccagcgcaaattcgatgatctccttttgccg tccccctggaccaaatggaaccgctttgcaatgggccttgagaacaattggctctgcaaatgagccatccttggactgggtgcaactcaccgacgacctctcccggtg ggaacaaccgcaagttcgcgaaaaatgggccgaagtctttttaggcaggtaa
	KOON_pfam09344 Cas7 (SEQ ID NO:1198)	atgaacaatcaaaatacacacgcatacaaatccatctgcttcaaaaccacgcaccggccaacttgaacagagacgatttgggagcaccaaaaacctgctacttggg cggcgtcctccgctcccgcatctccagccagtgcctcaaacgcagcatccgattgagccccctacttcaaagacctctgcggcggtgttcgcacacgccgcctcgcaa aactcatcactgaccaggttgatggcaacaatgcgacgaagaaaaaagtggccgccatcctgaagaaatgtggtttggaggccaaagagaacgacgaaacaaatat ggtcgttacaccacaaaagaagccatcagggaaatggcaacagccttgagaaacggactcgaagaacccgacattatcgccagattcaaagagcttttgagcacg aatactgctgtcccggacatggcactgtgtggacgcatgttggagacaggagaaatcaaggacaccaacgtcgaagccgccctgcaagccgcccacgccatttcca cacacgaggctcgccccgaagtggattactatgtgacggcggatgacatccccggaactgatgccggcgctggttatgtggacgaagccatattcgtttccgcctgtt

		attacaaatacttttccattgattgggcacaattgctccacaacctcaagggctacccgggcaacagaatccatctcgcagcccataccgtcggcgcctttatcaaggca gctgccttggtgactccatccggcaaacaaaacagctatgccgcccacaatccacccgatggtatcctcgtggaaatctccccacacccatcaattatgccaatgcct ttgccaaacccgtggacattggccaacgtgatttaatcggccaaagcattgctcaactcgcccagtacatccacgagatggatgccggtttcaaagtcgttgacccgtc caagcgcttctggttttcgccgggcatgcgcaatcgcttggagacctcacagcgcgccggtgatacagcaggcgaccgccaaaccatccgagtccaggcggtgga ggttgatcaactcgaaaacctgatccatcaggtcatcaaggccattgacccgatcctcgactggcaggaagtccaaagaaccattctcaataaccatgaagcctga
	Cas1 (SEQ ID NO:1199)	atgaaactggacgacctccatatcctccccaaattctccgatggttggagctacctctatgtcgagcattgtcgcgtcgaacgagaacaaaacgccatcgccatctatg acacagatggcatgatccccgtcccttgcgccaacctggctctcttgatgcttggcccgggcacctccatctcccatgcagccataacagttctggcggaccacggct gtttaatcatctggtgcggagaacaaggcgtccgcttctacgccaccggtatgggtgaaacccgcaatgccgcccgcctcctccatcaagccttcatgtggagccacg aaaacctccgtttcgaagtcgtccgccgtctctacgcaatgcgttttacagaaaccctgccatccaacctgtcattatctcaaatccgcggactcgaaggccgccgcgtc cgcctggcctatgacaaagcggcccgtgccagcggcatcccctggaaaggccgccagtatgaccgccatcattggacccactccgacccccgtcaaccgtgcccttt ccaccgccaattcctgcctctacggcctctgtcactccgctattgtctcctcaggtttctccccgccataggcttttatccatactggcaaaatgctctccttcgtctatgata tcgccgacctctacaaaaccgaaatcacagtccccatagctttccaagcggccgcccaaggagacaccaacctcgaacgccgtgtccgcatctcatgccgtgacgc ctttgccgaatctcacctcttggcccgtattatccctgacattcaaaaagctcttgcttggcaaccatcccaaaacgaaaactccccccttgattcagaccccgcccttccc tctcacctttgggacccggatttaggggccgtgcaaggcggaatccaacaacctgccacatcccctttgcccccaactatacagccccagccccggaatctcatgat ggtgctaatccttga
	Cas2 (SEQ ID NO:1200)	atgatggtgctaatccttgaatcagtcccaccttcactgcgtggagaactcacccgctggctcttggaactcaaagcgggtgtcttcgtcggccgcgtctccgccctcgt ccgtgagcgcctttggaataaatcctgccaccacgcccgcaatggcggctgcctcctcatctaccacaccaataacgaacaaggctttaacatcctctcctatggtaat acatccagaactatcgccaacttcgatggactcctcctcgtcaaaatcccaaaataa
c5eH_5-0315288_10006183_organized	DR (SEQ ID NO:1201)	GTCTTCCCCACGCATGTGGGGGTGAACCG
	KOON_pfam09481 - Cas 11 Fusion Cas8e (SEQ ID NO:1202)	ttgaaatacaacctgcttgaggagcactggattcctgttctcggaaccgatgggaagcttggctacgtgggtatcaaggatacgctgacccatgcagggcgcatccgtt gtatctgcttagcaagcccactcgacatgttcgccgtgcaccggttcctgcttactttgctttattggaaagcagacttagcgggtggggtggaagaggttcgggcgtcg ttgctgaagagcgaattgccgagcgcggtgctcggcggtatcaacaaggaagaccattgttttgacctcttcgataaagaatcgccattcctgcaggacacgtctgtat ccatctccaaggaaaaagaaaagaagtcggtcggctctttgtttgctgaattagctgtggctactaacatagcccacttttcatcacggcgacgatgagagtatgcggcta tgcctgccgtgcgcaaccatcggcctgctccgtgtggttccgtggagccaatcaggcggccaaggaaagactccctcggtacacaatgcgccgccaatttatggcaat ggcaactggcgataacttggcggccacgcttggtttgaaccttgtaccgttggatactaaagcgggcaaagcgcagtggtccggctactttacgccgactgattcgac gagtccaatccccctatctcgaggcgctaacttggaatccccggcgcatccttctcccctcagcccaagaggcacttgctggtactgcggtgaaagcaatgtacccac ggtcgggaaaatcgtgtacatgaaaaacgaaaacacaaaaaagcgtgcagacgagaaattgttcgaatggcaggatccggctgccttttatgctctatacgatgtgag taagaaatatgacaaagtcaagaacgtgccgtataaaacgataaaaagttcccgggagaacttggctaccgatggccgagacttgcagcgcctggcgcatgacgat ccgtttcccgaatctaccgtagtcgtgaagaaccccaagcatgagggttggctcctggtcgttccatgcacaaatcccgcgaacaacaaaacctacgatcatcggcag gttgagctggcacaactttcaccagagacgattcgctccttaccgcctgccgtcccccgtccaggagacagggcatcgacggttggaaagaacccgaaagagtcg agcgcaaggacgttgagcattttgtgccggctgcggtgaggcttcttacgcataccgattgggttacgctatcgaacgctgcctacatggatatgcacgattcaccagc ggcgttcgatgtttttagcggcttgtattggggacttcgtgacaagaagaccaaaggacttccgtcgcggaatgtggcgtggttgatgctcaaactgatggcatcagtcc ccgcacgtgcccgcactcttcatcccaacgccgaattctgcccgctgcagagactgccgaaacgacagagcaatgagcgcgacatggcgtcccctacccttgtttct ttcccacgcggtcgtcggcttgaggccgcacttagcgccgcacttgttgacaatctgcgcaagagaaatcccgaacctgttgactggacacgtctatgttacggactag accaattactcgactaa
	KOON_pfam09344 Cas7 (SEQ ID NO:1203)	atgtttgtacagattcatatcctccagagcatgccaccaggaaacttaaaccgagatgacgaagggcagccgaaaaagtgtatttttggcggcgtcacgcgcggccg catcagcagtcagtgcctcaagcgaaatatccgccattctatacaattcaaagaagtgttcggcgatgacctcgcaaccagaacgacatacctgccacgtatggtggc cgacgaattgaggaagagcgacccaagcattaccgatgacgaattgaacaaagtcaaggccgcgcttgccagcaaattcaaggctgaaaagcgcggaggcaattc agacgaagcagcagagggtgaggaggcacaagcgcagcccgaaagctcttctgctgtgagcgatggagcagaccggacgggtcaactggtctttttcccaccacc ctttgctaaagaaattgctgagcttgtcgtagaactgcgtgggcgcacccctccggcatacaaccggttcattgggaccaaagcaaagctctctaaagctgaagagaa ggctctagacaaaacggttgatgagtttgtccagcgagcttccgcggcgagcaagaagctcacggttgacatcggccttttcggcagaatgactaccagtgacctggt cgtgaacgtggaagcagcctgtcaggtggcgcatgccattagcacgcatgagacgctaatcgagagcgactacttcaccgcgatggatgaaaagaagcaggaattc gcatcaactcagaccgagaaaacaggcgctgccttttctcggctctggcgataccgaaacgtttttcaattcagccgtctactacaagtatctcaatctcgacatcgatgc cttgaagaagcacctttcctggcaaagcgataaggcagctcttgcagccggggtgcttgtccgcgccgcggcgctatccaatcccacaggcaaacagaacagctttg ccgcgcacggcgttcctgaacttatcctcgtcgaagtatcggaagcaaagtgtccaatcagctatgcaaatgccttcctagaacccgttgacggtattaacctcatgcgt gaatcagcggcggcaatgagtaattacatcgacaccgtagtagcagccttctcgccgacagacatgagccgagcgctcctggcagtggggtctgcttcggtagatat cgcaaaagctaaaagagttaacacactcgaagaccttatcaaatccgtggttaaactcgtggagggtaccgcatga
	KOON_pfam09704 (SEQ ID NO:1204) Cas5	atgagcagtgaccacaatactttgtttctgcgcttggccggtcccatgcaggcgtggggcacctcatcccgttttcagttgcgccgcactgacctttacccgagcaagtc cggcatactgggaatgttgctctgtgccatgggtgttaatcgtgaagattcacggcagccgctcaagctgctgacaccactactcatgggagtacgcgtcgaccggac tggaaatcttgattgggattaccagacagcgggagccaagttaggcattcgaagcgccgacggcaaaatcaagataaccgcatcgaccggcgaatacgaaaccttg ctttcgcgccgtcaatatcttttgatgcctcatttctggttgcgttacatggcgattcaaccataatcgcctcttgcgccaacgcccgagaaatccgacctggccttgttc cttggaagaaagtgctgcgtcccggccgagcctgtttacgccggcaccgaaagttttgacacgctgaccccaagcgctttcgtcggtcccctggcgccccagaatcaa cgccattgaccgcgagggtagcggggcgtctgtaacactcgatacatatttggagcagccgccagagagacttgtgcccatgggtgcccggctcgtttccgatgtgc ccgtagccttcggtttctataattatgctccgcgtttggttatcaagaatcaggttacggtgccaatgggacaggcaacccaaacacagcgactgagcgcacgtcatgg caaccctacagtgaacacttctatacagtcgcgaggccgagtcgtctgaaacttgataacggcctttgcgtcttctgcaagtcccctgctgtggaagtgcatcatgtgtc ctacgaaaatgccggacacgagaccgactctgatctccggtcgttatgcaagacctgtcacgatgcgtgtacgatgctggaatacggtcatggccgaaaagaaagag ttgacccttgcaacccggctcaacgcacgaccattctcgaacaaatcaatcgtaatctcaaagagcaacatcctggccggcggcgcgaattgcttaaagctgtgcggt cgacctga
	KOON_pfam08798 Cas6 (SEQ ID NO:1205)	atgtttctttctcttctaaagttgacttggacattgccccggacggctctggctccgtaaccttaccgcgttcatcagcgcctgtgcatggcgtttccatcctccgcgcg caaagagggcgacgcctccgttcttggaacccttcgtcctgatgacttccagcatgttcacggcccgcgcacggaggaacaagcatttctttttccgcattgaccctctac ccggcgggaatgctgcgattgtgatgcaatctggaatgaagccggactgggactatgccttaacaacgctggctatcttcttgccgcgccacccccagttgaagactta tgaaccacaattccagaaccagcagcttctcaggttccgtctgctggcaaaccccacacggaagatagatacgaagacccgccaggatggcaccaaaagcaacgg taagcgagtccccgtggacaaggctaacctcgatgaatggctcatccgccatgccgaaaagtccgggttcaccgtcgagcagttgctcaacgtgcaaacgggctac gtcggcgcttccaatggcaaagatgacgaactcaagcggttcttctacgcccgctacgacggcatactccgtgtggcggatgctgagcatctgcacgattccgtcattc agggcattggctccgccaaaggcttcggcttcgggctgctttctctagcgcgtatcaaggagtag
	Cas1 (SEQ ID NO:1206)	atgcgtacattacatgaattgccgcgtttccgtgatagatggagccacctctacatggaaatggggcggctggatgttgatgcgaatgggttggtattccagcagggtg agagcttcgtgcgggtgccaattgaccagttaagcatcattatgctgggacccggctccacggttactcatgcggcggtgaaatccctttcccagaataactgcttattat catggacagggcaggatggtattcggctctatgcggcaagcattggtggcacctacagcgcacggaggatgattcgccaggcgaaactggttagtgacgaggaga aacgtctggaagttgcctggcgtatgtaccgtttccgctttaaagaagcggttcctcaggtggttagtctggagcaaatccggggtatggaaggcatacgcgtgcgtcg cacctatgccgaggcttcagctaaatacggggtggaatggaaaggtcgccagtatgaccaggatgattggaacaaaggtgacccaatcaaccgcgctctatccgcg gcgaacgcttgtctctatggagtttgtcacgcaggtattctttcagctggctattcatccgctctgggatttgttcacacgggtaaaatgctaagcttcgtttatgacatcgga gacctctacaaaaccgaggttacaatccctactgctttcaggatagctgctcaaaatccaccggaattagaaagagtagtacgcatggagtgccgaaaggtgttccac

		gaattcagactgatggataggctcttgcctgacatcgcggaggtgctgggtgttagtgatgatactggagaaaacgccgacgagcttgaagggcgcattgtcacgctg gctgttggaaccgaagacgggggttttttcttgggaacccgagcgcgagagttag
	Cas2 (SEQ ID NO:1207)	gtgaaaaaggcaaaggataccggcgtggtaatgcaaatctggacagaccagaacccgcaaggtttttcttaccggcaaactggtgaaagggatcggagatttattga cctggatggtctggcattggttgagttcaaacgcgctcaagagattgtcaaaagcaatttttctgacactagaacaccctag
c5eH_6-Draft_10005650_organized	KOON_pfam09481 Cas8e (SEQ ID NO:1208)	gtgaaatacaacttgcttcaagaaccatggattccggtgcttcgtatggatggctcccagactcatgtcggcatacgtgatgccctggcaagatcggggctgctcttga gtctcggcgagcccagtgcgacggacaatctggctctgcttcgattgctgattgccgtcactctgtggtgcaggtctgaacctggctcctcggagattgaagctctgcg ggagtcgcgcggaatccctgaaggctggcttgcacgtctggatgaagactcgggcgcatttgaactctttggagatggcccacggttctatcaagacgggagggtga agtccgaggtgccgggcagaccggcgtccgaccttctctcatacttcccggcggcaactgaaatcaaccacttcagacacatccacgacgacgatgtcgccatctgc ccagcgtgttgtgccttggggctcatacgcctcccggtgcttgccattcagggcggcgctgggaagtcgccgtccatcaataatgcgcctcctgtctacttcatacggc gcggtagaacgctgttagacacactcatgttgaatatgcctcgtggcgggccggctgcggatgacagtcccgtgtggaagcccggcccaaggtcgagcagcgaga tcggtacactcgaaggattcacgtggcaaccacgagctgtgtggttgggtccgacgagggcgacctcagatagtcggtgcagcaactgtggaagaagcggccccg tcgtatcaagcctggtattcgccaagggtcaggcaagaaaggatgacacacgctcatggagagacccgcacgttcactacacgaaggagcgcacgggcaacgata acgcagatccagacgcccgcgataacgttctgcgcgcatcgcatccgctcaagtacgcggccttggaggcgtcctctggaggaggacagcctcagcgattatgtt gtccgtcgaaccacatgagtgcggcacgttgccagccggcgatagctacgacggaacaggaagcgacgcagtcgttgtcgagtgtattgcgccattcacatcccag ggaaaggtcttcgatgagcaccggattgagtggcggataccacgcaaggttctccacgacgatgtcctgcgtcgctctgccttcgaggaactcgattggcttggccg gttcaacttgaggaagctcctgggaagggcaagaggtgacggccgcatcaggaagcttgcggctaggtgcctagtcaatcccctcatctgcgagacagagcgccg actgttcggttcattcgtgactctgttgtggtcactggcgcaggagacttcgtcagagcagcgaggctcggtggaacgctggcgacgagaagtacttgacgtgctcgc ctcgatgattcagcttgcccttacaacctgtgaccccggatcggctctgaaacgcaggcagatgcaacttcgggcggtggaagcgctccggcatgacgccgggatg actgtcacatcgcgtgggcgcgcttccgtcagacacgagcaagggcaattggagggagcgacaagatga
	KOON_pfam09485 Cse2 (SEQ ID NO:1209)	atgaatccaaccgaggtctacgtgcggcgtcttctggctctaggggacggcgaacgagcagaactccgtcacaaacagggtgcctctctcagtgagtccctggagg gatttgatttgttcaccggcctctggtggccattgcgacagcgcagtccggcagcaccgcagagagaaacaagttggctggcggcaaagctgatctgtcagtggccg attccgctcgcttcgtccgatgacagcgtgctacctcggttgctcggcagcgccgagcggcgactgtccgctcgctacgcacttcgtttcctgctgcggttcgataggc tcttgcaatccccaatggctttgctggaaccgcagcttcaatgggcattgaaggtggtggcacaggcggtagttgccggttcagccaaagggttgaactgggtgcagc ttcttgatgacctcgcggtatgggacggaagacccggcagggaatcaaagcacaatgtgcgcgacgcatgggctcgcgcctatctggcgtctagaaacagggagg gaaagcgatgttag
	KOON_pfam09344 Cas7 (SEQ ID NO:1210)	atgttagtcgagattcacatgattcagaaccacagtccgtcgaacctgaaccgagatgaccttggtgcccccaagagctgcacatttggtggcgtaactcgcgcgaga atttcaagccagtgtctgaagcgcagtatccgcaacccaggaaattctaacgacgtacacaacaaaggggcaagccttttcgtcgaagcgatgcgcgaacacacgg gaatcaggacgaagtactttcctgaaatggtgcggcaggcgctgaacgactcaagcgtaccagacgaggaatgggacagagtcgtgcaggcttgcacgagtatcg ccaagaaggaggtccagatagcgacgaggatgtacaggagcccggcgacagcgatgagcggcacaaactgccgcagctagtcctgttggagcgtactgaggc aaatgcgtttgttcatgtgttggaggcgatgcgtaccagtgacgagtggagcgccagctaccgccagtggatcgcaggcaaactgcagaagcggcctctcgcagctt tcctgcgtgaactctcgcgcgcctattcgagacgatcagtggacatcgccatgttcggacgcatgactacttccgatgcgttcgaggacgtcgaagccgcaatgcagg tagctcacgccatatcaacccatcaggcggtttcggatgtggactacttcacggctgtggacgacctgggcgagttgggcggtggtgcaggccacgttgacgaggc catgttcaattccgcctgcttctacaagtacttcgctctagactgggaccaactcatcagtaacctcggtggctcttcctcttccgcactcgcagcgtccgcactgggcca tttcattcgtgcggctgccatgacgacaccaaccggcaagcagaactcattcgcagcccacaacctccagacgggatttctcgtagaagtcaagcggaccggcaaa atccctgtcaactacgccaacgcgtttgtcgagcctgtcgcaatgaatgctccgcggggagtcgtgggtatgagcatcgctcaactgggccaatacgttcatcagattg cagagggttacggcattgactctcgtcgcttctggttctcaccgggggggcgctatcaattgacatggattgagcgctgtcagaaaggagagccgaccgagcgggа agttgttgcaggtgaagattcgcacgcgagattcgacgacttcgtgagtgtggttgtgcaagaggtttctggcctcgaatggaacagggtgaaggatccggacggttt cttgccggaggcgacatag
	Cas1 (SEQ ID NO:1211)	atgcgcacgctcttcgaccttccacggttccaggatcgctggagctacgtgtacctcgagacggggcgactcgatgtagatgatctcggcctgtgcttccgccaggac acccgaacaacgcgagtgcccatcgaccagctctcggtgttgatgctcggacccggcactacggtgacacacgctgcggtagcggcgctctcgcggaacaactgt ctcctggcgtggacagggcaggatggcacgcgactgtatgccgccagtacaggtggaacacacagcgccaaaaggctcatcaggcaggcgtcgctggtgagcg acgatgaatcccgcctgcatattgcctggcgcatgtaccgtttcagattcgcggaatccgtaccagacgtcgtgagtctggagagcataagagggatggagggccta cgtgtgcgcaaagcatacgccgacgccgcccgcgcatatggcatcgactggaagggcaggcgatacaaccagggcaactgggaagctgcggacccggtcaatc gtgcgctatctgctgcaaacgcctgcttgtacggtgtgtgtcacgccgcgatcctctcggccggatactctccgggactcggtttcatccacacgggcaagatgctcag cttcgtctacgacatagcagacctgtacaagacaactctctcacgattccgctggccttttcaaaccgcggccgagatgcccgagagtatcgaacgcggcacgcgaatgc gttgccgtacggcgttccacgaattcagactcatggagagacttctgccagatatcgcggaggtgcttggtgttggtaatgatgctggggagagtgccgacgaacttg aggggcgcattgtcgcgatggctcctggagccggcgactggaacctttcttgggaacccgagtgggcgagtgagggacgagttgtggatgaaggccgtcaagaca acgaaggagaccgggtgggtgctgcagatatggtcagacaggaatccacagggggttctcctatcgtcagcacggcatccaaacgaggcaactggtggacttcgag ggtctggcgctggtgacagtctctcctcaaccgcggactgttcaggatag
	Cas2 (SEQ ID NO:1212)	atgatgctggggagagtgccgacgaacttgaggggcgcattgtcgcgatggctcctggagccggcgactggaacctttcttgggaacccgagtgggcgagtgagg gacgagttgtggatgaaggccgtcaagacaacgaaggagaccgggtgggtgctgcagatatggtcagacaggaatccacagggggttctcctatcgtcagcacggc atccaaacgaggcaactggtggacttcgagggtctggcgctggtgacagtctctcctcaaccgcggactgttcaggatagcgacactgaaacccgttga

CRISPR-CAS SYSTEMS IN GENERAL

[0069] In general, a Cas protein (used interchangeably herein with CRISPR protein, CRISPR enzyme, CRISPR-Cas protein, CRISPR-Cas enzyme, Cas, or CRISPR effector) and/or a guide sequence is a component of a CRISPR-Cas system. A CRISPR-Cas system 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 Cas, 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 an engineered system of the invention, the direct repeat may encompass naturally occurring sequences or non-naturally occurring sequences. The direct repeat of the invention is not limited to naturally occurring lengths and sequences. A direct repeat can be 36 nucleotides (nt) in length, but a longer or shorter direct repeat can vary. For example, a direct repeat can be 20nt or longer, such as 30-100 nt or longer. For example, a direct repeat can be, 20nt, 30 nt, 40nt, 50nt, 60nt, 70nt, 70nt, 80nt, 90nt, 100nt or longer in length. In some embodiments, a direct repeat of the invention can include synthetic nucleotide sequences inserted between the 5' and 3' ends of naturally occurring direct repeats. In certain embodiments, the inserted sequence may be self-complementary, for example, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% self-complementary. Furthermore, a direct repeat of the invention may include insertions of nucleotides such as an aptamer or sequences that bind to an adapter protein (for association with functional domains). In certain embodiments, one end of a direct repeat containing such an insertion is roughly the first half of a short DR and the end is roughly the second half of the short DR.

[0070] 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. 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 2Kb 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.

[0071] In general, a guide sequence (or spacer sequence) may be 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, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). In some embodiments, a guide sequence (or spacer 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. Preferably the guide sequence is 10-40 nucleotides long, such as 20-30 or 20-40 nucleotides long or longer, such as 30 nucleotides long or about 30 nucleotides long.

[0072] In certain embodiments, the guide sequence is 10-30 nucleotides long, such as 20-30 nucleotides long, such as 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.

[0073] In some CRISPR-Cas systems, 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 crRNA 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 crRNA 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.

[0074] In certain embodiments, modulations of cleavage efficiency can be exploited by introduction of mismatches, e.g., 1 or more mismatches, such as 1 or 2 mismatches between spacer sequence and target sequence, including the position of the mismatch along the spacer/target. The more central (i.e., not 3' or 5') for instance a double mismatch is, the more cleavage efficiency is affected. Accordingly, by choosing mismatch position along the spacer, cleavage efficiency can be modulated. By means of example, if less than 100 % cleavage of targets is desired (e.g., in a cell population), 1 or more, such as preferably 2 mismatches between spacer and target sequence may be introduced in the spacer sequences. The more central along the spacer of the mismatch position, the lower the cleavage percentage.

[0075] A CRISPR-Cas system or components thereof may be used for introducing one or more mutations in a target locus or nucleic acid sequence. 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) . 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). 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) . 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). 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). 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).

[0076] Typically, in the context of an endogenous CRISPR-Cas 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 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, but may depend on for instance secondary structure, in particular in the case of RNA targets. In some cases, 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 (if applicable) 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.

[0077] In particularly preferred embodiments according to the invention, the guide RNA (capable of guiding Cas to a target locus) may comprise (1) a guide sequence capable of hybridizing to a target locus (a polynucleotide target locus, such as an RNA target locus) in the eukaryotic cell; (2) a direct repeat (DR) sequence) which reside in a single RNA, i.e., an sgRNA (arranged in a 5' to 3' orientation) or crRNA.

[0078] With respect to general information on CRISPR-Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, AAV, and making and using thereof, including as to amounts and formulations, all useful in the practice of the instant invention, reference is made to: US Patents Nos. 8,999,641, 8,993,233, 8,945,839, 8,932,814, 8,906,616, 8,895,308, 8,889,418, 8,889,356, 8,871,445, 8,865,406, 8,795,965, 8,771,945 and 8,697,359; US Patent Publications US 2014-0310830 (US APP. Ser. No. 14/105,031), US 2014-0287938 A1 (U.S. App. Ser. No. 14/213,991), US 2014-0273234 A1 (U.S. App. Ser. No. 14/293,674), US2014-0273232 A1 (U.S. App. Ser. No. 14/290,575), US 2014-0273231 (U.S. App. Ser. No. 14/259,420), US 2014-0256046 A1 (U.S. App. Ser. No. 14/226,274), US 2014-0248702 A1 (U.S. App. Ser. No. 14/258,458), US 2014-0242700 A1 (U.S. App. Ser. No. 14/222,930), US 2014-0242699 A1 (U.S. App. Ser. No.

14/183,512), US 2014-0242664 A1 (U.S. App. Ser. No. 14/104,990), US 2014-0234972 A1 (U.S. App. Ser. No. 14/183,471), US 2014-0227787 A1 (U.S. App. Ser. No. 14/256,912), US 2014-0189896 A1 (U.S. App. Ser. No. 14/105,035), US 2014-0186958 (U.S. App. Ser. No. 14/105,017), US 2014-0186919 A1 (U.S. App. Ser. No. 14/104,977), US 2014-0186843 A1 (U.S. App. Ser. No. 14/104,900), US 2014-0179770 A1 (U.S. App. Ser. No. 14/104,837) and US 2014-0179006 A1 (U.S. App. Ser. No. 14/183,486), US 2014-0170753 (US App Ser No 14/183,429); European Patents EP 2 784 162 B1 and EP 2 771 468 B1; 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 PCT Patent Publications WO 2014/093661 (PCT/US2013/074743), WO 2014/093694 (PCT/US2013/074790), WO 2014/093595 (PCT/US2013/074611), WO 2014/093718 (PCT/US2013/074825), WO 2014/093709 (PCT/US2013/074812), WO 2014/093622 (PCT/US2013/074667), WO 2014/093635 (PCT/US2013/074691), WO 2014/093655 (PCT/US2013/074736), WO 2014/093712 (PCT/US2013/074819), WO 2014/093701 (PCT/US2013/074800), WO 2014/018423 (PCT/US2013/051418), WO 2014/204723 (PCT/US2014/041790), WO 2014/204724 (PCT/US2014/041800), WO 2014/204725 (PCT/US2014/041803), WO 2014/204726 (PCT/US2014/041804), WO 2014/204727 (PCT/US2014/041806), WO 2014/204728 (PCT/US2014/041808), WO 2014/204729 (PCT/US2014/041809). Reference is also made to US provisional patent applications 61/758,468; 61/802,174; 61/806,375; 61/814,263; 61/819,803 and 61/828,130, filed on January 30, 2013; March 15, 2013; March 28, 2013; April 20, 2013; May 6, 2013 and May 28, 2013 respectively. Reference is also made to US provisional patent application 61/836,123, filed on June 17, 2013. Reference is additionally made to US provisional patent applications 61/835,931, 61/835,936, 61/836,127, 61/836, 101, 61/836,080 and 61/835,973, each filed June 17, 2013. Further reference is made to US provisional patent applications 61/862,468 and 61/862,355 filed on August 5, 2013; 61/871,301 filed on August 28, 2013; 61/960,777 filed on September 25, 2013 and 61/961,980 filed on October 28, 2013. Reference is yet further made to: PCT Patent applications Nos: PCT/US2014/041803, PCT/US2014/041800, PCT/US2014/041809, PCT/US2014/041804 and PCT/US2014/041806, each filed June 10, 2014 6/10/14; PCT/US2014/041808 filed June 11, 2014; and PCT/US2014/62558 filed October 28, 2014, and US Provisional Patent Applications Serial Nos.: 61/915,150, 61/915,301, 61/915,267 and 61/915,260, each filed December 12, 2013; 61/757,972 and 61/768,959, filed on January 29, 2013 and February 25, 2013; 61/835,936, 61/836,127, 61/836,101, 61/836,080, 61/835,973, and 61/835,931, filed June 17, 2013; 62/010,888 and 62/010,879, both filed June 11, 2014; 62/010,329 and

62/010,441, each filed June 10, 2014; 61/939,228 and 61/939,242, each filed February 12, 2014; 61/980,012, filed April 15,2014; 62/038,358, filed August 17, 2014; 62/054,490, 62/055,484, 62/055,460 and 62/055,487, each filed September 25, 2014; and 62/069,243, filed October 27, 2014. Reference is also made to US provisional patent applications Nos. 62/055,484, 62/055,460, and 62/055,487, filed September 25, 2014; US provisional patent application 61/980,012, filed April 15, 2014; and US provisional patent application 61/939,242 filed February 12, 2014. Reference is made to PCT application designating, inter alia, the United States, application No. PCT/US14/41806, filed June 10, 2014. Reference is made to US provisional patent application 61/930,214 filed on January 22, 2014. Reference is made to US provisional patent applications 61/915,251; 61/915,260 and 61/915,267, each filed on December 12, 2013. Reference is made to US provisional patent application USSN 61/980,012 filed April 15, 2014. Reference is made to PCT application designating, inter alia, the United States, application No. PCT/US14/41806, filed June 10, 2014. Reference is made to US provisional patent application 61/930,214 filed on January 22, 2014. Reference is made to US provisional patent applications 61/915,251; 61/915,260 and 61/915,267, each filed on December 12, 2013.

[0079] Mention is also made of US application 62/091,455, filed, 12-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US application 62/096,708, 24-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US application 62/091,462, 12-Dec-14, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US application 62/096,324, 23-Dec-14, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US application 62/091,456, 12-Dec-14, ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS; US application 62/091,461, 12-Dec-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOIETIC STEM CELLS (HSCs); US application 62/094,903, 19-Dec-14, UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE SEQUENCING; US application 62/096,761, 24-Dec-14, ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION; US application 62/098,059, 30-Dec-14, RNA-TARGETING SYSTEM; US application 62/096,656, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS; US application 62/096,697, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH AAV; US application 62/098,158, 30-Dec-14, ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS; US

application 62/151,052, 22-Apr-15, CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING; US application 62/054,490, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS; US application 62/055,484, 25-Sep-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,537, 4-Dec-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/054,651, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US application 62/067,886, 23-Oct-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US application 62/054,675, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES; US application 62/054,528, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS; US application 62/055,454, 25-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP); US application 62/055,460, 25-Sep-14, MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; US application 62/087,475, 4-Dec-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,487, 25-Sep-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,546, 4-Dec-14, MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; and US application 62/098,285, 30-Dec-14, CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS.

[0080] The systems may comprise templates. Delivery of templates may be via the cotemporaneous or separate from delivery of any or all the Cas protein or guide or crRNA and via the same delivery mechanism or different.

[0081] In certain embodiments, the methods as described herein may comprise providing a Cas transgenic cell in which one or more nucleic acids encoding one or more guide RNAs are provided or introduced operably connected in the cell with a regulatory element comprising a promoter of one or more gene of interest. As used herein, the term "Cas transgenic cell" refers to a cell, such as a eukaryotic cell, in which a Cas gene has been genomically integrated. The nature, type, or origin of the cell are not particularly limiting according to the present invention. Also, the way how the Cas transgene is introduced in the cell is may vary and can be any method as is known in the art. In certain embodiments, the Cas transgenic cell is obtained by introducing the Cas transgene in an isolated cell. In certain other embodiments, the Cas transgenic cell is obtained by isolating cells from a Cas transgenic organism. By means of example, and without limitation, the Cas transgenic cell as referred to herein may be derived from a Cas transgenic eukaryote, such as a Cas knock-in eukaryote. Reference is made to WO 2014/093622 (PCT/US13/74667), incorporated herein by reference. Methods of US Patent Publication Nos. 20120017290 and 20110265198 assigned to Sangamo BioSciences, Inc. directed to targeting the Rosa locus may be modified to utilize the CRISPR Cas system of the present invention. Methods of US Patent Publication No. 20130236946 assigned to Cellectis directed to targeting the Rosa locus may also be modified to utilize the CRISPR Cas system of the present invention. By means of further example reference is made to Platt et. al. (Cell; 159(2):440-455 (2014)), describing a Cas9 knock-in mouse, which is incorporated herein by reference. The Cas transgene can further comprise a Lox-Stop-polyA-Lox(LSL) cassette thereby rendering Cas expression inducible by Cre recombinase. Alternatively, the Cas transgenic cell may be obtained by introducing the Cas transgene in an isolated cell. Delivery systems for transgenes are well known in the art. By means of example, the Cas transgene may be delivered in for instance eukaryotic cell by means of vector (e.g., AAV, adenovirus, lentivirus) and/or particle and/or particle delivery, as also described herein elsewhere.

[0082] It will be understood by the skilled person that the cell, such as the Cas transgenic cell, as referred to herein may comprise further genomic alterations besides having an integrated Cas gene or the mutations arising from the sequence specific action of Cas when complexed with RNA capable of guiding Cas to a target locus, such as for instance one or more oncogenic mutations, as for instance and without limitation described in Platt et al. (2014), Chen et al., (2014) or Kumar et al. (2009).

[0083] The nucleic acid molecule encoding a Cas may be codon optimized. 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. 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. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g. about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the "Codon Usage Database" available at kazusa.orjp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. In some embodiments, one or more codons (e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a Cas correspond to the most frequently used codon for a particular amino acid.

[0084] In some embodiments, the Cas proteins may have nucleic acid cleavage activity. The Cas proteins may have RNA binding and DNA cleaving function. In some embodiments,

Cas may direct cleavage of one or two 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, e.g., 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, the Cas protein may direct more than one cleavage (such as one, two three, four, five, or more cleavages) of one or two strands within the target sequence and/or within the complement of the target sequence or at sequences associated with the target sequence and/or 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, 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, 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 two strands of a target polynucleotide containing a target sequence, e.g., alteration or mutation in a HNH domain to produce a mutated Cas substantially lacking all DNA cleavage activity, e.g., 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 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. 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.

[0085] Typically, in the context of an endogenous nucleic acid-targeting system, formation of a nucleic acid-targeting complex (comprising a guide RNA or crRNA hybridized to a target sequence and complexed with one or more nucleic acid-targeting effector proteins) results in cleavage of DNA strand(s) 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).

[0086] It will be appreciated that the effector protein is based on or derived from an enzyme, so the term 'effector protein' certainly includes 'enzyme' in some embodiments. However, it will also be appreciated that the effector protein may, as required in some

embodiments, have DNA or RNA binding, but not necessarily cutting or nicking, activity, including a dead-Cas protein function.

[0087] In some embodiments, a Cas protein may form a component of an inducible system. The inducible nature of the system would allow for spatiotemporal control of gene editing or gene expression using a form of energy. The form of energy may include but is not limited to electromagnetic radiation, sound energy, chemical energy and thermal energy. Examples of inducible system include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc.), or light inducible systems (Phytochrome, LOV domains, or cryptochrome). In one embodiment, the CRISPR effector protein may be a part of a Light Inducible Transcriptional Effector (LITE) to direct changes in transcriptional activity in a sequence-specific manner. The components of a light may include a CRISPR effector protein, a light-responsive cytochrome heterodimer (e.g. from Arabidopsis thaliana), and a transcriptional activation/repression domain. Further examples of inducible DNA binding proteins and methods for their use are provided in US 61/736465 and US 61/721,283,and WO 2014018423 A2 which is hereby incorporated by reference in its entirety.

[0088] In one aspect, the invention provides a mutated Cas as described herein elsewhere, having one or more mutations resulting in reduced off-target effects, e.g., improved CRISPR enzymes for use in effecting modifications to target loci but which reduce or eliminate activity towards off-targets, such as when complexed to guide RNAs, as well as improved CRISPR enzymes for increasing the activity of CRISPR enzymes, such as when complexed with guide RNAs. It is to be understood that mutated enzymes as described herein below may be used in any of the methods according to the invention as described herein elsewhere. Any of the methods, products, compositions and uses as described herein elsewhere are equally applicable with the mutated CRISPR enzymes as further detailed below.

[0089] Slaymaker et al. recently described a method for the generation of Cas orthologues with enhanced specificity (Slaymaker et al. 2015 “Rationally engineered Cas nucleases with improved specificity”). This strategy can be used to enhance the specificity of the Cas protein. Primary residues for mutagenesis are preferably all positive charges residues within the HNH domain and/or DinG domain. Additional residues are positive charged residues that are conserved between different orthologues.

[0090] In an aspect, the invention also provides methods and mutations for modulating Cas binding activity and/or binding specificity. In certain embodiments Cas proteins lacking nuclease activity are used. In certain embodiments, modified guide RNAs are employed that promote binding but not nuclease activity of a Cas nuclease. In such embodiments, on-target

binding can be increased or decreased. Also, in such embodiments off-target binding can be increased or decreased. Moreover, there can be increased or decreased specificity as to on-target binding vs. off-target binding.

[0091] The methods and mutations which can be employed in various combinations to increase or decrease activity and/or specificity of on-target vs. off-target activity, or increase or decrease binding and/or specificity of on-target vs. off-target binding, can be used to compensate or enhance mutations or modifications made to promote other effects. Such mutations or modifications made to promote other effects in include mutations or modification to the Cas and or mutation or modification made to a guide RNA. The methods and mutations of the invention are used to modulate Cas nuclease activity and/or binding with chemically modified guide RNAs.

[0092] In an aspect, the invention provides methods and mutations for modulating binding and/or binding specificity of Cas proteins according to the invention as defined herein comprising functional domains such as nucleases, transcriptional activators, transcriptional repressors, and the like. For example, a Cas protein can be made nuclease-null, or having altered or reduced nuclease activity by introducing mutations such as for instance Cas mutations described herein elsewhere. Nuclease deficient Cas proteins are useful for RNA-guided target sequence dependent delivery of functional domains. The invention provides methods and mutations for modulating binding of Cas proteins. In one embodiment, the functional domain comprises VP64, providing an RNA-guided transcription factor. In another embodiment, the functional domain comprises Fok I, providing an RNA-guided nuclease activity. Mention is made of U.S. Pat. Pub. 2014/0356959, U.S. Pat. Pub. 2014/0342456, U.S. Pat. Pub. 2015/0031132, and Mali, P. et al., 2013, Science 339(6121):823-6, doi: 10.1126/science.1232033, published online 3 January 2013 and through the teachings herein the invention comprehends methods and materials of these documents applied in conjunction with the teachings herein. In certain embodiments, on-target binding is increased. In certain embodiments, off-target binding is decreased. In certain embodiments, on-target binding is decreased. In certain embodiments, off-target binding is increased. Accordingly, the invention also provides for increasing or decreasing specificity of on-target binding vs. off-target binding of functionalized Cas binding proteins.

[0093] The use of Cas as an RNA-guided binding protein is not limited to nuclease-null Cas. Cas enzymes comprising nuclease activity can also function as RNA-guided binding proteins when used with certain guide RNAs. For example, short guide RNAs and guide RNAs comprising nucleotides mismatched to the target can promote RNA directed Cas binding to a

target sequence with little or no target cleavage. (See, e.g., Dahlman, 2015, Nat Biotechnol. 33(11):1159-1161, doi: 10.1038/nbt.3390, published online 05 October 2015). In an aspect, the invention provides methods and mutations for modulating binding of Cas proteins that comprise nuclease activity. In certain embodiments, on-target binding is increased. In certain embodiments, off-target binding is decreased. In certain embodiments, on-target binding is decreased. In certain embodiments, off-target binding is increased. In certain embodiments, there is increased or decreased specificity of on-target binding vs. off-target binding. In certain embodiments, nuclease activity of guide RNA-Cas enzyme is also modulated.

[0094] RNA–RNA duplex formation is important for cleavage activity and specificity throughout the target region, not only the seed region sequence closest to the PAM. Thus, truncated guide RNAs show reduced cleavage activity and specificity. In an aspect, the invention provides method and mutations for increasing activity and specificity of cleavage using altered guide RNAs.

[0095] In certain embodiments, the catalytic activity of the Cas protein of the invention is altered or modified. It is to be understood that mutated Cas has an altered or modified catalytic activity if the catalytic activity is different than the catalytic activity of the corresponding wild type Cas protein (e.g., unmutated Cas protein). Catalytic activity can be determined by means known in the art. By means of example, and without limitation, catalytic activity can be determined in vitro or in vivo by determination of indel percentage (for instance after a given time, or at a given dose). In certain embodiments, catalytic activity is increased. In certain embodiments, catalytic activity is increased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100%. In certain embodiments, catalytic activity is decreased. In certain embodiments, catalytic activity is decreased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or (substantially) 100%. The one or more mutations herein may inactivate the catalytic activity, which may substantially all catalytic activity, below detectable levels, or no measurable catalytic activity.

[0096] One or more characteristics of the engineered Cas protein may be different from a corresponding wiled type Cas protein. Examples of such characteristics include catalytic activity, gRNA binding, specificity of the Cas protein (e.g., specificity of editing a defined target), stability of the Cas protein, off-target binding, target binding, protease activity, nickase activity, PFS recognition. In some examples, a engineered Cas protein may comprise one or more mutations of the corresponding wild type Cas protein. In some embodiments, the catalytic

activity of the engineered Cas protein is increased as compared to a corresponding wildtype Cas protein. In some embodiments, the catalytic activity of the engineered Cas protein is decreased as compared to a corresponding wildtype Cas protein. In some embodiments, the gRNA binding of the engineered Cas protein is increased as compared to a corresponding wildtype Cas protein. In some embodiments, the gRNA binding of the engineered Cas protein is decreased as compared to a corresponding wildtype Cas protein. In some embodiments, the specificity of the Cas protein is increased as compared to a corresponding wildtype Cas protein. In some embodiments, the specificity of the Cas protein is decreased as compared to a corresponding wildtype Cas protein. In some embodiments, the stability of the Cas protein is increased as compared to a corresponding wildtype Cas protein. In some embodiments, the stability of the Cas protein is decreased as compared to a corresponding wildtype Cas protein. In some embodiments, the engineered Cas protein further comprises one or more mutations which inactivate catalytic activity. In some embodiments, the off-target binding of the Cas protein is increased as compared to a corresponding wildtype Cas protein. In some embodiments, the off-target binding of the Cas protein is decreased as compared to a corresponding wildtype Cas protein. In some embodiments, the target binding of the Cas protein is increased as compared to a corresponding wildtype Cas protein. In some embodiments, the target binding of the Cas protein is decreased as compared to a corresponding wildtype Cas protein. In some embodiments, the engineered Cas protein has a higher protease activity or polynucleotide-binding capability compared with a corresponding wildtype Cas protein. In some embodiments, the PFS recognition is altered as compared to a corresponding wildtype Cas protein.

[0097] In some embodiments, such as for Cas, a non-naturally occurring or engineered composition of the invention may comprise an accessory protein that enhances the Cas protein activity. In such embodiments, the Cas protein and the accessory protein may be from the same source or from a different source. In some embodiments, a non-naturally occurring or engineered composition of the invention comprises an accessory protein that represses Cas protein activity. In some embodiments, a non-naturally occurring or engineered composition of the invention comprises two or more crRNAs. In some embodiments, a non-naturally occurring or engineered composition of the invention comprises a guide sequence that hybridizes to a target RNA sequence in a prokaryotic cell. In some embodiments, a non-naturally occurring or engineered composition of the invention comprises a guide sequence that hybridizes to a target RNA sequence in a eukaryotic cell. In some embodiment, the Cas protein comprises one or more nuclear localization signals (NLSs).

[0098] In some embodiment of the non-naturally occurring or engineered composition of the invention, the Cas protein and the accessory protein are from the same organism.

[0099] In some embodiment of the non-naturally occurring or engineered composition of the invention, the Cas protein and the accessory protein are from different organisms.

[00100] In certain embodiments, the Cas proteins herein may be associated with a locus comprising short CRISPR repeats between 30 and 40 bp long, more typically between 34 and 38 bp long, even more typically between 36 and 37 bp long, e.g., 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, or 40 bp long. In certain embodiments the CRISPR repeats are long or dual repeats between 80 and 350 bp long such as between 80 and 200 bp long, even more typically between 86 and 88 bp long, e.g., 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, or 90 bp long

[00101] In certain embodiments, the guide RNA or mature crRNA comprises, consists essentially of, or consists of a direct repeat sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or mature crRNA comprises, consists essentially of, or consists of a direct repeat sequence linked to a guide sequence or spacer sequence. In preferred embodiments of the invention, the mature crRNA comprises a stem loop or an optimized stem loop structure or an optimized secondary structure. In preferred embodiments the mature crRNA comprises a stem loop or an optimized stem loop structure in the direct repeat sequence, wherein the stem loop or optimized stem loop structure is important for cleavage activity. In certain embodiments, the mature crRNA preferably comprises a single stem loop. In certain embodiments, the direct repeat sequence preferably comprises a single stem loop. In certain embodiments, the cleavage activity of the effector protein complex is modified by introducing mutations that affect the stem loop RNA duplex structure. In preferred embodiments, mutations which maintain the RNA duplex of the stem loop may be introduced, whereby the cleavage activity of the effector protein complex is maintained. In other preferred embodiments, mutations which disrupt the RNA duplex structure of the stem loop may be introduced, whereby the cleavage activity of the effector protein complex is completely abolished.

[00102] The CRISPR system as provided herein can make use of a crRNA or analogous polynucleotide comprising a guide sequence, wherein the polynucleotide is an RNA, a DNA or a mixture of RNA and DNA, and/or wherein the polynucleotide comprises one or more nucleotide analogs. The sequence can comprise any structure, including but not limited to a structure of a native crRNA, such as a bulge, a hairpin or a stem loop structure. In certain embodiments, the polynucleotide comprising the guide sequence forms a duplex with a second polynucleotide sequence which can be an RNA or a DNA sequence.

EXEMPLARY CAS PROTEINS

[00103] The Cas proteins herein, e.g. a Type IV protein, may comprise an HNH domain. In some examples, the Cas proteins further comprise a helicase domain. The helicase domain may be a DinG domain. In some embodiments, the Cas gene is found in several diverse bacterial genomes, typically in the same locus with Cas5, Cas6, and Cas7, and/or Cas8 genes and a CRISPR cassette. In some examples, the Cas protein contains an HNH domain and a helicase domain (e.g., a DinG domain). Nucleic acid sequences of exemplary Type IV Cas proteins include those in Table 1 and SEQ ID NOs 1-405.

[00104] In some embodiments, the HNH domain in the Cas protein may be less than 1000 amino acids, less than 900 amino acids, less than 800 amino acids, less than 700 amino acids, less than 600 amino acids, less than 500 amino acids, less than 400 amino acids, less than 300 amino acids, less than 200 amino acids, or less than 100 amino acids. In some examples, the HNH domain is less than 600 amino acids. In a particular example, the Cas protein comprises an HNH domain that is less than 600 amino acids in size in length, and at least one guide sequence capable of complexing with the Cas protein and directing binding of the guide-Cas protein complex to a target polynucleotide.

[00105] In some embodiments, the Cas proteins are a class 1, Type IV Cas, an ortholog thereof, or a homolog thereof. The systems and compositions may comprise orthologs and homologs of the Cas proteins. The terms “ortholog” and “homolog” are well known in the art. By means of further guidance, a “homologue” of a protein as used herein is a protein of the same species which performs the same or a similar function as the protein it is a homologue of. Homologous proteins may but need not be structurally related, or are only partially structurally related. An “ortholog” of a protein as used herein is a protein of a different species which performs the same or a similar function as the protein it is an orthologue of. Orthologous proteins may but need not be structurally related, or are only partially structurally related. Homologs and orthologs may be identified by homology modelling (see, e.g., Greer, Science vol. 228 (1985) 1055, and Blundell et al. Eur J Biochem vol 172 (1988), 513) or "structural BLAST" (Dey F, Cliff Zhang Q, Petrey D, Honig B. Toward a "structural BLAST": using structural relationships to infer function. Protein Sci. 2013 Apr;22(4):359-66. doi: 10.1002/pro.2225.). See also Shmakov et al. (2015) for application in the field of CRISPR-Cas loci. Homologous proteins may but need not be structurally related or are only partially structurally related. As used herein, when a Cas protein originates form a species, it may be the wild type Cas protein in the species, or a homolog of the wild type Cas protein in the species.

The Cas protein that is a homolog of the wild type Cas protein in the species may comprise one or more variations (e.g., mutations, truncations, etc.) of the wild type Cas protein.

[00106] It will be appreciated that any of the functionalities described herein may be engineered into Cas proteins from other orthologs, including chimeric enzymes comprising fragments from multiple orthologs. A chimeric enzyme can comprise a first fragment and a second fragment, and the fragments can be of CRISPR enzyme orthologs of organisms of genera herein mentioned or of species herein mentioned; advantageously, the fragments are from CRISPR enzyme orthologs of different species.

[00107] Orthologous proteins may but need not be structurally related, or are only partially structurally related. In particular embodiments, the homolog or ortholog of a Cas protein as referred to herein has a sequence homology or identity of at least 60%, preferably at least 70%, preferably at least 80%, more preferably at least 85%, even more preferably at least 90%, such as for instance at least 95% with a Cas proteins encoded by the nucleic acid sequences in Table 1 and SEQ ID NOs: 1-405.

CAS VARIANTS

[00108] The Cas proteins herein include variants and mutated forms of Cas proteins (comparing to wildtype or naturally occurring Cas proteins). In some examples, the present disclosure includes variants and mutated forms of the Cas proteins. The variants or mutated forms of Cas protein may be catalytically inactive, e.g., have no or reduced nuclease activity compared to a corresponding wildtype. In certain examples, the variants or mutated forms of Cas protein have nickase activity.

[00109] In some cases, the present disclosure provides for mutated Cas proteins comprising one or more modified of amino acids. The amino acids: (a) interact with a guide RNA that forms a complex with the mutated Cas protein; (b) are in an active site, an inter-domain linker domain, or a bridge helix domain of the mutated Cas protein; or (c) a combination thereof.

[00110] The term "corresponding amino acid" or "residue which corresponds to" refers to a particular amino acid or analogue thereof in a Cas homolog or ortholog that is identical or functionally equivalent to an amino acid in reference Cas protein. Accordingly, as used herein, referral to an "amino acid position corresponding to amino acid position [X]" of a specified Cas protein represents referral to a collection of equivalent positions in other recognized Cas and structural homologues and families.

[00111] In certain embodiments, the specificity of the Cas protein of the invention is altered or modified. It is to be understood that mutated Cas has an altered or modified specificity if the specificity is different than the specificity of the corresponding wild type Cas (i.e., unmutated

Cas). Specificity can be determined by means known in the art. By means of example, and without limitation, specificity can be determined by comparison of on-target activity and off-target activity. In certain embodiments, specificity is increased. In certain embodiments, specificity is increased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100%. In certain embodiments, specificity is decreased. In certain embodiments, specificity is decreased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or (substantially) 100%.

[00112] In certain embodiments, the stability of the Cas protein of the invention is altered or modified. It is to be understood that mutated Cas has an altered or modified stability if the stability is different than the stability of the corresponding wild type Cas (i.e., unmutated Cas). Stability can be determined by means known in the art. By means of example, and without limitation, stability can be determined by determining the half-life of the Cas protein. In certain embodiments, stability is increased. In certain embodiments, stability is increased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100%. In certain embodiments, stability is decreased. In certain embodiments, stability is decreased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or (substantially) 100%.

[00113] In certain embodiments, the target binding of the Cas protein of the invention is altered or modified. It is to be understood that mutated Cas has an altered or modified target binding if the target binding is different than the target binding of the corresponding wild type Cas (i.e. unmutated Cas). target binding can be determined by means known in the art. By means of example, and without limitation, target binding can be determined by calculating binding strength or affinity (such as based on equilibrium constants, Ka, Kd, etc.). In certain embodiments, target bindings increased. In certain embodiments, target binding is increased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100%. In certain embodiments, target binding is decreased. In certain embodiments, target binding is decreased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or (substantially) 100%.

[00114] In certain embodiments, the off-target binding of the Cas protein of the invention is altered or modified. It is to be understood that mutated Cas has an altered or modified off-target binding if the off-target binding is different than the off-target binding of the corresponding wild type Cas (i.e. unmutated Cas). Off-target binding can be determined by means known in the art. By means of example, and without limitation, off-target binding can be determined by calculating binding strength or affinity (such as based on equilibrium constants, Ka, Kd, etc.). In certain embodiments, off-target bindings increased. In certain embodiments, off-target binding is increased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100%. In certain embodiments, off-target binding is decreased. In certain embodiments, off-target binding is decreased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or (substantially) 100%.

[00115] The types of mutations in the Cas proteins can be conservative mutations or non-conservative mutations. In certain preferred embodiments, the amino acid which is mutated is mutated into alanine (A). In certain preferred embodiments, if the amino acid to be mutated is an aromatic amino acid, it is mutated into alanine or another aromatic amino acid (e.g., H, Y, W, or F). In certain preferred embodiments, if the amino acid to be mutated is a charged amino acid, it is mutated into alanine or another charged amino acid (e.g., H, K, R, D, or E). In certain preferred embodiments, if the amino acid to be mutated is a charged amino acid, it is mutated into alanine or another charged amino acid having the same charge. In certain preferred embodiments, if the amino acid to be mutated is a charged amino acid, it is mutated into alanine or another charged amino acid having the opposite charge.

[00116] The invention also provides for methods and compositions wherein one or more amino acid residues of the effector protein may be modified e.g., an engineered or non-naturally-occurring effector protein or Cas. In an embodiment, the modification may comprise mutation of one or more amino acid residues of the effector protein. The one or more mutations may be in one or more catalytically active domains of the effector protein, or a domain interacting with the crRNA (such as the guide sequence or direct repeat sequence). The effector protein may have reduced or abolished nuclease activity or alternatively increased nuclease activity compared with an effector protein lacking said one or more mutations. The effector protein may not direct cleavage of the RNA strand at the target locus of interest. In a preferred embodiment, the one or more mutations may comprise two mutations.

[00117] The Cas proteins herein may comprise one or more amino acids mutated. In some embodiments, the amino acid is mutated to A, P, or V, preferably A. In some embodiments, the amino acid is mutated to a hydrophobic amino acid. In some embodiments, the amino acid is mutated to an aromatic amino acid. In some embodiments, the amino acid is mutated to a charged amino acid. In some embodiments, the amino acid is mutated to a positively charged amino acid. In some embodiments, the amino acid is mutated to a negatively charged amino acid. In some embodiments, the amino acid is mutated to a polar amino acid. In some embodiments, the amino acid is mutated to an aliphatic amino acid.

Structural (sub)domains

[00118] In another aspect, the disclosure provides a mutated Cas protein comprising one or more mutations of amino acids, wherein the amino acids: interact with a guide RNA that forms a complex with the engineered Cas protein; or are in an active site, e.g., in HNH domain(s) and/or helicase domains(s), e.g., DinG domains.

[00119] In another aspect, the disclosure provides a mutated Cas protein comprising one or more mutations of amino acids, wherein the amino acids: interact with a guide RNA that forms a complex with the engineered Cas protein; or are in an active site, e.g., in HNH domain(s).

DESTABILIZED CAS AND FUSION PROTEINS

[00120] In certain embodiments, the Cas protein according to the invention as described herein is associated with or fused to a destabilization domain (DD). In some embodiments, the DD is ER50. A corresponding stabilizing ligand for this DD is, in some embodiments, 4HT. As such, in some embodiments, one of the at least one DDs is ER50 and a stabilizing ligand therefor is 4HT or CMP8. In some embodiments, the DD is DHFR50. A corresponding stabilizing ligand for this DD is, in some embodiments, TMP. As such, in some embodiments, one of the at least one DDs is DHFR50 and a stabilizing ligand therefor is TMP. In some embodiments, the DD is ER50. A corresponding stabilizing ligand for this DD is, in some embodiments, CMP8. CMP8 may therefore be an alternative stabilizing ligand to 4HT in the ER50 system. While it may be possible that CMP8 and 4HT can/should be used in a competitive matter, some cell types may be more susceptible to one or the other of these two ligands, and from this disclosure and the knowledge in the art the skilled person can use CMP8 and/or 4HT.

[00121] In some embodiments, one or two DDs may be fused to the N- terminal end of the Cas with one or two DDs fused to the C- terminal of the Cas. In some embodiments, the at least two DDs are associated with the Cas and the DDs are the same DD, i.e. the DDs are

homologous. Thus, both (or two or more) of the DDs could be ER50 DDs. This is preferred in some embodiments. Alternatively, both (or two or more) of the DDs could be DHFR50 DDs. This is also preferred in some embodiments. In some embodiments, the at least two DDs are associated with the Cas and the DDs are different DDs, i.e. the DDs are heterologous. Thus, one of the DDS could be ER50 while one or more of the DDs or any other DDs could be DHFR50. Having two or more DDs which are heterologous may be advantageous as it would provide a greater level of degradation control. A tandem fusion of more than one DD at the N or C-term may enhance degradation; and such a tandem fusion can be, for example ER50-ER50-Cas or DHFR-DHFR-Cas It is envisaged that high levels of degradation would occur in the absence of either stabilizing ligand, intermediate levels of degradation would occur in the absence of one stabilizing ligand and the presence of the other (or another) stabilizing ligand, while low levels of degradation would occur in the presence of both (or two of more) of the stabilizing ligands. Control may also be imparted by having an N-terminal ER50 DD and a C-terminal DHFR50 DD.

[00122] In some embodiments, the fusion of the Cas with the DD comprises a linker between the DD and the Cas. In some embodiments, the linker is a GlySer linker. In some embodiments, the DD-Cas further comprises at least one Nuclear Export Signal (NES). In some embodiments, the DD- Cas comprises two or more NESs. In some embodiments, the DD- Cas comprises at least one Nuclear Localization Signal (NLS). This may be in addition to an NES. In some embodiments, the Cas comprises or consists essentially of or consists of a localization (nuclear import or export) signal as, or as part of, the linker between the Cas and the DD. HA or Flag tags are also within the ambit of the invention as linkers. Applicants use NLS and/or NES as linker and also use Glycine Serine linkers as short as GS up to $(GGGGS)_3$.

[00123] Destabilizing domains have general utility to confer instability to a wide range of proteins; see, e.g., Miyazaki, J Am Chem Soc. Mar 7, 2012; 134(9): 3942–3945, incorporated herein by reference. CMP8 or 4-hydroxytamoxifen can be destabilizing domains. More generally, A temperature-sensitive mutant of mammalian DHFR (DHFRts), a destabilizing residue by the N-end rule, was found to be stable at a permissive temperature but unstable at 37 °C. The addition of methotrexate, a high-affinity ligand for mammalian DHFR, to cells expressing DHFRts inhibited degradation of the protein partially. This was an important demonstration that a small molecule ligand can stabilize a protein otherwise targeted for degradation in cells. A rapamycin derivative was used to stabilize an unstable mutant of the FRB domain of mTOR (FRB*) and restore the function of the fused kinase, GSK-3β.6,7 This system demonstrated that ligand-dependent stability represented an attractive strategy to

regulate the function of a specific protein in a complex biological environment. A system to control protein activity can involve the DD becoming functional when the ubiquitin complementation occurs by rapamycin induced dimerization of FK506-binding protein and FKBP12. Mutants of human FKBP12 or ecDHFR protein can be engineered to be metabolically unstable in the absence of their high-affinity ligands, Shield-1 or trimethoprim (TMP), respectively. These mutants are some of the possible destabilizing domains (DDs) useful in the practice of the invention and instability of a DD as a fusion with a Cas confers to the Cas degradation of the entire fusion protein by the proteasome. Shield-1 and TMP bind to and stabilize the DD in a dose-dependent manner. The estrogen receptor ligand binding domain (ERLBD, residues 305-549 of ERS1) can also be engineered as a destabilizing domain. Since the estrogen receptor signaling pathway is involved in a variety of diseases such as breast cancer, the pathway has been widely studied and numerous agonist and antagonists of estrogen receptor have been developed. Thus, compatible pairs of ERLBD and drugs are known. There are ligands that bind to mutant but not wild-type forms of the ERLBD. By using one of these mutant domains encoding three mutations (L384M, M421G, G521R)12, it is possible to regulate the stability of an ERLBD-derived DD using a ligand that does not perturb endogenous estrogen-sensitive networks. An additional mutation (Y537S) can be introduced to further destabilize the ERLBD and to configure it as a potential DD candidate. This tetra-mutant is an advantageous DD development. The mutant ERLBD can be fused to a Cas and its stability can be regulated or perturbed using a ligand, whereby the Cas has a DD. Another DD can be a 12-kDa (107-amino-acid) tag based on a mutated FKBP protein, stabilized by Shield1 ligand; see, e.g., Nature Methods 5, (2008). For instance a DD can be a modified FK506 binding protein 12 (FKBP12) that binds to and is reversibly stabilized by a synthetic, biologically inert small molecule, Shield-1; see, e.g., Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. Cell. 2006;126:995–1004; Banaszynski LA, Sellmyer MA, Contag CH, Wandless TJ, Thorne SH. Chemical control of protein stability and function in living mice. Nat Med. 2008;14:1123–1127; Maynard-Smith LA, Chen LC, Banaszynski LA, Ooi AG, Wandless TJ. A directed approach for engineering conditional protein stability using biologically silent small molecules. The Journal of biological chemistry. 2007;282:24866–24872; and Rodriguez, Chem Biol. Mar 23, 2012; 19(3): 391–398—all of which are incorporated herein by reference and may be employed in the practice of the invention in selected a DD to associate with a Cas in the practice of this invention. As can be seen, the knowledge in the art includes a number of DDs, and the DD can be associated with, e.g., fused

to, advantageously with a linker, to a Cas, whereby the DD can be stabilized in the presence of a ligand and when there is the absence thereof the DD can become destabilized, whereby the Cas is entirely destabilized, or the DD can be stabilized in the absence of a ligand and when the ligand is present the DD can become destabilized; the DD allows the Cas and hence the CRISPR-Cas complex or system to be regulated or controlled—turned on or off so to speak, to thereby provide means for regulation or control of the system, e.g., in an in vivo or in vitro environment. For instance, when a protein of interest is expressed as a fusion with the DD tag, it is destabilized and rapidly degraded in the cell, e.g., by proteasomes. Thus, absence of stabilizing ligand leads to a D associated Cas being degraded. When a new DD is fused to a protein of interest, its instability is conferred to the protein of interest, resulting in the rapid degradation of the entire fusion protein. Peak activity for Cas is sometimes beneficial to reduce off-target effects. Thus, short bursts of high activity are preferred. The present invention is able to provide such peaks. In some senses the system is inducible. In some other senses, the system repressed in the absence of stabilizing ligand and de-repressed in the presence of stabilizing ligand.

DEACTIVATED / INACTIVATED / DEAD CAS PROTEINS

[00124] In certain embodiments, the Cas protein herein is a catalytically inactive or dead Cas protein. In some cases, Cas protein herein is a catalytically inactive or dead Cas protein (dCas). In some cases, a dead Cas protein, e.g., a dead Cas protein has nickase activity. In some embodiments, the dCas protein comprises mutations in the nuclease domain. In some embodiments, the dCas protein has been truncated. In some cases, the dead Cas proteins may be fused with a deaminase herein, e.g., an adenosine deaminase.

[00125] Where the Cas protein has nuclease activity, the Cas protein may be modified to have diminished nuclease activity e.g., nuclease inactivation of at least 70%, at least 80%, at least 90%, at least 95%, at least 97%, or 100% as compared with the wild type enzyme; or to put in another way, a Cas enzyme having advantageously about 0% of the nuclease activity of the non-mutated or wild type Cas, or no more than about 3% or about 5% or about 10% of the nuclease activity of the non-mutated or wild type Cas. This is possible by introducing mutations into the nuclease domains of the Cas and orthologs thereof.

[00126] The inactivated Cas CRISPR enzyme may have associated (e.g., via fusion protein) one or more functional domains, including for example, one or more domains from the group comprising, consisting essentially of, or consisting of methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, DNA cleavage activity, nucleic

acid binding activity, and molecular switches (e.g., light inducible). Preferred domains are Fok1, VP64, P65, HSF1, MyoD1. In the event that Fok1 is provided, it is advantageous that multiple Fok1 functional domains are provided to allow for a functional dimer and that gRNAs are designed to provide proper spacing for functional use (Fok1) as specifically described in Tsai et al. Nature Biotechnology, Vol. 32, Number 6, June 2014). The adaptor protein may utilize known linkers to attach such functional domains. In some cases it is advantageous that additionally at least one NLS is provided. In some instances, it is advantageous to position the NLS at the N terminus. When more than one functional domain is included, the functional domains may be the same or different.

[00127] In general, the positioning of the one or more functional domain on the inactivated Cas enzyme is one which allows for correct spatial orientation for the functional domain to affect the target with the attributed functional effect. For example, if the functional domain is a transcription activator (e.g., VP64 or p65), the transcription activator is placed in a spatial orientation which allows it to affect the transcription of the target. Likewise, a transcription repressor will be advantageously positioned to affect the transcription of the target, and a nuclease (e.g., Fok1) will be advantageously positioned to cleave or partially cleave the target. This may include positions other than the N- / C- terminus of the CRISPR enzyme.

[00128] The dead or deactivated Cas proteins may be used as target-binding proteins, (e.g., DNA binding proteins). In these cases, the dead or deactivated Cas proteins may be fused with one or more functional domains.

FUNCTIONAL DOMAINS

[00129] The systems and compositions provided herein may comprise one or more of the Cas proteins associated with one or more functional domains. In certain embodiments, the systems and compositions comprise fusion proteins comprising the Cas proteins(s)/subunit(s) associated with the functional domain(s).

[00130] In one aspect, the CRISPR-Cas system may be a class I, Type IV CRISPR-Cas system comprising a fusion of multimeric subunit components.

[00131] In one embodiment, the class I, Type IV CRISPR-Cas system comprises a fusion of a functional domain at the N-terminus, the C-terminus, or any accessible part (e.g. loop) of the protein complex.

[00132] In one embodiment, the class I, Type IV CRISPR-Cas system comprises fusion of one or more of the DinG, Cas5, Cas6 (e.g., cas6b, cas6e, cas6e/f), Cas7, Cas8 and Cas8-like, Cas10-like, Cas11, Cas2 and Cas4 subunits in any combination thereof.

[00133] In one aspect, the CRISPR-Cas system may be a class I, Type I CRISPR-Cas system comprising a fusion of multimeric subunit components.

[00134] In one embodiment, the class I, Type I CRISPR-Cas system comprises a fusion of a functional domain at the N-terminus, the C-terminus, or any accessible part (e.g. loop) of the protein complex.

[00135] In one embodiment the class I, Type I CRISPR-Cas system comprises fusion of the Cas3 Cas5, Cas6, Cas7, Cas8 and Cas11 subunits, and in any combination thereof.

[00136] In some embodiments, one or more functional domains are associated with the Cas system protein subunits. In some embodiments, one or more functional domains are associated with an adaptor protein, for example as used with the modified guides of Konnerman et al. (Nature 517, 583–588, 29 January 2015). In some embodiments, one or more functional domains are associated with a dead gRNA (dRNA). In some embodiments, a dRNA complex with active Cas system/protein subunit(s) directs gene regulation by a functional domain at on gene locus while an gRNA directs DNA cleavage by the active Cas protein at another locus, for example as described analogously in CRISPR-Cas systems by Dahlman et al., 'Orthogonal gene control with a catalytically active Cas9 nuclease'. In some embodiments, dRNAs are selected to maximize selectivity of regulation for a gene locus of interest compared to off-target regulation. In some embodiments, dRNAs are selected to maximize target gene regulation and minimize target cleavage.

[00137] For the purposes of the following discussion, reference to a functional domain could be a functional domain associated with one or more Cas protein subunits of the Cas system or a functional domain associated with the adaptor protein.

[00138] In the practice of the invention, loops of the gRNA may be extended, without colliding with the Cas protein by the insertion of distinct RNA loop(s) or distinct sequence(s) that may recruit adaptor proteins that can bind to the distinct RNA loop(s) or distinct sequence(s). The adaptor proteins may include but are not limited to orthogonal RNA-binding protein / aptamer combinations that exist within the diversity of bacteriophage coat proteins. A list of such coat proteins includes, but is not limited to: Qβ, F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕCb5, ϕCb8r, ϕCb12r, ϕCb23r, 7s and PRR1. These adaptor proteins or orthogonal RNA binding proteins can further recruit effector proteins or fusions which comprise one or more functional domains.

[00139] In some embodiments, the functional domain may be selected from the group consisting of: transposase domain, integrase domain, recombinase domain, resolvase domain,

invertase domain, protease domain, DNA methyltransferase domain, DNA hydroxylmethylase domain, DNA demethylase domain, histone acetylase domain, histone deacetylases domain, nuclease domain, repressor domain, activator domain, nuclear-localization signal domains, transcription-regulatory protein (or transcription complex recruiting) domain, cellular uptake activity associated domain, nucleic acid binding domain, antibody presentation domain, histone modifying enzymes, recruiter of histone modifying enzymes; inhibitor of histone modifying enzymes, histone methyltransferase, histone demethylase, histone kinase, histone phosphatase, histone ribosylase, histone deribosylase, histone ubiquitinase, histone deubiquitinase, histone biotinase and histone tail protease. In some preferred embodiments, the functional domain is a transcriptional activation domain, such as, without limitation, VP64, p65, MyoD1, HSF1, RTA, SET7/9 or a histone acetyltransferase. In some embodiments, the functional domain is a transcription repression domain, preferably KRAB. In some embodiments, the transcription repression domain is SID, or concatemers of SID (eg SID4X). In some embodiments, the functional domain is an epigenetic modifying domain, such that an epigenetic modifying enzyme is provided. In some embodiments, the functional domain is an activation domain, which may be the P65 activation domain.

[00140] In some examples, the Cas is associated with a ligase or functional fragment thereof. The ligase may ligate a single-strand break (a nick) generated by the Cas. In certain cases, the ligase may ligate a double-strand break generated by the Cas. In certain examples, the Cas is associated with a reverse transcriptase or functional fragment thereof.

[00141] In some embodiments, the one or more functional domains is an NLS (Nuclear Localization Sequence) or an NES (Nuclear Export Signal). In some embodiments, the one or more functional domains is a transcriptional activation domain comprises VP64, p65, MyoD1, HSF1, RTA, SET7/9 and a histone acetyltransferase. Other references herein to activation (or activator) domains in respect of those associated with the CRISPR enzyme include any known transcriptional activation domain and specifically VP64, p65, MyoD1, HSF1, RTA, SET7/9 or a histone acetyltransferase.

[00142] In some embodiments, the one or more functional domains is a transcriptional repressor domain. In some embodiments, the transcriptional repressor domain is a KRAB domain. In some embodiments, the transcriptional repressor domain is a NuE domain, NcoR domain, SID domain or a SID4X domain.

[00143] In some embodiments, the one or more functional domains have one or more activities comprising methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification

activity, RNA cleavage activity, DNA cleavage activity, DNA integration activity or nucleic acid binding activity.

[00144] Histone modifying domains are also preferred in some embodiments. Exemplary histone modifying domains are discussed below. Transposase domains, HR (Homologous Recombination) machinery domains, recombinase domains, and/or integrase domains are also preferred as the present functional domains. In some embodiments, DNA integration activity includes HR machinery domains, integrase domains, recombinase domains and/or transposase domains. Histone acetyltransferases are preferred in some embodiments.

[00145] In some embodiments, the DNA cleavage activity is due to a nuclease. In some embodiments, the nuclease comprises a Fok1 nuclease. See, "Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing", Shengdar Q. Tsai, Nicolas Wyvekens, Cyd Khayter, Jennifer A. Foden, Vishal Thapar, Deepak Reyon, Mathew J. Goodwin, Martin J. Aryee, J. Keith Joung Nature Biotechnology 32(6): 569-77 (2014), relates to dimeric RNA-guided FokI Nucleases that recognize extended sequences and can edit endogenous genes with high efficiencies in human cells.

[00146] In some embodiments, the one or more functional domains is attached to the Cas protein so that upon binding to the sgRNA and target the functional domain is in a spatial orientation allowing for the functional domain to function in its attributed function.

[00147] Functional domains may be used to regulate transcription, e.g., transcriptional repression. Transcriptional repression is often mediated by chromatin modifying enzymes such as histone methyltransferases (HMTs) and deacetylases (HDACs). Repressive histone effector domains are known and an exemplary list is provided below. In the exemplary table, preference was given to proteins and functional truncations of small size to facilitate efficient viral packaging (for instance via AAV). In general, however, the domains may include HDACs, histone methyltransferases (HMTs), and histone acetyltransferase (HAT) inhibitors, as well as HDAC and HMT recruiting proteins. The functional domain may be or include, in some embodiments, HDAC Effector Domains, HDAC Recruiter Effector Domains, Histone Methyltransferase (HMT) Effector Domains, Histone Methyltransferase (HMT) Recruiter Effector Domains, or Histone Acetyltransferase Inhibitor Effector Domains.

[00148] Table 3. HDAC Effector Domains

Subtype/ Complex	Name	Substrate (if known)	Modification (if known)	Organism	Full size (aa)	Selected truncation (aa)	Final size (aa)	Catalytic domain
HDAC I	**HDAC8**	-	-	*X. laevis*	325	1-325	325	1-272: HDAC
HDAC I	**RPD3**	-	-	*S. cerevisiae*	433	19-340	322 (Vannier)	19-331: HDAC
HDAC IV	**MesoLo4**	-	-	*M. loti*	300	1-300 (Gregoretti)	300	-
HDAC IV	**HDAC11**	-	-	*H. sapiens*	347	1-347 (Gao)	347	14-326: HDAC
HD2	**HDT1**	-	-	*A. thaliana*	245	1-211 (Wu)	211	-
SIRT I	**SIRT3**	H3K9Ac H4K16Ac H3K56Ac	-	*H. sapiens*	399	143-399 (Scher)	257	126-382: SIRT
SIRT I	**HST2**	-	-	*C. albicans*	331	1-331 (Hnisz)	331	-
SIRT I	**CobB**	-	-	*E. coli (K12)*	242	1-242 (Landry)	242	-
SIRT I	**HST2**	-	-	*S. cerevisiae*	357	8-298 (Wilson)	291	-
SIRT III	**SIRT5**	H4K8Ac H4K16Ac	-	*H. sapiens*	310	37-310 (Gertz)	274	41-309: SIRT
SIRT III	**Sir2A**	-	-	*P. falciparum*	273	1-273 (Zhu)	273	19-273: SIRT
SIRT IV	**SIRT6**	H3K9Ac H3K56Ac	-	*H. sapiens*	355	1-289 (Tennen)	289	35-274: SIRT

[00149] Accordingly, the repressor domains of the present invention may be selected from histone methyltransferases (HMTs), histone deacetylases (HDACs), histone acetyltransferase (HAT) inhibitors, as well as HDAC and HMT recruiting proteins.

[00150] The HDAC domain may be any of those in the Table 3 above, namely: HDAC8, RPD3, MesoLo4, HDAC11, HDT1, SIRT3, HST2, CobB, HST2, SIRT5, Sir2A, or SIRT6.

[00151] In some embodiment, the functional domain may be a HDAC Recruiter Effector Domain. Preferred examples include those in the Table 4 below, namely MeCP2, MBD2b, Sin3a, NcoR, SALL1, RCOR1. NcoR is exemplified in the present Examples and, although preferred, it is envisaged that others in the class will also be useful.

[00152] Table 4. HDAC Recruiter Effector Domains

Subtype/ Complex	Name	Substrate (if known)	Modification (if known)	Organism	Full size (aa)	Selected truncation (aa)	Final size (aa)	Catalytic domain
Sin3a	**MeCP2**	-	-	*R. norvegicus*	492	207-492 (Nan)	286	-
Sin3a	**MBD2b**	-	-	*H. sapiens*	262	45-262 (Boeke)	218	-
Sin3a	**Sin3a**	-	-	*H. sapiens*	1273	524-851 (Laherty)	328	627-829: HDAC1 interaction
NcoR	**NcoR**	-	-	*H. sapiens*	2440	420-488 (Zhang)	69	-
NuRD	**SALL1**	-	-	*M. musculus*	1322	1-93 (Lauberth)	93	-
CoREST	**RCOR1**	-	-	*H. sapiens*	482	81-300 (Gu, Ouyang)	220	-

[00153] In some embodiment, the functional domain may be a Methyltransferase (HMT) Effector Domain. Preferred examples include those in the Table 5 below, namely NUE, vSET, EHMT2/G9A, SUV39H1, dim-5, KYP, SUVR4, SET4, SET1, SETD8, and TgSET8. NUE is exemplified in the present Examples and, although preferred, it is envisaged that others in the class will also be useful.

[00154] Table 5. Histone Methyltransferase (HMT) Effector Domains

Subtype/ Complex	Name	Substrate (if known)	Modification (if known)	Organism	Full size (aa)	Selected truncation (aa)	Final size (aa)	Catalytic domain
SET	**NUE**	H2B, H3, H4	-	*C. trachomatis*	219	1-219 (Pennini)	219	-
SET	**vSET**	-	H3K27me3	*P. bursaria chlorella virus*	119	1-119 (Mujtaba)	119	4-112: SET2
SUV39 family	**EHMT 2/G9A**	H1.4K2, H3K9, H3K27	H3K9me1/2, H1K25me1	*M. musculus*	1263	969-1263 (Tachibana)	295	1025-1233: preSET, SET, postSET
SUV39	**SUV39 H1**	-	H3K9me2/3	*H. sapiens*	412	79-412 (Snowden)	334	172-412: preSET, SET, postSET
Suvar3-9	**dim-5**	-	H3K9me3	*N. crassa*	331	1-331 (Rathert)	331	77-331: preSET, SET, postSET
Suvar3-9 (SUVH subfamily)	**KYP**	-	H3K9me1/2	*A. thaliana*	624	335-601	267 (Jackson)	-
Suvar3-9 (SUVR subfamily)	**SUVR4**	H3K9me1	H3K9me2/3	*A. thaliana*	492	180-492	313 (Thorstensen)	192-462: preSET, SET, postSET
Suvar4-20	**SET4**	-	H4K20me3	*C. elegans*	288	1-288 (Vielle)	288	-
SET8	**SET1**	-	H4K20me1	*C. elegans*	242	1-242 (Vielle)	242	-
SET8	**SETD8**	-	H4K20me1	*H. sapiens*	393	185-393	209 (Couture)	256-382: SET
SET8	**TgSET 8**	-	H4K20me1/ 2/3	*T. gondii*	1893	1590-1893 (Sautel)	304	1749-1884: SET

[00155] In some embodiment, the functional domain may be a Histone Methyltransferase (HMT) Recruiter Effector Domain. Preferred examples include those in the Table 4 below, namely Hp1a, PHF19, and NIPP1.

[00156] Table 6. Histone Methyltransferase (HMT) Recruiter Effector Domains

Subtype/ Complex	Name	Substrate (if known)	Modification (if known)	Organism	Full size (aa)	Selected truncation (aa)	Final size (aa)	Catalytic domain
-	**Hp1a**	-	H3K9me3	*M. musculus*	191	73-191	119 (Hathaway)	121-179: chromoshadow
-	**PHF19**	-	H3K27me3	*H. sapiens*	580	(1-250) + GGSG linker + (500-580)	335 (Ballaré)	163-250: PHD2
-	**NIPP1**	-	H3K27me3	*H. sapiens*	351	1-329 (Jin)	329	310-329: EED

[00157] In some embodiment, the functional domain may be Histone Acetyltransferase Inhibitor Effector Domain. Preferred examples include SET/TAF-1β listed in the Table 5 below.

[00158] Table 7. Histone Acetyltransferase Inhibitor Effector Domains

Subtype/ Complex	Name	Substrate (if known)	Modification (if known)	Organism	Full size (aa)	Selected truncation (aa)	Final size (aa)	Catalytic domain
-	**SET/TAF-1β**	-	-	*M. musculus*	289	1-289 (Cervoni)	289	-

[00159] It is also preferred to target endogenous (regulatory) control elements (such as enhancers and silencers) in addition to a promoter or promoter-proximal elements. Thus, the invention can also be used to target endogenous control elements (including enhancers and

silencers) in addition to targeting of the promoter. These control elements can be located upstream and downstream of the transcriptional start site (TSS), starting from 200bp from the TSS to 100kb away. Targeting of known control elements can be used to activate or repress the gene of interest. In some cases, a single control element can influence the transcription of multiple target genes. Targeting of a single control element could therefore be used to control the transcription of multiple genes simultaneously.

[00160] Targeting of putative control elements on the other hand (e.g. by tiling the region of the putative control element as well as 200bp up to 100kB around the element) can be used as a means to verify such elements (by measuring the transcription of the gene of interest) or to detect novel control elements (e.g. by tiling 100kb upstream and downstream of the TSS of the gene of interest). In addition, targeting of putative control elements can be useful in the context of understanding genetic causes of disease. Many mutations and common SNP variants associated with disease phenotypes are located outside coding regions. Targeting of such regions with either the activation or repression systems described herein can be followed by readout of transcription of either a) a set of putative targets (e.g. a set of genes located in closest proximity to the control element) or b) whole-transcriptome readout by e.g. RNAseq or microarray. This would allow for the identification of likely candidate genes involved in the disease phenotype. Such candidate genes could be useful as novel drug targets.

[00161] Histone acetyltransferase (HAT) inhibitors are mentioned herein. However, an alternative in some embodiments is for the one or more functional domains to comprise an acetyltransferase, preferably a histone acetyltransferase. These are useful in the field of epigenomics, for example in methods of interrogating the epigenome. Methods of interrogating the epigenome may include, for example, targeting epigenomic sequences. Targeting epigenomic sequences may include the guide being directed to an epigenomic target sequence. Epigenomic target sequence may include, in some embodiments, include a promoter, silencer or an enhancer sequence.

[00162] Examples of acetyltransferases are known but may include, in some embodiments, histone acetyltransferases. In some embodiments, the histone acetyltransferase may comprise the catalytic core of the human acetyltransferase p300 (Gerbasch & Reddy, Nature Biotech 6th April 2015).

Linkers

[00163] IThe term "linker" as used in reference to a fusion protein refers to a molecule which joins the proteins to form a fusion protein. Generally, such molecules have no specific

biological activity other than to join or to preserve some minimum distance or other spatial relationship between the proteins. However, in certain embodiments, the linker may be selected to influence some property of the linker and/or the fusion protein such as the folding, net charge, or hydrophobicity of the linker.

[00164] Suitable linkers for use in the methods of the present invention are well known to those of skill in the art and include, but are not limited to, straight or branched-chain carbon linkers, heterocyclic carbon linkers, or peptide linkers. However, as used herein the linker may also be a covalent bond (carbon-carbon bond or carbon-heteroatom bond). In particular embodiments, the linker is used to separate the Cas protein and the nucleotide deaminase by a distance sufficient to ensure that each protein retains its required functional property. Preferred peptide linker sequences adopt a flexible extended conformation and do not exhibit a propensity for developing an ordered secondary structure. In certain embodiments, the linker can be a chemical moiety which can be monomeric, dimeric, multimeric or polymeric. Preferably, the linker comprises amino acids. Typical amino acids in flexible linkers include Gly, Asn and Ser. Accordingly, in particular embodiments, the linker comprises a combination of one or more of Gly, Asn and Ser amino acids. Other near neutral amino acids, such as Thr and Ala, also may be used in the linker sequence. Exemplary linkers are disclosed in Maratea et al. (1985), Gene 40: 39-46; Murphy et al. (1986) Proc. Nat'l. Acad. Sci. USA 83: 8258-62; U.S. Pat. No. 4,935,233; and U.S. Pat. No. 4,751,180. For example, GlySer linkers GGS, GGGS (SEQ ID NO:413) or GSG can be used. GGS, GSG, GGGS or GGGGS (SEQ ID NO:414) linkers can be used in repeats of 3 (such as $(GGS)_3$ (SEQ ID NO:415), $(GGGGS)_3$ (SEQ ID NO:416)) or 5, 6, 7, 9 or even 12 or more, to provide suitable lengths. In some cases, the linker may be $(GGGGS)_{3\text{-}15}$, For example, in some cases, the linker may be $(GGGGS)_{3\text{-}11}$, e.g., GGGGS, $(GGGGS)_2$ (SEQ ID NO:417), $(GGGGS)_3$, $(GGGGS)_4$ (SEQ ID NO:418), $(GGGGS)_5$ (SEQ ID NO:419), $(GGGGS)_6$ (SEQ ID NO:420), $(GGGGS)_7$ (SEQ ID NO:421), $(GGGGS)_8$ (SEQ ID NO:422), $(GGGGS)_9$ (SEQ ID NO:423), $(GGGGS)_{10}$ (SEQ ID NO:424), or $(GGGGS)_{11}$ (SEQ ID NO:425).

[00165] In particular embodiments, linkers such as $(GGGGS)_3$ are preferably used herein. $(GGGGS)_6$ $(GGGGS)_9$ or $(GGGGS)_{12}$ (SEQ ID NO:426) may preferably be used as alternatives. Other preferred alternatives are $(GGGGS)_1$, $(GGGGS)_2$, $(GGGGS)_4$, $(GGGGS)_5$, $(GGGGS)_7$, $(GGGGS)_8$, $(GGGGS)_{10}$, or $(GGGGS)_{11}$. In yet a further embodiment, LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO:427) is used as a linker. In yet an additional embodiment, the linker is an XTEN linker. In particular embodiments, the Cas protein is linked to the deaminase protein or its catalytic domain by means of an

LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO:428) linker. In further particular embodiments, the Cas protein is linked C-terminally to the N-terminus of a deaminase protein or its catalytic domain by means of an LEPGEKPYKCPECGKSFSQSGALTRHQRTHTR (SEQ ID NO:429) linker. In addition, N- and C-terminal NLSs can also function as linker (e.g., PKKKRKVEASSPKKRKVEAS (SEQ ID NO:430)). Examples of suitable linkers are shown in Table 8.

Table 8. Examples of suitable linkers as disclosed herein.

GGS	GGTGGTAGT (SEQ ID NO:431)
GGSx3 (9)	GGTGGTAGTGGAGGGAGCGGCGGTTCA (SEQ ID NO:432)
GGSx7 (21)	ggtggaggaggctctggtggaggcggtagcggaggcggagggtcgGGTGGTAGTGGAGGG AGCGGCGGTTCA (SEQ ID NO:433)
XTEN	TCGGGATCTGAGACGCCTGGGACCTCGGAATCGGCTACGCCCGAA AGT (SEQ ID NO:434)
Z-EGFR_Short	Gtggataacaaatttaacaaagaaatgtgggcggcgtgggaagaaattcgtaacctgccgaacctgaacggc tggcagatgaccgcgtttattgcgagcctggtggatgatccgagccagagcgcgaacctgctggcggaagcg aaaaaactgaacgatgcgcaggcgccgaaaaccggcggtggttctggt (SEQ ID NO:435)
GSAT	Ggtggttctgccggtggctccggttctggctccagcggtggcagctctggtgcgtccggcacgggtactgcg ggtggcactggcagcggttccggtactggctctggc (SEQ ID NO:436)

[00166] Linkers may be used between the guide RNAs and the functional domain (activator or repressor), or between the Cas protein and the functional domain. The linkers may be used to engineer appropriate amounts of "mechanical flexibility".

[00167] In certain embodiments, the one or more functional domains are controllable, i.e. inducible.

SPLIT PROTEINS

[00168] It is noted that in this context, and more generally for the various applications as described herein, the use of a split version of the Cas protein can be envisaged. Indeed, this may not only allow increased specificity but may also be advantageous for delivery. The Cas is split in the sense that the two parts of the Cas enzyme substantially comprise a functioning Cas. The split may be so that the catalytic domain(s) are unaffected. That Cas may function as a nuclease or it may be a dead-Cas which is essentially an RNA-binding protein with very little or no catalytic activity, due to typically mutation(s) in its catalytic domains.

[00169] Each half of the split Cas may be fused to a dimerization partner. By means of example, and without limitation, employing rapamycin sensitive dimerization domains, allows to generate a chemically inducible split Cas for temporal control of Cas activity. Cas can thus be rendered chemically inducible by being split into two fragments and that rapamycin-sensitive dimerization domains may be used for controlled reassembly of the Cas. The two

parts of the split Cas can be thought of as the N' terminal part and the C' terminal part of the split Cas. The fusion is typically at the split point of the Cas. In other words, the C' terminal of the N' terminal part of the split Cas is fused to one of the dimer halves, whilst the N' terminal of the C' terminal part is fused to the other dimer half.

[00170] The Cas does not have to be split in the sense that the break is newly created. The split point is typically designed in silico and cloned into the constructs. Together, the two parts of the split Cas, the N' terminal and C' terminal parts, form a full Cas, comprising preferably at least 70% or more of the wildtype amino acids (or nucleotides encoding them), preferably at least 80% or more, preferably at least 90% or more, preferably at least 95% or more, and most preferably at least 99% or more of the wildtype amino acids (or nucleotides encoding them). Some trimming may be possible, and mutants are envisaged. Non-functional domains may be removed entirely. What is important is that the two parts may be brought together and that the desired Cas function is restored or reconstituted. The dimer may be a homodimer or a heterodimer.

[00171] The effector protein can moreover be fused to another functional RNase domain, such as a non-specific RNase or Argonaute 2, which acts in synergy to increase the RNase activity or to ensure further degradation of the message.

MODULATING CAS PROTEINS

[00172] The invention provides accessory proteins that modulate CRISPR protein function. In certain embodiments, the accessory protein modulates catalytic activity of a CRISPR protein. In an embodiment of the invention an accessory protein modulates targeted, or sequence specific, nuclease activity. In an embodiment of the invention, an accessory protein modulates collateral nuclease activity. In an embodiment of the invention, an accessory protein modulates binding to a target nucleic acid.

[00173] According to the invention, the nuclease activity to be modulated can be directed against nucleic acids comprising or consisting of RNA, including without limitation mRNA, miRNA, siRNA and nucleic acids comprising cleavable RNA linkages along with nucleotide analogs. In an embodiment of the invention, the nuclease activity to be modulated can be directed against nucleic acids comprising or consisting of DNA, including without limitation nucleic acids comprising cleavable DNA linkages and nucleic acid analogs.

[00174] In an embodiment of the invention, an accessory protein enhances an activity of a CRISPR protein. In certain embodiments, the accessory protein inhibits an activity of a CRISPR protein.

[00175] According to the invention, naturally occurring accessory proteins of Type IV CRISPR systems comprise small proteins encoded at or near a CRISPR locus that function to modify an activity of a CRISPR protein. In general, a CRISPR locus can be identified as comprising a putative CRISPR array and/or encoding a putative CRISPR effector protein. In an embodiment, an effector protein can be from 800 to 2000 amino acids, or from 900 to 1800 amino acids, or from 950 to 1300 amino acids. In an embodiment, an accessory protein can be encoded within 25 kb, or within 20 kb or within 15 kb, or within 10 kb of a putative CRISPR effector protein or array, or from 2 kb to 10 kb from a putative CRISPR effector protein or array.

[00176] In an embodiment of the invention, an accessory protein is from 50 to 300 amino acids, or from 100 to 300 amino acids or from 150 to 250 amino acids or about 200 amino acids.

[00177] Identification and use of a CRISPR accessory protein of the invention is independent of CRISPR effector protein classification. Accessory proteins of the invention can be found in association with or engineered to function with a variety of CRISPR effector proteins. Examples of accessory proteins identified and used herein are representative of CRISPR effector proteins generally. It is understood that CRISPR effector protein classification may involve homology, feature location, nucleic acid target (e.g. DNA or RNA), absence or presence of tracr RNA, location of guide / spacer sequence 5' or 3' of a direct repeat, or other criteria. In embodiments of the invention, accessory protein identification and use transcend such classifications.

[00178] According to the invention, in certain embodiments, enhancing activity of a Type IV or Type I Cas protein or complex thereof comprises contacting the Type IV or Type I Cas protein or complex thereof with an accessory protein from the same organism that activates the Cas protein. In other embodiments, enhancing activity of a Type IV or Type I Cas protein of complex thereof comprises contacting the Type IV or Type I Cas protein or complex thereof with an activator accessory protein from a different organism within the same subclass (e.g., Type IV or Type I). In other embodiments, enhancing activity of a Type IV or Type I Cas protein or complex thereof comprises contacting the Type IV or Type I Cas protein or complex thereof with an accessory protein not within the subclass.

[00179] According to the invention, in certain embodiments, repressing activity of a Type IV or Type I Cas protein or complex thereof comprises contacting the Type IV or Type I Cas protein or complex thereof with an accessory protein from the same organism that represses the Cas protein. In other embodiments, repressing activity of a Type IV or Type I Cas protein

or complex thereof comprises contacting the Type IV or Type I Cas protein or complex thereof with a repressor accessory protein from a different organism within the same subclass. In other embodiments, repressing activity of a Type IV or Type I Cas protein or complex thereof comprises contacting the Type IV or Type I Cas protein or complex thereof with a repressor accessory protein not within the subclass.

[00180] In certain embodiments where the Type IV or Type I Cas protein and the Type IV or Type I accessory protein are from the same organism, the two proteins will function together in an engineered CRISPR system. In certain embodiments, it will be desirable to alter the function of the engineered CRISPR system, for example by modifying either or both of the proteins or their expression. In embodiments where the Type IV or Type I Cas protein and the Type IV or Type I accessory protein are from different organisms which may be within the same class or different classes, the proteins may function together in an engineered CRISPR system but it will often be desired or necessary to modify either or both of the proteins to function together.

[00181] Accordingly, in certain embodiments of the invention either or both of a Cas protein and an accessory protein may be modified to adjust aspects of protein-protein interactions between the Cas protein and accessory protein. In certain embodiments, either or both of a Cas protein and an accessory protein may be modified to adjust aspects of protein-nucleic acid interactions. Ways to adjust protein-protein interactions and protein-nucleic acid interaction include without limitation, fitting molecular surfaces, polar interactions, hydrogen bonds, and modulating van der Waals interactions. In certain embodiments, adjusting protein-protein interactions or protein-nucleic acid binding comprises increasing or decreasing binding interactions. In certain embodiments, adjusting protein-protein interactions or protein-nucleic acid binding comprises modifications that favor or disfavor a conformation of the protein or nucleic acid.

[00182] By “fitting”, is meant determining including by automatic, or semi-automatic means, interactions between one or more atoms of a Cas protein (and optionally at least one atoms of a Cas accessory protein), or between one or more atoms of a Cas protein and one or more atoms of a nucleic acid, (or optionally between one or more atoms of a Cas accessory protein and a nucleic acid),and calculating the extent to which such interactions are stable. Interactions include attraction and repulsion, brought about by charge, steric considerations and the like.

[00183] The three-dimensional structure of Type IV CRISPR protein or complex thereof provides in the context of the instant invention an additional tool for identifying additional

mutations in orthologs of Cas. Further, the three-dimensional structure of Type I CRISPR protein or complex thereof provides in the context of the instant invention an additional tool for identifying additional mutations in orthologs of Cas. The crystal structure can also be basis for the design of new and specific Cass (and optionally Cas accessory proteins). Various computer-based methods for fitting are described further. Binding interactions of Cas (and optionally accessory proteins), and nucleic acids can be examined through the use of computer modeling using a docking program. Docking programs are known; for example, GRAM, DOCK or AUTODOCK (see Walters et al. Drug Discovery Today, vol. 3, no. 4 (1998), 160-178, and Dunbrack et al. Folding and Design 2 (1997), 27-42). This procedure can include computer fitting to ascertain how well the shape and the chemical structure of the binding partners. Computer-assisted, manual examination of the active site or binding site of a Type IV or Type I Cas protein system may be performed. Programs such as GRID (P. Goodford, J. Med. Chem, 1985, 28, 849-57)—a program that determines probable interaction sites between molecules with various functional groups—may also be used to analyze the active site or binding site to predict partial structures of binding compounds. Computer programs can be employed to estimate the attraction, repulsion or steric hindrance of the two binding partners, e.g., components of a Type IV or Type I CRISPR system, or a nucleic acid molecule and a component of a Type IV or Type I CRISPR system.

[00184] Amino acid substitutions may be made on the basis of differences or similarities in amino acid properties (such as polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues) and it is therefore useful to group amino acids together in functional groups. Amino acids may be grouped together based on the properties of their side chains alone. In comparing orthologs, there are likely to be residues conserved for structural or catalytic reasons. These sets may be described in the form of a Venn diagram (Livingstone C.D. and Barton G.J. (1993) "Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation" Comput. Appl. Biosci. 9: 745-756) (Taylor W.R. (1986) "The classification of amino acid conservation" J. Theor. Biol. 119; 205-218). Conservative substitutions may be made, for example according to the Table 9 below which describes a generally accepted Venn diagram grouping of amino acids.

[00185] Table 9. Generally accepted Venn diagram grouping of amino acids.

Set		Sub-set	
Hydrophobic	F W Y H K M I L V A G C	Aromatic	F W Y H
		Aliphatic	I L V
Polar	W Y H K R E D C S T N Q	Charged	H K R E D
		Positively charged	H K R
		Negatively charged	E D
Small	V C A G S P T N D	Tiny	A G S

[00186] In some embodiments, the modifications in Cas may comprise modification of one or more amino acid residues of the Cas protein. In some embodiments, the modifications in Cas may comprise modification of one or more amino acid residues located in a region which comprises residues which are positively charged in the unmodified Cas protein (and/or Cas accessory protein). In some embodiments, the modifications in Cas may comprise modification of one or more amino acid residues which are positively charged in the unmodified Cas protein (and/or Cas accessory protein). In some embodiments, the modifications in Cas may comprise modification of one or more amino acid residues which are not positively charged in the unmodified Cas protein (and/or Cas accessory protein). The modification may comprise modification of one or more amino acid residues which are uncharged in the unmodified Cas protein (and/or Cas accessory protein). The modification may comprise modification of one or more amino acid residues which are negatively charged in the unmodified Cas protein (and/or Cas accessory protein). The modification may comprise modification of one or more amino acid residues which are hydrophobic in the unmodified Cas protein (and/or Cas accessory protein). The modification may comprise modification of one or more amino acid residues which are polar in the unmodified Cas protein (and/or Cas accessory protein). The modification may comprise substitution of a hydrophobic amino acid or polar amino acid with a charged amino acid, which can be a negatively charged or positively charged amino acid. The modification may comprise substitution of a negatively charged amino acid with a positively charged or polar or hydrophobic amino acid. The modification may comprise substitution of a positively charged amino acid with a negatively charged or polar or hydrophobic amino acid.

[00187] Embodiments herein also include sequences (both polynucleotide or polypeptide) which may comprise homologous substitution (substitution and replacement are both used

herein to mean the interchange of an existing amino acid residue or nucleotide, with an alternative residue or nucleotide) that may occur i.e., like-for-like substitution in the case of amino acids such as basic for basic, acidic for acidic, polar for polar, etc. Non-homologous substitution may also occur i.e., from one class of residue to another or alternatively involving the inclusion of unnatural amino acids such as ornithine (hereinafter referred to as Z), diaminobutyric acid ornithine (hereinafter referred to as B), norleucine ornithine (hereinafter referred to as O), pyriylalanine, thienylalanine, naphthylalanine and phenylglycine. Variant amino acid sequences may include suitable spacer groups that may be inserted between any two amino acid residues of the sequence including alkyl groups such as methyl, ethyl or propyl groups in addition to amino acid spacers such as glycine or β-alanine residues. A further form of variation, which involves the presence of one or more amino acid residues in peptoid form, may be well understood by those skilled in the art. For the avoidance of doubt, "the peptoid form" is used to refer to variant amino acid residues wherein the α-carbon substituent group is on the residue's nitrogen atom rather than the α-carbon. Processes for preparing peptides in the peptoid form are known in the art, for example Simon RJ et al., PNAS (1992) 89(20), 9367-9371 and Horwell DC, Trends Biotechnol. (1995) 13(4), 132-134.

[00188] Homology modelling: Corresponding residues in other Cas orthologs can be identified by the methods of Zhang et al., 2012 (Nature; 490(7421): 556-60) and Chen et al., 2015 (PLoS Comput Biol; 11(5): e1004248)—a computational protein-protein interaction (PPI) method to predict interactions mediated by domain-motif interfaces. PrePPI (Predicting PPI), a structure based PPI prediction method, combines structural evidence with non-structural evidence using a Bayesian statistical framework. The method involves taking a pair a query proteins and using structural alignment to identify structural representatives that correspond to either their experimentally determined structures or homology models. Structural alignment is further used to identify both close and remote structural neighbors by considering global and local geometric relationships. Whenever two neighbors of the structural representatives form a complex reported in the Protein Data Bank, this defines a template for modelling the interaction between the two query proteins. Models of a complex are created by superimposing the representative structures on their corresponding structural neighbor in the template. This approach is in Dey et al., 2013 (Prot Sci; 22: 359-66).

[00189] In another aspect, the disclosure provides a method of altering activity of a Cas protein, comprising: identifying one or more candidate amino acids in the Cas protein based on a three-dimensional structure of at least a portion of the Cas protein, wherein the one or

more candidate amino acids interact with a guide RNA that forms a complex with the Cas protein, or are in an inter-domain linker domain, or a bridge helix domain of the Cas protein; and mutating the one or more candidate amino acids thereby generating a mutated Cas protein, wherein activity the mutated Cas protein is different than the Cas protein.

CRISPR-CAS SYSTEM PROMOTES NON-HOMOLOGOUS END-JOINING

[00190] In certain embodiments, nuclease-induced non-homologous end-joining (NHEJ) can be used to target gene-specific knockouts. Nuclease-induced NHEJ can also be used to remove (e.g., delete) sequence in a gene of interest. Generally, NHEJ repairs a double-strand break in the DNA by joining together the two ends; however, generally, the original sequence is restored only if two compatible ends, exactly as they were formed by the double-strand break, are perfectly ligated. The DNA ends of the double-strand break are frequently the subject of enzymatic processing, resulting in the addition or removal of nucleotides, at one or both strands, prior to rejoining of the ends. This results in the presence of insertion and/or deletion (indel) mutations in the DNA sequence at the site of the NHEJ repair. Two-thirds of these mutations typically alter the reading frame and, therefore, produce a non-functional protein. Additionally, mutations that maintain the reading frame, but which insert or delete a significant amount of sequence, can destroy functionality of the protein. This is locus dependent as mutations in critical functional domains are likely less tolerable than mutations in non-critical regions of the protein. The indel mutations generated by NHEJ are unpredictable in nature; however, at a given break site certain indel sequences are favored and are overrepresented in the population, likely due to small regions of microhomology. The lengths of deletions can vary widely; most commonly in the 1-50 bp range, but they can easily be greater than 50 bp, e.g., they can easily reach greater than about 100-200 bp. Insertions tend to be shorter and often include short duplications of the sequence immediately surrounding the break site. However, it is possible to obtain large insertions, and in these cases, the inserted sequence has often been traced to other regions of the genome or to plasmid DNA present in the cells.

[00191] Because NHEJ is a mutagenic process, it may also be used to delete small sequence motifs as long as the generation of a specific final sequence is not required. If a double-strand break is targeted near to a short target sequence, the deletion mutations caused by the NHEJ repair often span, and therefore remove, the unwanted nucleotides. For the deletion of larger DNA segments, introducing two double-strand breaks, one on each side of the sequence, can result in NHEJ between the ends with removal of the entire intervening sequence. Both of these approaches can be used to delete specific DNA sequences; however, the error-prone nature of NHEJ may still produce indel mutations at the site of repair.

[00192] Both double strand cleaving Cas molecules and single strand, or nickase, Cas molecules can be used in the methods and compositions described herein to generate NHEJ-mediated indels. NHEJ-mediated indels targeted to the gene, e.g., a coding region, e.g., an early coding region of a gene of interest can be used to knockout (i.e., eliminate expression of) a gene of interest. For example, early coding region of a gene of interest includes sequence immediately following a transcription start site, within a first exon of the coding sequence, or within 500 bp of the transcription start site (e.g., less than 500, 450, 400, 350, 300, 250, 200, 150, 100 or 50 bp).

[00193] In an embodiment, in which a guide RNA and Cas nuclease generate a double strand break for the purpose of inducing NHEJ-mediated indels, a guide RNA may be configured to position one double-strand break in close proximity to a nucleotide of the target position. In an embodiment, the cleavage site may be between 0-500 bp away from the target position (e.g., less than 500, 400, 300, 200, 100, 50, 40, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 bp from the target position).

[00194] In an embodiment, in which two guide RNAs complexing with Cas nickases induce two single strand breaks for the purpose of inducing NHEJ-mediated indels, two guide RNAs may be configured to position two single-strand breaks to provide for NHEJ repair a nucleotide of the target position.

GUIDE SEQUENCES

[00195] The systems and compositions herein may further comprise one or more guide sequences. The guide sequences may hybridize or be capable of hybridizing with a target sequence. In embodiments of the invention the terms guide sequence and guide RNA and crRNA are used interchangeably as in foregoing cited documents such as WO 2014/093622 (PCT/US2013/074667). 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, CA), 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. Preferably the guide sequence is 10 - 30 nucleotides long, such as 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. 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.

[00196] 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 IV 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.

[00197] In certain embodiments, the CRISPR system as provided herein can make use of a crRNA or analogous polynucleotide comprising a guide sequence, wherein the polynucleotide is an RNA, a DNA or a mixture of RNA and DNA, and/or wherein the polynucleotide comprises one or more nucleotide analogs. The sequence can comprise any structure, including but not limited to a structure of a native crRNA, such as a bulge, a hairpin or a stem loop structure. In certain embodiments, the polynucleotide comprising the guide sequence forms a duplex with a second polynucleotide sequence which can be an RNA or a DNA sequence.

[00198] In certain embodiments, guides of the invention comprise non-naturally occurring nucleic acids and/or non-naturally occurring nucleotides and/or nucleotide analogs, and/or

chemically modifications. Non-naturally occurring nucleic acids can include, for example, mixtures of naturally and non-naturally occurring nucleotides. Non-naturally occurring nucleotides and/or nucleotide analogs may be modified at the ribose, phosphate, and/or base moiety. In an embodiment of the invention, a guide nucleic acid comprises ribonucleotides and non-ribonucleotides. In one such embodiment, a guide comprises one or more ribonucleotides and one or more deoxyribonucleotides. In an embodiment of the invention, the guide comprises one or more non-naturally occurring nucleotide or nucleotide analog such as a nucleotide with phosphorothioate linkage, boranophosphate linkage, a locked nucleic acid (LNA) nucleotides comprising a methylene bridge between the 2′ and 4′ carbons of the ribose ring, or bridged nucleic acids (BNA). Other examples of modified nucleotides include 2'-O-methyl analogs, 2'-deoxy analogs, 2-thiouridine analogs, N6-methyladenosine analogs, or 2'-fluoro analogs. Further examples of modified bases include, but are not limited to, 2-aminopurine, 5-bromo-uridine, pseudouridine (Ψ), N1-methylpseudouridine (me1Ψ), 5-methoxyuridine(5moU), inosine, 7-methylguanosine. Examples of guide RNA chemical modifications include, without limitation, incorporation of 2′-O-methyl (M), 2′-O-methyl 3′phosphorothioate (MS), S-constrained ethyl (cEt), or 2′-O-methyl 3′thioPACE (MSP) at one or more terminal nucleotides. Such chemically modified guide RNAs can comprise increased stability and increased activity as compared to unmodified guide RNAs, though on-target vs. off-target specificity is not predictable. (See, Hendel, 2015, Nat Biotechnol. 33(9):985-9, doi: 10.1038/nbt.3290, published online 29 June 2015; Allerson et al., J. Med. Chem. 2005, 48:901-904; Bramsen et al., Front. Genet., 2012, 3:154; Deng et al., PNAS, 2015, 112:11870-11875; Sharma et al., MedChemComm., 2014, 5:1454-1471; Li et al., Nature Biomedical Engineering, 2017, 1, 0066 DOI:10.1038/s41551-017-0066).

[00199] In certain embodiments, the gRNA (crRNA) binding of the Cas protein of the invention is altered or modified. It is to be understood that mutated Cas has an altered or modified gRNA binding if the gRNA binding is different than the gRNA binding of the corresponding wild type Cas (i.e. unmutated Cas).gRNA binding can be determined by means known in the art. By means of example, and without limitation, gRNA binding can be determined by calculating binding strength or affinity (such as based on equilibrium constants, Ka, Kd, etc.). In certain embodiments, gRNA binding is increased. In certain embodiments, gRNA binding is increased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100%. In certain embodiments, gRNA binding is decreased. In certain embodiments, gRNA binding is decreased by at least 5%, preferably at least 10%, more

preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or (substantially) 100%.

[00200] In some embodiments, the 5' and/or 3' end of a guide RNA is modified by a variety of functional moieties including fluorescent dyes, polyethylene glycol, cholesterol, proteins, or detection tags. (See Kelly et al., 2016, J. Biotech. 233:74-83). In certain embodiments, a guide comprises ribonucleotides in a region that binds to a target DNA and one or more deoxyribonucleotides and/or nucleotide analogs in a region that binds to Cas. In an embodiment of the invention, deoxyribonucleotides and/or nucleotide analogs are incorporated in engineered guide structures, such as, without limitation, 5' and/or 3' end, stem-loop regions, and the seed region. In certain embodiments, the modification is not in the 5'-handle of the stem-loop regions. Chemical modification in the 5'-handle of the stem-loop region of a guide may abolish its function (see Li, et al., Nature Biomedical Engineering, 2017, 1:0066). In certain embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides of a guide is chemically modified. In some embodiments, 3-5 nucleotides at either the 3' or the 5' end of a guide is chemically modified. In some embodiments, only minor modifications are introduced in the seed region, such as 2'-F modifications. In some embodiments, 2'-F modification is introduced at the 3' end of a guide. In certain embodiments, three to five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-O-methyl (M), 2'-O-methyl-3'-phosphorothioate (MS), S-constrained ethyl(cEt), or 2'-O-methyl-3'-thioPACE (MSP). Such modification can enhance genome editing efficiency (see Hendel et al., Nat. Biotechnol. (2015) 33(9): 985-989). In certain embodiments, all of the phosphodiester bonds of a guide are substituted with phosphorothioates (PS) for enhancing levels of gene disruption. In certain embodiments, more than five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-O-Me, 2'-F or S-constrained ethyl(cEt). Such chemically modified guide can mediate enhanced levels of gene disruption (see Ragdarm et al., 0215, PNAS, E7110-E7111). In an embodiment of the invention, a guide is modified to comprise a chemical moiety at its 3' and/or 5' end. Such moieties include, but are not limited to amine, azide, alkyne, thio, dibenzocyclooctyne (DBCO), or Rhodamine. In certain embodiment, the chemical moiety is conjugated to the guide by a linker, such as an alkyl chain. In certain embodiments, the chemical moiety of the modified guide can be used to attach the guide to another molecule, such as DNA, RNA, protein, or nanoparticles. Such chemically modified guide can be used to identify or enrich cells generically edited by a CRISPR system (see Lee et al., eLife, 2017, 6:e25312, DOI:10.7554)

[00201] In some embodiments, the modification to the guide is a chemical modification, an insertion, a deletion or a split. In some embodiments, the chemical modification includes, but is not limited to, incorporation of 2'-O-methyl (M) analogs, 2'-deoxy analogs, 2-thiouridine analogs, N6-methyladenosine analogs, 2'-fluoro analogs, 2-aminopurine, 5-bromo-uridine, pseudouridine (Ψ), N1-methylpseudouridine (me1Ψ), 5-methoxyuridine(5moU), inosine, 7-methylguanosine, 2'-O-methyl-3'-phosphorothioate (MS), S-constrained ethyl(cEt), phosphorothioate (PS), or 2'-O-methyl-3'-thioPACE (MSP). In some embodiments, the guide comprises one or more of phosphorothioate modifications. In certain embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 25 nucleotides of the guide are chemically modified. In certain embodiments, one or more nucleotides in the seed region are chemically modified. In certain embodiments, one or more nucleotides in the 3'-terminus are chemically modified. In certain embodiments, none of the nucleotides in the 5'-handle is chemically modified. In some embodiments, the chemical modification in the seed region is a minor modification, such as incorporation of a 2'-fluoro analog. In a specific embodiment, one nucleotide of the seed region is replaced with a 2'-fluoro analog. In some embodiments, 5 or 10 nucleotides in the 3'-terminus are chemically modified. Such chemical modifications at the 3'-terminus of the Cpf1 CrRNA improve gene cutting efficiency (see Li, et al., Nature Biomedical Engineering, 2017, 1:0066). In a specific embodiment, 5 nucleotides in the 3'-terminus are replaced with 2'-fluoro analogues. In a specific embodiment, 10 nucleotides in the 3'-terminus are replaced with 2'-fluoro analogues. In a specific embodiment, 5 nucleotides in the 3'-terminus are replaced with 2'- O-methyl (M) analogs.

[00202] In some embodiments, the loop of the 5'-handle of the guide is modified. In some embodiments, the loop of the 5'-handle of the guide is modified to have a deletion, an insertion, a split, or chemical modifications. In certain embodiments, the loop comprises 3, 4, or 5 nucleotides. In certain embodiments, the loop comprises the sequence of UCUU, UUUU, UAUU, or UGUU.

[00203] In one aspect, the guide comprises portions that are chemically linked or conjugated via a non-phosphodiester bond. In one aspect, the guide comprises, in non-limiting examples, direct repeat sequence portion and a targeting sequence portion that are chemically linked or conjugated via a non-nucleotide loop. In some embodiments, the portions are joined via a non-phosphodiester covalent linker. Examples of the covalent linker include but are not limited to a chemical moiety selected from the group consisting of carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, fulfones, sulfoxides, ureas,

thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[00204] In some embodiments, portions of the guide are first synthesized using the standard phosphoramidite synthetic protocol (Herdewijn, P., ed., Methods in Molecular Biology Col 288, Oligonucleotide Synthesis: Methods and Applications, Humana Press, New Jersey (2012)). In some embodiments, the non-targeting guide portions can be functionalized to contain an appropriate functional group for ligation using the standard protocol known in the art (Hermanson, G. T., Bioconjugate Techniques, Academic Press (2013)). Examples of functional groups include, but are not limited to, hydroxyl, amine, carboxylic acid, carboxylic acid halide, carboxylic acid active ester, aldehyde, carbonyl, chlorocarbonyl, imidazolylcarbonyl, hydrozide, semicarbazide, thio semicarbazide, thiol, maleimide, haloalkyl, sufonyl, ally, propargyl, diene, alkyne, and azide. Once a non-targeting portions of a guide is functionalized, a covalent chemical bond or linkage can be formed between the two oligonucleotides. Examples of chemical bonds include, but are not limited to, those based on carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, fulfones, sulfoxides, ureas, thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[00205] In some embodiments, one or more portions of a guide can be chemically synthesized. In some embodiments, the chemical synthesis uses automated, solid-phase oligonucleotide synthesis machines with 2'-acetoxyethyl orthoester (2'-ACE) (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., J. Am. Chem. Soc. (2011) 133: 11540-11546; Hendel et al., Nat. Biotechnol. (2015) 33:985-989).

[00206] In some embodiments, the guide portions can be covalently linked using various bioconjugation reactions, loops, bridges, and non-nucleotide links via modifications of sugar, internucleotide phosphodiester bonds, purine and pyrimidine residues. Sletten et al., Angew. Chem. Int. Ed. (2009) 48:6974-6998; Manoharan, M. Curr. Opin. Chem. Biol. (2004) 8: 570-9; Behlke et al., Oligonucleotides (2008) 18: 305-19; Watts, et al., Drug. Discov. Today (2008) 13: 842-55; Shukla, et al., ChemMedChem (2010) 5: 328-49.

[00207] In some embodiments, the guide portions can be covalently linked using click chemistry. In some embodiments, guide portions can be covalently linked using a triazole linker. In some embodiments, guide portions can be covalently linked using Huisgen 1,3-

dipolar cycloaddition reaction involving an alkyne and azide to yield a highly stable triazole linker (He et al., ChemBioChem (2015) 17: 1809-1812; WO 2016/186745). In some embodiments, guide portions are covalently linked by ligating a 5'-hexyne portion and a 3'-azide portion. In some embodiments, either or both of the 5'-hexyne guide portion and a 3'-azide guide portion can be protected with 2'-acetoxyethl orthoester (2'-ACE) group, which can be subsequently removed using Dharmacon protocol (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18).

[00208] In some embodiments, guide portions can be covalently linked via a linker (e.g., a non-nucleotide loop) that comprises a moiety such as spacers, attachments, bioconjugates, chromophores, reporter groups, dye labeled RNAs, and non-naturally occurring nucleotide analogues. More specifically, suitable spacers for purposes of this invention include, but are not limited to, polyethers (e.g., polyethylene glycols, polyalcohols, polypropylene glycol or mixtures of efhylene and propylene glycols), polyamines group (e.g., spennine, spermidine and polymeric derivatives thereof), polyesters (e.g., poly(ethyl acrylate)), polyphosphodiesters, alkylenes, and combinations thereof. Suitable attachments include any moiety that can be added to the linker to add additional properties to the linker, such as but not limited to, fluorescent labels. Suitable bioconjugates include, but are not limited to, peptides, glycosides, lipids, cholesterol, phospholipids, diacyl glycerols and dialkyl glycerols, fatty acids, hydrocarbons, enzyme substrates, steroids, biotin, digoxigenin, carbohydrates, polysaccharides. Suitable chromophores, reporter groups, and dye-labeled RNAs include, but are not limited to, fluorescent dyes such as fluorescein and rhodamine, chemiluminescent, electrochemiluminescent, and bioluminescent marker compounds. The design of example linkers conjugating two RNA components are also described in WO 2004/015075.

[00209] The linker (e.g., a non-nucleotide loop) can be of any length. In some embodiments, the linker has a length equivalent to about 0-16 nucleotides. In some embodiments, the linker has a length equivalent to about 0-8 nucleotides. In some embodiments, the linker has a length equivalent to about 0-4 nucleotides. In some embodiments, the linker has a length equivalent to about 2 nucleotides. Example linker design is also described in WO2011/008730.

[00210] 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. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT,

Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a guide RNA or crRNA) 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 CRISPR-Cas 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. A guide sequence, and hence a guide RNA or crRNA 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 preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[00211] In some embodiments, a guide RNA or crRNA is selected to reduce the degree secondary structure within the guide RNA or crRNA. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the guide RNA participates 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).

[00212] In some embodiments, a nucleic acid-targeting guide is designed or selected to modulate intermolecular interactions among guide molecules, such as among stem-loop regions of different guide molecules. It will be appreciated that nucleotides within a guide that base-pair to form a stem-loop are also capable of base-pairing to form an intermolecular duplex with a second guide and that such an intermolecular duplex would not have a secondary structure compatible with CRISPR complex formation. Accordingly, is useful to select or design DR sequences in order to modulate stem-loop formation and CRISPR complex formation. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of nucleic acid-targeting guides are in intermolecular duplexes. It will be appreciated that stem-loop variation will often be within limits imposed by DR-CRISPR effector interactions. One way to modulate stem-loop formation or change the equilibrium between stem-loop and intermolecular duplex is to vary nucleotide pairs in the stem of the stem-loop of a DR. For example, in one embodiment, a G-C pair is replaced by an A-U or U-A pair. In another embodiment, an A-U pair is substituted for a G-C or a C-G pair. In another embodiment, a naturally occurring nucleotide is replaced by a nucleotide analog. Another way to modulate stem-loop formation or change the equilibrium between stem-loop and intermolecular duplex is to modify the loop of the stem-loop of a DR. Without be bound by theory, the loop can be viewed as an intervening sequence flanked by two sequences that are complementary to each other. When that intervening sequence is not self-complementary, its effect will be to destabilize intermolecular duplex formation. The same principle applies when guides are multiplexed: while the targeting sequences may differ, it may be advantageous to modify the stem-loop region in the DRs of the different guides. Moreover, when guides are multiplexed, the relative activities of the different guides can be modulated by balancing the activity of each individual guide. In certain embodiments, the equilibrium between intermolecular stem-loops vs. intermolecular duplexes is determined. The determination may be made by physical or biochemical means and can be in the presence or absence of a CRISPR effector.

[00213] 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. In other embodiments, multiple DRs (such as dual DRs) may be present.

[00214] In certain embodiments, the crRNA comprises a stem loop, preferably a single stem loop. In certain embodiments, the direct repeat sequence forms a stem loop, preferably a single stem loop.

[00215] In certain embodiments, the spacer length of the guide RNA is from 15 to 35 nt. In certain embodiments, the spacer length of the guide RNA is at least 15 nucleotides. In certain embodiments, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27-30 nt, e.g., 27, 28, 29, or 30 nt, from 30-35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer.

[00216] 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 sca 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 sca sequence or tracr sequence. In some embodiments, the degree of complementarity between the tracr sequence and sca 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 certain embodiments, the tracrRNA may not be required. Indeed, the CRISPR-Cas protein from *Bergeyella zoohelcum* and orthologs thereof do not require a tracrRNA to ensure cleavage of an RNA target.

[00217] In further detail, the assay is as follows for a RNA target, provided that a PAM sequence is required to direct recognition. Two E. coli strains are used in this assay. One carries a plasmid that encodes the endogenous effector protein locus from the bacterial strain. The other strain carries an empty plasmid (e.g., pACYC184, control strain). All possible 7 or 8 bp PAM sequences are presented on an antibiotic resistance plasmid (pUC19 with ampicillin resistance gene). The PAM is located next to the sequence of proto-spacer 1 (the RNA target to the first spacer in the endogenous effector protein locus). Two PAM libraries were cloned.

One has a 8 random bp 5' of the proto-spacer (e.g., total of 65536 different PAM sequences = complexity). The other library has 7 random bp 3' of the proto-spacer (e.g., total complexity is 16384 different PAMs). Both libraries were cloned to have in average 500 plasmids per possible PAM. Test strain and control strain were transformed with 5'PAM and 3'PAM library in separate transformations and transformed cells were plated separately on ampicillin plates. Recognition and subsequent cutting/interference with the plasmid renders a cell vulnerable to ampicillin and prevents growth. Approximately 12h after transformation, all colonies formed by the test and control strains where harvested and plasmid RNA was isolated. Plasmid RNA was used as template for PCR amplification and subsequent deep sequencing. Representation of all PAMs in the untransformed libraries showed the expected representation of PAMs in transformed cells. Representation of all PAMs found in control strains showed the actual representation. Representation of all PAMs in test strain showed which PAMs are not recognized by the enzyme and comparison to the control strain allows extracting the sequence of the depleted PAM. In particular embodiments, the cleavage, such as the RNA cleavage is not PAM dependent.

[00218] For minimization of toxicity and off-target effect, it will be important to control the concentration of guide RNA delivered. Optimal concentrations of nucleic acid –targeting 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. The concentration that gives the highest level of on-target modification while minimizing the level of off-target modification should be chosen for *in vivo* delivery. The system is derived advantageously from a CRISPR-Cas system. Dead guide sequences

[00219] In one aspect, the invention provides guide sequences which are modified in a manner which allows for formation of the CRISPR Cas complex and successful binding to the target, while at the same time, not either allowing for or not allowing for successful nuclease activity (i.e., without nuclease activity / without indel activity). For matters of explanation such modified guide sequences are referred to as "dead guides" or "dead guide sequences". These dead guides or dead guide sequences can be thought of as catalytically inactive or conformationally inactive with regard to nuclease activity. Indeed, dead guide sequences may not sufficiently engage in productive base pairing with respect to the ability to promote catalytic activity or to distinguish on-target and off-target binding activity. Briefly, the assay involves synthesizing a CRISPR target RNA and guide RNAs comprising mismatches with the target RNA, combining these with the enzyme and analyzing cleavage based on gels based on

the presence of bands generated by cleavage products, and quantifying cleavage based upon relative band intensities.

[00220] Hence, in a related aspect, the invention provides a non-naturally occurring or engineered composition CRISPR-Cas system comprising a functional enzyme as described herein, and guide RNA (gRNA) or crRNA wherein the gRNA or crRNA comprises a dead guide sequence whereby the gRNA is capable of hybridizing to a target sequence such that the CRISPR-Cas system is directed to a genomic locus of interest in a cell without detectable RNA cleavage activity of a non-mutant enzyme of the system. It is to be understood that any of the gRNAs or crRNAs according to the invention as described herein elsewhere may be used as dead gRNAs / crRNAs comprising a dead guide sequence.

[00221] The ability of a dead guide sequence to direct sequence-specific binding of a CRISPR complex to an RNA target sequence may be assessed by any suitable assay. For example, the components of a CRISPR-Cas system sufficient to form a CRISPR-Cas complex, including the dead 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 system, followed by an assessment of preferential cleavage within the target sequence.

[00222] As explained further herein, several structural parameters allow for a proper framework to arrive at such dead guides. Dead guide sequences can be typically shorter than respective guide sequences which result in active RNA cleavage. In particular embodiments, dead guides are 5%, 10%, 20%, 30%, 40%, 50%, shorter than respective guides directed to the same.

[00223] As explained below and known in the art, one aspect of gRNA or crRNA – specificity is the direct repeat sequence, which is to be appropriately linked to such guides. In particular, this implies that the direct repeat sequences are designed dependent on the origin of the enzyme. Structural data available for validated dead guide sequences may be used for designing CRISPR-Cas specific equivalents. Structural similarity between, e.g., the orthologous nuclease domains of two or more CRISPR-Cas proteins may be used to transfer design equivalent dead guides. Thus, the dead guide herein may be appropriately modified in length and sequence to reflect such CRISPR-Cas specific equivalents, allowing for formation of the CRISPR-Cas complex and successful binding to the target RNA, while at the same time, not allowing for successful nuclease activity.

[00224] Dead guides allow one to use gRNA or crRNA as a means for gene targeting, without the consequence of nuclease activity, while at the same time providing directed means for activation or repression. Guide RNA or crRNA comprising a dead guide may be modified

to further include elements in a manner which allow for activation or repression of gene activity, in particular protein adaptors (e.g. aptamers) as described herein elsewhere allowing for functional placement of gene effectors (e.g. activators or repressors of gene activity). One example is the incorporation of aptamers, as explained herein and in the state of the art. By engineering the gRNA or crRNA comprising a dead guide to incorporate protein-interacting aptamers (Konermann et al., "Genome-scale transcription activation by an engineered CRISPR-Cas9 complex," doi:10.1038/nature14136, incorporated herein by reference), one may assemble multiple distinct effector domains. Such may be modeled after natural processes.

[00225] The use of two different aptamers (each associated with a distinct nucleic acid-targeting guide RNAs) allows an activator-adaptor protein fusion and a repressor-adaptor protein fusion to be used, with different nucleic acid-targeting guide RNAs or crRNAs, to activate expression of RNA, whilst repressing another. They, along with their different guide RNAs or crRNAs can be administered together, or substantially together, in a multiplexed approach. A large number of such modified RNA-targeting guide RNAs or crRNAs can be used all at the same time, for example 10 or 20 or 30 and so forth, whilst only one (or at least a minimal number) of Cas protein molecules need to be delivered, as a comparatively small number of effector protein molecules can be used with a large number of modified guides. The adaptor protein may be associated (preferably linked or fused to) one or more activators or one or more repressors. For example, the adaptor protein may be associated with a first activator and a second activator. The first and second activators may be the same, but they are preferably different activators. Three or more or even four or more activators (or repressors) may be used, but package size may limit the number being higher than 5 different functional domains. Linkers are preferably used, over a direct fusion to the adaptor protein, where two or more functional domains are associated with the adaptor protein. Suitable linkers might include the GlySer linker.

Mismatches

[00226] In some embodiments, at least one guide polynucleotide comprises a mismatch. The mismatch may be up- or downstream of a single nucleotide variation on the one or more guide sequences. In certain embodiments, modulations of cleavage efficiency can be exploited by introduction of mismatches, e.g. 1 or more mismatches, such as 1 or 2 mismatches between spacer sequence and target sequence, including the position of the mismatch along the spacer/target. The more central (i.e. not 3' or 5') for instance a double mismatch is, the more cleavage efficiency is affected. Accordingly, by choosing mismatch position along the spacer,

cleavage efficiency can be modulated. By means of example, if less than 100 % cleavage of targets is desired (e.g. in a cell population), 1 or more, such as preferably 2 mismatches between spacer and target sequence may be introduced in the spacer sequences. The more central along the spacer of the mismatch position, the lower the cleavage percentage. In certain example embodiments, the cleavage efficiency may be exploited to design single guides that can distinguish two or more targets that vary by a single nucleotide, such as a single nucleotide polymorphism (SNP), variation, or (point) mutation. The CRISPR effector may have reduced sensitivity to SNPs (or other single nucleotide variations) and continue to cleave SNP targets with a certain level of efficiency. Thus, for two targets, or a set of targets, a guide RNA may be designed with a nucleotide sequence that is complementary to one of the targets i.e. the on-target SNP. The guide RNA is further designed to have a synthetic mismatch. As used herein a "synthetic mismatch" refers to a non-naturally occurring mismatch that is introduced upstream or downstream of the naturally occurring SNP, such as at most 5 nucleotides upstream or downstream, for instance 4, 3, 2, or 1 nucleotide upstream or downstream, preferably at most 3 nucleotides upstream or downstream, more preferably at most 2 nucleotides upstream or downstream, most preferably 1 nucleotide upstream or downstream (i.e. adjacent the SNP). When the CRISPR effector binds to the on-target SNP, only a single mismatch will be formed with the synthetic mismatch and the CRISPR effector will continue to be activated and a detectable signal produced. When the guide RNA hybridizes to an off-target SNP, two mismatches will be formed, the mismatch from the SNP and the synthetic mismatch, and no detectable signal generated. Thus, the systems disclosed herein may be designed to distinguish SNPs within a population. For, example the systems may be used to distinguish pathogenic strains that differ by a single SNP or detect certain disease specific SNPs, such as but not limited to, disease associated SNPs, such as without limitation cancer associated SNPs.

[00227] In certain embodiments, the guide RNA is designed such that the mismatch (e.g. The synthetic mismatch, e.g., an additional mutation besides a SNP) is located on position 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, or 30 of the spacer sequence (starting at the 5' end). In certain embodiments, the guide RNA is designed such that the mismatch is located on position 1, 2, 3, 4, 5, 6, 7, 8, or 9 of the spacer sequence (starting at the 5' end). In certain embodiments, the guide RNA is designed such that the mismatch is located on position 4, 5, 6, or 7of the spacer sequence (starting at the 5' end. In certain embodiments, the guide RNA is designed such that the mismatch is located on position 5 of the spacer sequence (starting at the 5' end).

[00228] In certain embodiments, the guide RNA is designed such that the mismatch is located 2 nucleotides upstream of the SNP (i.e. one intervening nucleotide). In certain embodiments, the guide RNA is designed such that the mismatch is located 2 nucleotides downstream of the SNP (i.e. one intervening nucleotide). In certain embodiments, the guide RNA is designed such that the mismatch is located on position 5 of the spacer sequence (starting at the 5' end) and the SNP is located on position 3 of the spacer sequence (starting at the 5' end).

[00229] In an aspect the invention provides a system for specific delivery of functional components to the RNA environment. This can be ensured using the CRISPR systems comprising the Cas proteins of the present invention which allow specific targeting of different components to RNA. More particularly such components include activators or repressors, such as activators or repressors of RNA translation, degradation, etc. Applications of this system are described elsewhere herein.

DETERMINATION OF PAM

[00230] In certain embodiments, a protospacer adjacent motif (PAM) or PAM-like motif directs binding of the Cas protein complex as disclosed herein to the target locus of interest. In some embodiments, the PAM may be a 5' PAM (i.e., located upstream of the 5' end of the protospacer). In other embodiments, the PAM may be a 3' PAM (i.e., located downstream of the 5' end of the protospacer). In other embodiments, both a 5' PAM and a 3' PAM are required. In one example, the PAM comprises or is CC. In certain examples, the PAM is or comprises (C/T)CN (e.g., a 5' CCN or TCN). In certain examples, the PAM is or comprises GG.

[00231] In certain embodiments, a PAM or PAM-like motif may not be required for directing binding of the effector protein (e.g. a Cas protein). In certain embodiments, a 5' PAM is D (e.g., A, G, or U). In certain embodiments of the invention, cleavage at repeat sequences may generate crRNAs (e.g. short or long crRNAs) containing a full spacer sequence flanked by a short nucleotide (e.g. 5, 6, 7, 8, 9, or 10 nt or longer if it is a dual repeat) repeat sequence at the 5' end (this may be referred to as a crRNA ''tag'') and the rest of the repeat at the 3'end. In certain embodiments, targeting by the effector proteins described herein may require the lack of homology between the crRNA tag and the target 5' flanking sequence. This requirement may be similar to that described further in Samai et al. "Co-transcriptional DNA and RNA Cleavage during Type IVI CRISPR-Cas Immunity" Cell 161, 1164–1174, May 21, 2015, where the requirement is thought to distinguish between bona fide targets on invading nucleic

acids from the CRISPR array itself, and where the presence of repeat sequences will lead to full homology with the crRNA tag and prevent autoimmunity.

[00232] In certain embodiments, the PFS (or PAM) recognition or specificity of the Cas protein of the invention is altered or modified. It is to be understood that mutated Cas has an altered or modified PFS recognition or specificity if the PFS recognition or specificity is different than the PFS recognition or specificity of the corresponding wild type Cas (i.e. unmutated Cas). PFS recognition or specificity can be determined by means known in the art. By means of example, and without limitation, PFS recognition or specificity can be determined by PFS (PAM) screens. In certain embodiments, at least one different PFS is recognized by the Cas. In certain embodiments, at least one PFS is recognized by the mutated Cas which is not recognized by the corresponding wild type Cas. In certain embodiments, at least one PFS is recognized by the mutated Cas which is not recognized by the corresponding wild type Cas, in addition to the wild type PFS. In certain embodiments, at least one PFS is recognized by the mutated Cas which is not recognized by the corresponding wild type Cas, and the wild type PFS is not anymore recognized. In certain embodiments, the PFS recognized by the mutated Cas is longer than the PFS recognized by the wild type Cas, such as 1, 2, or 3 nucleotides longer. In certain embodiments, the PFS recognized by the mutated Cas is shorter than the PFS recognized by the wild type Cas, such as 1, 2, or 3 nucleotides shorter.

[00233] In one aspect, determining the PFS sequence for suitable guide sequence of the nucleic acid-targeting protein is by comparison of sequences targeted by guides in depleted cells. In one aspect of such method, the method further comprises comparing the guide abundance for the different conditions in different replicate experiments. In one aspect of such method, the control guides are selected in that they are determined to show limited deviation in guide depletion in replicate experiments. In one aspect of such method, the significance of depletion is determined as (a) a depletion which is more than the most depleted control guide; or (b) a depletion which is more than the average depletion plus two times the standard deviation for the control guides. In one aspect of such method, the host cell is a bacterial host cell. In one aspect of such method, the step of co-introducing the plasmids is by electroporation and the host cell is an electro-competent host cell.

[00234] In some examples, determination of PAM can be performed as follows. This experiment closely parallels similar work in *E. coli* for the heterologous expression of StCas9 (Sapranauskas, R. et al. Nucleic Acids Res 39, 9275–9282 (2011)). Applicants introduce a plasmid containing both a PAM and a resistance gene into the heterologous *E. coli*, and then

plate on the corresponding antibiotic. If there is DNA cleavage of the plasmid, Applicants observe no viable colonies.

[00235] In further detail, the assay is as follows for a DNA target. Two *E.coli* strains are used in this assay. One carries a plasmid that encodes the endogenous effector protein locus from the bacterial strain. The other strain carries an empty plasmid (e.g.pACYC184, control strain). All possible 7 or 8 bp PAM sequences are presented on an antibiotic resistance plasmid (pUC19 with ampicillin resistance gene). The PAM is located next to the sequence of proto-spacer 1 (the DNA target to the first spacer in the endogenous effector protein locus). Two PAM libraries were cloned. One has a 8 random bp 5' of the proto-spacer (e.g. total of 65536 different PAM sequences = complexity). The other library has 7 random bp 3' of the proto-spacer (e.g. total complexity is 16384 different PAMs). Both libraries were cloned to have in average 500 plasmids per possible PAM. Test strain and control strain were transformed with 5'PAM and 3'PAM library in separate transformations and transformed cells were plated separately on ampicillin plates. Recognition and subsequent cutting/interference with the plasmid renders a cell vulnerable to ampicillin and prevents growth. Approximately 12h after transformation, all colonies formed by the test and control strains where harvested and plasmid DNA was isolated. Plasmid DNA was used as template for PCR amplification and subsequent deep sequencing. Representation of all PAMs in the untransfomed libraries showed the expected representation of PAMs in transformed cells. Representation of all PAMs found in control strains showed the actual representation. Representation of all PAMs in test strain showed which PAMs are not recognized by the enzyme and comparison to the control strain allows extracting the sequence of the depleted PAM.

NUCLEAR LOCALIZATION SEQUENCES

[00236] 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. In a preferred embodiment of the invention, the Cas protein 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 (SEQ ID NO:437); the NLS from nucleoplasmin (e.g. the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK (SEQ ID NO:438); the c-myc NLS having the amino acid sequence PAAKRVKLD (SEQ ID NO:439) or RQRRNELKRSP (SEQ ID NO:440); the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY (SEQ ID NO:441); the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV (SEQ ID NO:442) of the IBB domain from importin-alpha; the sequences VSRKRPRP (SEQ ID NO:443) and PPKKARED (SEQ ID NO:444) of the myoma T protein; the sequence POPKKKPL (SEQ ID NO:445) of human p53; the sequence SALIKKKKKMAP (SEQ ID NO:446) of mouse c-abl IV; the sequences DRLRR (SEQ ID NO:447) and PKQKKRK (SEQ ID NO:448) of the influenza virus NS1; the sequence RKLKKKIKKL (SEQ ID NO:449) of the Hepatitis virus delta antigen; the sequence REKKKFLKRR (SEQ ID NO:450) of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK (SEQ ID NO:451) of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK (SEQ ID NO:452) of the steroid hormone receptors (human) glucocorticoid. 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 certain embodiments, other localization tags may be fused to the Cas protein, such as without limitation for localizing the Cas to particular sites in a cell, such as organells, such mitochondria, plastids, chloroplast, vesicles, golgi, (nuclear or cellular)

membranes, ribosomes, nucleoluse, ER, cytoskeleton, vacuoles, centrosome, nucleosome, granules, centrioles, etc.

[00237] In certain embodiments of the invention, at least one nuclear localization signal (NLS) is attached to the nucleic acid sequences encoding the Cas proteins. In preferred embodiments at least one or more C-terminal or N-terminal NLSs are attached (and hence nucleic acid molecule(s) coding for the Cas protein can include coding for NLS(s) so that the expressed product has the NLS(s) attached or connected). In a preferred embodiment a C-terminal NLS is attached for optimal expression and nuclear targeting in eukaryotic cells, preferably human cells. The invention also encompasses methods for delivering multiple nucleic acid components, wherein each nucleic acid component is specific for a different target locus of interest thereby modifying multiple target loci of interest. The nucleic acid component of the complex may comprise one or more protein-binding RNA aptamers. The one or more aptamers may be capable of binding a bacteriophage coat protein.

MULTIPLEX TARGETING APPROACH

[00238] The Cas proteins herein can employ more than one RNA guide without losing activity. This may enable the use of the Cas proteins, CRISPR-Cas systems or complexes as defined herein for targeting multiple targets (e.g., DNA targets), genes or gene loci, with a single enzyme, system or complex as defined herein. The guide RNAs may be tandemly arranged, optionally separated by a nucleotide sequence such as a direct repeat as defined herein. The position of the different guide RNAs is the tandem does not influence the activity.

[00239] In any of the described methods the complex may be delivered with multiple guides for multiplexed use. In any of the described methods more than one protein(s) may be used. In some examples, one Cas protein may be delivered with multiple guides, e.g., at least 2, at least 5, at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 120, at least 140, at least 160, at least 180, at least 200, at least 220, at least 240, at least 260, at least 280, at least 300, at least 350, at least 400, or at least 500 guides. In some examples, a system herein may comprise a Cas protein and multiple guides, e.g., at least 2, at least 5, at least 10, at least 15, at least 20, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 120, at least 140, at least 160, at least 180, at least 200, at least 220, at least 240, at least 260, at least 280, at least 300, at least 350, at least 400, or at least 500 guides.

[00240] The Cas enzyme may form part of a CRISPR system or complex, which further comprises tandemly arranged guide RNAs (gRNAs) comprising a series of 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 25, 25, 30, or more than 30 guide sequences, each capable of specifically hybridizing

to a target sequence in a genomic locus of interest in a cell. In some embodiments, the functional Cas CRISPR system or complex binds to the multiple target sequences. In some embodiments, the functional CRISPR system or complex may edit the multiple target sequences, e.g., the target sequences may comprise a genomic locus, and in some embodiments there may be an alteration of gene expression. In some embodiments, the functional CRISPR system or complex may comprise further functional domains. In some embodiments, the invention provides a method for altering or modifying expression of multiple gene products. The method may comprise introducing into a cell containing said target nucleic acids, e.g., DNA molecules, or containing and expressing target nucleic acid, e.g., DNA molecules; for instance, the target nucleic acids may encode gene products or provide for expression of gene products (e.g., regulatory sequences). In some general embodiments, the Cas enzyme used for multiplex targeting is associated with one or more functional domains. In some more specific embodiments, the CRISPR enzyme used for multiplex targeting is a deadCas as defined herein elsewhere. In some embodiments, each of the guide sequence is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length. Examples of multiplex genome engineering using CRISPR effector proteins are provided in Cong et al. (Science Feb 15;339(6121):819-23 (2013) and other publications cited herein.

[00241] In any of the described methods the strand break may be a single strand break or a double strand break. In preferred embodiments the double strand break may refer to the breakage of two sections of RNA, such as the two sections of RNA formed when a single strand RNA molecule has folded onto itself or putative double helices that are formed with an RNA molecule which contains self-complementary sequences allows parts of the RNA to fold and pair with itself.

BASE EDITING

[00242] The present disclosure also provides for a base editing system. In general, such a system may comprise a deaminase (e.g., an adenosine deaminase or cytidine deaminase) fused with a Cas protein (e.g., a Type IV Cas protein herein). The Cas protein may be a dead Cas protein or a Cas nickase protein. In certain examples, the system comprises a mutated form of an adenosine deaminase fused with a dead CRISPR-Cas or CRISPR-Cas nickase. The mutated form of the adenosine deaminase may have both adenosine deaminase and cytidine deaminase activities.

[00243] In one aspect, the present disclosure provides an engineered adenosine deaminase. The engineered adenosine deaminase may comprise one or more mutations herein. In some embodiments, the engineered adenosine deaminase has cytidine deaminase activity. In certain

examples, the engineered adenosine deaminase has both cytidine deaminase activity and adenosine deaminase. In some cases, the modifications by base editors herein may be used for targeting post-translational signaling or catalysis. In some embodiments, compositions herein comprise nucleotide sequence comprising encoding sequences for one or more components of a base editing system. A base-editing system may comprise a deaminase (e.g., an adenosine deaminase or cytidine deaminase) fused with a Cas protein or a variant thereof.

[00244] In certain examples, the system comprises a mutated form of an adenosine deaminase fused with a dead CRISPR-Cas or CRISPR-Cas nickase. The mutated form of the adenosine deaminase may have both adenosine deaminase and cytidine deaminase activities. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, based on amino acid sequence positions of hADAR2-D, and mutations in a homologous ADAR protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, based on amino acid sequence

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

some examples, provided herein includes a mutated adenosine deaminase e.g., an adenosine deaminase comprising E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, and S661T, fused with a dead CRISPR-Cas protein or a CRISPR-Cas nickase. In some examples, provided herein includes a mutated adenosine deaminase e.g., an adenosine deaminase comprising E488Q, V351G, S486A, T375S, S370C, P462A, N597I, L332I, I398V, K350I, M383L, D619G, S582T, V440I, S495N, K418E, S661T, and S375N fused with a dead CRISPR-Cas protein or a CRISPR-Cas nickase.

[00245] In some embodiments, the adenosine deaminase may be a tRNA-specific adenosine deaminase or a variant thereof. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: W23L, W23R, R26G, H36L, N37S, P48S, P48T, P48A, I49V, R51L, N72D, L84F, S97C, A106V, D108N, H123Y, G125A, A142N, S146C, D147Y, R152H, R152P, E155V, I156F, K157N, K161T, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: D108N based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, A142N, based on amino acid sequence positions of E. coli TadA, and mutations in a homologous deaminase protein corresponding to the above. In some embodiments, the adenosine deaminase may comprise one or more of the mutations: A106V, D108N, D147Y, E155V, L84F, H123Y, I156F, H36L, R51L, S146C, K157N, based on amino acid sequence positions of E. coli TadA, and mutations in a

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

[00246] In some examples, the base editing systems may comprise an intein-mediated trans-splicing system that enables in vivo delivery of a base editor, e.g., a split-intein cytidine base editors (CBE) or adenine base editor (ABE) engineered to trans-splice. Examples of the such base editing systems include those described in Colin K.W. Lim et al., Treatment of a Mouse Model of ALS by In Vivo Base Editing, Mol Ther. 2020 Jan 14. pii: S1525-0016(20)30011-3. doi: 10.1016/j.ymthe.2020.01.005; and Jonathan M. Levy et al., Cytosine and adenine base editing of the brain, liver, retina, heart and skeletal muscle of mice via adeno-associated viruses, Nature Biomedical Engineering volume 4, pages97–110(2020), which are incorporated by reference herein in their entireties.

[00247] Examples of base editing systems include those described in WO2019071048 (e.g. paragraphs [0933]-0938]), WO2019084063 (e.g., paragraphs [0173]-[0186], [0323]-[0475],

[0893]-[1094]), WO2019126716 (e.g., paragraphs [0290]-[0425], [1077]-[1084]), WO2019126709 (e.g., paragraphs [0294]-[0453]), WO2019126762 (e.g., paragraphs [0309]-[0438]), WO2019126774 (e.g., paragraphs [0511]-[0670]), Cox DBT, et al., RNA editing with CRISPR-Cas13, Science. 2017 Nov 24;358(6366):1019-1027; Abudayyeh OO, et al., A cytosine deaminase for programmable single-base RNA editing, Science 26 Jul 2019: Vol. 365, Issue 6451, pp. 382-386; Gaudelli NM et al., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage, Nature volume 551, pages 464–471 (23 November 2017); Komor AC, et al., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016 May 19;533(7603):420-4; Jordan L. Doman et al., Evaluation and minimization of Cas9-independent off-target DNA editing by cytosine base editors, Nat Biotechnol (2020). https://doi.org/10.1038/s41587-020-0414-6, which are incorporated by reference herein in their entireties.

PRIME EDITING

[00248] In some embodiments, the Cas protein herein may be used for prime editing. In some cases, the Cas protein may be a nickase, e.g., a DNA nickase. The Cas may be a dCas. In some cases, the Cas has one or more mutations.

[00249] The Cas protein may be associated with a reverse transcriptase. The reverse transcriptase may be fused to the C-terminus of a Cas protein. Alternatively or additionally, the reverse transcriptase may be fused to the N-terminus of a Cas protein. The fusion may be via a linker and/or an adaptor protein. In some examples, the reverse transcriptase may be an M-MLV reverse transcriptase or variant thereof. The M-MLV reverse transcriptase variant may comprise one or more mutations. For the examples, the M-MLV reverse transcriptase may comprise D200N, L603W, and T330P. In another example, the M-MLV reverse transcriptase may comprise D200N, L603W, T330P, T306K, and W313F. In a particular example, the fusion of Cas and reverse transcriptase is Cas (H840A) fused with M-MLV reverse transcriptase (D200N+L603W+T330P+T306K+W313F).

[00250] In some embodiments, the Cas protein herein may target DNA using a guide RNA containing a binding sequence that hybridizes to the target sequence on the DNA. The guide RNA may further comprise an editing sequence that contains new genetic information that replaces target DNA nucleotides.

[00251] A single-strand break (a nick) may be generated on the target DNA by the Cas protein at the target site to expose a 3'-hydroxyl group, thus priming the reverse transcription of an edit-encoding extension on the guide directly into the target site. These steps may result in a branched intermediate with two redundant single-stranded DNA flaps: a 5' flap that

contains the unedited DNA sequence, and a 3' flap that contains the edited sequence copied from the guide RNA. The 5' flaps may be removed by a structure-specific endonuclease, e.g., FEN122, which excises 5' flaps generated during lagging-strand DNA synthesis and long-patch base excision repair. The non-edited DNA strand may be nicked to induce bias DNA repair to preferentially replace the non-edited strand. Examples of prime editing systems and methods include those described in Anzalone AV *et al.*, Search-and-replace genome editing without double-strand breaks or donor DNA, Nature. 2019 Oct 21. doi: 10.1038/s41586-019-1711-4, which is incorporated by reference herein in its entirety.

[00252] The Cas proteins may be used to prime-edit a single nucleotide on a target DNA. Alternatively or additionally, the Cas proteins may be used to prime-edit at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 10, at least 11, at least 12, at least 13, at least 14, at least 15, at least 16, at least 17, at least 18, at least 19, at least 20, at least 21, at least 22, at least 23, at least 24, at least 25, at least 26, at least 27, at least 28, at least 29, at least 30, at least 40, at least 50, at least 60, at least 70, at least 80, at least 90, at least 100, at least 200, at least 300, at least 400, at least 500, at least 600, at least 700, at least 800, at least 900, or at least 10000 nucleotides on a target DNA.

CELLS AND ORGANISMS

[00253] In a further aspect, the invention provides a eukaryotic cell comprising a modified target locus of interest, wherein the target locus of interest has been modified according to in any of the herein described methods. A further aspect provides a cell line of said cell. Another aspect provides a multicellular organism comprising one or more said cells.

[00254] The present disclosure provides cells, tissues, organisms comprising the engineered Cas protein, the CRISPR-Cas systems, the polynucleotides encoding one or more components of the CRISPR-Cas systems, and/or vectors comprising the polynucleotides. The invention also provides for the nucleotide sequence encoding the effector protein being codon optimized for expression in a eukaryote or eukaryotic cell in any of the herein described methods or compositions. In an embodiment of the invention, the codon optimized effector protein is any Cas protein discussed herein and is codon optimized for operability in a eukaryotic cell or organism, e.g., such cell or organism as elsewhere herein mentioned, for instance, without limitation, a yeast cell, or a mammalian cell or organism, including a mouse cell, a rat cell, and a human cell or non-human eukaryote organism, e.g., plant.

[00255] In certain embodiments, the modification of the target locus of interest may result in: the eukaryotic cell comprising altered expression of at least one gene product; the eukaryotic cell comprising altered expression of at least one gene product, wherein the expression of the

at least one gene product is increased; the eukaryotic cell comprising altered expression of at least one gene product, wherein the expression of the at least one gene product is decreased; or the eukaryotic cell comprising an edited genome.

[00256] In certain embodiments, the eukaryotic cell may be a mammalian cell or a human cell.

[00257] In further embodiments, the non-naturally occurring or engineered compositions, the vector systems, or the delivery systems as described in the present specification may be used for: site-specific gene knockout; site-specific genome editing; RNA sequence-specific interference; or multiplexed genome engineering.

[00258] Also provided is a gene product from the cell, the cell line, or the organism as described herein. In certain embodiments, the amount of gene product expressed may be greater than or less than the amount of gene product from a cell that does not have altered expression or edited genome. In certain embodiments, the gene product may be altered in comparison with the gene product from a cell that does not have altered expression or edited genome.

DELIVERY

[00259] The present disclosure also provides delivery systems for introducing components of the systems and compositions herein to cells, tissues, organs, or organisms. A delivery system may comprise one or more delivery vehicles and/or cargos. Exemplary delivery systems and methods include those described in paragraphs [00117] to [00278] of Feng Zhang et al., (WO2016106236A1), and pages 1241-1251 and Table 1 of Lino CA et al., Delivering CRISPR: a review of the challenges and approaches, DRUG DELIVERY, 2018, VOL. 25, NO. 1, 1234–1257, which are incorporated by reference herein in their entireties.

[00260] The methods, systems, and tools provided herein may be designed for use with Class 1 CRISPR proteins or multi-subunit proteins. In an embodiment, the delivery is tailors for a Type I or Type IV Cas proteins detailed herein. Delivery, as further detailed herein, may comprise delivery of one or more subunits or CRISPR associated proteins separately, as one or more fusion proteins, or as polynucleotides encoding the proteins. Because the Class 1 systems typically comprise a multi-protein effector complex, which can, in embodiments, include one or more of ancillary proteins, such as one or more proteins in a complex referred to as a CRISPR-associated complex for antiviral defense (Cascade), one or more adaptation proteins and/or one or more accessory, CRISPR associated Rossman fold (CARF) domain containing proteins, delivery of such Class I systems may be modified to effect delivery, which may be further modified for successful delivery in eukaryotes. Delivery of multimeric Class I

complexes is known in the art. See, e.g. Pickar-Oliver et al., Nat Biotechnol. 2019 Dec; 37(12): 1493–1501; doi: 10.1038/s41587-019-0235-7. Briefly, Pcikar-Oliver utilized a CMV promoter for each subunit of the system and further included N-terminal Flag epitope tags and nuclear localization systems. While Pickar-Olivier delivered each subunit of the complex on a separate vector delivery of more than one subunit on the same construct. Dolan et al. delivered *T. fusca* Type I-E for genome editing in hESCs via RNP electroporation utilizing C-terminal NLSs on Cas3 and to the C-terminus of each of the six Cas7 subunits delivered via electroporation. Dolan et al., Mol Cell. 2019 Jun 6; 74(5): 936–950.e5; doi: 10.1016/j.molcel.2019.03.014; see also Morisaka, et al. Nat. Commun. 10, 5302 (2019); Cameron et al, Nat Biotechnol. 2019 Dec;37(12):1471-147;. doi: 10.1038/s41587-019-0310-0 (fusion of multi-subunit cascade to Fok1 nuclease domain for delivery via polycistronic vector with guide RNA delivered on separate plasmid for eukaryotic application); and Young et al., Commun Biol. 2019 Oct 18;2:383. doi: 10.1038/s42003-019-0637-6 (delivery of class 1 type 1-E S. thermophilus system in *Zea mays* by tethering a plant transcriptional activation domain to 3 different subunits of the Cascade complex). Codon optimization based on human codon usage and/or further codon optimization by optimization tools such as ATUM/DNA2.0 can be performed to further optimize expression.

CARGOS

[00261] The delivery systems may comprise one or more cargos. The cargos may comprise one or more components of the systems and compositions herein. A cargo may comprise one or more of the following: i) a plasmid encoding one or more Cas proteins; ii) a plasmid encoding one or more guide RNAs, iii) mRNA of one or more Cas proteins; iv) one or more guide RNAs; v) one or more Cas proteins; vi) any combination thereof. In some examples, a cargo may comprise a plasmid encoding one or more Cas protein and one or more (e.g., a plurality of) guide RNAs. In some embodiments, a cargo may comprise mRNA encoding one or more Cas proteins and one or more guide RNAs.

[00262] In some examples, a cargo may comprise one or more Cas proteins and one or more guide RNAs, e.g., in the form of ribonucleoprotein complexes (RNP). The ribonucleoprotein complexes may be delivered by methods and systems herein. In some cases, the ribonucleoprotein may be delivered by way of a polypeptide-based shuttle agent. In one example, the ribonucleoprotein may be delivered using synthetic peptides comprising an endosome leakage domain (ELD) operably linked to a cell penetrating domain (CPD), to a histidine-rich domain and a CPD, e.g., as describe in WO2016161516.

PHYSICAL DELIVERY

[00263] In some embodiments, the cargos may be introduced to cells by physical delivery methods. Examples of physical methods include microinjection, electroporation, and hydrodynamic delivery.

Microinjection

[00264] Microinjection of the cargo directly to cells can achieve high efficiency, e.g., above 90% or about 100%. In some embodiments, microinjection may be performed using a microscope and a needle (e.g., with 0.5–5.0 μm in diameter) to pierce a cell membrane and deliver the cargo directly to a target site within the cell. Microinjection may be used for *in vitro* and *ex vivo* delivery.

[00265] Plasmids comprising coding sequences for Cas proteins and/or guide RNAs, mRNAs, and/or guide RNAs, may be microinjected. In some cases, microinjection may be used i) to deliver DNA directly to a cell nucleus, and/or ii) to deliver mRNA (e.g., in vitro transcribed) to a cell nucleus or cytoplasm. In certain examples, microinjection may be used to delivery sgRNA directly to the nucleus and Cas-encoding mRNA to the cytoplasm, e.g., facilitating translation and shuttling of Cas to the nucleus.

[00266] Microinjection may be used to generate genetically modified animals. For example, gene editing cargos may be injected into zygotes to allow for efficient germline modification. Such approach can yield normal embryos and full-term mouse pups harboring the desired modification(s). Microinjection can also be used to provide transiently up- or down- regulate a specific gene within the genome of a cell, e.g., using CRISPRa and CRISPRi.

Electroporation

[00267] In some embodiments, the cargos and/or delivery vehicles may be delivered by electroporation. Electroporation may use pulsed high-voltage electrical currents to transiently open nanometer-sized pores within the cellular membrane of cells suspended in buffer, allowing for components with hydrodynamic diameters of tens of nanometers to flow into the cell. In some cases, electroporation may be used on various cell types and efficiently transfer cargo into cells. Electroporation may be used for *in vitro* and *ex vivo* delivery.

[00268] Electroporation may also be used to deliver the cargo to into the nuclei of mammalian cells by applying specific voltage and reagents, e.g., by nucleofection. Such approaches include those described in Wu Y, et al. (2015). Cell Res 25:67–79; Ye L, et al. (2014). Proc Natl Acad Sci USA 111:9591–6; Choi PS, Meyerson M. (2014). Nat Commun 5:3728; Wang J, Quake SR. (2014). Proc Natl Acad Sci 111:13157–62. Electroporation may

also be used to deliver the cargo in vivo, e.g., with methods described in Zuckermann M, et al. (2015). Nat Commun 6:7391.

Hydrodynamic delivery

[00269] Hydrodynamic delivery may also be used for delivering the cargos, e.g., for *in vivo* delivery. In some examples, hydrodynamic delivery may be performed by rapidly pushing a large volume (8–10% body weight) solution containing the gene editing cargo into the bloodstream of a subject (e.g., an animal or human), e.g., for mice, via the tail vein. As blood is incompressible, the large bolus of liquid may result in an increase in hydrodynamic pressure that temporarily enhances permeability into endothelial and parenchymal cells, allowing for cargo not normally capable of crossing a cellular membrane to pass into cells. This approach may be used for delivering naked DNA plasmids and proteins. The delivered cargos may be enriched in liver, kidney, lung, muscle, and/or heart.

Transfection

[00270] The cargos, e.g., nucleic acids, may be introduced to cells by transfection methods for introducing nucleic acids into cells. Examples of transfection methods include calcium phosphate-mediated transfection, cationic transfection, liposome transfection, dendrimer transfection, heat shock transfection, magnetofection, lipofection, impalefection, optical transfection, proprietary agent-enhanced uptake of nucleic acid.

DELIVERY VEHICLES

[00271] The delivery systems may comprise one or more delivery vehicles. The delivery vehicles may deliver the cargo into cells, tissues, organs, or organisms (e.g., animals or plants). The cargos may be packaged, carried, or otherwise associated with the delivery vehicles. The delivery vehicles may be selected based on the types of cargo to be delivered, and/or the delivery is *in vitro* and/or *in vivo*. Examples of delivery vehicles include vectors, viruses, non-viral vehicles, and other delivery reagents described herein.

[00272] The delivery vehicles in accordance with the present invention may a greatest dimension (e.g. diameter) of less than 100 microns (μm). In some embodiments, the delivery vehicles have a greatest dimension of less than 10 μm. In some embodiments, the delivery vehicles may have a greatest dimension of less than 2000 nanometers (nm). In some embodiments, the delivery vehicles may have a greatest dimension of less than 1000 nanometers (nm). In some embodiments, the delivery vehicles may have a greatest dimension (e.g., diameter) of less than 900 nm, less than 800 nm, less than 700 nm, less than 600 nm, less than 500 nm, less than 400 nm, less than 300 nm, less than 200 nm, less than 150nm, or less

than 100nm, less than 50nm. In some embodiments, the delivery vehicles may have a greatest dimension ranging between 25 nm and 200 nm.

[00273] In some embodiments, the delivery vehicles may be or comprise particles. For example, the delivery vehicle may be or comprise nanoparticles (e.g., particles with a greatest dimension (e.g., diameter) no greater than 1000nm. The particles may be provided in different forms, e.g., as solid particles (e.g., metal such as silver, gold, iron, titanium), non-metal, lipid-based solids, polymers), suspensions of particles, or combinations thereof. Metal, dielectric, and semiconductor particles may be prepared, as well as hybrid structures (e.g., core–shell particles).

Vectors

[00274] The systems, compositions, and/or delivery systems may comprise one or more vectors. The present disclosure also include vector systems. A vector system may comprise one or more vectors. In some embodiments, a vector refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. A vector may be a plasmid, e.g., a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Certain vectors may be capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Some vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. In certain examples, vectors may be expression vectors, e.g., capable of directing the expression of genes to which they are operatively-linked. In some cases, the expression vectors may be for expression in eukaryotic cells. Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[00275] Examples of vectors include pGEX, pMAL, pRIT5, E. coli expression vectors (e.g., pTrc, pET 11d, yeast expression vectors (e.g., pYepSec1, pMFa, pJRY88, pYES2, and picZ, Baculovirus vectors (e.g., for expression in insect cells such as SF9 cells) (e.g., pAc series and the pVL series), mammalian expression vectors (e.g., pCDM8 and pMT2PC.

[00276] A vector may comprise i) Cas encoding sequence(s), and/or ii) a single, or at least 2, at least 3, at least 4, at least 5, at least 6, at least 7, at least 8, at least 9, at least 10, at least

12, at least 14, at least 16, at least 32, at least 48, at least 50 guide RNA(s) encoding sequences. In a single vector there can be a promoter for each RNA coding sequence. Alternatively or additionally, in a single vector, there may be a promoter controlling (e.g., driving transcription and/or expression) multiple RNA encoding sequences.

Regulatory elements

[00277] A vector may comprise one or more regulatory elements. The regulatory element(s) may be operably linked to coding sequences of Cas proteins, accessary proteins, guide RNAs (e.g., a single guide RNA, crRNA, and/or tracrRNA), or combination thereof. The term "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). In certain examples, a vector may comprise: a first regulatory element operably linked to a nucleotide sequence encoding a Cas protein, and a second regulatory element operably linked to a nucleotide sequence encoding a guide RNA.

[00278] Examples of regulatory elements include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g., transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g., liver, pancreas), or particular cell types (e.g., lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific.

[00279] Examples of promoters include one or more pol III promoter (e.g., 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g., 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g., 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus

(CMV) promoter (optionally with the CMV enhancer), the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter.

Viral vectors

[00280] The cargos may be delivered by viruses. In some embodiments, viral vectors are used. A viral vector may comprise virally-derived DNA or RNA sequences for packaging into a virus (e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Viruses and viral vectors may be used for *in vitro*, *ex vivo*, and/or *in vivo* deliveries.

Adeno associated virus (AAV)

[00281] The systems and compositions herein may be delivered by adeno associated virus (AAV). AAV vectors may be used for such delivery. AAV, of the Dependovirus genus and Parvoviridae family, is a single stranded DNA virus. In some embodiments, AAV may provide a persistent source of the provided DNA, as AAV delivered genomic material can exist indefinitely in cells, e.g., either as exogenous DNA or, with some modification, be directly integrated into the host DNA. In some embodiments, AAV do not cause or relate with any diseases in humans. The virus itself is able to efficiently infect cells while provoking little to no innate or adaptive immune response or associated toxicity.

[00282] Examples of AAV that can be used herein include AAV-1, AAV-2, AAV-3, AAV-4, AAV-5, AAV-6, AAV-8, and AAV-9. The type of AAV may be selected with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. AAV-2-based vectors were originally proposed for CFTR delivery to CF airways, other serotypes such as AAV-1, AAV-5, AAV-6, and AAV-9 exhibit improved gene transfer efficiency in a variety of models of the lung epithelium. Examples of cell types targeted by AAV are described in Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)), and shown as follows:

Cell Line	AAV-1	AAV-2	AAV-3	AAV-4	AAV-5	AAV-6	AAV-8	AAV-9
Huh-7	13	100	2.5	0.0	0.1	10	0.7	0.0
HEK293	25	100	2.5	0.1	0.1	5	0.7	0.1
HeLa	3	100	2.0	0.1	6.7	1	0.2	0.1
HepG2	3	100	16.7	0.3	1.7	5	0.3	ND
Hep1A	20	100	0.2	1.0	0.1	1	0.2	0.0

911	17	100	11	0.2	0.1	17	0.1	ND
CHO	100	100	14	1.4	333	50	10	1.0
COS	33	100	33	3.3	5.0	14	2.0	0.5
MeWo	10	100	20	0.3	6.7	10	1.0	0.2
NIH3T3	10	100	2.9	2.9	0.3	10	0.3	ND
A549	14	100	20	ND	0.5	10	0.5	0.1
HT1180	20	100	10	0.1	0.3	33	0.5	0.1
Monocytes	1111	100	ND	ND	125	1429	ND	ND
Immature DC	2500	100	ND	ND	222	2857	ND	ND
Mature DC	2222	100	ND	ND	333	3333	ND	ND

[00283] CRISPR-Cas AAV particles may be created in HEK 293 T cells. Once particles with specific tropism have been created, they are used to infect the target cell line much in the same way that native viral particles do. This may allow for persistent presence of CRISPR-Cas components in the infected cell type, and what makes this version of delivery particularly suited to cases where long-term expression is desirable. Examples of doses and formulations for AAV that can be used include those describe in US Patent Nos. 8,454,972 and 8,404,658.

[00284] Various strategies may be used for delivery the systems and compositions herein with AAVs. In some examples, coding sequences of Cas and gRNA may be packaged directly onto one DNA plasmid vector and delivered via one AAV particle. In some examples, AAVs may be used to deliver gRNAs into cells that have been previously engineered to express Cas. In some examples, coding sequences of Cas and gRNA may be made into two separate AAV particles, which are used for co-transfection of target cells. In some examples, markers, tags, and other sequences may be packaged in the same AAV particles as coding sequences of Cas and/or gRNAs.

Lentiviruses

[00285] The systems and compositions herein may be delivered by lentiviruses. Lentiviral vectors may be used for such delivery. Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells.

[00286] Examples of lentiviruses include human immunodeficiency virus (HIV), which may use its envelope glycoproteins of other viruses to target a broad range of cell types; minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV), which may be used for ocular therapies. In certain embodiments, self-inactivating lentiviral vectors with an siRNA targeting a common exon shared by HIV tat/rev, a nucleolar-localizing TAR decoy, and an anti–CCR5-specific hammerhead ribozyme (see, e.g., DiGiusto et al.

(2010) Sci Transl Med 2:36ra43) may be used/and or adapted to the nucleic acid-targeting system herein.

[00287] Lentiviruses may be pseudo-typed with other viral proteins, such as the G protein of vesicular stomatitis virus. In doing so, the cellular tropism of the lentiviruses can be altered to be as broad or narrow as desired. In some cases, to improve safety, second- and third-generation lentiviral systems may split essential genes across three plasmids, which may reduce the likelihood of accidental reconstitution of viable viral particles within cells.

[00288] In some examples, leveraging the integration ability, lentiviruses may be used to create libraries of cells comprising various genetic modifications, e.g., for screening and/or studying genes and signaling pathways.

Adenoviruses

[00289] The systems and compositions herein may be delivered by adenoviruses. Adenoviral vectors may be used for such delivery. Adenoviruses include nonenveloped viruses with an icosahedral nucleocapsid containing a double stranded DNA genome. Adenoviruses may infect dividing and non-dividing cells. In some embodiments, adenoviruses do not integrate into the genome of host cells, which may be used for limiting off-target effects of CRISPR-Cas systems in gene editing applications.

Non-viral vehicles

[00290] The delivery vehicles may comprise non-viral vehicles. In general, methods and vehicles capable of delivering nucleic acids and/or proteins may be used for delivering the systems compositions herein. Examples of non-viral vehicles include lipid nanoparticles, cell-penetrating peptides (CPPs), DNA nanoclews, gold nanoparticles, streptolysin O, multifunctional envelope-type nanodevices (MENDs), lipid-coated mesoporous silica particles, and other inorganic nanoparticles.

Lipid particles

[00291] The delivery vehicles may comprise lipid particles, e.g., lipid nanoparticles (LNPs) and liposomes.

Lipid nanoparticles (LNPs)

[00292] LNPs may encapsulate nucleic acids within cationic lipid particles (e.g., liposomes), and may be delivered to cells with relative ease. In some examples, lipid nanoparticles do not contain any viral components, which helps minimize safety and immunogenicity concerns. Lipid particles may be used for *in vitro*, *ex vivo*, and *in vivo* deliveries. Lipid particles may be used for various scales of cell populations.

[00293] In some examples. LNPs may be used for delivering DNA molecules (e.g., those comprising coding sequences of Cas and/or gRNA) and/or RNA molecules (e.g., mRNA of Cas, gRNAs). In certain cases, LNPs may be use for delivering RNP complexes of Cas/gRNA.

[00294] Components in LNPs may comprise cationic lipids 1,2- dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N- dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA), (3- o-[2"-(methoxypolyethyleneglycol 2000) succinoyl]-1,2-dimyristoyl-sn-glycol (PEG-S-DMG), R-3-[(ro-methoxy-poly(ethylene glycol)2000) carbamoyl]-1,2-dimyristyloxlpropyl-3-amine (PEG-C-DOMG, and any combination thereof. Preparation of LNPs and encapsulation may be adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011).

Liposomes

[00295] In some embodiments, a lipid particle may be liposome. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. In some embodiments, liposomes are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB).

[00296] Liposomes can be made from several different types of lipids, e.g., phospholipids. A liposome may comprise natural phospholipids and lipids such as 1,2-distearoryl-sn-glycero-3 -phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines, monosialoganglioside, or any combination thereof.

[00297] Several other additives may be added to liposomes in order to modify their structure and properties. For instance, liposomes may further comprise cholesterol, sphingomyelin, and/or 1,2-dioleoyl-sn-glycero-3- phosphoethanolamine (DOPE), e.g., to increase stability and/or to prevent the leakage of the liposomal inner cargo.

Stable nucleic-acid-lipid particles (SNALPs)

[00298] In some embodiments, the lipid particles may be stable nucleic acid lipid particles (SNALPs). SNALPs may comprise an ionizable lipid (DLinDMA) (e.g., cationic at low pH), a neutral helper lipid, cholesterol, a diffusible polyethylene glycol (PEG)-lipid, or any combination thereof. In some examples, SNALPs may comprise synthetic cholesterol, dipalmitoylphosphatidylcholine, 3-N-[(w-methoxy polyethylene glycol)2000)carbamoyl]-1,2-dimyrestyloxypropylamine, and cationic 1,2-dilinoleyloxy-3-N,Ndimethylaminopropane. In some examples, SNALPs may comprise synthetic cholesterol, 1,2-distearoyl-sn-glycero-3-

phosphocholine, PEG- cDMA, and l,2-dilinoleyloxy-3-(N;N-dimethyl)aminopropane (DLinDMA)

Other lipids

[00299] The lipid particles may also comprise one or more other types of lipids, e.g., cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[l,3]- dioxolane (DLin-KC2-DMA), DLin-KC2-DMA4, C12- 200 and colipids disteroylphosphatidyl choline, cholesterol, and PEG-DMG.

Lipoplexes/polyplexes

[00300] In some embodiments, the delivery vehicles comprise lipoplexes and/or polyplexes. Lipoplexes may bind to negatively charged cell membrane and induce endocytosis into the cells. Examples of lipoplexes may be complexes comprising lipid(s) and non-lipid components. Examples of lipoplexes and polyplexes include FuGENE-6 reagent, a non-liposomal solution containing lipids and other components, zwitterionic amino lipids (ZALs), Ca2þ (e.g., forming DNA/Ca^{2+} microcomplexes), polyethenimine (PEI) (e.g., branched PEI), and poly(L-lysine) (PLL).

Cell penetrating peptides

[00301] In some embodiments, the delivery vehicles comprise cell penetrating peptides (CPPs). CPPs are short peptides that facilitate cellular uptake of various molecular cargo (e.g., from nanosized particles to small chemical molecules and large fragments of DNA).

[00302] CPPs may be of different sizes, amino acid sequences, and charges. In some examples, CPPs can translocate the plasma membrane and facilitate the delivery of various molecular cargoes to the cytoplasm or an organelle. CPPs may be introduced into cells via different mechanisms, e.g., direct penetration in the membrane, endocytosis-mediated entry, and translocation through the formation of a transitory structure.

[00303] CPPs may have an amino acid composition that either contains a high relative abundance of positively charged amino acids such as lysine or arginine or has sequences that contain an alternating pattern of polar/charged amino acids and non-polar, hydrophobic amino acids. These two types of structures are referred to as polycationic or amphipathic, respectively. A third class of CPPs are the hydrophobic peptides, containing only apolar residues, with low net charge or have hydrophobic amino acid groups that are crucial for cellular uptake. Another type of CPPs is the trans-activating transcriptional activator (Tat) from Human Immunodeficiency Virus 1 (HIV-1). Examples of CPPs include to Penetratin, Tat (48-60),

Transportan, and (R-AhX-R4) (Ahx refers to aminohexanoyl). Examples of CPPs and related applications also include those described in US Patent 8,372,951.

[00304] CPPs can be used for *in vitro* and *ex vivo* work quite readily, and extensive optimization for each cargo and cell type is usually required. In some examples, CPPs may be covalently attached to the Cas protein directly, which is then complexed with the gRNA and delivered to cells. In some examples, separate delivery of CPP–Cas and CPP–gRNA to multiple cells may be performed. CPP may also be used to delivery RNPs.

DNA nanoclews

[00305] In some embodiments, the delivery vehicles comprise DNA nanoclews. A DNA nanoclew refers to a sphere-like structure of DNA (e.g., with a shape of a ball of yarn). The nanoclew may be synthesized by rolling circle amplification with palindromic sequences that aide in the self-assembly of the structure. The sphere may then be loaded with a payload. An example of DNA nanoclew is described in Sun W et al, J Am Chem Soc. 2014 Oct 22;136(42):14722-5; and Sun W et al, Angew Chem Int Ed Engl. 2015 Oct 5;54(41):12029-33. DNA nanoclew may have a palindromic sequences to be partially complementary to the gRNA within the Cas:gRNA ribonucleoprotein complex. A DNA nanoclew may be coated, e.g., coated with PEI to induce endosomal escape.

Gold nanoparticles

[00306] In some embodiments, the delivery vehicles comprise gold nanoparticles (also referred to AuNPs or colloidal gold). Gold nanoparticles may form complex with cargos, e.g., Cas:gRNA RNP. Gold nanoparticles may be coated, e.g., coated in a silicate and an endosomal disruptive polymer, PAsp(DET). Examples of gold nanoparticles include AuraSense Therapeutics' Spherical Nucleic Acid (SNA™) constructs, and those described in Mout R, et al. (2017). ACS Nano 11:2452–8; Lee K, et al. (2017). Nat Biomed Eng 1:889–901.

iTOP

[00307] In some embodiments, the delivery vehicles comprise iTOP. iTOP refers to a combination of small molecules drives the highly efficient intracellular delivery of native proteins, independent of any transduction peptide. iTOP may be used for induced transduction by osmocytosis and propanebetaine, using NaCl-mediated hyperosmolality together with a transduction compound (propanebetaine) to trigger macropinocytotic uptake into cells of extracellular macromolecules. Examples of iTOP methods and reagents include those described in D'Astolfo DS, Pagliero RJ, Pras A, et al. (2015). Cell 161:674–690.

Polymer-based particles

[00308] In some embodiments, the delivery vehicles may comprise polymer-based particles (e.g., nanoparticles). In some embodiments, the polymer-based particles may mimic a viral mechanism of membrane fusion. The polymer-based particles may be a synthetic copy of Influenza virus machinery and form transfection complexes with various types of nucleic acids ((siRNA, miRNA, plasmid DNA or shRNA, mRNA) that cells take up via the endocytosis pathway, a process that involves the formation of an acidic compartment. The low pH in late endosomes acts as a chemical switch that renders the particle surface hydrophobic and facilitates membrane crossing. Once in the cytosol, the particle releases its payload for cellular action. This Active Endosome Escape technology is safe and maximizes transfection efficiency as it is using a natural uptake pathway. In some embodiments, the polymer-based particles may comprise alkylated and carboxyalkylated branched polyethylenimine. In some examples, the polymer-based particles are VIROMER, e.g., VIROMER RNAi, VIROMER RED, VIROMER mRNA, VIROMER CRISPR. Example methods of delivering the systems and compositions herein include those described in Bawage SS et al., Synthetic mRNA expressed Cas13a mitigates RNA virus infections, www.biorxiv.org/content/10.1101/370460v1.full doi: doi.org/10.1101/370460, Viromer® RED, a powerful tool for transfection of keratinocytes. doi: 10.13140/RG.2.2.16993.61281, Viromer® Transfection - Factbook 2018: technology, product overview, users' data., doi:10.13140/RG.2.2.23912.16642.

Streptolysin O (SLO)

[00309] The delivery vehicles may be streptolysin O (SLO). SLO is a toxin produced by Group A streptococci that works by creating pores in mammalian cell membranes. SLO may act in a reversible manner, which allows for the delivery of proteins (e.g., up to 100 kDa) to the cytosol of cells without compromising overall viability. Examples of SLO include those described in Sierig G, et al. (2003). Infect Immun 71:446–55; Walev I, et al. (2001). Proc Natl Acad Sci U S A 98:3185–90; Teng KW, et al. (2017). Elife 6:e25460.

Multifunctional envelope-type nanodevice (MEND)

[00310] The delivery vehicles may comprise multifunctional envelope-type nanodevice (MENDs). MENDs may comprise condensed plasmid DNA, a PLL core, and a lipid film shell. A MEND may further comprise cell-penetrating peptide (e.g., stearyl octaarginine). The cell penetrating peptide may be in the lipid shell. The lipid envelope may be modified with one or more functional components, e.g., one or more of: polyethylene glycol (e.g., to increase vascular circulation time), ligands for targeting of specific tissues/cells, additional cell-

penetrating peptides (e.g., for greater cellular delivery), lipids to enhance endosomal escape, and nuclear delivery tags. In some examples, the MEND may be a tetra-lamellar MEND (T-MEND), which may target the cellular nucleus and mitochondria. In certain examples, a MEND may be a PEG-peptide-DOPE-conjugated MEND (PPD-MEND), which may target bladder cancer cells. Examples of MENDs include those described in Kogure K, et al. (2004). J Control Release 98:317–23; Nakamura T, et al. (2012). Acc Chem Res 45:1113–21.

Lipid-coated mesoporous silica particles

[00311] The delivery vehicles may comprise lipid-coated mesoporous silica particles. Lipid-coated mesoporous silica particles may comprise a mesoporous silica nanoparticle core and a lipid membrane shell. The silica core may have a large internal surface area, leading to high cargo loading capacities. In some embodiments, pore sizes, pore chemistry, and overall particle sizes may be modified for loading different types of cargos. The lipid coating of the particle may also be modified to maximize cargo loading, increase circulation times, and provide precise targeting and cargo release. Examples of lipid-coated mesoporous silica particles include those described in Du X, et al. (2014). Biomaterials 35:5580–90; Durfee PN, et al. (2016). ACS Nano 10:8325–45.

Inorganic nanoparticles

[00312] The delivery vehicles may comprise inorganic nanoparticles. Examples of inorganic nanoparticles include carbon nanotubes (CNTs) (e.g., as described in Bates K and Kostarelos K. (2013). Adv Drug Deliv Rev 65:2023–33.), bare mesoporous silica nanoparticles (MSNPs) (e.g., as described in Luo GF, et al. (2014). Sci Rep 4:6064), and dense silica nanoparticles (SiNPs) (as described in Luo D and Saltzman WM. (2000). Nat Biotechnol 18:893–5).

METHODS OF USE IN GENERAL

[00313] In another aspect, the present disclosure discloses methods of using the compositions and systems herein. In general, the methods include modifying a target nucleic acid by introducing in a cell or organism that comprises the target nucleic acid the engineered Cas protein, polynucleotide(s) encoding engineered Cas protein, the CRISPR-Cas system, or the vector or vector system comprising the polynucleotide(s), such that the engineered Cas protein modifies the target nucleic acid in the cell or organism.

[00314] In some embodiments, the target nucleic acid comprises a genomic locus, and the engineered Cas protein modifies gene product encoded at the genomic locus or expression of the gene product. The target nucleic acid is DNA or RNA and wherein one or more nucleotides

in the target nucleic acid may be base edited. The target nucleic acid may be DNA or RNA and wherein the target nucleic acid is cleaved. The engineered Cas protein may further cleave non-target nucleic acid.

[00315] In one embodiment, the systems and methods herein may be used for cleaving a target nucleic acid. The methods may comprise modifying a target nucleic acid using a nucleic acid-targeting complex that binds to the target nucleic acid and effect cleavage of said target nucleic acid. In an embodiment, the systems or compositions herein, when introduced into a cell, may create a break (e.g., a single or a double strand break) in the nucleic acid sequence. For example, the systems and methods can be used to cleave a disease nucleic acid in a cell. For example, an exogenous nucleic acid template comprising a sequence to be integrated flanked by an upstream sequence and a downstream sequence may be introduced into a cell. The upstream and downstream sequences share sequence similarity with either side of the site of integration in the nucleic acid. Where desired, a donor nucleic acid can be mRNA. The exogenous nucleic acid template comprises a sequence to be integrated (e.g., a mutated nucleic acid). The sequence for integration may be a sequence endogenous or exogenous to the cell. 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 nucleic acid. A template are selected to promote recombination between the nucleic acid sequence of interest and the donor nucleic acid. The upstream sequence may be a nucleic acid sequence that shares sequence similarity with the nucleic acid sequence upstream of the targeted site for integration. Similarly, the downstream sequence may be a nucleic acid sequence that shares sequence similarity with the nucleic acid sequence downstream of the targeted site of integration. The upstream and downstream sequences in the exogenous nucleic acid template can have 75%, 80%, 85%, 90%, 95%, or 100% sequence identity with the targeted nucleic acid sequence. Preferably, the upstream and downstream sequences in the exogenous nucleic acid template have about 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with the targeted sequence. In some cases, the upstream and downstream sequences in the exogenous nucleic acid template have about 99% or 100% sequence identity with the targeted nucleic acid 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 nucleic acid template may further comprise a marker. Such a marker may make it easy to screen for targeted integrations. Examples of suitable markers include restriction sites, fluorescent proteins, or selectable markers. The exogenous nucleic acid template of the invention can be constructed using recombinant techniques (see, for example, Sambrook et al., 2001 and Ausubel et al., 1996). In a method for modifying a target nucleic acid by integrating an exogenous nucleic acid template, a break (e.g., double or single stranded break in double or single stranded nucleic acid) is introduced into the nucleic acid sequence by the nucleic acid-targeting complex, the break is repaired via homologous recombination with an exogenous nucleic acid template such that the template is integrated into the nucleic acid target. The presence of a double-stranded break facilitates integration of the template. In other embodiments, this invention provides a method of modifying expression of a nucleic acid in a eukaryotic cell. The method comprises increasing or decreasing expression of a target polynucleotide by using a nucleic acid-targeting complex that binds to the DNA or RNA (e.g., mRNA or pre-mRNA).

[00316] In some methods, a target nucleic acid can be inactivated to affect the modification of the expression in a cell. For example, upon the binding of a nucleic acid-targeting complex to a target sequence in a cell, the target nucleic acid is inactivated such that the sequence is not translated, the coded protein is not produced, or the sequence does not function as the wild-type sequence does. For example, a protein or microRNA coding sequence may be inactivated such that the protein or microRNA or pre-microRNA transcript is not produced. The target nucleic acid of a nucleic acid-targeting complex can be any nucleic acid endogenous or exogenous to the eukaryotic cell. For example, the target nucleic acid can be a nucleic acid residing in the nucleus of the eukaryotic cell. The target nucleic acid can be a sequence (e.g., mRNA or pre-mRNA) coding a gene product (e.g., a protein) or a non-coding sequence (e.g., ncRNA, lncRNA, tRNA, or rRNA). Examples of target nucleic acid include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated nucleic acid. Examples of target nucleic acid include a disease associated nucleic acid. A "disease-associated" nucleic acid refers to any nucleic acid which is yielding translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non-disease control. It may be a nucleic acid transcribed from a gene that becomes expressed at an abnormally high level; it may be a RNA transcribed from 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 nucleic acid also refers to a nucleic acid transcribed from a gene possessing

mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The translated products may be known or unknown, and may be at a normal or abnormal level. The target nucleic acid of a nucleic acid-targeting complex can be any nucleic acid endogenous or exogenous to the eukaryotic cell. For example, the target nucleic acid can be a nucleic acid residing in the nucleus of the eukaryotic cell. The target nucleic acid can be a sequence (e.g., mRNA or pre-mRNA) coding a gene product (e.g., a protein) or a non-coding sequence (e.g., ncRNA, lncRNA, tRNA, or rRNA).

[00317] In some embodiments, the methods may further comprise visualizing activity and, optionally, using a detectable label. The method may also comprise detecting binding of one or more components of the CRISPR-Cas system to the target nucleic acid.

[00318] According to one aspect the invention provides non-naturally occurring or engineered composition comprising a guide RNA comprising a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell, wherein the guide RNA is modified by the insertion of one or more distinct RNA sequence(s) that bind an adaptor protein. In particular embodiments, the RNA sequences may bind to two or more adaptor proteins (e.g. aptamers), and wherein each adaptor protein is associated with one or more functional domains. The guide RNAs of the CRISPR-Cas enzymes described herein are shown to be amenable to modification of the guide sequence. In particular embodiments, the guide RNA is modified by the insertion of distinct RNA sequence(s) 5' of the direct repeat, within the direct repeat, or 3' of the guide sequence. When there is more than one functional domain, the functional domains can be same or different, e.g., two of the same or two different activators or repressors. In an aspect the invention provides a herein-discussed composition, wherein the one or more functional domains are attached to the Cas protein so that upon binding to the target RNA the functional domain is in a spatial orientation allowing for the functional domain to function in its attributed function; In an aspect the invention provides a herein-discussed composition, wherein the composition comprises a CRISPR-Cas complex having at least three functional domains, at least one of which is associated with the Cas protein and at least two of which are associated with the gRNA.

[00319] Accordingly, in an aspect the invention provides non-naturally occurring or engineered CRISPR-Cas complex composition comprising the guide RNA as herein-discussed and a CRISPR-Cas which is an Cas protein, wherein optionally the Cas protein comprises at least one mutation, such that the Cas protein has no more than 5% of the nuclease activity of the enzyme not having the at least one mutation, and optionally one or more comprising at least

one or more nuclear localization sequences. In particular embodiments, the guide RNA is additionally or alternatively modified so as to still ensure binding of the Cas protein but to prevent cleavage by the Cas protein (as detailed elsewhere herein).

[00320] In particular embodiments, the Cas protein is a Cas protein which has a diminished nuclease activity of at least 97%, or 100% as compared with the CRISPR-Cas enzyme not having the at least one mutation. In an aspect the invention provides a herein-discussed composition, wherein the CRISPR-Cas enzyme comprises two or more mutations as otherwise herein-discussed.

[00321] In particular embodiments, an system is provided as described herein above comprising two or more functional domains. In particular embodiments, the two or more functional domains are heterologous functional domain. In particular embodiments, the system comprises an adaptor protein which is a fusion protein comprising a functional domain, the fusion protein optionally comprising a linker between the adaptor protein and the functional domain. In particular embodiments, the linker includes a GlySer linker. Additionally or alternatively, one or more functional domains are attached to the RNA effector protein by way of a linker, optionally a GlySer linker.

[00322] In an aspect the invention provides a herein-discussed composition, wherein the one or more functional domains associated with the adaptor protein or the Cas protein is a domain capable of activating or repressing RNA translation. In an aspect the invention provides a herein-discussed composition, wherein at least one of the one or more functional domains associated with the adaptor protein have one or more activities comprising methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, DNA integration activity RNA cleavage activity, DNA cleavage activity or nucleic acid binding activity, or molecular switch activity or chemical inducibility or light inducibility.

[00323] In an aspect the invention provides a herein-discussed composition comprising an aptamer sequence. In particular embodiments, the aptamer sequence is two or more aptamer sequences specific to the same adaptor protein. In an aspect the invention provides a herein-discussed composition, wherein the aptamer sequence is two or more aptamer sequences specific to different adaptor protein. In an aspect the invention provides a herein-discussed composition, wherein the adaptor protein comprises bacteriophage coat proteins. Accordingly, in particular embodiments, the aptamer is selected from a binding protein specifically binding any one of the adaptor proteins listed above. In an aspect the invention provides a herein-discussed composition, wherein the cell is a eukaryotic cell. In an aspect the invention provides

a herein-discussed composition, wherein the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell, whereby the mammalian cell is optionally a mouse cell. In an aspect the invention provides a herein-discussed composition, wherein the mammalian cell is a human cell.

[00324] In an aspect the invention provides a herein above-discussed composition wherein there is more than one guide RNA or gRNA or crRNA, and these target different sequences whereby when the composition is employed, there is multiplexing. In an aspect the invention provides a composition wherein there is more than one guide RNA or gRNA or crRNA modified by the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins.

[00325] In an aspect the invention provides a herein-discussed composition wherein one or more adaptor proteins associated with one or more functional domains is present and bound to the distinct RNA sequence(s) inserted into the guide RNA(s).

[00326] In an aspect the invention provides a herein-discussed composition wherein the guide RNA is modified to have at least one non-coding functional loop; e.g., wherein the at least one non-coding functional loop is repressive; for instance, wherein at least one non-coding functional loop comprises Alu.

[00327] In an aspect the invention provides a method for modifying gene expression comprising the administration to a host or expression in a host *in vivo* of one or more of the compositions as herein discussed.

[00328] In an aspect the invention provides a herein-discussed method comprising the delivery of the composition or nucleic acid molecule(s) coding therefor, wherein said nucleic acid molecule(s) are operatively linked to regulatory sequence(s) and expressed *in vivo*. In an aspect the invention provides a herein-discussed method wherein the expression *in vivo* is via a lentivirus, an adenovirus, or an AAV.

[00329] In an aspect the invention provides a mammalian cell line of cells as herein-discussed, wherein the cell line is, optionally, a human cell line or a mouse cell line. In an aspect the invention provides a transgenic mammalian model, optionally a mouse, wherein the model has been transformed with a herein-discussed composition or is a progeny of said transformant.

[00330] In an aspect the invention provides a nucleic acid molecule(s) encoding guide RNA or the CRISPR-Cas complex or the composition as herein-discussed. In an aspect the invention provides a vector comprising: a nucleic acid molecule encoding a guide RNA (gRNA) or crRNA comprising a guide sequence capable of hybridizing to an RNA target sequence in a

cell, wherein the direct repeat of the gRNA or crRNA is modified by the insertion of distinct RNA sequence(s) that bind(s) to two or more adaptor proteins, and wherein each adaptor protein is associated with one or more functional domains; or, wherein the gRNA is modified to have at least one non-coding functional loop. In an aspect the invention provides vector(s) comprising nucleic acid molecule(s) encoding: non-naturally occurring or engineered CRISPR-Cas complex composition comprising the gRNA or crRNA herein-discussed, and an Cas protein, wherein optionally the Cas protein comprises at least one mutation, such that the Cas protein has no more than 5% of the nuclease activity of the Cas protein not having the at least one mutation, and optionally one or more comprising at least one or more nuclear localization sequences. In an aspect a vector can further comprise regulatory element(s) operable in a eukaryotic cell operably linked to the nucleic acid molecule encoding the guide RNA (gRNA) or crRNA and/or the nucleic acid molecule encoding the Cas protein and/or the optional nuclear localization sequence(s).

[00331] In one aspect, the invention provides a kit comprising one or more of the components described herein. In some embodiments, the kit comprises a vector system as described herein and instructions for using the kit.

[00332] In an aspect the invention provides a method of screening for gain of function (GOF) or loss of function (LOF) or for screening non-coding RNAs or potential regulatory regions (e.g. enhancers, repressors) comprising the cell line of as herein-discussed or cells of the model herein-discussed containing or expressing the Cas protein and introducing a composition as herein-discussed into cells of the cell line or model, whereby the gRNA or crRNA includes either an activator or a repressor, and monitoring for GOF or LOF respectively as to those cells as to which the introduced gRNA or crRNA includes an activator or as to those cells as to which the introduced gRNA or crRNA includes a repressor.

[00333] In an aspect the invention provides a library of non-naturally occurring or engineered compositions, each comprising a CRISPR guide RNA (gRNA) or crRNA comprising a guide sequence capable of hybridizing to a target RNA sequence of interest in a cell, an Cas protein, wherein the Cas protein comprises at least one mutation, such that the Cas protein has no more than 5% of the nuclease activity of the Cas protein not having the at least one mutation, wherein the gRNA or crRNA is modified by the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins, and wherein the adaptor protein is associated with one or more functional domains, wherein the composition comprises one or more or two or more adaptor proteins, wherein the each protein is associated with one or more functional domains, and wherein the gRNAs or crRNAs comprise a genome wide library

comprising a plurality of guide RNAs (gRNAs) or crRNAs. In an aspect the invention provides a library as herein-discussed, wherein the Cas protein has a diminished nuclease activity of at least 97%, or 100% as compare with the Cas protein not having the at least one mutation. In an aspect the invention provides a library as herein-discussed, wherein the adaptor protein is a fusion protein comprising the functional domain. In an aspect the invention provides a library as herein discussed, wherein the gRNA or crRNA is not modified by the insertion of distinct RNA sequence(s) that bind to the one or two or more adaptor proteins. In an aspect the invention provides a library as herein discussed, wherein the one or two or more functional domains are associated with the Cas protein. In an aspect the invention provides a library as herein discussed, wherein the cell population of cells is a population of eukaryotic cells. In an aspect the invention provides a library as herein discussed, wherein the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell. In an aspect the invention provides a library as herein discussed, wherein the mammalian cell is a human cell. In an aspect the invention provides a library as herein discussed, wherein the population of cells is a population of embryonic stem (ES) cells.

[00334] In an aspect the invention provides a library as herein discussed, wherein the targeting is of about 100 or more RNA sequences. In an aspect the invention provides a library as herein discussed, wherein the targeting is of about 1000 or more RNA sequences. In an aspect the invention provides a library as herein discussed, wherein the targeting is of about 20,000 or more sequences. In an aspect the invention provides a library as herein discussed, wherein the targeting is of the entire transcriptome. In an aspect the invention provides a library as herein discussed, wherein the targeting is of a panel of target sequences focused on a relevant or desirable pathway. In an aspect the invention provides a library as herein discussed, wherein the pathway is an immune pathway. In an aspect the invention provides a library as herein discussed, wherein the pathway is a cell division pathway.

[00335] In one aspect, the invention provides a method of generating a model eukaryotic cell comprising a gene with modified expression. In some embodiments, a disease gene is any gene associated an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) introducing one or more vectors encoding the components of the system described herein above into a eukaryotic cell, and (b) allowing a CRISPR complex to bind to a target polynucleotide so as to modify expression of a gene, thereby generating a model eukaryotic cell comprising modified gene expression.

[00336] The structural information provided herein allows for interrogation of guide RNA or crRNA interaction with the target RNA and the Cas protein permitting engineering or

alteration of guide RNA structure to optimize functionality of the entire CRISPR-Cas system. For example, the guide RNA or crRNA may be extended, without colliding with the Cas protein by the insertion of adaptor proteins that can bind to RNA. These adaptor proteins can further recruit effector proteins or fusions which comprise one or more functional domains.

[00337] An aspect of the invention is that the above elements are comprised in a single composition or comprised in individual compositions. These compositions may advantageously be applied to a host to elicit a functional effect on the genomic level.

[00338] The skilled person will understand that modifications to the guide RNA or crRNA which allow for binding of the adapter + functional domain but not proper positioning of the adapter + functional domain (e.g. due to steric hindrance within the three dimension structure of the CRISPR-Cas complex) are modifications which are not intended. The one or more modified guide RNA or crRNA may be modified, by introduction of a distinct RNA sequence(s) 5' of the direct repeat, within the direct repeat, or 3' of the guide sequence.

[00339] The modified guide RNA or crRNA, the inactivated Cas protein (with or without functional domains), and the binding protein with one or more functional domains, may each individually be comprised in a composition and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or described herein for delivery to a host (e.g. lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g. for lentiviral gRNA or crRNA selection) and concentration of gRNA or crRNA (e.g. dependent on whether multiple gRNAs or crRNAs are used) may be advantageous for eliciting an improved effect.

[00340] Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic events. The compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling *in vivo* (e.g. gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling; the use the compositions of the invention to establish cell lines and transgenic animals for optimization and screening purposes).

[00341] The current invention comprehends the use of the compositions of the current invention to establish and utilize conditional or inducible CRISPR-Cas events. (See, e.g., Platt et al., Cell (2014), dx.doi.org/10.1016/j.cell.2014.09.014, or PCT patent publications cited herein, such as WO 2014/093622 (PCT/US2013/074667), which are not believed prior to the present invention or application).

[00342] The invention provides a method of modifying expression of a target gene of interest, the method comprising contacting a target RNA with one or more non-naturally occurring or engineered compositions comprising i) a mutated Cas protein according to the invention as described herein, and ii) a crRNA, wherein the crRNA comprises a) a guide sequence that hybridizes to a target RNA sequence in a cell, and b) a direct repeat sequence, wherein the Cas protein forms a complex with the crRNA, wherein the guide sequence directs sequence-specific binding to the target RNA sequence in a cell, whereby there is formed a CRISPR complex comprising the Cas protein complexed with the guide sequence that is hybridized to the target RNA sequence, whereby expression of the target locus of interest is modified. The complex can be formed in vitro or ex vivo and introduced into a cell or contacted with RNA; or can be formed in vivo.

[00343] In some embodiment of the method of modifying expression of a target gene of interest, the target gene is in a prokaryotic cell.

[00344] In some embodiment of the method of modifying expression of a target gene of interest, the target gene is in a eukaryotic cell.

[00345] The invention provides a cell comprising a modified target of interest, wherein the target of interest has been modified according to any of the method disclosed herein.

[00346] In some embodiment of the invention, the cell is a prokaryotic cell.

[00347] In some embodiment of the invention, the cell is a eukaryotic cell.

[00348] In some embodiment, modification of the target of interest in a cell results in: a cell comprising altered expression of at least one gene product; a cell comprising altered expression of at least one gene product, wherein the expression of the at least one gene product is increased; or a cell comprising altered expression of at least one gene product, wherein the expression of the at least one gene product is decreased.

[00349] In some embodiment, the cell is a mammalian cell or a human cell.

[00350] The invention provides a cell line of or comprising a cell disclosed herein or a cell modified by any of the methods disclosed herein, or progeny thereof.

[00351] The invention provides a multicellular organism comprising one or more cells disclosed herein or one or more cells modified according to any of the methods disclosed herein.

[00352] The invention provides a plant or animal model comprising one or more cells disclosed herein or one or more cells modified according to any of the methods disclosed herein.

[00353] The invention provides a gene product from a cell or the cell line or the organism or the plant or animal model disclosed herein.

[00354] In some embodiment, the amount of gene product expressed is greater than or less than the amount of gene product from a cell that does not have altered expression.

[00355] The invention provides a method of identifying the requirements of a suitable guide sequence for the Cas protein of the invention, said method comprising: (a) selecting a set of essential genes within an organism, (b) designing a library of targeting guide sequences capable of hybridizing to regions the coding regions of these genes as well as 5' and 3' UTRs of these genes, (c) generating randomized guide sequences that do not hybridize to any region within the genome of said organism as control guides, (d) preparing a plasmid comprising the nucleic acid-targeting protein and a first resistance gene and a guide plasmid library comprising said library of targeting guides and said control guides and a second resistance gene, (e) co-introducing said plasmids into a host cell, (f) introducing said host cells on a selective medium for said first and second resistance genes, (g) sequencing essential genes of growing host cells, (h) determining significance of depletion of cells transformed with targeting guides by comparing depletion of cells with control guides; and, (i) determining based on the depleted guide sequences the requirements of a suitable guide sequence.

[00356] The invention provides a method of modifying sequences associated with or at a target locus of interest, the method comprising delivering to said locus a non-naturally occurring or engineered composition comprising a Cas protein and one or more nucleic acid components, wherein the effector protein forms a complex with the one or more nucleic acid components and upon binding of the said complex to the locus of interest the effector protein induces the modification of the sequences associated with or at the target locus of interest. In a preferred embodiment, the modification is the introduction of a strand break. In a preferred embodiment, the sequences associated with or at the target locus of interest comprises RNA or consists of RNA.

[00357] The invention provides a method of modifying sequences associated with or at a target locus of interest, the method comprising delivering to said locus a non-naturally occurring or engineered composition comprising a Cas protein, optionally a small accessory protein, and one or more nucleic acid components, wherein the effector protein forms a complex with the one or more nucleic acid components and upon binding of the said complex to the locus of interest the effector protein induces the modification of the sequences associated with or at the target locus of interest. In a preferred embodiment, the modification is the

introduction of a strand break. In a preferred embodiment, the sequences associated with or at the target locus of interest comprises RNA or consists of RNA.

[00358] The invention provides a method of modifying sequences associated with or at a target locus of interest, the method comprising delivering to said sequences associated with or at the locus a non-naturally occurring or engineered composition comprising a Cas loci effector protein and one or more nucleic acid components, wherein the Cas protein forms a complex with the one or more nucleic acid components and upon binding of the said complex to the locus of interest the effector protein induces the modification of sequences associated with or at the target locus of interest. In a preferred embodiment, the modification is the introduction of a strand break. In a preferred embodiment the Cas protein forms a complex with one nucleic acid component; advantageously an engineered or non-naturally occurring nucleic acid component. The induction of modification of sequences associated with or at the target locus of interest can be Cas protein-nucleic acid guided. In a preferred embodiment the one nucleic acid component is a CRISPR RNA (crRNA). In a preferred embodiment the one nucleic acid component is a mature crRNA or guide RNA, wherein the mature crRNA or guide RNA comprises a spacer sequence (or guide sequence) and a direct repeat (DR) sequence or derivatives thereof. In a preferred embodiment the spacer sequence or the derivative thereof comprises a seed sequence, wherein the seed sequence is critical for recognition and/or hybridization to the sequence at the target locus. In a preferred embodiment of the invention the crRNA is a short crRNA that may be associated with a short DR sequence. In another embodiment of the invention the crRNA is a long crRNA that may be associated with a long DR sequence (or dual DR). Aspects of the invention relate to Cas protein complexes having one or more non-naturally occurring or engineered or modified or optimized nucleic acid components. In a preferred embodiment the nucleic acid component comprises RNA. In a preferred embodiment the nucleic acid component of the complex may comprise a guide sequence linked to a direct repeat sequence, wherein the direct repeat sequence comprises one or more stem loops or optimized secondary structures. In preferred embodiments of the invention, the direct repeat may be a short DR or a long DR (dual DR). In a preferred embodiment the direct repeat may be modified to comprise one or more protein-binding RNA aptamers. In a preferred embodiment, one or more aptamers may be included such as part of optimized secondary structure. Such aptamers may be capable of binding a bacteriophage coat protein. In a preferred embodiment the bacteriophage coat protein is MS2. The invention also provides for the nucleic acid component of the complex being 30 or more, 40 or more or 50 or more nucleotides in length.

[00359] The invention provides methods of genome editing or modifying sequences associated with or at a target locus of interest wherein the method comprises introducing a Cas complex into any desired cell type, prokaryotic or eukaryotic cell, whereby the Cas protein complex effectively functions to interfere with RNA in the eukaryotic or prokaryotic cell. In preferred embodiments, the cell is a eukaryotic cell and the RNA is transcribed from a mammalian genome or is present in a mammalian cell. In preferred methods of RNA editing or genome editing in human cells, the Cas proteins may include but are not limited to the specific species of Cas proteins disclosed herein.

[00360] The invention also provides a method of modifying a target locus of interest, the method comprising delivering to said locus a non-naturally occurring or engineered composition comprising a Cas protein and one or more nucleic acid components, wherein the Cas protein forms a complex with the one or more nucleic acid components and upon binding of the said complex to the locus of interest the effector protein induces the modification of the target locus of interest. In a preferred embodiment, the modification is the introduction of a strand break.

[00361] In such methods the target locus of interest may be comprised within a RNA molecule. In such methods the target locus of interest may be comprised in a RNA molecule in vitro.

[00362] In such methods the target locus of interest may be comprised in a RNA molecule within a cell. The cell may be a prokaryotic cell or a eukaryotic cell. The cell may be a mammalian cell. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell are altered for improved production of biologic products such as an antibody, starch, alcohol or other desired cellular output. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell include an alteration that changes the biologic product produced.

[00363] The mammalian cell many be a non-human mammal, e.g., primate, bovine, ovine, porcine, canine, rodent, Leporidae such as monkey, cow, sheep, pig, dog, rabbit, rat or mouse cell. The cell may be a non-mammalian eukaryotic cell such as poultry bird (e.g., chicken), vertebrate fish (e.g., salmon) or shellfish (e.g., oyster, claim, lobster, shrimp) cell. The cell may also be a plant cell. The plant cell may be of a monocot or dicot or of a crop or grain plant such as cassava, corn, sorghum, soybean, wheat, oat or rice. The plant cell may also be of an algae, tree or production plant, fruit or vegetable (e.g., trees such as citrus trees, e.g., orange, grapefruit or lemon trees; peach or nectarine trees; apple or pear trees; nut trees such as almond or walnut or pistachio trees; nightshade plants; plants of the genus Brassica; plants of the genus

Lectica; plants of the genus Spinalis; plants of the genus Capsicum; cotton, tobacco, asparagus, carrot, cabbage, broccoli, cauliflower, tomato, eggplant, pepper, lettuce, spinach, strawberry, blueberry, raspberry, blackberry, grape, coffee, cocoa).

[00364] The invention provides a method of modifying a target locus of interest, the method comprising delivering to said locus a non-naturally occurring or engineered composition comprising a Cas protein and one or more nucleic acid components, wherein the effector protein forms a complex with the one or more nucleic acid components and upon binding of the said complex to the locus of interest the effector protein induces the modification of the target locus of interest. In a preferred embodiment, the modification is the introduction of a strand break.

[00365] In such methods the target locus of interest may be comprised within an RNA molecule. In a preferred embodiment, the target locus of interest comprises or consists of RNA.

[00366] The invention also provides a method of modifying a target locus of interest, the method comprising delivering to said locus a non-naturally occurring or engineered composition comprising a Cas protein and one or more nucleic acid components, wherein the Cas protein forms a complex with the one or more nucleic acid components and upon binding of the said complex to the locus of interest the effector protein induces the modification of the target locus of interest. In a preferred embodiment, the modification is the introduction of a strand break.

[00367] Preferably, in such methods the target locus of interest may be comprised in a RNA molecule in vitro. Also preferably, in such methods the target locus of interest may be comprised in a RNA molecule within a cell. The cell may be a prokaryotic cell or a eukaryotic cell. The cell may be a mammalian cell. The cell may be a rodent cell. The cell may be a mouse cell.

[00368] In any of the described methods the target locus of interest may be a genomic or epigenomic locus of interest. In any of the described methods the complex may be delivered with multiple guides for multiplexed use. In any of the described methods more than one protein(s) may be used.

[00369] In further aspects of the invention the nucleic acid components may comprise a CRISPR RNA (crRNA) sequence. As the effector protein is a Cas protein, the nucleic acid components may comprise a CRISPR RNA (crRNA) sequence and generally may not comprise any trans-activating crRNA (tracr RNA) sequence.

[00370] In any of the described methods the effector protein and nucleic acid components may be provided via one or more polynucleotide molecules encoding the protein and/or nucleic

acid component(s), and wherein the one or more polynucleotide molecules are operably configured to express the protein and/or the nucleic acid component(s). The one or more polynucleotide molecules may comprise one or more regulatory elements operably configured to express the protein and/or the nucleic acid component(s). The one or more polynucleotide molecules may be comprised within one or more vectors. In any of the described methods the target locus of interest may be a genomic, epigenomic, or transcriptomic locus of interest.

TRANSCRIPT TRACKING

[00371] In another aspect, the present disclosure provides compositions and methods for transcript tracking. In some embodiments, transcript tracking allows researchers to visualize transcripts in cells, tissues, organs or animals, providing important spatio-temporal information regarding RNA dynamics and function.

[00372] In some embodiments, the compositions may be a Cas protein herein with one or more labels, or a CRISPR-Cas system comprising such labeled Cas protein. The Cas protein or system may bind to one or more transcripts such that the transcripts may be detected (e.g., visualized) using the label on the Cas protein.

[00373] In some embodiments, the present disclosure includes a system for expressing a Cas protein with one or more polypeptides or polynucleotide labels. The system may comprise polynucleotides encoding the Cas protein and/or the labels. The system may further include vector systems comprising such polynucleotides. For example, a Cas protein may be fused with a fluorescent protein or a fragment thereof. Examples of fluorescent proteins include GFP proteins, EGFP, Azami-Green, Kaede, ZsGreen1 and CopGFP; CFP proteins, such as Cerulean, mCFP, AmCyan1, MiCy, and CyPet; BFP proteins such as EBFP; YFP proteins such as EYFP, YPet, Venus, ZsYellow, and mCitrine; OFP proteins such as cOFP, mKO, and mOrange; red fluorescent protein, or RFP; red or far-red fluorescent proteins from any other species, such as *Heteractis* reef coral and *Actinia* or *Entacmaea* sea anemone, as well as variants thereof. RFPs include, for example, *Discosoma*variants, such as mRFP1, mCherry, tdTomato, mStrawberry, mTangerine, DsRed2, and DsRed-T1, *Anthomedusa* J-Red and Anemonia AsRed2. Far-red fluorescent proteins include, for example, *Actinia* AQ143, *Entacmaea* eqFP611, *Discosoma* variants such as mPlum and mRasberry, and *Heteractis* HcRed1 and t-HcRed.

[00374] In some cases, the systems for expressing the labeled Cas protein may be inducible. For example, the systems may comprise polynucleotides encoding the Cas protein and/or labels under control of a regulatory element herein, e.g., inducible promoters. Such systems may

allow spatial and/or temporal control of the expression of the labels, thus enabling spatial and/or temporal control of transcript tracking.

[00375] In certain cases, the CRISPR-Cas may be labeled with a detectable tag. The labeling may be performed in cells. Alternatively or additionally, the labeling may be performed first and the labeled Cas protein is then delivered into cells, tissues, organs, or organs.

[00376] The detectable tags may be detected (e.g., visualized by imaging, ultrasound, or MRI). Examples of such detectable tags include detectable oligonucleotide tags may be, but are not limited to, oligonucleotides comprising unique nucleotide sequences, oligonucleotides comprising detectable moieties, and oligonucleotides comprising both unique nucleotide sequences and detectable moieties. In some cases, the detectable tag comprises a labeling substance, which is detectable by spectroscopic, photochemical, biochemical, immunochemical, electrical, optical or chemical means. Such tags include biotin for staining with labeled streptavidin conjugate, magnetic beads (e.g., Dynabeads®), fluorescent dyes (e.g., fluorescein, texas red, rhodamine, green fluorescent protein, and the like), radiolabels (e.g., ^{3}H, ^{125}I, ^{35}S, ^{14}C, or ^{32}P), enzymes (e.g., horse radish peroxidase, alkaline phosphatase and others commonly used in an ELISA), and calorimetric labels such as colloidal gold or colored glass or plastic (e.g., polystyrene, polypropylene, latex, etc.) beads. Detectable tags may be detected by many methods. For example, radiolabels may be detected using photographic film or scintillation counters, fluorescent markers may be detected using a photodetector to detect emitted light. Enzymatic labels are typically detected by providing the enzyme with a substrate and detecting, the reaction product produced by the action of the enzyme on the substrate, and calorimetric labels are detected by simply visualizing the colored label. Examples of the labeling substance which may be employed include labeling substances known to those skilled in the art, such as fluorescent dyes, enzymes, coenzymes, chemiluminescent substances, and radioactive substances. Specific examples include radioisotopes (e.g., ^{32}P, ^{14}C, ^{125}I, ^{3}H, and ^{131}I), fluorescein, rhodamine, dansyl chloride, umbelliferone, luciferase, peroxidase, alkaline phosphatase, β-galactosidase, β-glucosidase, horseradish peroxidase, glucoamylase, lysozyme, saccharide oxidase, microperoxidase, biotin, and ruthenium. In the case where biotin is employed as a labeling substance, preferably, after addition of a biotin-labeled antibody, streptavidin bound to an enzyme (e.g., peroxidase) is further added. Advantageously, the label is a fluorescent label. Examples of fluorescent labels include, but are not limited to, Atto dyes, 4-acetamido-4'-isothiocyanatostilbene-2,2'disulfonic acid; acridine and derivatives: acridine, acridine isothiocyanate; 5-(2'-aminoethyl)aminonaphthalene-1-sulfonic acid (EDANS); 4-amino-N-[3-vinylsulfonyl)phenyl]naphthalimide-3,5 disulfonate; N-(4-anilino-1-

naphthyl)maleimide; anthranilamide; BODIPY; Brilliant Yellow; coumarin and derivatives; coumarin, 7-amino-4-methylcoumarin (AMC, Coumarin 120), 7-amino-4-trifluoromethylcouluarin (Coumaran 151); cyanine dyes; cyanosine; 4',6-diaminidino-2-phenylindole (DAPI); 5'5"-dibromopyrogallol-sulfonaphthalein (Bromopyrogallol Red); 7-diethylamino-3-(4'-isothiocyanatophenyl)-4-methylcoumarin; diethylenetriamine pentaacetate; 4,4'-diisothiocyanatodihydro-stilbene-2,2'-disulfonic acid; 4,4'-diisothiocyanatostilbene-2,2'-disulfonic acid; 5-[dimethylamino]naphthalene-1-sulfonyl chloride (DNS, dansylchloride); 4-dimethylaminophenylazophenyl-4'-isothiocyanate (DABITC); eosin and derivatives; eosin, eosin isothiocyanate, erythrosin and derivatives; erythrosin B, erythrosin, isothiocyanate; ethidium; fluorescein and derivatives; 5-carboxyfluorescein (FAM), 5-(4,6-dichlorotriazin-2-yl)aminofluorescein (DTAF), 2',7'-dimethoxy-4'5'-dichloro-6-carboxyfluorescein, fluorescein, fluorescein isothiocyanate, QFITC, (XRITC); fluorescamine; IR144; IR1446; Malachite Green isothiocyanate; 4-methylumbelliferoneortho cresolphthalein; nitrotyrosine; pararosaniline; Phenol Red; B-phycoerythrin; o-phthaldialdehyde; pyrene and derivatives: pyrene, pyrene butyrate, succinimidyl 1-pyrene; butyrate quantum dots; Reactive Red 4 (Cibacron.TM. Brilliant Red 3B-A) rhodamine and derivatives: 6-carboxy-X-rhodamine (ROX), 6-carboxyrhodamine (R6G), lissamine rhodamine B sulfonyl chloride rhodamine (Rhod), rhodamine B, rhodamine 123, rhodamine X isothiocyanate, sulforhodamine B, sulforhodamine 101, sulfonyl chloride derivative of sulforhodamine 101 (Texas Red); N,N,N',N' tetramethyl-6-carboxyrhodamine (TAMRA); tetramethyl rhodamine; tetramethyl rhodamine isothiocyanate (TRITC); riboflavin; rosolic acid; terbium chelate derivatives; Cy3; Cy5; Cy5.5; Cy7; IRD 700; IRD 800; La Jolta Blue; phthalo cyanine; and naphthalo cyanine. A fluorescent label may be a fluorescent protein, such as blue fluorescent protein, cyan fluorescent protein, green fluorescent protein, red fluorescent protein, yellow fluorescent protein or any photoconvertible protein. Colorimetric labeling, bioluminescent labeling and/or chemiluminescent labeling may further accomplish labeling. Labeling further may include energy transfer between molecules in the hybridization complex by perturbation analysis, quenching, or electron transport between donor and acceptor molecules, the latter of which may be facilitated by double stranded match hybridization complexes. The fluorescent label may be a perylene or a terrylen. In the alternative, the fluorescent label may be a fluorescent bar code. Advantageously, the label may be light sensitive, wherein the label is light-activated and/or light cleaves the one or more linkers to release the molecular cargo. The light-activated molecular cargo may be a major

light-harvesting complex (LHCII). In another embodiment, the fluorescent label may induce free radical formation. In some embodiments, the detectable moieties may be quantum dots.

[00377] In some embodiments, the present disclosure provides for a system for delivery the labeled Cas proteins or labeled CRISPR-Cas systems. The delivery system may comprise any delivery vehicles, e.g., those described herein such as RNP, liposomes, nanoparticles, exosomes, microvesicles, nucleic acid nanoassemblies, a gene gun, an implantable device, or the vector systems herein.

NUCLEIC ACID TARGETING

[00378] In certain embodiments, the Cas protein herein is, or in, or comprises, or consists essentially of, or consists of, or involves or relates to Cas protein herein, wherein one or more amino acids are mutated, as described herein elsewhere. Thus, in some embodiments, the effector protein may be a RNA-binding protein, such as a dead-Cas type effector protein, which may be optionally functionalized as described herein for instance with an transcriptional activator or repressor domain, NLS or other functional domain. In some embodiments, the effector protein may be a RNA-binding protein that cleaves a single strand of RNA. If the RNA bound is ssRNA, then the ssRNA is fully cleaved. In some embodiments, the effector protein may be a RNA-binding protein that cleaves a double strand of RNA, for example if it comprises two RNase domains. If the RNA bound is dsRNA, then the dsRNA is fully cleaved. In some embodiments, the effector protein may be a RNA-binding protein that has nickase activity, i.e. it binds dsRNA, but only cleaves one of the RNA strands.

[00379] RNase function in CRISPR systems is known, for example mRNA targeting has been reported for certain type III CRISPR-Cas systems (Hale et al., 2014, Genes Dev, vol. 28, 2432-2443; Hale et al., 2009, Cell, vol. 139, 945-956; Peng et al., 2015, Nucleic acids research, vol. 43, 406-417) and provides significant advantages. A CRISPR-Cas system, composition or method targeting RNA via the present effector proteins is thus provided.

[00380] The target RNA, i.e. the RNA of interest, is the RNA to be targeted by the present invention leading to the recruitment to, and the binding of the effector protein at, the target site of interest on the target RNA. The target RNA may be any suitable form of RNA. This may include, in some embodiments, mRNA. In other embodiments, the target RNA may include tRNA or rRNA.

DNA CLEAVAGE AND REPAIR

[00381] 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 template can be introduced into the genome sequence. In some methods, the HDR process is used modify genome sequence. For example, an exogenous polynucleotide template 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. The exogenous polynucleotide template comprises a sequence to be integrated (e.g., a mutated gene). The sequence for integration may be a sequence endogenous or exogenous to the cell. Examples of a sequence to be integrated include polynucleotides encoding a protein or a non-coding RNA (e.g., a microRNA). Thus, the sequence for integration may be operably linked to an appropriate control sequence or sequences. Alternatively, the sequence to be integrated may provide a regulatory function. The upstream and downstream sequences in the exogenous polynucleotide template are selected to promote recombination between the chromosomal sequence of interest and the donor polynucleotide. The upstream sequence is a nucleic acid sequence that shares sequence similarity with the genome sequence upstream of the targeted site for integration. Similarly, the downstream sequence is a nucleic acid sequence that shares sequence similarity with the chromosomal sequence downstream of the targeted site of integration. The upstream and downstream sequences in the exogenous polynucleotide template can have 75%, 80%, 85%, 90%, 95%, or 100% sequence identity with the targeted genome sequence. Preferably, the upstream and downstream sequences in the exogenous polynucleotide template have about 95%, 96%, 97%, 98%, 99%, or 100% sequence identity with the targeted genome sequence. In some methods, the upstream and downstream sequences in the exogenous polynucleotide template have about 99% or 100% sequence identity with the targeted genome sequence. 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 template may further comprise a marker. Such a marker may make it easy to screen for targeted integrations. Examples of suitable markers include restriction sites, fluorescent proteins, or selectable markers. The exogenous polynucleotide template of the invention can be constructed using recombinant techniques (see, for example, Sambrook et al., 2001 and Ausubel et al., 1996). In a method for modifying a target polynucleotide by integrating an exogenous polynucleotide template, a double stranded break is introduced into the genome sequence by the CRISPR complex, the break is repaired via homologous recombination an exogenous polynucleotide template such that the template is integrated into the genome. The presence of a double-stranded break facilitates integration of the template. 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. 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. 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).

Gene Editing or Altering a Target Loci with Cas; HDR and templates

[00382] The double strand break or single strand break in one of the strands advantageously should be sufficiently close to target position such that correction occurs. In an embodiment, the distance is not more than 50, 100, 200, 300, 350 or 400 nucleotides. While not wishing to be bound by theory, it is believed that the break should be sufficiently close to target position such that the break is within the region that is subject to exonuclease-mediated removal during end resection. If the distance between the target position and a break is too great, the mutation may not be included in the end resection and, therefore, may not be corrected, as the template nucleic acid sequence may only be used to correct sequence within the end resection region.

[00383] In an embodiment, in which a guide RNA and a Type IV molecule, in particular Cas or an ortholog or homolog thereof, preferably a Cas nuclease induce a double strand break for the purpose of inducing HDR-mediated correction, the cleavage site is between 0-200 bp (e.g., 0 to 175, 0 to 150, 0 to 125, 0 to 100, 0 to 75, 0 to 50, 0 to 25, 25 to 200, 25 to 175, 25 to 150, 25 to 125, 25 to 100, 25 to 75, 25 to 50, 50 to 200, 50 to 175, 50 to 150, 50 to 125, 50 to 100, 50 to 75, 75 to 200, 75 to 175, 75 to 150, 75 to 1 25, 75 to 100 bp) away from the target position. In an embodiment, the cleavage site is between 0- 100 bp (e.g., 0 to 75, 0 to 50, 0 to 25, 25 to 100, 25 to 75, 25 to 50, 50 to 100, 50 to 75 or 75 to 100 bp) away from the target position. In a further embodiment, two or more guide RNAs complexing with Cas or an ortholog or homolog thereof, may be used to induce multiplexed breaks for purpose of inducing HDR-mediated correction.

[00384] The homology arm should extend at least as far as the region in which end resection may occur, e.g., in order to allow the resected single stranded overhang to find a

complementary region within the donor template. The overall length could be limited by parameters such as plasmid size or viral packaging limits. In an embodiment, a homology arm may not extend into repeated elements. Exemplary homology arm lengths include a least 50, 100, 250, 500, 750 or 1000 nucleotides.

[00385] Target position, as used herein, refers to a site on a target nucleic acid or target gene (e.g., the chromosome) that is modified by a Type IV, in particular Cas or an ortholog or homolog thereof, preferably Cas molecule-dependent process. For example, the target position can be a modified Cas molecule cleavage of the target nucleic acid and template nucleic acid directed modification, e.g., correction, of the target position. In an embodiment, a target position can be a site between two nucleotides, e.g., adjacent nucleotides, on the target nucleic acid into which one or more nucleotides is added. The target position may comprise one or more nucleotides that are altered, e.g., corrected, by a template nucleic acid. In an embodiment, the target position is within a target sequence (e.g., the sequence to which the guide RNA binds). In an embodiment, a target position is upstream or downstream of a target sequence (e.g., the sequence to which the guide RNA binds).

[00386] A template nucleic acid, as that term is used herein, refers to a nucleic acid sequence which can be used in conjunction with a Type IV molecule, in particular Cas or an ortholog or homolog thereof, preferably a Cas molecule and a guide RNA molecule to alter the structure of a target position. In an embodiment, the target nucleic acid is modified to have some or all of the sequence of the template nucleic acid, typically at or near cleavage site(s). In an embodiment, the template nucleic acid is single stranded. In an alternate embodiment, the template nucleic acid is double stranded. In an embodiment, the template nucleic acid is DNA, e.g., double stranded DNA. In an alternate embodiment, the template nucleic acid is single stranded DNA.

[00387] In an embodiment, the template nucleic acid alters the structure of the target position by participating in homologous recombination. In an embodiment, the template nucleic acid alters the sequence of the target position. In an embodiment, the template nucleic acid results in the incorporation of a modified, or non-naturally occurring base into the target nucleic acid.

[00388] The template sequence may undergo a breakage mediated or catalyzed recombination with the target sequence. In an embodiment, the template nucleic acid may include sequence that corresponds to a site on the target sequence that is cleaved by a Cas mediated cleavage event. In an embodiment, the template nucleic acid may include sequence

that corresponds to both, a first site on the target sequence that is cleaved in a first Cas mediated event, and a second site on the target sequence that is cleaved in a second Cas mediated event.

[00389] In certain embodiments, the template nucleic acid can include sequence which results in an alteration in the coding sequence of a translated sequence, e.g., one which results in the substitution of one amino acid for another in a protein product, e.g., transforming a mutant allele into a wild type allele, transforming a wild type allele into a mutant allele, and/or introducing a stop codon, insertion of an amino acid residue, deletion of an amino acid residue, or a nonsense mutation. In certain embodiments, the template nucleic acid can include sequence which results in an alteration in a non-coding sequence, e.g., an alteration in an exon or in a 5' or 3' non-translated or non-transcribed region. Such alterations include an alteration in a control element, e.g., a promoter, enhancer, and an alteration in a cis-acting or trans-acting control element.

[00390] A template nucleic acid having homology with a target position in a target gene may be used to alter the structure of a target sequence. The template sequence may be used to alter an unwanted structure, e.g., an unwanted or mutant nucleotide. The template nucleic acid may include sequence which, when integrated, results in: decreasing the activity of a positive control element; increasing the activity of a positive control element; decreasing the activity of a negative control element; increasing the activity of a negative control element; decreasing the expression of a gene; increasing the expression of a gene; increasing resistance to a disorder or disease; increasing resistance to viral entry; correcting a mutation or altering an unwanted amino acid residue conferring, increasing, abolishing or decreasing a biological property of a gene product, e.g., increasing the enzymatic activity of an enzyme, or increasing the ability of a gene product to interact with another molecule.

[00391] The template nucleic acid may include sequence which results in: a change in sequence of 1 , 2, 3, 4, 5, 6, 7, 8, 9, 10, 1 1 , 12 or more nucleotides of the target sequence. In an embodiment, the template nucleic acid may be 20+/- 10, 30+/- 10, 40+/- 10, 50+/- 10, 60+/- 10, 70+/- 10, 80+/- 10, 90+/- 10, 100+/- 10, 1 10+/- 10, 120+/- 10, 130+/- 10, 140+/- 10, 150+/- 10, 160+/- 10, 170+/- 10, 1 80+/- 10, 190+/- 10, 200+/- 10, 210+/-10, of 220+/- 10 nucleotides in length. In an embodiment, the template nucleic acid may be 30+/-20, 40+/-20, 50+/-20, 60+/- 20, 70+/- 20, 80+/-20, 90+/-20, 100+/-20, 1 10+/-20, 120+/-20, 130+/-20, 140+/-20, I 50+/-20, 160+/-20, 170+/-20, 180+/-20, 190+/-20, 200+/-20, 210+/-20, of 220+/-20 nucleotides in length. In an embodiment, the template nucleic acid is 10 to 1 ,000, 20 to 900, 30 to 800, 40 to 700, 50 to 600, 50 to 500, 50 to 400, 50 to300, 50 to 200, or 50 to 100 nucleotides in length.

[00392] A template nucleic acid comprises the following components: [5' homology arm]-[replacement sequence]-[3' homology arm]. The homology arms provide for recombination into the chromosome, thus replacing the undesired element, e.g., a mutation or signature, with the replacement sequence. In an embodiment, the homology arms flank the most distal cleavage sites. In an embodiment, the 3' end of the 5' homology arm is the position next to the 5' end of the replacement sequence. In an embodiment, the 5' homology arm can extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' from the 5' end of the replacement sequence. In an embodiment, the 5' end of the 3' homology arm is the position next to the 3' end of the replacement sequence. In an embodiment, the 3' homology arm can extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 3' from the 3' end of the replacement sequence.

[00393] In certain embodiments, one or both homology arms may be shortened to avoid including certain sequence repeat elements. For example, a 5' homology arm may be shortened to avoid a sequence repeat element. In other embodiments, a 3' homology arm may be shortened to avoid a sequence repeat element. In some embodiments, both the 5' and the 3' homology arms may be shortened to avoid including certain sequence repeat elements.

[00394] In certain embodiments, a template nucleic acids for correcting a mutation may designed for use as a single-stranded oligonucleotide. When using a single-stranded oligonucleotide, 5' and 3' homology arms may range up to about 200 base pairs (bp) in length, e.g., at least 25, 50, 75, 100, 125, 150, 175, or 200 bp in length.

DNA repair and NHEJ

[00395] In certain embodiments, nuclease-induced non-homologous end-joining (NHEJ) can be used to target gene-specific knockouts. Nuclease-induced NHEJ can also be used to remove (e.g., delete) sequence in a gene of interest. Generally, NHEJ repairs a double-strand break in the DNA by joining together the two ends; however, generally, the original sequence is restored only if two compatible ends, exactly as they were formed by the double-strand break, are perfectly ligated. The DNA ends of the double-strand break are frequently the subject of enzymatic processing, resulting in the addition or removal of nucleotides, at one or both strands, prior to rejoining of the ends. This results in the presence of insertion and/or deletion (indel) mutations in the DNA sequence at the site of the NHEJ repair. Two-thirds of these mutations typically alter the reading frame and, therefore, produce a non-functional protein. Additionally, mutations that maintain the reading frame, but which insert or delete a significant amount of sequence, can destroy functionality of the protein. This is locus dependent as

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

[00396] Because NHEJ is a mutagenic process, it may also be used to delete small sequence motifs as long as the generation of a specific final sequence is not required. If a double-strand break is targeted near to a short target sequence, the deletion mutations caused by the NHEJ repair often span, and therefore remove, the unwanted nucleotides. For the deletion of larger DNA segments, introducing two double-strand breaks, one on each side of the sequence, can result in NHEJ between the ends with removal of the entire intervening sequence. Both of these approaches can be used to delete specific DNA sequences; however, the error-prone nature of NHEJ may still produce indel mutations at the site of repair.

[00397] Both double strand cleaving Type IV molecule, in particular Cas or an ortholog or homolog thereof, preferably Cas molecules and single strand, or nickase, Type IV molecule, in particular Cas or an ortholog or homolog thereof, preferably Cas molecules can be used in the methods and compositions described herein to generate NHEJ- mediated indels. NHEJ-mediated indels targeted to the gene, e.g., a coding region, e.g., an early coding region of a gene of interest can be used to knockout (i.e., eliminate expression of) a gene of interest. For example, early coding region of a gene of interest includes sequence immediately following a transcription start site, within a first exon of the coding sequence, or within 500 bp of the transcription start site (e.g., less than 500, 450, 400, 350, 300, 250, 200, 150, 100 or 50 bp).

[00398] In an embodiment, in which a guide RNA and Type IV molecule, in particular Cas or an ortholog or homolog thereof, preferably Cas nuclease generate a double strand break for the purpose of inducing NHEJ-mediated indels, a guide RNA may be configured to position one double-strand break in close proximity to a nucleotide of the target position. In an embodiment, the cleavage site may be between 0-500 bp away from the target position (e.g., less than 500, 400, 300, 200, 100, 50, 40, 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 bp from the target position).

[00399] In an embodiment, in which two guide RNAs complexing with Type IV molecules, in particular Cas or an ortholog or homolog thereof, preferably Cas nickases induce two single strand breaks for the purpose of inducing NHEJ-mediated indels, two guide RNAs may be configured to position two single-strand breaks to provide for NHEJ repair a nucleotide of the target position.

SELF-INACTIVATING SYSTEMS

[00400] Once all copies of RNA in a cell have been edited, continued a CRISPR-Cas protein expression or activity in that cell is no longer necessary. A Self-Inactivating system that relies on the use of RNA as to the CRISPR-Cas or crRNA as the guide target sequence can shut down the system by preventing expression of CRISPR-Cas or complex formation.

MODULATION OF CELLULAR STATUS

[00401] In certain embodiments CRISPR-Cas in a complex with crRNA is activated upon binding to target RNA and subsequently cleaves any nearby ssRNA targets (i.e, "collateral" or "bystander" effects). CRISPR-Cas, once primed by the cognate target, can cleave other (non-complementary) RNA molecules. Such promiscuous RNA cleavage could potentially cause cellular toxicity, or otherwise affect cellular physiology or cell status.

[00402] Accordingly, in certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of cell dormancy. In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of cell cycle arrest. In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in reduction of cell growth and/or cell proliferation, In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of cell anergy. In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of cell apoptosis. In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of cell necrosis. In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of cell death. In certain embodiments, the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein are used for or are for use in induction of programmed cell death.

[00403] In certain embodiments, the invention relates to a method for induction of cell dormancy comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for induction of cell cycle arrest comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for reduction of cell growth and/or cell proliferation comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for induction of cell anergy comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for induction of cell apoptosis comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for induction of cell necrosis comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for induction of cell death comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates to a method for induction of programmed cell death comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein.

[00404] The methods and uses as described herein may be therapeutic or prophylactic and may target particular cells, cell (sub)populations, or cell/tissue types. In particular, the methods and uses as described herein may be therapeutic or prophylactic and may target particular cells, cell (sub)populations, or cell/tissue types expressing one or more target sequences, such as one or more particular target RNA (e.g., ss RNA). Without limitation, target cells may for instance be cancer cells expressing a particular transcript, e.g. neurons of a given class, (immune) cells causing e.g. autoimmunity, or cells infected by a specific (e.g., viral) pathogen, etc.

[00405] Accordingly, in certain embodiments, the invention relates to a method for treating a pathological condition characterized by the presence of undesirable cells (host cells), comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. In certain embodiments, the invention relates the use of the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for treating a pathological condition characterized by the

presence of undesirable cells (host cells). In certain embodiments, the invention relates the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for use in treating a pathological condition characterized by the presence of undesirable cells (host cells). It is to be understood that preferably the CRISPR-Cas system targets a target specific for the undesirable cells. In certain embodiments, the invention relates to the use of the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for treating, preventing, or alleviating cancer. In certain embodiments, the invention relates to the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for use in treating, preventing, or alleviating cancer. In certain embodiments, the invention relates to a method for treating, preventing, or alleviating cancer comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. It is to be understood that preferably the CRISPR-Cas system targets a target specific for the cancer cells. In certain embodiments, the invention relates to the use of the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for treating, preventing, or alleviating infection of cells by a pathogen. In certain embodiments, the invention relates to the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for use in treating, preventing, or alleviating infection of cells by a pathogen. In certain embodiments, the invention relates to a method for treating, preventing, or alleviating infection of cells by a pathogen comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. It is to be understood that preferably the CRISPR-Cas system targets a target specific for the cells infected by the pathogen (e.g. a pathogen derived target). In certain embodiments, the invention relates to the use of the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for treating, preventing, or alleviating an autoimmune disorder. In certain embodiments, the invention relates to the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein for use in treating, preventing, or alleviating an autoimmune disorder. In certain embodiments, the invention relates to a method for treating, preventing, or alleviating an autoimmune disorder comprising introducing or inducing the non-naturally occurring or engineered composition, vector system, or delivery systems as described herein. It is to be understood that preferably the CRISPR-Cas system targets a target specific for the cells responsible for the autoimmune disorder (e.g. specific immune cells).

IN VITRO APEX LABELING

[00406] Cellular processes depend on a network of molecular interactions among protein, RNA, and DNA. Accurate detection of protein−DNA and protein−RNA interactions is key to understanding such processes. In vitro proximity labeling technology employs an affinity tag combined with e.g. a photoactivatable probe to label polypeptides and RNAs in the vicinity of a protein or RNA of interest in vitro. After UV irradiation the photoactivatable group reacts with proteins and other molecules that are in close proximity to the tagged molecule, thereby labelling them. Labelled interacting molecules can subsequently be recovered and identified. The Cas protein of the invention can for instance be used to target a probe to a selected RNA sequence.

[00407] These applications could also be applied in animal models for in vivo imaging of disease relevant applications or difficult-to culture cell types.

SYNTHETIC BIOLOGY

[00408] The development of biological systems has a wide utility, including in clinical applications. It is envisaged that the programmable Cas proteins of the invention can be used fused to split proteins of toxic domains for targeted cell death, for instance using cancer-linked RNA as target transcript. Further, pathways involving protein-protein interaction can be influenced in synthetic biological systems with e.g. fusion complexes with the appropriate effectors such as kinases or other enzymes.

PROTEIN SPLICING: INTEINS

[00409] Protein splicing is a post-translational process in which an intervening polypeptide, referred to as an intein, catalyzes its own excision from the polypeptides flacking it, referred to as exteins, as well as subsequent ligation of the exteins. The assembly of two or more Cas proteins as described herein on a target transcript could be used to direct the release of a split intein (Topilina and Mills Mob DNA. 2014 Feb 4;5(1):5), thereby allowing for direct computation of the existence of a mRNA transcript and subsequent release of a protein product, such as a metabolic enzyme or a transcription factor (for downstream actuation of transcription pathways). This application may have significant relevance in synthetic biology (see above) or large-scale bioproduction (only produce product under certain conditions).

INDUCIBLE, DOSED AND SELF-INACTIVATING SYSTEMS

[00410] In one embodiment, fusion complexes comprising an Cas protein of the invention and an effector component are designed to be inducible, for instance light inducible or chemically inducible. Such inducibility allows for activation of the effector component at a desired moment in time.

[00411] Light inducibility is for instance achieved by designing a fusion complex wherein CRY2PHR/CIBN pairing is used for fusion. This system is particularly useful for light induction of protein interactions in living cells (Konermann S, et al. Nature. 2013;500:472–476).

[00412] Chemical inducibility is for instance provided for by designing a fusion complex wherein FKBP/FRB (FK506 binding protein / FKBP rapamycin binding) pairing is used for fusion. Using this system rapamycin is required for binding of proteins (Zetsche et al. Nat Biotechnol. 2015;33(2):139-42 describes the use of this system for Cas).

[00413] Further, when introduced in the cell as DNA, the Cas protein of the inventions can be modulated by inducible promoters, such as tetracycline or doxycycline controlled transcriptional activation (Tet-On and Tet-Off expression system), hormone inducible gene expression system such as for instance an ecdysone inducible gene expression system and an arabinose-inducible gene expression system. When delivered as RNA, expression of the Cas protein can be modulated via a riboswitch, which can sense a small molecule like tetracycline (as described in Goldfless et al. Nucleic Acids Res. 2012;40(9):e64).

[00414] In one embodiment, the delivery of the Cas protein of the invention can be modulated to change the amount of protein or crRNA in the cell, thereby changing the magnitude of the desired effect or any undesired off-target effects.

[00415] In one embodiment, the Cas proteins described herein can be designed to be self-inactivating. When delivered to a cell as RNA, either mRNA or as a replication RNA therapeutic (Wrobleska et al Nat Biotechnol. 2015 Aug; 33(8): 839–841), they can self-inactivate expression and subsequent effects by destroying the own RNA, thereby reducing residency and potential undesirable effects.

[00416] For further *in vivo* applications of Cas proteins as described herein, reference is made to Mackay JP et al (Nat Struct Mol Biol. 2011 Mar;18(3):256-61), Nelles et al (Bioessays. 2015 Jul;37(7):732-9) and Abil Z and Zhao H (Mol Biosyst. 2015 Oct;11(10):2658-65), which are incorporated herein by reference. In particular, the following applications are envisaged in certain embodiments of the invention, preferably in certain embodiments by using catalytically inactive CRISPR-Cas: enhancing translation (e.g. CRISPR-Cas – translation promotion factor fusions (e.g. eIF4 fusions)); repressing translation (e.g. gRNA targeting ribosome binding sites); exon skipping (e.g. gRNAs targeting splice donor and/or acceptor sites); exon inclusion (e.g. gRNA targeting a particular exon splice donor and/or acceptor site to be included or CRISPR-Cas fused to or recruiting spliceosome components (e.g. U1 snRNA)); accessing RNA localization (e.g. CRISPR-Cas – marker fusions (e.g. EGFP fusions)); altering RNA

localization (e.g. CRISPR-Cas – localization signal fusions (e.g. NLS or NES fusions)); RNA degradation (in this case no catalytically inactive CRISPR-Cas is to be used if relied on the activity of CRISPR-Cas, alternatively and for increased specificity, a split CRISPR-Cas may be used); inhibition of non-coding RNA function (e.g. miRNA), such as by degradation or binding of gRNA to functional sites (possibly titrating out at specific sites by relocalization by CRISPR-Cas-signal sequence fusions).

[00417] As described herein before and demonstrated in the Examples, CRISPR-Cas function is robust to 5'or 3' extensions of the crRNA and to extension of the crRNA loop. It is therefore envisaging that MS2 loops and other recruitment domains can be added to the crRNA without affecting complex formation and binding to target transcripts. Such modifications to the crRNA for recruitment of various effector domains are applicable in the uses of a RNA targeted effector proteins described above.

[00418] CRISPR-Cas is capable of mediating resistance to RNA phages. It is therefore envisaged that CRISPR-Cas can be used to immunize, e.g. animals, humans and plants, against RNA-only pathogens, including but not limited to Ebola virus and Zika virus.

[00419] In certain embodiments, CRISPR-Cas can process (cleave) its own array. This applies to both the wildtype Cas protein and the mutated Cas protein containing one or more mutated amino acid residues as herein discussed. It is therefore envisaged that multiple crRNAs designed for different target transcripts and/or applications can be delivered as a single pre-crRNA or as a single transcript driven by one promotor. Such method of delivery has the advantages that it is substantially more compact, easier to synthesize and easier to delivery in viral systems. It will be understood that exact amino acid positions may vary for orthologues of a herein CRISPR-Cas can be adequately determined by protein alignment, as is known in the art, and as described herein elsewhere. Aspects of the invention also encompass methods and uses of the compositions and systems described herein in genome engineering, e.g. for altering or manipulating the expression of one or more genes or the one or more gene products, in prokaryotic or eukaryotic cells, *in vitro*, *in vivo* or *ex vivo*.

[00420] In an aspect, the invention provides methods and compositions for modulating, e.g., reducing, expression of a target RNA in cells. In the subject methods, a CRISPR-Cas system of the invention is provided that interferes with transcription, stability, and / or translation of an RNA.

[00421] In certain embodiments, an effective amount of CRISPR-Cas system is used to cleave RNA or otherwise inhibit RNA expression. In this regard, the system has uses similar to siRNA and shRNA, thus can also be substituted for such methods. The method includes,

without limitation, use of a CRISPR-Cas system as a substitute for e.g., an interfering ribonucleic acid (such as an siRNA or shRNA) or a transcription template thereof, e.g., a DNA encoding an shRNA. The CRISPR-Cas system is introduced into a target cell, e.g., by being administered to a mammal that includes the target cell.

[00422] Advantageously, a CRISPR-Cas system of the invention is specific. For example, whereas interfering ribonucleic acid (such as an siRNA or shRNA) polynucleotide systems are plagued by design and stability issues and off-target binding, a CRISPR-Cas system of the invention can be designed with high specificity.

[00423] In an aspect of the invention, novel systems also referred to as RNA- or CRISPR systems of the present application are based on herein-identified Cas proteins which do not require the generation of customized proteins to target specific RNA sequences but rather a single enzyme can be programmed by a RNA molecule to recognize a specific RNA target, in other words the enzyme can be recruited to a specific RNA target using said RNA molecule.

[00424] In some embodiments, one or more elements of a nucleic acid-targeting system is derived from a particular organism comprising an endogenous CRISPR system. In certain embodiments, the CRISPR system is found in Eubacterium and Ruminococcus. In certain embodiments, the effector protein comprises targeted and collateral ssRNA cleavage activity.

[00425] In certain embodiments, the effector protein locus structures include a WYL domain containing accessory protein (so denoted after three amino acids that were conserved in the originally identified group of these domains; see, e.g., WYL domain IPR026881). In certain embodiments, the WYL domain accessory protein comprises at least one helix-turn-helix (HTH) or ribbon-helix-helix (RHH) DNA-binding domain. In certain embodiments, the WYL domain containing accessory protein increases both the targeted and the collateral ssRNA cleavage activity of the Cas protein. In certain embodiments, the WYL domain containing accessory protein comprises an N-terminal RHH domain, as well as a pattern of primarily hydrophobic conserved residues, including an invariant tyrosine-leucine doublet corresponding to the original WYL motif. In certain embodiments, the WYL domain containing accessory protein is WYL1. WYL1 is a single WYL-domain protein associated primarily with Ruminococcus.

GENOME-WIDE KNOCK-OUT SCREENING

[00426] The Cas proteins and systems described herein can be used to perform efficient and cost effective functional genomic screens. Such screens can utilize CRISPR-Cas genome wide libraries. Such screens and libraries can provide for determining the function of genes, cellular pathways genes are involved in, and how any alteration in gene expression can result in a

particular biological process. An advantage of the present invention is that the CRISPR system 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.

[00427] A genome wide library may comprise a plurality of CRISPR-Cas system guide RNAs, as described herein, comprising guide sequences that are capable of targeting a plurality of target sequences in a plurality of genomic loci in a population of eukaryotic cells. The population of cells may be a population of embryonic stem (ES) cells. The target sequence in the genomic locus may be a non-coding sequence. The non-coding sequence may be an intron, regulatory sequence, splice site, 3' UTR, 5' UTR, or polyadenylation signal. Gene function of one or more gene products may be altered by said targeting. The targeting may result in a knockout of gene function. The targeting of a gene product may comprise more than one guide RNA. A gene product may be targeted by 2, 3, 4, 5, 6, 7, 8, 9, or 10 guide RNAs, preferably 3 to 4 per gene. Off-target modifications may be minimized (See, e.g., DNA targeting specificity of RNA-guided Cas9 nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. Nat Biotechnol doi:10.1038/nbt.2647 (2013)), incorporated herein by reference. The targeting may be of about 100 or more sequences. The targeting may be of about 1000 or more sequences. The targeting may be of about 20,000 or more sequences. The targeting may be of the entire genome. The targeting may be of a panel of target sequences focused on a relevant or desirable pathway. The pathway may be an immune pathway. The pathway may be a cell division pathway.

[00428] One aspect of the invention comprehends a genome wide library that may comprise a plurality of CRISPR-Cas system guide RNAs that may comprise guide sequences that are capable of targeting a plurality of target sequences in a plurality of genomic loci, wherein said targeting results in a knockout of gene function. This library may potentially comprise guide RNAs that target each and every gene in the genome of an organism.

[00429] In some embodiments of the invention the organism or subject is a eukaryote (including mammal including human) or a non-human eukaryote or a non-human animal or a non-human mammal. In some embodiments, the organism or subject is a non-human animal, and may be an arthropod, for example, an insect, or may be a nematode. In some methods of the invention the organism or subject is a plant. In some methods of the invention the organism or subject is a mammal or a non-human mammal. A non-human mammal may be for example a rodent (preferably a mouse or a rat), an ungulate, or a primate. In some methods of the invention the organism or subject is algae, including microalgae, or is a fungus.

[00430] The knockout of gene function may comprise: introducing into each cell in the population of cells a vector system of one or more vectors comprising an engineered, non-naturally occurring CRISPR-Cas system comprising I. a Cas protein, and II. one or more guide RNAs, wherein components I and II may be same or on different vectors of the system, integrating components I and II into each cell, wherein the guide sequence targets a unique gene in each cell, wherein the Cas protein is operably linked to a regulatory element, wherein when transcribed, the guide RNA comprising the guide sequence directs sequence-specific binding of a CRISPR-Cas system to a target sequence in the genomic loci of the unique gene, inducing cleavage of the genomic loci by the Cas protein, and confirming different knockout mutations in a plurality of unique genes in each cell of the population of cells thereby generating a gene knockout cell library. The invention comprehends that the population of cells is a population of eukaryotic cells, and in a preferred embodiment, the population of cells is a population of embryonic stem (ES) cells.

[00431] The one or more vectors may be plasmid vectors. The vector may be a single vector comprising Cas, a sgRNA, and optionally, a selection marker into target cells. Not being bound by a theory, the ability to simultaneously deliver Cas and sgRNA through a single vector enables application to any cell type of interest, without the need to first generate cell lines that express Cas. The regulatory element may be an inducible promoter. The inducible promoter may be a doxycycline inducible promoter. In some methods of the invention the expression of the guide sequence is under the control of the T7 promoter and is driven by the expression of T7 polymerase. The confirming of different knockout mutations may be by whole exome sequencing. The knockout mutation may be achieved in 100 or more unique genes. The knockout mutation may be achieved in 1000 or more unique genes. The knockout mutation may be achieved in 20,000 or more unique genes. The knockout mutation may be achieved in the entire genome. The knockout of gene function may be achieved in a plurality of unique genes which function in a particular physiological pathway or condition. The pathway or condition may be an immune pathway or condition. The pathway or condition may be a cell division pathway or condition.

[00432] The invention also provides kits that comprise the genome wide libraries mentioned herein. The kit may comprise a single container comprising vectors or plasmids comprising the library of the invention. The kit may also comprise a panel comprising a selection of unique CRISPR-Cas system guide RNAs comprising guide sequences from the library of the invention, wherein the selection is indicative of a particular physiological condition. The invention comprehends that the targeting is of about 100 or more sequences, about 1000 or

more sequences or about 20,000 or more sequences or the entire genome. Furthermore, a panel of target sequences may be focused on a relevant or desirable pathway, such as an immune pathway or cell division.

APPLICATIONS IN PLANTS AND FUNGI

[00433] In general, the term "plant" relates to any various photosynthetic, eukaryotic, unicellular or multicellular organism of the kingdom Plantae characteristically growing by cell division, containing chloroplasts, and having cell walls comprised of cellulose. The term plant encompasses monocotyledonous and dicotyledonous plants. Specifically, the plants are intended to comprise without limitation angiosperm and gymnosperm plants such as acacia, alfalfa, amaranth, apple, apricot, artichoke, ash tree, asparagus, avocado, banana, barley, beans, beet, birch, beech, blackberry, blueberry, broccoli, Brussel's sprouts, cabbage, canola, cantaloupe, carrot, cassava, cauliflower, cedar, a cereal, celery, chestnut, cherry, Chinese cabbage, citrus, clementine, clover, coffee, corn, cotton, cowpea, cucumber, cypress, eggplant, elm, endive, eucalyptus, fennel, figs, fir, geranium, grape, grapefruit, groundnuts, ground cherry, gum hemlock, hickory, kale, kiwifruit, kohlrabi, larch, lettuce, leek, lemon, lime, locust, pine, maidenhair, maize, mango, maple, melon, millet, mushroom, mustard, nuts, oak, oats, oil palm, okra, onion, orange, an ornamental plant or flower or tree, papaya, palm, parsley, parsnip, pea, peach, peanut, pear, peat, pepper, persimmon, pigeon pea, pine, pineapple, plantain, plum, pomegranate, potato, pumpkin, radicchio, radish, rapeseed, raspberry, rice, rye, sorghum, safflower, sallow, soybean, spinach, spruce, squash, strawberry, sugar beet, sugarcane, sunflower, sweet potato, sweet corn, tangerine, tea, tobacco, tomato, trees, triticale, turf grasses, turnips, vine, walnut, watercress, watermelon, wheat, yams, yew, and zucchini. The term plant also encompasses Algae, which are mainly photoautotrophs unified primarily by their lack of roots, leaves and other organs that characterize higher plants.

[00434] The methods for modulating gene expression using the system as described herein can be used to confer desired traits on essentially any plant. A wide variety of plants and plant cell systems may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation methods mentioned above. In preferred embodiments, target plants and plant cells for engineering include, but are not limited to, those monocotyledonous and dicotyledonous plants, such as crops including grain crops (e.g., wheat, maize, rice, millet, barley), fruit crops (e.g., tomato, apple, pear, strawberry, orange), forage crops (e.g., alfalfa), root vegetable crops (e.g., carrot, potato, sugar beets, yam), leafy vegetable crops (e.g., lettuce, spinach); flowering plants (e.g., petunia, rose, chrysanthemum), conifers and pine trees (e.g.,

pine fir, spruce); plants used in phytoremediation (e.g., heavy metal accumulating plants); oil crops (e.g., sunflower, rape seed) and plants used for experimental purposes (e.g., Arabidopsis). Thus, the methods and CRISPR-Cas systems can be used over a broad range of plants, such as for example with dicotyledonous plants belonging to the orders Magniolales, Illiciales, Laurales, Piperales, Aristochiales, Nymphaeales, Ranunculales, Papeverales, Sarraceniaceae, Trochodendrales, Hamamelidales, Eucomiales, Leitneriales, Myricales, Fagales, Casuarinales, Caryophyllales, Batales, Polygonales, Plumbaginales, Dilleniales, Theales, Malvales, Urticales, Lecythidales, Violales, Salicales, Capparales, Ericales, Diapensales, Ebenales, Primulales, Rosales, Fabales, Podostemales, Haloragales, Myrtales, Cornales, Proteales, San tales, Rafflesiales, Celastrales, Euphorbiales, Rhamnales, Sapindales, Juglandales, Geraniales, Polygalales, Umbellales, Gentianales, Polemoniales, Lamiales, Plantaginales, Scrophulariales, Campanulales, Rubiales, Dipsacales, and Asterales; the methods and CRISPR-Cas systems can be used with monocotyledonous plants such as those belonging to the orders Alismatales, Hydrocharitales, Najadales, Triuridales, Commelinales, Eriocaulales, Restionales, Poales, Juncales, Cyperales, Typhales, Bromeliales, Zingiberales, Arecales, Cyclanthales, Pandanales, Arales, Lilliales, and Orchid ales, or with plants belonging to Gymnospermae, e.g those belonging to the orders Pinales, Ginkgoales, Cycadales, Araucariales, Cupressales and Gnetales.

[00435] The systems and methods of use described herein can be used over a broad range of plant species, included in the non-limitative list of dicot, monocot or gymnosperm genera hereunder: *Atropa, Alseodaphne, Anacardium, Arachis, Beilschmiedia, Brassica, Carthamus, Cocculus, Croton, Cucumis, Citrus, Citrullus, Capsicum, Catharanthus, Cocos, Coffea, Cucurbita, Daucus, Duguetia, Eschscholzia, Ficus, Fragaria, Glaucium, Glycine, Gossypium, Helianthus, Hevea, Hyoscyamus, Lactuca, Landolphia, Linum, Litsea, Lycopersicon, Lupinus, Manihot, Majorana, Malus, Medicago, Nicotiana, Olea, Parthenium, Papaver, Persea, Phaseolus, Pistacia, Pisum, Pyrus, Prunus, Raphanus, Ricinus, Senecio, Sinomenium, Stephania, Sinapis, Solanum, Theobroma, Trifolium, Trigonella, Vicia, Vinca, Vilis, and Vigna; and the genera Allium, Andropogon, Aragrostis, Asparagus, Avena, Cynodon, Elaeis, Festuca, Festulolium, Heterocallis, Hordeum, Lemna, Lolium, Musa, Oryza, Panicum, Pannesetum, Phleum, Poa, Secale, Sorghum, Triticum, Zea, Abies, Cunninghamia, Ephedra, Picea, Pinus,* and *Pseudotsuga.*

[00436] The systems and methods of use can also be used over a broad range of "algae" or "algae cells"; including for example algea selected from several eukaryotic phyla, including the Rhodophyta (red algae), Chlorophyta (green algae), Phaeophyta (brown algae),

Bacillariophyta (diatoms), Eustigmatophyta and dinoflagellates as well as the prokaryotic phylum Cyanobacteria (blue-green algae). The term "algae" includes for example algae selected from : *Amphora*, *Anabaena*, *Anikstrodesmis*, *Botryococcus*, *Chaetoceros*, *Chlamydomonas*, *Chlorella*, *Chlorococcum*, *Cyclotella*, *Cylindrotheca*, *Dunaliella*, *Emiliana*, *Euglena*, *Hematococcus*, *Isochrysis*, *Monochrysis*, *Monoraphidium*, *Nannochloris*, *Nannnochloropsis*, *Navicula*, *Nephrochloris*, *Nephroselmis*, *Nitzschia*, *Nodularia*, *Nostoc*, *Oochromonas*, *Oocystis*, *Oscillartoria*, *Pavlova*, *Phaeodactylum*, *Playtmonas*, *Pleurochrysis*, *Porhyra*, *Pseudoanabaena*, *Pyramimonas*, *Stichococcus*, *Synechococcus*, *Synechocystis*, *Tetraselmis*, *Thalassiosira*, *and Trichodesmium*.

[00437] A part of a plant, e.g., a "plant tissue" may be treated according to the methods of the present invention to produce an improved plant. Plant tissue also encompasses plant cells. The term “plant cell” as used herein refers to individual units of a living plant, either in an intact whole plant or in an isolated form grown in in vitro tissue cultures, on media or agar, in suspension in a growth media or buffer or as a part of higher organized unites, such as, for example, plant tissue, a plant organ, or a whole plant.

[00438] A “protoplast” refers to a plant cell that has had its protective cell wall completely or partially removed using, for example, mechanical or enzymatic means resulting in an intact biochemical competent unit of living plant that can reform their cell wall, proliferate and regenerate grow into a whole plant under proper growing conditions.

[00439] The term "transformation" broadly refers to the process by which a plant host is genetically modified by the introduction of DNA by means of Agrobacteria or one of a variety of chemical or physical methods. As used herein, the term "plant host" refers to plants, including any cells, tissues, organs, or progeny of the plants. Many suitable plant tissues or plant cells can be transformed and include, but are not limited to, protoplasts, somatic embryos, pollen, leaves, seedlings, stems, calli, stolons, microtubers, and shoots. A plant tissue also refers to any clone of such a plant, seed, progeny, propagule whether generated sexually or asexually, and descendants of any of these, such as cuttings or seed.

[00440] The term "transformed" as used herein, refers to a cell, tissue, organ, or organism into which a foreign DNA molecule, such as a construct, has been introduced. The introduced DNA molecule may be integrated into the genomic DNA of the recipient cell, tissue, organ, or organism such that the introduced DNA molecule is transmitted to the subsequent progeny. In these embodiments, the "transformed" or “transgenic” cell or plant may also include progeny of the cell or plant and progeny produced from a breeding program employing such a transformed plant as a parent in a cross and exhibiting an altered phenotype resulting from the

presence of the introduced DNA molecule. Preferably, the transgenic plant is fertile and capable of transmitting the introduced DNA to progeny through sexual reproduction.

[00441] The term "progeny", such as the progeny of a transgenic plant, is one that is born of, begotten by, or derived from a plant or the transgenic plant. The introduced DNA molecule may also be transiently introduced into the recipient cell such that the introduced DNA molecule is not inherited by subsequent progeny and thus not considered "transgenic". Accordingly, as used herein, a "non-transgenic" plant or plant cell is a plant which does not contain a foreign DNA stably integrated into its genome.

[00442] The term "plant promoter" as used herein is a promoter capable of initiating transcription in plant cells, whether or not its origin is a plant cell. Exemplary suitable plant promoters include, but are not limited to, those that are obtained from plants, plant viruses, and bacteria such as Agrobacterium or Rhizobium which comprise genes expressed in plant cells.

[00443] As used herein, a "fungal cell" refers to any type of eukaryotic cell within the kingdom of fungi. Phyla within the kingdom of fungi include Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota, Glomeromycota, Microsporidia, and Neocallimastigomycota. Fungal cells may include yeasts, molds, and filamentous fungi. In some embodiments, the fungal cell is a yeast cell.

[00444] As used herein, the term "yeast cell" refers to any fungal cell within the phyla Ascomycota and Basidiomycota. Yeast cells may include budding yeast cells, fission yeast cells, and mold cells. Without being limited to these organisms, many types of yeast used in laboratory and industrial settings are part of the phylum Ascomycota. In some embodiments, the yeast cell is an *S. cerevisiae, Kluyveromyces marxianus,* or *Issatchenkia orientalis* cell. Other yeast cells may include without limitation *Candida* spp. (e.g., *Candida albicans*), *Yarrowia* spp. (e.g., *Yarrowia lipolytica*), *Pichia* spp. (e.g., *Pichia pastoris*), *Kluyveromyces* spp. (e.g., *Kluyveromyces lactis* and *Kluyveromyces marxianus*), *Neurospora* spp. (e.g., *Neurospora crassa*), Fusarium spp. (e.g., *Fusarium oxysporum*), and *Issatchenkia* spp. (e.g., *Issatchenkia orientalis*, a.k.a. *Pichia kudriavzevii* and *Candida acidothermophilum*). In some embodiments, the fungal cell is a filamentous fungal cell. As used herein, the term "filamentous fungal cell" refers to any type of fungal cell that grows in filaments, i.e., hyphae or mycelia. Examples of filamentous fungal cells may include without limitation *Aspergillus* spp. (e.g., *Aspergillus niger*), Trichoderma spp. (e.g., Trichoderma reesei), Rhizopus spp. (e.g., *Rhizopus oryzae*), and *Mortierella* spp. (e.g., *Mortierella isabellina*).

[00445] In some embodiments, the fungal cell is an industrial strain. As used herein, "industrial strain" refers to any strain of fungal cell used in or isolated from an industrial

process, e.g., production of a product on a commercial or industrial scale. Industrial strain may refer to a fungal species that is typically used in an industrial process, or it may refer to an isolate of a fungal species that may be also used for non-industrial purposes (e.g., laboratory research). Examples of industrial processes may include fermentation (e.g., in production of food or beverage products), distillation, biofuel production, production of a compound, and production of a polypeptide. Examples of industrial strains may include, without limitation, JAY270 and ATCC4124.

[00446] In some embodiments, the fungal cell is a polyploid cell. As used herein, a "polyploid" cell may refer to any cell whose genome is present in more than one copy. A polyploid cell may refer to a type of cell that is naturally found in a polyploid state, or it may refer to a cell that has been induced to exist in a polyploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A polyploid cell may refer to a cell whose entire genome is polyploid, or it may refer to a cell that is polyploid in a particular genomic locus of interest. Without wishing to be bound to theory, it is thought that the abundance of guide RNA may more often be a rate-limiting component in genome engineering of polyploid cells than in haploid cells, and thus the methods using the CRISPR-Cas CRISPR system described herein may take advantage of using a certain fungal cell type.

[00447] In some embodiments, the fungal cell is a diploid cell. As used herein, a "diploid" cell may refer to any cell whose genome is present in two copies. A diploid cell may refer to a type of cell that is naturally found in a diploid state, or it may refer to a cell that has been induced to exist in a diploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the S. cerevisiae strain S228C may be maintained in a haploid or diploid state. A diploid cell may refer to a cell whose entire genome is diploid, or it may refer to a cell that is diploid in a particular genomic locus of interest. In some embodiments, the fungal cell is a haploid cell. As used herein, a "haploid" cell may refer to any cell whose genome is present in one copy. A haploid cell may refer to a type of cell that is naturally found in a haploid state, or it may refer to a cell that has been induced to exist in a haploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the S. cerevisiae strain S228C may be maintained in a haploid or diploid state. A haploid cell may refer to a cell whose entire genome is haploid, or it may refer to a cell that is haploid in a particular genomic locus of interest.

[00448] As used herein, a "yeast expression vector" refers to a nucleic acid that contains one or more sequences encoding an RNA and/or polypeptide and may further contain any desired elements that control the expression of the nucleic acid(s), as well as any elements that enable the replication and maintenance of the expression vector inside the yeast cell. Many suitable yeast expression vectors and features thereof are known in the art; for example, various vectors and techniques are illustrated in in Yeast Protocols, 2nd edition, Xiao, W., ed. (Humana Press, New York, 2007) and Buckholz, R.G. and Gleeson, M.A. (1991) Biotechnology (NY) 9(11): 1067-72. Yeast vectors may contain, without limitation, a centromeric (CEN) sequence, an autonomous replication sequence (ARS), a promoter, such as an RNA Polymerase III promoter, operably linked to a sequence or gene of interest, a terminator such as an RNA polymerase III terminator, an origin of replication, and a marker gene (e.g., auxotrophic, antibiotic, or other selectable markers). Examples of expression vectors for use in yeast may include plasmids, yeast artificial chromosomes, 2μ plasmids, yeast integrative plasmids, yeast replicative plasmids, shuttle vectors, and episomal plasmids.

STABLE INTEGRATION OF SYSTEMS IN THE GENOME OF PLANTS AND PLANT CELLS

[00449] In particular embodiments, it is envisaged that the polynucleotides encoding the components of the CRISPR system are introduced for stable integration into the genome of a plant cell. In these embodiments, the design of the transformation vector or the expression system can be adjusted depending on when, where and under what conditions the guide RNA and/or the gene(s) are expressed.

[00450] In particular embodiments, it is envisaged to introduce the components of the CRISPR system stably into the genomic DNA of a plant cell. Additionally or alternatively, it is envisaged to introduce the components of the CRISPR system for stable integration into the DNA of a plant organelle such as, but not limited to a plastid, e mitochondrion or a chloroplast.

[00451] The expression system for stable integration into the genome of a plant cell may contain one or more of the following elements: a promoter element that can be used to express the guide RNA and/or Cas protein in a plant cell; a 5' untranslated region to enhance expression ; an intron element to further enhance expression in certain cells, such as monocot cells; a multiple-cloning site to provide convenient restriction sites for inserting the one or more guide RNAs and/or the gene sequences and other desired elements; and a 3' untranslated region to provide for efficient termination of the expressed transcript.

[00452] The elements of the expression system may be on one or more expression constructs which are either circular such as a plasmid or transformation vector, or non-circular such as linear double stranded DNA.

In a particular embodiment, a CRISPR expression system comprises at least:

(a) a nucleotide sequence encoding a guide RNA (gRNA) that hybridizes with a target sequence in a plant, and wherein the guide RNA comprises a guide sequence and a direct repeat sequence, and

(b) a nucleotide sequence encoding a Cas protein,

wherein components (a) or (b) are located on the same or on different constructs, and whereby the different nucleotide sequences can be under control of the same or a different regulatory element operable in a plant cell.

[00453] DNA construct(s) containing the components of the CRISPR system, may be introduced into the genome of a plant, plant part, or plant cell by a variety of conventional techniques. The process generally comprises the steps of selecting a suitable host cell or host tissue, introducing the construct(s) into the host cell or host tissue, and regenerating plant cells or plants therefrom.

[00454] In particular embodiments, the DNA construct may be introduced into the plant cell using techniques such as but not limited to electroporation, microinjection, aerosol beam injection of plant cell protoplasts, or the DNA constructs can be introduced directly to plant tissue using biolistic methods, such as DNA particle bombardment (see also Fu et al., Transgenic Res. 2000 Feb;9(1):11-9). The basis of particle bombardment is the acceleration of particles coated with gene/s of interest toward cells, resulting in the penetration of the protoplasm by the particles and typically stable integration into the genome. (see e.g. Klein et al, Nature (1987), Klein et a1, Bio/Technology (1992), Casas et a1, Proc. Natl. Acad. Sci. USA (1993).).

[00455] In particular embodiments, the DNA constructs containing components of the CRISPR system may be introduced into the plant by *Agrobacterium*-mediated transformation. The DNA constructs may be combined with suitable T-DNA flanking regions and introduced into a conventional *Agrobacterium tumefaciens* host vector. The foreign DNA can be incorporated into the genome of plants by infecting the plants or by incubating plant protoplasts with Agrobacterium bacteria, containing one or more Ti (tumor-inducing) plasmids. (see e.g. Fraley et al., (1985), Rogers et al., (1987) and U.S. Pat. No. 5,563,055).

PLANT PROMOTERS

[00456] In order to ensure appropriate expression in a plant cell, the components of the CRISPR-Cas CRISPR system described herein are typically placed under control of a plant promoter, i.e. a promoter operable in plant cells. The use of different types of promoters is envisaged.

[00457] A constitutive plant promoter is a promoter that is able to express the open reading frame (ORF) that it controls in all or nearly all of the plant tissues during all or nearly all developmental stages of the plant (referred to as "constitutive expression"). One non-limiting example of a constitutive promoter is the cauliflower mosaic virus 35S promoter. The present invention envisages methods for modifying RNA sequences and as such also envisages regulating expression of plant biomolecules. In particular embodiments of the present invention it is thus advantageous to place one or more elements of the CRISPR system under the control of a promoter that can be regulated. “Regulated promoter" refers to promoters that direct gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and includes tissue-specific, tissue-preferred and inducible promoters. Different promoters may direct the expression of a gene in different tissues or cell types, or at different stages of development, or in response to different environmental conditions. In particular embodiments, one or more of the CRISPR components are expressed under the control of a constitutive promoter, such as the cauliflower mosaic virus 35S promoter issue-preferred promoters can be utilized to target enhanced expression in certain cell types within a particular plant tissue, for instance vascular cells in leaves or roots or in specific cells of the seed. Examples of particular promoters for use in the CRISPR system–are found in Kawamata et al., (1997) Plant Cell Physiol 38:792-803; Yamamoto et al., (1997) Plant J 12:255-65; Hire et al, (1992) Plant Mol Biol 20:207-18,Kuster et al, (1995) Plant Mol Biol 29:759-72, and Capana et al., (1994) Plant Mol Biol 25:681 -91.

[00458] Examples of promoters that are inducible and that allow for spatiotemporal control of gene editing or gene expression may use a form of energy. The form of energy may include but is not limited to sound energy, electromagnetic radiation, chemical energy and/or thermal energy. Examples of inducible systems include tetracycline inducible promoters (Tet-On or Tet-Off), small molecule two-hybrid transcription activations systems (FKBP, ABA, etc), or light inducible systems (Phytochrome, LOV domains, or cryptochrome)., such as a Light Inducible Transcriptional Effector (LITE) that direct changes in transcriptional activity in a sequence-specific manner. The components of a light inducible system may include a CRISPR-Cas, a light-responsive cytochrome heterodimer (e.g. from Arabidopsis thaliana), and a transcriptional activation/repression domain. Further examples of inducible DNA binding proteins and methods for their use are provided in US 61/736465 and US 61/721,283, which is hereby incorporated by reference in its entirety.

[00459] In particular embodiments, transient or inducible expression can be achieved by using, for example, chemical-regulated promotors, i.e. whereby the application of an

exogenous chemical induces gene expression. Modulating of gene expression can also be obtained by a chemical-repressible promoter, where application of the chemical represses gene expression. Chemical-inducible promoters include, but are not limited to, the maize ln2-2 promoter, activated by benzene sulfonamide herbicide safeners (De Veylder et al., (1997) Plant Cell Physiol 38:568-77), the maize GST promoter (GST-ll-27, WO93/01294), activated by hydrophobic electrophilic compounds used as pre-emergent herbicides, and the tobacco PR-1 a promoter (Ono et al., (2004) Biosci Biotechnol Biochem 68:803-7) activated by salicylic acid. Promoters which are regulated by antibiotics, such as tetracycline-inducible and tetracycline-repressible promoters (Gatz et al., (1991) Mol Gen Genet 227:229-37; U.S. Patent Nos. 5,814,618 and 5,789,156) can also be used herein.

TRANSLOCATION TO AND/OR EXPRESSION IN SPECIFIC PLANT ORGANELLES

[00460] The expression system may comprise elements for translocation to and/or expression in a specific plant organelle.

Chloroplast targeting

[00461] In particular embodiments, it is envisaged that the CRISPR system is used to specifically modify expression and/or translation of chloroplast genes or to ensure expression in the chloroplast. For this purpose use is made of chloroplast transformation methods or compartmentalization of the CRISPR components to the chloroplast. For instance, the introduction of genetic modifications in the plastid genome can reduce biosafety issues such as gene flow through pollen.

[00462] Methods of chloroplast transformation are known in the art and include Particle bombardment, PEG treatment, and microinjection. Additionally, methods involving the translocation of transformation cassettes from the nuclear genome to the plastid can be used as described in WO2010061186.

[00463] Alternatively, it is envisaged to target one or more of the CRISPR components to the plant chloroplast. This is achieved by incorporating in the expression construct a sequence encoding a chloroplast transit peptide (CTP) or plastid transit peptide, operably linked to the 5' region of the sequence encoding the protein. The CTP is removed in a processing step during translocation into the chloroplast. Chloroplast targeting of expressed proteins is well known to the skilled artisan (see for instance Protein Transport into Chloroplasts, 2010, Annual Review of Plant Biology, Vol. 61: 157-180) . In such embodiments it is also desired to target the one or more guide RNAs to the plant chloroplast. Methods and constructs which can be used for translocating guide RNA into the chloroplast by means of a chloroplast localization

sequence are described, for instance, in US 20040142476, incorporated herein by reference. Such variations of constructs can be incorporated into the expression systems of the invention to efficiently translocate the RNA targeting -guide RNA(s).

INTRODUCTION OF POLYNUCLEOTIDES ENCODING THE SYSTEMS IN ALGAL CELLS.

[00464] Transgenic algae (or other plants such as rape) may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol) or other products. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[00465] US 8945839 describes a method for engineering Micro-Algae (*Chlamydomonas reinhardtii* cells) species) using Cas. Using similar tools, the methods of the CRISPR system described herein can be applied on *Chlamydomonas* species and other algae. In particular embodiments, protein and guide RNA(s) are introduced in algae expressed using a vector that expresses protein under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2 -tubulin. Guide RNA is optionally delivered using a vector containing T7 promoter. Alternatively, mRNA and *in vitro* transcribed guide RNA can be delivered to algal cells. Electroporation protocols are available to the skilled person such as the standard recommended protocol from the GeneArt Chlamydomonas Engineering kit.

INTRODUCTION OF POLYNUCLEOTIDES ENCODING THE SYSTEMS IN YEAST CELLS

[00466] In particular embodiments, the invention relates to the use of the CRISPR system for RNA editing in yeast cells. Methods for transforming yeast cells which can be used to introduce polynucleotides encoding the CRISPR system components are well known to the artisan and are reviewed by Kawai et al., 2010, Bioeng Bugs. 2010 Nov-Dec; 1(6): 395–403). Non-limiting examples include transformation of yeast cells by lithium acetate treatment (which may further include carrier DNA and PEG treatment), bombardment or by electroporation.

TRANSIENT EXPRESSION OF SYSTEMS IN PLANTS AND PLANT CELL

[00467] In particular embodiments, it is envisaged that the guide RNA and/or gene are transiently expressed in the plant cell. In these embodiments, the CRISPR system can ensure modification of RNA target molecules only when both the guide RNA and the Cas protein is present in a cell, such that gene expression can further be controlled. As the expression of the Cas protein is transient, plants regenerated from such plant cells typically contain no foreign DNA. In particular embodiments the Cas protein is stably expressed by the plant cell and the guide sequence is transiently expressed.

[00468] In particularly preferred embodiments, the CRISPR system components can be introduced in the plant cells using a plant viral vector (Scholthof et al. 1996, Annu Rev Phytopathol. 1996;34:299-323). In further particular embodiments, said viral vector is a vector from a DNA virus. For example, geminivirus (e.g., cabbage leaf curl virus, bean yellow dwarf virus, wheat dwarf virus, tomato leaf curl virus, maize streak virus, tobacco leaf curl virus, or tomato golden mosaic virus) or nanovirus (e.g., Faba bean necrotic yellow virus). In other particular embodiments, said viral vector is a vector from an RNA virus. For example, tobravirus (e.g., tobacco rattle virus, tobacco mosaic virus), potexvirus (e.g., potato virus X), or hordeivirus (e.g., barley stripe mosaic virus). The replicating genomes of plant viruses are non-integrative vectors, which is of interest in the context of avoiding the production of GMO plants.

[00469] In particular embodiments, the vector used for transient expression of CRISPR constructs is for instance a pEAQ vector, which is tailored for Agrobacterium-mediated transient expression (Sainsbury F. et al., Plant Biotechnol J. 2009 Sep;7(7):682-93) in the protoplast. Precise targeting of genomic locations was demonstrated using a modified Cabbage Leaf Curl virus (CaLCuV) vector to express gRNAs in stable transgenic plants expressing a Cas (see Scientific Reports 5, Article number: 14926 (2015), doi:10.1038/srep14926).

[00470] In particular embodiments, double-stranded DNA fragments encoding the guide RNA or crRNA and/or the Cas gene can be transiently introduced into the plant cell. In such embodiments, the introduced double-stranded DNA fragments are provided in sufficient quantity to modify RNA molecule(s) in the cell but do not persist after a contemplated period of time has passed or after one or more cell divisions. Methods for direct DNA transfer in plants are known by the skilled artisan (see for instance Davey et al. Plant Mol Biol. 1989 Sep;13(3):273-85.)

[00471] In other embodiments, an RNA polynucleotide encoding the Cas protein is introduced into the plant cell, which is then translated and processed by the host cell generating the protein in sufficient quantity to modify the RNA molecule(s) cell (in the presence of at least one guide RNA) but which does not persist after a contemplated period of time has passed or after one or more cell divisions. Methods for introducing mRNA to plant protoplasts for transient expression are known by the skilled artisan (see for instance in Gallie, Plant Cell Reports (1993), 13;119-122). Combinations of the different methods described above are also envisaged.

DELIVERY OF CRISPR COMPONENTS TO THE PLANT CELL

[00472] In particular embodiments, it is of interest to deliver one or more components of the CRISPR system directly to the plant cell. This is of interest, inter alia, for the generation of non-transgenic plants. In particular embodiments, one or more of the components is prepared outside the plant or plant cell and delivered to the cell. For instance in particular embodiments, the Cas protein is prepared in vitro prior to introduction to the plant cell. Cas protein can be prepared by various methods known by one of skill in the art and include recombinant production. After expression, the Cas protein is isolated, refolded if needed, purified and optionally treated to remove any purification tags, such as a His-tag. Once crude, partially purified, or more completely purified Cas protein is obtained, the protein may be introduced to the plant cell.

[00473] In particular embodiments, the Cas protein is mixed with guide RNA targeting the nucleic acid of interest to form a pre-assembled ribonucleoprotein.

[00474] The individual components or pre-assembled ribonucleoprotein can be introduced into the plant cell via electroporation, by bombardment with nucleic acid targeting -associated gene product coated particles, by chemical transfection or by some other means of transport across a cell membrane. For instance, transfection of a plant protoplast with a pre-assembled CRISPR ribonucleoprotein has been demonstrated to ensure targeted modification of the plant genome (as described by Woo et al. Nature Biotechnology, 2015; DOI: 10.1038/nbt.3389). These methods can be modified to achieve targeted modification of RNA molecules in the plants.

[00475] In particular embodiments, the CRISPR system components are introduced into the plant cells using nanoparticles. The components, either as protein or nucleic acid or in a combination thereof, can be uploaded onto or packaged in nanoparticles and applied to the plants (such as for instance described in WO 2008042156 and US 20130185823). In particular, embodiments of the invention comprise nanoparticles uploaded with or packed with DNA molecule(s) encoding the Cas protein, DNA molecules encoding the guide RNA and/or isolated guide RNA as described in WO2015089419.

[00476] Further means of introducing one or more components of the CRISPR system to the plant cell is by using cell penetrating peptides (CPP). Accordingly, in particular, embodiments the invention comprises compositions comprising a cell penetrating peptide linked to an Cas protein. In particular embodiments of the present invention, an RNA targeting protein and/or guide RNA(s) is coupled to one or more CPPs to effectively transport them inside plant protoplasts (Ramakrishna (2014, Genome Res. 2014 Jun;24(6):1020-7 for Cas9 in human cells). In other embodiments, the Cas gene and/or guide RNA(s) are encoded by one or

more circular or non-circular DNA molecule(s) which are coupled to one or more CPPs for plant protoplast delivery. The plant protoplasts are then regenerated to plant cells and further to plants. CPPs are generally described as short peptides of fewer than 35 amino acids either derived from proteins or from chimeric sequences which are capable of transporting biomolecules across cell membrane in a receptor independent manner. CPP can be cationic peptides, peptides having hydrophobic sequences, amphipatic peptides, peptides having proline-rich and anti-microbial sequence, and chimeric or bipartite peptides (Pooga and Langel 2005). CPPs are able to penetrate biological membranes and as such trigger the movement of various biomolecules across cell membranes into the cytoplasm and to improve their intracellular routing, and hence facilitate interaction of the biolomolecule with the target. Examples of CPP include amongst others: Tat, a nuclear transcriptional activator protein required for viral replication by HIV type1, penetratin, Kaposi fibroblast growth factor (FGF) signal peptide sequence, integrin β3 signal peptide sequence; polyarginine peptide Args sequence, Guanine rich-molecular transporters, sweet arrow peptide, etc.

FURTHER APPLICATIONS IN PLANTS AND YEASTS

Biofuel production

[00477] The term "biofuel" as used herein is an alternative fuel made from plant and plant-derived resources. Renewable biofuels can be extracted from organic matter whose energy has been obtained through a process of carbon fixation or are made through the use or conversion of biomass. This biomass can be used directly for biofuels or can be converted to convenient energy containing substances by thermal conversion, chemical conversion, and biochemical conversion. This biomass conversion can result in fuel in solid, liquid, or gas form. There are two types of biofuels: bioethanol and biodiesel. Bioethanol is mainly produced by the sugar fermentation process of cellulose (starch), which is mostly derived from maize and sugar cane. Biodiesel on the other hand is mainly produced from oil crops such as rapeseed, palm, and soybean. Biofuels are used mainly for transportation.

Enhancing plant properties for biofuel production

[00478] In particular embodiments, the methods using the CRISPR system as described herein are used to alter the properties of the cell wall in order to facilitate access by key hydrolysing agents for a more efficient release of sugars for fermentation. In particular embodiments, the biosynthesis of cellulose and/or lignin are modified. Cellulose is the major component of the cell wall. The biosynthesis of cellulose and lignin are co-regulated. By reducing the proportion of lignin in a plant the proportion of cellulose can be increased. In

particular embodiments, the methods described herein are used to downregulate lignin biosynthesis in the plant so as to increase fermentable carbohydrates. More particularly, the methods described herein are used to downregulate at least a first lignin biosynthesis gene selected from the group consisting of 4-coumarate 3-hydroxylase (C3H), phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), hydroxycinnamoyl transferase (HCT), caffeic acid O-methyltransferase (COMT), caffeoyl CoA 3-O-methyltransferase (CCoAOMT), ferulate 5- hydroxylase (F5H), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl CoA-reductase (CCR), 4- coumarate-CoA ligase (4CL), monolignol-lignin-specific glycosyltransferase, and aldehyde dehydrogenase (ALDH) as disclosed in WO 2008064289 A2.

[00479] In particular embodiments, the methods described herein are used to produce plant mass that produces lower levels of acetic acid during fermentation (see also WO 2010096488).

Modifying yeast for Biofuel production

[00480] In particular embodiments, the Cas protein provided herein is used for bioethanol production by recombinant micro-organisms. For instance, Cas proteins can be used to engineer micro-organisms, such as yeast, to generate biofuel or biopolymers from fermentable sugars and optionally to be able to degrade plant-derived lignocellulose derived from agricultural waste as a source of fermentable sugars. More particularly, the invention provides methods whereby the CRISPR complex is used to modify the expression of endogenous genes required for biofuel production and/or to modify endogenous genes why may interfere with the biofuel synthesis. More particularly the methods involve stimulating the expression in a micro-organism such as a yeast of one or more nucleotide sequence encoding enzymes involved in the conversion of pyruvate to ethanol or another product of interest. In particular embodiments the methods ensure the stimulation of expression of one or more enzymes which allows the micro-organism to degrade cellulose, such as a cellulase. In yet further embodiments, the CRISPR complex is used to suppress endogenous metabolic pathways which compete with the biofuel production pathway.

<u>Modifying Algae and plants for production of vegetable oils or biofuels</u>

[00481] Transgenic algae or other plants such as rape may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol), for instance. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[00482] US 8945839 describes a method for engineering Micro-Algae (Chlamydomonas reinhardtii cells) species) using Cas. Using similar tools, the methods of the CRISPR system described herein can be applied on Chlamydomonas species and other algae. In particular embodiments, the Cas protein and guide RNA are introduced in algae expressed using a vector that expresses the Cas protein under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2 -tubulin. Guide RNA will be delivered using a vector containing T7 promoter. Alternatively, *in vitro* transcribed guide RNA can be delivered to algae cells. Electroporation protocol follows standard recommended protocol from the GeneArt Chlamydomonas Engineering kit.

Particular applications of the Cas proteins in plants

[00483] In particular embodiments, present invention can be used as a therapy for virus removal in plant systems as it is able to cleave viral RNA. Previous studies in human systems have demonstrated the success of utilizing CRISPR in targeting the single strand RNA virus, hepatitis C (A. Price, et al., Proc. Natl. Acad. Sci, 2015). These methods may also be adapted for using the CRISPR system in plants.

Improved plants

[00484] The present invention also provides plants and yeast cells obtainable and obtained by the methods provided herein. The improved plants obtained by the methods described herein may be useful in food or feed production through the modified expression of genes which, for instance ensure tolerance to plant pests, herbicides, drought, low or high temperatures, excessive water, etc.

[00485] The improved plants obtained by the methods described herein, especially crops and algae may be useful in food or feed production through expression of, for instance, higher protein, carbohydrate, nutrient or vitamin levels than would normally be seen in the wildtype. In this regard, improved plants, especially pulses and tubers are preferred.

[00486] Improved algae or other plants such as rape may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol), for instance. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[00487] The invention also provides for improved parts of a plant. Plant parts include, but are not limited to, leaves, stems, roots, tubers, seeds, endosperm, ovule, and pollen. Plant parts as envisaged herein may be viable, nonviable, regeneratable, and/or non- regeneratable.

[00488] It is also encompassed herein to provide plant cells and plants generated according to the methods of the invention. Gametes, seeds, embryos, either zygotic or somatic, progeny or hybrids of plants comprising the genetic modification, which are produced by traditional breeding methods, are also included within the scope of the present invention. Such plants may contain a heterologous or foreign DNA sequence inserted at or instead of a target sequence. Alternatively, such plants may contain only an alteration (mutation, deletion, insertion, substitution) in one or more nucleotides. As such, such plants will only be different from their progenitor plants by the presence of the particular modification.

[00489] In an embodiment of the invention, a CRISPR-Cas system is used to engineer pathogen resistant plants, for example by creating resistance against diseases caused by bacteria, fungi or viruses. In certain embodiments, pathogen resistance can be accomplished by engineering crops to produce a CRISPR-Cas system that will be ingested by an insect pest, leading to mortality. In an embodiment of the invention, a CRISPR-Cas system is used to engineer abiotic stress tolerance. In another embodiment, a CRISPR-Cas system is used to engineer drought stress tolerance or salt stress tolerance, or cold or heat stress tolerance. Younis et al. 2014, Int. J. Biol. Sci. 10;1150 reviewed potential targets of plant breeding methods, all of which are amenable to correction or improvement through use of a CRISPR-Cas system described herein. Some non-limiting target crops include Arabidops Zea mays is thaliana, Oryza sativa L, Prunus domestica L., Gossypium hirsutum, Nicotiana rustica, Zea mays, Medicago sativa, Nicotiana benthamiana and Arabidopsis thaliana

[00490] In an embodiment of the invention, a CRISPR-Cas system is used for management of crop pests. For example, a CRISPR-Cas system operable in a crop pest can be expressed from a plant host or transferred directly to the target, for example using a viral vector.

[00491] In an embodiment, the invention provides a method of efficiently producing homozygous organisms from a heterozygous non-human starting organism. In an embodiment, the invention is used in plant breeding. In another embodiment, the invention is used in animal breeding. In such embodiments, a homozygous organism such as a plant or animal is made by preventing or suppressing recombination by interfering with at least one target gene involved in double strand breaks, chromosome pairing and/or strand exchange.

[00492] The invention in some embodiments comprehends a method of modifying an cell or organism. The cell may be a prokaryotic cell or a eukaryotic cell. The cell may be a mammalian cell. The mammalian cell many be a non-human primate, bovine, porcine, rodent or mouse cell. The cell may be a non-mammalian eukaryotic cell such as poultry, fish or shrimp. The cell may also be a plant cell. The plant cell may be of a crop plant such as cassava,

corn, sorghum, wheat, or rice. The plant cell may also be of an algae, tree or vegetable. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell are altered for improved production of biologic products such as an antibody, starch, alcohol or other desired cellular output. The modification introduced to the cell by the present invention may be such that the cell and progeny of the cell include an alteration that changes the biologic product produced. The system may comprise one or more different vectors. In an aspect of the invention, the effector protein is codon optimized for expression the desired cell type, preferentially a eukaryotic cell, preferably a mammalian cell or a human cell. CRISPR-Cas system(s) (e.g., single or multiplexed) can be used in conjunction with recent advances in crop genomics. Such CRISPR system(s) can be used to perform efficient and cost effective plant gene or genome or transcriptome interrogation or editing or manipulation—for instance, for rapid investigation and/or selection and/or interrogations and/or comparison and/or manipulations and/or transformation of plant genes or genomes; e.g., to create, identify, develop, optimize, or confer trait(s) or characteristic(s) to plant(s) or to transform a plant genome. There can accordingly be improved production of plants, new plants with new combinations of traits or characteristics or new plants with enhanced traits. Such CRISPR system(s) can be used with regard to plants in Site-Directed Integration (SDI) or Gene Editing (GE) or any Near Reverse Breeding (NRB) or Reverse Breeding (RB) techniques. Accordingly, reference herein to animal cells may also apply, mutatis mutandis, to plant cells unless otherwise apparent; and, the enzymes herein having reduced off-target effects and systems employing such enzymes can be used in plant applications, including those mentioned herein. Engineered plants modified by the effector protein and suitable guide (crRNA), and progeny thereof, as provided. These may include disease or drought resistant crops, such as wheat, barley, rice, soybean or corn; plants modified to remove or reduce the ability to self-pollinate (but which can instead, optionally, hybridise instead); and allergenic foods such as peanuts and nuts where the immunogenic proteins have been disabled, destroyed or disrupted by targeting via a effector protein and suitable guide. Any aspect of using classical CRISPR-Cas systems may be adapted to use in CRISPR systems that are Cas protein agnostic.

MODELS OF CONDITIONS

[00493] A method of the invention may be used to create a plant, an animal or cell that may be used to model and/or study genetic or epigenetic conditions of interest, such as a through a model of mutations of interest or a disease model. As used herein, "disease" refers to a disease, disorder, or indication in a subject. For example, a method of the invention may be used to create an animal or cell that comprises a modification in one or more nucleic acid sequences

associated with a disease, or a plant, animal or cell in which expression of one or more nucleic acid sequences associated with a disease are altered. Such a nucleic acid sequence may encode or be translated a disease associated protein sequence or may be a disease associated control sequence. Accordingly, it is understood that in embodiments of the invention, a plant, subject, patient, organism or cell can be a non-human subject, patient, organism or cell. Thus, the invention provides a plant, animal or cell, produced by the present methods, or a progeny thereof. The progeny may be a clone of the produced plant or animal, or may result from sexual reproduction by crossing with other individuals of the same species to introgress further desirable traits into their offspring. The cell may be in vivo or ex vivo in the cases of multicellular organisms, particularly animals or plants. In the instance where the cell is in cultured, a cell line may be established if appropriate culturing conditions are met and preferably if the cell is suitably adapted for this purpose (for instance a stem cell). Bacterial cell lines produced by the invention are also envisaged. Hence, cell lines are also envisaged. In some methods, the disease model can be used to study the effects of mutations, or more general altered, such as reduced, expression of genes or gene products on the animal or cell and development and/or progression of the disease using measures commonly used in the study of the disease. Alternatively, such a disease model is useful for studying the effect of a pharmaceutically active compound on the disease. In some methods, the disease model can be used to assess the efficacy of a potential gene therapy strategy. That is, a disease-associated RNA can be modified such that the disease development and/or progression is displayed or inhibited or reduced and then effects of a compound on the progression or inhibition or reduction are tested.

[00494] Useful in the practice of the instant invention utilizing CRISPR-Cas proteins and complexes thereof and nucleic acid molecules encoding same and methods using same, reference is made to: Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. Science Dec 12. (2013); Published in final edited form as: Science. 2014 Jan 3; 343(6166): 84–87. Shalem et al. involves 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. Reference is also made to US patent publication number US20140357530; and PCT Patent Publication WO2014093701, hereby incorporated herein by reference.

[00495] The term "associated with" is used here in relation to the association of the functional domain to the CRISPR-Cas protein or the adaptor protein. It is used in respect of how one molecule 'associates' with respect to another, for example between an adaptor protein and a functional domain, or between the CRISPR-Cas protein and a functional domain. In the case of such protein-protein interactions, this association may be viewed in terms of recognition in the way an antibody recognizes an epitope. Alternatively, one protein may be associated with another protein via a fusion of the two, for instance one subunit being fused to another subunit. Fusion typically occurs by addition of the amino acid sequence of one to that of the other, for instance via splicing together of the nucleotide sequences that encode each protein or subunit. Alternatively, this may essentially be viewed as binding between two molecules or direct linkage, such as a fusion protein. In any event, the fusion protein may include a linker between the two subunits of interest (i.e. between the enzyme and the functional domain or between the adaptor protein and the functional domain). Thus, in some embodiments, the CRISPR-Cas protein or adaptor protein is associated with a functional domain by binding thereto. In other embodiments, the CRISPR-Cas protein or adaptor protein is associated with a functional domain because the two are fused together, optionally via an intermediate linker.

APPLICATIONS IN NON-HUMAN ANIMALS

[00496] In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example a mammal. Also, the organism may be an arthropod such as an insect. The present invention may also be extended to other agricultural applications such as, for example, farm and production animals. For example, pigs have many features that make them attractive as biomedical models, especially in regenerative medicine. In particular, pigs with severe combined immunodeficiency (SCID) may provide useful models for regenerative

medicine, xenotransplantation (discussed also elsewhere herein), and tumor development and will aid in developing therapies for human SCID patients. Lee et al., (Proc Natl Acad Sci U S A. 2014 May 20;111(20):7260-5) utilized a reporter-guided transcription activator-like effector nuclease (TALEN) system to generated targeted modifications of recombination activating gene (RAG) 2 in somatic cells at high efficiency, including some that affected both alleles.

[00497] The methods of Lee et al., (Proc Natl Acad Sci U S A. 2014 May 20;111(20):7260-5) may be applied to the present invention analogously as follows. Mutated pigs are produced by targeted modification of RAG2 in fetal fibroblast cells followed by SCNT and embryo transfer. Constructs coding for CRISPR Cas and a reporter are electroporated into fetal-derived fibroblast cells. After 48 h, transfected cells expressing the green fluorescent protein are sorted into individual wells of a 96-well plate at an estimated dilution of a single cell per well. Targeted modification of RAG2 are screened by amplifying a genomic DNA fragment flanking any CRISPR Cas cutting sites followed by sequencing the PCR products. After screening and ensuring lack of off-site mutations, cells carrying targeted modification of RAG2 are used for SCNT. The polar body, along with a portion of the adjacent cytoplasm of oocyte, presumably containing the metaphase II plate, are removed, and a donor cell are placed in the perivitelline. The reconstructed embryos are then electrically porated to fuse the donor cell with the oocyte and then chemically activated. The activated embryos are incubated in Porcine Zygote Medium 3 (PZM3) with 0.5 μM Scriptaid (S7817; Sigma-Aldrich) for 14–16 h. Embryos are then washed to remove the Scriptaid and cultured in PZM3 until they were transferred into the oviducts of surrogate pigs.

[00498] The present invention is also applicable to modifying SNPs of other animals, such as cows. Tan et al. (Proc Natl Acad Sci U S A. 2013 Oct 8; 110(41): 16526–16531) expanded the livestock gene editing toolbox to include transcription activator-like (TAL) effector nuclease (TALEN)- and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9- stimulated homology-directed repair (HDR) using plasmid, rAAV, and oligonucleotide templates. Gene specific gRNA sequences were cloned into the Church lab gRNA vector (Addgene ID: 41824) according to their methods (Mali P, et al. (2013) RNA-Guided Human Genome Engineering via Cas9. Science 339(6121):823-826). The Cas9 nuclease was provided either by co-transfection of the hCas9 plasmid (Addgene ID: 41815) or mRNA synthesized from RCIScript-hCas9. This RCIScript-hCas9 was constructed by sub-cloning the XbaI-AgeI fragment from the hCas9 plasmid (encompassing the hCas9 cDNA) into the RCIScript plasmid.

[00499] Heo et al. (Stem Cells Dev. 2015 Feb 1;24(3):393-402. doi: 10.1089/scd.2014.0278. Epub 2014 Nov 3) reported highly efficient gene targeting in the bovine genome using bovine pluripotent cells and clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 nuclease. First, Heo et al. generate induced pluripotent stem cells (iPSCs) from bovine somatic fibroblasts by the ectopic expression of yamanaka factors and GSK3β and MEK inhibitor (2i) treatment. Heo et al. observed that these bovine iPSCs are highly similar to naïve pluripotent stem cells with regard to gene expression and developmental potential in teratomas. Moreover, CRISPR-Cas nuclease, which was specific for the bovine NANOG locus, showed highly efficient editing of the bovine genome in bovine iPSCs and embryos.

[00500] Igenity® provides a profile analysis of animals, such as cows, to perform and transmit traits of economic traits of economic importance, such as carcass composition, carcass quality, maternal and reproductive traits and average daily gain. The analysis of a comprehensive Igenity® profile begins with the discovery of DNA markers (most often single nucleotide polymorphisms or SNPs). All the markers behind the Igenity® profile were discovered by independent scientists at research institutions, including universities, research organizations, and government entities such as USDA. Markers are then analyzed at Igenity® in validation populations. Igenity® uses multiple resource populations that represent various production environments and biological types, often working with industry partners from the seedstock, cow-calf, feedlot and/or packing segments of the beef industry to collect phenotypes that are not commonly available. Cattle genome databases are widely available, see, e.g., the NAGRP Cattle Genome Coordination Program (www.animalgenome.org/cattle/maps/db.html). Thus, the present invention maybe applied to target bovine SNPs. One of skill in the art may utilize the above protocols for targeting SNPs and apply them to bovine SNPs as described, for example, by Tan et al. or Heo et al.

[00501] Qingjian Zou et al. (Journal of Molecular Cell Biology Advance Access published October 12, 2015) demonstrated increased muscle mass in dogs by targeting targeting the first exon of the dog Myostatin (MSTN) gene (a negative regulator of skeletal muscle mass). First, the efficiency of the sgRNA was validated, using cotransfection of the the sgRNA targeting MSTN with a Cas vector into canine embryonic fibroblasts (CEFs). Thereafter, MSTN KO dogs were generated by micro-injecting embryos with normal morphology with a mixture of Cas mRNA and MSTN sgRNA and auto-transplantation of the zygotes into the oviduct of the same female dog. The knock-out puppies displayed an obvious muscular phenotype on thighs compared with its wild-type littermate sister.

LIVESTOCK - PIGS

[00502] Viral targets in livestock may include, in some embodiments, porcine CD163, for example on porcine macrophages. CD163 is associated with infection (thought to be through viral cell entry) by PRRSv (Porcine Reproductive and Respiratory Syndrome virus, an arterivirus). Infection by PRRSv, especially of porcine alveolar macrophages (found in the lung), results in a previously incurable porcine syndrome ("Mystery swine disease" or "blue ear disease") that causes suffering, including reproductive failure, weight loss and high mortality rates in domestic pigs. Opportunistic infections, such as enzootic pneumonia, meningitis and ear oedema, are often seen due to immune deficiency through loss of macrophage activity. It also has significant economic and environmental repercussions due to increased antibiotic use and financial loss (an estimated $660m per year).

[00503] As reported by Kristin M Whitworth and Dr Randall Prather et al. (Nature Biotech 3434 published online 07 December 2015) at the University of Missouri and in collaboration with Genus Plc, CD163 was targeted using CRISPR-Cas and the offspring of edited pigs were resistant when exposed to PRRSv. One founder male and one founder female, both of whom had mutations in exon 7 of CD163, were bred to produce offspring. The founder male possessed an 11-bp deletion in exon 7 on one allele, which results in a frameshift mutation and missense translation at amino acid 45 in domain 5 and a subsequent premature stop codon at amino acid 64. The other allele had a 2-bp addition in exon 7 and a 377-bp deletion in the preceding intron, which were predicted to result in the expression of the first 49 amino acids of domain 5, followed by a premature stop code at amino acid 85. The sow had a 7 bp addition in one allele that when translated was predicted to express the first 48 amino acids of domain 5, followed by a premature stop codon at amino acid 70. The sow's other allele was unamplifiable. Selected offspring were predicted to be a null animal (CD163–/–), i.e. a CD163 knock out.

[00504] Accordingly, in some embodiments, porcine alveolar macrophages may be targeted by the CRISPR protein. In some embodiments, porcine CD163 may be targeted by the CRISPR protein. In some embodiments, porcine CD163 may be knocked out through induction of a DSB or through insertions or deletions, for example targeting deletion or modification of exon 7, including one or more of those described above, or in other regions of the gene, for example deletion or modification of exon 5.

[00505] An edited pig and its progeny are also envisaged, for example a CD163 knock out pig. This may be for livestock, breeding or modelling purposes (i.e. a porcine model). Semen comprising the gene knock out is also provided.

[00506] CD163 is a member of the scavenger receptor cysteine-rich (SRCR) superfamily. Based on in vitro studies SRCR domain 5 of the protein is the domain responsible for unpackaging and release of the viral genome. As such, other members of the SRCR superfamily may also be targeted in order to assess resistance to other viruses. PRRSV is also a member of the mammalian arterivirus group, which also includes murine lactate dehydrogenase-elevating virus, simian hemorrhagic fever virus and equine arteritis virus. The arteriviruses share important pathogenesis properties, including macrophage tropism and the capacity to cause both severe disease and persistent infection. Accordingly, arteriviruses, and in particular murine lactate dehydrogenase-elevating virus, simian hemorrhagic fever virus and equine arteritis virus, may be targeted, for example through porcine CD163 or homologues thereof in other species, and murine, simian and equine models and knockout also provided.

[00507] Indeed, this approach may be extended to viruses or bacteria that cause other livestock diseases that may be transmitted to humans, such as Swine Influenza Virus (SIV) strains which include influenza C and the subtypes of influenza A known as H1N1, H1N2, H2N1, H3N1, H3N2, and H2N3, as well as pneumonia, meningitis and oedema mentioned above.

[00508] The methods for genome editing using the Cas system as described herein can be used to confer desired traits on essentially any plant, algae, fungus, yeast, etc. A wide variety of plants, algae, fungus, yeast, etc and plant algae, fungus, yeast cell or tissue systems may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation methods mentioned above.

[00509] In particular embodiments, the methods described herein are used to modify endogenous genes or to modify their expression without the permanent introduction into the genome of the plant, algae, fungus, yeast, etc of any foreign gene, including those encoding CRISPR components, so as to avoid the presence of foreign DNA in the genome of the plant. This can be of interest as the regulatory requirements for non-transgenic plants are less rigorous.

[00510] The CRISPR systems provided herein can be used to introduce targeted double-strand or single-strand breaks and/or to introduce gene activator and or repressor systems and without being limitative, can be used for gene targeting, gene replacement, targeted mutagenesis, targeted deletions or insertions, targeted inversions and/or targeted translocations. By co-expression of multiple targeting RNAs directed to achieve multiple modifications in a single cell, multiplexed genome modification can be ensured. This

technology can be used to high-precision engineering of plants with improved characteristics, including enhanced nutritional quality, increased resistance to diseases and resistance to biotic and abiotic stress, and increased production of commercially valuable plant products or heterologous compounds.

[00511] The methods described herein generally result in the generation of "improved plants, algae, fungi, yeast, etc" in that they have one or more desirable traits compared to the wildtype plant. In particular embodiments, the plants, algae, fungi, yeast, etc., cells or parts obtained are transgenic plants, comprising an exogenous DNA sequence incorporated into the genome of all or part of the cells. In particular embodiments, non-transgenic genetically modified plants, algae, fungi, yeast, etc., parts or cells are obtained, in that no exogenous DNA sequence is incorporated into the genome of any of the cells of the plant. In such embodiments, the improved plants, algae, fungi, yeast, etc. are non-transgenic. Where only the modification of an endogenous gene is ensured and no foreign genes are introduced or maintained in the plant, algae, fungi, yeast, etc. genome, the resulting genetically modified crops contain no foreign genes and can thus basically be considered non-transgenic. The different applications of the CRISPR-Cas system for plant, algae, fungi, yeast, etc. genome editing include, but are not limited to: introduction of one or more foreign genes to confer an agricultural trait of interest; editing of endogenous genes to confer an agricultural trait of interest; modulating of endogenous genes by the CRISPR-Cas system to confer an agricultural trait of interest. Exemplary genes conferring agronomic traits include, but are not limited to genes that confer resistance to pests or diseases; genes involved in plant diseases, such as those listed in WO 2013046247; genes that confer resistance to herbicides, fungicides, or the like; genes involved in (abiotic) stress tolerance. Other aspects of the use of the CRISPR-Cas system include, but are not limited to: create (male) sterile plants; increasing the fertility stage in plants/algae etc.; generate genetic variation in a crop of interest; affect fruit-ripening; increasing storage life of plants/algae etc.; reducing allergen in plants/algae etc.; ensure a value added trait (e.g. nutritional improvement); Screening methods for endogenous genes of interest; biofuel, fatty acid, organic acid, etc. production.

THERAPEUTIC APPLICATIONS

[00512] The systems of the invention can be applied in various therapeutic applications, e.g., in areas of former RNA cutting technologies, without undue experimentation, from this disclosure, including therapeutic, assay and other applications, because the present application provides the foundation for informed engineering of the system. The present invention provides for therapeutic treatment of a disease caused by overexpression of nucleic acids, toxic nucleic

acids and/or mutated nucleic acids (such as, for example, splicing defects or truncations). Expression of the toxic RNA may be associated with formation of nuclear inclusions and late-onset degenerative changes in brain, heart or skeletal muscle. In the best studied example, myotonic dystrophy, it appears that the main pathogenic effect of the toxic RNA is to sequester binding proteins and compromise the regulation of alternative splicing (Hum. Mol. Genet. (2006) 15 (suppl 2): R162-R169). Myotonic dystrophy [dystrophia myotonica (DM)] is of particular interest to geneticists because it produces an extremely wide range of clinical features. A partial listing would include muscle wasting, cataracts, insulin resistance, testicular atrophy, slowing of cardiac conduction, cutaneous tumors and effects on cognition. The classical form of DM, which is now called DM type 1 (DM1), is caused by an expansion of CTG repeats in the 3′-untranslated region (UTR) of DMPK, a gene encoding a cytosolic protein kinase.

[00513] In some examples, the disease is caused by a G→A or C→T point mutation or a pathogenic SNP. In some examples, the disease caused by a T→C or A→G point mutation or a pathogenic SNP. For example, the disease may be cancer, haemophilia, beta-thalassemia, Marfan syndrome and Wiskott-Aldrich syndrome.

EXEMPLARY THERAPIES

[00514] The present invention also contemplates use of the CRISPR-Cas system and the base editor described herein, for treatment in a variety of diseases and disorders. In some embodiments, the invention described herein relates to a method for therapy in which cells are edited ex vivo by CRISPR or the base editor to modulate at least one gene, with subsequent administration of the edited cells to a patient in need thereof. In some embodiments, the editing involves knocking in, knocking out or knocking down expression of at least one target gene in a cell. In particular embodiments, the editing inserts an exogenous, gene, minigene or sequence, which may comprise one or more exons and introns or natural or synthetic introns into the locus of a target gene, a hot-spot locus, a safe harbor locus of the gene genomic locations where new genes or genetic elements can be introduced without disrupting the expression or regulation of adjacent genes, or correction by insertions or deletions one or more mutations in DNA sequences that encode regulatory elements of a target gene. In some embodiment, the editing comprise introducing one or more point mutations in a nucleic acid (e.g., a genomic DNA) in a target cell.

[00515] In embodiments, the treatment is for disease/disorder of an organ, including liver disease, eye disease, muscle disease, heart disease, blood disease, brain disease, kidney disease, or may comprise treatment for an autoimmune disease, central nervous system disease, cancer

and other proliferative diseases, neurodegenerative disorders, inflammatory disease, metabolic disorder, musculoskeletal disorder and the like.

[00516] Particular diseases/disorders include chondroplasia, achromatopsia, acid maltase deficiency, adrenoleukodystrophy, aicardi syndrome, alpha- 1 antitrypsin deficiency, alpha-thalassemia, androgen insensitivity syndrome, apert syndrome, arrhythmogenic right ventricular, dysplasia, ataxia telangictasia, barth syndrome, beta-thalassemia, blue rubber bleb nevus syndrome, canavan disease, chronic granulomatous diseases (CGD), cri du chat syndrome, cystic fibrosis, dercum's disease, ectodermal dysplasia, fanconi anemia, fibrodysplasia ossificans progressive, fragile X syndrome, galactosemis, Gaucher's disease, generalized gangliosidoses (e.g., GM1), hemochromatosis, the hemoglobin C mutation in the 6th codon of beta-globin (HbC), hemophilia, Huntington's disease, Hurler Syndrome, hypophosphatasia, Klinefleter syndrome, Krabbes Disease, Langer-Giedion Syndrome, leukodystrophy, long QT syndrome, Marfan syndrome, Moebius syndrome, mucopolysaccharidosis (MPS), nail patella syndrome, nephrogenic diabetes insipdius, neurofibromatosis, Neimann-Pick disease, osteogenesis imperfecta, porphyria, Prader- Willi syndrome, progeria, Proteus syndrome, retinoblastoma, Rett syndrome, Rubinstein-Taybi syndrome, Sanfilippo syndrome, severe combined immunodeficiency (SCID), Shwachman syndrome, sickle cell disease (sickle cell anemia), Smith-Magenis syndrome, Stickler syndrome, Tay-Sachs disease, Thrombocytopenia Absent Radius (TAR) syndrome, Treacher Collins syndrome, trisomy, tuberous sclerosis, Turner's syndrome, urea cycle disorder, von Hippel- Landau disease, Waardenburg syndrome, Williams syndrome, Wilson's disease, and Wiskott- Aldrich syndrome.

[00517] In embodiments, the disease is associated with expression of a tumor antigen, e.g., a proliferative disease, a precancerous condition, a cancer, or a non-cancer related indication associated with expression of the tumor antigen, which may in some embodiments comprise a target selected from B2M, CD247, CD3D, CD3E, CD3G, TRAC, TRBC1, TRBC2, HLA-A, HLA-B, HLA-C, DCK, CD52, FKBP1A, CIITA, NLRC5, RFXANK, RFX5, RFXAP, or NR3C1, HAVCR2, LAG3, PDCD1, PD-L2, CTLA4, CEACAM (CEACAM-1, CEACAM-3 and/or CEACAM-5), VISTA, BTLA, TIGIT, LAIR1, CD160, 2B4, CD80, CD86, B7-H3 (CD113), B7-H4 (VTCN1), HVEM (TNFRSF14 or CD107), KIR, A2aR, MHC class I, MHC class II, GAL9, adenosine, and TGF beta, or PTPN11 DCK, CD52, NR3C1, LILRB1, CD19; CD123; CD22; CD30; CD171; CS-1 (also referred to as CD2 subset 1, CRACC, SLAMF7, CD319, and 19A24); C-type lectin-like molecule-1 (CLL-1 or CLECL1); CD33; epidermal growth factor receptor variant III (EGFRvIII); ganglioside G2 (GD2); ganglioside GD3

(aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-1)Cer); TNF receptor family member B cell maturation (BCMA); Tn antigen ((Tn Ag) or (GalNAca-Ser/Thr)); prostate-specific membrane antigen (PSMA); Receptor tyrosine kinase-like orphan receptor 1 (ROR1); Fms-Like Tyrosine Kinase 3 (FLT3); Tumor-associated glycoprotein 72 (TAG72); CD38; CD44v6; Carcinoembryonic antigen (CEA); Epithelial cell adhesion molecule (EPCAM); B7H3 (CD276); KIT (CD117); Interleukin-13 receptor subunit alpha-2 (IL-13Ra2 or CD213A2); Mesothelin; Interleukin 11 receptor alpha (IL-11Ra); prostate stem cell antigen (PSCA); Protease Serine 21 (Testisin or PRSS21); vascular endothelial growth factor receptor 2 (VEGFR2); Lewis(Y) antigen; CD24; Platelet-derived growth factor receptor beta (PDGFR-beta); Stage-specific embryonic antigen-4 (SSEA-4); CD20; Folate receptor alpha; Receptor tyrosine-protein kinase ERBB2 (Her2/neu); n kinase ERBB2 (Her2/neu); Mucin 1, cell surface associated (MUC1); epidermal growth factor receptor (EGFR); neural cell adhesion molecule (NCAM); Prostase; prostatic acid phosphatase (PAP); elongation factor 2 mutated (ELF2M); Ephrin B2; fibroblast activation protein alpha (FAP); insulin-like growth factor 1 receptor (IGF-I receptor), carbonic anhydrase IX (CAIX); Proteasome (Prosome, Macropain) Subunit, Beta Type, 9 (LMP2); glycoprotein 100 (gp100); oncogene fusion protein consisting of breakpoint cluster region (BCR) and Abelson murine leukemia viral oncogene homolog 1 (Abl) (bcr-abl); tyrosinase; ephrin type-A receptor 2 (EphA2); Fucosyl GM1; sialyl Lewis adhesion molecule (sLe); ganglioside GM3 (aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-1)Cer); transglutaminase 5 (TGS5); high molecular weight-melanoma-associated antigen (HMWMAA); o-acetyl-GD2 ganglioside (OAcGD2); Folate receptor beta; tumor endothelial marker 1 (TEM1/CD248); tumor endothelial marker 7-related (TEM7R); claudin 6 (CLDN6); thyroid stimulating hormone receptor (TSHR); G protein-coupled receptor class C group 5, member D (GPRC5D); chromosome X open reading frame 61 (CXORF61); CD97; CD179a; anaplastic lymphoma kinase (ALK); Polysialic acid; placenta-specific 1 (PLAC1); hexasaccharide portion of globoH glycoceramide (GloboH); mammary gland differentiation antigen (NY-BR-1); uroplakin 2 (UPK2); Hepatitis A virus cellular receptor 1 (HAVCR1); adrenoceptor beta 3 (ADRB3); pannexin 3 (PANX3); G protein-coupled receptor 20 (GPR20); lymphocyte antigen 6 complex, locus K 9 (LY6K); Olfactory receptor 51E2 (OR51E2); TCR Gamma Alternate Reading Frame Protein (TARP); Wilms tumor protein (WT1); Cancer/testis antigen 1 (NY-ESO-1); Cancer/testis antigen 2 (LAGE-1a); Melanoma-associated antigen 1 (MAGE-A1); ETS translocation-variant gene 6, located on chromosome 12p (ETV6-AML); sperm protein 17 (SPA17); X Antigen Family, Member 1A (XAGE1); angiopoietin-binding cell surface receptor 2 (Tie 2); melanoma cancer testis antigen-1 (MAD-CT-1); melanoma

cancer testis antigen-2 (MAD-CT-2); Fos-related antigen 1; tumor protein p53 (p53); p53 mutant; prostein; surviving; telomerase; prostate carcinoma tumor antigen-1 (PCTA-1 or Galectin 8), melanoma antigen recognized by T cells 1 (MelanA or MART1); Rat sarcoma (Ras) mutant; human Telomerase reverse transcriptase (hTERT); sarcoma translocation breakpoints; melanoma inhibitor of apoptosis (ML-IAP); ERG (transmembrane protease, serine 2 (TMPRSS2) ETS fusion gene); N-Acetyl glucosaminyl-transferase V (NA17); paired box protein Pax-3 (PAX3); Androgen receptor; Cyclin B1; v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN); Ras Homolog Family Member C (RhoC); Tyrosinase-related protein 2 (TRP-2); Cytochrome P450 1B1 (CYP1B1); CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS or Brother of the Regulator of Imprinted Sites), Squamous Cell Carcinoma Antigen Recognized By T Cells 3 (SART3); Paired box protein Pax-5 (PAX5); proacrosin binding protein sp32 (OY-TES1); lymphocyte-specific protein tyrosine kinase (LCK); A kinase anchor protein 4 (AKAP-4); synovial sarcoma, X breakpoint 2 (SSX2); Receptor for Advanced Glycation Endproducts (RAGE-1); renal ubiquitous 1 (RU1); renal ubiquitous 2 (RU2); legumain; human papilloma virus E6 (HPV E6); human papilloma virus E7 (HPV E7); intestinal carboxyl esterase; heat shock protein 70-2 mutated (mut hsp70-2); CD79a; CD79b; CD72; Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1); Fc fragment of IgA receptor (FCAR or CD89); Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2); CD300 molecule-like family member f (CD300LF); C-type lectin domain family 12 member A (CLEC12A); bone marrow stromal cell antigen 2 (BST2); EGF-like module-containing mucin-like hormone receptor-like 2 (EMR2); lymphocyte antigen 75 (LY75); Glypican-3 (GPC3); Fc receptor-like 5 (FCRLS); and immunoglobulin lambda-like polypeptide 1 (IGLL1), CD19, BCMA, CD70, G6PC, Dystrophin, including modification of exon 51 by deletion or excision, DMPK, CFTR (cystic fibrosis transmembrane conductance regulator). In embodiments, the targets comprise CD70, or a Knock-in of CD33 and Knock-out of B2M. In embodiments, the targets comprise a knockout of TRAC and B2M, or TRAC B2M and PD1, with or without additional target genes. In certain embodiments, the disease is cystic fibrosis with targeting of the SCNN1A gene, e.g., the non-coding or coding regions, e.g., a promoter region, or a transcribed sequence, e.g., intronic or exonic sequence, targeted knock-in at CFTR sequence within intron 2, into which, e.g., can be introduced CFTR sequence that codes for CFTR exons 3-27; and sequence within CFTR intron 10, into which sequence that codes for CFTR exons 11-27 can be introduced.

[00518] In embodiments, the disease is Metachromatic Leukodystrophy, and the target is Arylsulfatase A, the disease is Wiskott-Aldrich Syndrome and the target is Wiskott-Aldrich

Syndrome protein, the disease is Adreno leukodystrophy and the target is ATP-binding cassette DI, the disease is Human Immunodeficiency Virus and the target is receptor type 5- C-C chemokine or CXCR4 gene, the disease is Beta-thalassemia and the target is Hemoglobin beta subunit, the disease is X-linked Severe Combined ID receptor subunit gamma and the target is interelukin-2 receptor subunit gamma, the disease is Multisystemic Lysosomal Storage Disorder cystinosis and the target is cystinosin, the disease is Diamon-Blackfan anemia and the target is Ribosomal protein S19, the disease is Fanconi Anemia and the target is Fanconi anemia complementation groups (e.g. FNACA, FNACB, FANCC, FANCD1, FANCD2, FANCE, FANCF, RAD51C), the disease is Shwachman-Bodian-Diamond Bodian-Diamond syndrome and the target is Shwachman syndrome gene, the disease is Gaucher's disease and the target is Glucocerebrosidase, the disease is Hemophilia A and the target is Anti-hemophiliac factor OR Factor VIII, Christmas factor, Serine protease, Factor Hemophilia B IX, the disease is Adenosine deaminase deficiency (ADA-SCID) and the target is Adenosine deaminase, the disease is GM1 gangliosidoses and the target is beta-galactosidase, the disease is Glycogen storage disease type II, Pompe disease, the disease is acid maltase deficiency acid and the target is alpha-glucosidase, the disease is Niemann-Pick disease, SMPDl -associated (Types Sphingomyelin phosphodiesterase 1 OR A and B) acid and the target is sphingomyelinase, the disease is Krabbe disease, globoid cell leukodystrophy and the target is Galactosylceramidase or galactosylceramide lipidosis and the target is galactercerebrosidease, Human leukocyte antigens DR-15, DQ-6, the disease is Multiple Sclerosis (MS) DRB1, the disease is Herpes Simplex Virus 1 or 2 and the target is knocking down of one, two or three of RS1, RL2 and/or LAT genes. In embodiments, the disease is an HPV associated cancer with treatment including edited cells comprising binding molecules, such as TCRs or antigen binding fragments thereof and antibodies and antigen-binding fragments thereof, such as those that recognize or bind human papilloma virus. The disease can be Hepatitis B with a target of one or more of PreC, C, X, PreS1, PreS2, S, P and/or SP gene(s).

[00519] In embodiments, the immune disease is severe combined immunodeficiency (SCID), Omenn syndrome, and in one aspect the target is Recombination Activating Gene 1 (RAG1) or an interleukin-7 receptor (IL7R). In particular embodiments, the disease is Transthyretin Amyloidosis (ATTR), Familial amyloid cardiomyopathy, and in one aspect, the target is the TTR gene, including one or more mutations in the TTR gene. In embodiments, the disease is Alpha-1 Antitrypsin Deficiency (AATD) or another disease in which Alpha-1 Antitrypsin is implicated, for example GvHD, Organ transplant rejection, diabetes, liver

disease, COPD, Emphysema and Cystic Fibrosis, in particular embodiments, the target is SERPINA1.

[00520] In embodiments, the disease is primary hyperoxaluria, which, in certain embodiments, the target comprises one or more of Lactate dehydrogenase A (LDHA) and hydroxy Acid Oxidase 1 (HAO 1). In embodiments, the disease is primary hyperoxaluria type 1 (ph1) and other alanine-glyoxylate aminotransferase (agxt) gene related conditions or disorders, such as Adenocarcinoma, Chronic Alcoholic Intoxication, Alzheimer's Disease, Cooley's anemia, Aneurysm, Anxiety Disorders, Asthma, Malignant neoplasm of breast, Malignant neoplasm of skin, Renal Cell Carcinoma, Cardiovascular Diseases, Malignant tumor of cervix, Coronary Arteriosclerosis, Coronary heart disease, Diabetes, Diabetes Mellitus, Diabetes Mellitus Non- Insulin-Dependent, Diabetic Nephropathy, Eclampsia, Eczema, Subacute Bacterial Endocarditis, Glioblastoma, Glycogen storage disease type II, Sensorineural Hearing Loss (disorder), Hepatitis, Hepatitis A, Hepatitis B, Homocystinuria, Hereditary Sensory Autonomic Neuropathy Type 1, Hyperaldosteronism, Hypercholesterolemia, Hyperoxaluria, Primary Hyperoxaluria, Hypertensive disease, Inflammatory Bowel Diseases, Kidney Calculi, Kidney Diseases, Chronic Kidney Failure, leiomyosarcoma, Metabolic Diseases, Inborn Errors of Metabolism, Mitral Valve Prolapse Syndrome, Myocardial Infarction, Neoplasm Metastasis, Nephrotic Syndrome, Obesity, Ovarian Diseases, Periodontitis, Polycystic Ovary Syndrome, Kidney Failure, Adult Respiratory Distress Syndrome, Retinal Diseases, Cerebrovascular accident, Turner Syndrome, Viral hepatitis, Tooth Loss, Premature Ovarian Failure, Essential Hypertension, Left Ventricular Hypertrophy, Migraine Disorders, Cutaneous Melanoma, Hypertensive heart disease, Chronic glomerulonephritis, Migraine with Aura, Secondary hypertension, Acute myocardial infarction, Atherosclerosis of aorta, Allergic asthma, pineoblastoma, Malignant neoplasm of lung, Primary hyperoxaluria type I, Primary hyperoxaluria type 2, Inflammatory Breast Carcinoma, Cervix carcinoma, Restenosis, Bleeding ulcer, Generalized glycogen storage disease of infants, Nephrolithiasis, Chronic rejection of renal transplant, Urolithiasis, pricking of skin, Metabolic Syndrome X, Maternal hypertension, Carotid Atherosclerosis, Carcinogenesis, Breast Carcinoma, Carcinoma of lung, Nephronophthisis, Microalbuminuria, Familial Retinoblastoma, Systolic Heart Failure Ischemic stroke, Left ventricular systolic dysfunction, Cauda Equina Paraganglioma, Hepatocarcinogenesis, Chronic Kidney Diseases, Glioblastoma Multiforme, Non-Neoplastic Disorder, Calcium Oxalate Nephrolithiasis, Ablepharon-Macrostomia Syndrome, Coronary Artery Disease, Liver carcinoma, Chronic kidney disease stage 5, Allergic rhinitis (disorder), Crigler Najjar syndrome type 2, and

Ischemic Cerebrovascular Accident. In certain embodiments, treatment is targeted to the liver. In embodiments, the gene is AGXT, with a a cytogenetic location of 2q37.3 and the genomic coordinate are on Chromosome 2 on the forward strand at position 240,868,479-240,880,502.

[00521] Treatment can also target collagen type vii alpha 1 chain (col7a1) gene related conditions or disorders, such as Malignant neoplasm of skin, Squamous cell carcinoma, Colorectal Neoplasms, Crohn Disease, Epidermolysis Bullosa, Indirect Inguinal Hernia, Pruritus, Schizophrenia, Dermatologic disorders, Genetic Skin Diseases, Teratoma, Cockayne-Touraine Disease, Epidermolysis Bullosa Acquisita, Epidermolysis Bullosa Dystrophica, Junctional Epidermolysis Bullosa, Hallopeau- Siemens Disease, Bullous Skin Diseases, Agenesis of corpus callosum, Dystrophia unguium, Vesicular Stomatitis, Epidermolysis Bullosa With Congenital Localized Absence Of Skin And Deformity Of Nails, Juvenile Myoclonic Epilepsy, Squamous cell carcinoma of esophagus, Poikiloderma of Kindler, pretibial Epidermolysis bullosa, Dominant dystrophic epidermolysis bullosa albopapular type (disorder), Localized recessive dystrophic epidermolysis bullosa, Generalized dystrophic epidermolysis bullosa, Squamous cell carcinoma of skin, Epidermolysis Bullosa Pruriginosa, Mammary Neoplasms, Epidermolysis Bullosa Simplex Superficialis, Isolated Toenail Dystrophy, Transient bullous dermolysis of the newborn, Autosomal Recessive Epidermolysis Bullosa Dystrophica Localisata Variant, and Autosomal Recessive Epidermolysis Bullosa Dystrophica Inversa.

[00522] In embodiments, the disease is acute myeloid leukemia (AML), targeting Wilms Tumor I (WTI) and HLA expressing cells. In embodiments, the therapy is T cell therapy, as described elsewhere herein, comprising engineered T cells with WTI specific TCRs. In certain embodiments, the target is CD157 in AML.

[00523] In embodiments, the disease is a blood disease. In certain embodiments, the disease is hemophilia, in one aspect the target is Factor XI. In other embodiments, the disease is a hemoglobinopathy, such as sickle cell disease, sickle cell trait, hemoglobin C disease, hemoglobin C trait, hemoglobin S/C disease, hemoglobin D disease, hemoglobin E disease, a thalassemia, a condition associated with hemoglobin with increased oxygen affinity, a condition associated with hemoglobin with decreased oxygen affinity, unstable hemoglobin disease, methemoglobinemia. Hemostasis and Factor X and XII deficiencies can also be treated. In embodiments, the target is BCL11A gene (e.g., a human BCL11a gene), a BCL11a enhancer (e.g., a human BCL11a enhancer), or a HFPH region (e.g., a human HPFH region), beta globulin, fetal hemoglobin, γ–globin genes (e.g., HBG1, HBG2, or HBG1 and HBG2), the erythroid specific enhancer of the BCL11A gene (BCL11Ae), or a combination thereof.

[00524] In embodiments, the target locus can be one or more of RAC, TRBCl, TRBC2, CD3E, CD3G, CD3D, B2M, CIITA, CD247, HLA-A, HLA-B, HLA-C, DCK, CD52, FKBP1A, NLRC5, RFXANK, RFX5, RFXAP, NR3C1, CD274, HAVCR2, LAG3, PDCD1, PD-L2, HCF2, PAI, TFPI, PLAT, PLAU, PLG, RPOZ, F7, F8, F9, F2, F5, F7, F10, F11, F12, F13A1, F13B, STAT1, FOXP3, IL2RG, DCLRE1C, ICOS, MHC2TA, GALNS, HGSNAT, ARSB, RFXAP, CD20, CD81, TNFRSF13B, SEC23B, PKLR, IFNG, SPTB, SPTA, SLC4A1, EPO, EPB42, CSF2 CSF3, VFW, SERPINCA1, CTLA4, CEACAM (e.g., CEACAM-1, CEACAM-3 and/or CEACAM-5), VISTA, BTLA, TIGIT, LAIR1, CD160, 2B4, CD80, CD86, B7-H3 (CD113), B7-H4 (VTCNl), HVEM (TNFRSF14 or CD107), KIR, A2aR, MHC class I, MHC class II, GAL9, adenosine, and TGF beta, PTPN11, and combinations thereof. In embodiments, the target sequence within the genomic nucleic acid sequence at Chrl 1:5,250,094-5,250,237, - strand, hg38; Chrl 1:5,255,022-5,255,164, - strand, hg38; nondeletional HFPH region; Chrl 1:5,249,833 to Chrl 1:5,250,237, - strand, hg38; Chrl 1:5,254,738 to Chrl 1:5,255, 164, - strand, hg38; Chrl 1 : 5,249,833-5,249,927, - strand, hg3; Chrl 1 : 5,254,738-5,254,851, - strand, hg38; Chrl 1:5,250, 139-5,250,237, - strand, hg38.

[00525] In embodiments, the disease is associated with high cholesterol, and regulation of cholesterol is provided, in some embodiments, regulation is affected by modification in the target PCSK9. Other diseases in which PCSK9 can be implicated, and thus would be a target for the systems and methods described herein include Abetaiipoproteinemia, Adenoma, Arteriosclerosis, Atherosclerosis, Cardiovascular Diseases, Cholelithiasis, Coronary Arteriosclerosis, Coronary heart disease, Non-Insulin-Dependent Diabetes Meliitus, Hypercholesterolemia, Familial Hypercholesterolemia, Hyperinsuiinism, Hyperlipidemia, Familial Combined Hyperlipidemia, Hypobetalipoproteinemias, Chronic Kidney Failure, Liver diseases, Liver neoplasms, melanoma, Myocardial Infarction, Narcolepsy, Neoplasm Metastasis, Nephroblastoma, Obesity, Peritonitis, Pseudoxanthoma Elasticum, Cerebrovascular accident, Vascular Diseases, Xanthomatosis, Peripheral Vascular Diseases, Myocardial Ischemia, Dyslipidemias, Impaired glucose tolerance, Xanthoma, Polygenic hypercholesterolemia, Secondary malignant neoplasm of liver, Dementia, Overweight, Hepatitis C, Chronic, Carotid Atherosclerosis, Hyperlipoproteinemia Type Ha, Intracranial Atherosclerosis, Ischemic stroke, Acute Coronary Syndrome, Aortic calcification, Cardiovascular morbidity, Hyperlipoproteinemia Type lib, Peripheral Arterial Diseases, Familial Hyperaldosteronism Type II, Familial hypobetalipoproteinemia, Autosomal Recessive Hypercholesterolemia, Autosomal Dominant Hypercholesterolemia 3, Coronary Artery Disease, Liver carcinoma, Ischemic Cerebrovascular Accident, and Arteriosclerotic

cardiovascular disease NOS. In embodiments, the treatment can be targeted to the liver, the primary location of activity of PCSK9.

[00526] In embodiments, the disease or disorder is Hyper IGM syndrome or a disorder characterized by defective CD40 signaling. In certain embodiments, the insertion of CD40L exons are used to restore proper CD40 signaling and B cell class switch recombination. In particular embodiments, the target is CD40 ligand (CD40L)-edited at one or more of exons 2-5 of the CD40L gene, in cells, e.g., T cells or hematopoietic stem cells (HSCs).

[00527] In embodiments, the disease is merosin-deficient congenital muscular dystrophy (mdcmd) and other laminin, alpha 2 (lama2) gene related conditions or disorders. The therapy can be targeted to the muscle, for example, skeletal muscle, smooth muscle, and/or cardiac muscle. In certain embodiments, the target is Laminin, Alpha 2 (LAMA2) which may also be referred to as Laminin- 12 Subunit Alpha, Laminin-2 Subunit Alpha, Laminin-4 Subunit Alpha 3, Merosin Heavy Chain, Laminin M Chain, LAMM, Congenital Muscular Dystrophy and Merosin. LAMA2 has a cytogenetic location of 6q22.33 and the genomic coordinate are on Chromosome 6 on the forward strand at position 128,883, 141-129,516,563. In embodiments, the disease treated can be Merosin-Deficient Congenital Muscular Dystrophy (MDCMD), Amyotrophic Lateral Sclerosis, Bladder Neoplasm, Charcot-Marie-Tooth Disease, Colorectal Carcinoma, Contracture, Cyst, Duchenne Muscular Dystrophy, Fatigue, Hyperopia, Renovascular Hypertension, melanoma, Mental Retardation, Myopathy, Muscular Dystrophy, Myopia, Myositis, Neuromuscular Diseases, Peripheral Neuropathy, Refractive Errors, Schizophrenia, Severe mental retardation (I.Q. 20-34), Thyroid Neoplasm, Tobacco Use Disorder, Severe Combined Immunodeficiency, Synovial Cyst, Adenocarcinoma of lung (disorder), Tumor Progression, Strawberry nevus of skin, Muscle degeneration, Microdontia (disorder), Walker-Warburg congenital muscular dystrophy, Chronic Periodontitis, Leukoencephalopathies, Impaired cognition, Fukuyama Type Congenital Muscular Dystrophy, Scleroatonic muscular dystrophy, Eichsfeld type congenital muscular dystrophy, Neuropathy, Muscle eye brain disease, Limb-Muscular Dystrophies, Girdle, Congenital muscular dystrophy (disorder), Muscle fibrosis, cancer recurrence, Drug Resistant Epilepsy, Respiratory Failure, Myxoid cyst, Abnormal breathing, Muscular dystrophy congenital merosin negative, Colorectal Cancer, Congenital Muscular Dystrophy due to Partial LAMA2 Deficiency, and Autosomal Dominant Craniometaphyseal Dysplasia.

[00528] In certain embodiments, the target is an AAVS1 (PPPIR12C), an ALB gene, an Angptl3 gene, an ApoC3 gene, an ASGR2 gene, a CCR5 gene, a FIX (F9) gene, a G6PC gene, a Gys2 gene, an HGD gene, a Lp(a) gene, a Pcsk9 gene, a Serpinal gene, a TF gene, and a TTR

gene). Assessment of efficiency of HDR/NHEJ mediated knock-in of cDNA into the first exon can utilize cDNA knock-in into "safe harbor" sites such as: single-stranded or double-stranded DNA having homologous arms to one of the following regions, for example: ApoC3 (chr11:116829908-116833071), Angptl3 (chr1:62,597,487-62,606,305), Serpinal (chr14:94376747-94390692), Lp(a) (chr6:160531483-160664259), Pcsk9 (chr1:55,039,475-55,064,852), FIX (chrX:139,530,736-139,563,458), ALB (chr4:73,404,254-73,421,411), TTR (chr1 8:31,591,766-31,599,023), TF (chr3:133,661,997-133,779,005), G6PC (chr17:42,900,796-42,914,432), Gys2 (chr12:21,536,188-21,604,857), AAVS1 (PPP1R12C) (chr19:55,090,912-55,117,599), HGD (chr3:120,628,167-120,682,570), CCR5 (chr3:46,370,854-46,376,206), or ASGR2 (chr17:7,101,322-7,114,310).

[00529] In one aspect, the target is superoxide dismutase 1, soluble (SOD1), which can aid in treatment of a disease or disorder associated with the gene. In particular embodiments, the disease or disorder is associated with SOD1, and can be, for example, Adenocarcinoma, Albuminuria, Chronic Alcoholic Intoxication, Alzheimer's Disease, Amnesia, Amyloidosis, Amyotrophic Lateral Sclerosis, Anemia, Autoimmune hemolytic anemia, Sickle Cell Anemia, Anoxia, Anxiety Disorders, Aortic Diseases, Arteriosclerosis, Rheumatoid Arthritis, Asphyxia Neonatorum, Asthma, Atherosclerosis, Autistic Disorder, Autoimmune Diseases, Barrett Esophagus, Behcet Syndrome, Malignant neoplasm of urinary bladder, Brain Neoplasms, Malignant neoplasm of breast, Oral candidiasis, Malignant tumor of colon, Bronchogenic Carcinoma, Non-Small Cell Lung Carcinoma, Squamous cell carcinoma, Transitional Cell Carcinoma, Cardiovascular Diseases, Carotid Artery Thrombosis, Neoplastic Cell Transformation, Cerebral Infarction, Brain Ischemia, Transient Ischemic Attack, Charcot-Marie-Tooth Disease, Cholera, Colitis, Colorectal Carcinoma, Coronary Arteriosclerosis, Coronary heart disease, Infection by Cryptococcus neoformans, Deafness, Cessation of life, Deglutition Disorders, Presenile dementia, Depressive disorder, Contact Dermatitis, Diabetes, Diabetes Mellitus, Experimental Diabetes Mellitus, Insulin-Dependent Diabetes Mellitus, Non-Insulin-Dependent Diabetes Mellitus, Diabetic Angiopathies, Diabetic Nephropathy, Diabetic Retinopathy, Down Syndrome, Dwarfism, Edema, Japanese Encephalitis, Toxic Epidermal Necrolysis, Temporal Lobe Epilepsy, Exanthema, Muscular fasciculation, Alcoholic Fatty Liver, Fetal Growth Retardation, Fibromyalgia, Fibrosarcoma, Fragile X Syndrome, Giardiasis, Glioblastoma, Glioma, Headache, Partial Hearing Loss, Cardiac Arrest, Heart failure, Atrial Septal Defects, Helminthiasis, Hemochromatosis, Hemolysis (disorder), Chronic Hepatitis, HIV Infections, Huntington Disease, Hypercholesterolemia, Hyperglycemia, Hyperplasia, Hypertensive disease, Hyperthyroidism, Hypopituitarism,

Hypoproteinemia, Hypotension, natural Hypothermia, Hypothyroidism, Immunologic Deficiency Syndromes, Immune System Diseases, Inflammation, Inflammatory Bowel Diseases, Influenza, Intestinal Diseases, Ischemia, Kearns-Sayre syndrome, Keratoconus, Kidney Calculi, Kidney Diseases, Acute Kidney Failure, Chronic Kidney Failure, Polycystic Kidney Diseases, leukemia, Myeloid Leukemia, Acute Promyelocytic Leukemia, Liver Cirrhosis, Liver diseases, Liver neoplasms, Locked-In Syndrome, Chronic Obstructive Airway Disease, Lung Neoplasms, Systemic Lupus Erythematosus, Non-Hodgkin Lymphoma, Machado- Joseph Disease, Malaria, Malignant neoplasm of stomach, Animal Mammary Neoplasms, Marfan Syndrome, Meningomyelocele, Mental Retardation, Mitral Valve Stenosis, Acquired Dental Fluorosis, Movement Disorders, Multiple Sclerosis, Muscle Rigidity, Muscle Spasticity, Muscular Atrophy, Spinal Muscular Atrophy, Myopathy, Mycoses, Myocardial Infarction, Myocardial Reperfusion Injury, Necrosis, Nephrosis, Nephrotic Syndrome, Nerve Degeneration, nervous system disorder, Neuralgia, Neuroblastoma, Neuroma, Neuromuscular Diseases, Obesity, Occupational Diseases, Ocular Hypertension, Oligospermia, Degenerative polyarthritis, Osteoporosis, Ovarian Carcinoma, Pain, Pancreatitis, Papillon-Lefevre Disease, Paresis, Parkinson Disease, Phenylketonurias, Pituitary Diseases, Pre-Eclampsia, Prostatic Neoplasms, Protein Deficiency, Proteinuria, Psoriasis, Pulmonary Fibrosis, Renal Artery Obstruction, Reperfusion Injury, Retinal Degeneration, Retinal Diseases, Retinoblastoma, Schistosomiasis, Schistosomiasis mansoni, Schizophrenia, Scrapie, Seizures, Age-related cataract, Compression of spinal cord, Cerebrovascular accident, Subarachnoid Hemorrhage, Progressive supranuclear palsy, Tetanus, Trisomy, Turner Syndrome, Unipolar Depression, Urticaria, Vitiligo, Vocal Cord Paralysis, Intestinal Volvulus, Weight Gain, HMN (Hereditary Motor Neuropathy) Proximal Type I, Holoprosencephaly, Motor Neuron Disease, Neurofibrillary degeneration (morphologic abnormality), Burning sensation, Apathy, Mood swings, Synovial Cyst, Cataract, Migraine Disorders, Sciatic Neuropathy, Sensory neuropathy, Atrophic condition of skin, Muscle Weakness, Esophageal carcinoma, Lingual-Facial-Buccal Dyskinesia, Idiopathic pulmonary hypertension, Lateral Sclerosis, Migraine with Aura, Mixed Conductive-Sensorineural Hearing Loss, Iron deficiency anemia, Malnutrition, Prion Diseases, Mitochondrial Myopathies, MELAS Syndrome, Chronic progressive external ophthalmoplegia, General Paralysis, Premature aging syndrome, Fibrillation, Psychiatric symptom, Memory impairment, Muscle degeneration, Neurologic Symptoms, Gastric hemorrhage, Pancreatic carcinoma, Pick Disease of the Brain, Liver Fibrosis, Malignant neoplasm of lung, Age related macular degeneration, Parkinsonian Disorders, Disease

Progression, Hypocupremia, Cytochrome-c Oxidase Deficiency, Essential Tremor, Familial Motor Neuron Disease, Lower Motor Neuron Disease, Degenerative myelopathy, Diabetic Polyneuropathies, Liver and Intrahepatic Biliary Tract Carcinoma, Persian Gulf Syndrome, Senile Plaques, Atrophic, Frontotemporal dementia, Semantic Dementia, Common Migraine, Impaired cognition, Malignant neoplasm of liver, Malignant neoplasm of pancreas, Malignant neoplasm of prostate, Pure Autonomic Failure, Motor symptoms, Spastic, Dementia, Neurodegenerative Disorders, Chronic Hepatitis C, Guam Form Amyotrophic Lateral Sclerosis, Stiff limbs, Multisystem disorder, Loss of scalp hair, Prostate carcinoma, Hepatopulmonary Syndrome, Hashimoto Disease, Progressive Neoplastic Disease, Breast Carcinoma, Terminal illness, Carcinoma of lung, Tardive Dyskinesia, Secondary malignant neoplasm of lymph node, Colon Carcinoma, Stomach Carcinoma, Central neuroblastoma, Dissecting aneurysm of the thoracic aorta, Diabetic macular edema, Microalbuminuria, Middle Cerebral Artery Occlusion, Middle Cerebral Artery Infarction, Upper motor neuron signs, Frontotemporal Lobar Degeneration, Memory Loss, Classical phenylketonuria, CADASIL Syndrome, Neurologic Gait Disorders, Spinocerebellar Ataxia Type 2, Spinal Cord Ischemia, Lewy Body Disease, Muscular Atrophy, Spinobulbar, Chromosome 21 monosomy, Thrombocytosis, Spots on skin, Drug-Induced Liver Injury, Hereditary Leber Optic Atrophy, Cerebral Ischemia, ovarian neoplasm, Tauopathies, Macroangiopathy, Persistent pulmonary hypertension, Malignant neoplasm of ovary, Myxoid cyst, Drusen, Sarcoma, Weight decreased, Major Depressive Disorder, Mild cognitive disorder, Degenerative disorder, Partial Trisomy, Cardiovascular morbidity, hearing impairment, Cognitive changes, Ureteral Calculi, Mammary Neoplasms, Colorectal Cancer, Chronic Kidney Diseases, Minimal Change Nephrotic Syndrome, Non-Neoplastic Disorder, X-Linked Bulbo- Spinal Atrophy, Mammographic Density, Normal Tension Glaucoma Susceptibility To Finding), Vitiligo-Associated Multiple Autoimmune Disease Susceptibility 1 (Finding), Amyotrophic Lateral Sclerosis And/Or Frontotemporal Dementia 1, Amyotrophic Lateral Sclerosis 1, Sporadic Amyotrophic Lateral Sclerosis, monomelic Amyotrophy, Coronary Artery Disease, Transformed migraine, Regurgitation, Urothelial Carcinoma, Motor disturbances, Liver carcinoma, Protein Misfolding Disorders, TDP-43 Proteinopathies, Promyelocytic leukemia, Weight Gain Adverse Event, Mitochondrial cytopathy, Idiopathic pulmonary arterial hypertension, Progressive cGVHD, Infection, GRN-related frontotemporal dementia, Mitochondrial pathology, and Hearing Loss.

[00530] In particular embodiments, the disease is associated with the gene ATXN1, ATXN2, or ATXN3, which may be targeted for treatment. In some embodiments, the CAG

repeat region located in exon 8 of ATXN1, exon 1 of ATXN2, or exon 10 of the ATXN3 is targeted. In embodiments, the disease is spinocerebellar ataxia 3 (sca3), sca1, or sca2 and other related disorders, such as Congenital Abnormality, Alzheimer's Disease, Amyotrophic Lateral Sclerosis, Ataxia, Ataxia Telangiectasia, Cerebellar Ataxia, Cerebellar Diseases, Chorea, Cleft Palate, Cystic Fibrosis, Mental Depression, Depressive disorder, Dystonia, Esophageal Neoplasms, Exotropia, Cardiac Arrest, Huntington Disease, Machado- Joseph Disease, Movement Disorders, Muscular Dystrophy, Myotonic Dystrophy, Narcolepsy, Nerve Degeneration, Neuroblastoma, Parkinson Disease, Peripheral Neuropathy, Restless Legs Syndrome, Retinal Degeneration, Retinitis Pigmentosa, Schizophrenia, Shy-Drager Syndrome, Sleep disturbances, Hereditary Spastic Paraplegia, Thromboembolism, Stiff-Person Syndrome, Spinocerebellar Ataxia, Esophageal carcinoma, Polyneuropathy, Effects of heat, Muscle twitch, Extrapyramidal sign, Ataxic, Neurologic Symptoms, Cerebral atrophy, Parkinsonian Disorders, Protein S Deficiency, Cerebellar degeneration, Familial Amyloid Neuropathy Portuguese Type, Spastic syndrome, Vertical Nystagmus, Nystagmus End-Position, Antithrombin III Deficiency, Atrophic, Complicated hereditary spastic paraplegia, Multiple System Atrophy, Pallidoluysian degeneration, Dystonia Disorders, Pure Autonomic Failure, Thrombophilia, Protein C, Deficiency, Congenital Myotonic Dystrophy, Motor symptoms, Neuropathy, Neurodegenerative Disorders, Malignant neoplasm of esophagus, Visual disturbance, Activated Protein C Resistance, Terminal illness, Myokymia, Central neuroblastoma, Dyssomnias, Appendicular Ataxia, Narcolepsy-Cataplexy Syndrome, Machado- Joseph Disease Type I, Machado- Joseph Disease Type II, Machado- Joseph Disease Type III, Dentatorubral-Pallidoluysian Atrophy, Gait Ataxia, Spinocerebellar Ataxia Type 1, Spinocerebellar Ataxia Type 2, Spinocerebellar Ataxia Type 6 (disorder), Spinocerebellar Ataxia Type 7, Muscular Spinobulbar Atrophy, Genomic Instability, Episodic ataxia type 2 (disorder), Bulbo-Spinal Atrophy X-Linked, Fragile X Tremor/ Ataxia Syndrome, Thrombophilia Due to Activated Protein C Resistance (Disorder), Amyotrophic Lateral Sclerosis 1, Neuronal Intranuclear Inclusion Disease, Hereditary Antithrombin Iii Deficiency, and Late-Onset Parkinson Disease.

[00531] In embodiments, the disease is associated with expression of a tumor antigen-cancer or non-cancer related indication, for example acute lymphoid leukemia, diffuse large B cell lymphoma, follicular lymphoma, chronic lymphocytic leukemia, Hodgkin lymphoma, non-Hodgkin lymphoma. In embodiments, the target can be TET2 intron, a TET2 intron-exon junction, a sequence within a genomic region of chr4.

[00532] In embodiments, neurodegenerative diseases can be treated. In particular embodiments, the target is Synuclein, Alpha (SNCA). In certain embodiments, the disorder treated is a pain related disorder, including congenital pain insensitivity, Compressive Neuropathies, Paroxysmal Extreme Pain Disorder, High grade atrioventricular block, Small Fiber Neuropathy, and Familial Episodic Pain Syndrome 2. In certain embodiments, the target is Sodium Channel, Voltage Gated, Type X Alpha Subunit (SCNIOA).

[00533] In certain embodiments, hematopoietic stem cells and progenitor stem cells are edited, including knock-ins. In particular embodiments, the knock-in is for treatment of lysosomal storage diseases, glycogen storage diseases, mucopolysaccharoidoses, or any disease in which the secretion of a protein will ameliorate the disease. In one embodiment, the disease is sickle cell disease (SCD). In another embodiment, the disease is β-thalassemia.

[00534] In certain embodiments, the T cell or NK cell is used for cancer treatment and may include T cells comprising the recombinant receptor (e.g. CAR) and one or more phenotypic markers selected from CCR7+, 4-1BB+ (CD137+), TIM3+, CD27+, CD62L+, CD127+, CD45RA+, CD45RO-, t-betl'w, IL-7Ra+, CD95+, IL-2RP+, CXCR3+ or LFA-1+. In certain embodiments the editing of a T cell for caner immunotherapy comprises altering one or more T-cell expressed gene, e.g., one or more of FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, B2M, TRAC and TRBC gene. In some embodiments, editing includes alterations introduced into, or proximate to, the CBLB target sites to reduce CBLB gene expression in T cells for treatment of proliferative diseases and may include larger insertions or deletions at one or more CBLB target sites. T cell editing of TGFBR2 target sequence can be, for example, located in exon 3, 4, or 5 of the TGFBR2 gene and utilized for cancers and lymphoma treatment.

[00535] Cells for transplantation can be edited and may include allele-specific modification of one or more immunogenicity genes (e.g., an HLA gene) of a cell, e.g., HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DRB3/4/5, HLA-DQ, and HLA-DP MiHAs, and any other MHC Class I or Class II genes or loci, which may include delivery of one or more matched recipient HLA alleles into the original position(s) where the one or more mismatched donor HLA alleles are located, and may include inserting one or more matched recipient HLA alleles into a “safe harbor” locus. In an embodiment, the method further includes introducing a chemotherapy resistance gene for in vivo selection in a gene.

[00536] Methods and systems can target Dystrophia Myotonica-Protein Kinase (DMPK) for editing, in particular embodiments, the target is the CTG trinucleotide repeat in the 3' untranslated region (UTR) of the DMPK gene. Disorders or diseases associated with DMPK include Atherosclerosis, Azoospermia, Hypertrophic Cardiomyopathy, Celiac Disease,

Congenital chromosomal disease, Diabetes Mellitus, Focal glomerulosclerosis, Huntington Disease, Hypogonadism, Muscular Atrophy, Myopathy, Muscular Dystrophy, Myotonia, Myotonic Dystrophy, Neuromuscular Diseases, Optic Atrophy, Paresis, Schizophrenia, Cataract, Spinocerebellar Ataxia, Muscle Weakness, Adrenoleukodystrophy, Centronuclear myopathy, Interstitial fibrosis, myotonic muscular dystrophy, Abnormal mental state, X-linked Charcot- Marie-Tooth disease 1, Congenital Myotonic Dystrophy, Bilateral cataracts (disorder), Congenital Fiber Type Disproportion, Myotonic Disorders, Multisystem disorder, 3- Methylglutaconic aciduria type 3, cardiac event, Cardiogenic Syncope, Congenital Structural Myopathy, Mental handicap, Adrenomyeloneuropathy, Dystrophia myotonica 2, and Intellectual Disability.

[00537] In embodiments, the disease is an inborn error of metabolism. The disease may be selected from Disorders of Carbohydrate Metabolism (glycogen storage disease, G6PD deficiency), Disorders of Amino Acid Metabolism (phenylketonuria, maple syrup urine disease, glutaric acidemia type 1), Urea Cycle Disorder or Urea Cycle Defects (carbamoyl phosphate synthease I deficiency), Disorders of Organic Acid Metabolism (alkaptonuria, 2-hydroxyglutaric acidurias), Disorders of Fatty Acid Oxidation/Mitochondrial Metabolism (Medium-chain acyl-coenzyme A dehydrogenase deficiency), Disorders of Porphyrin metabolism (acute intermittent porphyria), Disorders of Purine/Pyrimidine Metabolism (Lesch-Nynan syndrome), Disorders of Steroid Metabolism (lipoid congenital adrenal hyperplasia, congenital adrenal hyperplasia), Disorders of Mitochondrial Function (Kearns-Sayre syndrome), Disorders of Peroxisomal function (Zellweger syndrome), or Lysosomal Storage Disorders (Gaucher's disease, Niemann-Pick disease).

[00538] In embodiments, the target can comprise Recombination Activating Gene 1 (RAG1), BCL11 A, PCSK9, laminin, alpha 2 (lama2), ATXN3, alanine-glyoxylate aminotransferase (AGXT), collagen type vii alpha 1 chain (COL7a1), spinocerebellar ataxia type 1 protein (ATXN1), Angiopoietin-like 3 (ANGPTL3), Frataxin (FXN), Superoxidase Dismutase 1, soluble (SOD1), Synuclein, Alpha (SNCA), Sodium Channel, Voltage Gated, Type X Alpha Subunit (SCN10A), Spinocerebellar Ataxia Type 2 Protein (ATXN2), Dystrophia Myotonica-Protein Kinase (DMPK), beta globin locus on chromosome 11, acyl-coenzyme A dehydrogenase for medium chain fatty acids (ACADM), long- chain 3-hydroxyl-coenzyme A dehydrogenase for long chain fatty acids (HADHA), acyl-coenzyme A dehydrogenase for very long-chain fatty acids (ACADVL), Apolipoprotein C3 (APOCIII), Transthyretin (TTR), Angiopoietin-like 4 (ANGPTL4), Sodium Voltage-Gated Channel Alpha

Subunit 9 (SCN9A), Interleukin-7 receptor (IL7R), glucose-6-phosphatase, catalytic (G6PC), haemochromatosis (HFE), SERPINA1, C9ORF72, β-globin, dystrophin, γ-globin.

[00539] In certain embodiments, the disease or disorder is associated with Apolipoprotein C3 (APOCIII), which can be targeted for editing. In embodiments, the disease or disorder may be Dyslipidemias, Hyperalphalipoproteinemia Type 2, Lupus Nephritis, Wilms Tumor 5, Morbid obesity and spermatogenic, Glaucoma, Diabetic Retinopathy, Arthrogryposis renal dysfunction cholestasis syndrome, Cognition Disorders, Altered response to myocardial infarction, Glucose Intolerance, Positive regulation of triglyceride biosynthetic process, Renal Insufficiency, Chronic, Hyperlipidemias, Chronic Kidney Failure, Apolipoprotein C-III Deficiency, Coronary Disease, Neonatal Diabetes Mellitus, Neonatal, with Congenital Hypothyroidism, Hypercholesterolemia Autosomal Dominant 3, Hyperlipoproteinemia Type III, Hyperthyroidism, Coronary Artery Disease, Renal Artery Obstruction, Metabolic Syndrome X, Hyperlipidemia, Familial Combined, Insulin Resistance, Transient infantile hypertriglyceridemia, Diabetic Nephropathies, Diabetes Mellitus (Type 1), Nephrotic Syndrome Type 5 with or without ocular abnormalities, and Hemorrhagic Fever with renal syndrome.

[00540] In certain embodiments, the target is Angiopoietin-like 4(ANGPTL4). Diseases or disorders associated with ANGPTL4 that can be treated include ANGPTL4 is associated with dyslipidemias, low plasma triglyceride levels, regulator of angiogenesis and modulate tumorigenesis, and severe diabetic retinopathy. both proliferative diabetic retinopathy and non-proliferative diabetic retinopathy.

[00541] In embodiments, editing can be used for the treatment of fatty acid disorders. In certain embodiments, the target is one or more of ACADM, HADHA, ACADVL. In embodiments, the targeted edit is the activity of a gene in a cell selected from the acyl-coenzyme A dehydrogenase for medium chain fatty acids (ACADM) gene, the long- chain 3-hydroxyl-coenzyme A dehydrogenase for long chain fatty acids (HADHA) gene, and the acyl-coenzyme A dehydrogenase for very long-chain fatty acids (ACADVL) gene. In one aspect, the disease is medium chain acyl-coenzyme A dehydrogenase deficiency (MCADD), long-chain 3-hydroxyl-coenzyme A dehydrogenase deficiency (LCHADD), and/or very long-chain acyl-coenzyme A dehydrogenase deficiency (VLCADD).

IMMUNE ORTHOGONAL ORTHOLOGS

[00542] In some embodiments, when CRISPR enzymes need to be expressed or administered in a subject, immunogenicity of CRISPR enzymes may be reduced by sequentially expressing or administering immune orthogonal orthologs of the CRISPR

enzymes to the subject. As used herein, the term "immune orthogonal orthologs" refer to orthologous proteins that have similar or substantially the same function or activity, but have no or low cross-reactivity with the immune response generated by one another. In some embodiments, sequential expression or administration of such orthologs elicits low or no secondary immune response. The immune orthogonal orthologs can avoid being neutralized by antibodies (e.g., existing antibodies in the host before the orthologs are expressed or administered). Cells expressing the orthologs can avoid being cleared by the host's immune system (e.g., by activated CTLs). In some examples, CRISPR enzyme orthologs from different species may be immune orthogonal orthologs.

[00543] Immune orthogonal orthologs may be identified by analyzing the sequences, structures, and/or immunogenicity of a set of candidates orthologs. In an example method, a set of immune orthogonal orthologs may be identified by a) comparing the sequences of a set of candidate orthologs (e.g., orthologs from different species) to identify a subset of candidates that have low or no sequence similarity; b) assessing immune overlap among the members of the subset of candidates to identify candidates that have no or low immune overlap. In some cases, immune overlap among candidates may be assessed by determining the binding (e.g., affinity) between a candidate ortholog and MHC (e.g., MHC type I and/or MHC II) of the host. Alternatively or additionally, immune overlap among candidates may be assessed by determining B-cell epitopes for the candidate orthologs. In one example, immune orthogonal orthologs may be identified using the method described in Moreno AM et al., BioRxiv, published online January 10, 2018, doi: doi.org/10.1101/245985.

[00544] As will be apparent, it is envisaged that the present system can be used to target any polynucleotide sequence of interest. The invention provides a non-naturally occurring or engineered composition, or one or more polynucleotides encoding components of said composition, or vector or delivery systems comprising one or more polynucleotides encoding components of said composition for use in a modifying a target cell in vivo, ex vivo or in vitro and, may be conducted in a manner alters the cell such that once modified the progeny or cell line of the CRISPR modified cell retains the altered phenotype. The modified cells and progeny may be part of a multi-cellular organism such as a plant or animal with ex vivo or in vivo application of CRISPR system to desired cell types. The CRISPR invention may be a therapeutic method of treatment. The therapeutic method of treatment may comprise gene or genome editing, or gene therapy.

[00545] Treating pathogens, like bacterial, fungal and parasitic pathogens

[00546] The present invention may also be applied to treat bacterial, fungal and parasitic pathogens. Most research efforts have focused on developing new antibiotics, which once developed, would nevertheless be subject to the same problems of drug resistance. The invention provides novel CRISPR-based alternatives which overcome those difficulties. Furthermore, unlike existing antibiotics, CRISPR-based treatments can be made pathogen specific, inducing bacterial cell death of a target pathogen while avoiding beneficial bacteria.

[00547] Jiang et al. ("RNA-guided editing of bacterial genomes using CRISPR-Cas systems," Nature Biotechnology vol. 31, p. 233-9, March 2013) used a CRISPR-Cas system to mutate or kill S. pneumoniae and E. coli. The work, which introduced precise mutations into the genomes, relied on dual-RNA:Cas-directed cleavage at the targeted genomic site to kill unmutated cells and circumvented the need for selectable markers or counter-selection systems. CRISPR systems have be used to reverse antibiotic resistance and eliminate the transfer of resistance between strains. Bickard et al. showed that Cas9, reprogrammed to target virulence genes, kills virulent, but not avirulent, S. aureus. Reprogramming the nuclease to target antibiotic resistance genes destroyed staphylococcal plasmids that harbor antibiotic resistance genesand immunized against the spread of plasmid-borne resistance genes. (see, Bikard et al., "Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials," Nature Biotechnology vol. 32, 1146–1150, doi:10.1038/nbt.3043, published online 05 October 2014.) Bikard showed that CRISPR-Cas9 antimicrobials function in vivo to kill S. aureus in a mouse skin colonization model. Similarly, Yosef et al used a CRISPR system to target genes encoding enzymes that confer resistance to β-lactam antibiotics (see Yousef et al., "Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria," Proc. Natl. Acad. Sci. USA, vol. 112, p. 7267–7272, doi: 10.1073/pnas.1500107112 published online May 18, 2015).

[00548] CRISPR systems can be used to edit genomes of parasites that are resistant to other genetic approaches. For example, a CRISPR-Cas system was shown to introduce double-stranded breaks into the in the Plasmodium yoelii genome (see, Zhang et al., "Efficient Editing of Malaria Parasite Genome Using the CRISPR/Cas9 System," mBio. vol. 5, e01414-14, Jul-Aug 2014). Ghorbal et al. ("Genome editing in the human malaria parasite Plasmodium falciparumusing the CRISPR-Cas9 system," Nature Biotechnology, vol. 32, p. 819-821, doi: 10.1038/nbt.2925, published online June 1, 2014) modified the sequences of two genes, orc1 and kelch13, which have putative roles in gene silencing and emerging resistance to artemisinin, respectively. Parasites that were altered at the appropriate sites were recovered with very high efficiency, despite there being no direct selection for the modification,

indicating that neutral or even deleterious mutations can be generated using this system. CRISPR-Cas is also used to modify the genomes of other pathogenic parasites, including Toxoplasma gondii (see Shen et al., "Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9," mBio vol. 5:e01114-14, 2014; and Sidik et al., "Efficient Genome Engineering of Toxoplasma gondii Using CRISPR/Cas9," PLoS One vol. 9, e100450, doi: 10.1371/journal.pone.0100450, published online June 27, 2014).

[00549] Vyas et al. ("A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families," Science Advances, vol. 1, e1500248, DOI: 10.1126/sciadv.1500248, April 3, 2015) employed a CRISPR system to overcome long-standing obstacles to genetic engineering in C. albicans and efficiently mutate in a single experiment both copies of several different genes. In an organism where several mechanisms contribute to drug resistance, Vyas produced homozygous double mutants that no longer displayed the hyper-resistance to fluconazole or cycloheximide displayed by the parental clinical isolate Can90. Vyas also obtained homozygous loss-of-function mutations in essential genes of C. albicans by creating conditional alleles. Null alleles of DCR1, which is required for ribosomal RNA processing, are lethal at low temperature but viable at high temperature. Vyas used a repair template that introduced a nonsense mutation and isolated dcr1/dcr1 mutants that failed to grow at 16°C.

[00550] The CRISPR system of the present invention for use in P. falciparum by disrupting chromosomal loci. Ghorbal et al. ("Genome editing in the human malaria parasite Plasmodium falciparum using the CRISPR-Cas9 system", Nature Biotechnology, 32, 819-821 (2014), DOI: 10.1038/nbt.2925, June 1, 2014) employed a CRISPR system to introduce specific gene knockouts and single-nucleotide substtitions in the malaria genome. To adapt the CRISPR-Cas system to P. falciparum, Ghorbal et al. generated expression vectors for under the control of plasmodial regulatory elements in the pUF1-Cas episome that also carries the drug-selectable marker ydhodh, which gives resistance to DSM1, a P. falciparum dihydroorotate dehydrogenase (PfDHODH) inhibitor and for transcription of the sgRNA, used P. falciparum U6 small nuclear (sn)RNA regulatory elements placing the guide RNA and the donor DNA template for homologous recombination repair on the same plasmid, pL7. See also, Zhang C. et al. ("Efficient editing of malaria parasite genome using the CRISPR/Cas9 system", MBio, 2014 Jul 1; 5(4):E01414-14, doi: 10.1128/MbIO.01414-14) and Wagner et al. ("Efficient CRISPR-Cas9-mediated genome editing in Plasmodium falciparum, Nature Methods 11, 915-918 (2014), DOI: 10.1038/nmeth.3063).

[00551] Treating pathogens, like viral pathogens such as HIV

[00552] Cas-mediated genome editing might be used to introduce protective mutations in somatic tissues to combat nongenetic or complex diseases. For example, NHEJ-mediated inactivation of the CCR5 receptor in lymphocytes (Lombardo et al., Nat Biotechnol. 2007 Nov; 25(11):1298-306) may be a viable strategy for circumventing HIV infection, whereas deletion of PCSK9 (Cohen et al., Nat Genet. 2005 Feb; 37(2):161-5) orangiopoietin (Musunuru et al., N Engl J Med. 2010 Dec 2; 363(23):2220-7) may provide therapeutic effects against statin-resistant hypercholesterolemia or hyperlipidemia. Although these targets may be also addressed using siRNA-mediated protein knockdown, a unique advantage of NHEJ-mediated gene inactivation is the ability to achieve permanent therapeutic benefit without the need for continuing treatment. As with all gene therapies, it will of course be important to establish that each proposed therapeutic use has a favorable benefit-risk ratio.

[00553] Hydrodynamic delivery of plasmid DNA encoding Cas and guide RNA along with a repair template into the liver of an adult mouse model of tyrosinemia was shown to be able to correct the mutant Fah gene and rescue expression of the wild-type Fah protein in ~1 out of 250 cells (Nat Biotechnol. 2014 Jun; 32(6):551-3). In addition, clinical trials successfully used ZF nucleases to combat HIV infection by ex vivo knockout of the CCR5 receptor. In all patients, HIV DNA levels decreased, and in one out of four patients, HIV RNA became undetectable (Tebas et al., N Engl J Med. 2014 Mar 6; 370(10):901-10). Both of these results demonstrate the promise of programmable nucleases as a new therapeutic platform.

[00554] In another embodiment, self-inactivating lentiviral vectors with an siRNA targeting a common exon shared by HIV tat/rev, a nucleolar-localizing TAR decoy, and an anti–CCR5-specific hammerhead ribozyme (see, e.g., DiGiusto et al. (2010) Sci Transl Med 2:36ra43) may be used/and or adapted to the CRISPR-Cas system of the present invention. A minimum of 2.5 × 106 CD34+ cells per kilogram patient weight may be collected and prestimulated for 16 to 20 hours in X-VIVO 15 medium (Lonza) containing 2 μmol/L-glutamine, stem cell factor (100 ng/ml), Flt-3 ligand (Flt-3L) (100 ng/ml), and thrombopoietin (10 ng/ml) (CellGenix) at a density of 2 × 106 cells/ml. Prestimulated cells may be transduced with lentiviral at a multiplicity of infection of 5 for 16 to 24 hours in 75-cm2 tissue culture flasks coated with fibronectin (25 mg/cm2) (RetroNectin, Takara Bio Inc.).

[00555] With the knowledge in the art and the teachings in this disclosure the skilled person can correct HSCs as to immunodeficiency condition such as HIV / AIDS comprising contacting an HSC with a CRISPR-Cas system that targets and knocks out CCR5. An guide RNA (and advantageously a dual guide approach, e.g., a pair of different guide RNAs; for instance, guide RNAs targeting of two clinically relevant genes, B2M and CCR5, in primary human CD4+ T

cells and CD34+ hematopoietic stem and progenitor cells (HSPCs)) that targets and knocks out CCR5-and-Cas protein containing particle is contacted with HSCs. The so contacted cells can be administered; and optionally treated / expanded; cf. Cartier. See also Kiem, "Hematopoietic stem cell-based gene therapy for HIV disease," Cell Stem Cell. Feb 3, 2012; 10(2): 137–147; incorporated herein by reference along with the documents it cites; Mandal et al, "Efficient Ablation of Genes in Human Hematopoietic Stem and Effector Cells using CRISPR/Cas9," Cell Stem Cell, Volume 15, Issue 5, p643–652, 6 November 2014; incorporated herein by reference along with the documents it cites. Mention is also made of Ebina, "CRISPR/Cas9 system to suppress HIV-1 expression by editing HIV-1 integrated proviral DNA" SCIENTIFIC REPORTS | 3 : 2510 | DOI: 10.1038/srep02510, incorporated herein by reference along with the documents it cites, as another means for combatting HIV/AIDS using a CRISPR-Cas system.

[00556] The rationale for genome editing for HIV treatment originates from the observation that individuals homozygous for loss of function mutations in CCR5, a cellular co-receptor for the virus, are highly resistant to infection and otherwise healthy, suggesting that mimicking this mutation with genome editing could be a safe and effective therapeutic strategy [Liu, R., et al. Cell 86, 367-377 (1996)]. This idea was clinically validated when an HIV infected patient was given an allogeneic bone marrow transplant from a donor homozygous for a loss of function CCR5 mutation, resulting in undetectable levels of HIV and restoration of normal CD4 T-cell counts [Hutter, G., et al. The New England journal of medicine 360, 692-698 (2009)]. Although bone marrow transplantation is not a realistic treatment strategy for most HIV patients, due to cost and potential graft vs. host disease, HIV therapies that convert a patient's own T-cells into CCR5 are desirable.

[00557] Early studies using ZFNs and NHEJ to knockout CCR5 in humanized mouse models of HIV showed that transplantation of CCR5 edited CD4 T cells improved viral load and CD4 T-cell counts [Perez, E.E., et al. Nature biotechnology 26, 808-816 (2008)]. Importantly, these models also showed that HIV infection resulted in selection for CCR5 null cells, suggesting that editing confers a fitness advantage and potentially allowing a small number of edited cells to create a therapeutic effect.

[00558] As a result of this and other promising preclinical studies, genome editing therapy that knocks out CCR5 in patient T cells has now been tested in humans [Holt, N., et al. Nature biotechnology 28, 839-847 (2010); Li, L., et al. Molecular therapy : the journal of the American Society of Gene Therapy 21, 1259-1269 (2013)]. In a recent phase I clinical trial, CD4+ T cells from patients with HIV were removed, edited with ZFNs designed to knockout the CCR5 gene,

and autologously transplanted back into patients [Tebas, P., et al. The New England journal of medicine 370, 901-910 (2014)].

[00559] In another study (Mandal et al., Cell Stem Cell, Volume 15, Issue 5, p643–652, 6 November 2014), CRISPR-Cas9 has targeted two clinically relevant genes, B2M and CCR5, in human CD4+ T cells and CD34+ hematopoietic stem and progenitor cells (HSPCs). Use of single RNA guides led to highly efficient mutagenesis in HSPCs but not in T cells. A dual guide approach improved gene deletion efficacy in both cell types. HSPCs that had undergone genome editing with CRISPR-Cas retained multilineage potential. Predicted on- and off-target mutations were examined via target capture sequencing in HSPCs and low levels of off-target mutagenesis were observed at only one site. These results demonstrate that CRISPR-Cas9 can efficiently ablate genes in HSPCs with minimal off-target mutagenesis, which have broad applicability for hematopoietic cell-based therapy.

[00560] Wang et al. (PLoS One. 2014 Dec 26;9(12):e115987. doi: 10.1371/journal.pone.0115987) silenced CCR5 via CRISPR associated protein 9 (Cas9) and single guided RNAs (guide RNAs) with lentiviral vectors expressing Cas9 and CCR5 guide RNAs. Wang et al. showed that a single round transduction of lentiviral vectors expressing Cas9 and CCR5 guide RNAs into HIV-1 susceptible human CD4+ cells yields high frequencies of CCR5 gene disruption. CCR5 gene-disrupted cells are not only resistant to R5-tropic HIV-1, including transmitted/founder (T/F) HIV-1 isolates, but also have selective advantage over CCR5 gene-undisrupted cells during R5-tropic HIV-1 infection. Genome mutations at potential off-target sites that are highly homologous to these CCR5 guide RNAs in stably transduced cells even at 84 days post transduction were not detected by a T7 endonuclease I assay.

[00561] Fine et al. (Sci Rep. 2015 Jul 1;5:10777. doi: 10.1038/srep10777) identified a two-cassette system expressing pieces of the S. pyogenes Cas9 (SpCas9) protein which splice together in cellula to form a functional protein capable of site-specific DNA cleavage. With specific CRISPR guide strands, Fine et al. demonstrated the efficacy of this system in cleaving the HBB and CCR5 genes in human HEK-293T cells as a single Cas9 and as a pair of Cas9 nickases. The trans-spliced SpCas9 (tsSpCas9) displayed ~35% of the nuclease activity compared with the wild-type SpCas9 (wtSpCas9) at standard transfection doses, but had substantially decreased activity at lower dosing levels. The greatly reduced open reading frame length of the tsSpCas9 relative to wtSpCas9 potentially allows for more complex and longer genetic elements to be packaged into an AAV vector including tissue-specific promoters, multiplexed guide RNA expression, and effector domain fusions to SpCas9.

[00562] Li et al. (J Gen Virol. 2015 Aug;96(8):2381-93. doi: 10.1099/vir.0.000139. Epub 2015 Apr 8) demonstrated that CRISPR-Cas can efficiently mediate the editing of the CCR5 locus in cell lines, resulting in the knockout of CCR5 expression on the cell surface. Next-generation sequencing revealed that various mutations were introduced around the predicted cleavage site of CCR5. For each of the three most effective guide RNAs that were analyzed, no significant off-target effects were detected at the 15 top-scoring potential sites. By constructing chimeric Ad5F35 adenoviruses carrying CRISPR-Cas components, Li et al. efficiently transduced primary CD4+ T-lymphocytes and disrupted CCR5 expression, and the positively transduced cells were conferred with HIV-1 resistance.

[00563] One of skill in the art may utilize the above studies of, for example, Holt, N., et al. Nature biotechnology 28, 839-847 (2010), Li, L., et al. Molecular therapy : the journal of the American Society of Gene Therapy 21, 1259-1269 (2013), Mandal et al., Cell Stem Cell, Volume 15, Issue 5, p643–652, 6 November 2014, Wang et al. (PLoS One. 2014 Dec 26;9(12):e115987. doi: 10.1371/journal.pone.0115987), Fine et al. (Sci Rep. 2015 Jul 1;5:10777. doi: 10.1038/srep10777) and Li et al. (J Gen Virol. 2015 Aug;96(8):2381-93. doi: 10.1099/vir.0.000139. Epub 2015 Apr 8) for targeting CCR5 with the CRISPR Cas system of the present invention.

[00564] Treating pathogens, like viral pathogens, such as HBV

[00565] The present invention may also be applied to treat hepatitis B virus (HBV). However, the CRISPR Cas system must be adapted to avoid the shortcomings of RNAi, such as the risk of oversatring endogenous small RNA pathways, by for example, optimizing dose and sequence (see, e.g., Grimm et al., Nature vol. 441, 26 May 2006). For example, low doses, such as about 1-10 x 1014 particles per human are contemplated. In another embodiment, the CRISPR Cas system directed against HBV may be administered in liposomes, such as a stable nucleic-acid-lipid particle (SNALP) (see, e.g., Morrissey et al., Nature Biotechnology, Vol. 23, No. 8, August 2005). Daily intravenous injections of about 1, 3 or 5 mg/kg/day of CRISPR Cas targeted to HBV RNA in a SNALP are contemplated. The daily treatment may be over about three days and then weekly for about five weeks. In another embodiment, the system of Chen et al. (Gene Therapy (2007) 14, 11–19) may be used/and or adapted for the CRISPR Cas system of the present invention. Chen et al. use a double-stranded adeno-associated virus 8-pseudotyped vector (dsAAV2/8) to deliver shRNA. A single administration of dsAAV2/8 vector (1 x 1012 vector genomes per mouse), carrying HBV-specific shRNA, effectively suppressed the steady level of HBV protein, mRNA and replicative DNA in liver of HBV transgenic mice, leading to up to 2–3 log10 decrease in HBV load in the circulation. Significant

HBV suppression sustained for at least 120 days after vector administration. The therapeutic effect of shRNA was target sequence dependent and did not involve activation of interferon. For the present invention, a CRISPR Cas system directed to HBV may be cloned into an AAV vector, such as a dsAAV2/8 vector and administered to a human, for example, at a dosage of about 1 x 1015 vector genomes to about 1 x 1016 vector genomes per human. In another embodiment, the method of Wooddell et al. (Molecular Therapy vol. 21 no. 5, 973–985 May 2013) may be used/and or adapted to the CRISPR Cas system of the present invention. Woodell et al. show that simple coinjection of a hepatocyte-targeted, N-acetylgalactosamine-conjugated melittin-like peptide (NAG-MLP) with a liver-tropic cholesterol-conjugated siRNA (chol-siRNA) targeting coagulation factor VII (F7) results in efficient F7 knockdown in mice and nonhuman primates without changes in clinical chemistry or induction of cytokines. Using transient and transgenic mouse models of HBV infection, Wooddell et al. show that a single coinjection of NAG-MLP with potent chol-siRNAs targeting conserved HBV sequences resulted in multilog repression of viral RNA, proteins, and viral DNA with long duration of effect. Intraveinous coinjections, for example, of about 6 mg/kg of NAG-MLP and 6 mg/kg of HBV specific CRISPR Cas may be envisioned for the present invention. In the alternative, about 3 mg/kg of NAG-MLP and 3 mg/kg of HBV specific CRISPR Cas may be delivered on day one, followed by administration of about about 2-3 mg/kg of NAG-MLP and 2-3 mg/kg of HBV specific CRISPR Cas two weeks later.

[00566] In some embodiments, the target sequence is an HBV sequence. In some embodiments, the target sequences is comprised in an episomal viral nucleic acid molecule which is not integrated into the genome of the organism to thereby manipulate the episomal viral nucleic acid molecule. In some embodiments, the episomal nucleic acid molecule is a double-stranded DNA polynucleotide molecule or is a covalently closed circular DNA (cccDNA). In some embodiments, the CRISPR complex is capable of reducing the amount of episomal viral nucleic acid molecule in a cell of the organism compared to the amount of episomal viral nucleic acid molecule in a cell of the organism in the absence of providing the complex, or is capable of manipulating the episomal viral nucleic acid molecule to promote degradation of the episomal nucleic acid molecule. In some embodiments, the target HBV sequence is integrated into the genome of the organism. In some embodiments, when formed within the cell, the CRISPR complex is capable of manipulating the integrated nucleic acid to promote excision of all or part of the target HBV nucleic acid from the genome of the organism. In some embodiments, said at least one target HBV nucleic acid is comprised in a double-stranded DNA polynucleotide cccDNA molecule and/or viral DNA integrated into the genome

of the organism and wherein the CRISPR complex manipulates at least one target HBV nucleic acid to cleave viral cccDNA and/or integrated viral DNA. In some embodiments, said cleavage comprises one or more double-strand break(s) introduced into the viral cccDNA and/or integrated viral DNA, optionally at least two double-strand break(s). In some embodiments, said cleavage is via one or more single-strand break(s) introduced into the viral cccDNA and/or integrated viral DNA, optionally at least two single-strand break(s). In some embodiments, said one or more double-strand break(s) or said one or more single-strand break(s) leads to the formation of one or more insertion or deletion mutations (INDELs) in the viral cccDNA sequences and/or integrated viral DNA sequences.

[00567] Lin et al. (Mol Ther Nucleic Acids. 2014 Aug 19;3:e186. doi: 10.1038/mtna.2014.38) designed eight gRNAs against HBV of genotype A. With the HBV-specific gRNAs, the CRISPR-Cas system significantly reduced the production of HBV core and surface proteins in Huh-7 cells transfected with an HBV-expression vector. Among eight screened gRNAs, two effective ones were identified. One gRNA targeting the conserved HBV sequence acted against different genotypes. Using a hydrodynamics-HBV persistence mouse model, Lin et al. further demonstrated that this system could cleave the intrahepatic HBV genome-containing plasmid and facilitate its clearance in vivo, resulting in reduction of serum surface antigen levels. These data suggest that the CRISPR-Cas system could disrupt the HBV-expressing templates both in vitro and in vivo, indicating its potential in eradicating persistent HBV infection.

[00568] Dong et al. (Antiviral Res. 2015 Jun;118:110-7. doi: 10.1016/j.antiviral.2015.03.015. Epub 2015 Apr 3) used the CRISPR-Cas system to target the HBV genome and efficiently inhibit HBV infection. Dong et al. synthesized four single-guide RNAs (guide RNAs) targeting the conserved regions of HBV. The expression of these guide RNAS with Cas reduced the viral production in Huh7 cells as well as in HBV-replication cell HepG2.2.15. Dong et al. further demonstrated that CRISPR-Cas direct cleavage and cleavage-mediated mutagenesis occurred in HBV cccDNA of transfected cells. In the mouse model carrying HBV cccDNA, injection of guide RNA-Cas plasmids via rapid tail vein resulted in the low level of cccDNA and HBV protein.

[00569] Liu et al. (J Gen Virol. 2015 Aug;96(8):2252-61. doi: 10.1099/vir.0.000159. Epub 2015 Apr 22) designed eight guide RNAs (gRNAs) that targeted the conserved regions of different HBV genotypes, which could significantly inhibit HBV replication both in vitro and in vivo to investigate the possibility of using the CRISPR-Cas system to disrupt the HBV DNA templates. The HBV-specific gRNA/Cas system could inhibit the replication of HBV of

different genotypes in cells, and the viral DNA was significantly reduced by a single gRNA/Cas system and cleared by a combination of different gRNA/Cas systems.

[00570] Wang et al. (World J Gastroenterol. 2015 Aug 28;21(32):9554-65. doi: 10.3748/wjg.v21.i32.9554) designed 15 gRNAs against HBV of genotypes A-D. Eleven combinations of two above gRNAs (dual-gRNAs) covering the regulatory region of HBV were chosen. The efficiency of each gRNA and 11 dual-gRNAs on the suppression of HBV (genotypes A-D) replication was examined by the measurement of HBV surface antigen (HBsAg) or e antigen (HBeAg) in the culture supernatant. The destruction of HBV-expressing vector was examined in HuH7 cells co-transfected with dual-gRNAs and HBV-expressing vector using polymerase chain reaction (PCR) and sequencing method, and the destruction of cccDNA was examined in HepAD38 cells using KCl precipitation, plasmid-safe ATP-dependent DNase (PSAD) digestion, rolling circle amplification and quantitative PCR combined method. The cytotoxicity of these gRNAs was assessed by a mitochondrial tetrazolium assay. All of gRNAs could significantly reduce HBsAg or HBeAg production in the culture supernatant, which was dependent on the region in which gRNA against. All of dual gRNAs could efficiently suppress HBsAg and/or HBeAg production for HBV of genotypes A-D, and the efficacy of dual gRNAs in suppressing HBsAg and/or HBeAg production was significantly increased when compared to the single gRNA used alone. Furthermore, by PCR direct sequencing we confirmed that these dual gRNAs could specifically destroy HBV expressing template by removing the fragment between the cleavage sites of the two used gRNAs. Most importantly, gRNA-5 and gRNA-12 combination not only could efficiently suppress HBsAg and/or HBeAg production, but also destroy the cccDNA reservoirs in HepAD38 cells.

[00571] Karimova et al. (Sci Rep. 2015 Sep 3;5:13734. doi: 10.1038/srep13734) identified cross-genotype conserved HBV sequences in the S and X region of the HBV genome that were targeted for specific and effective cleavage by a Cas nickase. This approach disrupted not only episomal cccDNA and chromosomally integrated HBV target sites in reporter cell lines, but also HBV replication in chronically and de novo infected hepatoma cell lines.

[00572] One of skill in the art may utilize the above studies of, for example, Lin et al. (Mol Ther Nucleic Acids. 2014 Aug 19;3:e186. doi: 10.1038/mtna.2014.38), Dong et al. (Antiviral Res. 2015 Jun;118:110-7. doi: 10.1016/j.antiviral.2015.03.015. Epub 2015 Apr 3), Liu et al. (J Gen Virol. 2015 Aug;96(8):2252-61. doi: 10.1099/vir.0.000159. Epub 2015 Apr 22), Wang et al. (World J Gastroenterol. 2015 Aug 28;21(32):9554-65. doi: 10.3748/wjg.v21.i32.9554) and

Karimova et al. (Sci Rep. 2015 Sep 3;5:13734. doi: 10.1038/srep13734) for targeting HBV with the CRISPR Cas system of the present invention.

[00573] Chronic hepatitis B virus (HBV) infection is prevalent, deadly, and seldom cured due to the persistence of viral episomal DNA (cccDNA) in infected cells. Ramanan et al. (Ramanan V, Shlomai A, Cox DB, Schwartz RE, Michailidis E, Bhatta A, Scott DA, Zhang F, Rice CM, Bhatia SN, .Sci Rep. 2015 Jun 2;5:10833. doi: 10.1038/srep10833, published online 2nd June 2015.) showed that the CRISPR/Cas system can specifically target and cleave conserved regions in the HBV genome, resulting in robust suppression of viral gene expression and replication. Upon sustained expression of Cas and appropriately chosen guide RNAs, they demonstrated cleavage of cccDNA by Cas and a dramatic reduction in both cccDNA and other parameters of viral gene expression and replication. Thus, they showed that directly targeting viral episomal DNA is a novel therapeutic approach to control the virus and possibly cure patients. This is also described in WO2015089465 A1, in the name of The Broad Institute et al., the contents of which are hereby incorporated by reference

[00574] As such targeting viral episomal DNA in HBV is preferred in some embodiments.

[00575] The present invention may also be applied to treat pathogens, e.g. bacterial, fungal and parasitic pathogens. Most research efforts have focused on developing new antibiotics, which once developed, would nevertheless be subject to the same problems of drug resistance. The invention provides novel CRISPR-based alternatives which overcome those difficulties. Furthermore, unlike existing antibiotics, CRISPR-based treatments can be made pathogen specific, inducing bacterial cell death of a target pathogen while avoiding beneficial bacteria.

[00576] The present invention may also be applied to treat hepatitis C virus (HCV). The methods of Roelvinki et al. (Molecular Therapy vol. 20 no. 9, 1737-1749 Sep 2012) may be applied to the CRISPR Cas system. For example, an AAV vector such as AAV8 may be a contemplated vector and for example a dosage of about 1.25 × 1011 to 1.25 × 1013 vector genomes per kilogram body weight (vg/kg) may be contemplated. The present invention may also be applied to treat pathogens, e.g. bacterial, fungal and parasitic pathogens. Most research efforts have focused on developing new antibiotics, which once developed, would nevertheless be subject to the same problems of drug resistance. The invention provides novel CRISPR-based alternatives which overcome those difficulties. Furthermore, unlike existing antibiotics, CRISPR-based treatments can be made pathogen specific, inducing bacterial cell death of a target pathogen while avoiding beneficial bacteria.

[00577] Jiang et al. ("RNA-guided editing of bacterial genomes using CRISPR-Cas systems," Nature Biotechnology vol. 31, p. 233-9, March 2013) used a CRISPR-Cas system to

mutate or kill S. pneumoniae and E. coli. The work, which introduced precise mutations into the genomes, relied on dual-RNA:Cas-directed cleavage at the targeted genomic site to kill unmutated cells and circumvented the need for selectable markers or counter-selection systems. CRISPR systems have be used to reverse antibiotic resistance and eliminate the transfer of resistance between strains. Bickard et al. showed that Cas, reprogrammed to target virulence genes, kills virulent, but not avirulent, S. aureus. Reprogramming the nuclease to target antibiotic resistance genes destroyed staphylococcal plasmids that harbor antibiotic resistance genes and immunized against the spread of plasmid-borne resistance genes. (see, Bikard et al., "Exploiting CRISPR-Cas nucleases to produce sequence-specific antimicrobials," Nature Biotechnology vol. 32, 1146–1150, doi:10.1038/nbt.3043, published online 05 October 2014.) Bikard showed that CRISPR-Cas antimicrobials function in vivo to kill S. aureus in a mouse skin colonization model. Similarly, Yosef et al used a CRISPR system to target genes encoding enzymes that confer resistance to β-lactam antibiotics (see Yousef et al., "Temperate and lytic bacteriophages programmed to sensitize and kill antibiotic-resistant bacteria," Proc. Natl. Acad. Sci. USA, vol. 112, p. 7267–7272, doi: 10.1073/pnas.1500107112 published online May 18, 2015).

[00578] CRISPR systems can be used to edit genomes of parasites that are resistant to other genetic approaches. For example, a CRISPR-Cas system was shown to introduce double-stranded breaks into the in the Plasmodium yoelii genome (see, Zhang et al., "Efficient Editing of Malaria Parasite Genome Using the CRISPR/Cas System," mBio. vol. 5, e01414-14, Jul-Aug 2014). Ghorbal et al. ("Genome editing in the human malaria parasite Plasmodium falciparumusing the CRISPR-Cas system," Nature Biotechnology, vol. 32, p. 819-821, doi: 10.1038/nbt.2925, published online June 1, 2014) modified the sequences of two genes, orc1 and kelch13, which have putative roles in gene silencing and emerging resistance to artemisinin, respectively. Parasites that were altered at the appropriate sites were recovered with very high efficiency, despite there being no direct selection for the modification, indicating that neutral or even deleterious mutations can be generated using this system. CRISPR-Cas is also used to modify the genomes of other pathogenic parasites, including Toxoplasma gondii (see Shen et al., "Efficient gene disruption in diverse strains of Toxoplasma gondii using CRISPR/CAS9," mBio vol. 5:e01114-14, 2014; and Sidik et al., "Efficient Genome Engineering of Toxoplasma gondii Using CRISPR/Cas," PLoS One vol. 9, e100450, doi: 10.1371/journal.pone.0100450, published online June 27, 2014).

[00579] Vyas et al. ("A Candida albicans CRISPR system permits genetic engineering of essential genes and gene families," Science Advances, vol. 1, e1500248, DOI:

10.1126/sciadv.1500248, April 3, 2015) employed a CRISPR system to overcome long-standing obstacles to genetic engineering in C. albicans and efficiently mutate in a single experiment both copies of several different genes. In an organism where several mechanisms contribute to drug resistance, Vyas produced homozygous double mutants that no longer displayed the hyper-resistance to fluconazole or cycloheximide displayed by the parental clinical isolate Can90. Vyas also obtained homozygous loss-of-function mutations in essential genes of C. albicans by creating conditional alleles. Null alleles of DCR1, which is required for ribosomal RNA processing, are lethal at low temperature but viable at high temperature. Vyas used a repair template that introduced a nonsense mutation and isolated dcr1/dcr1 mutants that failed to grow at 16°C.

[00580] Treating Diseases with Genetic or Epigenetic Aspects

[00581] The CRISPR-Cas systems of the present invention can be used to correct genetic mutations that were previously attempted with limited success using TALEN and ZFN and have been identified as potential targets for Cas systems, including as in published applications of Editas Medicine describing methods to use Cas systems to target loci to therapeutically address disesaes with gene therapy, including, WO 2015/048577 CRISPR-RELATED METHODS AND COMPOSITIONS of Gluckmann et al.; WO 2015/070083 CRISPR-RELATED METHODS AND COMPOSITIONS WITH GOVERNING gRNAS of Glucksmann et al.; In some embodiments, the treatment, prophylaxis or diagnosis of Primary Open Angle Glaucoma (POAG) is provided. The target is preferably the MYOC gene. This is described in WO2015153780, the disclosure of which is hereby incorporated by reference.

[00582] Mention is made of WO2015/134812 CRISPR/CAS-RELATED METHODS AND COMPOSITIONS FOR TREATING USHER SYNDROME AND RETINITIS PIGMENTOSA of Maeder et al. Through the teachings herein the invention comprehends methods and materials of these documents applied in conjunction with the teachings herein. In an aspect of ocular and auditory gene therapy, methods and compositions for treating Usher Syndrome and Retinis-Pigmentosa may be adapted to the CRISPR-Cas system of the present invention (see, e.g., WO 2015/134812). In an embodiment, the WO 2015/134812 involves a treatment or delaying the onset or progression of Usher Syndrome type IIA (USH2A, USH11A) and retinitis pigmentosa 39 (RP39) by gene editing, e.g., using CRISPR-Cas mediated methods to correct the guanine deletion at position 2299 in the USH2A gene (e.g., replace the deleted guanine residue at position 2299 in the USH2A gene). In a related aspect, a mutation is targeted by cleaving with either one or more nuclease, one or more nickase, or a combination thereof, e.g., to induce HDR with a donor template that corrects the point mutation

(e.g., the single nucleotide, e.g., guanine, deletion). The alteration or correction of the mutant USH2A gene can be mediated by any mechanism. Exemplary mechanisms that can be associated with the alteration (e.g., correction) of the mutant HSH2A gene include, but are not limited to, non-homologous end joining, microhomology-mediated end joining (MMEJ), homology-directed repair (e.g., endogenous donor template mediated), SDSA (synthesis dependent strand annealing), single-strand annealing or single strand invasion. In an embodiment, the method used for treating Usher Syndrome and Retinis-Pigmentosa can include acquiring knowledge of the mutation carried by the subject, e.g., by sequencing the appropriate portion of the USH2A gene.

[00583] Accordingly, in some embodiments, the treatment, prophylaxis or diagnosis of Retinitis Pigmentosa is provided. A number of different genes are known to be associated with or result in Retinitis Pigmentosa, such as RP1, RP2 and so forth. These genes are targeted in some embodiments and either knocked out or repaired through provision of suitable a template. In some embodiments, delivery is to the eye by injection.

[00584] One or more Retinitis Pigmentosa genes can, in some embodiments, be selected from: RP1 (Retinitis pigmentosa-1), RP2 (Retinitis pigmentosa-2), RPGR (Retinitis pigmentosa-3), PRPH2 (Retinitis pigmentosa-7), RP9 (Retinitis pigmentosa-9), IMPDH1 (Retinitis pigmentosa-10), PRPF31 (Retinitis pigmentosa-11), CRB1 (Retinitis pigmentosa-12, autosomal recessive), PRPF8 (Retinitis pigmentosa-13), TULP1 (Retinitis pigmentosa-14), CA4 (Retinitis pigmentosa-17), HPRPF3 (Retinitis pigmentosa-18), ABCA4 (Retinitis pigmentosa-19), EYS (Retinitis pigmentosa-25), CERKL (Retinitis pigmentosa-26), FSCN2 (Retinitis pigmentosa-30), TOPORS (Retinitis pigmentosa-31), SNRNP200 (Retinitis pigmentosa 33), SEMA4A (Retinitis pigmentosa-35), PRCD (Retinitis pigmentosa-36), NR2E3 (Retinitis pigmentosa-37), MERTK (Retinitis pigmentosa-38), USH2A (Retinitis pigmentosa-39), PROM1 (Retinitis pigmentosa-41), KLHL7 (Retinitis pigmentosa-42), CNGB1 (Retinitis pigmentosa-45), BEST1 (Retinitis pigmentosa-50), TTC8 (Retinitis pigmentosa 51), C2orf71 (Retinitis pigmentosa 54), ARL6 (Retinitis pigmentosa 55), ZNF513 (Retinitis pigmentosa 58), DHDDS (Retinitis pigmentosa 59), BEST1 (Retinitis pigmentosa, concentric), PRPH2 (Retinitis pigmentosa, digenic), LRAT (Retinitis pigmentosa, juvenile), SPATA7 (Retinitis pigmentosa, juvenile, autosomal recessive), CRX (Retinitis pigmentosa, late-onset dominant), and/or RPGR (Retinitis pigmentosa, X-linked, and sinorespiratory infections, with or without deafness).

[00585] In some embodiments, the Retinitis Pigmentosa gene is MERTK (Retinitis pigmentosa-38) or USH2A (Retinitis pigmentosa-39).

[00586] Mention is also made of WO 2015/138510 and through the teachings herein the invention (using a CRISPR-Cas system) comprehends providing a treatment or delaying the onset or progression of Leber's Congenital Amaurosis 10 (LCA 10). LCA 10 is caused by a mutation in the CEP290 gene, e.g., a c.2991+1655, adenine to guanine mutation in the CEP290 gene which gives rise to a cryptic splice site in intron 26. This is a mutation at nucleotide 1655 of intron 26 of CEP290, e.g., an A to G mutation. CEP290 is also known as: CT87; MKS4; POC3; rd16; BBS14; JBTS5; LCAJO; NPHP6; SLSN6; and 3H11Ag (see, e.g., WO 2015/138510). In an aspect of gene therapy, the invention involves introducing one or more breaks near the site of the LCA target position (e.g., c.2991 + 1655; A to G) in at least one allele of the CEP290 gene. Altering the LCA10 target position refers to (1) break-induced introduction of an indel (also referred to herein as NHEJ-mediated introduction of an indel) in close proximity to or including a LCA10 target position (e.g., c.2991+1655A to G), or (2) break-induced deletion (also referred to herein as NHEJ-mediated deletion) of genomic sequence including the mutation at a LCA10 target position (e.g., c.2991+1655A to G). Both approaches give rise to the loss or destruction of the cryptic splice site resulting from the mutation at the LCA 10 target position. Accordingly, the use of Cas in the treatment of LCA is specifically envisaged.

[00587] Treating Diseases of the Circulatory System

[00588] The present invention also contemplates delivering the CRISPR-Cas system, specifically the novel CRISPR effector protein systems described herein, to the blood or hematopoetic stem cells. The plasma exosomes of Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130) were previously described and may be utilized to deliver the CRISPR Cas system to the blood. The nucleic acid-targeting system of the present invention is also contemplated to treat hemoglobinopathies, such as thalassemias and sickle cell disease. See, e.g., International Patent Publication No. WO 2013/126794 for potential targets that may be targeted by the CRISPR Cas system of the present invention.

[00589] Drakopoulou, "Review Article, The Ongoing Challenge of Hematopoietic Stem Cell-Based Gene Therapy for β-Thalassemia," Stem Cells International, Volume 2011, Article ID 987980, 10 pages, doi:10.4061/2011/987980, incorporated herein by reference along with the documents it cites, as if set out in full, discuss modifying HSCs using a lentivirus that delivers a gene for β-globin or γ-globin. In contrast to using lentivirus, with the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to β-Thalassemia using a CRISPR-Cas system that targets and corrects the mutation (e.g., with a suitable HDR template that delivers a coding sequence for β-globin or γ-globin,

advantageously non-sickling β-globin or γ-globin); specifically, the guide RNA can target mutation that give rise to β-Thalassemia, and the HDR can provide coding for proper expression of β-globin or γ-globin. An guide RNA that targets the mutation-and-Cas protein containing particle is contacted with HSCs carrying the mutation. The particle also can contain a suitable HDR template to correct the mutation for proper expression of β-globin or γ-globin; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. The so contacted cells can be administered; and optionally treated / expanded; cf. Cartier. In this regard mention is made of: Cavazzana, "Outcomes of Gene Therapy for β-Thalassemia Major via Transplantation of Autologous Hematopoietic Stem Cells Transduced Ex Vivo with a Lentiviral βA-T87Q-Globin Vector." Cavazzana-Calvo, "Transfusion independence and HMGA2 activation after gene therapy of human β-thalassemia", Nature 467, 318–322 (16 September 2010) doi:10.1038/nature09328; Nienhuis, "Development of Gene Therapy for Thalassemia, Cold Spring Harbor Perpsectives in Medicine, doi: 10.1101/cshperspect.a011833 (2012), LentiGlobin BB305, a lentiviral vector containing an engineered β-globin gene (βA-T87Q); and Xie et al., "Seamless gene correction of β-thalassaemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyback" Genome Research gr.173427.114 (2014) www.genome.org/cgi/doi/10.1101/gr.173427.114 (Cold Spring Harbor Laboratory Press); that is the subject of Cavazzana work involving human β-thalassaemia and the subject of the Xie work, are all incorporated herein by reference, together with all documents cited therein or associated therewith. In the instant invention, the HDR template can provide for the HSC to express an engineered β-globin gene (e.g., βA-T87Q), or β-globin as in Xie.

[00590] Xu et al. (Sci Rep. 2015 Jul 9;5:12065. doi: 10.1038/srep12065) have designed TALENs and CRISPR-Cas to directly target the intron2 mutation site IVS2-654 in the globin gene. Xu et al. observed different frequencies of double-strand breaks (DSBs) at IVS2-654 loci using TALENs and CRISPR-Cas, and TALENs mediated a higher homologous gene targeting efficiency compared to CRISPR-Cas when combined with the piggyBac transposon donor. In addition, more obvious off-target events were observed for CRISPR-Cas compared to TALENs. Finally, TALENs-corrected iPSC clones were selected for erythroblast differentiation using the OP9 co-culture system and detected relatively higher transcription of HBB than the uncorrected cells.

[00591] Song et al. (Stem Cells Dev. 2015 May 1;24(9):1053-65. doi: 10.1089/scd.2014.0347. Epub 2015 Feb 5) used CRISPR/ Cas to correct β-Thal iPSCs; gene-corrected cells exhibit normal karyotypes and full pluripotency as human embryonic stem cells

(hESCs) showed no off-targeting effects. Then, Song et al. evaluated the differentiation efficiency of the gene-corrected β-Thal iPSCs. Song et al. found that during hematopoietic differentiation, gene-corrected β-Thal iPSCs showed an increased embryoid body ratio and various hematopoietic progenitor cell percentages. More importantly, the gene-corrected β-Thal iPSC lines restored HBB expression and reduced reactive oxygen species production compared with the uncorrected group. Song et al.'s study suggested that hematopoietic differentiation efficiency of β-Thal iPSCs was greatly improved once corrected by the CRISPR-Cas system. Similar methods may be performed utilizing the CRISPR-Cas systems described herein. Sickle cell anemia is an autosomal recessive genetic disease in which red blood cells become sickle-shaped. It is caused by a single base substitution in the β-globin gene, which is located on the short arm of chromosome 11. As a result, valine is produced instead of glutamic acid causing the production of sickle hemoglobin (HbS). This results in the formation of a distorted shape of the erythrocytes. Due to this abnormal shape, small blood vessels can be blocked, causing serious damage to the bone, spleen and skin tissues. This may lead to episodes of pain, frequent infections, hand-foot syndrome or even multiple organ failure. The distorted erythrocytes are also more susceptible to hemolysis, which leads to serious anemia. As in the case of β-thalassemia, sickle cell anemia can be corrected by modifying HSCs with the CRISPR-Cas system. The system allows the specific editing of the cell's genome by cutting its DNA and then letting it repair itself. The Cas protein is inserted and directed by a RNA guide to the mutated point and then it cuts the DNA at that point. Simultaneously, a healthy version of the sequence is inserted. This sequence is used by the cell's own repair system to fix the induced cut. In this way, the CRISPR-Cas allows the correction of the mutation in the previously obtained stem cells. With the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to sickle cell anemia using a CRISPR-Cas system that targets and corrects the mutation (e.g., with a suitable HDR template that delivers a coding sequence for β-globin, advantageously non-sickling β-globin); specifically, the guide RNA can target mutation that give rise to sickle cell anemia, and the HDR can provide coding for proper expression of β-globin. An guide RNA that targets the mutation-and-Cas protein containing particle is contacted with HSCs carrying the mutation. The particle also can contain a suitable HDR template to correct the mutation for proper expression of β-globin; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. The so contacted cells can be administered; and optionally treated / expanded; cf. Cartier. The HDR template can provide for the HSC to express an engineered β-globin gene (e.g., βA-T87Q), or β-globin as in Xie.

[00592] Williams, "Broadening the Indications for Hematopoietic Stem Cell Genetic Therapies," Cell Stem Cell 13:263-264 (2013), incorporated herein by reference along with the documents it cites, as if set out in full, report lentivirus-mediated gene transfer into HSC/P cells from patients with the lysosomal storage disease metachromatic leukodystrophy disease (MLD), a genetic disease caused by deficiency of arylsulfatase A (ARSA), resulting in nerve demyelination; and lentivirus-mediated gene transfer into HSCs of patients with Wiskott-Aldrich syndrome (WAS) (patients with defective WAS protein, an effector of the small GTPase CDC42 that regulates cytoskeletal function in blood cell lineages and thus suffer from immune deficiency with recurrent infections, autoimmune symptoms, and thrombocytopenia with abnormally small and dysfunctional platelets leading to excessive bleeding and an increased risk of leukemia and lymphoma). In contrast to using lentivirus, with the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to MLD (deficiency of arylsulfatase A (ARSA)) using a CRISPR-Cas system that targets and corrects the mutation (deficiency of arylsulfatase A (ARSA)) (e.g., with a suitable HDR template that delivers a coding sequence for ARSA); specifically, the guide RNA can target mutation that gives rise to MLD (deficient ARSA), and the HDR can provide coding for proper expression of ARSA. An guide RNA that targets the mutation-and-Cas protein containing particle is contacted with HSCs carrying the mutation. The particle also can contain a suitable HDR template to correct the mutation for proper expression of ARSA; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. The so contacted cells can be administered; and optionally treated / expanded; cf. Cartier. In contrast to using lentivirus, with the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to WAS using a CRISPR-Cas system that targets and corrects the mutation (deficiency of WAS protein) (e.g., with a suitable HDR template that delivers a coding sequence for WAS protein); specifically, the guide RNA can target mutation that gives rise to WAS (deficient WAS protein), and the HDR can provide coding for proper expression of WAS protein. An guide RNA that targets the mutation-and-Cas protein containing particle is contacted with HSCs carrying the mutation. The particle also can contain a suitable HDR template to correct the mutation for proper expression of WAS protein; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. The so contacted cells can be administered; and optionally treated / expanded; cf. Cartier.

[00593] Watts, "Hematopoietic Stem Cell Expansion and Gene Therapy" Cytotherapy 13(10):1164–1171. doi:10.3109/14653249.2011.620748 (2011), incorporated herein by reference along with the documents it cites, as if set out in full, discusses hematopoietic stem

cell (HSC) gene therapy, e.g., virus-mediated HSC gene therapy, as an highly attractive treatment option for many disorders including hematologic conditions, immunodeficiencies including HIV/AIDS, and other genetic disorders like lysosomal storage diseases, including SCID-X1, ADA-SCID, β-thalassemia, X-linked CGD, Wiskott-Aldrich syndrome, Fanconi anemia, adrenoleukodystrophy (ALD), and metachromatic leukodystrophy (MLD).

[00594] US Patent Publication Nos. 20110225664, 20110091441, 20100229252, 20090271881 and 20090222937 assigned to Cellectis, relates to CREI variants , wherein at least one of the two I-CreI monomers has at least two substitutions, one in each of the two functional subdomains of the LAGLIDADG (SEQ ID NO:494) core domain situated respectively from positions 26 to 40 and 44 to 77 of I-CreI, said variant being able to cleave a DNA target sequence from the human interleukin-2 receptor gamma chain (IL2RG) gene also named common cytokine receptor gamma chain gene or gamma C gene. The target sequences identified in US Patent Publication Nos. 20110225664, 20110091441, 20100229252, 20090271881 and 20090222937 may be utilized for the nucleic acid-targeting system of the present invention.

[00595] Severe Combined Immune Deficiency (SCID) results from a defect in lymphocytes T maturation, always associated with a functional defect in lymphocytes B (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). Overall incidence is estimated to 1 in 75 000 births. Patients with untreated SCID are subject to multiple opportunist micro-organism infections, and do generally not live beyond one year. SCID can be treated by allogenic hematopoietic stem cell transfer, from a familial donor. Histocompatibility with the donor can vary widely. In the case of Adenosine Deaminase (ADA) deficiency, one of the SCID forms, patients can be treated by injection of recombinant Adenosine Deaminase enzyme.

[00596] Since the ADA gene has been shown to be mutated in SCID patients (Giblett et al., Lancet, 1972, 2, 1067-1069), several other genes involved in SCID have been identified (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). There are four major causes for SCID: (i) the most frequent form of SCID, SCID-X1 (X-linked SCID or X-SCID), is caused by mutation in the IL2RG gene, resulting in the absence of mature T lymphocytes and NK cells. IL2RG encodes the gamma C protein (Noguchi, et al., Cell, 1993, 73, 147-157), a common component of at least five interleukin receptor complexes. These receptors activate several targets through the JAK3 kinase (Macchi et al., Nature, 1995, 377, 65-68), which inactivation results in the same syndrome as gamma C inactivation; (ii) mutation in the ADA gene results in a defect in purine metabolism that is

lethal for lymphocyte precursors, which in turn results in the quasi absence of B, T and NK cells; (iii) V(D)J recombination is an essential step in the maturation of immunoglobulins and T lymphocytes receptors (TCRs). Mutations in Recombination Activating Gene 1 and 2 (RAG1 and RAG2) and Artemis, three genes involved in this process, result in the absence of mature T and B lymphocytes; and (iv) Mutations in other genes such as CD45, involved in T cell specific signaling have also been reported, although they represent a minority of cases (Cavazzana-Calvo et al., Annu. Rev. Med., 2005, 56, 585-602; Fischer et al., Immunol. Rev., 2005, 203, 98-109). Since when their genetic bases have been identified, the different SCID forms have become a paradigm for gene therapy approaches (Fischer et al., Immunol. Rev., 2005, 203, 98-109) for two major reasons. First, as in all blood diseases, an ex vivo treatment can be envisioned. Hematopoietic Stem Cells (HSCs) can be recovered from bone marrow, and keep their pluripotent properties for a few cell divisions. Therefore, they can be treated in vitro, and then reinjected into the patient, where they repopulate the bone marrow. Second, since the maturation of lymphocytes is impaired in SCID patients, corrected cells have a selective advantage. Therefore, a small number of corrected cells can restore a functional immune system. This hypothesis was validated several times by (i) the partial restoration of immune functions associated with the reversion of mutations in SCID patients (Hirschhorn et al., Nat. Genet., 1996, 13, 290-295; Stephan et al., N. Engl. J. Med., 1996, 335, 1563-1567; Bousso et al., Proc. Natl., Acad. Sci. USA, 2000, 97, 274-278; Wada et al., Proc. Natl. Acad. Sci. USA, 2001, 98, 8697-8702; Nishikomori et al., Blood, 2004, 103, 4565-4572), (ii) the correction of SCID-X1 deficiencies in vitro in hematopoietic cells (Candotti et al., Blood, 1996, 87, 3097-3102; Cavazzana-Calvo et al., Blood, 1996, Blood, 88, 3901-3909; Taylor et al., Blood, 1996, 87, 3103-3107; Hacein-Bey et al., Blood, 1998, 92, 4090-4097), (iii) the correction of SCID-X1 (Soudais et al., Blood, 2000, 95, 3071-3077; Tsai et al., Blood, 2002, 100, 72-79), JAK-3 (Bunting et al., Nat. Med., 1998, 4, 58-64; Bunting et al., Hum. Gene Ther., 2000, 11, 2353-2364) and RAG2 (Yates et al., Blood, 2002, 100, 3942-3949) deficiencies in vivo in animal models and (iv) by the result of gene therapy clinical trials (Cavazzana-Calvo et al., Science, 2000, 288, 669-672; Aiuti et al., Nat. Med., 2002; 8, 423-425; Gaspar et al., Lancet, 2004, 364, 2181-2187).

[00597] US Patent Publication No. 20110182867 assigned to the Children's Medical Center Corporation and the President and Fellows of Harvard College relates to methods and uses of modulating fetal hemoglobin expression (HbF) in a hematopoietic progenitor cells via inhibitors of BCL11A expression or activity, such as RNAi and antibodies. The targets disclosed in US Patent Publication No. 20110182867, such as BCL11A, may be targeted by

the CRISPR Cas system of the present invention for modulating fetal hemoglobin expression. See also Bauer et al. (Science 11 October 2013: Vol. 342 no. 6155 pp. 253-257) and Xu et al. (Science 18 November 2011: Vol. 334 no. 6058 pp. 993-996) for additional BCL11A targets.

[00598] With the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to a genetic hematologic disorder, e.g., β-Thalassemia, Hemophilia, or a genetic lysosomal storage disease.

[00599] HSC—Delivery to and Editing of Hematopoietic Stem Cells; and Particular Conditions.

[00600] The term "Hematopoietic Stem Cell" or "HSC" is meant to include broadly those cells considered to be an HSC, e.g., blood cells that give rise to all the other blood cells and are derived from mesoderm; located in the red bone marrow, which is contained in the core of most bones. HSCs of the invention include cells having a phenotype of hematopoietic stem cells, identified by small size, lack of lineage (lin) markers, and markers that belong to the cluster of differentiation series, like: CD34, CD38, CD90, CD133, CD105, CD45, and also c-kit, - the receptor for stem cell factor. Hematopoietic stem cells are negative for the markers that are used for detection of lineage commitment, and are, thus, called Lin-; and, during their purification by FACS, a number of up to 14 different mature blood-lineage markers, e.g., CD13 & CD33 for myeloid, CD71 for erythroid, CD19 for B cells, CD61 for megakaryocytic, etc. for humans; and, B220 (murine CD45) for B cells, Mac-1 (CD11b/CD18) for monocytes, Gr-1 for Granulocytes, Ter119 for erythroid cells, Il7Ra, CD3, CD4, CD5, CD8 for T cells, etc. Mouse HSC markers: CD34lo/-, SCA-1+, Thy1.1+/lo, CD38+, C-kit+, lin-, and Human HSC markers: CD34+, CD59+, Thy1/CD90+, CD38lo/-, C-kit/CD117+, and lin-. HSCs are identified by markers. Hence in embodiments discussed herein, the HSCs can be CD34+ cells. HSCs can also be hematopoietic stem cells that are CD34-/CD38-. Stem cells that may lack c-kit on the cell surface that are considered in the art as HSCs are within the ambit of the invention, as well as CD133+ cells likewise considered HSCs in the art.

[00601] The CRISPR-Cas system may be engineered to target genetic locus or loci in HSCs. Cas protein, advantageously codon-optimized for a eukaryotic cell and especially a mammalian cell, e.g., a human cell, for instance, HSC, and sgRNA targeting a locus or loci in HSC, e.g., the gene EMX1, may be prepared. These may be delivered via particles. The particles may be formed by the Cas protein and the gRNA being admixed. The gRNA and Cas protein mixture may for example be admixed with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol,

whereby particles containing the gRNA and Cas protein may be formed. The invention comprehends so making particles and particles from such a method as well as uses thereof.

[00602] More generally, particles may be formed using an efficient process. First, Cas protein and gRNA targeting the gene EMX1 or the control gene LacZ may be mixed together at a suitable, e.g.,3:1 to 1:3 or 2:1 to 1:2 or 1:1 molar ratio, at a suitable temperature, e.g., 15-30C, e.g., 20-25C, e.g., room temperature, for a suitable time, e.g., 15-45, such as 30 minutes, advantageously in sterile, nuclease free buffer, e.g., 1X PBS. Separately, particle components such as or comprising: a surfactant, e.g., cationic lipid, e.g., 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP); phospholipid, e.g., dimyristoylphosphatidylcholine (DMPC); biodegradable polymer, such as an ethylene-glycol polymer or PEG, and a lipoprotein, such as a low-density lipoprotein, e.g., cholesterol may be dissolved in an alcohol, advantageously a C1-6 alkyl alcohol, such as methanol, ethanol, isopropanol, e.g., 100% ethanol. The two solutions may be mixed together to form particles containing the Cas-gRNA complexes. In certain embodiments the particle can contain an HDR template. That can be a particle co-administered with gRNA+Cas protein-containing particle, or i.e., in addition to contacting an HSC with an gRNA+Cas protein-containing particle, the HSC is contacted with a particle containing an HDR template; or the HSC is contacted with a particle containing all of the gRNA, Cas and the HDR template. The HDR template can be administered by a separate vector, whereby in a first instance the particle penetrates an HSC cell and the separate vector also penetrates the cell, wherein the HSC genome is modified by the gRNA+Cas and the HDR template is also present, whereby a genomic loci is modified by the HDR; for instance, this may result in correcting a mutation.

[00603] After the particles form, HSCs in 96 well plates may be transfected with 15ug Cas protein per well. Three days after transfection, HSCs may be harvested, and the number of insertions and deletions (indels) at the EMX1 locus may be quantified.

[00604] This illustrates how HSCs can be modified using CRISPR-Cas targeting a genomic locus or loci of interest in the HSC. The HSCs that are to be modified can be in vivo, i.e., in an organism, for example a human or a non-human eukaryote, e.g., animal, such as fish, e.g., zebra fish, mammal, e.g., primate, e.g., ape, chimpanzee, macaque, rodent, e.g., mouse, rabbit, rat, canine or dog, livestock (cow / bovine, sheep / ovine, goat or pig), fowl or poultry, e.g., chicken. The HSCs that are to be modified can be in vitro, i.e., outside of such an organism. And, modified HSCs can be used ex vivo, i.e., one or more HSCs of such an organism can be obtained or isolated from the organism, optionally the HSC(s) can be expanded, the HSC(s) are modified by a composition comprising a CRISPR-Cas that targets a genetic locus or loci in

the HSC, e.g., by contacting the HSC(s) with the composition, for instance, wherein the composition comprises a particle containing the CRISPR enzyme and one or more gRNA that targets the genetic locus or loci in the HSC, such as a particle obtained or obtainable from admixing an gRNA and Cas protein mixture with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol (wherein one or more gRNA targets the genetic locus or loci in the HSC), optionally expanding the resultant modified HSCs and administering to the organism the resultant modified HSCs. In some instances the isolated or obtained HSCs can be from a first organism, such as an organism from a same species as a second organism, and the second organism can be the organism to which the resultant modified HSCs are administered, e.g., the first organism can be a donor (such as a relative as in a parent or sibling) to the second organism. Modified HSCs can have genetic modifications to address or alleviate or reduce symptoms of a disease or condition state of an individual or subject or patient. Modified HSCs, e.g., in the instance of a first organism donor to a second organism, can have genetic modifications to have the HSCs have one or more proteins e.g. surface markers or proteins more like that of the second organism. Modified HSCs can have genetic modifications to simulate a disease or condition state of an individual or subject or patient and would be re-administered to a non-human organism so as to prepare an animal model. Expansion of HSCs is within the ambit of the skilled person from this disclosure and knowledge in the art, see e.g., Lee, "Improved ex vivo expansion of adult hematopoietic stem cells by overcoming CUL4-mediated degradation of HOXB4." Blood. 2013 May 16;121(20):4082-9. doi: 10.1182/blood-2012-09-455204. Epub 2013 Mar 21.

[00605] As indicated to improve activity, gRNA may be pre-complexed with the Cas protein, before formulating the entire complex in a particle. Formulations may be made with a different molar ratio of different components known to promote delivery of nucleic acids into cells (e.g. 1,2-dioleoyl-3-trimethylammonium-propane (DOTAP), 1,2-ditetradecanoyl-sn-glycero-3-phosphocholine (DMPC), polyethylene glycol (PEG), and cholesterol) For example DOTAP : DMPC : PEG : Cholesterol Molar Ratios may be DOTAP 100, DMPC 0, PEG 0, Cholesterol 0; or DOTAP 90, DMPC 0, PEG 10, Cholesterol 0; or DOTAP 90, DMPC 0, PEG 5, Cholesterol 5. DOTAP 100, DMPC 0, PEG 0, Cholesterol 0. The invention accordingly comprehends admixing gRNA, Cas protein and components that form a particle; as well as particles from such admixing.

[00606] In a preferred embodiment, particles containing the Cas-gRNA complexes may be formed by mixing Cas protein and one or more gRNAs together, preferably at a 1:1 molar ratio,

enzyme: guide RNA. Separately, the different components known to promote delivery of nucleic acids (e.g. DOTAP, DMPC, PEG, and cholesterol) are dissolved, preferably in ethanol. The two solutions are mixed together to form particles containing the Cas-gRNA complexes. After the particles are formed, Cas-gRNA complexes may be transfected into cells (e.g. HSCs). Bar coding may be applied. The particles, the Cas-9 and/or the gRNA may be barcoded.

[00607] The invention in an embodiment comprehends a method of preparing an gRNA-and-Cas protein containing particle comprising admixing an gRNA and Cas protein mixture with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol. An embodiment comprehends an gRNA-and-Cas protein containing particle from the method. The invention in an embodiment comprehends use of the particle in a method of modifying a genomic locus of interest, or an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest, comprising contacting a cell containing the genomic locus of interest with the particle wherein the gRNA targets the genomic locus of interest; or a method of modifying a genomic locus of interest, or an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest, comprising contacting a cell containing the genomic locus of interest with the particle wherein the gRNA targets the genomic locus of interest. In these embodiments, the genomic locus of interest is advantageously a genomic locus in an HSC.

[00608] Considerations for Therapeutic Applications: A consideration in genome editing therapy is the choice of sequence-specific nuclease, such as a variant of a Cas nuclease. Each nuclease variant may possess its own unique set of strengths and weaknesses, many of which must be balanced in the context of treatment to maximize therapeutic benefit. Thus far, two therapeutic editing approaches with nucleases have shown significant promise: gene disruption and gene correction. Gene disruption involves stimulation of NHEJ to create targeted indels in genetic elements, often resulting in loss of function mutations that are beneficial to patients. In contrast, gene correction uses HDR to directly reverse a disease causing mutation, restoring function while preserving physiological regulation of the corrected element. HDR may also be used to insert a therapeutic transgene into a defined 'safe harbor' locus in the genome to recover missing gene function. For a specific editing therapy to be efficacious, a sufficiently high level of modification must be achieved in target cell populations to reverse disease symptoms. This therapeutic modification 'threshold' is determined by the fitness of edited cells following treatment and the amount of gene product necessary to reverse symptoms. With regard to fitness, editing creates three potential outcomes for treated cells relative to their unedited

counterparts: increased, neutral, or decreased fitness. In the case of increased fitness, for example in the treatment of SCID-X1, modified hematopoietic progenitor cells selectively expand relative to their unedited counterparts. SCID-X1 is a disease caused by mutations in the IL2RG gene, the function of which is required for proper development of the hematopoietic lymphocyte lineage [Leonard, W.J., et al. Immunological reviews 138, 61-86 (1994); Kaushansky, K. & Williams, W.J. Williams hematology, (McGraw-Hill Medical, New York, 2010)]. In clinical trials with patients who received viral gene therapy for SCID-X1, and a rare example of a spontaneous correction of SCID-X1 mutation, corrected hematopoietic progenitor cells may be able to overcome this developmental block and expand relative to their diseased counterparts to mediate therapy [Bousso, P., et al. Proceedings of the National Academy of Sciences of the United States of America 97, 274-278 (2000); Hacein-Bey-Abina, S., et al. The New England journal of medicine 346, 1185-1193 (2002); Gaspar, H.B., et al. Lancet 364, 2181-2187 (2004)]. In this case, where edited cells possess a selective advantage, even low numbers of edited cells can be amplified through expansion, providing a therapeutic benefit to the patient. In contrast, editing for other hematopoietic diseases, like chronic granulomatous disorder (CGD), would induce no change in fitness for edited hematopoietic progenitor cells, increasing the therapeutic modification threshold. CGD is caused by mutations in genes encoding phagocytic oxidase proteins, which are normally used by neutrophils to generate reactive oxygen species that kill pathogens [Mukherjee, S. & Thrasher, A.J. Gene 525, 174-181 (2013)]. As dysfunction of these genes does not influence hematopoietic progenitor cell fitness or development, but only the ability of a mature hematopoietic cell type to fight infections, there would be likely no preferential expansion of edited cells in this disease. Indeed, no selective advantage for gene corrected cells in CGD has been observed in gene therapy trials, leading to difficulties with long-term cell engraftment [Malech, H.L., et al. Proceedings of the National Academy of Sciences of the United States of America 94, 12133-12138 (1997); Kang, H.J., et al. Molecular therapy : the journal of the American Society of Gene Therapy 19, 2092-2101 (2011)]. As such, significantly higher levels of editing would be required to treat diseases like CGD, where editing creates a neutral fitness advantage, relative to diseases where editing creates increased fitness for target cells. If editing imposes a fitness disadvantage, as would be the case for restoring function to a tumor suppressor gene in cancer cells, modified cells would be outcompeted by their diseased counterparts, causing the benefit of treatment to be low relative to editing rates. This latter class of diseases would be particularly difficult to treat with genome editing therapy.

[00609] In addition to cell fitness, the amount of gene product necessary to treat disease also influences the minimal level of therapeutic genome editing that must be achieved to reverse symptoms. Haemophilia B is one disease where a small change in gene product levels can result in significant changes in clinical outcomes. This disease is caused by mutations in the gene encoding factor IX, a protein normally secreted by the liver into the blood, where it functions as a component of the clotting cascade. Clinical severity of haemophilia B is related to the amount of factor IX activity. Whereas severe disease is associated with less than 1% of normal activity, milder forms of the diseases are associated with greater than 1% of factor IX activity [Kaushansky, K. & Williams, W.J. Williams hematology, (McGraw-Hill Medical, New York, 2010); Lofqvist, T., et al. Journal of internal medicine 241, 395-400 (1997)]. This suggests that editing therapies that can restore factor IX expression to even a small percentage of liver cells could have a large impact on clinical outcomes. A study using ZFNs to correct a mouse model of haemophilia B shortly after birth demonstrated that 3-7% correction was sufficient to reverse disease symptoms, providing preclinical evidence for this hypothesis [Li, H., et al. Nature 475, 217-221 (2011)].

[00610] Disorders where a small change in gene product levels can influence clinical outcomes and diseases where there is a fitness advantage for edited cells, are ideal targets for genome editing therapy, as the therapeutic modification threshold is low enough to permit a high chance of success given the current technology. Targeting these diseases has now resulted in successes with editing therapy at the preclinical level and a phase I clinical trial. Improvements in DSB repair pathway manipulation and nuclease delivery are needed to extend these promising results to diseases with a neutral fitness advantage for edited cells, or where larger amounts of gene product are needed for treatment. Table 10 below shows some examples of applications of genome editing to therapeutic models, and the references shown in Table 10 and the documents cited in those references are hereby incorporated herein by reference as if set out in full.

[00611] Table 10. Examples of applications of genome editing to therapeutic models.

Disease Type	Nuclease Platform Employed	Therapeutic Strategy	References
Hemophilia B	ZFN	HDR-mediated insertion of correct gene sequence	Li, H., et al. Nature 475, 217-221 (2011)
SCID	ZFN	HDR-mediated insertion of correct gene sequence	Genovese, P., et al. Nature 510, 235-240 (2014)

Disease Type	Nuclease Platform Employed	Therapeutic Strategy	References
Hereditary tyrosinemia	CRISPR	HDR-mediated correction of mutation in liver	Yin, H., et al. Nature biotechnology 32, 551-553 (2014)

[00612] Addressing each of the conditions of the foregoing table, using the CRISPR-Cas system to target by either HDR-mediated correction of mutation, or HDR-mediated insertion of correct gene sequence, advantageously via a delivery system as herein, e.g., a particle delivery system, is within the ambit of the skilled person from this disclosure and the knowledge in the art. Thus, an embodiment comprehends contacting a Hemophilia B, SCID (e.g., SCID-X1, ADA-SCID) or Hereditary tyrosinemia mutation-carrying HSC with an gRNA-and-Cas protein containing particle targeting a genomic locus of interest as to Hemophilia B, SCID (e.g., SCID-X1, ADA-SCID) or Hereditary tyrosinemia (e.g., as in Li, Genovese or Yin). The particle also can contain a suitable HDR template to correct the mutation; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. In this regard, it is mentioned that Haemophilia B is an X-linked recessive disorder caused by loss-of-function mutations in the gene encoding Factor IX, a crucial component of the clotting cascade. Recovering Factor IX activity to above 1% of its levels in severely affected individuals can transform the disease into a significantly milder form, as infusion of recombinant Factor IX into such patients prophylactically from a young age to achieve such levels largely ameliorates clinical complications. With the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to Haemophilia B using a CRISPR-Cas system that targets and corrects the mutation (X-linked recessive disorder caused by loss-of-function mutations in the gene encoding Factor IX) (e.g., with a suitable HDR template that delivers a coding sequence for Factor IX); specifically, the gRNA can target mutation that give rise to Haemophilia B, and the HDR can provide coding for proper expression of Factor IX. An gRNA that targets the mutation-and-Cas protein containing particle is contacted with HSCs carrying the mutation. The particle also can contain a suitable HDR template to correct the mutation for proper expression of Factor IX; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. The so contacted cells can be administered; and optionally treated / expanded; cf. Cartier, discussed herein.

[00613] In Cartier, "MINI-SYMPOSIUM: X-Linked Adrenoleukodystrophypa, Hematopoietic Stem Cell Transplantation and Hematopoietic Stem Cell Gene Therapy in X-

Linked Adrenoleukodystrophy," Brain Pathology 20 (2010) 857–862, incorporated herein by reference along with the documents it cites, as if set out in full, there is recognition that allogeneic hematopoietic stem cell transplantation (HSCT) was utilized to deliver normal lysosomal enzyme to the brain of a patient with Hurler's disease, and a discussion of HSC gene therapy to treat ALD. In two patients, peripheral CD34+cells were collected after granulocyte-colony stimulating factor (G-CSF) mobilization and transduced with an myeloproliferative sarcoma virus enhancer, negative control region deleted, dl587rev primer binding site substituted (MND)-ALD lentiviral vector. CD34+ cells from the patients were transduced with the MND-ALD vector during 16 h in the presence of cytokines at low concentrations. Transduced CD34+ cells were frozen after transduction to perform on 5% of cells various safety tests that included in particular three replication-competent lentivirus (RCL) assays. Transduction efficacy of CD34+ cells ranged from 35% to 50% with a mean number of lentiviral integrated copy between 0.65 and 0.70. After the thawing of transduced CD34+ cells, the patients were reinfused with more than 4.106 transduced CD34+ cells/kg following full myeloablation with busulfan and cyclophos-phamide. The patient's HSCs were ablated to favor engraftment of the gene-corrected HSCs. Hematological recovery occurred between days 13 and 15 for the two patients. Nearly complete immunological recovery occurred at 12 months for the first patient, and at 9 months for the second patient. In contrast to using lentivirus, with the knowledge in the art and the teachings in this disclosure, the skilled person can correct HSCs as to ALD using a CRISPR-Cas system that targets and corrects the mutation (e.g., with a suitable HDR template); specifically, the gRNA can target mutations in ABCD1, a gene located on the X chromosome that codes for ALD, a peroxisomal membrane transporter protein, and the HDR can provide coding for proper expression of the protein. An gRNA that targets the mutation-and-Cas protein containing particle is contacted with HSCs, e.g., CD34+ cells carrying the mutation as in Cartier. The particle also can contain a suitable HDR template to correct the mutation for expression of the peroxisomal membrane transporter protein; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. The so contacted cells optionally can be treated as in Cartier. The so contacted cells can be administered as in Cartier.

[00614] Mention is made of WO 2015/148860, through the teachings herein the invention comprehends methods and materials of these documents applied in conjunction with the teachings herein. In an aspect of blood-related disease gene therapy, methods and compositions for treating beta thalassemia may be adapted to the CRISPR-Cas system of the present invention (see, e.g., WO 2015/148860). In an embodiment, WO 2015/148860 involves

the treatment or prevention of beta thalassemia, or its symptoms, e.g., by altering the gene for B-cell CLL/lymphoma 11A (BCL11A). The BCL11A gene is also known as B-cell CLL/lymphoma 11A, BCL11A -L, BCL11A -S, BCL11AXL, CTIP 1, HBFQTL5 and ZNF. BCL11A encodes a zinc-finger protein that is involved in the regulation of globin gene expression. By altering the BCL11A gene (e.g., one or both alleles of the BCL11A gene), the levels of gamma globin can be increased. Gamma globin can replace beta globin in the hemoglobin complex and effectively carry oxygen to tissues, thereby ameliorating beta thalassemia disease phenotypes.

[00615] Mention is also made of WO 2015/148863 and through the teachings herein the invention comprehends methods and materials of these documents which may be adapted to the CRISPR-Cas system of the present invention. In an aspect of treating and preventing sickle cell disease, which is an inherited hematologic disease, WO 2015/148863 comprehends altering the BCL11A gene. By altering the BCL11A gene (e.g., one or both alleles of the BCL11A gene), the levels of gamma globin can be increased. Gamma globin can replace beta globin in the hemoglobin complex and effectively carry oxygen to tissues, thereby ameliorating sickle cell disease phenotypes.

[00616] In an aspect of the invention, methods and compositions which involve editing a target nucleic acid sequence, or modulating expression of a target nucleic acid sequence, and applications thereof in connection with cancer immunotherapy are comprehended by adapting the CRISPR-Cas system of the present invention. Reference is made to the application of gene therapy in WO 2015/161276 which involves methods and compositions which can be used to affect T-cell proliferation, survival and/or function by altering one or more T-cell expressed genes, e.g., one or more of FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, TRAC and/or TRBC genes. In a related aspect, T-cell proliferation can be affected by altering one or more T -cell expressed genes, e.g., the CBLB and/or PTPN6 gene, FAS and/ or BID gene, CTLA4 and/or PDCDI and/or TRAC and/or TRBC gene.

[00617] Chimeric antigen receptor (CAR)19 T-cells exhibit anti-leukemic effects in patient malignancies. However, leukemia patients often do not have enough T-cells to collect, meaning that treatment must involve modified T cells from donors. Accordingly, there is interest in establishing a bank of donor T-cells. Qasim et al. ("First Clinical Application of Talen Engineered Universal CAR19 T Cells in B-ALL" ASH 57th Annual Meeting and Exposition, Dec. 5-8, 2015, Abstract 2046 (ash.confex.com/ash/2015/webprogram/Paper81653.html published online November 2015) discusses modifying CAR19 T cells to eliminate the risk of graft-versus-host disease through

the disruption of T-cell receptor expression and CD52 targeting. Furthermore, CD52 cells were targeted such that they became insensitive to Alemtuzumab, and thus allowed Alemtuzumab to prevent host-mediated rejection of human leukocyte antigen (HLA) mismatched CAR19 T-cells. Investigators used third generation self-inactivating lentiviral vector encoding a 4g7 CAR19 (CD19 scFv-4-1BB-CD3ζ) linked to RQR8, then electroporated cells with two pairs of TALEN mRNA for multiplex targeting for both the T-cell receptor (TCR) alpha constant chain locus and the CD52 gene locus. Cells which were still expressing TCR following ex vivo expansion were depleted using CliniMacs α/β TCR depletion, yielding a T-cell product (UCART19) with <1% TCR expression, 85% of which expressed CAR19, and 64% becoming CD52 negative. The modified CAR19 T cells were administered to treat a patient's relapsed acute lymphoblastic leukemia. The teachings provided herein provide effective methods for providing modified hematopoietic stem cells and progeny thereof, including but not limited to cells of the myeloid and lymphoid lineages of blood, including T cells, B cells, monocytes, macrophages, neutrophils, basophils, eosinophils, erythrocytes, dendritic cells, and megakaryocytes or platelets, and natural killer cells and their precursors and progenitors. Such cells can be modified by knocking out, knocking in, or otherwise modulating targets, for example to remove or modulate CD52 as described above, and other targets, such as, without limitation, CXCR4, and PD-1. Thus compositions, cells, and method of the invention can be used to modulate immune responses and to treat, without limitation, malignancies, viral infections, and immune disorders, in conjunction with modification of administration of T cells or other cells to patients.

[00618] Mention is made of WO 2015/148670 and through the teachings herein the invention comprehends methods and materials of this document applied in conjunction with the teachings herein. In an aspect of gene therapy, methods and compositions for editing of a target sequence related to or in connection with Human Immunodeficiency Virus (HIV) and Acquired Immunodeficiency Syndrome (AIDS) are comprehended. In a related aspect, the invention described herein comprehends prevention and treatment of HIV infection and AIDS, by introducing one or more mutations in the gene for C-C chemokine receptor type 5 (CCR5). The CCR5 gene is also known as CKR5, CCR-5, CD195, CKR-5, CCCKR5, CMKBR5, IDDM22, and CC-CKR-5. In a further aspect, the invention described herein comprehends provide for prevention or reduction of HIV infection and/or prevention or reduction of the ability for HIV to enter host cells, e.g., in subjects who are already infected. Exemplary host cells for HIV include, but are not limited to, CD4 cells, T cells, gut associated lymphatic tissue (GALT), macrophages, dendritic cells, myeloid precursor cell, and microglia. Viral entry into

the host cells requires interaction of the viral glycoproteins gp41 and gp120 with both the CD4 receptor and a co-receptor, e.g., CCR5. If a co-receptor, e.g., CCR5, is not present on the surface of the host cells, the virus cannot bind and enter the host cells. The progress of the disease is thus impeded. By knocking out or knocking down CCR5 in the host cells, e.g., by introducing a protective mutation (such as a CCR5 delta 32 mutation), entry of the HIV virus into the host cells is prevented.

[00619] X-linked Chronic granulomatous disease (CGD) is a hereditary disorder of host defense due to absent or decreased activity of phagocyte NADPH oxidase. Using a CRISPR-Cas system that targets and corrects the mutation (absent or decreased activity of phagocyte NADPH oxidase) (e.g., with a suitable HDR template that delivers a coding sequence for phagocyte NADPH oxidase); specifically, the gRNA can target mutation that gives rise to CGD (deficient phagocyte NADPH oxidase), and the HDR can provide coding for proper expression of phagocyte NADPH oxidase. An gRNA that targets the mutation-and-Cas protein containing particle is contacted with HSCs carrying the mutation. The particle also can contain a suitable HDR template to correct the mutation for proper expression of phagocyte NADPH oxidase; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. The so contacted cells can be administered; and optionally treated / expanded; cf. Cartier.

[00620] Fanconi anemia: Mutations in at least 15 genes (FANCA, FANCB, FANCC, FANCD1/BRCA2, FANCD2, FANCE, FANCF, FANCG, FANCI, FANCJ/BACH1/BRIP1, FANCL/PHF9/POG, FANCM, FANCN/PALB2, FANCO/Rad51C, and FANCP/SLX4/BTBD12) can cause Fanconi anemia. Proteins produced from these genes are involved in a cell process known as the FA pathway. The FA pathway is turned on (activated) when the process of making new copies of DNA, called DNA replication, is blocked due to DNA damage. The FA pathway sends certain proteins to the area of damage, which trigger DNA repair so DNA replication can continue. The FA pathway is particularly responsive to a certain type of DNA damage known as interstrand cross-links (ICLs). ICLs occur when two DNA building blocks (nucleotides) on opposite strands of DNA are abnormally attached or linked together, which stops the process of DNA replication. ICLs can be caused by a buildup of toxic substances produced in the body or by treatment with certain cancer therapy drugs. Eight proteins associated with Fanconi anemia group together to form a complex known as the FA core complex. The FA core complex activates two proteins, called FANCD2 and FANCI. The activation of these two proteins brings DNA repair proteins to the area of the ICL so the cross-link can be removed and DNA replication can continue. the FA core complex. More in

particular, the FA core complex is a nuclear multiprotein complex consisting of FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL, and FANCM, functions as an E3 ubiquitin ligase and mediates the activation of the ID complex, which is a heterodimer composed of FANCD2 and FANCI. Once monoubiquitinated, it interacts with classical tumor suppressors downstream of the FA pathway including FANCD1/BRCA2, FANCN/PALB2, FANCJ/BRIP1, and FANCO/Rad51C and thereby contributes to DNA repair via homologous recombination (HR). Eighty to 90 percent of FA cases are due to mutations in one of three genes, FANCA, FANCC, and FANCG. These genes provide instructions for producing components of the FA core complex. Mutations in such genes associated with the FA core complex will cause the complex to be nonfunctional and disrupt the entire FA pathway. As a result, DNA damage is not repaired efficiently and ICLs build up over time. Geiselhart, "Review Article, Disrupted Signaling through the Fanconi Anemia Pathway Leads to Dysfunctional Hematopoietic Stem Cell Biology: Underlying Mechanisms and Potential Therapeutic Strategies," Anemia Volume 2012 (2012), Article ID 265790, dx.doi.org/10.1155/2012/265790 discussed FA and an animal experiment involving intrafemoral injection of a lentivirus encoding the FANCC gene resulting in correction of HSCs in vivo. Using a CRISPR-Cas system that targets and one or more of the mutations associated with FA, for instance a CRISPR-Cas system having gRNA(s) and HDR template(s) that respectively targets one or more of the mutations of FANCA, FANCC, or FANCG that give rise to FA and provide corrective expression of one or more of FANCA, FANCC or FANCG; e.g., the gRNA can target a mutation as to FANCC, and the HDR can provide coding for proper expression of FANCC. An gRNA that targets the mutation(s) (e.g., one or more involved in FA, such as mutation(s) as to any one or more of FANCA, FANCC or FANCG)-and-Cas protein containing particle is contacted with HSCs carrying the mutation(s). The particle also can contain a suitable HDR template(s) to correct the mutation for proper expression of one or more of the proteins involved in FA, such as any one or more of FANCA, FANCC or FANCG; or the HSC can be contacted with a second particle or a vector that contains or delivers the HDR template. The so contacted cells can be administered; and optionally treated / expanded; cf. Cartier.

[00621] The particle in the herein discussion (e.g., as to containing gRNA(s) and Cas, optionally HDR template(s), or HDR template(s); for instance as to Hemophilia B, SCID, SCID-X1, ADA-SCID, Hereditary tyrosinemia, β-thalassemia, X-linked CGD, Wiskott-Aldrich syndrome, Fanconi anemia, adrenoleukodystrophy (ALD), metachromatic leukodystrophy (MLD), HIV/AIDS, Immunodeficiency disorder, Hematologic condition, or

genetic lysosomal storage disease) is advantageously obtained or obtainable from admixing an gRNA(s) and Cas protein mixture (optionally containing HDR template(s) or such mixture only containing HDR template(s) when separate particles as to template(s) is desired) with a mixture comprising or consisting essentially of or consisting of surfactant, phospholipid, biodegradable polymer, lipoprotein and alcohol (wherein one or more gRNA targets the genetic locus or loci in the HSC).

[00622] Indeed, the invention is especially suited for treating hematopoietic genetic disorders with genome editing, and immunodeficiency disorders, such as genetic immunodeficiency disorders, especially through using the particle technology herein-discussed. Genetic immunodeficiencies are diseases where genome editing interventions of the instant invention can successful. The reasons include: Hematopoietic cells, of which immune cells are a subset, are therapeutically accessible. They can be removed from the body and transplanted autologously or allogenically. Further, certain genetic immunodeficiencies, e.g., severe combined immunodeficiency (SCID), create a proliferative disadvantage for immune cells. Correction of genetic lesions causing SCID by rare, spontaneous 'reverse' mutations indicates that correcting even one lymphocyte progenitor may be sufficient to recover immune function in patients.../../../Users/t_kowalski/AppData/Local/Microsoft/Windows/Temporary Internet Files/Content.Outlook/GA8VY8LK/Treating SCID for Ellen.docx - _ENREF_1 See Bousso, P., et al. Diversity, functionality, and stability of the T cell repertoire derived in vivo from a single human T cell precursor. Proceedings of the National Academy of Sciences of the United States of America 97, 274-278 (2000). The selective advantage for edited cells allows for even low levels of editing to result in a therapeutic effect. This effect of the instant invention can be seen in SCID, Wiskott-Aldrich Syndrome, and the other conditions mentioned herein, including other genetic hematopoietic disorders such as alpha- and beta- thalassemia, where hemoglobin deficiencies negatively affect the fitness of erythroid progenitors.

[00623] The activity of NHEJ and HDR DSB repair varies significantly by cell type and cell state. NHEJ is not highly regulated by the cell cycle and is efficient across cell types, allowing for high levels of gene disruption in accessible target cell populations. In contrast, HDR acts primarily during S/G2 phase, and is therefore restricted to cells that are actively dividing, limiting treatments that require precise genome modifications to mitotic cells [Ciccia, A. & Elledge, S.J. Molecular cell 40, 179-204 (2010); Chapman, J.R., et al. Molecular cell 47, 497-510 (2012)].

[00624] The efficiency of correction via HDR may be controlled by the epigenetic state or sequence of the targeted locus, or the specific repair template configuration (single vs. double

stranded, long vs. short homology arms) used [Hacein-Bey-Abina, S., et al. The New England journal of medicine 346, 1185-1193 (2002); Gaspar, H.B., et al. Lancet 364, 2181-2187 (2004); Beumer, K.J., et al. G3 (2013)]. The relative activity of NHEJ and HDR machineries in target cells may also affect gene correction efficiency, as these pathways may compete to resolve DSBs [Beumer, K.J., et al. Proceedings of the National Academy of Sciences of the United States of America 105, 19821-19826 (2008)]. HDR also imposes a delivery challenge not seen with NHEJ strategies, as it requires the concurrent delivery of nucleases and repair templates. In practice, these constraints have so far led to low levels of HDR in therapeutically relevant cell types. Clinical translation has therefore largely focused on NHEJ strategies to treat disease, although proof-of-concept preclinical HDR treatments have now been described for mouse models of haemophilia B and hereditary tyrosinemia [Li, H., et al. Nature 475, 217-221 (2011); Yin, H., et al. Nature biotechnology 32, 551-553 (2014)].

[00625] Any given genome editing application may comprise combinations of proteins, small RNA molecules, and/or repair templates, making delivery of these multiple parts substantially more challenging than small molecule therapeutics. Two main strategies for delivery of genome editing tools have been developed: ex vivo and in vivo. In ex vivo treatments, diseased cells are removed from the body, edited and then transplanted back into the patient. Ex vivo editing has the advantage of allowing the target cell population to be well defined and the specific dosage of therapeutic molecules delivered to cells to be specified. The latter consideration may be particularly important when off-target modifications are a concern, as titrating the amount of nuclease may decrease such mutations (Hsu et al., 2013). Another advantage of ex vivo approaches is the typically high editing rates that can be achieved, due to the development of efficient delivery systems for proteins and nucleic acids into cells in culture for research and gene therapy applications.

[00626] There may be drawbacks with ex vivo approaches that limit application to a small number of diseases. For instance, target cells must be capable of surviving manipulation outside the body. For many tissues, like the brain, culturing cells outside the body is a major challenge, because cells either fail to survive, or lose properties necessary for their function in vivo. Thus, in view of this disclosure and the knowledge in the art, ex vivo therapy as to tissues with adult stem cell populations amenable to ex vivo culture and manipulation, such as the hematopoietic system, by the CRISPR-Cas system are enabled. [Bunn, H.F. & Aster, J. Pathophysiology of blood disorders, (McGraw-Hill, New York, 2011)]

[00627] In vivo genome editing involves direct delivery of editing systems to cell types in their native tissues. In vivo editing allows diseases in which the affected cell population is not

amenable to ex vivo manipulation to be treated. Furthermore, delivering nucleases to cells in situ allows for the treatment of multiple tissue and cell types. These properties probably allow in vivo treatment to be applied to a wider range of diseases than ex vivo therapies.

[00628] To date, in vivo editing has largely been achieved through the use of viral vectors with defined, tissue-specific tropism. Such vectors are currently limited in terms of cargo carrying capacity and tropism, restricting this mode of therapy to organ systems where transduction with clinically useful vectors is efficient, such as the liver, muscle and eye [Kotterman, M.A. & Schaffer, D.V. Nature reviews. Genetics 15, 445-451 (2014); Nguyen, T.H. & Ferry, N. Gene therapy 11 Suppl 1, S76-84 (2004); Boye, S.E., et al. Molecular therapy, the journal of the American Society of Gene Therapy 21, 509-519 (2013)].

[00629] A potential barrier for in vivo delivery is the immune response that may be created in response to the large amounts of virus necessary for treatment, but this phenomenon is not unique to genome editing and is observed with other virus based gene therapies [Bessis, N., et al. Gene therapy 11 Suppl 1, S10-17 (2004)]. It is also possible that peptides from editing nucleases themselves are presented on MHC Class I molecules to stimulate an immune response, although there is little evidence to support this happening at the preclinical level. Another major difficulty with this mode of therapy is controlling the distribution and consequently the dosage of genome editing nucleases in vivo, leading to off-target mutation profiles that may be difficult to predict. However, in view of this disclosure and the knowledge in the art, including the use of virus- and particle-based therapies being used in the treatment of cancers, in vivo modification of HSCs, for instance by delivery by either particle or virus, is within the ambit of the skilled person.

[00630] Ex Vivo Editing Therapy: The long-standing clinical expertise with the purification, culture and transplantation of hematopoietic cells has made diseases affecting the blood system such as SCID, Fanconi anemia, Wiskott-Aldrich syndrome and sickle cell anemia the focus of ex vivo editing therapy. Another reason to focus on hematopoietic cells is that, thanks to previous efforts to design gene therapy for blood disorders, delivery systems of relatively high efficiency already exist. With these advantages, this mode of therapy can be applied to diseases where edited cells possess a fitness advantage, so that a small number of engrafted, edited cells can expand and treat disease. One such disease is HIV, where infection results in a fitness disadvantage to CD4+ T cells.

[00631] Ex vivo editing therapy has been recently extended to include gene correction strategies. The barriers to HDR ex vivo were overcome in a recent paper from Genovese and colleagues, who achieved gene correction of a mutated IL2RG gene in hematopoietic stem

cells (HSCs) obtained from a patient suffering from SCID-X1 [Genovese, P., et al. Nature 510, 235-240 (2014)]. Genovese et. al. accomplished gene correction in HSCs using a multimodal strategy. First, HSCs were transduced using integration-deficient lentivirus containing an HDR template encoding a therapeutic cDNA for IL2RG. Following transduction, cells were electroporated with mRNA encoding ZFNs targeting a mutational hotspot in IL2RG to stimulate HDR based gene correction. To increase HDR rates, culture conditions were optimized with small molecules to encourage HSC division. With optimized culture conditions, nucleases and HDR templates, gene corrected HSCs from the SCID-X1 patient were obtained in culture at therapeutically relevant rates. HSCs from unaffected individuals that underwent the same gene correction procedure could sustain long-term hematopoiesis in mice, the gold standard for HSC function. HSCs are capable of giving rise to all hematopoietic cell types and can be autologously transplanted, making them an extremely valuable cell population for all hematopoietic genetic disorders [Weissman, I.L. & Shizuru, J.A. Blood 112, 3543-3553 (2008)]. Gene corrected HSCs could, in principle, be used to treat a wide range of genetic blood disorders making this study an exciting breakthrough for therapeutic genome editing.

[00632] In Vivo Editing Therapy: In vivo editing can be used advantageously from this disclosure and the knowledge in the art. For organ systems where delivery is efficient, there have already been a number of exciting preclinical therapeutic successes. The first example of successful in vivo editing therapy was demonstrated in a mouse model of haemophilia B [Li, H., et al. Nature 475, 217-221 (2011)]. As noted earlier, Haemophilia B is an X-linked recessive disorder caused by loss-of-function mutations in the gene encoding Factor IX, a crucial component of the clotting cascade. Recovering Factor IX activity to above 1% of its levels in severely affected individuals can transform the disease into a significantly milder form, as infusion of recombinant Factor IX into such patients prophylactically from a young age to achieve such levels largely ameliorates clinical complications [Lofqvist, T., et al. Journal of internal medicine 241, 395-400 (1997)]. Thus, only low levels of HDR gene correction are necessary to change clinical outcomes for patients. In addition, Factor IX is synthesized and secreted by the liver, an organ that can be transduced efficiently by viral vectors encoding editing systems.

[00633] Using hepatotropic adeno-associated viral (AAV) serotypes encoding ZFNs and a corrective HDR template, up to 7% gene correction of a mutated, humanized Factor IX gene in the murine liver was achieved [Li, H., et al. Nature 475, 217-221 (2011)]. This resulted in improvement of clot formation kinetics, a measure of the function of the clotting cascade, demonstrating for the first time that in vivo editing therapy is not only feasible, but also

efficacious. As discussed herein, the skilled person is positioned from the teachings herein and the knowledge in the art, e.g., Li to address Haemophilia B with a particle-containing HDR template and a CRISPR-Cas system that targets the mutation of the X-linked recessive disorder to reverse the loss-of-function mutation.

[00634] Building on this study, other groups have recently used in vivo genome editing of the liver with CRISPR-Cas to successfully treat a mouse model of hereditary tyrosinemia and to create mutations that provide protection against cardiovascular disease. These two distinct applications demonstrate the versatility of this approach for disorders that involve hepatic dysfunction [Yin, H., et al. Nature biotechnology 32, 551-553 (2014); Ding, Q., et al. Circulation research 115, 488-492 (2014)]. Application of in vivo editing to other organ systems are necessary to prove that this strategy is widely applicable. Currently, efforts to optimize both viral and non-viral vectors are underway to expand the range of disorders that can be treated with this mode of therapy [Kotterman, M.A. & Schaffer, D.V. Nature reviews. Genetics 15, 445-451 (2014); Yin, H., et al. Nature reviews. Genetics 15, 541-555 (2014)]. As discussed herein, the skilled person is positioned from the teachings herein and the knowledge in the art, e.g., Yin to address hereditary tyrosinemia with a particle-containing HDR template and a CRISPR-Cas system that targets the mutation.

[00635] Targeted deletion, therapeutic applications: Targeted deletion of genes may be preferred. Preferred are, therefore, genes involved in immunodeficiency disorder, hematologic condition, or genetic lysosomal storage disease, e.g., Hemophilia B, SCID, SCID-X1, ADA-SCID, Hereditary tyrosinemia, β-thalassemia, X-linked CGD, Wiskott-Aldrich syndrome, Fanconi anemia, adrenoleukodystrophy (ALD), metachromatic leukodystrophy (MLD), HIV/AIDS, other metabolic disorders, genes encoding mis-folded proteins involved in diseases, genes leading to loss-of-function involved in diseases; generally, mutations that can be targeted in an HSC, using any herein-discussed delivery system, with the particle system considered advantageous.

[00636] In the present invention, the immunogenicity of the CRISPR enzyme in particular may be reduced following the approach first set out in Tangri et al with respect to erythropoietin and subsequently developed. Accordingly, directed evolution or rational design may be used to reduce the immunogenicity of the Cas in the host species (human or other species).

[00637] Genome editing: The CRISPR/Cas systems of the present invention can be used to correct genetic mutations that were previously attempted with limited success using TALEN and ZFN and lentiviruses, including as herein discussed; see also Internation Patent Publication No. WO2013163628.

Treating Disease of the Brain, Central Nervous and Immune Systems

[00638] The present invention also contemplates delivering the CRISPR-Cas system to the brain or neurons. For example, RNA interference (RNAi) offers therapeutic potential for this disorder by reducing the expression of HTT, the disease-causing gene of Huntington's disease (see, e.g., McBride et al., Molecular Therapy vol. 19 no. 12 Dec. 2011, pp. 2152-2162), therefore Applicant postulates that it may be used/and or adapted to the CRISPR-Cas system. The CRISPR-Cas system may be generated using an algorithm to reduce the off-targeting potential of antisense sequences. The CRISPR-Cas sequences may target either a sequence in exon 52 of mouse, rhesus or human huntingtin and expressed in a viral vector, such as AAV. Animals, including humans, may be injected with about three microinjections per hemisphere (six injections total): the first 1 mm rostral to the anterior commissure (12 μl) and the two remaining injections (12 μl and 10 μl, respectively) spaced 3 and 6 mm caudal to the first injection with 1e12 vg/ml of AAV at a rate of about 1 μl/minute, and the needle was left in place for an additional 5 minutes to allow the injectate to diffuse from the needle tip.

[00639] DiFiglia et al. (PNAS, October 23, 2007, vol. 104, no. 43, 17204–17209) observed that single administration into the adult striatum of an siRNA targeting Htt can silence mutant Htt, attenuate neuronal pathology, and delay the abnormal behavioral phenotype observed in a rapid-onset, viral transgenic mouse model of HD. DiFiglia injected mice intrastriatally with 2 μl of Cy3-labeled cc-siRNA-Htt or unconjugated siRNA-Htt at 10 μM. A similar dosage of CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 5-10 ml of 10 μM CRISPR Cas targeted to Htt may be injected intrastriatally.

[00640] In another example, Boudreau et al. (Molecular Therapy vol. 17 no. 6 june 2009) injects 5 μl of recombinant AAV serotype 2/1 vectors expressing htt-specific RNAi virus (at 4 x 1012 viral genomes/ml) into the striatum. A similar dosage of CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 10-20 ml of 4 x 1012 viral genomes/ml) CRISPR Cas targeted to Htt may be injected intrastriatally.

[00641] In another example, a CRISPR Cas targeted to HTT may be administered continuously (see, e.g., Yu et al., Cell 150, 895–908, August 31, 2012). Yu et al. utilizes osmotic pumps delivering 0.25 ml/hr (Model 2004) to deliver 300 mg/day of ss-siRNA or phosphate-buffered saline (PBS) (Sigma Aldrich) for 28 days, and pumps designed to deliver 0.5 μl/hr (Model 2002) were used to deliver 75 mg/day of the positive control MOE ASO for 14 days. Pumps (Durect Corporation) were filled with ss-siRNA or MOE diluted in sterile PBS and then incubated at 37 C for 24 or 48 (Model 2004) hours prior to implantation. Mice were anesthetized with 2.5% isofluorane, and a midline incision was made at the base of the skull.

Using stereotaxic guides, a cannula was implanted into the right lateral ventricle and secured with Loctite adhesive. A catheter attached to an Alzet osmotic mini pump was attached to the cannula, and the pump was placed subcutaneously in the midscapular area. The incision was closed with 5.0 nylon sutures. A similar dosage of CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 500 to 1000 g/day CRISPR Cas targeted to Htt may be administered.

[00642] In another example of continuous infusion, Stiles et al. (Experimental Neurology 233 (2012) 463–471) implanted an intraparenchymal catheter with a titanium needle tip into the right putamen. The catheter was connected to a SynchroMed® II Pump (Medtronic Neurological, Minneapolis, MN) subcutaneously implanted in the abdomen. After a 7 day infusion of phosphate buffered saline at 6 μL/day, pumps were re-filled with test article and programmed for continuous delivery for 7 days. About 2.3 to 11.52 mg/d of siRNA were infused at varying infusion rates of about 0.1 to 0.5 μL/min. A similar dosage of CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 20 to 200 mg/day CRISPR Cas targeted to Htt may be administered. In another example, the methods of US Patent Publication No. 20130253040 assigned to Sangamo may also be also be adapted from TALES to the nucleic acid-targeting system of the present invention for treating Huntington's Disease.

[00643] In another example, the methods of US Patent Publication No. 20130253040 (WO2013130824) assigned to Sangamo may also be also be adapted from TALES to the CRISPR Cas system of the present invention for treating Huntington's Disease.

[00644] WO2015089354 A1 in the name of The Broad Institute et al., hereby incorporated by reference, describes a targets for Huntington's Disease (HP). Possible target genes of CRISPR complex in regard to Huntington's Disease: PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; and TGM2. Accordingly, one or more of PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; and TGM2 may be selected as targets for Huntington's Disease in some embodiments of the present invention.

[00645] Other trinucleotide repeat disorders. These may include any of the following: Category I includes Huntington's disease (HD) and the spinocerebellar ataxias; Category II expansions are phenotypically diverse with heterogeneous expansions that are generally small in magnitude, but also found in the exons of genes; and Category III includes fragile X syndrome, myotonic dystrophy, two of the spinocerebellar ataxias, juvenile myoclonic epilepsy, and Friedreich's ataxia.

[00646] A further aspect of the invention relates to utilizing the CRISPR-Cas system for correcting defects in the EMP2A and EMP2B genes that have been identified to be associated with Lafora disease. Lafora disease is an autosomal recessive condition which is characterized by progressive myoclonus epilepsy which may start as epileptic seizures in adolescence. A few cases of the disease may be caused by mutations in genes yet to be identified. The disease causes seizures, muscle spasms, difficulty walking, dementia, and eventually death. There is currently no therapy that has proven effective against disease progression. Other genetic abnormalities associated with epilepsy may also be targeted by the CRISPR-Cas system and the underlying genetics is further described in Genetics of Epilepsy and Genetic Epilepsies, edited by Giuliano Avanzini, Jeffrey L. Noebels, Mariani Foundation Paediatric Neurology:20; 2009).

[00647] The methods of US Patent Publication No. 20110158957 assigned to Sangamo BioSciences, Inc. involved in inactivating T cell receptor (TCR) genes may also be modified to the CRISPR Cas system of the present invention. In another example, the methods of US Patent Publication No. 20100311124 assigned to Sangamo BioSciences, Inc. and US Patent Publication No. 20110225664 assigned to Cellectis, which are both involved in inactivating glutamine synthetase gene expression genes may also be modified to the CRISPR Cas system of the present invention.

[00648] Delivery options for the brain include encapsulation of CRISPR enzyme and guide RNA in the form of either DNA or RNA into liposomes and conjugating to molecular Trojan horses for trans-blood brain barrier (BBB) delivery. Molecular Trojan horses have been shown to be effective for delivery of B-gal expression vectors into the brain of non-human primates. The same approach can be used to delivery vectors containing CRISPR enzyme and guide RNA. For instance, Xia CF and Boado RJ, Pardridge WM ("Antibody-mediated targeting of siRNA via the human insulin receptor using avidin-biotin technology." Mol Pharm. 2009 May-Jun;6(3):747-51. doi: 10.1021/mp800194) describes how delivery of short interfering RNA (siRNA) to cells in culture, and in vivo, is possible with combined use of a receptor-specific monoclonal antibody (mAb) and avidin-biotin technology. The authors also report that because the bond between the targeting mAb and the siRNA is stable with avidin-biotin technology, and RNAi effects at distant sites such as brain are observed in vivo following an intravenous administration of the targeted siRNA.

[00649] Zhang et al. (Mol Ther. 2003 Jan;7(1):11-8.)) describe how expression plasmids encoding reporters such as luciferase were encapsulated in the interior of an "artificial virus" comprised of an 85 nm pegylated immunoliposome, which was targeted to the rhesus monkey

brain in vivo with a monoclonal antibody (MAb) to the human insulin receptor (HIR). The HIRMAb enables the liposome carrying the exogenous gene to undergo transcytosis across the blood-brain barrier and endocytosis across the neuronal plasma membrane following intravenous injection. The level of luciferase gene expression in the brain was 50-fold higher in the rhesus monkey as compared to the rat. Widespread neuronal expression of the beta-galactosidase gene in primate brain was demonstrated by both histochemistry and confocal microscopy. The authors indicate that this approach makes feasible reversible adult transgenics in 24 hours. Accordingly, the use of immunoliposome is preferred. These may be used in conjunction with antibodies to target specific tissues or cell surface proteins.

Alzheimer's Disease

[00650] US Patent Publication No. 20110023153, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with Alzheimer's Disease. Once modified cells and animals may be further tested using known methods to study the effects of the targeted mutations on the development and/or progression of AD using measures commonly used in the study of AD – such as, without limitation, learning and memory, anxiety, depression, addiction, and sensory motor functions as well as assays that measure behavioral, functional, pathological, metabolic and biochemical function.

[00651] The present disclosure comprises editing of any chromosomal sequences that encode proteins associated with AD. The AD-related proteins are typically selected based on an experimental association of the AD-related protein to an AD disorder. For example, the production rate or circulating concentration of an AD-related protein may be elevated or depressed in a population having an AD disorder relative to a population lacking the AD disorder. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the AD-related proteins may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[00652] Examples of Alzheimer's disease associated proteins may include the very low density lipoprotein receptor protein (VLDLR) encoded by the VLDLR gene, the ubiquitin-like modifier activating enzyme 1 (UBA1) encoded by the UBA1 gene, or the NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) encoded by the UBA3 gene, for example.

[00653] By way of non-limiting example, proteins associated with AD include but are not limited to the proteins listed as follows: Chromosomal Sequence Encoded Protein ALAS2

Delta-aminolevulinate synthase 2 (ALAS2) ABCA1 ATP-binding cassette transporter (ABCA1) ACE Angiotensin I-converting enzyme (ACE) APOE Apolipoprotein E precursor (APOE) APP amyloid precursor protein (APP) AQP1 aquaporin 1 protein (AQP1) BIN1 Myc box-dependent-interacting protein 1 or bridging integrator 1 protein (BIN1) BDNF brain-derived neurotrophic factor (BDNF) BTNL8 Butyrophilin-like protein 8 (BTNL8) C1ORF49 chromosome 1 open reading frame 49 CDH4 Cadherin-4 CHRNB2 Neuronal acetylcholine receptor subunit beta-2 CKLFSF2 CKLF-like MARVEL transmembrane domain- containing protein 2 (CKLFSF2) CLEC4E C-type lectin domain family 4, member e (CLEC4E) CLU clusterin protein (also known as apoplipoprotein J) CR1 Erythrocyte complement receptor 1 (CR1, also known as CD35, C3b/C4b receptor and immune adherence receptor) CR1L Erythrocyte complement receptor 1 (CR1L) CSF3R granulocyte colony-stimulating factor 3 receptor (CSF3R) CST3 Cystatin C or cystatin 3 CYP2C Cytochrome P450 2C DAPK1 Death-associated protein kinase 1 (DAPK1) ESR1 Estrogen receptor 1 FCAR Fc fragment of IgA receptor (FCAR, also known as CD89) FCGR3B Fc fragment of IgG, low affinity IIIb, receptor (FCGR3B or CD16b) FFA2 Free fatty acid receptor 2 (FFA2) FGA Fibrinogen (Factor I) GAB2 GRB2-associated-binding protein 2 (GAB2) GAB2 GRB2-associated-binding protein 2 (GAB2) GALP Galanin-like peptide GAPDHS Glyceraldehyde-3-phosphate dehydrogenase, spermatogenic (GAPDHS) GMPB GMBP HP Haptoglobin (HP) HTR7 5-hydroxytryptamine (serotonin) receptor 7 (adenylate cyclase-coupled) IDE Insulin degrading enzyme IF127 IF127 IFI6 Interferon, alpha-inducible protein 6 (IFI6) IFIT2 Interferon-induced protein with tetratricopeptide repeats 2 (IFIT2) IL1RN interleukin-1 receptor antagonist (IL-1RA) IL8RA Interleukin 8 receptor, alpha (IL8RA or CD181) IL8RB Interleukin 8 receptor, beta (IL8RB) JAG1 Jagged 1 (JAG1) KCNJ15 Potassium inwardly-rectifying channel, subfamily J, member 15 (KCNJ15) LRP6 Low-density lipoprotein receptor-related protein 6 (LRP6) MAPT microtubule-associated protein tau (MAPT) MARK4 MAP/microtubule affinity-regulating kinase 4 (MARK4) MPHOSPH1 M-phase phosphoprotein 1 MTHFR 5,10-methylenetetrahydrofolate reductase MX2 Interferon-induced GTP-binding protein Mx2 NBN Nibrin, also known as NBN NCSTN Nicastrin NIACR2 Niacin receptor 2 (NIACR2, also known as GPR109B) NMNAT3 nicotinamide nucleotide adenylyltransferase 3 NTM Neurotrimin (or HNT) ORM1 Orosmucoid 1 (ORM1) or Alpha-1-acid glycoprotein 1 P2RY13 P2Y purinoceptor 13 (P2RY13) PBEF1 Nicotinamide phosphoribosyltransferase (NAmPRTase or Nampt) also known as pre-B-cell colony-enhancing factor 1 (PBEF1) or visfatin PCK1 Phosphoenolpyruvate carboxykinase PICALM phosphatidylinositol binding clathrin assembly protein (PICALM) PLAU Urokinase-type plasminogen activator (PLAU)

PLXNC1 Plexin C1 (PLXNC1) PRNP Prion protein PSEN1 presenilin 1 protein (PSEN1) PSEN2 presenilin 2 protein (PSEN2) PTPRA protein tyrosine phosphatase receptor type A protein (PTPRA) RALGPS2 Ral GEF with PH domain and SH3 binding motif 2 (RALGPS2) RGSL2 regulator of G-protein signaling like 2 (RGSL2) SELENBP1 Selenium binding protein 1 (SELNBP1) SLC25A37 Mitoferrin-1 SORL1 sortilin-related receptor L(DLR class) A repeats-containing protein (SORL1) TF Transferrin TFAM Mitochondrial transcription factor A TNF Tumor necrosis factor TNFRSF10C Tumor necrosis factor receptor superfamily member 10C (TNFRSF10C) TNFSF10 Tumor necrosis factor receptor superfamily, (TRAIL) member 10a (TNFSF10) UBA1 ubiquitin-like modifier activating enzyme 1 (UBA1) UBA3 NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) UBB ubiquitin B protein (UBB) UBQLN1 Ubiquilin-1 UCHL1 ubiquitin carboxyl-terminal esterase L1 protein (UCHL1) UCHL3 ubiquitin carboxyl-terminal hydrolase isozyme L3 protein (UCHL3) VLDLR very low density lipoprotein receptor protein (VLDLR)

[00654] In exemplary embodiments, the proteins associated with AD whose chromosomal sequence is edited may be the very low density lipoprotein receptor protein (VLDLR) encoded by the VLDLR gene, the ubiquitin-like modifier activating enzyme 1 (UBA1) encoded by the UBA1 gene, the NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) encoded by the UBA3 gene, the aquaporin 1 protein (AQP1) encoded by the AQP1 gene, the ubiquitin carboxyl-terminal esterase L1 protein (UCHL1) encoded by the UCHL1 gene, the ubiquitin carboxyl-terminal hydrolase isozyme L3 protein (UCHL3) encoded by the UCHL3 gene, the ubiquitin B protein (UBB) encoded by the UBB gene, the microtubule-associated protein tau (MAPT) encoded by the MAPT gene, the protein tyrosine phosphatase receptor type A protein (PTPRA) encoded by the PTPRA gene, the phosphatidylinositol binding clathrin assembly protein (PICALM) encoded by the PICALM gene, the clusterin protein (also known as apoplipoprotein J) encoded by the CLU gene, the presenilin 1 protein encoded by the PSEN1 gene, the presenilin 2 protein encoded by the PSEN2 gene, the sortilin-related receptor L(DLR class) A repeats-containing protein (SORL1) protein encoded by the SORL1 gene, the amyloid precursor protein (APP) encoded by the APP gene, the Apolipoprotein E precursor (APOE) encoded by the APOE gene, or the brain-derived neurotrophic factor (BDNF) encoded by the BDNF gene. In an exemplary embodiment, the genetically modified animal is a rat, and the edited chromosomal sequence encoding the protein associated with AD is as as follows: APP amyloid precursor protein (APP) NM_019288 AQP1 aquaporin 1 protein (AQP1) NM_012778 BDNF Brain-derived neurotrophic factor NM_012513 CLU clusterin protein (also known as NM_053021 apoplipoprotein J) MAPT microtubule-associated protein NM_017212 tau

(MAPT) PICALM phosphatidylinositol binding NM_053554 clathrin assembly protein (PICALM) PSEN1 presenilin 1 protein (PSEN1) NM_019163 PSEN2 presenilin 2 protein (PSEN2) NM_031087 PTPRA protein tyrosine phosphatase NM_012763 receptor type A protein (PTPRA) SORL1 sortilin-related receptor L(DLR NM_053519, class) A repeats-containing XM_001065506, protein (SORL1) XM_217115 UBA1 ubiquitin-like modifier activating NM_001014080 enzyme 1 (UBA1) UBA3 NEDD8-activating enzyme E1 NM_057205 catalytic subunit protein (UBE1C) UBB ubiquitin B protein (UBB) NM_138895 UCHL1 ubiquitin carboxyl-terminal NM_017237 esterase L1 protein (UCHL1) UCHL3 ubiquitin carboxyl-terminal NM_001110165 hydrolase isozyme L3 protein (UCHL3) VLDLR very low density lipoprotein NM_013155 receptor protein (VLDLR)

[00655] The animal or cell may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9,10, 11, 12, 13, 14, 15 or more disrupted chromosomal sequences encoding a protein associated with AD and zero, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15 or more chromosomally integrated sequences encoding a protein associated with AD.

[00656] The edited or integrated chromosomal sequence may be modified to encode an altered protein associated with AD. A number of mutations in AD-related chromosomal sequences have been associated with AD. For instance, the V7171 (i.e. valine at position 717 is changed to isoleucine) missense mutation in APP causes familial AD. Multiple mutations in the presenilin-1 protein, such as H163R (i.e. histidine at position 163 is changed to arginine), A246E (i.e. alanine at position 246 is changed to glutamate), L286V (i.e. leucine at position 286 is changed to valine) and C410Y (i.e. cysteine at position 410 is changed to tyrosine) cause familial Alzheimer's type 3. Mutations in the presenilin-2 protein, such as N141 I (i.e. asparagine at position 141 is changed to isoleucine), M239V (i.e. methionine at position 239 is changed to valine), and D439A (i.e. aspartate at position 439 is changed to alanine) cause familial Alzheimer's type 4. Other associations of genetic variants in AD-associated genes and disease are known in the art. See, for example, Waring et al. (2008) Arch. Neurol. 65:329-334, the disclosure of which is incorporated by reference herein in its entirety.

[00657] Secretase Disorders

[00658] US Patent Publication No. 20110023146, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with secretase-associated disorders. Secretases are essential for processing pre-proteins into their biologically active forms. Defects in various components of the secretase pathways contribute to many disorders, particularly those with hallmark amyloid genesis or amyloid plaques, such as Alzheimer's disease (AD).

[00659] A secretase disorder and the proteins associated with these disorders are a diverse set of proteins that effect susceptibility for numerous disorders, the presence of the disorder, the severity of the disorder, or any combination thereof. The present disclosure comprises editing of any chromosomal sequences that encode proteins associated with a secretase disorder. The proteins associated with a secretase disorder are typically selected based on an experimental association of the secretase--related proteins with the development of a secretase disorder. For example, the production rate or circulating concentration of a protein associated with a secretase disorder may be elevated or depressed in a population with a secretase disorder relative to a population without a secretase disorder. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the protein associated with a secretase disorder may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[00660] By way of non-limiting example, proteins associated with a secretase disorder include PSENEN (presenilin enhancer 2 homolog (C. elegans)), CTSB (cathepsin B), PSEN1 (presenilin 1), APP (amyloid beta (A4) precursor protein), APH1B (anterior pharynx defective 1 homolog B (C. elegans)), PSEN2 (presenilin 2 (Alzheimer disease 4)), BACE1 (beta-site APP-cleaving enzyme 1), ITM2B (integral membrane protein 2B), CTSD (cathepsin D), NOTCH1 (Notch homolog 1, translocation-associated (Drosophila)), TNF (tumor necrosis factor (TNF superfamily, member 2)), INS (insulin), DYT10 (dystonia 10), ADAM17 (ADAM metallopeptidase domain 17), APOE (apolipoprotein E), ACE (angiotensin I converting enzyme (peptidyl-dipeptidase A) 1), STN (statin), TP53 (tumor protein p53), IL6 (interleukin 6 (interferon, beta 2)), NGFR (nerve growth factor receptor (TNFR superfamily, member 16)), IL1B (interleukin 1, beta), ACHE (acetylcholinesterase (Yt blood group)), CTNNB1 (catenin (cadherin-associated protein), beta 1, 88kDa), IGF1 (insulin-like growth factor 1 (somatomedin C)), IFNG (interferon, gamma), NRG1 (neuregulin 1), CASP3 (caspase 3, apoptosis-related cysteine peptidase), MAPK1 (mitogen-activated protein kinase 1), CDH1 (cadherin 1, type 1, E-cadherin (epithelial)), APBB1 (amyloid beta (A4) precursor protein-binding, family B, member 1 (Fe65)), HMGCR (3-hydroxy-3-methylglutaryl-Coenzyme A reductase), CREB1 (cAMP responsive element binding protein 1), PTGS2 (prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)), HES1 (hairy and enhancer of split 1, (Drosophila)), CAT (catalase), TGFB1 (transforming growth factor, beta 1), ENO2 (enolase 2

(gamma, neuronal)), ERBB4 (v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)), TRAPPC10 (trafficking protein particle complex 10), MAOB (monoamine oxidase B), NGF (nerve growth factor (beta polypeptide)), MMP12 (matrix metallopeptidase 12 (macrophage elastase)), JAG1 (jagged 1 (Alagille syndrome)), CD40LG (CD40 ligand), PPARG (peroxisome proliferator-activated receptor gamma), FGF2 (fibroblast growth factor 2 (basic)), IL3 (interleukin 3 (colony-stimulating factor, multiple)), LRP1 (low density lipoprotein receptor-related protein 1), NOTCH4 (Notch homolog 4 (Drosophila)), MAPK8 (mitogen-activated protein kinase 8), PREP (prolyl endopeptidase), NOTCH3 (Notch homolog 3 (Drosophila)), PRNP (prion protein), CTSG (cathepsin G), EGF (epidermal growth factor (beta-urogastrone)), REN (renin), CD44 (CD44 molecule (Indian blood group)), SELP (selectin P (granule membrane protein 140 kDa, antigen CD62)), GHR (growth hormone receptor), ADCYAP1 (adenylate cyclase activating polypeptide 1 (pituitary)), INSR (insulin receptor), GFAP (glial fibrillary acidic protein), MMP3 (matrix metallopeptidase 3 (stromelysin 1, progelatinase)), MAPK10 (mitogen-activated protein kinase 10), SP1 (Sp1 transcription factor), MYC (v-myc myelocytomatosis viral oncogene homolog (avian)), CTSE (cathepsin E), PPARA (peroxisome proliferator-activated receptor alpha), JUN (jun oncogene), TIMP1 (TIMP metallopeptidase inhibitor 1), IL5 (interleukin 5 (colony-stimulating factor, eosinophil)), IL1A (interleukin 1, alpha), MMP9 (matrix metallopeptidase 9 (gelatinase B, 92 kDa gelatinase, 92 kDa type IV collagenase)), HTR4 (5-hydroxytryptamine (serotonin) receptor 4), HSPG2 (heparan sulfate proteoglycan 2), KRAS (v-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog), CYCS (cytochrome c, somatic), SMG1 (SMG1 homolog, phosphatidylinositol 3-kinase-related kinase (C. elegans)), IL1R1 (interleukin 1 receptor, type I), PROK1 (prokineticin 1), MAPK3 (mitogen-activated protein kinase 3), NTRK1 (neurotrophic tyrosine kinase, receptor, type 1), IL13 (interleukin 13), MME (membrane metallo-endopeptidase), TKT (transketolase), CXCR2 (chemokine (C-X-C motif) receptor 2), IGF1R (insulin-like growth factor 1 receptor), RARA (retinoic acid receptor, alpha), CREBBP (CREB binding protein), PTGS1 (prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)), GALT (galactose-1-phosphate uridylyltransferase), CHRM1 (cholinergic receptor, muscarinic 1), ATXN1 (ataxin 1), PAWR (PRKC, apoptosis, WT1, regulator), NOTCH2 (Notch homolog 2 (Drosophila)), M6PR (mannose-6-phosphate receptor (cation dependent)), CYP46A1 (cytochrome P450, family 46, subfamily A, polypeptide 1), CSNK1 D (casein kinase 1, delta), MAPK14 (mitogen-activated protein kinase 14), PRG2 (proteoglycan 2, bone marrow (natural killer cell activator, eosinophil granule major basic protein)), PRKCA (protein kinase C, alpha), L1 CAM (L1 cell adhesion molecule), CD40

(CD40 molecule, TNF receptor superfamily member 5), NR1I2 (nuclear receptor subfamily 1, group I, member 2), JAG2 (jagged 2), CTNND1 (catenin (cadherin-associated protein), delta 1), CDH2 (cadherin 2, type 1, N-cadherin (neuronal)), CMA1 (chymase 1, mast cell), SORT1 (sortilin 1), DLK1 (delta-like 1 homolog (Drosophila)), THEM4 (thioesterase superfamily member 4), JUP (junction plakoglobin), CD46 (CD46 molecule, complement regulatory protein), CCL11 (chemokine (C-C motif) ligand 11), CAV3 (caveolin 3), RNASE3 (ribonuclease, RNase A family, 3 (eosinophil cationic protein)), HSPA8 (heat shock 70kDa protein 8), CASP9 (caspase 9, apoptosis-related cysteine peptidase), CYP3A4 (cytochrome P450, family 3, subfamily A, polypeptide 4), CCR3 (chemokine (C-C motif) receptor 3), TFAP2A (transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)), SCP2 (sterol carrier protein 2), CDK4 (cyclin-dependent kinase 4), HIF1A (hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)), TCF7L2 (transcription factor 7-like 2 (T-cell specific, HMG-box)), IL1R2 (interleukin 1 receptor, type II), B3GALTL (beta 1,3-galactosyltransferase-like), MDM2 (Mdm2 p53 binding protein homolog (mouse)), RELA (v-rel reticuloendotheliosis viral oncogene homolog A (avian)), CASP7 (caspase 7, apoptosis-related cysteine peptidase), IDE (insulin-degrading enzyme), FABP4 (fatty acid binding protein 4, adipocyte), CASK (calcium/calmodulin-dependent serine protein kinase (MAGUK family)), ADCYAP1R1 (adenylate cyclase activating polypeptide 1 (pituitary) receptor type I), ATF4 (activating transcription factor 4 (tax-responsive enhancer element B67)), PDGFA (platelet-derived growth factor alpha polypeptide), C21 or f33 (chromosome 21 open reading frame 33), SCG5 (secretogranin V (7B2 protein)), RNF123 (ring finger protein 123), NFKB1 (nuclear factor of kappa light polypeptide gene enhancer in B-cells 1), ERBB2 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 2, neuro/glioblastoma derived oncogene homolog (avian)), CAV1 (caveolin 1, caveolae protein, 22 kDa), MMP7 (matrix metallopeptidase 7 (matrilysin, uterine)), TGFA (transforming growth factor, alpha), RXRA (retinoid X receptor, alpha), STX1A (syntaxin 1A (brain)), PSMC4 (proteasome (prosome, macropain) 26S subunit, ATPase, 4), P2RY2 (purinergic receptor P2Y, G-protein coupled, 2), TNFRSF21 (tumor necrosis factor receptor superfamily, member 21), DLG1 (discs, large homolog 1 (Drosophila)), NUMBL (numb homolog (Drosophila)-like), SPN (sialophorin), PLSCR1 (phospholipid scramblase 1), UBQLN2 (ubiquilin 2), UBQLN1 (ubiquilin 1), PCSK7 (proprotein convertase subtilisin/kexin type 7), SPON1 (spondin 1, extracellular matrix protein), SILV (silver homolog (mouse)), QPCT (glutaminyl-peptide cyclotransferase), HESS (hairy and enhancer of split 5 (Drosophila)), GCC1 (GRIP and coiled-coil domain containing 1), and any combination thereof.

[00661] The genetically modified animal or cell may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more disrupted chromosomal sequences encoding a protein associated with a secretase disorder and zero, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more chromosomally integrated sequences encoding a disrupted protein associated with a secretase disorder.

ALS

[00662] US Patent Publication No. 20110023144, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with amyotrophyic lateral sclerosis (ALS) disease. ALS is characterized by the gradual steady degeneration of certain nerve cells in the brain cortex, brain stem, and spinal cord involved in voluntary movement.

[00663] Motor neuron disorders and the proteins associated with these disorders are a diverse set of proteins that effect susceptibility for developing a motor neuron disorder, the presence of the motor neuron disorder, the severity of the motor neuron disorder or any combination thereof. The present disclosure comprises editing of any chromosomal sequences that encode proteins associated with ALS disease, a specific motor neuron disorder. The proteins associated with ALS are typically selected based on an experimental association of ALS--related proteins to ALS. For example, the production rate or circulating concentration of a protein associated with ALS may be elevated or depressed in a population with ALS relative to a population without ALS. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the proteins associated with ALS may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[00664] By way of non-limiting example, proteins associated with ALS include but are not limited to the following proteins: SOD1 superoxide dismutase 1, ALS3 amyotrophic lateral soluble sclerosis 3 SETX senataxin ALS5 amyotrophic lateral sclerosis 5 FUS fused in sarcoma ALS7 amyotrophic lateral sclerosis 7 ALS2 amyotrophic lateral DPP6 Dipeptidyl-peptidase 6 sclerosis 2 NEFH neurofilament, heavy PTGS1 prostaglandin- polypeptide endoperoxide synthase 1 SLC1A2 solute carrier family 1 TNFRSF10B tumor necrosis factor (glial high affinity receptor superfamily, glutamate transporter), member 10b member 2 PRPH peripherin HSP90AA1 heat shock protein 90 kDa alpha (cytosolic), class A member 1 GRIA2 glutamate receptor, IFNG interferon, gamma ionotropic, AMPA 2 S100B S100 calcium binding FGF2 fibroblast growth factor 2 protein B AOX1 aldehyde oxidase 1 CS citrate synthase TARDBP

TAR DNA binding protein TXN thioredoxin RAPH1 Ras association MAP3K5 mitogen-activated protein (RaIGDS/AF-6) and kinase 5 pleckstrin homology domains 1 NBEAL1 neurobeachin-like 1 GPX1 glutathione peroxidase 1 ICA1L islet cell autoantigen RAC1 ras-related C3 botulinum 1.69 kDa-like toxin substrate 1 MAPT microtubule-associated ITPR2 inositol 1,4,5- protein tau triphosphate receptor, type 2 ALS2CR4 amyotrophic lateral GLS glutaminase sclerosis 2 (juvenile) chromosome region, candidate 4 ALS2CR8 amyotrophic lateral CNTFR ciliary neurotrophic factor sclerosis 2 (juvenile) receptor chromosome region, candidate 8 ALS2CR11 amyotrophic lateral FOLH1 folate hydrolase 1 sclerosis 2 (juvenile) chromosome region, candidate 11 FAM117B family with sequence P4HB prolyl 4-hydroxylase, similarity 117, member B beta polypeptide CNTF ciliary neurotrophic factor SQSTM1 sequestosome 1 STRADB STE20-related kinase NAIP NLR family, apoptosis adaptor beta inhibitory protein YWHAQ tyrosine 3- SLC33A1 solute carrier family 33 monooxygenase/tryptoph (acetyl-CoA transporter), an 5-monooxygenase member 1 activation protein, theta polypeptide TRAK2 trafficking protein, FIG. 4 FIG. 4 homolog, SAC1 kinesin binding 2 lipid phosphatase domain containing NIF3L1 NIF3 NGG1 interacting INA internexin neuronal factor 3-like 1 intermediate filament protein, alpha PARD3B par-3 partitioning COX8A cytochrome c oxidase defective 3 homolog B subunit VIIIA CDK15 cyclin-dependent kinase HECW1 HECT, C2 and WW 15 domain containing E3 ubiquitin protein ligase 1 NOS1 nitric oxide synthase 1 MET met proto-oncogene SOD2 superoxide dismutase 2, HSPB1 heat shock 27 kDa mitochondrial protein 1 NEFL neurofilament, light CTSB cathepsin B polypeptide ANG angiogenin, HSPA8 heat shock 70 kDa ribonuclease, RNase A protein 8 family, 5 VAPB VAMP (vesicle- ESR1 estrogen receptor 1 associated membrane protein)-associated protein B and C SNCA synuclein, alpha HGF hepatocyte growth factor CAT catalase ACTB actin, beta NEFM neurofilament, medium TH tyrosine hydroxylase polypeptide BCL2 B-cell CLL/lymphoma 2 FAS Fas (TNF receptor superfamily, member 6) CASP3 caspase 3, apoptosis- CLU clusterin related cysteine peptidase SMN1 survival of motor neuron G6PD glucose-6-phosphate 1, telomeric dehydrogenase BAX BCL2-associated X HSF1 heat shock transcription protein factor 1 RNF19A ring finger protein 19A JUN jun oncogene ALS2CR12 amyotrophic lateral HSPA5 heat shock 70 kDa sclerosis 2 (juvenile) protein 5 chromosome region, candidate 12 MAPK14 mitogen-activated protein IL10 interleukin 10 kinase 14 APEX1 APEX nuclease TXNRD1 thioredoxin reductase 1 (multifunctional DNA repair enzyme) 1 NOS2 nitric oxide synthase 2, TIMP1 TIMP metallopeptidase inducible inhibitor 1 CASP9 caspase 9, apoptosis- XIAP X-linked inhibitor of related cysteine apoptosis peptidase GLG1 golgi glycoprotein 1 EPO erythropoietin VEGFA

vascular endothelial ELN elastin growth factor A GDNF glial cell derived NFE2L2 nuclear factor (erythroid- neurotrophic factor derived 2)-like 2 SLC6A3 solute carrier family 6 HSPA4 heat shock 70 kDa (neurotransmitter protein 4 transporter, dopamine), member 3 APOE apolipoprotein E PSMB8 proteasome (prosome, macropain) subunit, beta type, 8 DCTN1 dynactin 1 TIMP3 TIMP metallopeptidase inhibitor 3 KIFAP3 kinesin-associated SLC1A1 solute carrier family 1 protein 3 (neuronal/epithelial high affinity glutamate transporter, system Xag), member 1 SMN2 survival of motor neuron CCNC cyclin C 2, centromeric MPP4 membrane protein, STUB1 STIP1 homology and U- palmitoylated 4 box containing protein 1 ALS2 amyloid beta (A4) PRDX6 peroxiredoxin 6 precursor protein SYP synaptophysin CABIN1 calcineurin binding protein 1 CASP1 caspase 1, apoptosis- GART phosphoribosylglycinami related cysteine de formyltransferase, peptidase phosphoribosylglycinami de synthetase, phosphoribosylaminoimi dazole synthetase CDK5 cyclin-dependent kinase 5 ATXN3 ataxin 3 RTN4 reticulon 4 C1QB complement component 1, q subcomponent, B chain VEGFC nerve growth factor HTT huntingtin receptor PARK7 Parkinson disease 7 XDH xanthine dehydrogenase GFAP glial fibrillary acidic MAP2 microtubule-associated protein protein 2 CYCS cytochrome c, somatic FCGR3B Fc fragment of IgG, low affinity IIIb, CCS copper chaperone for UBL5 ubiquitin-like 5 superoxide dismutase MMP9 matrix metallopeptidase SLC18A3 solute carrier family 18 9 ((vesicular acetylcholine), member 3 TRPM7 transient receptor HSPB2 heat shock 27 kDa potential cation channel, protein 2 subfamily M, member 7 AKT1 v-akt murine thymoma DERL1 Der1-like domain family, viral oncogene homolog 1 member 1 CCL2 chemokine (C--C motif) NGRN neugrin, neurite ligand 2 outgrowth associated GSR glutathione reductase TPPP3 tubulin polymerization- promoting protein family member 3 APAF1 apoptotic peptidase BTBD10 BTB (POZ) domain activating factor 1 containing 10 GLUD1 glutamate CXCR4 chemokine (C--X--C motif) dehydrogenase 1 receptor 4 SLC1A3 solute carrier family 1 FLT1 fms-related tyrosine (glial high affinity glutamate transporter), member 3 kinase 1 PON1 paraoxonase 1 AR androgen receptor LIF leukemia inhibitory factor ERBB3 v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 LGALS1 lectin, galactoside- CD44 CD44 molecule binding, soluble, 1 TP53 tumor protein p53 TLR3 toll-like receptor 3 GRIA1 glutamate receptor, GAPDH glyceraldehyde-3- ionotropic, AMPA 1 phosphate dehydrogenase GRIK1 glutamate receptor, DES desmin ionotropic, kainate 1 CHAT choline acetyltransferase FLT4 fms-related tyrosine kinase 4 CHMP2B chromatin modifying BAG1 BCL2-associated protein 2B athanogene MT3 metallothionein 3 CHRNA4 cholinergic receptor, nicotinic, alpha 4 GSS glutathione synthetase BAK1 BCL2-antagonist/killer 1 KDR kinase insert domain GSTP1

glutathione S-transferase receptor (a type III pi 1 receptor tyrosine kinase) OGG1 8-oxoguanine DNA IL6 interleukin 6 (interferon, glycosylase beta 2).

[00665] The animal or cell may comprise 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more disrupted chromosomal sequences encoding a protein associated with ALS and zero, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 or more chromosomally integrated sequences encoding the disrupted protein associated with ALS. Preferred proteins associated with ALS include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

Autism

[00666] US Patent Publication No. 20110023145, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with autism spectrum disorders (ASD). Autism spectrum disorders (ASDs) are a group of disorders characterized by qualitative impairment in social interaction and communication, and restricted repetitive and stereotyped patterns of behavior, interests, and activities. The three disorders, autism, Asperger syndrome (AS) and pervasive developmental disorder-not otherwise specified (PDD-NOS) are a continuum of the same disorder with varying degrees of severity, associated intellectual functioning and medical conditions. ASDs are predominantly genetically determined disorders with a heritability of around 90%.

[00667] US Patent Publication No. 20110023145 comprises editing of any chromosomal sequences that encode proteins associated with ASD which may be applied to the CRISPR Cas system of the present invention. The proteins associated with ASD are typically selected based on an experimental association of the protein associated with ASD to an incidence or indication of an ASD. For example, the production rate or circulating concentration of a protein associated with ASD may be elevated or depressed in a population having an ASD relative to a population lacking the ASD. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the proteins associated with ASD may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[00668] Non limiting examples of disease states or disorders that may be associated with proteins associated with ASD include autism, Asperger syndrome (AS), pervasive developmental disorder-not otherwise specified (PDD-NOS), Rett's syndrome, tuberous sclerosis, phenylketonuria, Smith-Lemli-Opitz syndrome and fragile X syndrome. By way of non-limiting example, proteins associated with ASD include but are not limited to the following proteins: ATP10C aminophospholipid- MET MET receptor transporting ATPase tyrosine kinase (ATP10C) BZRAP1 MGLUR5 (GRM5) Metabotropic glutamate receptor 5 (MGLUR5) CDH10 Cadherin-10 MGLUR6 (GRM6) Metabotropic glutamate receptor 6 (MGLUR6) CDH9 Cadherin-9 NLGN1 Neuroligin-1 CNTN4 Contactin-4 NLGN2 Neuroligin-2 CNTNAP2 Contactin-associated SEMA5A Neuroligin-3 protein-like 2 (CNTNAP2) DHCR7 7-dehydrocholesterol NLGN4X Neuroligin-4 X- reductase (DHCR7) linked DOC2A Double C2-like domain- NLGN4Y Neuroligin-4 Y- containing protein alpha linked DPP6 Dipeptidyl NLGN5 Neuroligin-5 aminopeptidase-like protein 6 EN2 engrailed 2 (EN2) NRCAM Neuronal cell adhesion molecule (NRCAM) MDGA2 fragile X mental retardation NRXN1 Neurexin-1 1 (MDGA2) FMR2 (AFF2) AF4/FMR2 family member 2 OR4M2 Olfactory receptor (AFF2) 4M2 FOXP2 Forkhead box protein P2 OR4N4 Olfactory receptor (FOXP2) 4N4 FXR1 Fragile X mental OXTR oxytocin receptor retardation, autosomal (OXTR) homolog 1 (FXR1) FXR2 Fragile X mental PAH phenylalanine retardation, autosomal hydroxylase (PAH) homolog 2 (FXR2) GABRA1 Gamma-aminobutyric acid PTEN Phosphatase and receptor subunit alpha-1 tensin homologue (GABRA1) (PTEN) GABRA5 GABAA (.gamma.-aminobutyric PTPRZ1 Receptor-type acid) receptor alpha 5 tyrosine-protein subunit (GABRA5) phosphatase zeta (PTPRZ1) GABRB1 Gamma-aminobutyric acid RELN Reelin receptor subunit beta-1 (GABRB1) GABRB3 GABAA (.gamma.-aminobutyric RPL10 60S ribosomal acid) receptor .beta.3 subunit protein L10 (GABRB3) GABRG1 Gamma-aminobutyric acid SEMA5A Semaphorin-5A receptor subunit gamma-1 (SEMA5A) (GABRG1) HIRIP3 HIRA-interacting protein 3 SEZ6L2 seizure related 6 homolog (mouse)- like 2 HOXA1 Homeobox protein Hox-A1 SHANK3 SH3 and multiple (HOXA1) ankyrin repeat domains 3 (SHANK3) IL6 Interleukin-6 SHBZRAP1 SH3 and multiple ankyrin repeat domains 3 (SHBZRAP1) LAMB1 Laminin subunit beta-1 SLC6A4 Serotonin (LAMB1) transporter (SERT) MAPK3 Mitogen-activated protein TAS2R1 Taste receptor kinase 3 type 2 member 1 TAS2R1 MAZ Myc-associated zinc finger TSC1 Tuberous sclerosis protein protein 1 MDGA2 MAM domain containing TSC2 Tuberous sclerosis glycosylphosphatidylinositol protein 2 anchor 2 (MDGA2) MECP2 Methyl CpG binding UBE3A Ubiquitin protein protein 2 (MECP2) ligase E3A (UBE3A) MECP2 methyl

CpG binding WNT2 Wingless-type protein 2 (MECP2) MMTV integration site family, member 2 (WNT2)

[00669] The identity of the protein associated with ASD whose chromosomal sequence is edited can and will vary. In preferred embodiments, the proteins associated with ASD whose chromosomal sequence is edited may be the benzodiazapine receptor (peripheral) associated protein 1 (BZRAP1) encoded by the BZRAP1 gene, the AF4/FMR2 family member 2 protein (AFF2) encoded by the AFF2 gene (also termed MFR2), the fragile X mental retardation autosomal homolog 1 protein (FXR1) encoded by the FXR1 gene, the fragile X mental retardation autosomal homolog 2 protein (FXR2) encoded by the FXR2 gene, the MAM domain containing glycosylphosphatidylinositol anchor 2 protein (MDGA2) encoded by the MDGA2 gene, the methyl CpG binding protein 2 (MECP2) encoded by the MECP2 gene, the metabotropic glutamate receptor 5 (MGLUR5) encoded by the MGLUR5-1 gene (also termed GRM5), the neurexin 1 protein encoded by the NRXN1 gene, or the semaphorin-5A protein (SEMA5A) encoded by the SEMA5A gene. In an exemplary embodiment, the genetically modified animal is a rat, and the edited chromosomal sequence encoding the protein associated with ASD is as listed below: BZRAP1 benzodiazapine receptor XM_002727789, (peripheral) associated XM_213427, protein 1 (BZRAP1) XM_002724533, XM_001081125 AFF2 (FMR2) AF4/FMR2 family member 2 XM_219832, (AFF2) XM_001054673 FXR1 Fragile X mental NM_001012179 retardation, autosomal homolog 1 (FXR1) FXR2 Fragile X mental NM_001100647 retardation, autosomal homolog 2 (FXR2) MDGA2 MAM domain containing NM_199269 glycosylphosphatidylinositol anchor 2 (MDGA2) MECP2 Methyl CpG binding NM_022673 protein 2 (MECP2) MGLUR5 Metabotropic glutamate NM_017012 (GRM5) receptor 5 (MGLUR5) NRXN1 Neurexin-1 NM_021767 SEMA5A Semaphorin-5A (SEMA5A) NM_001107659.

Trinucleotide Repeat Expansion Disorders

[00670] US Patent Publication No. 20110016540, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with trinucleotide repeat expansion disorders. Trinucleotide repeat expansion disorders are complex, progressive disorders that involve developmental neurobiology and often affect cognition as well as sensori-motor functions.

[00671] Trinucleotide repeat expansion proteins are a diverse set of proteins associated with susceptibility for developing a trinucleotide repeat expansion disorder, the presence of a trinucleotide repeat expansion disorder, the severity of a trinucleotide repeat expansion disorder or any combination thereof. Trinucleotide repeat expansion disorders are divided into

two categories determined by the type of repeat. The most common repeat is the triplet CAG, which, when present in the coding region of a gene, codes for the amino acid glutamine (Q). Therefore, these disorders are referred to as the polyglutamine (polyQ) disorders and comprise the following diseases: Huntington Disease (HD); Spinobulbar Muscular Atrophy (SBMA); Spinocerebellar Ataxias (SCA types 1, 2, 3, 6, 7, and 17); and Dentatorubro-Pallidoluysian Atrophy (DRPLA). The remaining trinucleotide repeat expansion disorders either do not involve the CAG triplet or the CAG triplet is not in the coding region of the gene and are, therefore, referred to as the non-polyglutamine disorders. The non-polyglutamine disorders comprise Fragile X Syndrome (FRAXA); Fragile XE Mental Retardation (FRAXE); Friedreich Ataxia (FRDA); Myotonic Dystrophy (DM); and Spinocerebellar Ataxias (SCA types 8, and 12).

[00672] The proteins associated with trinucleotide repeat expansion disorders are typically selected based on an experimental association of the protein associated with a trinucleotide repeat expansion disorder to a trinucleotide repeat expansion disorder. For example, the production rate or circulating concentration of a protein associated with a trinucleotide repeat expansion disorder may be elevated or depressed in a population having a trinucleotide repeat expansion disorder relative to a population lacking the trinucleotide repeat expansion disorder. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the proteins associated with trinucleotide repeat expansion disorders may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[00673] Non-limiting examples of proteins associated with trinucleotide repeat expansion disorders include AR (androgen receptor), FMR1 (fragile X mental retardation 1), HTT (huntingtin), DMPK (dystrophia myotonica-protein kinase), FXN (frataxin), ATXN2 (ataxin 2), ATN1 (atrophin 1), FEN1 (flap structure-specific endonuclease 1), TNRC6A (trinucleotide repeat containing 6A), PABPN1 (poly(A) binding protein, nuclear 1), JPH3 (junctophilin 3), MED15 (mediator complex subunit 15), ATXN1 (ataxin 1), ATXN3 (ataxin 3), TBP (TATA box binding protein), CACNA1A (calcium channel, voltage-dependent, P/Q type, alpha 1A subunit), ATXN80S (ATXN8 opposite strand (non-protein coding)), PPP2R2B (protein phosphatase 2, regulatory subunit B, beta), ATXN7 (ataxin 7), TNRC6B (trinucleotide repeat containing 6B), TNRC6C (trinucleotide repeat containing 6C), CELF3 (CUGBP, Elav-like

family member 3), MAB21L1 (mab-21-like 1 (C. elegans)), MSH2 (mutS homolog 2, colon cancer, nonpolyposis type 1 (E. coli)), TMEM185A (transmembrane protein 185A), SIX5 (SIX homeobox 5), CNPY3 (canopy 3 homolog (zebrafish)), FRAXE (fragile site, folic acid type, rare, fra(X)(q28) E), GNB2 (guanine nucleotide binding protein (G protein), beta polypeptide 2), RPL14 (ribosomal protein L14), ATXN8 (ataxin 8), INSR (insulin receptor), TTR (transthyretin), EP400 (E1A binding protein p400), GIGYF2 (GRB10 interacting GYF protein 2), OGG1 (8-oxoguanine DNA glycosylase), STC1 (stanniocalcin 1), CNDP1 (carnosine dipeptidase 1 (metallopeptidase M20 family)), C10orf2 (chromosome 10 open reading frame 2), MAML3 mastermind-like 3 (Drosophila), DKC1 (dyskeratosis congenita 1, dyskerin), PAXIP1 (PAX interacting (with transcription-activation domain) protein 1), CASK (calcium/calmodulin-dependent serine protein kinase (MAGUK family)), MAPT (microtubule-associated protein tau), SP1 (Sp1 transcription factor), POLG (polymerase (DNA directed), gamma), AFF2 (AF4/FMR2 family, member 2), THBS1 (thrombospondin 1), TP53 (tumor protein p53), ESR1 (estrogen receptor 1), CGGBP1 (CGG triplet repeat binding protein 1), ABT1 (activator of basal transcription 1), KLK3 (kallikrein-related peptidase 3), PRNP (prion protein), JUN (jun oncogene), KCNN3 (potassium intermediate/small conductance calcium-activated channel, subfamily N, member 3), BAX (BCL2-associated X protein), FRAXA (fragile site, folic acid type, rare, fra(X)(q27.3) A (macroorchidism, mental retardation)), KBTBD10 (kelch repeat and BTB (POZ) domain containing 10), MBNL1 (muscleblind-like (Drosophila)), RAD51 (RAD51 homolog (RecA homolog, E. coli) (S. cerevisiae)), NCOA3 (nuclear receptor coactivator 3), ERDA1 (expanded repeat domain, CAG/CTG 1), TSC1 (tuberous sclerosis 1), COMP (cartilage oligomeric matrix protein), GCLC (glutamate-cysteine ligase, catalytic subunit), RRAD (Ras-related associated with diabetes), MSH3 (mutS homolog 3 (E. coli)), DRD2 (dopamine receptor D2), CD44 (CD44 molecule (Indian blood group)), CTCF (CCCTC-binding factor (zinc finger protein)), CCND1 (cyclin D1), CLSPN (claspin homolog (Xenopus laevis)), MEF2A (myocyte enhancer factor 2A), PTPRU (protein tyrosine phosphatase, receptor type, U), GAPDH (glyceraldehyde-3-phosphate dehydrogenase), TRIM22 (tripartite motif-containing 22), WT1 (Wilms tumor 1), AHR (aryl hydrocarbon receptor), GPX1 (glutathione peroxidase 1), TPMT (thiopurine S-methyltransferase), NDP (Norrie disease (pseudoglioma)), ARX (aristaless related homeobox), MUS81 (MUS81 endonuclease homolog (S. cerevisiae)), TYR (tyrosinase (oculocutaneous albinism IA)), EGR1 (early growth response 1), UNG (uracil-DNA glycosylase), NUMBL (numb homolog (Drosophila)-like), FABP2 (fatty acid binding protein 2, intestinal), EN2 (engrailed homeobox 2), CRYGC (crystallin, gamma C), SRP14 (signal recognition particle

14 kDa (homologous Alu RNA binding protein)), CRYGB (crystallin, gamma B), PDCD1 (programmed cell death 1), HOXA1 (homeobox A1), ATXN2L (ataxin 2-like), PMS2 (PMS2 postmeiotic segregation increased 2 (S. cerevisiae)), GLA (galactosidase, alpha), CBL (Cas-Br-M (murine) ecotropic retroviral transforming sequence), FTH1 (ferritin, heavy polypeptide 1), IL12RB2 (interleukin 12 receptor, beta 2), OTX2 (orthodenticle homeobox 2), HOXA5 (homeobox A5), POLG2 (polymerase (DNA directed), gamma 2, accessory subunit), DLX2 (distal-less homeobox 2), SIRPA (signal-regulatory protein alpha), OTX1 (orthodenticle homeobox 1), AHRR (aryl-hydrocarbon receptor repressor), MANF (mesencephalic astrocyte-derived neurotrophic factor), TMEM158 (transmembrane protein 158 (gene/pseudogene)), and ENSG00000078687.

[00674] Preferred proteins associated with trinucleotide repeat expansion disorders include HTT (Huntingtin), AR (androgen receptor), FXN (frataxin), Atxn3 (ataxin), Atxn1 (ataxin), Atxn2 (ataxin), Atxn7 (ataxin), Atxn10 (ataxin), DMPK (dystrophia myotonica-protein kinase), Atn1 (atrophin 1), CBP (creb binding protein), VLDLR (very low density lipoprotein receptor), and any combination thereof.

Treating Hearing Diseases

[00675] The present invention also contemplates delivering the CRISPR-Cas system to one or both ears.

[00676] Researchers are looking into whether gene therapy could be used to aid current deafness treatments - namely, cochlear implants. Deafness is often caused by lost or damaged hair cells that cannot relay signals to auditory neurons. In such cases, cochlear implants may be used to respond to sound and transmit electrical signals to the nerve cells. But these neurons often degenerate and retract from the cochlea as fewer growth factors are released by impaired hair cells.

[00677] US patent application 20120328580 describes injection of a pharmaceutical composition into the ear (e.g., auricular administration), such as into the luminae of the cochlea (e.g., the Scala media, Sc vestibulae, and Sc tympani), e.g., using a syringe, e.g., a single-dose syringe. For example, one or more of the compounds described herein can be administered by intratympanic injection (e.g., into the middle ear), and/or injections into the outer, middle, and/or inner ear. Such methods are routinely used in the art, for example, for the administration of steroids and antibiotics into human ears. Injection can be, for example, through the round window of the ear or through the cochlear capsule. Other inner ear administration methods are known in the art (see, e.g., Salt and Plontke, Drug Discovery Today, 10:1299-1306, 2005).

[00678] In another mode of administration, the pharmaceutical composition can be administered in situ, via a catheter or pump. A catheter or pump can, for example, direct a pharmaceutical composition into the cochlear luminae or the round window of the ear and/or the lumen of the colon. Exemplary drug delivery apparatus and methods suitable for administering one or more of the compounds described herein into an ear, e.g., a human ear, are described by McKenna et al., (U.S. Publication No. 2006/0030837) and Jacobsen et al., (U.S. Pat. No. 7,206,639). In some embodiments, a catheter or pump can be positioned, e.g., in the ear (e.g., the outer, middle, and/or inner ear) of a patient during a surgical procedure. In some embodiments, a catheter or pump can be positioned, e.g., in the ear (e.g., the outer, middle, and/or inner ear) of a patient without the need for a surgical procedure.

[00679] Alternatively or in addition, one or more of the compounds described herein can be administered in combination with a mechanical device such as a cochlear implant or a hearing aid, which is worn in the outer ear. An exemplary cochlear implant that is suitable for use with the present invention is described by Edge et al., (U.S. Publication No. 2007/0093878).

[00680] In some embodiments, the modes of administration described above may be combined in any order and can be simultaneous or interspersed.

[00681] Alternatively or in addition, the present invention may be administered according to any of the Food and Drug Administration approved methods, for example, as described in CDER Data Standards Manual, version number 004 (which is available at fda.give/cder/dsm/DRG/drg00301.htm).

[00682] In general, the cell therapy methods described in US patent application 20120328580 can be used to promote complete or partial differentiation of a cell to or towards a mature cell type of the inner ear (e.g., a hair cell) in vitro. Cells resulting from such methods can then be transplanted or implanted into a patient in need of such treatment. The cell culture methods required to practice these methods, including methods for identifying and selecting suitable cell types, methods for promoting complete or partial differentiation of selected cells, methods for identifying complete or partially differentiated cell types, and methods for implanting complete or partially differentiated cells are described below.

[00683] Cells suitable for use in the present invention include, but are not limited to, cells that are capable of differentiating completely or partially into a mature cell of the inner ear, e.g., a hair cell (e.g., an inner and/or outer hair cell), when contacted, e.g., in vitro, with one or more of the compounds described herein. Exemplary cells that are capable of differentiating into a hair cell include, but are not limited to stem cells (e.g., inner ear stem cells, adult stem cells, bone marrow derived stem cells, embryonic stem cells, mesenchymal stem cells, skin

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

[00684] The CRISPR Cas molecules of the present invention may be delivered to the ear by direct application of pharmaceutical composition to the outer ear, with compositions modified from US Published application, 20110142917. In some embodiments the pharmaceutical composition is applied to the ear canal. Delivery to the ear may also be referred to as aural or otic delivery.

[00685] Qi et al. discloses methods for efficient siRNA transfection to the inner ear through the intact round window by a novel proteidic delivery technology which may be applied to the nucleic acid-targeting system of the present invention (see, e.g., Qi et al., Gene Therapy (2013), 1-9). In particular, a TAT double stranded RNA-binding domains (TAT-DRBDs), which can transfect Cy3-labeled siRNA into cells of the inner ear, including the inner and outer hair cells, crista ampullaris, macula utriculi and macula sacculi, through intact round-window permeation was successful for delivering double stranded siRNAs in vivo for treating various inner ear

ailments and preservation of hearing function. About 40 µl of 10mM RNA may be contemplated as the dosage for administration to the ear.

[00686] According to Rejali et al. (Hear Res. 2007 Jun;228(1-2):180-7), cochlear implant function can be improved by good preservation of the spiral ganglion neurons, which are the target of electrical stimulation by the implant and brain derived neurotrophic factor (BDNF) has previously been shown to enhance spiral ganglion survival in experimentally deafened ears. Rejali et al. tested a modified design of the cochlear implant electrode that includes a coating of fibroblast cells transduced by a viral vector with a BDNF gene insert. To accomplish this type of ex vivo gene transfer, Rejali et al. transduced guinea pig fibroblasts with an adenovirus with a BDNF gene cassette insert, and determined that these cells secreted BDNF and then attached BDNF-secreting cells to the cochlear implant electrode via an agarose gel, and implanted the electrode in the scala tympani. Rejali et al. determined that the BDNF expressing electrodes were able to preserve significantly more spiral ganglion neurons in the basal turns of the cochlea after 48 days of implantation when compared to control electrodes and demonstrated the feasibility of combining cochlear implant therapy with ex vivo gene transfer for enhancing spiral ganglion neuron survival. Such a system may be applied to the nucleic acid-targeting system of the present invention for delivery to the ear.

[00687] Mukherjea et al. (Antioxidants & Redox Signaling, Volume 13, Number 5, 2010) document that knockdown of NOX3 using short interfering (si) RNA abrogated cisplatin ototoxicity, as evidenced by protection of OHCs from damage and reduced threshold shifts in auditory brainstem responses (ABRs). Different doses of siNOX3 (0.3, 0.6, and 0.9 µg) were administered to rats and NOX3 expression was evaluated by real time RT-PCR. The lowest dose of NOX3 siRNA used (0.3 µg) did not show any inhibition of NOX3 mRNA when compared to transtympanic administration of scrambled siRNA or untreated cochleae. However, administration of the higher doses of NOX3 siRNA (0.6 and 0.9 µg) reduced NOX3 expression compared to control scrambled siRNA. Such a system may be applied to the CRISPR Cas system of the present invention for transtympanic administration with a dosage of about 2 mg to about 4 mg of CRISPR Cas for administration to a human.

[00688] Jung et al. (Molecular Therapy, vol. 21 no. 4, 834–841 apr. 2013) demonstrate that Hes5 levels in the utricle decreased after the application of siRNA and that the number of hair cells in these utricles was significantly larger than following control treatment. The data suggest that siRNA technology may be useful for inducing repair and regeneration in the inner ear and that the Notch signaling pathway is a potentially useful target for specific gene expression inhibition. Jung et al. injected 8 µg of Hes5 siRNA in 2 µl volume, prepared by

adding sterile normal saline to the lyophilized siRNA to a vestibular epithelium of the ear. Such a system may be applied to the nucleic acid-targeting system of the present invention for administration to the vestibular epithelium of the ear with a dosage of about 1 to about 30 mg of CRISPR Cas for administration to a human.

[00689] Gene Targeting in Non-Dividing Cells (Neurones & Muscle)

[00690] Non-dividing (especially non-dividing, fully differentiated) cell types present issues for gene targeting or genome engineering, for example because homologous recombination (HR) is generally supressed in the G1 cell-cycle phase. However, while studying the mechanisms by which cells control normal DNA repair systems, Durocher discovered a previously unknown switch that keeps HR "off" in non-dividing cells and devised a strategy to toggle this switch back on. Orthwein et al. (Daniel Durocher's lab at the Mount Sinai Hospital in Ottawa, Canada) recently reported (Nature 16142, published online 9 Dec 2015) have shown that the suppression of HR can be lifted and gene targeting successfully concluded in both kidney (293T) and osteosarcoma (U2OS) cells. Tumor suppressors, BRCA1, PALB2 and BRAC2 are known to promote DNA DSB repair by HR. They found that formation of a complex of BRCA1 with PALB2 - BRAC2 is governed by a ubiquitin site on PALB2, such that action on the site by an E3 ubiquitin ligase. This E3 ubiquitin ligase is composed of KEAP1 (a PALB2 -interacting protein) in complex with cullin-3 (CUL3)–RBX1. PALB2 ubiquitylation suppresses its interaction with BRCA1 and is counteracted by the deubiquitylase USP11, which is itself under cell cycle control. Restoration of the BRCA1–PALB2 interaction combined with the activation of DNA-end resection is sufficient to induce homologous recombination in G1, as measured by a number of methods including a CRISPR–Cas9-based gene-targeting assay directed at USP11 or KEAP1 (expressed from a pX459 vector). However, when the BRCA1–PALB2 interaction was restored in resection-competent G1 cells using either KEAP1 depletion or expression of the PALB2-KR mutant, a robust increase in gene-targeting events was detected.

[00691] Thus, reactivation of HR in cells, especially non-dividing, fully differentiated cell types is preferred, in some embodiments. In some embodiments, promotion of the BRCA1–PALB2 interaction is preferred in some embodiments. In some embodiments, the target ell is a non-dividing cell. In some embodiments, the target cell is a neurone or muscle cell. In some embodiments, the target cell is targeted in vivo. In some embodiments, the cell is in G1 and HR is suppressed. In some embodiments, use of KEAP1 depletion, for example inhibition of expression of KEAP1 activity, is preferred. KEAP1 depletion may be achieved through siRNA, for example as shown in Orthwein et al. Alternatively, expression of the PALB2-KR

mutant (lacking all eight Lys residues in the BRCA1-interaction domain is preferred, either in combination with KEAP1 depletion or alone. PALB2-KR interacts with BRCA1 irrespective of cell cycle position. Thus, promotion or restoration of the BRCA1-PALB2 interaction, especially in G1 cells, is preferred in some embodiments, especially where the target cells are non-dividing, or where removal and return (ex vivo gene targeting) is problematic, for example neurone or muscle cells. KEAP1 siRNA is available from ThermoFischer. In some embodiments, a BRCA1–PALB2 complex may be delivered to the G1 cell. In some embodiments, PALB2 deubiquitylation may be promoted for example by increased expression of the deubiquitylase USP11, so it is envisaged that a construct may be provided to promote or up-regulate expression or activity of the deubiquitylase USP11.

Treating Diseases of the Eye

[00692] The present invention also contemplates delivering the CRISPR-Cas system to one or both eyes.

[00693] In particular embodiments of the invention, the CRISPR-Cas system may be used to correct ocular defects that arise from several genetic mutations further described in Genetic Diseases of the Eye, Second Edition, edited by Elias I. Traboulsi, Oxford University Press, 2012.

[00694] In some embodiments, the condition to be treated or targeted is an eye disorder. In some embodiments, the eye disorder may include glaucoma. In some embodiments, the eye disorder includes a retinal degenerative disease. In some embodiments, the retinal degenerative disease is selected from Stargardt disease, Bardet-Biedl Syndrome, Best disease, Blue Cone Monochromacy, Choroidermia, Cone-rod dystrophy, Congenital Stationary Night Blindness, Enhanced S-Cone Syndrome, Juvenile X-Linked Retinoschisis, Leber Congenital Amaurosis, Malattia Leventinesse, Norrie Disease or X-linked Familial Exudative Vitreoretinopathy, Pattern Dystrophy, Sorsby Dystrophy, Usher Syndrome, Retinitis Pigmentosa, Achromatopsia or Macular dystrophies or degeneration, Retinitis Pigmentosa, Achromatopsia, and age related macular degeneration. In some embodiments, the retinal degenerative disease is Leber Congenital Amaurosis (LCA) or Retinitis Pigmentosa. In some embodiments, the CRISPR system is delivered to the eye, optionally via intravitreal injection or subretinal injection.

[00695] For administration to the eye, lentiviral vectors, in particular equine infectious anemia viruses (EIAV) are particularly preferred.

[00696] In another embodiment, minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV) are also contemplated, especially for ocular gene therapy (see, e.g., Balagaan, J Gene Med 2006; 8: 275 – 285, Published online 21 November 2005 in Wiley

InterScience (www.interscience.wiley.com). DOI: 10.1002/jgm.845). The vectors are contemplated to have cytomegalovirus (CMV) promoter driving expression of the target gene. Intracameral, subretinal, intraocular and intravitreal injections are all contemplated (see, e.g., Balagaan, J Gene Med 2006; 8: 275 – 285, Published online 21 November 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/jgm.845). Intraocular injections may be performed with the aid of an operating microscope. For subretinal and intravitreal injections, eyes may be prolapsed by gentle digital pressure and fundi visualized using a contact lens system consisting of a drop of a coupling medium solution on the cornea covered with a glass microscope slide coverslip. For subretinal injections, the tip of a 10-mm 34-gauge needle, mounted on a 5-μl Hamilton syringe may be advanced under direct visualization through the superior equatorial sclera tangentially towards the posterior pole until the aperture of the needle was visible in the subretinal space. Then, 2 μl of vector suspension may be injected to produce a superior bullous retinal detachment, thus confirming subretinal vector administration. This approach creates a self-sealing sclerotomy allowing the vector suspension to be retained in the subretinal space until it is absorbed by the RPE, usually within 48 h of the procedure. This procedure may be repeated in the inferior hemisphere to produce an inferior retinal detachment. This technique results in the exposure of approximately 70% of neurosensory retina and RPE to the vector suspension. For intravitreal injections, the needle tip may be advanced through the sclera 1 mm posterior to the corneoscleral limbus and 2 μl of vector suspension injected into the vitreous cavity. For intracameral injections, the needle tip may be advanced through a corneoscleral limbal paracentesis, directed towards the central cornea, and 2 μl of vector suspension may be injected. For intracameral injections, the needle tip may be advanced through a corneoscleral limbal paracentesis, directed towards the central cornea, and 2 μl of vector suspension may be injected. These vectors may be injected at titres of either 1.0–1.4 × 1010 or 1.0–1.4 × 109 transducing units (TU)/ml.

[00697] In another embodiment, RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostain and angiostatin that is delivered via a subretinal injection for the treatment of the web form of age-related macular degeneration is also contemplated (see, e.g., Binley et al., HUMAN GENE THERAPY 23:980–991 (September 2012)). Such a vector may be modified for the CRISPR-Cas system of the present invention. Each eye may be treated with either RetinoStat® at a dose of 1.1 x 105 transducing units per eye (TU/eye) in a total volume of 100 μl.

[00698] In another embodiment, an E1-, partial E3-, E4-deleted adenoviral vector may be contemplated for delivery to the eye. Twenty-eight patients with advanced neovascular

agerelated macular degeneration (AMD) were given a single intravitreous injection of an E1-, partial E3-, E4-deleted adenoviral vector expressing human pigment ep- ithelium-derived factor (AdPEDF.ll) (see, e.g., Campochiaro et al., Human Gene Therapy 17:167-176 (February 2006)). Doses ranging from 106 to 109.5 particle units (PU) were investigated and there were no serious adverse events related to AdPEDF.ll and no dose-limiting toxicities (see, e.g., Campochiaro et al., Human Gene Therapy 17:167-176 (February 2006)). Adenoviral vectormediated ocular gene transfer appears to be a viable approach for the treatment of ocular disorders and could be applied to the CRISPR Cas system.

[00699] In another embodiment, the sd-rxRNA® system of RXi Pharmaceuticals may be used/and or adapted for delivering CRISPR Cas to the eye. In this system, a single intravitreal administration of 3 μg of sd-rxRNA results in sequence-specific reduction of PPIB mRNA levels for 14 days. The sd-rxRNA® system may be applied to the nucleic acid-targeting system of the present invention, contemplating a dose of about 3 to 20 mg of CRISPR administered to a human.

[00700] Millington-Ward et al. (Molecular Therapy, vol. 19 no. 4, 642–649 apr. 2011) describes adeno-associated virus (AAV) vectors to deliver an RNA interference (RNAi)-based rhodopsin suppressor and a codon-modified rhodopsin replacement gene resistant to suppression due to nucleotide alterations at degenerate positions over the RNAi target site. An injection of either 6.0 x 108 vp or 1.8 x 1010 vp AAV were subretinally injected into the eyes by Millington-Ward et al. The AAV vectors of Millington-Ward et al. may be applied to the CRISPR Cas system of the present invention, contemplating a dose of about 2 x 1011 to about 6 x 1013 vp administered to a human.

[00701] Dalkara et al. (Sci Transl Med 5, 189ra76 (2013)) also relates to in vivo directed evolution to fashion an AAV vector that delivers wild-type versions of defective genes throughout the retina after noninjurious injection into the eyes' vitreous humor. Dalkara describes a a 7mer peptide display library and an AAV library constructed by DNA shuffling of cap genes from AAV1, 2, 4, 5, 6, 8, and 9. The rcAAV libraries and rAAV vectors expressing GFP under a CAG or Rho promoter were packaged and and deoxyribonuclease-resistant genomic titers were obtained through quantitative PCR. The libraries were pooled, and two rounds of evolution were performed, each consisting of initial library diversification followed by three in vivo selection steps. In each such step, P30 rho-GFP mice were intravitreally injected with 2 ml of iodixanol-purified, phosphate-buffered saline (PBS)–dialyzed library with a genomic titer of about 1×10^{12} vg/ml. The AAV vectors of Dalkara et al. may be applied

to the nucleic acid-targeting system of the present invention, contemplating a dose of about 1 x 10^{15} to about 1 x 10^{16} vg/ml administered to a human.

[00702] In a particular embodiment, the rhodopsin gene may be targeted for the treatment of retinitis pigmentosa (RP), wherein the system of US Patent Publication No. 20120204282 assigned to Sangamo BioSciences, Inc. may be modified in accordance of the CRISPR Cas system of the present invention.

[00703] In another embodiment, the methods of US Patent Publication No. 20130183282 assigned to Cellectis, which is directed to methods of cleaving a target sequence from the human rhodopsin gene, may also be modified to the nucleic acid-targeting system of the present invention.

[00704] US Patent Publication No. 20130202678 assigned to Academia Sinica relates to methods for treating retinopathies and sight-threatening ophthalmologic disorders relating to delivering of the Puf-A gene (which is expressed in retinal ganglion and pigmented cells of eye tissues and displays a unique anti-apoptotic activity) to the sub-retinal or intravitreal space in the eye. In particular, desirable targets are zgc:193933, prdm1a, spata2, tex10, rbb4, ddx3, zp2.2, Blimp-1 and HtrA2, all of which may be targeted by the nucleic acid-targeting system of the present invention.

[00705] Wu (Cell Stem Cell,13:659–62, 2013) designed a guide RNA that led Cas9 to a single base pair mutation that causes cataracts in mice, where it induced DNA cleavage. Then using either the other wild-type allele or oligos given to the zygotes repair mechanisms corrected the sequence of the broken allele and corrected the cataract-causing genetic defect in mutant mouse.

[00706] US Patent Publication No. 20120159653, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with macular degeration (MD). Macular degeneration (MD) is the primary cause of visual impairment in the elderly, but is also a hallmark symptom of childhood diseases such as Stargardt disease, Sorsby fundus, and fatal childhood neurodegenerative diseases, with an age of onset as young as infancy. Macular degeneration results in a loss of vision in the center of the visual field (the macula) because of damage to the retina. Currently existing animal models do not recapitulate major hallmarks of the disease as it is observed in humans. The available animal models comprising mutant genes encoding proteins associated with MD also produce highly variable phenotypes, making translations to human disease and therapy development problematic.

[00707] One aspect of US Patent Publication No. 20120159653 relates to editing of any chromosomal sequences that encode proteins associated with MD which may be applied to the

nucleic acid-targeting system of the present invention. The proteins associated with MD are typically selected based on an experimental association of the protein associated with MD to an MD disorder. For example, the production rate or circulating concentration of a protein associated with MD may be elevated or depressed in a population having an MD disorder relative to a population lacking the MD disorder. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the proteins associated with MD may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[00708] By way of non-limiting example, proteins associated with MD include but are not limited to the following proteins: (ABCA4) ATP-binding cassette, sub-family A (ABC1), member 4 ACHM1 achromatopsia (rod monochromacy) 1 ApoE Apolipoprotein E (ApoE) C1QTNF5 (CTRP5) C1q and tumor necrosis factor related protein 5 (C1QTNF5) C2 Complement component 2 (C2) C3 Complement components (C3) CCL2 Chemokine (C-C motif) Ligand 2 (CCL2) CCR2 Chemokine (C-C motif) receptor 2 (CCR2) CD36 Cluster of Differentiation 36 CFB Complement factor B CFH Complement factor CFH H CFHR1 complement factor H-related 1 CFHR3 complement factor H-related 3 CNGB3 cyclic nucleotide gated channel beta 3 CP ceruloplasmin (CP) CRP C reactive protein (CRP) CST3 cystatin C or cystatin 3 (CST3) CTSD Cathepsin D (CTSD) CX3CR1 chemokine (C-X3-C motif) receptor 1 ELOVL4 Elongation of very long chain fatty acids 4 ERCC6 excision repair crosscomplementing rodent repair deficiency, complementation group 6 FBLN5 Fibulin-5 FBLN5 Fibulin 5 FBLN6 Fibulin 6 FSCN2 fascin (FSCN2) HMCN1 Hemicentrin 1 HMCN1 hemicentin 1 HTRA1 HtrA serine peptidase 1 (HTRA1) HTRA1 HtrA serine peptidase 1 IL-6 Interleukin 6 IL-8 Interleukin 8 LOC387715 Hypothetical protein PLEKHA1 Pleckstrin homology domaincontaining family A member 1 (PLEKHA1) PROM1 Prominin 1(PROM1 or CD133) PRPH2 Peripherin-2 RPGR retinitis pigmentosa GTPase regulator SERPING1 serpin peptidase inhibitor, clade G, member 1 (C1- inhibitor) TCOF1 Treacle TIMP3 Metalloproteinase inhibitor 3 (TIMP3) TLR3 Toll-like receptor 3.

[00709] The identity of the protein associated with MD whose chromosomal sequence is edited can and will vary. In preferred embodiments, the proteins associated with MD whose chromosomal sequence is edited may be the ATP-binding cassette, sub-family A (ABC1) member 4 protein (ABCA4) encoded by the ABCR gene, the apolipoprotein E protein (APOE)

encoded by the APOE gene, the chemokine (C-C motif) Ligand 2 protein (CCL2) encoded by the CCL2 gene, the chemokine (C-C motif) receptor 2 protein (CCR2) encoded by the CCR2 gene, the ceruloplasmin protein (CP) encoded by the CP gene, the cathepsin D protein (CTSD) encoded by the CTSD gene, or the metalloproteinase inhibitor 3 protein (TIMP3) encoded by the TIMP3 gene. In an exemplary embodiment, the genetically modified animal is a rat, and the edited chromosomal sequence encoding the protein associated with MD may be: (ABCA4) ATPbinding cassette, NM_000350 sub-family A (ABC1), member 4 APOE Apolipoprotein E NM_138828 (APOE) CCL2 Chemokine (C-C NM_031530 motif) Ligand 2 (CCL2) CCR2 Chemokine (C-C NM_021866 motif) receptor 2 (CCR2) CP ceruloplasmin (CP) NM_012532 CTSD Cathepsin D (CTSD) NM_134334 TIMP3 Metalloproteinase NM_012886 inhibitor 3 (TIMP3) The animal or cell may comprise 1, 2, 3, 4, 5, 6, 7 or more disrupted chromosomal sequences encoding a protein associated with MD and zero, 1, 2, 3, 4, 5, 6, 7 or more chromosomally integrated sequences encoding the disrupted protein associated with MD.

[00710] The edited or integrated chromosomal sequence may be modified to encode an altered protein associated with MD. Several mutations in MD-related chromosomal sequences have been associated with MD. Non-limiting examples of mutations in chromosomal sequences associated with MD include those that may cause MD including in the ABCR protein, E471K (i.e. glutamate at position 471 is changed to lysine), R1129L (i.e. arginine at position 1129 is changed to leucine), T1428M (i.e. threonine at position 1428 is changed to methionine), R1517S (i.e. arginine at position 1517 is changed to serine), I1562T (i.e. isoleucine at position 1562 is changed to threonine), and G1578R (i.e. glycine at position 1578 is changed to arginine); in the CCR2 protein, V64I (i.e. valine at position 192 is changed to isoleucine); in CP protein, G969B (i.e. glycine at position 969 is changed to asparagine or aspartate); in TIMP3 protein, S156C (i.e. serine at position 156 is changed to cysteine), G166C (i.e. glycine at position 166 is changed to cysteine), G167C (i.e. glycine at position 167 is changed to cysteine), Y168C (i.e. tyrosine at position 168 is changed to cysteine), S170C (i.e. serine at position 170 is changed to cysteine), Y172C (i.e. tyrosine at position 172 is changed to cysteine) and S181C (i.e. serine at position 181 is changed to cysteine). Other associations of genetic variants in MD-associated genes and disease are known in the art.

[00711] CRISPR systems are useful to correct diseases resulting from autosomal dominant genes. For example, CRISPR/Cas9 was used to remove an autosomal dominant gene that causes receptor loss in the eye. Bakondi, B. et al., In Vivo CRISPR/Cas9 Gene Editing Corrects Retinal Dystrophy in the S334ter-3 Rat Model of Autosomal Dominant Retinitis Pigmentosa. Molecular Therapy, 2015; DOI: 10.1038/mt.2015.220.

[00712] Treating Circulatory and Muscular Diseases

[00713] The present invention also contemplates delivering the CRISPR-Cas system described herein to the heart. For the heart, a myocardium tropic adeno-associated virus (AAVM) is preferred, in particular AAVM41 which showed preferential gene transfer in the heart (see, e.g., Lin-Yanga et al., PNAS, March 10, 2009, vol. 106, no. 10). Administration may be systemic or local. A dosage of about 1-10 x 1014 vector genomes are contemplated for systemic administration. See also, e.g., Eulalio et al. (2012) Nature 492: 376 and Somasuntharam et al. (2013) Biomaterials 34: 7790.

[00714] For example, US Patent Publication No. 20110023139, describes use of zinc finger nucleases to genetically modify cells, animals and proteins associated with cardiovascular disease. Cardiovascular diseases generally include high blood pressure, heart attacks, heart failure, and stroke and TIA. Any chromosomal sequence involved in cardiovascular disease or the protein encoded by any chromosomal sequence involved in cardiovascular disease may be utilized in the methods described in this disclosure. The cardiovascular-related proteins are typically selected based on an experimental association of the cardiovascular-related protein to the development of cardiovascular disease. For example, the production rate or circulating concentration of a cardiovascular-related protein may be elevated or depressed in a population having a cardiovascular disorder relative to a population lacking the cardiovascular disorder. Differences in protein levels may be assessed using proteomic techniques including but not limited to Western blot, immunohistochemical staining, enzyme linked immunosorbent assay (ELISA), and mass spectrometry. Alternatively, the cardiovascular-related proteins may be identified by obtaining gene expression profiles of the genes encoding the proteins using genomic techniques including but not limited to DNA microarray analysis, serial analysis of gene expression (SAGE), and quantitative real-time polymerase chain reaction (Q-PCR).

[00715] By way of example, the chromosomal sequence may comprise, but is not limited to, IL1B (interleukin 1, beta), XDH (xanthine dehydrogenase), TP53 (tumor protein p53), PTGIS (prostaglandin 12 (prostacyclin) synthase), MB (myoglobin), IL4 (interleukin 4), ANGPT1 (angiopoietin 1), ABCG8 (ATP-binding cassette, sub-family G (WHITE), member 8), CTSK (cathepsin K), PTGIR (prostaglandin 12 (prostacyclin) receptor (IP)), KCNJ11 (potassium inwardly-rectifying channel, subfamily J, member 11), INS (insulin), CRP (C-reactive protein, pentraxin-related), PDGFRB (platelet-derived growth factor receptor, beta polypeptide), CCNA2 (cyclin A2), PDGFB (platelet-derived growth factor beta polypeptide (simian sarcoma viral (v-sis) oncogene homolog)), KCNJ5 (potassium inwardly-rectifying channel, subfamily J, member 5), KCNN3 (potassium intermediate/small conductance

calcium-activated channel, subfamily N, member 3), CAPN10 (calpain 10), PTGES (prostaglandin E synthase), ADRA2B (adrenergic, alpha-2B-, receptor), ABCG5 (ATP-binding cassette, sub-family G (WHITE), member 5), PRDX2 (peroxiredoxin 2), CAPN5 (calpain 5), PARP14 (poly (ADP-ribose) polymerase family, member 14), MEX3C (mex-3 homolog C (C. elegans)), ACE angiotensin I converting enzyme (peptidyl-dipeptidase A) 1), TNF (tumor necrosis factor (TNF superfamily, member 2)), IL6 (interleukin 6 (interferon, beta 2)), STN (statin), SERPINE1 (serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1), ALB (albumin), ADIPOQ (adiponectin, C1Q and collagen domain containing), APOB (apolipoprotein B (including Ag(x) antigen)), APOE (apolipoprotein E), LEP (leptin), MTHFR (5,10-methylenetetrahydrofolate reductase (NADPH)), APOA1 (apolipoprotein A-I), EDN1 (endothelin 1), NPPB (natriuretic peptide precursor B), NOS3 (nitric oxide synthase 3 (endothelial cell)), PPARG (peroxisome proliferator-activated receptor gamma), PLAT (plasminogen activator, tissue), PTGS2 (prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)), CETP (cholesteryl ester transfer protein, plasma), AGTR1 (angiotensin II receptor, type 1), HMGCR (3-hydroxy-3-methylglutaryl-Coenzyme A reductase), IGF1 (insulin-like growth factor 1 (somatomedin C)), SELE (selectin E), REN (renin), PPARA (peroxisome proliferator-activated receptor alpha), PON1 (paraoxonase 1), KNG1 (kininogen 1), CCL2 (chemokine (C-C motif) ligand 2), LPL (lipoprotein lipase), VWF (von Willebrand factor), F2 (coagulation factor II (thrombin)), ICAM1 (intercellular adhesion molecule 1), TGFB1 (transforming growth factor, beta 1), NPPA (natriuretic peptide precursor A), IL10 (interleukin 10), EPO (erythropoietin), SOD1 (superoxide dismutase 1, soluble), VCAM1 (vascular cell adhesion molecule 1), IFNG (interferon, gamma), LPA (lipoprotein, Lp(a)), MPO (myeloperoxidase), ESR1 (estrogen receptor 1), MAPK1 (mitogen-activated protein kinase 1), HP (haptoglobin), F3 (coagulation factor III (thromboplastin, tissue factor)), CST3 (cystatin C), COG2 (component of oligomeric golgi complex 2), MMP9 (matrix metallopeptidase 9 (gelatinase B, 92 kDa gelatinase, 92 kDa type IV collagenase)), SERPINC1 (serpin peptidase inhibitor, clade C (antithrombin), member 1), F8 (coagulation factor VIII, procoagulant component), HMOX1 (heme oxygenase (decycling) 1), APOC3 (apolipoprotein C-III), IL8 (interleukin 8), PROK1 (prokineticin 1), CBS (cystathionine-beta-synthase), NOS2 (nitric oxide synthase 2, inducible), TLR4 (toll-like receptor 4), SELP (selectin P (granule membrane protein 140 kDa, antigen CD62)), ABCA1 (ATP-binding cassette, sub-family A (ABC1), member 1), AGT (angiotensinogen (serpin peptidase inhibitor, clade A, member 8)), LDLR (low density lipoprotein receptor), GPT (glutamic-pyruvate transaminase (alanine aminotransferase)),

VEGFA (vascular endothelial growth factor A), NR3C2 (nuclear receptor subfamily 3, group C, member 2), IL18 (interleukin 18 (interferon-gamma-inducing factor)), NOS1 (nitric oxide synthase 1 (neuronal)), NR3C1 (nuclear receptor subfamily 3, group C, member 1 (glucocorticoid receptor)), FGB (fibrinogen beta chain), HGF (hepatocyte growth factor (hepapoietin A; scatter factor)), IL1A (interleukin 1, alpha), RETN (resistin), AKT1 (v-akt murine thymoma viral oncogene homolog 1), LIPC (lipase, hepatic), HSPD1 (heat shock 60 kDa protein 1 (chaperonin)), MAPK14 (mitogen-activated protein kinase 14), SPP1 (secreted phosphoprotein 1), ITGB3 (integrin, beta 3 (platelet glycoprotein 111a, antigen CD61)), CAT (catalase), UTS2 (urotensin 2), THBD (thrombomodulin), F10 (coagulation factor X), CP (ceruloplasmin (ferroxidase)), TNFRSF11B (tumor necrosis factor receptor superfamily, member 11b), EDNRA (endothelin receptor type A), EGFR (epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homolog, avian)), MMP2 (matrix metallopeptidase 2 (gelatinase A, 72 kDa gelatinase, 72 kDa type IV collagenase)), PLG (plasminogen), NPY (neuropeptide Y), RHOD (ras homolog gene family, member D), MAPK8 (mitogen-activated protein kinase 8), MYC (v-myc myelocytomatosis viral oncogene homolog (avian)), FN1 (fibronectin 1), CMA1 (chymase 1, mast cell), PLAU (plasminogen activator, urokinase), GNB3 (guanine nucleotide binding protein (G protein), beta polypeptide 3), ADRB2 (adrenergic, beta-2-, receptor, surface), APOA5 (apolipoprotein A-V), SOD2 (superoxide dismutase 2, mitochondrial), F5 (coagulation factor V (proaccelerin, labile factor)), VDR (vitamin D (1,25-dihydroxyvitamin D3) receptor), ALOX5 (arachidonate 5-lipoxygenase), HLA-DRB1 (major histocompatibility complex, class II, DR beta 1), PARP1 (poly (ADP-ribose) polymerase 1), CD40LG (CD40 ligand), PON2 (paraoxonase 2), AGER (advanced glycosylation end product-specific receptor), IRS1 (insulin receptor substrate 1), PTGS1 (prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)), ECE1 (endothelin converting enzyme 1), F7 (coagulation factor VII (serum prothrombin conversion accelerator)), URN (interleukin 1 receptor antagonist), EPHX2 (epoxide hydrolase 2, cytoplasmic), IGFBP1 (insulin-like growth factor binding protein 1), MAPK10 (mitogen-activated protein kinase 10), FAS (Fas (TNF receptor superfamily, member 6)), ABCB1 (ATP-binding cassette, sub-family B (MDR/TAP), member 1), JUN (jun oncogene), IGFBP3 (insulin-like growth factor binding protein 3), CD14 (CD14 molecule), PDE5A (phosphodiesterase 5A, cGMP-specific), AGTR2 (angiotensin II receptor, type 2), CD40 (CD40 molecule, TNF receptor superfamily member 5), LCAT (lecithin-cholesterol acyltransferase), CCR5 (chemokine (C-C motif) receptor 5), MMP1 (matrix metallopeptidase 1 (interstitial collagenase)), TIMP1 (TIMP metallopeptidase inhibitor 1), ADM

(adrenomedullin), DYT10 (dystonia 10), STAT3 (signal transducer and activator of transcription 3 (acute-phase response factor)), MMP3 (matrix metallopeptidase 3 (stromelysin 1, progelatinase)), ELN (elastin), USF1 (upstream transcription factor 1), CFH (complement factor H), HSPA4 (heat shock 70 kDa protein 4), MMP12 (matrix metallopeptidase 12 (macrophage elastase)), MME (membrane metallo-endopeptidase), F2R (coagulation factor II (thrombin) receptor), SELL (selectin L), CTSB (cathepsin B), ANXA5 (annexin A5), ADRB1 (adrenergic, beta-1-, receptor), CYBA (cytochrome b-245, alpha polypeptide), FGA (fibrinogen alpha chain), GGT1 (gamma-glutamyltransferase 1), LIPG (lipase, endothelial), HIF1A (hypoxia inducible factor 1, alpha subunit (basic helix-loop-helix transcription factor)), CXCR4 (chemokine (C-X-C motif) receptor 4), PROC (protein C (inactivator of coagulation factors Va and VIIIa)), SCARB1 (scavenger receptor class B, member 1), CD79A (CD79a molecule, immunoglobulin-associated alpha), PLTP (phospholipid transfer protein), ADD1 (adducin 1 (alpha)), FGG (fibrinogen gamma chain), SAA1 (serum amyloid A1), KCNH2 (potassium voltage-gated channel, subfamily H (eag-related), member 2), DPP4 (dipeptidyl-peptidase 4), G6PD (glucose-6-phosphate dehydrogenase), NPR1 (natriuretic peptide receptor A/guanylate cyclase A (atrionatriuretic peptide receptor A)), VTN (vitronectin), KIAA0101 (KIAA0101), FOS (FBJ murine osteosarcoma viral oncogene homolog), TLR2 (toll-like receptor 2), PPIG (peptidylprolyl isomerase G (cyclophilin G)), IL1R1 (interleukin 1 receptor, type I), AR (androgen receptor), CYP1A1 (cytochrome P450, family 1, subfamily A, polypeptide 1), SERPINA1 (serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1), MTR (5-methyltetrahydrofolate-homocysteine methyltransferase), RBP4 (retinol binding protein 4, plasma), APOA4 (apolipoprotein A-IV), CDKN2A (cyclin-dependent kinase inhibitor 2A (melanoma, p16, inhibits CDK4)), FGF2 (fibroblast growth factor 2 (basic)), EDNRB (endothelin receptor type B), ITGA2 (integrin, alpha 2 (CD49B, alpha 2 subunit of VLA-2 receptor)), CABIN1 (calcineurin binding protein 1), SHBG (sex hormone-binding globulin), HMGB1 (high-mobility group box 1), HSP90B2P (heat shock protein 90 kDa beta (Grp94), member 2 (pseudogene)), CYP3A4 (cytochrome P450, family 3, subfamily A, polypeptide 4), GJA1 (gap junction protein, alpha 1, 43 kDa), CAV1 (caveolin 1, caveolae protein, 22 kDa), ESR2 (estrogen receptor 2 (ER beta)), LTA (lymphotoxin alpha (TNF superfamily, member 1)), GDF15 (growth differentiation factor 15), BDNF (brain-derived neurotrophic factor), CYP2D6 (cytochrome P450, family 2, subfamily D, polypeptide 6), NGF (nerve growth factor (beta polypeptide)), SP1 (Sp1 transcription factor), TGIF1 (TGFB-induced factor homeobox 1), SRC (v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog (avian)), EGF (epidermal growth factor (beta-urogastrone)), PIK3CG

(phosphoinositide-3-kinase, catalytic, gamma polypeptide), HLA-A (major histocompatibility complex, class I, A), KCNQ1 (potassium voltage-gated channel, KQT-like subfamily, member 1), CNR1 (cannabinoid receptor 1 (brain)), FBN1 (fibrillin 1), CHKA (choline kinase alpha), BEST1 (bestrophin 1), APP (amyloid beta (A4) precursor protein), CTNNB1 (catenin (cadherin-associated protein), beta 1, 88 kDa), IL2 (interleukin 2), CD36 (CD36 molecule (thrombospondin receptor)), PRKAB1 (protein kinase, AMP-activated, beta 1 non-catalytic subunit), TPO (thyroid peroxidase), ALDH7A1 (aldehyde dehydrogenase 7 family, member A1), CX3CR1 (chemokine (C-X3-C motif) receptor 1), TH (tyrosine hydroxylase), F9 (coagulation factor IX), GH1 (growth hormone 1), TF (transferrin), HFE (hemochromatosis), IL17A (interleukin 17A), PTEN (phosphatase and tensin homolog), GSTM1 (glutathione S-transferase mu 1), DMD (dystrophin), GATA4 (GATA binding protein 4), F13A1 (coagulation factor XIII, A1 polypeptide), TTR (transthyretin), FABP4 (fatty acid binding protein 4, adipocyte), PON3 (paraoxonase 3), APOC1 (apolipoprotein C-I), INSR (insulin receptor), TNFRSF1B (tumor necrosis factor receptor superfamily, member 1B), HTR2A (5-hydroxytryptamine (serotonin) receptor 2A), CSF3 (colony stimulating factor 3 (granulocyte)), CYP2C9 (cytochrome P450, family 2, subfamily C, polypeptide 9), TXN (thioredoxin), CYP11B2 (cytochrome P450, family 11, subfamily B, polypeptide 2), PTH (parathyroid hormone), CSF2 (colony stimulating factor 2 (granulocyte-macrophage)), KDR (kinase insert domain receptor (a type III receptor tyrosine kinase)), PLA2G2A (phospholipase A2, group IIA (platelets, synovial fluid)), B2M (beta-2-microglobulin), THBS1 (thrombospondin 1), GCG (glucagon), RHOA (ras homolog gene family, member A), ALDH2 (aldehyde dehydrogenase 2 family (mitochondrial)), TCF7L2 (transcription factor 7-like 2 (T-cell specific, HMG-box)), BDKRB2 (bradykinin receptor B2), NFE2L2 (nuclear factor (erythroid-derived 2)-like 2), NOTCH1 (Notch homolog 1, translocation-associated (Drosophila)), UGT1A1 (UDP glucuronosyltransferase 1 family, polypeptide A1), IFNA1 (interferon, alpha 1), PPARD (peroxisome proliferator-activated receptor delta), SIRT1 (sirtuin (silent mating type information regulation 2 homolog) 1 (S. cerevisiae)), GNRH1 (gonadotropin-releasing hormone 1 (luteinizing-releasing hormone)), PAPPA (pregnancy-associated plasma protein A, pappalysin 1), ARR3 (arrestin 3, retinal (X-arrestin)), NPPC (natriuretic peptide precursor C), AHSP (alpha hemoglobin stabilizing protein), PTK2 (PTK2 protein tyrosine kinase 2), IL13 (interleukin 13), MTOR (mechanistic target of rapamycin (serine/threonine kinase)), ITGB2 (integrin, beta 2 (complement component 3 receptor 3 and 4 subunit)), GSTT1 (glutathione S-transferase theta 1), IL6ST (interleukin 6 signal transducer (gp130, oncostatin M receptor)), CPB2 (carboxypeptidase B2 (plasma)), CYP1A2 (cytochrome P450, family 1, subfamily A,

polypeptide 2), HNF4A (hepatocyte nuclear factor 4, alpha), SLC6A4 (solute carrier family 6 (neurotransmitter transporter, serotonin), member 4), PLA2G6 (phospholipase A2, group VI (cytosolic, calcium-independent)), TNFSF11 (tumor necrosis factor (ligand) superfamily, member 11), SLC8A1 (solute carrier family 8 (sodium/calcium exchanger), member 1), F2RL1 (coagulation factor II (thrombin) receptor-like 1), AKR1A1 (aldo-keto reductase family 1, member A1 (aldehyde reductase)), ALDH9A1 (aldehyde dehydrogenase 9 family, member A1), BGLAP (bone gamma-carboxyglutamate (gla) protein), MTTP (microsomal triglyceride transfer protein), MTRR (5-methyltetrahydrofolate-homocysteine methyltransferase reductase), SULT1A3 (sulfotransferase family, cytosolic, 1A, phenol-preferring, member 3), RAGE (renal tumor antigen), C4B (complement component 4B (Chido blood group), P2RY12 (purinergic receptor P2Y, G-protein coupled, 12), RNLS (renalase, FAD-dependent amine oxidase), CREB1 (cAMP responsive element binding protein 1), POMC (proopiomelanocortin), RAC1 (ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1)), LMNA (lamin NC), CD59 (CD59 molecule, complement regulatory protein), SCN5A (sodium channel, voltage-gated, type V, alpha subunit), CYP1B1 (cytochrome P450, family 1, subfamily B, polypeptide 1), MIF (macrophage migration inhibitory factor (glycosylation-inhibiting factor)), MMP13 (matrix metallopeptidase 13 (collagenase 3)), TIMP2 (TIMP metallopeptidase inhibitor 2), CYP19A1 (cytochrome P450, family 19, subfamily A, polypeptide 1), CYP21A2 (cytochrome P450, family 21, subfamily A, polypeptide 2), PTPN22 (protein tyrosine phosphatase, non-receptor type 22 (lymphoid)), MYH14 (myosin, heavy chain 14, non-muscle), MBL2 (mannose-binding lectin (protein C) 2, soluble (opsonic defect)), SELPLG (selectin P ligand), AOC3 (amine oxidase, copper containing 3 (vascular adhesion protein 1)), CTSL1 (cathepsin L1), PCNA (proliferating cell nuclear antigen), IGF2 (insulin-like growth factor 2 (somatomedin A)), ITGB1 (integrin, beta 1 (fibronectin receptor, beta polypeptide, antigen CD29 includes MDF2, MSK12)), CAST (calpastatin), CXCL12 (chemokine (C-X-C motif) ligand 12 (stromal cell-derived factor 1)), IGHE (immunoglobulin heavy constant epsilon), KCNE1 (potassium voltage-gated channel, Isk-related family, member 1), TFRC (transferrin receptor (p90, CD71)), COL1A1 (collagen, type I, alpha 1), COL1A2 (collagen, type I, alpha 2), IL2RB (interleukin 2 receptor, beta), PLA2G10 (phospholipase A2, group X), ANGPT2 (angiopoietin 2), PROCR (protein C receptor, endothelial (EPCR)), NOX4 (NADPH oxidase 4), HAMP (hepcidin antimicrobial peptide), PTPN11 (protein tyrosine phosphatase, non-receptor type 11), SLC2A1 (solute carrier family 2 (facilitated glucose transporter), member 1), IL2RA (interleukin 2 receptor, alpha), CCL5 (chemokine (C-C motif) ligand 5), IRF1 (interferon regulatory factor 1), CFLAR

(CASP8 and FADD-like apoptosis regulator), CALCA (calcitonin-related polypeptide alpha), EIF4E (eukaryotic translation initiation factor 4E), GSTP1 (glutathione S-transferase pi 1), JAK2 (Janus kinase 2), CYP3A5 (cytochrome P450, family 3, subfamily A, polypeptide 5), HSPG2 (heparan sulfate proteoglycan 2), CCL3 (chemokine (C-C motif) ligand 3), MYD88 (myeloid differentiation primary response gene (88)), VIP (vasoactive intestinal peptide), SOAT1 (sterol O-acyltransferase 1), ADRBK1 (adrenergic, beta, receptor kinase 1), NR4A2 (nuclear receptor subfamily 4, group A, member 2), MMP8 (matrix metallopeptidase 8 (neutrophil collagenase)), NPR2 (natriuretic peptide receptor B/guanylate cyclase B (atrionatriuretic peptide receptor B)), GCH1 (GTP cyclohydrolase 1), EPRS (glutamyl-prolyl-tRNA synthetase), PPARGC1A (peroxisome proliferator-activated receptor gamma, coactivator 1 alpha), F12 (coagulation factor XII (Hageman factor)), PECAM1 (platelet/endothelial cell adhesion molecule), CCL4 (chemokine (C-C motif) ligand 4), SERPINA3 (serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 3), CASR (calcium-sensing receptor), GJA5 (gap junction protein, alpha 5, 40 kDa), FABP2 (fatty acid binding protein 2, intestinal), TTF2 (transcription termination factor, RNA polymerase II), PROS1 (protein S (alpha)), CTF1 (cardiotrophin 1), SGCB (sarcoglycan, beta (43 kDa dystrophin-associated glycoprotein)), YME1L1 (YME1-like 1 (S. cerevisiae)), CAMP (cathelicidin antimicrobial peptide), ZC3H12A (zinc finger CCCH-type containing 12A), AKR1B1 (aldo-keto reductase family 1, member B1 (aldose reductase)), DES (desmin), MMP7 (matrix metallopeptidase 7 (matrilysin, uterine)), AHR (aryl hydrocarbon receptor), CSF1 (colony stimulating factor 1 (macrophage)), HDAC9 (histone deacetylase 9), CTGF (connective tissue growth factor), KCNMA1 (potassium large conductance calcium-activated channel, subfamily M, alpha member 1), UGT1A (UDP glucuronosyltransferase 1 family, polypeptide A complex locus), PRKCA (protein kinase C, alpha), COMT (catechol-.beta.-methyltransferase), S100B (S100 calcium binding protein B), EGR1 (early growth response 1), PRL (prolactin), IL15 (interleukin 15), DRD4 (dopamine receptor D4), CAMK2G (calcium/calmodulin-dependent protein kinase II gamma), SLC22A2 (solute carrier family 22 (organic cation transporter), member 2), CCL11 (chemokine (C-C motif) ligand 11), PGF (B321 placental growth factor), THPO (thrombopoietin), GP6 (glycoprotein VI (platelet)), TACR1 (tachykinin receptor 1), NTS (neurotensin), HNF1A (HNF1 homeobox A), SST (somatostatin), KCND1 (potassium voltage-gated channel, Shal-related subfamily, member 1), LOC646627 (phospholipase inhibitor), TBXAS1 (thromboxane A synthase 1 (platelet)), CYP2J2 (cytochrome P450, family 2, subfamily J, polypeptide 2), TBXA2R (thromboxane A2 receptor), ADH1C (alcohol dehydrogenase 1C (class I), gamma polypeptide), ALOX12

(arachidonate 12-lipoxygenase), AHSG (alpha-2-HS-glycoprotein), BHMT (betaine-homocysteine methyltransferase), GJA4 (gap junction protein, alpha 4, 37 kDa), SLC25A4 (solute carrier family 25 (mitochondrial carrier; adenine nucleotide translocator), member 4), ACLY (ATP citrate lyase), ALOX5AP (arachidonate 5-lipoxygenase-activating protein), NUMA1 (nuclear mitotic apparatus protein 1), CYP27B1 (cytochrome P450, family 27, subfamily B, polypeptide 1), CYSLTR2 (cysteinyl leukotriene receptor 2), SOD3 (superoxide dismutase 3, extracellular), LTC4S (leukotriene C4 synthase), UCN (urocortin), GHRL (ghrelin/obestatin prepropeptide), APOC2 (apolipoprotein C-II), CLEC4A (C-type lectin domain family 4, member A), KBTBD10 (kelch repeat and BTB (POZ) domain containing 10), TNC (tenascin C), TYMS (thymidylate synthetase), SHCl (SHC (Src homology 2 domain containing) transforming protein 1), LRP1 (low density lipoprotein receptor-related protein 1), SOCS3 (suppressor of cytokine signaling 3), ADH1B (alcohol dehydrogenase 1B (class I), beta polypeptide), KLK3 (kallikrein-related peptidase 3), HSD11B1 (hydroxysteroid (11-beta) dehydrogenase 1), VKORC1 (vitamin K epoxide reductase complex, subunit 1), SERPINB2 (serpin peptidase inhibitor, clade B (ovalbumin), member 2), TNS1 (tensin 1), RNF19A (ring finger protein 19A), EPOR (erythropoietin receptor), ITGAM (integrin, alpha M (complement component 3 receptor 3 subunit)), PITX2 (paired-like homeodomain 2), MAPK7 (mitogen-activated protein kinase 7), FCGR3A (Fc fragment of IgG, low affinity 111a, receptor (CD16a)), LEPR (leptin receptor), ENG (endoglin), GPX1 (glutathione peroxidase 1), GOT2 (glutamic-oxaloacetic transaminase 2, mitochondrial (aspartate aminotransferase 2)), HRH1 (histamine receptor H1), NR112 (nuclear receptor subfamily 1, group I, member 2), CRH (corticotropin releasing hormone), HTR1A (5-hydroxytryptamine (serotonin) receptor 1A), VDAC1 (voltage-dependent anion channel 1), HPSE (heparanase), SFTPD (surfactant protein D), TAP2 (transporter 2, ATP-binding cassette, sub-family B (MDR/TAP)), RNF123 (ring finger protein 123), PTK2B (PTK2B protein tyrosine kinase 2 beta), NTRK2 (neurotrophic tyrosine kinase, receptor, type 2), IL6R (interleukin 6 receptor), ACHE (acetylcholinesterase (Yt blood group)), GLP1R (glucagon-like peptide 1 receptor), GHR (growth hormone receptor), GSR (glutathione reductase), NQO1 (NAD(P)H dehydrogenase, quinone 1), NR5A1 (nuclear receptor subfamily 5, group A, member 1), GJB2 (gap junction protein, beta 2, 26 kDa), SLC9A1 (solute carrier family 9 (sodium/hydrogen exchanger), member 1), MAOA (monoamine oxidase A), PCSK9 (proprotein convertase subtilisin/kexin type 9), FCGR2A (Fc fragment of IgG, low affinity IIa, receptor (CD32)), SERPINF1 (serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 1), EDN3 (endothelin 3), DHFR (dihydrofolate reductase), GAS6 (growth arrest-specific 6), SMPD1

(sphingomyelin phosphodiesterase 1, acid lysosomal), UCP2 (uncoupling protein 2 (mitochondrial, proton carrier)), TFAP2A (transcription factor AP-2 alpha (activating enhancer binding protein 2 alpha)), C4BPA (complement component 4 binding protein, alpha), SERPINF2 (serpin peptidase inhibitor, clade F (alpha-2 antiplasmin, pigment epithelium derived factor), member 2), TYMP (thymidine phosphorylase), ALPP (alkaline phosphatase, placental (Regan isozyme)), CXCR2 (chemokine (C-X-C motif) receptor 2), SLC39A3 (solute carrier family 39 (zinc transporter), member 3), ABCG2 (ATP-binding cassette, sub-family G (WHITE), member 2), ADA (adenosine deaminase), JAK3 (Janus kinase 3), HSPA1A (heat shock 70 kDa protein 1A), FASN (fatty acid synthase), FGF1 (fibroblast growth factor 1 (acidic)), F11 (coagulation factor XI), ATP7A (ATPase, Cu++ transporting, alpha polypeptide), CR1 (complement component (3b/4b) receptor 1 (Knops blood group)), GFAP (glial fibrillary acidic protein), ROCK1 (Rho-associated, coiled-coil containing protein kinase 1), MECP2 (methyl CpG binding protein 2 (Rett syndrome)), MYLK (myosin light chain kinase), BCHE (butyrylcholinesterase), LIPE (lipase, hormone-sensitive), PRDX5 (peroxiredoxin 5), ADORA1 (adenosine A1 receptor), WRN (Werner syndrome, RecQ helicase-like), CXCR3 (chemokine (C-X-C motif) receptor 3), CD81 (CD81 molecule), SMAD7 (SMAD family member 7), LAMC2 (laminin, gamma 2), MAP3K5 (mitogen-activated protein kinase kinase kinase 5), CHGA (chromogranin A (parathyroid secretory protein 1)), IAPP (islet amyloid polypeptide), RHO (rhodopsin), ENPP1 (ectonucleotide pyrophosphatase/phosphodiesterase 1), PTHLH (parathyroid hormone-like hormone), NRG1 (neuregulin 1), VEGFC (vascular endothelial growth factor C), ENPEP (glutamyl aminopeptidase (aminopeptidase A)), CEBPB (CCAAT/enhancer binding protein (C/EBP), beta), NAGLU (N-acetylglucosaminidase, alpha-), F2RL3 (coagulation factor II (thrombin) receptor-like 3), CX3CL1 (chemokine (C-X3-C motif) ligand 1), BDKRB1 (bradykinin receptor B1), ADAMTS13 (ADAM metallopeptidase with thrombospondin type 1 motif, 13), ELANE (elastase, neutrophil expressed), ENPP2 (ectonucleotide pyrophosphatase/phosphodiesterase 2), CISH (cytokine inducible SH2-containing protein), GAST (gastrin), MYOC (myocilin, trabecular meshwork inducible glucocorticoid response), ATP1A2 (ATPase, Na+/K+ transporting, alpha 2 polypeptide), NF1 (neurofibromin 1), GJB1 (gap junction protein, beta 1, 32 kDa), MEF2A (myocyte enhancer factor 2A), VCL (vinculin), BMPR2 (bone morphogenetic protein receptor, type II (serine/threonine kinase)), TUBB (tubulin, beta), CDC42 (cell division cycle 42 (GTP binding protein, 25 kDa)), KRT18 (keratin 18), HSF1 (heat shock transcription factor 1), MYB (v-myb myeloblastosis viral oncogene homolog (avian)), PRKAA2 (protein kinase, AMP-activated, alpha 2 catalytic subunit),

ROCK2 (Rho-associated, coiled-coil containing protein kinase 2), TFPI (tissue factor pathway inhibitor (lipoprotein-associated coagulation inhibitor)), PRKG1 (protein kinase, cGMP-dependent, type I), BMP2 (bone morphogenetic protein 2), CTNND1 (catenin (cadherin-associated protein), delta 1), CTH (cystathionase (cystathionine gamma-lyase)), CTSS (cathepsin S), VAV2 (vav 2 guanine nucleotide exchange factor), NPY2R (neuropeptide Y receptor Y2), IGFBP2 (insulin-like growth factor binding protein 2, 36 kDa), CD28 (CD28 molecule), GSTA1 (glutathione S-transferase alpha 1), PPIA (peptidylprolyl isomerase A (cyclophilin A)), APOH (apolipoprotein H (beta-2-glycoprotein I)), S100A8 (S100 calcium binding protein A8), IL11 (interleukin 11), ALOX15 (arachidonate 15-lipoxygenase), FBLN1 (fibulin 1), NR1H3 (nuclear receptor subfamily 1, group H, member 3), SCD (stearoyl-CoA desaturase (delta-9-desaturase)), GIP (gastric inhibitory polypeptide), CHGB (chromogranin B (secretogranin 1)), PRKCB (protein kinase C, beta), SRD5A1 (steroid-5-alpha-reductase, alpha polypeptide 1 (3-oxo-5 alpha-steroid delta 4-dehydrogenase alpha 1)), HSD11B2 (hydroxysteroid (11-beta) dehydrogenase 2), CALCRL (calcitonin receptor-like), GALNT2 (UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 2 (GalNAc-T2)), ANGPTL4 (angiopoietin-like 4), KCNN4 (potassium intermediate/small conductance calcium-activated channel, subfamily N, member 4), PIK3C2A (phosphoinositide-3-kinase, class 2, alpha polypeptide), HBEGF (heparin-binding EGF-like growth factor), CYP7A1 (cytochrome P450, family 7, subfamily A, polypeptide 1), HLA-DRB5 (major histocompatibility complex, class II, DR beta 5), BNIP3 (BCL2/adenovirus E1B 19 kDa interacting protein 3), GCKR (glucokinase (hexokinase 4) regulator), S100A12 (S100 calcium binding protein A12), PADI4 (peptidyl arginine deiminase, type IV), HSPA14 (heat shock 70 kDa protein 14), CXCR1 (chemokine (C-X-C motif) receptor 1), H19 (H19, imprinted maternally expressed transcript (non-protein coding)), KRTAP19-3 (keratin associated protein 19-3), IDDM2 (insulin-dependent diabetes mellitus 2), RAC2 (ras-related C3 botulinum toxin substrate 2 (rho family, small GTP binding protein Rac2)), RYR1 (ryanodine receptor 1 (skeletal)), CLOCK (clock homolog (mouse)), NGFR (nerve growth factor receptor (TNFR superfamily, member 16)), DBH (dopamine beta-hydroxylase (dopamine beta-monooxygenase)), CHRNA4 (cholinergic receptor, nicotinic, alpha 4), CACNA1C (calcium channel, voltage-dependent, L type, alpha 1C subunit), PRKAG2 (protein kinase, AMP-activated, gamma 2 non-catalytic subunit), CHAT (choline acetyltransferase), PTGDS (prostaglandin D2 synthase 21 kDa (brain)), NR1H2 (nuclear receptor subfamily 1, group H, member 2), TEK (TEK tyrosine kinase, endothelial), VEGFB (vascular endothelial growth factor B), MEF2C (myocyte enhancer factor 2C), MAPKAPK2 (mitogen-activated

protein kinase-activated protein kinase 2), TNFRSF11A (tumor necrosis factor receptor superfamily, member 11a, NFKB activator), HSPA9 (heat shock 70 kDa protein 9 (mortalin)), CYSLTR1 (cysteinyl leukotriene receptor 1), MAT1A (methionine adenosyltransferase I, alpha), OPRL1 (opiate receptor-like 1), IMPA1 (inositol(myo)-1(or 4)-monophosphatase 1), CLCN2 (chloride channel 2), DLD (dihydrolipoamide dehydrogenase), PSMA6 (proteasome (prosome, macropain) subunit, alpha type, 6), PSMB8 (proteasome (prosome, macropain) subunit, beta type, 8 (large multifunctional peptidase 7)), CHI3L1 (chitinase 3-like 1 (cartilage glycoprotein-39)), ALDH1B1 (aldehyde dehydrogenase 1 family, member B1), PARP2 (poly (ADP-ribose) polymerase 2), STAR (steroidogenic acute regulatory protein), LBP (lipopolysaccharide binding protein), ABCC6 (ATP-binding cassette, sub-family C(CFTR/MRP), member 6), RGS2 (regulator of G-protein signaling 2, 24 kDa), EFNB2 (ephrin-B2), GJB6 (gap junction protein, beta 6, 30 kDa), APOA2 (apolipoprotein A-II), AMPD1 (adenosine monophosphate deaminase 1), DYSF (dysferlin, limb girdle muscular dystrophy 2B (autosomal recessive)), FDFT1 (farnesyl-diphosphate farnesyltransferase 1), EDN2 (endothelin 2), CCR6 (chemokine (C-C motif) receptor 6), GJB3 (gap junction protein, beta 3, 31 kDa), IL1RL1 (interleukin 1 receptor-like 1), ENTPD1 (ectonucleoside triphosphate diphosphohydrolase 1), BBS4 (Bardet-Biedl syndrome 4), CELSR2 (cadherin, EGF LAG seven-pass G-type receptor 2 (flamingo homolog, Drosophila)), F11R (F11 receptor), RAPGEF3 (Rap guanine nucleotide exchange factor (GEF) 3), HYAL1 (hyaluronoglucosaminidase 1), ZNF259 (zinc finger protein 259), ATOX1 (ATX1 antioxidant protein 1 homolog (yeast)), ATF6 (activating transcription factor 6), KHK (ketohexokinase (fructokinase)), SAT1 (spermidine/spermine N1-acetyltransferase 1), GGH (gamma-glutamyl hydrolase (conjugase, folylpolygammaglutamyl hydrolase)), TIMP4 (TIMP metallopeptidase inhibitor 4), SLC4A4 (solute carrier family 4, sodium bicarbonate cotransporter, member 4), PDE2A (phosphodiesterase 2A, cGMP-stimulated), PDE3B (phosphodiesterase 3B, cGMP-inhibited), FADS1 (fatty acid desaturase 1), FADS2 (fatty acid desaturase 2), TMSB4X (thymosin beta 4, X-linked), TXNIP (thioredoxin interacting protein), LIMS1 (LIM and senescent cell antigen-like domains 1), RHOB (ras homolog gene family, member B), LY96 (lymphocyte antigen 96), FOXO1 (forkhead box O1), PNPLA2 (patatin-like phospholipase domain containing 2), TRH (thyrotropin-releasing hormone), GJC1 (gap junction protein, gamma 1, 45 kDa), SLC17A5 (solute carrier family 17 (anion/sugar transporter), member 5), FTO (fat mass and obesity associated), GJD2 (gap junction protein, delta 2, 36 kDa), PSRC1 (proline/serine-rich coiled-coil 1), CASP12 (caspase 12 (gene/pseudogene)), GPBAR1 (G protein-coupled bile acid receptor 1), PXK (PX domain containing serine/threonine kinase),

IL33 (interleukin 33), TRIB1 (tribbles homolog 1 (Drosophila)), PBX4 (pre-B-cell leukemia homeobox 4), NUPR1 (nuclear protein, transcriptional regulator, 1), 15-Sep(15 kDa selenoprotein), CILP2 (cartilage intermediate layer protein 2), TERC (telomerase RNA component), GGT2 (gamma-glutamyltransferase 2), MT-CO1 (mitochondrially encoded cytochrome c oxidase I), and UOX (urate oxidase, pseudogene). Any of these sequences, may be a target for the CRISPR-Cas system, e.g., to address mutation.

[00716] In an additional embodiment, the chromosomal sequence may further be selected from Pon1 (paraoxonase 1), LDLR (LDL receptor), ApoE (Apolipoprotein E), Apo B-100 (Apolipoprotein B-100), ApoA (Apolipoprotein(a)), ApoA1 (Apolipoprotein A1), CBS (Cystathione B-synthase), Glycoprotein IIb/IIb, MTHRF (5,10-methylenetetrahydrofolate reductase (NADPH), and combinations thereof. In one iteration, the chromosomal sequences and proteins encoded by chromosomal sequences involved in cardiovascular disease may be chosen from Cacna1C, Sod1, Pten, Ppar(alpha), Apo E, Leptin, and combinations thereof as target(s) for the CRISPR-Cas system.

Treating Diseases of the Liver and Kidney

[00717] The present invention also contemplates delivering the CRISPR-Cas system described herein effector protein systems, to the liver and/or kidney. Delivery strategies to induce cellular uptake of the therapeutic nucleic acid include physical force or vector systems such as viral-, lipid- or complex- based delivery, or nanocarriers. From the initial applications with less possible clinical relevance, when nucleic acids were addressed to renal cells with hydrodynamic high pressure injection systemically, a wide range of gene therapeutic viral and non-viral carriers have been applied already to target posttranscriptional events in different animal kidney disease models in vivo (Csaba Révész and Péter Hamar (2011). Delivery Methods to Target RNAs in the Kidney, Gene Therapy Applications, Prof. Chunsheng Kang (Ed.), ISBN: 978-953-307-541-9, InTech, Available from: www.intechopen.com/books/gene-therapy-applications/delivery-methods-to-target-rnas-inthe-kidney). Delivery methods to the kidney may include those in Yuan et al. (Am J Physiol Renal Physiol 295: F605–F617, 2008) investigated whether in vivo delivery of small interfering RNAs (siRNAs) targeting the 12/15-lipoxygenase (12/15-LO) pathway of arachidonate acid metabolism can ameliorate renal injury and diabetic nephropathy (DN) in a streptozotocininjected mouse model of type 1 diabetes. To achieve greater in vivo access and siRNA expression in the kidney, Yuan et al. used double-stranded 12/15-LO siRNA oligonucleotides conjugated with cholesterol. About 400 μg of siRNA was injected subcutaneously into mice. The method of Yuang et al. may be applied to

the CRISPR Cas system of the present invention contemplating a 1-2 g subcutaneous injection of CRISPR Cas conjugated with cholesterol to a human for delivery to the kidneys.

[00718] Molitoris et al. (J Am Soc Nephrol 20: 1754–1764, 2009) exploited proximal tubule cells (PTCs), as the site of oligonucleotide reabsorption within the kidney to test the efficacy of siRNA targeted to p53, a pivotal protein in the apoptotic pathway, to prevent kidney injury. Naked synthetic siRNA to p53 injected intravenously 4 h after ischemic injury maximally protected both PTCs and kidney function. Molitoris et al.'s data indicates that rapid delivery of siRNA to proximal tubule cells follows intravenous administration. For dose-response analysis, rats were injected with doses of siP53, 0.33; 1, 3, or 5mg/kg, given at the same four time points, resulting in cumulative doses of 1.32; 4, 12, and 20 mg/kg, respectively. All siRNA doses tested produced a SCr reducing effect on day one with higher doses being effective over approximately five days compared with PBS-treated ischemic control rats. The 12 and 20 mg/kg cumulative doses provided the best protective effect. The method of Molitoris et al. may be applied to the nucleic acid-targeting system of the present invention contemplating 12 and 20 mg/kg cumulative doses to a human for delivery to the kidneys.

[00719] Thompson et al. (Nucleic Acid Therapeutics, Volume 22, Number 4, 2012) reports the toxicological and pharmacokinetic properties of the synthetic, small interfering RNA I5NP following intravenous administration in rodents and nonhuman primates. I5NP is designed to act via the RNA interference (RNAi) pathway to temporarily inhibit expression of the pro-apoptotic protein p53 and is being developed to protect cells from acute ischemia/reperfusion injuries such as acute kidney injury that can occur during major cardiac surgery and delayed graft function that can occur following renal transplantation. Doses of 800mg/kg I5NP in rodents, and 1,000 mg/kg I5NP in nonhuman primates, were required to elicit adverse effects, which in the monkey were isolated to direct effects on the blood that included a sub-clinical activation of complement and slightly increased clotting times. In the rat, no additional adverse effects were observed with a rat analogue of I5NP, indicating that the effects likely represent class effects of synthetic RNA duplexes rather than toxicity related to the intended pharmacologic activity of I5NP. Taken together, these data support clinical testing of intravenous administration of I5NP for the preservation of renal function following acute ischemia/reperfusion injury. The no observed adverse effect level (NOAEL) in the monkey was 500 mg/kg. No effects on cardiovascular, respiratory, and neurologic parameters were observed in monkeys following i.v. administration at dose levels up to 25 mg/kg. Therefore, a similar dosage may be contemplated for intravenous administration of CRISPR Cas to the kidneys of a human.

[00720] Shimizu et al. (J Am Soc Nephrol 21: 622–633, 2010) developed a system to target delivery of siRNAs to glomeruli via poly(ethylene glycol)-poly(L-lysine)-based vehicles. The siRNA/nanocarrier complex was approximately 10 to 20 nm in diameter, a size that would allow it to move across the fenestrated endothelium to access to the mesangium. After intraperitoneal injection of fluorescence-labeled siRNA/nanocarrier complexes, Shimizu et al. detected siRNAs in the blood circulation for a prolonged time. Repeated intraperitoneal administration of a mitogen-activated protein kinase 1 (MAPK1) siRNA/nanocarrier complex suppressed glomerular MAPK1 mRNA and protein expression in a mouse model of glomerulonephritis. For the investigation of siRNA accumulation, Cy5-labeled siRNAs complexed with PIC nanocarriers (0.5 ml, 5 nmol of siRNA content), naked Cy5-labeled siRNAs (0.5 ml, 5 nmol), or Cy5-labeled siRNAs encapsulated in HVJ-E (0.5 ml, 5 nmol of siRNA content) were administrated to BALBc mice. The method of Shimizu et al. may be applied to the nucleic acid-targeting system of the present invention contemplating a dose of about of 10-20 μmol CRISPR Cas complexed with nanocarriers in about 1-2 liters to a human for intraperitoneal administration and delivery to the kidneys.

[00721] Table 11. Delivery methods to the kidney are summarized as follows:

Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
Hydrodynamic / Lipid	TransIT In Vivo Gene Delivery System, DOTAP	p85α	Acute renal injury	Ischemia-reperfusion	Uptake, biodistribution	Larson et al., Surgery, (Aug 2007), Vol. 142, No. 2, pp. (262-269)
Hydrodynamic / Lipid	Lipofectamine 2000	Fas	Acute renal injury	Ischemia-reperfusion	Blood urea nitrogen, Fas Immunohistochemistry, apoptosis, histological scoring	Hamar et al., Proc Natl Acad Sci, (Oct 2004), Vol. 101, No. 41, pp. (14883-14888)
Hydrodynamic	n.a.	Apoptosis cascade elements	Acute renal injury	Ischemia-reperfusion	n.a.	Zheng et al., Am J Pathol, (Oct 2008), Vol. 173, No. 4, pp. (973–980)

Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
Hydrodynamic	n.a.	Nuclear factor kappa-b (NFkB)	Acute renal injury	Ischemia-reperfusion	n.a.	Feng et al., Transplantation, (May 2009), Vol. 87, No. 9, pp. (1283–1289)
Hydrodynamic /Viral	Lipofectamine 2000	Apoptosis antagonizing transcription factor (AATF)	Acute renal injury	Ischemia-reperfusion	Apoptosis, oxidative stress, caspase activation, membrane lipid peroxidation	Xie & Guo, Am Soc Nephrol, (Dec 2006), Vol. 17, No. 12, pp. (3336–3346)
Hydrodynamic	pBAsi mU6 Neo/ TransIT-EE Hydrodynamic Delivery System	Gremlin	Diabetic nephropathy	Streptozotozin - induced diabetes	Proteinuria, serum creatinine, glomerular and tubular diameter, collagen type IV/BMP7 expression	Q. Zhang et al., PloS ONE, (Jul 2010), Vol. 5, No. 7, e11709, pp. (1-13)
Viral/Lipid	pSUPER vector/Lipofectamine	TGF-β type II receptor	Interstitial renal fibrosis	Unilateral urethral obstruction	α-SMA expression, collagen content,	Kushibikia et al., J Controlled Release, (Jul 2005), Vol. 105, No. 3, pp. (318-331)
Viral	Adeno-associated virus-2	Mineral corticoid receptor	Hyper-tension caused renal damage	Cold-induced hypertension	blood pressure, serum albumin, serum urea nitrogen, serum creatinine, kidney weight, urinary sodium	Wang et al., Gene Therapy, (Jul 2006), Vol. 13, No. 14, pp. (1097-1103)
Hydrodynamic /Viral	pU6 vector	Luciferase	n.a.	n.a.	uptake	Kobayashi et al., Journal of Pharmacology and Experimental

Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
						Therapeutics, (Feb 2004), Vol. 308, No. 2, pp. (688-693)
Lipid	Lipoproteins, albumin	apoB1, apoM	n.a.	n.a.	Uptake, binding affinity to lipoproteins and albumin	Wolfrum et al., Nature Biotechnology, (Sep 2007), Vol. 25, No. 10, pp. (1149-1157)
Lipid	Lipofectamine 2000	p53	Acute renal injury	Ischemic and cisplatin-induced acute injury	Histological scoring, apoptosis	Molitoris et al., J Am Soc Nephrol, (Aug 2009), Vol. 20, No. 8, pp. (1754-1764)
Lipid	DOTAP/DOPE, DOTAP/DOPE/DOPE-PEG2000	COX-2	Breast adeno-carcinoma	MDA-MB-231 breast cancer xenograft-bearing mouse	Cell viability, uptake	Mikhaylova et al., Cancer Gene Therapy, (Mar 2011), Vol. 16, No. 3, pp. (217-226)
Lipid	Cholesterol	12/15-lipoxygenase	Diabetic nephro-pathy	Streptozotocin - induced diabetes	Albuminuria, urinary creatinine, histology, type I and IV collagen, TGF-β, fibronectin, plasminogen activator inhibitor 1	Yuan et al., Am J Physiol Renal Physiol, (Jun 2008), Vol. 295, pp. (F605-F617)
Lipid	Lipofectamine 2000	Mitochondrial membrane 44 (TIM44)	Diabetic nephro-pathy	Streptozotocin - induced diabetes	Cell proliferation and apoptosis, histology, ROS, mitochondrial import of Mn-SOD and glutathione peroxidase,	Y. Zhang et al., J Am Soc Nephrol, (Apr 2006), Vol. 17, No. 4, pp. (1090–1101)

Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
					cellular membrane polarization	
Hydrodynamic / Lipid	Proteolipo-some	RLIP76	Renal carcinoma	Caki-2 kidney cancer xenograft-bearing mouse	uptake	Singhal et al., Cancer Res, (May 2009), Vol. 69, No. 10, pp. (4244-4251)
Polymer	PEGylated PEI	Luciferase pGL3	n.a.	n.a.	Uptake, biodistribution, erythrocyte aggregation	Malek et al., Toxicology and Applied Pharmacology, (Apr 2009), Vol. 236, No. 1, pp. (97-108)
Polymer	PEGylated poly-L-lysine	MAPK1	Lupus glomerulo- nephritis	Glomerulo-nephritis	Proteinuria, glomerulosclerosis, TGF- β, fibronectin, plasminogen activator inhibitor 1	Shimizu et al., J Am Soc Nephrology, (Apr 2010), Vol. 21, No. 4, pp. (622-633)
Polymer/Nano particle	Hyaluronic acid/ Quantum dot/ PEI	VEGF	Kidney cancer/ melanoma	B16F1 melanoma tumor-bearing mouse	Biodistribution, citotoxicity, tumor volume, endocytosis	Jiang et al., Molecular Pharmaceutics, (May-Jun 2009), Vol. 6, No. 3, pp. (727-737)
Polymer/Nano particle	PEGylated polycapro-lactone nanofiber	GAPDH	n.a.	n.a.	cell viability, uptake	Cao et al, J Controlled Release, (Jun 2010), Vol. 144, No. 2, pp. (203-212)
Aptamer	Spiegelmer mNOX-E36	CC chemokine ligand 2	Glomerulo sclerosis	Uninephrecto- mized mouse	urinary albumin, urinary creatinine, histopathology, glomerular filtration rate, macrophage count, serum Ccl2, Mac-2+, Ki-67+	Ninichuk et al., Am J Pathol, (Mar 2008), Vol. 172, No. 3, pp. (628-637)

Delivery method	Carrier	Target RNA	Disease	Model	Functional assays	Author
Aptamer	Aptamer NOX-F37	vasopressin (AVP)	Congestive heart failure	n.a.	Binding affinity to D-AVP, Inhibition of AVP Signaling, Urine osmolality and sodium concentration,	Purschke et al., Proc Natl Acad Sci, (Mar 2006), Vol. 103, No. 13, pp. (5173-5178)

Targeting the Liver or Liver Cells

[00722] Targeting liver cells is provided. This may be in vitro or in vivo. Hepatocytes are preferred. Delivery of the Cas protein may be via viral vectors, especially AAV (and in particular AAV2/6) vectors. These may be administered by intravenous injection.

[00723] A preferred target for liver, whether in vitro or in vivo, is the albumin gene. This is a so-called 'safe harbor" as albumin is expressed at very high levels and so some reduction in the production of albumin following successful gene editing is tolerated. It is also preferred as the high levels of expression seen from the albumin promoter/enhancer allows for useful levels of correct or transgene production (from the inserted donor template) to be achieved even if only a small fraction of hepatocytes are edited.

[00724] Intron 1 of albumin has been shown by Wechsler et al. (reported at the 57th Annual Meeting and Exposition of the American Society of Hematology - abstract available online at ash.confex.com/ash/2015/webprogram/Paper86495.html and presented on 6th December 2015) to be a suitable target site. Their work used Zn Fingers to cut the DNA at this target site, and suitable guide sequences can be generated to guide cleavage at the same site by a CRISPR protein.

[00725] The use of targets within highly–expressed genes (genes with highly active enhancers/promoters) such as albumin may also allow a promoterless donor template to be used, as reported by Wechsler et al. and this is also broadly applicable outside liver targeting. Other examples of highly-expressed genes are known.

Other disease of the liver

[00726] In particular embodiments, the CRISPR proteins of the present invention are used in the treatment of liver disorders such as transthyretin amyloidosis (ATTR), alpha-1

antitrypsin deficiency and other hepatic-based inborn errors of metabolism. FAP is caused by a mutation in the gene that encodes transthyretin (TTR). While it ia an autosomal dominant disease, not al carriers develop the disease. There are over 100 mutations in the TTR gene known to be associated with the disease. Examples of common mutations include V30M. The principle of treatment of TTR based on gene silencing has been demonstrated by studies with iRNA (Ueda et al. 2014 Transl Neurogener. 3:19). Wilson's Disease (WD) is caused by mutations in the gene encoding ATP7B, which is found exclusively in the hepatocyte. There are over 500 mutations associated with WD, with increased prevalence in specific regions such as East Asia. Other examples are A1ATD (an autosomal recessive disease caused by mutations in the SERPINA1 gene) and PKU (an autosomal recessive disease caused by mutations in the phenylalanine hydroxylase (PAH) gene).

[00727] Liver –Associated Blood Disorders, especially Hemophilia and in particular Hemophilia B

[00728] Successful gene editing of hepatocytes has been achieved in mice (both in vitro and in vivo) and in non-human primates (in vivo), showing that treatment of blood disorders through gene editing/genome engineering in hepatocytes is feasible. In particular, expression of the human F9 (hF9) gene in hepatocytes has been shown in non-human primates indicating a treatment for Hemophillia B in humans.

[00729] Wechsler et al. reported at the 57th Annual Meeting and Exposition of the American Society of Hematology (abstract presented 6th December 2015 and available online at ash.confex.com/ash/2015/webprogram/Paper86495.html) that they has successfully expressed human F9 (hF9) from hepatocytes in non-human primates through in vivo gene editing. This was achieved using 1) two zinc finger nucleases (ZFNs) targeting intron 1 of the albumin locus, and 2) a human F9 donor template construct. The ZFNs and donor template were encoded on separate hepatotropic adeno-associated virus serotype 2/6 (AAV2/6) vectors injected intravenously, resulting in targeted insertion of a corrected copy of the hF9 gene into the albumin locus in a proportion of liver hepatocytes.

[00730] The albumin locus was selected as a "safe harbor" as production of this most abundant plasma protein exceeds 10 g/day, and moderate reductions in those levels are well-tolerated. Genome edited hepatocytes produced normal hFIX (hF9) in therapeutic quantities, rather than albumin, driven by the highly active albumin enhancer/promoter. Targeted integration of the hF9 transgene at the albumin locus and splicing of this gene into the albumin transcript was shown.

[00731] Mice studies: C57BL/6 mice were administered vehicle (n=20) or AAV2/6 vectors (n=25) encoding mouse surrogate reagents at 1.0 x1013 vector genome (vg)/kg via tail vein injection. ELISA analysis of plasma hFIX in the treated mice showed peak levels of 50-1053 ng/mL that were sustained for the duration of the 6-month study. Analysis of FIX activity from mouse plasma confirmed bioactivity commensurate with expression levels.

[00732] Non-human primate (NHP) studies: a single intravenous co-infusion of AAV2/6 vectors encoding the NHP targeted albumin-specific ZFNs and a human F9 donor at 1.2x1013 vg/kg (n=5/group) resulted in >50 ng/mL (>1% of normal) in this large animal model. The use of higher AAV2/6 doses (up to 1.5x1014 vg/kg) yielded plasma hFIX levels up to 1000 ng/ml (or 20% of normal) in several animals and up to 2000 ng/ml (or 50% of normal) in a single animal, for the duration of the study (3 months).

[00733] The treatment was well tolerated in mice and NHPs, with no significant toxicological findings related to AAV2/6 ZFN + donor treatment in either species at therapeutic doses. Sangamo (CA, USA) has since applied to the FDA, and been granted, permission to conduct the world's first human clinical trial for an in vivo genome editing application. This follows on the back of the EMEA's approval of the Glybera gene therapy treatment of lipoprotein lipase deficiency.

[00734] Accordingly, it is preferred, in some embodiments, that any or all of the following are used:

[00735] AAV (especially AAV2/6) vectors, preferably administered by intravenous injection;

[00736] Albumin as target for gene editing/insertion of transgene/template– especially at intron 1 of albumin;

[00737] human F9 donor template; and/or

[00738] a promoterless donor template.

Hemophilia B

[00739] Accordingly, in some embodiments, it is preferred that the present invention is used to treat Hemophilia B. As such it is preferred that F9 (Factor IX) is targeted through provision of a suitable guide RNA. The enzyme and the guide may ideally be targeted to the liver where F9 is produced, although they can be delivered together or separately. A template is provided, in some embodiments, and that this is the human F9 gene. It will be appreciated that the hF9 template comprises the wt or 'correct' version of hF9 so that the treatment is effective. In some embodiments, a two-vector system may be used– one vector for the Cas and one vector for the repair template(s). The repair template may include two or more repair templates, for example,

two F9 sequences from different mammalian species. In some embodiments, both a mouse and human F9 sequence are provided. This is may be delivered to mice. Yang Yang, John White, McMenamin Deirdre, and Peter Bell, PhD, presenting at 58th Annual American Society of Hematology Meeting (Nov 2016), report that this increases potency and accuracy. The second vector inserted the human sequence of factor IX into the mouse genome. In some embodiments, the targeted insertion leads to the expression of a chimeric hyperactive factor IX protein. In some embodiments, this is under the control of the native mouse factor IX promoter. Injecting this two-component system (vector 1 and vector 2) into newborn and adult "knock-out" mice at increasing doses led to expression and activity of stable factor IX activity at normal (or even higher) levels for over four months. In the case of treating humans, a native human F9 promoter may be used instead. In some embodiments, the wt phenotype is restored.

[00740] In an alternative embodiment, the hemophilia B version of F9 may be delivered so as to create a model organism, cell or cell line (for example a murine or non-human primate model organism, cell or cell line), the model organism, cell or cell line having or carrying the Hemophilia B phenotype, i.e. an inability to produce wt F9.

Hemophilia A

[00741] In some embodiments, the F9 (factor IX) gene may be replaced by the F8 (factor VIII) gene described above, leading to treatment of Hemophilia A (through provision of a correct F8 gene) and/or creation of a Hemophilia A model organism, cell or cell line (through provision of an incorrect, Hemophilia A version of the F8 gene).

Hemophilia C

[00742] In some embodiments, the F9 (factor IX) gene may be replaced by the F11 (factor XI) gene described above, leading to treatment of Hemophilia C (through provision of a correct F11 gene) and/or creation of a Hemophilia C model organism, cell or cell line (through provision of an incorrect, Hemophilia C version of the F11 gene).

Transthyretin Amyloidosis

[00743] Transthyretin is a protein, mainly produced in the liver, present in the serum and CSF which carries thyroxin hormone and retinol binding protein bound to retinol (Vitamin A). Over 120 different mutations can cause Transthyretin amyloidosis (ATTR), a heritable genetic disorder wherein mutant forms of the protein aggregate in tissues, particularly the peripheral nervous system, causing polyneuropathy. Familial amyloid polyneuropathy (FAP) is the most common TTR disorder and, in 2014, was thought to affect 47 per 100,000 people in Europe. A mutation in the TTR gene of Val30Met is thought be the most common mutation, causing an

estimated 50% of FAP cases. In the absence a liver transplant, the only known cure to date, the disease is usually fatal within a decade of diagnosis. The majority of cases are monogenic.

[00744] In mouse models of ATTR, the TTR gene may be edited in a dose dependent manner by the delivery of CRISPR/Cas. In some embodiments, the Cas is provided as mRNA. In some embodiments, Cas mRNA and guide RNA are packaged in LNPs. A system comprising Cas mRNA and guide RNA packaged in LNPs achieved up to 60% editing efficiency in the liver, with serum TTR levels being reduced by up to 80%. In some embodiments, therefore, Transthyretin is targeted, in particular correcting for the Val30Met mutation. In some embodiments, therefore, ATTR is treated.

Alpha-1 Antitrypsin Deficiency

[00745] Alpha-1 Antitrypsin (A1AT) is a protein produced in the liver which primarily functions to decrease the activity of neutrophil elastase, an enzyme which degrades connective tissue, in the lungs. Alpha-1 Antitrypsin Deficiency (ATTD) is a disease caused by mutation of the SERPINA1 gene, which encodes A1AT. Impaired production of A1AT leads to a gradual degredation of the connective tissue of the lung resulting in emphysema like symptoms.

[00746] Several mutations can cause ATTD, though the most common mutations are Glu342Lys (referred to as Z allele, wild-type is referred to as M) or Glu264Val (referred to as the S allele), and each allele contributes equally to the disease state, with two affected alleles resulting in more pronounced pathophysiology. These results not only resulted in degradation of the connective tissue of sensitive organs, such as the lung, but accumulation of the mutants in the liver can result in proteotoxicity. Current treatments focus on the replacement of A1AT by injection of protein retrieved from donated human plasma. In severe cases a lung and/or liver transplant may be considered.

[00747] The common variants of the disease are again monogenic. In some embodiments, the SERPINA1 gene is targeted. In some embodiments, the Glu342Lys mutation (referred to as Z allele, wild-type is referred to as M) or the Glu264Val mutation (referred to as the S allele) are corrected for. In some embodiments, therefore, the faulty gene would require replacement by the wild-type functioning gene. In some embodiments, a knockout and repair approach is required, so a repair template is provided. In the case of bi-allelic mutations, in some embodiments only one guide RNA would be required for homozygous mutations, but in the case of heterozygous mutations two guide RNAs may be required. Delivery is, in some embodiments, to the lung or liver.

Inborn errors of metabolism

[00748] Inborn errors of metabolism (IEMs) are an umbrella group of diseases which affect metabolic processes. In some embodiments, an IEM is to be treated. The majority of these diseases are monogenic in nature (e.g. phenylketonuria) and the pathophysiology results from either the abnormal accumulation of substances which are inherently toxic, or mutations which result in an inability to synthesize essential substances. Depending on the nature of the IEM, CRISPR/Cas may be used to facilitate a knock-out alone, or in combination with replacement of a faulty gene via a repair template. Exemplary diseases that may benefit from CRISPR/Cas technology are, in some embodiments: primary hyperoxaluria type 1 (PH1), argininosuccinic lyase deficiency, ornithine transcarbamylase deficiency, phenylketonuria, or PKU, and maple syrup urine disease.

Treating Epithelial and Lung Diseases

[00749] The present invention also contemplates delivering the CRISPR-Cas system described herein to one or both lungs.

[00750] Although AAV-2-based vectors were originally proposed for CFTR delivery to CF airways, other serotypes such as AAV-1, AAV-5, AAV-6, and AAV-9 exhibit improved gene transfer efficiency in a variety of models of the lung epithelium (see, e.g., Li et al., Molecular Therapy, vol. 17 no. 12, 2067-2077 Dec 2009). AAV-1 was demonstrated to be ~100-fold more efficient than AAV-2 and AAV-5 at transducing human airway epithelial cells in vitro,5 although AAV-1 transduced murine tracheal airway epithelia in vivo with an efficiency equal to that of AAV-5. Other studies have shown that AAV-5 is 50-fold more efficient than AAV-2 at gene delivery to human airway epithelium (HAE) in vitro and significantly more efficient in the mouse lung airway epithelium in vivo. AAV-6 has also been shown to be more efficient than AAV-2 in human airway epithelial cells in vitro and murine airways in vivo.8 The more recent isolate, AAV-9, was shown to display greater gene transfer efficiency than AAV-5 in murine nasal and alveolar epithelia in vivo with gene expression detected for over 9 months suggesting AAV may enable long-term gene expression in vivo, a desirable property for a CFTR gene delivery vector. Furthermore, it was demonstrated that AAV-9 could be re-administered to the murine lung with no loss of CFTR expression and minimal immune consequences. CF and non- CF HAE cultures may be inoculated on the apical surface with 100 μl of AAV vectors for hours (see, e.g., Li et al., Molecular Therapy, vol. 17 no. 12, 2067-2077 Dec 2009). The MOI may vary from 1 × 103 to 4 × 105 vector genomes/cell, depending on virus concentration and purposes of the experiments. The above cited vectors are contemplated for the delivery and/or administration of the invention.

[00751] Zamora et al. (Am J Respir Crit Care Med Vol 183. pp 531–538, 2011) reported an example of the application of an RNA interference therapeutic to the treatment of human infectious disease and also a randomized trial of an antiviral drug in respiratory syncytial virus (RSV)-infected lung transplant recipients. Zamora et al. performed a randomized, double-blind, placebocontrolled trial in LTX recipients with RSV respiratory tract infection. Patients were permitted to receive standard of care for RSV. Aerosolized ALN-RSV01 (0.6 mg/kg) or placebo was administered daily for 3 days. This study demonstrates that an RNAi therapeutic targeting RSV can be safely administered to LTX recipients with RSV infection. Three daily doses of ALN-RSV01 did not result in any exacerbation of respiratory tract symptoms or impairment of lung function and did not exhibit any systemic proinflammatory effects, such as induction of cytokines or CRP. Pharmacokinetics showed only low, transient systemic exposure after inhalation, consistent with preclinical animal data showing that ALN-RSV01, administered intravenously or by inhalation, is rapidly cleared from the circulation through exonucleasemediated digestion and renal excretion. The method of Zamora et al. may be applied to the nucleic acid-targeting system of the present invention and an aerosolized CRISPR Cas, for example with a dosage of 0.6 mg/kg, may be contemplated for the present invention.

[00752] Subjects treated for a lung disease may for example receive pharmaceutically effective amount of aerosolized AAV vector system per lung endobronchially delivered while spontaneously breathing. As such, aerosolized delivery is preferred for AAV delivery in general. An adenovirus or an AAV particle may be used for delivery. Suitable gene constructs, each operably linked to one or more regulatory sequences, may be cloned into the delivery vector. In this instance, the following constructs are provided as examples: Cbh or EF1a promoter for Cas, U6 or H1 promoter for guide RNA): A preferred arrangement is to use a CFTRdelta508 targeting guide, a repair template for deltaF508 mutation and a codon optimized Cas, with optionally one or more nuclear localization signal or sequence(s) (NLS(s)), e.g., two (2) NLSs. Constructs without NLS are also envisaged.

Treating Diseases of the Muscular System

[00753] The present invention also contemplates delivering the CRISPR-Cas system described herein to muscle(s).

[00754] Bortolanza et al. (Molecular Therapy vol. 19 no. 11, 2055–2064 Nov. 2011) shows that systemic delivery of RNA interference expression cassettes in the FRG1 mouse, after the onset of facioscapulohumeral muscular dystrophy (FSHD), led to a dose-dependent long-term FRG1 knockdown without signs of toxicity. Bortolanza et al. found that a single intravenous

injection of 5 × 1012 vg of rAAV6-sh1FRG1 rescues muscle histopathology and muscle function of FRG1 mice. In detail, 200 μl containing 2 × 1012 or 5 × 1012 vg of vector in physiological solution were injected into the tail vein using a 25-gauge Terumo syringe. The method of Bortolanza et al. may be applied to an AAV expressing CRISPR Cas and injected into humans at a dosage of about 2 × 1015 or 2 × 1016 vg of vector.

[00755] Dumonceaux et al. (Molecular Therapy vol. 18 no. 5, 881–887 May 2010) inhibit the myostatin pathway using the technique of RNA interference directed against the myostatin receptor AcvRIIb mRNA (sh-AcvRIIb). The restoration of a quasi-dystrophin was mediated by the vectorized U7 exon-skipping technique (U7-DYS). Adeno-associated vectors carrying either the sh-AcvrIIb construct alone, the U7-DYS construct alone, or a combination of both constructs were injected in the tibialis anterior (TA) muscle of dystrophic mdx mice. The injections were performed with 1011 AAV viral genomes. The method of Dumonceaux et al. may be applied to an AAV expressing CRISPR Cas and injected into humans, for example, at a dosage of about 1014 to about 1015 vg of vector.

[00756] Kinouchi et al. (Gene Therapy (2008) 15, 1126–1130) report the effectiveness of in vivo siRNA delivery into skeletal muscles of normal or diseased mice through nanoparticle formation of chemically unmodified siRNAs with atelocollagen (ATCOL). ATCOL-mediated local application of siRNA targeting myostatin, a negative regulator of skeletal muscle growth, in mouse skeletal muscles or intravenously, caused a marked increase in the muscle mass within a few weeks after application. These results imply that ATCOL-mediated application of siRNAs is a powerful tool for future therapeutic use for diseases including muscular atrophy. MstsiRNAs (final concentration, 10 mM) were mixed with ATCOL (final concentration for local administration, 0.5%) (AteloGene, Kohken, Tokyo, Japan) according to the manufacturer's instructions. After anesthesia of mice (20-week-old male C57BL/6) by Nembutal (25 mg/kg, i.p.), the Mst-siRNA/ATCOL complex was injected into the masseter and biceps femoris muscles. The method of Kinouchi et al. may be applied to CRISPR Cas and injected into a human, for example, at a dosage of about 500 to 1000 ml of a 40 μM solution into the muscle. Hagstrom et al. (Molecular Therapy Vol. 10, No. 2, August 2004) describe an intravascular, nonviral methodology that enables efficient and repeatable delivery of nucleic acids to muscle cells (myofibers) throughout the limb muscles of mammals. The procedure involves the injection of naked plasmid DNA or siRNA into a distal vein of a limb that is transiently isolated by a tourniquet or blood pressure cuff. Nucleic acid delivery to myofibers is facilitated by its rapid injection in sufficient volume to enable extravasation of the nucleic acid solution into muscle tissue. High levels of transgene expression in skeletal muscle were

achieved in both small and large animals with minimal toxicity. Evidence of siRNA delivery to limb muscle was also obtained. For plasmid DNA intravenous injection into a rhesus monkey, a threeway stopcock was connected to two syringe pumps (Model PHD 2000; Harvard Instruments), each loaded with a single syringe. Five minutes after a papaverine injection, pDNA (15.5 to 25.7 mg in 40 –100 ml saline) was injected at a rate of 1.7 or 2.0 ml/s. This could be scaled up for plasmid DNA expressing CRISPR Cas of the present invention with an injection of about 300 to 500 mg in 800 to 2000 ml saline for a human. For adenoviral vector injections into a rat, 2 x 109 infectious particles were injected in 3 ml of normal saline solution (NSS). This could be scaled up for an adenoviral vector expressing CRISPR Cas of the present invention with an injection of about 1 x 1013 infectious particles were injected in 10 liters of NSS for a human. For siRNA, a rat was injected into the great saphenous vein with 12.5 μg of a siRNA and a primate was injected injected into the great saphenous vein with 750 μg of a siRNA. This could be scaled up for a CRISPR Cas of the present invention, for example, with an injection of about 15 to about 50 mg into the great saphenous vein of a human.

[00757] See also, for example, WO2013163628 A2, Genetic Correction of Mutated Genes, published application of Duke University describes efforts to correct, for example, a frameshift mutation which causes a premature stop codon and a truncated gene product that can be corrected via nuclease mediated non-homologous end joining such as those responsible for Duchenne Muscular Dystrophy, ("DMD") a recessive, fatal, X-linked disorder that results in muscle degeneration due to mutations in the dystrophin gene. The majority of dystrophin mutations that cause DMD are deletions of exons that disrupt the reading frame and cause premature translation termination in the dystrophin gene. Dystrophin is a cytoplasmic protein that provides structural stability to the dystroglycan complex of the cell membrane that is responsible for regulating muscle cell integrity and function. The dystrophin gene or "DMD gene" as used interchangeably herein is 2.2 megabases at locus Xp21. The primary transcription measures about 2,400 kb with the mature mRNA being about 14 kb. 79 exons code for the protein which is over 3500 amino acids. Exon 51 is frequently adjacent to frame-disrupting deletions in DMD patients and has been targeted in clinical trials for oligonucleotide-based exon skipping. A clinical trial for the exon 51 skipping compound eteplirsen recently reported a significant functional benefit across 48 weeks, with an average of 47% dystrophin positive fibers compared to baseline. Mutations in exon 51 are ideally suited for permanent correction by NHEJ-based genome editing.

[00758] The methods of US Patent Publication No. 20130145487 assigned to Cellectis, which relates to meganuclease variants to cleave a target sequence from the human dystrophin

gene (DMD), may also be modified to for the nucleic acid-targeting system of the present invention.

Treating Diseases of the Skin

[00759] The present invention also contemplates delivering the CRISPR-Cas system described herein to the skin.

[00760] Hickerson et al. (Molecular Therapy—Nucleic Acids (2013) 2, e129) relates to a motorized microneedle array skin delivery device for delivering self-delivery (sd)-siRNA to human and murine skin. The primary challenge to translating siRNA-based skin therapeutics to the clinic is the development of effective delivery systems. Substantial effort has been invested in a variety of skin delivery technologies with limited success. In a clinical study in which skin was treated with siRNA, the exquisite pain associated with the hypodermic needle injection precluded enrollment of additional patients in the trial, highlighting the need for improved, more "patient-friendly" (i.e., little or no pain) delivery approaches. Microneedles represent an efficient way to deliver large charged cargos including siRNAs across the primary barrier, the stratum corneum, and are generally regarded as less painful than conventional hypodermic needles. Motorized "stamp type" microneedle devices, including the motorized microneedle array (MMNA) device used by Hickerson et al., have been shown to be safe in hairless mice studies and cause little or no pain as evidenced by (i) widespread use in the cosmetic industry and (ii) limited testing in which nearly all volunteers found use of the device to be much less painful than a flushot, suggesting siRNA delivery using this device will result in much less pain than was experienced in the previous clinical trial using hypodermic needle injections. The MMNA device (marketed as Triple-M or Tri-M by Bomtech Electronic Co, Seoul, South Korea) was adapted for delivery of siRNA to mouse and human skin. sd-siRNA solution (up to 300 μl of 0.1 mg/ml RNA) was introduced into the chamber of the disposable Tri-M needle cartridge (Bomtech), which was set to a depth of 0.1 mm. For treating human skin, deidentified skin (obtained immediately following surgical procedures) was manually stretched and pinned to a cork platform before treatment. All intradermal injections were performed using an insulin syringe with a 28-gauge 0.5-inch needle. The MMNA device and method of Hickerson et al. could be used and/or adapted to deliver the CRISPR Cas of the present invention, for example, at a dosage of up to 300 μl of 0.1 mg/ml CRISPR Cas to the skin.

[00761] Leachman et al. (Molecular Therapy, vol. 18 no. 2, 442–446 Feb. 2010) relates to a phase Ib clinical trial for treatment of a rare skin disorder pachyonychia congenita (PC), an autosomal dominant syndrome that includes a disabling plantar keratoderma, utilizing the first

short-interfering RNA (siRNA)-based therapeutic for skin. This siRNA, called TD101, specifically and potently targets the keratin 6a (K6a) N171K mutant mRNA without affecting wild-type K6a mRNA.

[00762] Zheng et al. (PNAS, July 24, 2012, vol. 109, no. 30, 11975–11980) show that spherical nucleic acid nanoparticle conjugates (SNA-NCs), gold cores surrounded by a dense shell of highly oriented, covalently immobilized siRNA, freely penetrate almost 100% of keratinocytes in vitro, mouse skin, and human epidermis within hours after application. Zheng et al. demonstrated that a single application of 25 nM epidermal growth factor receptor (EGFR) SNA-NCs for 60 h demonstrate effective gene knockdown in human skin. A similar dosage may be contemplated for CRISPR Cas immobilized in SNA-NCs for administration to the skin.

Cancer

[00763] In some embodiments, the treatment, prophylaxis or diagnosis of cancer is provided. The target is preferably one or more of the FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, TRAC or TRBC genes. The cancer may be one or more of lymphoma, chronic lymphocytic leukemia (CLL), B cell acute lymphocytic leukemia (B-ALL), acute lymphoblastic leukemia, acute myeloid leukemia, non-Hodgkin's lymphoma (NHL), diffuse large cell lymphoma (DLCL), multiple myeloma, renal cell carcinoma (RCC), neuroblastoma, colorectal cancer, breast cancer, ovarian cancer, melanoma, sarcoma, prostate cancer, lung cancer, esophageal cancer, hepatocellular carcinoma, pancreatic cancer, astrocytoma, mesothelioma, head and neck cancer, and medulloblastoma. This may be implemented with engineered chimeric antigen receptor (CAR) T cell. This is described in WO2015161276, the disclosure of which is hereby incorporated by reference and described herein below.

[00764] Target genes suitable for the treatment or prophylaxis of cancer may include, in some embodiments, those described in WO2015048577 the disclosure of which is hereby incorporated by reference.

Usher Syndrome or retinitis pigmentosa-39

[00765] In some embodiments, the treatment, prophylaxis or diagnosis of Usher Syndrome or retinitis pigmentosa-39 is provided. The target is preferably the USH2A gene. In some embodiments, correction of a G deletion at position 2299 (2299delG) is provided. This is described in WO2015134812A1, the disclosure of which is hereby incorporated by reference.

Autoimmune and inflammatory disorders

[00766] In some embodiments, autoimmune and inflammatory disorders are treated. These include Multiple Sclerosis (MS) or Rheumatoid Arthritis (RA), for example.

Cystic Fibrosis (CF)

[00767] In some embodiments, the treatment, prophylaxis or diagnosis of cystic fibrosis is provided. The target is preferably the SCNN1A or the CFTR gene. This is described in WO2015157070, the disclosure of which is hereby incorporated by reference.

[00768] Schwank et al. (Cell Stem Cell, 13:653–58, 2013) used CRISPR-Cas9 to correct a defect associated with cystic fibrosis in human stem cells. The team's target was the gene for an ion channel, cystic fibrosis transmembrane conductor receptor (CFTR). A deletion in CFTR causes the protein to misfold in cystic fibrosis patients. Using cultured intestinal stem cells developed from cell samples from two children with cystic fibrosis, Schwank et al. were able to correct the defect using CRISPR along with a donor plasmid containing the reparative sequence to be inserted. The researchers then grew the cells into intestinal "organoids," or miniature guts, and showed that they functioned normally. In this case, about half of clonal organoids underwent the proper genetic correction.

[00769] In some embodiments, Cystic fibrosis is treated, for example. Delivery to the lungs is therefore preferred. The F508 mutation (delta-F508, full name CFTRΔF508 or F508del-CFTR) is preferably corrected. In some embodiments, the targets may be ABCC7, CF or MRP7.

Duchenne's muscular dystrophy

[00770] Duchenne's muscular dystrophy (DMD) is a recessive, sex-linked muscle wasting disease that affects approximately 1 in 5000 males at birth. Mutations of the dystrophin gene result in an absence of dystrophin in skeletal muscle, where it normally functions to connect the cytoskeleton of the muscle fiber to the basal lamina. The absence of dystrophin caused be these mutations results in excessive calcium entry into the soma which causes the mitochondria to rupture, destroying the cell. Current treatments are focused on easing the symptoms of DMD, and the average life expetency is approximately 26 years.

[00771] CRISPR/Cas efficacy as a treatment for certain types of DMD has been demonstrated in mouse models. In one such study, the muscular dystrophy phenotype was partially corrected in the mouse by knocking-out a mutant exon resulting in a functional protein (see Nelson et al. (2016) Science, Long et al. (2016) Science, and Tabebordbar et al. (2016) Science).

[00772] In some embodiments, DMD is treated. In some embodiments, delivery is to the muscle by injection.

Glycogen Storage Diseases, including 1a

[00773] Glycogen Storage Disease 1a is a genetic disease resulting from deficiency of the enzyme glucose-6-phosphatase. The deficiency impairs the ability of the liver to produce free glucose from glycogen and from gluconeogenesis. In some embodiments, the gene encoding the glucose-6-phosphatase enzyme is targeted. In some embodiments, Glycogen Storage Disease 1a is treated. In some embodiments, delivery is to the liver by encapsulation of the Cas (in protein or mRNA form) in a lipid particle, such as an LNP.

[00774] In some embodiments, Glycogen Storage Diseases, including 1a, are targeted and preferably treated, for example by targeting polynucleotides associated with the condition/disease/infection. The associated polynucleotides include DNA, which may include genes (where genes include any coding sequence and regulatory elements such as enhancers or promoters). In some embodiments, the associated polynucleotides may include the SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, or PFKM genes.

Hurler Syndrome

[00775] Hurler syndrome, also known as mucopolysaccharidosis type I (MPS I), Hurler's disease, is a genetic disorder that results in the buildup of glycosaminoglycans (formerly known as mucopolysaccharides) due to a deficiency of alpha-L iduronidase, an enzyme responsible for the degradation of mucopolysaccharides in lysosomes. Hurler syndrome is often classified as a lysosomal storage disease, and is clinically related to Hunter Syndrome. Hunter syndrome is X-linked while Hurler syndrome is autosomal recessive. MPS I is divided into three subtypes based on severity of symptoms. All three types result from an absence of, or insufficient levels of, the enzyme α-L-iduronidase. MPS I H or Hurler syndrome is the most severe of the MPS I subtypes. The other two types are MPS I S or Scheie syndrome and MPS I H-S or Hurler-Scheie syndrome. Children born to an MPS I parent carry a defective IDUA gene, which has been mapped to the 4p16.3 site on chromosome 4. The gene is named IDUA because of its iduronidase enzyme protein product. As of 2001, 52 different mutations in the IDUA gene have been shown to cause Hurler syndrome. Successful treatment of the mouse, dog, and cat models of MPS I by delivery of the iduronidase gene through retroviral, lentiviral, AAV, and even nonviral vectors.

[00776] In some embodiments, the α-L-iduronidase gene is targeted and a repair template preferably provided.

HIV and AIDS

[00777] In some embodiments, the treatment, prophylaxis or diagnosis of HIV and AIDS is provided. The target is preferably the CCR5 gene in HIV. This is described in WO2015148670A1, the disclosure of which is hereby incorporated by reference.

Beta Thalassaemia

[00778] In some embodiments, the treatment, prophylaxis or diagnosis of Beta Thalassaemia is provided. The target is preferably the BCL11A gene. This is described in WO2015148860, the disclosure of which is hereby incorporated by reference.

Sickle Cell Disease (SCD)

[00779] In some embodiments, the treatment, prophylaxis or diagnosis of Sickle Cell Disease (SCD) is provided. The target is preferably the HBB or BCL11A gene. This is described in WO2015148863, the disclosure of which is hereby incorporated by reference.

Herpes Simplex Virus 1 and 2

[00780] Herpesviridae are a family of viruses composed of linear double-stranded DNA genomes with 75-200 genes. For the purposes of gene editing, the most commonly studied family member is Herpes Simplex Virus – 1 (HSV-1), a virus which has a distinct number of advantages over other viral vectors (reviewed in Vannuci et al. (2003)). Thus, in some embodiments, the viral vector is an HSV viral vector. In some embodiments, the HSV viral vector is HSV-1.

[00781] HSV-1 has a large genome of approximately 152 kb of double stranded DNA. This genome comprises of more than 80 genes, many of which can be replaced or removed, allowing a gene insert of between 30-150 kb. The viral vectors derived from HSV-1 are generally separated into 3 groups: replication-competent attenuated vectors, replication-incompetent recombinant vectors, and defective helper-dependent vectors known as amplicons. Gene transfer using HSV-1 as a vector has been demonstrated previously, for instance for the treatment of neuropathic pain (see, e.g., Wolfe et al. (2009) Gene Ther) and rheumatoid arthritis (see e.g., Burton et al. (2001) Stem Cells).

[00782] Thus, in some embodiments, the viral vector is an HSV viral vector. In some embodiments, the HSV viral vector is HSV-1. In some embodiments, the vector is used for delivery of one or more CRISPR components. It may be particularly useful for delivery of the Cas9 and one or more guide RNAs, for example 2 or more, 3 or more, or 4 or more guide RNAs. In some embodiments, the vector is theretofore useful in a multiplex system. In some embodiments, this delivery is for the treatment of treatment of neuropathic pain or rheumatoid arthritis.

[00783] In some embodiments, the treatment, prophylaxis or diagnosis of HSV-1 (Herpes Simplex Virus 1) is provided. The target is preferably the UL19, UL30, UL48 or UL50 gene in HSV-1. This is described in WO2015153789, the disclosure of which is hereby incorporated by reference.

[00784] In other embodiments, the treatment, prophylaxis or diagnosis of HSV-2 (Herpes Simplex Virus 2) is provided. The target is preferably the UL19, UL30, UL48 or UL50 gene in HSV-2. This is described in WO2015153791, the disclosure of which is hereby incorporated by reference.

[00785] In some embodiments, the treatment, prophylaxis or diagnosis of Primary Open Angle Glaucoma (POAG) is provided. The target is preferably the MYOC gene. This is described in WO2015153780, the disclosure of which is hereby incorporated by reference.

Adoptive Cell Therapies

[00786] The present invention also contemplates use of the CRISPR-Cas system described herein to modify cells for adoptive therapies. Aspects of the invention accordingly involve the adoptive transfer of immune system cells, such as T cells, specific for selected antigens, such as tumor associated antigens (see Maus et al., 2014, Adoptive Immunotherapy for Cancer or Viruses, Annual Review of Immunology, Vol. 32: 189-225; Rosenberg and Restifo, 2015, Adoptive cell transfer as personalized immunotherapy for human cancer, Science Vol. 348 no. 6230 pp. 62-68; and, Restifo et al., 2015, Adoptive immunotherapy for cancer: harnessing the T cell response. Nat. Rev. Immunol. 12(4): 269-281; and Jenson and Riddell, 2014, Design and implementation of adoptive therapy with chimeric antigen receptor-modified T cells. Immunol Rev. 257(1): 127–144). Various strategies may for example be employed to genetically modify T cells by altering the specificity of the T cell receptor (TCR) for example by introducing new TCR α and β chains with selected peptide specificity (see U.S. Patent No. 8,697,854; PCT Patent Publications: WO2003020763, WO2004033685, WO2004044004, WO2005114215, WO2006000830, WO2008038002, WO2008039818, WO2004074322, WO2005113595, WO2006125962, WO2013166321, WO2013039889, WO2014018863, WO2014083173; U.S. Patent No. 8,088,379).

[00787] As an alternative to, or addition to, TCR modifications, chimeric antigen receptors (CARs) may be used in order to generate immunoresponsive cells, such as T cells, specific for selected targets, such as malignant cells, with a wide variety of receptor chimera constructs having been described (see U.S. Patent Nos. 5,843,728; 5,851,828; 5,912,170; 6,004,811; 6,284,240; 6,392,013; 6,410,014; 6,753,162; 8,211,422; and, PCT Publication WO9215322). Alternative CAR constructs may be characterized as belonging to successive generations. First-

generation CARs typically consist of a single-chain variable fragment of an antibody specific for an antigen, for example comprising a VL linked to a VH of a specific antibody, linked by a flexible linker, for example by a CD8α hinge domain and a CD8α transmembrane domain, to the transmembrane and intracellular signaling domains of either CD3ζ or FcRγ (scFv-CD3ζ or scFv-FcRγ; see U.S. Patent No. 7,741,465; U.S. Patent No. 5,912,172; U.S. Patent No. 5,906,936). Second-generation CARs incorporate the intracellular domains of one or more costimulatory molecules, such as CD28, OX40 (CD134), or 4-1BB (CD137) within the endodomain (for example scFv-CD28/OX40/4-1BB-CD3ζ; see U.S. Patent Nos. 8,911,993; 8,916,381; 8,975,071; 9,101,584; 9,102,760; 9,102,761). Third-generation CARs include a combination of costimulatory endodomains, such a CD3ζ-chain, CD97, GDI la-CD18, CD2, ICOS, CD27, CD154, CDS, OX40, 4-1BB, or CD28 signaling domains (for example scFv-CD28-4-1BB-CD3ζ or scFv-CD28-OX40-CD3ζ; see U.S. Patent No. 8,906,682; U.S. Patent No. 8,399,645; U.S. Pat. No. 5,686,281; PCT Publication No. WO2014134165; PCT Publication No. WO2012079000). Alternatively, costimulation may be orchestrated by expressing CARs in antigen-specific T cells, chosen so as to be activated and expanded following engagement of their native αβTCR, for example by antigen on professional antigen-presenting cells, with attendant costimulation. In addition, additional engineered receptors may be provided on the immunoresponsive cells, for example to improve targeting of a T-cell attack and/or minimize side effects.

[00788] Alternative techniques may be used to transform target immunoresponsive cells, such as protoplast fusion, lipofection, transfection or electroporation. A wide variety of vectors may be used, such as retroviral vectors, lentiviral vectors, adenoviral vectors, adeno-associated viral vectors, plasmids or transposons, such as a Sleeping Beauty transposon (see U.S. Patent Nos. 6,489,458; 7,148,203; 7,160,682; 7,985,739; 8,227,432), may be used to introduce CARs, for example using 2nd generation antigen-specific CARs signaling through CD3ζ and either CD28 or CD137. Viral vectors may for example include vectors based on HIV, SV40, EBV, HSV or BPV.

[00789] Cells that are targeted for transformation may for example include T cells, Natural Killer (NK) cells, cytotoxic T lymphocytes (CTL), regulatory T cells, human embryonic stem cells, tumor-infiltrating lymphocytes (TIL) or a pluripotent stem cell from which lymphoid cells may be differentiated. T cells expressing a desired CAR may for example be selected through co-culture with γ-irradiated activating and propagating cells (AaPC), which co-express the cancer antigen and co-stimulatory molecules. The engineered CAR T-cells may be expanded, for example by co-culture on AaPC in presence of soluble factors, such as IL-2 and

IL-21. This expansion may for example be carried out so as to provide memory CAR+ T cells (which may for example be assayed by non-enzymatic digital array and/or multi-panel flow cytometry). In this way, CAR T cells may be provided that have specific cytotoxic activity against antigen-bearing tumors (optionally in conjunction with production of desired chemokines such as interferon-γ). CAR T cells of this kind may for example be used in animal models, for example to threat tumor xenografts.

[00790] Approaches such as the foregoing may be adapted to provide methods of treating and/or increasing survival of a subject having a disease, such as a neoplasia, for example by administering an effective amount of an immunoresponsive cell comprising an antigen recognizing receptor that binds a selected antigen, wherein the binding activates the immunoreponsive cell, thereby treating or preventing the disease (such as a neoplasia, a pathogen infection, an autoimmune disorder, or an allogeneic transplant reaction). Dosing in CAR T cell therapies may for example involve administration of from 106 to 109 cells/kg, with or without a course of lymphodepletion, for example with cyclophosphamide.

[00791] In one embodiment, the treatment can be administrated into patients undergoing an immunosuppressive treatment. The cells or population of cells, may be made resistant to at least one immunosuppressive agent due to the inactivation of a gene encoding a receptor for such immunosuppressive agent. Not being bound by a theory, the immunosuppressive treatment should help the selection and expansion of the immunoresponsive or T cells according to the invention within the patient.

[00792] The administration of the cells or population of cells according to the present invention may be carried out in any convenient manner, including by aerosol inhalation, injection, ingestion, transfusion, implantation or transplantation. The cells or population of cells may be administered to a patient subcutaneously, intradermally, intratumorally, intranodally, intramedullary, intramuscularly, by intravenous or intralymphatic injection, or intraperitoneally. In one embodiment, the cell compositions of the present invention are preferably administered by intravenous injection.

[00793] The administration of the cells or population of cells can consist of the administration of 104- 109 cells per kg body weight, preferably 105 to 106 cells/kg body weight including all integer values of cell numbers within those ranges. Dosing in CAR T cell therapies may for example involve administration of from 106 to 109 cells/kg, with or without a course of lymphodepletion, for example with cyclophosphamide. The cells or population of cells can be administrated in one or more doses. In another embodiment, the effective amount of cells are administrated as a single dose. In another embodiment, the effective amount of cells are

administrated as more than one dose over a period time. Timing of administration is within the judgment of managing physician and depends on the clinical condition of the patient. The cells or population of cells may be obtained from any source, such as a blood bank or a donor. While individual needs vary, determination of optimal ranges of effective amounts of a given cell type for a particular disease or conditions are within the skill of one in the art. An effective amount means an amount which provides a therapeutic or prophylactic benefit. The dosage administrated will be dependent upon the age, health and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment and the nature of the effect desired.

[00794] In another embodiment, the effective amount of cells or composition comprising those cells are administrated parenterally. The administration can be an intravenous administration. The administration can be directly done by injection within a tumor.

[00795] To guard against possible adverse reactions, engineered immunoresponsive cells may be equipped with a transgenic safety switch, in the form of a transgene that renders the cells vulnerable to exposure to a specific signal. For example, the herpes simplex viral thymidine kinase (TK) gene may be used in this way, for example by introduction into allogeneic T lymphocytes used as donor lymphocyte infusions following stem cell transplantation (Greco, et al., Improving the safety of cell therapy with the TK-suicide gene. Front. Pharmacol. 2015; 6: 95). In such cells, administration of a nucleoside prodrug such as ganciclovir or acyclovir causes cell death. Alternative safety switch constructs include inducible caspase 9, for example triggered by administration of a small-molecule dimerizer that brings together two nonfunctional icasp9 molecules to form the active enzyme. A wide variety of alternative approaches to implementing cellular proliferation controls have been described (see U.S. Patent Publication No. 20130071414; PCT Patent Publication WO2011146862; PCT Patent Publication WO2014011987; PCT Patent Publication WO2013040371; Zhou et al. BLOOD, 2014, 123/25:3895 – 3905; Di Stasi et al., The New England Journal of Medicine 2011; 365:1673-1683; Sadelain M, The New England Journal of Medicine 2011; 365:1735-173; Ramos et al., Stem Cells 28(6):1107-15 (2010)).

[00796] In a further refinement of adoptive therapies, genome editing with a CRISPR-Cas system as described herein may be used to tailor immunoresponsive cells to alternative implementations, for example providing edited CAR T cells (see Poirot et al., 2015, Multiplex genome edited T-cell manufacturing platform for "off-the-shelf" adoptive T-cell immunotherapies, Cancer Res 75 (18): 3853). For example, immunoresponsive cells may be edited to delete expression of some or all of the class of HLA type II and/or type I molecules,

or to knockout selected genes that may inhibit the desired immune response, such as the PD1 gene.

[00797] Cells may be edited using any CRISPR system and method of use thereof as described herein. CRISPR systems may be delivered to an immune cell by any method described herein. In preferred embodiments, cells are edited ex vivo and transferred to a subject in need thereof. Immunoresponsive cells, CAR T cells or any cells used for adoptive cell transfer may be edited. Editing may be performed to eliminate potential alloreactive T-cell receptors (TCR), disrupt the target of a chemotherapeutic agent, block an immune checkpoint, activate a T cell, and/or increase the differentiation and/or proliferation of functionally exhausted or dysfunctional CD8+ T-cells (see PCT Patent Publications: WO2013176915, WO2014059173, WO2014172606, WO2014184744, and WO2014191128). Editing may result in inactivation of a gene.

[00798] By inactivating a gene it is intended that the gene of interest is not expressed in a functional protein form. In a particular embodiment, the CRISPR system specifically catalyzes cleavage in one targeted gene thereby inactivating said targeted gene. The nucleic acid strand breaks caused are commonly repaired through the distinct mechanisms of homologous recombination or non-homologous end joining (NHEJ). However, NHEJ is an imperfect repair process that often results in changes to the DNA sequence at the site of the cleavage. Repair via non-homologous end joining (NHEJ) often results in small insertions or deletions (Indel) and can be used for the creation of specific gene knockouts. Cells in which a cleavage induced mutagenesis event has occurred can be identified and/or selected by well-known methods in the art.

[00799] T cell receptors (TCR) are cell surface receptors that participate in the activation of T cells in response to the presentation of antigen. The TCR is generally made from two chains, α and β, which assemble to form a heterodimer and associates with the CD3-transducing subunits to form the T cell receptor complex present on the cell surface. Each α and β chain of the TCR consists of an immunoglobulin-like N-terminal variable (V) and constant (C) region, a hydrophobic transmembrane domain, and a short cytoplasmic region. As for immunoglobulin molecules, the variable region of the α and β chains are generated by V(D)J recombination, creating a large diversity of antigen specificities within the population of T cells. However, in contrast to immunoglobulins that recognize intact antigen, T cells are activated by processed peptide fragments in association with an MHC molecule, introducing an extra dimension to antigen recognition by T cells, known as MHC restriction. Recognition of MHC disparities between the donor and recipient through the T cell receptor leads to T cell proliferation and the

potential development of graft versus host disease (GVHD). The inactivation of TCRα or TCRβ can result in the elimination of the TCR from the surface of T cells preventing recognition of alloantigen and thus GVHD. However, TCR disruption generally results in the elimination of the CD3 signaling component and alters the means of further T cell expansion.

[00800] Allogeneic cells are rapidly rejected by the host immune system. It has been demonstrated that, allogeneic leukocytes present in non-irradiated blood products will persist for no more than 5 to 6 days (Boni, Muranski et al. 2008 Blood 1;112(12):4746-54). Thus, to prevent rejection of allogeneic cells, the host's immune system usually has to be suppressed to some extent. However, in the case of adoptive cell transfer the use of immunosuppressive drugs also have a detrimental effect on the introduced therapeutic T cells. Therefore, to effectively use an adoptive immunotherapy approach in these conditions, the introduced cells would need to be resistant to the immunosuppressive treatment. Thus, in a particular embodiment, the present invention further comprises a step of modifying T cells to make them resistant to an immunosuppressive agent, preferably by inactivating at least one gene encoding a target for an immunosuppressive agent. An immunosuppressive agent is an agent that suppresses immune function by one of several mechanisms of action. An immunosuppressive agent can be, but is not limited to a calcineurin inhibitor, a target of rapamycin, an interleukin-2 receptor α-chain blocker, an inhibitor of inosine monophosphate dehydrogenase, an inhibitor of dihydrofolic acid reductase, a corticosteroid or an immunosuppressive antimetabolite. The present invention allows conferring immunosuppressive resistance to T cells for immunotherapy by inactivating the target of the immunosuppressive agent in T cells. As non-limiting examples, targets for an immunosuppressive agent can be a receptor for an immunosuppressive agent such as: CD52, glucocorticoid receptor (GR), a FKBP family gene member and a cyclophilin family gene member.

[00801] Immune checkpoints are inhibitory pathways that slow down or stop immune reactions and prevent excessive tissue damage from uncontrolled activity of immune cells. In certain embodiments, the immune checkpoint targeted is the programmed death-1 (PD-1 or CD279) gene (PDCD1). In other embodiments, the immune checkpoint targeted is cytotoxic T-lymphocyte-associated antigen (CTLA-4). In additional embodiments, the immune checkpoint targeted is another member of the CD28 and CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. In further additional embodiments, the immune checkpoint targeted is a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3.

[00802] Additional immune checkpoints include Src homology 2 domain-containing protein tyrosine phosphatase 1 (SHP-1) (Watson HA, et al., SHP-1: the next checkpoint target for cancer immunotherapy? Biochem Soc Trans. 2016 Apr 15;44(2):356-62). SHP-1 is a widely expressed inhibitory protein tyrosine phosphatase (PTP). In T-cells, it is a negative regulator of antigen-dependent activation and proliferation. It is a cytosolic protein, and therefore not amenable to antibody-mediated therapies, but its role in activation and proliferation makes it an attractive target for genetic manipulation in adoptive transfer strategies, such as chimeric antigen receptor (CAR) T cells. Immune checkpoints may also include T cell immunoreceptor with Ig and ITIM domains (TIGIT/Vstm3/WUCAM/VSIG9) and VISTA (Le Mercier I, et al., (2015) Beyond CTLA-4 and PD-1, the generation Z of negative checkpoint regulators. Front. Immunol. 6:418).

[00803] WO2014172606 relates to the use of MT1 and/or MT1 inhibitors to increase proliferation and/or activity of exhausted CD8+ T-cells and to decrease CD8+ T-cell exhaustion (e.g., decrease functionally exhausted or unresponsive CD8+ immune cells). In certain embodiments, metallothioneins are targeted by gene editing in adoptively transferred T cells.

[00804] In certain embodiments, targets of gene editing may be at least one targeted locus involved in the expression of an immune checkpoint protein. Such targets may include, but are not limited to CTLA4, PPP2CA, PPP2CB, PTPN6, PTPN22, PDCD1, ICOS (CD278), PDL1, KIR, LAG3, HAVCR2, BTLA, CD160, TIGIT, CD96, CRTAM, LAIR1, SIGLEC7, SIGLEC9, CD244 (2B4), TNFRSF10B, TNFRSF10A, CASP8, CASP10, CASP3, CASP6, CASP7, FADD, FAS, TGFBRII, TGFRBRI, SMAD2, SMAD3, SMAD4, SMAD10, SKI, SKIL, TGIF1, IL10RA, IL10RB, HMOX2, IL6R, IL6ST, EIF2AK4, CSK, PAG1, SIT1, FOXP3, PRDM1, BATF, VISTA, GUCY1A2, GUCY1A3, GUCY1B2, GUCY1B3, MT1, MT2, CD40, OX40, CD137, GITR, CD27, SHP-1 or TIM-3. In preferred embodiments, the gene locus involved in the expression of PD-1 or CTLA-4 genes is targeted. In other preferred embodiments, combinations of genes are targeted, such as but not limited to PD-1 and TIGIT.

[00805] In other embodiments, at least two genes are edited. Pairs of genes may include, but are not limited to PD1 and TCRα, PD1 and TCRβ, CTLA-4 and TCRα, CTLA-4 and TCRβ, LAG3 and TCRα, LAG3 and TCRβ, Tim3 and TCRα, Tim3 and TCRβ, BTLA and TCRα, BTLA and TCRβ, BY55 and TCRα, BY55 and TCRβ, TIGIT and TCRα, TIGIT and TCRβ, B7H5 and TCRα, B7H5 and TCRβ, LAIR1 and TCRα, LAIR1 and TCRβ, SIGLEC10 and TCRα, SIGLEC10 and TCRβ, 2B4 and TCRα, 2B4 and TCRβ.

[00806] Whether prior to or after genetic modification of the T cells, the T cells can be activated and expanded generally using methods as described, for example, in U.S. Patents 6,352,694; 6,534,055; 6,905,680; 5,858,358; 6,887,466; 6,905,681; 7,144,575; 7,232,566; 7,175,843; 5,883,223; 6,905,874; 6,797,514; 6,867,041; and 7,572,631. T cells can be expanded in vitro or in vivo.

[00807] The practice of the present invention employs, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989) (Sambrook, Fritsch and Maniatis); MOLECULAR CLONING: A LABORATORY MANUAL, 4th edition (2012) (Green and Sambrook); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (1987) (F. M. Ausubel, et al. eds.); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.); PCR 2: A PRACTICAL APPROACH (1995) (M.J. MacPherson, B.D. Hames and G.R. Taylor eds.); ANTIBODIES, A LABORATORY MANUAL (1988) (Harlow and Lane, eds.); ANTIBODIES A LABORATORY MANUAL, 2nd edition (2013) (E.A. Greenfield ed.); and ANIMAL CELL CULTURE (1987) (R.I. Freshney, ed.).

[00808] The practice of the present invention employs, unless otherwise indicated, conventional techniques for generation of genetically modified mice. See Marten H. Hofker and Jan van Deursen, TRANSGENIC MOUSE METHODS AND PROTOCOLS, 2nd edition (2011).

[00809] In some embodiments, the invention described herein relates to a method for adoptive immunotherapy, in which T cells are edited ex vivo by CRISPR to modulate at least one gene and subsequently administered to a patient in need thereof. In some embodiments, the CRISPR editing comprising knocking-out or knocking-down the expression of a target gene in the edited T cells. In some embodiments, in addition to modulating the target gene, the T cells are also edited ex vivo by CRISPR to (1) knock-in an exogenous gene encoding a chimeric antigen receptor (CAR) or a T-cell receptor (TCR), (2) knock-out or knock-down expression of an immune checkpoint receptor, (3) knock-out or knock-down expression of an endogenous TCR, (4) knock-out or knock-down expression of a human leukocyte antigen class I (HLA-I) proteins, and/or (5) knock-out or knock-down expression of an endogenous gene encoding an antigen targeted by an exogenous CAR or TCR.

[00810] In some embodiments, the T cells are contacted ex vivo with an adeno-associated virus (AAV) vector encoding a CRISPR effector protein, and a guide molecule comprising a

guide sequence hybridizable to a target sequence, a tracr mate sequence, and a tracr sequence hybridizable to the tracr mate sequence. In some embodiments, the T cells are contacted ex vivo (e.g., by electroporation) with a ribonucleoprotein (RNP) comprising a CRISPR effector protein complexed with a guide molecule, wherein the guide molecule comprising a guide sequence hybridizable to a target sequence, a tracr mate sequence, and a tracr sequence hybridizable to the tracr mate sequence. See Rupp et al., Scientific Reports 7:737 (2017); Liu et al., Cell Research 27:154-157 (2017). In some embodiments, the T cells are contacted ex vivo (e.g., by electroporation) with an mRNA encoding a CRISPR effector protein, and a guide molecule comprising a guide sequence hybridizable to a target sequence, a tracr mate sequence, and a tracr sequence hybridizable to the tracr mate sequence. See Eyquem et al., Nature 543:113-117 (2017). In some embodiments, the T cells are not contacted ex vivo with a lentivirus or retrovirus vector.

[00811] In some embodiments, the method comprises editing T cells ex vivo by CRISPR to knock-in an exogenous gene encoding a CAR, thereby allowing the edited T cells to recognize cancer cells based on the expression of specific proteins located on the cell surface. In some embodiments, T cells are edited ex vivo by CRISPR to knock-in an exogenous gene encoding a TCR, thereby allowing the edited T cells to recognize proteins derived from either the surface or inside of the cancer cells. In some embodiments, the method comprising providing an exogenous CAR-encoding or TCR-encoding sequence as a donor sequence, which can be integrated by homology-directed repair (HDR) into a genomic locus targeted by a CRISPR guide sequence. In some embodiments, targeting the exogenous CAR or TCR to an endogenous TCR α constant (TRAC) locus can reduce tonic CAR signaling and facilitate effective internalization and re-expression of the CAR following single or repeated exposure to antigen, thereby delaying effector T-cell differentiation and exhaustion. See Eyquem et al., Nature 543:113-117 (2017).

[00812] In some embodiments, the method comprises editing T cells ex vivo by CRISPR to block one or more immune checkpoint receptors to reduce immunosuppression by cancer cells. In some embodiments, T cells are edited ex vivo by CRISPR to knock-out or knock-down an endogenous gene involved in the programmed death-1 (PD-1) signaling pathway, such as PD-1 and PD-L1. In some embodiments, T cells are edited ex vivo by CRISPR to mutate the Pdcd1 locus or the CD274 locus. In some embodiments, T cells are edited ex vivo by CRISPR using one or more guide sequences targeting the first exon of PD-1. See Rupp et al., Scientific Reports 7:737 (2017); Liu et al., Cell Research 27:154-157 (2017).

[00813] In some embodiments, the method comprises editing T cells ex vivo by CRISPR to eliminate potential alloreactive TCRs to allow allogeneic adoptive transfer. In some embodiments, T cells are edited ex vivo by CRISPR to knock-out or knock-down an endogenous gene encoding a TCR (e.g., an αβ TCR) to avoid graft-versus-host-disease (GVHD). In some embodiments, T cells are edited ex vivo by CRISPR to mutate the TRAC locus. In some embodiments, T cells are edited ex vivo by CRISPR using one or more guide sequences targeting the first exon of TRAC. See Liu et al., Cell Research 27:154-157 (2017). In some embodiments, the method comprises use of CRISPR to knock-in an exogenous gene encoding a CAR or a TCR into the TRAC locus, while simultaneously knocking-out the endogenous TCR (e.g., with a donor sequence encoding a self-cleaving P2A peptide following the CAR cDNA). See Eyquem et al., Nature 543:113-117 (2017). In some embodiments, the exogenous gene comprises a promoter-less CAR-encoding or TCR-encoding sequence which is inserted operably downstream of an endogenous TCR promoter.

[00814] In some embodiments, the method comprises editing T cells ex vivo by CRISPR to knock-out or knock-down an endogenous gene encoding an HLA-I protein to minimize immunogenicity of the edited T cells. In some embodiments, T cells are edited ex vivo by CRISPR to mutate the beta-2 microglobulin (B2M) locus. In some embodiments, T cells are edited ex vivo by CRISPR using one or more guide sequences targeting the first exon of B2M. See Liu et al., Cell Research 27:154-157 (2017). In some embodiments, the method comprises use of CRISPR to knock-in an exogenous gene encoding a CAR or a TCR into the B2M locus, while simultaneously knocking-out the endogenous B2M (e.g., with a donor sequence encoding a self-cleaving P2A peptide following the CAR cDNA). See Eyquem et al., Nature 543:113-117 (2017). In some embodiments, the exogenous gene comprises a promoter-less CAR-encoding or TCR-encoding sequence which is inserted operably downstream of an endogenous B2M promoter.

[00815] In some embodiments, the method comprises editing T cells ex vivo by CRISPR to knock-out or knock-down an endogenous gene encoding an antigen targeted by an exogenous CAR or TCR. In some embodiments, the T cells are edited ex vivo by CRISPR to knock-out or knock-down the expression of a tumor antigen selected from human telomerase reverse transcriptase (hTERT), survivin, mouse double minute 2 homolog (MDM2), cytochrome P450 1B 1 (CYP1B), HER2/neu, Wilms' tumor gene 1 (WT1), livin, alphafetoprotein (AFP), carcinoembryonic antigen (CEA), mucin 16 (MUC16), MUC1, prostate-specific membrane antigen (PSMA), p53 or cyclin (DI) (see WO2016/011210). In some embodiments, the T cells are edited ex vivo by CRISPR to knock-out or knock-down the expression of an antigen

selected from B cell maturation antigen (BCMA), transmembrane activator and CAML Interactor (TACI), or B-cell activating factor receptor (BAFF-R), CD38, CD138, CS-1, CD33, CD26, CD30, CD53, CD92, CD100, CD148, CD150, CD200, CD261, CD262, or CD362 (see WO2017/011804).

Gene Drives

[00816] The present invention also contemplates use of the CRISPR-Cas system described herein to provide RNA-guided gene drives, for example in systems analogous to gene drives described in PCT Patent Publication WO 2015/105928. Systems of this kind may for example provide methods for altering eukaryotic germline cells, by introducing into the germline cell a nucleic acid sequence encoding an RNA-guided DNA nuclease and one or more guide RNAs. The guide RNAs may be designed to be complementary to one or more target locations on genomic DNA of the germline cell. The nucleic acid sequence encoding the RNA guided DNA nuclease and the nucleic acid sequence encoding the guide RNAs may be provided on constructs between flanking sequences, with promoters arranged such that the germline cell may express the RNA guided DNA nuclease and the guide RNAs, together with any desired cargo-encoding sequences that are also situated between the flanking sequences. The flanking sequences will typically include a sequence which is identical to a corresponding sequence on a selected target chromosome, so that the flanking sequences work with the components encoded by the construct to facilitate insertion of the foreign nucleic acid construct sequences into genomic DNA at a target cut site by mechanisms such as homologous recombination, to render the germline cell homozygous for the foreign nucleic acid sequence. In this way, gene-drive systems are capable of introgressing desired cargo genes throughout a breeding population (Gantz et al., 2015, Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito Anopheles stephensi, PNAS 2015, published ahead of print November 23, 2015, doi:10.1073/pnas.1521077112; Esvelt et al., 2014, Concerning RNA-guided gene drives for the alteration of wild populations eLife 2014;3:e03401). In select embodiments, target sequences may be selected which have few potential off-target sites in a genome. Targeting multiple sites within a target locus, using multiple guide RNAs, may increase the cutting frequency and hinder the evolution of drive resistant alleles. Truncated guide RNAs may reduce off-target cutting. Paired nickases may be used instead of a single nuclease, to further increase specificity. Gene drive constructs may include cargo sequences encoding transcriptional regulators, for example to activate homologous recombination genes and/or repress non-homologous end-joining. Target sites may be chosen within an essential gene, so that non-homologous end-joining events may cause

lethality rather than creating a drive-resistant allele. The gene drive constructs can be engineered to function in a range of hosts at a range of temperatures (Cho et al. 2013, Rapid and Tunable Control of Protein Stability in Caenorhabditis elegans Using a Small Molecule, PLoS ONE 8(8): e72393. doi:10.1371/journal.pone.0072393).

Xenotransplantation

[00817] The present invention also contemplates use of the CRISPR-Cas system described herein to provide RNA-guided DNA nucleases adapted to be used to provide modified tissues for transplantation. For example, RNA-guided DNA nucleases may be used to knockout, knockdown or disrupt selected genes in an animal, such as a transgenic pig (such as the human heme oxygenase-1 transgenic pig line), for example by disrupting expression of genes that encode epitopes recognized by the human immune system, i.e. xenoantigen genes. Candidate porcine genes for disruption may for example include α(1,3)-galactosyltransferase and cytidine monophosphate-N-acetylneuraminic acid hydroxylase genes (see PCT Patent Publication WO 2014/066505). In addition, genes encoding endogenous retroviruses may be disrupted, for example the genes encoding all porcine endogenous retroviruses (see Yang et al., 2015, Genome-wide inactivation of porcine endogenous retroviruses (PERVs), Science 27 November 2015: Vol. 350 no. 6264 pp. 1101-1104). In addition, RNA-guided DNA nucleases may be used to target a site for integration of additional genes in xenotransplant donor animals, such as a human CD55 gene to improve protection against hyperacute rejection.

General Gene Therapy Considerations

[00818] Examples of disease-associated genes and polynucleotides amd disease specific information is available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web.

[00819] Mutations in these genes and pathways can result in production of improper proteins or proteins in improper amounts which affect function. Further examples of genes, diseases and proteins are hereby incorporated by reference from US Provisional application 61/736,527 filed December 12, 2012. Such genes, proteins and pathways may be the target polynucleotide of a CRISPR complex of the present invention. Examples of disease-associated genes and polynucleotides are listed in Tables 10 and 11. Examples of signaling biochemical pathway-associated genes and polynucleotides are listed in Table 10.

[00820] Table 12. Examples of signaling biochemical pathway-associated genes and polynucleotides.

DISEASE/DISORDERS	GENE(S)

Neoplasia	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4;
	Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF;
	HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alpha; PPAR
	gamma; WT1 (Wilms Tumor); FGF Receptor Family
	members (5 members: 1, 2, 3, 4, 5); CDKN2a; APC; RB
	(retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR
	(Androgen Receptor); TSG101; IGF; IGF Receptor; Igf1 (4
	variants); Igf2 (3 variants); Igf 1 Receptor; Igf 2 Receptor;
	Bax; Bcl2; caspases family (9 members:
	1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
Age-related Macular	Abcr; Ccl2; Cc2; cp (ceruloplasmin); Timp3; cathepsinD;
Degeneration	Vldlr; Ccr2
Schizophrenia	Neuregulin1 (Nrg1); Erb4 (receptor for Neuregulin);
	Complexin1 (Cplx1); Tph1 Tryptophan hydroxylase; Tph2
	Tryptophan hydroxylase 2; Neurexin 1; GSK3; GSK3a;
	GSK3b
Disorders	5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA;
	DTNBP1; Dao (Dao1)
Trinucleotide Repeat	HTT (Huntington's Dx); SBMA/SMAX1/AR (Kennedy's
Disorders	Dx); FXN/X25 (Friedrich's Ataxia); ATX3 (Machado-
	Joseph's Dx); ATXN1 and ATXN2 (spinocerebellar
	ataxias); DMPK (myotonic dystrophy); Atrophin-1 and Atn1
	(DRPLA Dx); CBP (Creb-BP - global instability); VLDLR
	(Alzheimer's); Atxn7; Atxn10
Fragile X Syndrome	FMR2; FXR1; FXR2; mGLUR5
Secretase Related	APH-1 (alpha and beta); Presenilin (Psen1); nicastrin
Disorders	(Ncstn); PEN-2
Others	Nos1; Parp1; Nat1; Nat2
Prion - related disorders	Prp
ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a;
	VEGF-b; VEGF-c)
Drug addiction	Prkce (alcohol); Drd2; Drd4; ABAT (alcohol); GRIA2;
	Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1 (alcohol)
Autism	Mecp2; BZRAP1; MDGA2; Sema5A; Neurexin 1; Fragile X
	(FMR2 (AFF2); FXR1; FXR2; Mglur5)
Alzheimer's Disease	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterin; PS1;
	SORL1; CR1; Vldlr; Uba1; Uba3; CHIP28 (Aqp1,
	Aquaporin 1); Uchl1; Uchl3; APP
Inflammation	IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL-
	17b; IL-17c; IL-17d; IL-17f); II-23; Cx3cr1; ptpn22; TNFa;
	NOD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b);
	CTLA4; Cx3cl1
Parkinson's Disease	x-Synuclein; DJ-1; LRRK2; Parkin; PINK1

[00821] Table 12 (Cont'd)

Blood and coagulation diseases and disorders	Anemia (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); Bare lymphocyte syndrome (TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5), Bleeding disorders (TBXA2R, P2RX1, P2X1); Factor H and factor H-like 1 (HF1, CFH, HUS); Factor V and factor VIII (MCFD2); Factor VII deficiency (F7); Factor X deficiency (F10); Factor XI deficiency (F11); Factor XII deficiency (F12, HAF); Factor XIIIA deficiency (F13A1, F13A); Factor XIIIB deficiency (F13B); Fanconi anemia (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCJ, PHF9, FANCL, FANCM, KIAA1596); Hemophagocytic lymphohistiocytosis disorders (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemophilia A (F8, F8C, HEMA); Hemophilia B (F9, HEMB), Hemorrhagic disorders (PI, ATT, F5); Leukocyde deficiencies and disorders (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); Sickle cell anemia (HBB); Thalassemia (HBA2, HBB, HBD, LCRB, HBA1).
Cell dysregulation and oncology diseases and disorders	B-cell non-Hodgkin lymphoma (BCL7A, BCL7); Leukemia (TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL11, ARLTS1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN).
Inflammation and immune related diseases and disorders	AIDS (KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1); Autoimmune lymphoproliferative syndrome (TNFRSF6, APT1, FAS, CD95, ALPS1A); Combined immunodeficiency, (IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228), HIV susceptibility or infection (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCKR5 (CCR5)); Immunodeficiencies (CD3E, CD3G,

	AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI); Inflammation (IL-10, IL-1 (IL-1a, IL-1b), IL-13, IL-17 (IL-17a (CTLA8), IL-17b, IL-17c, IL-17d, IL-17f), II-23, Cx3cr1, ptpn22, TNFa, NOD2/CARD15 for IBD, IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3cl1); Severe combined immunodeficiencies (SCIDs)(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).
Metabolic, liver, kidney and protein diseases and disorders	Amyloid neuropathy (TTR, PALB); Amyloidosis (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); Cirrhosis (KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); Cystic fibrosis (CFTR, ABCC7, CF, MRP7); Glycogen storage diseases (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Hepatic adenoma, 142330 (TCF1, HNF1A, MODY3), Hepatic failure, early onset, and neurologic disorder (SCOD1, SCO1), Hepatic lipase deficiency (LIPC), Hepatoblastoma, cancer and carcinomas (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; Medullary cystic kidney disease (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); Phenylketonuria (PAH, PKU1, QDPR, DHPR, PTS); Polycystic kidney and hepatic disease (FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63).
Muscular / Skeletal diseases and disorders	Becker muscular dystrophy (DMD, BMD, MYF6), Duchenne Muscular Dystrophy (DMD, BMD); Emery-Dreifuss muscular dystrophy (LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); Facioscapulohumeral muscular dystrophy (FSHMD1A, FSHD1A); Muscular dystrophy (FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3, LGMD1C, SEPN1, SELN, RSMD1, PLEC1, PLTN, EBS1); Osteopetrosis (LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); Muscular atrophy (VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1).
Neurological and neuronal diseases and disorders	ALS (SOD1, ALS2, STEX, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c); Alzheimer disease (APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLAU, URK, ACE, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); Autism (Mecp2, BZRAP1, MDGA2, Sema5A, Neurexin 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); Fragile X Syndrome (FMR2, FXR1, FXR2, mGLUR5); Huntington's disease and disease like disorders (HD, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); Parkinson disease (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK5, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); Rett syndrome (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Synuclein, DJ-1); Schizophrenia (Neuregulin1 (Nrg1), Erb4 (receptor for Neuregulin), Complexin1 (Cplx1), Tph1 Tryptophan hydroxylase, Tph2, Tryptophan hydroxylase 2, Neurexin 1, GSK3, GSK3a, GSK3b, 5-HTT (Slc6a4), COMT, DRD (Drd1a), SLC6A3, DAOA, DTNBP1, Dao (Dao1)); Secretase Related Disorders (APH-1 (alpha and beta), Presenilin (Psen1), nicastrin, (Ncstn), PEN-2, Nos1, Parp1, Nat1, Nat2); Trinucleotide Repeat Disorders (HTT (Huntington's Dx), SBMA/SMAX1/AR (Kennedy's Dx), FXN/X25 (Friedrich's Ataxia), ATX3 (Machado- Joseph's Dx), ATXN1 and ATXN2 (spinocerebellar ataxias), DMPK (myotonic dystrophy), Atrophin-1 and Atn1 (DRPLA Dx), CBP (Creb-BP - global instability), VLDLR (Alzheimer's), Atxn7, Atxn10).
Occular diseases and disorders	Age-related macular degeneration (Abcr, Ccl2, Cc2, cp (ceruloplasmin), Timp3, cathepsinD, Vldlr, Ccr2); Cataract (CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); Corneal clouding and dystrophy (APOA1, TGFBI, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD); Cornea plana congenital (KERA, CNA2); Glaucoma (MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A); Leber congenital amaurosis (CRB1, RP12, CRX, CORD2, CRD, RPGRIP1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3); Macular dystrophy (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2).

[00822] Table 12 (Cont'd):

CELLULAR FUNCTION	GENES
PI3K/AKT Signaling	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2;
	PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1;
	AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2;
	PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2;
	ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3;
	PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7;
	YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A;
	CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1;
	CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1;
	PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2;

	TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK;
	HSP90AA1; RPS6KB1
ERK/MAPK Signaling	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2;
	EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6;
	MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1;
	PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A;
	PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN;
	EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC;
	CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ;
	PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1;
	PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1;
	CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
Glucocorticoid Receptor	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1;
Signaling	MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I;
	PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2;
	MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1;
	MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13;
	RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1;
	PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3;
	MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP;
	CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2;
	PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1;
	ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1;
	STAT1; IL6; HSP90AA1
Axonal Guidance Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12;
	IGF1; RAC1; RAP1A; EIF4E; PRKCZ; NRP1; NTRK2;
	ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2;
	PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2;
	CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11;
	PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1;
	FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1;
	GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3;
	CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B;
	AKT3; PRKCA
Ephrin Receptor Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1;
	PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2;
	MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2;
	DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14;
	CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1;
	KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2;
	PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10;
	MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2;
	EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4;
	AKT3; SGK
Actin Cytoskeleton	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1;
Signaling	PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6;
	ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8;
	PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8;
	F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD;
	PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7;
	PPP1CC; PXN; VIL2; RAF1; GSN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3;
	ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL;
	BRAF; VAV3; SGK
Huntington's Disease	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2;
Signaling	MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2;
	PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST;
	GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1;
	GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2;
	HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A;
	HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1;
	PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX;
	ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3
Apoptosis Signaling	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1;
	BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB;
	CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8;
	BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF;

	RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2;
	CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2;
	BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK;
	CASP3; BIRC3; PARP1
B Cell Receptor Signaling	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11;
	AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1;
	MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9;
	EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB;
	MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1;
	NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN;
	GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Leukocyte Extravasation	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA;
Signaling	RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11;
	MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12;
	PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB;
	MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK;
	MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2;
	CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK;
	CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
Integrin Signaling	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A;
	TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2;
	CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8;
	CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP;
	RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1;
	TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2;
	CRKL; BRAF; GSK3B; AKT3
Acute Phase Response	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11;
Signaling	AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14;
	PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS;
	MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1;
	TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1;
	IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1;
	CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN;
	AKT3; IL1R1; IL6
PTEN Signaling	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11;
	MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA;
	CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1;
	MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR;
	RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2;
	AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1;
	NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2;
	GSK3B; AKT3; FOXO1; CASP3; RPS6KB1
p53 Signaling	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A;
	BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2;
	PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1;
	PMAIP1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9;
	CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A;
	HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1;
	SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN;
	SNAI2; GSK3B; BAX; AKT3
Aryl Hydrocarbon Receptor	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1;
Signaling	NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1;
	SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1;
	MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1;
	SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF;
	CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1;
	CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1;
	HSP90AA1
Xenobiotic Metabolism	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1;
Signaling	NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A;
	PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1;
	ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD;
	GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL;
	NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1;
	CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1;
	NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1;
	HSP90AA1
SAPK/JNK Signaling	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1;
	GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA;

	FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1;
	GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A;
	TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2;
	PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1;
	CRKL; BRAF; SGK
PPAr/RXR Signaling	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN;
	RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2;
	ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8;
	IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A;
	NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7;
	CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1;
	TGFBR1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1;
	ADIPOQ
NF-KB Signaling	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAF6;
	TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2;
	KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF;
	INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1;
	PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10;
	GSK3B; AKT3; TNFAIP3; IL1R1
Neuregulin Signaling	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1;
	MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI;
	CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS;
	PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2;
	ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3;
	EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL;
	AKT3; PRKCA; HSP90AA1; RPS6KB1
Wnt & Beta catenin	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO;
Signaling	AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A;
	WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK;
	LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1;
	PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1; CCND1;
	GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B;
	AKT3; SOX2
Insulin Receptor Signaling	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1;
	PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3;
	MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1;
	SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN;
	MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1;
	GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK;
	RPS6KB1
IL-6 Signaling	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3;
	MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1;
	MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG;
	RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3;
	MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6
Hepatic Cholestasis	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA;
	RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8;
	PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1;
	TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8;
	CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4;
	JUN; IL1R1; PRKCA; IL6
IGF-1 Signaling	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2;
	PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8;
	IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A;
	YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1;
	PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3;
	FOXO1; SRF; CTGF; RPS6KB1
NRF2-mediated Oxidative	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1;
Stress Response	NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL;
	NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP;
	MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1;
	GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
Hepatic Fibrosis/Hepatic	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF;
Stellate Cell Activation	SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9;
	IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8;
	PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX;
	IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9

PPAR Signaling	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB;
	NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3;
	NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2;
	PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG;
	RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA;
	MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Fc Epsilon RI Signaling	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11;
	AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD;
	MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN;
	MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3;
	VAV3; PRKCA
G-Protein Coupled	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB;
Receptor Signaling	PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB;
	PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1;
	IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK;
	PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3;
	PRKCA
Inositol Phosphate	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6;
Metabolism	MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2;
	PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1;
	MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
PDGF Signaling	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB;
	PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC;
	PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2;
	PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC;
	JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
VEGF Signaling	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF;
	AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3;
	BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN;
	RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN;
	VEGFA; AKT3; FOXO1; PRKCA
Natural Killer Cell Signaling	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11;
	KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB;
	PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6;
	PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1;
	PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
Cell Cycle: G1/S	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC;
Checkpoint Regulation	ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11;
	HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1;
	E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1;
	GSK3B; RBL1; HDAC6
T Cell Receptor Signaling	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS;
	NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN;
	MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10;
	JUN; VAV3
Death Receptor Signaling	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD;
	FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8;
	DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB;
	CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3;
	BIRC3
FGF Signaling	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11;
	AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8;
	MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1;
	AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4;
	AKT3; PRKCA; HGF
GM-CSF Signaling	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A;
	STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3;
	ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3;
	STAT1
Amyotrophic Lateral	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2;
Sclerosis Signaling	PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1;
	PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1;
	APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
JAK/Stat Signaling	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B;
	PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A;
	PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3;

	STAT1
Nicotinate and Nicotinamide	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1;
Metabolism	PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1;
	PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2;
	MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
Chemokine Signaling	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ;
	CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13;
	RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1;
	MAP2K2; MAP2K1; JUN; CCL2; PRKCA
IL-2 Signaling	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS;
	STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2;
	JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
Synaptic Long Term	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS;
Depression	PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3;
	KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA;
	YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Estrogen Receptor	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2;
Signaling	SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1;
	HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP;
	MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Protein Ubiquitination	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4;
Pathway	CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7;
	USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8;
	USP1; VHL; HSP90AA1; BIRC3
IL-10 Signaling	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2;
	MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF;
	IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1;
	JUN; IL1R1; IL6
VDR/RXR Activation	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1;
	NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD;
	RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1;
	LRP5; CEBPB; FOXO1; PRKCA
TGF-beta Signaling	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1;
	FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2;
	SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2;
	MAP2K1; TGFBR1; SMAD4; JUN; SMAD5
Toll-like Receptor Signaling	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13;
	RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK;
	NFKB1; TLR2; JUN
p38 MAPK Signaling	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS;
	CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2;
	MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1;
	SRF; STAT1
Neurotrophin/TRK Signaling	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS;
	PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A;
	RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1;
	CDC42; JUN; ATF4
FXR/RXR Activation	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8;
	APOB; MAPK10; PPARG; MTTP; MAPK9; PPARGC1A;
	TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
Synaptic Long Term	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1;
Potentiation	PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS;
	PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1;
	ATF4; PRKCA
Calcium Signaling	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1;
	CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11;
	HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4;
	HDAC6
EGF Signaling	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1;
	STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Hypoxia Signaling in the	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT;
Cardiovascular System	HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM;
	VEGFA; JUN; ATF4; VHL; HSP90AA1
LPS/IL-1 Mediated Inhibition	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1;
of RXR Function	MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2;
	TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
LXR/RXR Activation	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA;
	NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1;

	SREBF1; IL1R1; CCL2; IL6; MMP9
Amyloid Processing	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2;
	CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1;
	PSEN1; CSNK1A1; GSK3B; AKT3; APP
IL-4 Signaling	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1;
	PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1;
	FRAP1; AKT3; RPS6KB1
Cell Cycle: G2/M DNA	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC;
Damage Checkpoint	CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A;
Regulation	PRKDC; ATM; SFN; CDKN2A
Nitric Oxide Signaling in the	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3;
Cardiovascular System	CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1;
	VEGFA; AKT3; HSP90AA1
Purine Metabolism	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4;
	PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C;
	NT5E; POLD1; NME1
cAMP-mediated Signaling	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3;
	SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Mitochondrial Dysfunction	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9;
	PARK7; PSEN1; PARK2; APP; CASP3
Notch Signaling	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2;
	PSEN1; NOTCH3; NOTCH1; DLL4
Endoplasmic Reticulum	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4;
Stress Pathway	EIF2AK3; CASP3
Pyrimidine Metabolism	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B;
	NT5E; POLD1; NME1
Parkinson's Signaling	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7;
	PARK2; CASP3
Cardiac & Beta Adrenergic	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC;
Signaling	PPP2R5C
Glycolysis/Gluconeogenesis	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Interferon Signaling	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Sonic Hedgehog Signaling	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B
Glycerophospholipid	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Metabolism	
Phospholipid Degradation	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Tryptophan Metabolism	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Lysine Degradation	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Nucleotide Excision Repair	ERCC5; ERCC4; XPA; XPC; ERCC1
Pathway	
Starch and Sucrose	UCHL1; HK2; GCK; GPI; HK1
Metabolism	
Aminosugars Metabolism	NQO1; HK2; GCK; HK1
Arachidonic Acid	PRDX6; GRN; YWHAZ; CYP1B1
Metabolism	
Circadian Rhythm Signaling	CSNK1E; CREB1; ATF4; NR1D1
Coagulation System	BDKRB1; F2R; SERPINE1; F3
Dopamine Receptor	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Signaling	
Glutathione Metabolism	IDH2; GSTP1; ANPEP; IDH1
Glycerolipid Metabolism	ALDH1A1; GPAM; SPHK1; SPHK2
Linoleic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Methionine Metabolism	DNMT1; DNMT3B; AHCY; DNMT3A
Pyruvate Metabolism	GLO1; ALDH1A1; PKM2; LDHA
Arginine and Proline	ALDH1A1; NOS3; NOS2A
Metabolism	
Eicosanoid Signaling	PRDX6; GRN; YWHAZ
Fructose and Mannose	HK2; GCK; HK1
Metabolism	
Galactose Metabolism	HK2; GCK; HK1
Stilbene, Coumarine and	PRDX6; PRDX1; TYR
Lignin Biosynthesis	
Antigen Presentation	CALR; B2M
Pathway	
Biosynthesis of Steroids	NQO1; DHCR7
Butanoate Metabolism	ALDH1A1; NLGN1
Citrate Cycle	IDH2; IDH1
Fatty Acid Metabolism	ALDH1A1; CYP1B1
Glycerophospholipid	PRDX6; CHKA
Metabolism	
Histidine Metabolism	PRMT5; ALDH1A1

Inositol Metabolism	ERO1L; APEX1
Metabolism of Xenobiotics	GSTP1; CYP1B1
by Cytochrome p450	
Methane Metabolism	PRDX6; PRDX1
Phenylalanine Metabolism	PRDX6; PRDX1
Propanoate Metabolism	ALDH1A1; LDHA
Selenoamino Acid	PRMT5; AHCY
Metabolism	
Sphingolipid Metabolism	SPHK1; SPHK2
Aminophosphonate	PRMT5
Metabolism	
Androgen and Estrogen	PRMT5
Metabolism	
Ascorbate and Aldarate	ALDH1A1
Metabolism	
Bile Acid Biosynthesis	ALDH1A1
Cysteine Metabolism	LDHA
Fatty Acid Biosynthesis	FASN
Glutamate Receptor	GNB2L1
Signaling	
NRF2-mediated Oxidative	PRDX1
Stress Response	
Pentose Phosphate	GPI
Pathway	
Pentose and Glucuronate	UCHL1
Interconversions	
Retinol Metabolism	ALDH1A1
Riboflavin Metabolism	TYR
Tyrosine Metabolism	PRMT5, TYR
Ubiquinone Biosynthesis	PRMT5
Valine, Leucine and	ALDH1A1
Isoleucine Degradation	
Glycine, Serine and	CHKA
Threonine Metabolism	
Lysine Degradation	ALDH1A1
Pain/Taste	TRPM5; TRPA1
Pain	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2;
	Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca;
	Prkacb; Prkar1a; Prkar2a
Mitochondrial Function	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Developmental Neurology	BMP-4; Chordin (Chrd); Noggin (Nog); WNT (Wnt2;
	Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b;
	Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenin;
	Dkk-1; Frizzled related proteins; Otx-2; Gbx2; FGF-8;
	Reelin; Dab1; unc-86 (Pou4f1 or Brn3a); Numb; Reln

[00823] Embodiments of the invention also relate to methods and compositions related to knocking out genes, amplifying genes and repairing particular mutations associated with DNA repeat instability and neurological disorders (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Second Edition, Academic Press, Oct 13, 2011 – Medical). Specific aspects of tandem repeat sequences have been found to be responsible for more than twenty human diseases (New insights into repeat instability: role of RNA•DNA hybrids. McIvor EI, Polak U, Napierala M. RNA Biol. 2010 Sep-Oct;7(5):551-8). The present effector protein systems may be harnessed to correct these defects of genomic instability.

[00824] Several further aspects of the invention relate to correcting defects associated with a wide range of genetic diseases which are further described on the website of the National Institutes of Health under the topic subsection Genetic Disorders (website at

health.nih.gov/topic/GeneticDisorders). The genetic brain diseases may include but are not limited to Adrenoleukodystrophy, Agenesis of the Corpus Callosum, Aicardi Syndrome, Alpers' Disease, Alzheimer's Disease, Barth Syndrome, Batten Disease, CADASIL, Cerebellar Degeneration, Fabry's Disease, Gerstmann-Straussler-Scheinker Disease, Huntington's Disease and other Triplet Repeat Disorders, Leigh's Disease, Lesch-Nyhan Syndrome, Menkes Disease, Mitochondrial Myopathies and NINDS Colpocephaly. These diseases are further described on the website of the National Institutes of Health under the subsection Genetic Brain Disorders.

Adoptive Cell Transfer

[00825] The systems herein may be used for adoptive cell transfer treatments. As used herein, "ACT", "adoptive cell therapy" and "adoptive cell transfer" may be used interchangeably. In certain embodiments, Adoptive cell therapy (ACT) can refer to the transfer of cells to a patient with the goal of transferring the functionality and characteristics into the new host by engraftment of the cells (see, e.g., Mettananda et al., Editing an α-globin enhancer in primary human hematopoietic stem cells as a treatment for β-thalassemia, Nat Commun. 2017 Sep 4;8(1):424). As used herein, the term "engraft" or "engraftment" refers to the process of cell incorporation into a tissue of interest *in vivo* through contact with existing cells of the tissue. Adoptive cell therapy (ACT) can refer to the transfer of cells, most commonly immune-derived cells, back into the same patient or into a new recipient host with the goal of transferring the immunologic functionality and characteristics into the new host. If possible, use of autologous cells helps the recipient by minimizing GVHD issues. The adoptive transfer of autologous tumor infiltrating lymphocytes (TIL) (Zacharakis et al., (2018) Nat Med. 2018 Jun;24(6):724-730; Besser et al., (2010) Clin. Cancer Res 16 (9) 2646–55; Dudley et al., (2002) Science 298 (5594): 850–4; and Dudley et al., (2005) Journal of Clinical Oncology 23 (10): 2346–57.) or genetically re-directed peripheral blood mononuclear cells (Johnson et al., (2009) Blood 114 (3): 535–46; and Morgan et al., (2006) Science 314(5796) 126-9) has been used to successfully treat patients with advanced solid tumors, including melanoma, metastatic breast cancer and colorectal carcinoma, as well as patients with CD19-expressing hematologic malignancies (Kalos et al., (2011) Science Translational Medicine 3 (95): 95ra73). In certain embodiments, allogenic cells immune cells are transferred (see, e.g., Ren et al., (2017) Clin Cancer Res 23 (9) 2255-2266). As described further herein, allogenic cells can be edited to reduce alloreactivity and prevent graft-versus-host disease. Thus, use of allogenic cells allows for cells to be obtained from healthy donors and prepared for use in patients as opposed to preparing autologous cells from a patient after diagnosis.

[00826] Aspects of the invention involve the adoptive transfer of immune system cells, such as T cells, specific for selected antigens, such as tumor associated antigens or tumor specific neoantigens (see, e.g., Maus et al., 2014, Adoptive Immunotherapy for Cancer or Viruses, Annual Review of Immunology, Vol. 32: 189-225; Rosenberg and Restifo, 2015, Adoptive cell transfer as personalized immunotherapy for human cancer, Science Vol. 348 no. 6230 pp. 62-68; Restifo et al., 2015, Adoptive immunotherapy for cancer: harnessing the T cell response. Nat. Rev. Immunol. 12(4): 269-281; and Jenson and Riddell, 2014, Design and implementation of adoptive therapy with chimeric antigen receptor-modified T cells. Immunol Rev. 257(1): 127–144; and Rajasagi et al., 2014, Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. Blood. 2014 Jul 17;124(3):453-62).

[00827] In certain embodiments, an antigen (such as a tumor antigen) to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) may be selected from a group consisting of: B cell maturation antigen (BCMA) (see, e.g., Friedman et al., Effective Targeting of Multiple BCMA-Expressing Hematological Malignancies by Anti-BCMA CAR T Cells, Hum Gene Ther. 2018 Mar 8; Berdeja JG, et al. Durable clinical responses in heavily pretreated patients with relapsed/refractory multiple myeloma: updated results from a multicenter study of bb2121 anti-Bcma CAR T cell therapy. Blood. 2017;130:740; and Mouhieddine and Ghobrial, Immunotherapy in Multiple Myeloma: The Era of CAR T Cell Therapy, Hematologist, May-June 2018, Volume 15, issue 3); PSA (prostate-specific antigen); prostate-specific membrane antigen (PSMA); PSCA (Prostate stem cell antigen); Tyrosine-protein kinase transmembrane receptor ROR1; fibroblast activation protein (FAP); Tumor-associated glycoprotein 72 (TAG72); Carcinoembryonic antigen (CEA); Epithelial cell adhesion molecule (EPCAM); Mesothelin; Human Epidermal growth factor Receptor 2 (ERBB2 (Her2/neu)); Prostase; Prostatic acid phosphatase (PAP); elongation factor 2 mutant (ELF2M); Insulin-like growth factor 1 receptor (IGF-1R); gplOO; BCR-ABL (breakpoint cluster region-Abelson); tyrosinase; New York esophageal squamous cell carcinoma 1 (NY-ESO-1); κ-light chain, LAGE (L antigen); MAGE (melanoma antigen); Melanoma-associated antigen 1 (MAGE-A1); MAGE A3; MAGE A6; legumain; Human papillomavirus (HPV) E6; HPV E7; prostein; survivin; PCTA1 (Galectin 8); Melan-A/MART-1; Ras mutant; TRP-1 (tyrosinase related protein 1, or gp75); Tyrosinase-related Protein 2 (TRP2); TRP-2/INT2 (TRP-2/intron 2); RAGE (renal antigen); receptor for advanced glycation end products 1 (RAGE1); Renal ubiquitous 1, 2 (RU1, RU2); intestinal carboxyl esterase (iCE); Heat shock protein 70-2 (HSP70-2) mutant; thyroid stimulating hormone receptor (TSHR); CD123; CD171; CD19;

CD20; CD22; CD26; CD30; CD33; CD44v7/8 (cluster of differentiation 44, exons 7/8); CD53; CD92; CD100; CD148; CD150; CD200; CD261; CD262; CD362; CS-1 (CD2 subset 1, CRACC, SLAMF7, CD319, and 19A24); C-type lectin-like molecule-1 (CLL-1); ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-1)Cer); Tn antigen (Tn Ag); Fms-Like Tyrosine Kinase 3 (FLT3); CD38; CD138; CD44v6; B7H3 (CD276); KIT (CD117); Interleukin-13 receptor subunit alpha-2 (IL-13Ra2); Interleukin 11 receptor alpha (IL-11Ra); prostate stem cell antigen (PSCA); Protease Serine 21 (PRSS21); vascular endothelial growth factor receptor 2 (VEGFR2); Lewis(Y) antigen; CD24; Platelet-derived growth factor receptor beta (PDGFR-beta); stage-specific embryonic antigen-4 (SSEA-4); Mucin 1, cell surface associated (MUC1); mucin 16 (MUC16); epidermal growth factor receptor (EGFR); epidermal growth factor receptor variant III (EGFRvIII); neural cell adhesion molecule (NCAM); carbonic anhydrase IX (CAIX); Proteasome (Prosome, Macropain) Subunit, Beta Type, 9 (LMP2); ephrin type-A receptor 2 (EphA2); Ephrin B2; Fucosyl GM1; sialyl Lewis adhesion molecule (sLe); ganglioside GM3 (aNeu5Ac(2-3)bDGalp(1-4)bDGlcp(1-1)Cer); TGS5; high molecular weight-melanoma-associated antigen (HMWMAA); o-acetyl-GD2 ganglioside (OAcGD2); Folate receptor alpha; Folate receptor beta; tumor endothelial marker 1 (TEM1/CD248); tumor endothelial marker 7-related (TEM7R); claudin 6 (CLDN6); G protein-coupled receptor class C group 5, member D (GPRC5D); chromosome X open reading frame 61 (CXORF61); CD97; CD179a; anaplastic lymphoma kinase (ALK); Polysialic acid; placenta-specific 1 (PLAC1); hexasaccharide portion of globoH glycoceramide (GloboH); mammary gland differentiation antigen (NY-BR-1); uroplakin 2 (UPK2); Hepatitis A virus cellular receptor 1 (HAVCR1); adrenoceptor beta 3 (ADRB3); pannexin 3 (PANX3); G protein-coupled receptor 20 (GPR20); lymphocyte antigen 6 complex, locus K 9 (LY6K); Olfactory receptor 51E2 (OR51E2); TCR Gamma Alternate Reading Frame Protein (TARP); Wilms tumor protein (WT1); ETS translocation-variant gene 6, located on chromosome 12p (ETV6-AML); sperm protein 17 (SPA17); X Antigen Family, Member 1A (XAGE1); angiopoietin-binding cell surface receptor 2 (Tie 2); CT (cancer/testis (antigen)); melanoma cancer testis antigen-1 (MAD-CT-1); melanoma cancer testis antigen-2 (MAD-CT-2); Fos-related antigen 1; p53; p53 mutant; human Telomerase reverse transcriptase (hTERT); sarcoma translocation breakpoints; melanoma inhibitor of apoptosis (ML-IAP); ERG (transmembrane protease, serine 2 (TMPRSS2) ETS fusion gene); N-Acetyl glucosaminyl-transferase V (NA17); paired box protein Pax-3 (PAX3); Androgen receptor; Cyclin B1; Cyclin D1; v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN); Ras Homolog Family Member C (RhoC); Cytochrome P450 1B1 (CYP1B1); CCCTC-Binding

Factor (Zinc Finger Protein)-Like (BORIS); Squamous Cell Carcinoma Antigen Recognized By T Cells-1 or 3 (SART1, SART3); Paired box protein Pax-5 (PAX5); proacrosin binding protein sp32 (OY-TES1); lymphocyte-specific protein tyrosine kinase (LCK); A kinase anchor protein 4 (AKAP-4); synovial sarcoma, X breakpoint-1, -2, -3 or -4 (SSX1, SSX2, SSX3, SSX4); CD79a; CD79b; CD72; Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1); Fc fragment of IgA receptor (FCAR); Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2); CD300 molecule-like family member f (CD300LF); C-type lectin domain family 12 member A (CLEC12A); bone marrow stromal cell antigen 2 (BST2); EGF-like module-containing mucin-like hormone receptor-like 2 (EMR2); lymphocyte antigen 75 (LY75); Glypican-3 (GPC3); Fc receptor-like 5 (FCRL5); mouse double minute 2 homolog (MDM2); livin; alphafetoprotein (AFP); transmembrane activator and CAML Interactor (TACI); B-cell activating factor receptor (BAFF-R); V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolog (KRAS); immunoglobulin lambda-like polypeptide 1 (IGLL1); 707-AP (707 alanine proline); ART-4 (adenocarcinoma antigen recognized by T4 cells); BAGE (B antigen; b-catenin/m, b-catenin/mutated); CAMEL (CTL-recognized antigen on melanoma); CAP1 (carcinoembryonic antigen peptide 1); CASP-8 (caspase-8); CDC27m (cell-division cycle 27 mutated); CDK4/m (cycline-dependent kinase 4 mutated); Cyp-B (cyclophilin B); DAM (differentiation antigen melanoma); EGP-2 (epithelial glycoprotein 2); EGP-40 (epithelial glycoprotein 40); Erbb2, 3, 4 (erythroblastic leukemia viral oncogene homolog-2, -3, 4); FBP (folate binding protein); , fAchR (Fetal acetylcholine receptor); G250 (glycoprotein 250); GAGE (G antigen); GnT-V (N-acetylglucosaminyltransferase V); HAGE (helicose antigen); ULA-A (human leukocyte antigen-A); HST2 (human signet ring tumor 2); KIAA0205; KDR (kinase insert domain receptor); LDLR/FUT (low density lipid receptor/GDP L-fucose: b-D-galactosidase 2-a-L fucosyltransferase); L1CAM (L1 cell adhesion molecule); MC1R (melanocortin 1 receptor); Myosin/m (myosin mutated); MUM-1, -2, -3 (melanoma ubiquitous mutated 1, 2, 3); NA88-A (NA cDNA clone of patient M88); KG2D (Natural killer group 2, member D) ligands; oncofetal antigen (h5T4); p190 minor bcr-abl (protein of 190KD bcr-abl); Pml/RARa (promyelocytic leukaemia/retinoic acid receptor a); PRAME (preferentially expressed antigen of melanoma); SAGE (sarcoma antigen); TEL/AML1 (translocation Ets-family leukemia/acute myeloid leukemia 1); TPI/m (triosephosphate isomerase mutated); CD70; and any combination thereof.

[00828] In certain embodiments, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a tumor-specific antigen (TSA).

[00829] In certain embodiments, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a neoantigen.

[00830] In certain embodiments, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a tumor-associated antigen (TAA).

[00831] In certain embodiments, an antigen to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) is a universal tumor antigen. In certain preferred embodiments, the universal tumor antigen is selected from the group consisting of: a human telomerase reverse transcriptase (hTERT), survivin, mouse double minute 2 homolog (MDM2), cytochrome P450 1B 1 (CYP1B), HER2/neu, Wilms' tumor gene 1 (WT1), livin, alphafetoprotein (AFP), carcinoembryonic antigen (CEA), mucin 16 (MUC16), MUC1, prostate-specific membrane antigen (PSMA), p53, cyclin (Dl), and any combinations thereof.

[00832] In certain embodiments, an antigen (such as a tumor antigen) to be targeted in adoptive cell therapy (such as particularly CAR or TCR T-cell therapy) of a disease (such as particularly of tumor or cancer) may be selected from a group consisting of: CD19, BCMA, CD70, CLL-1, MAGE A3, MAGE A6, HPV E6, HPV E7, WT1, CD22, CD171, ROR1, MUC16, and SSX2. In certain preferred embodiments, the antigen may be CD19. For example, CD19 may be targeted in hematologic malignancies, such as in lymphomas, more particularly in B-cell lymphomas, such as without limitation in diffuse large B-cell lymphoma, primary mediastinal b-cell lymphoma, transformed follicular lymphoma, marginal zone lymphoma, mantle cell lymphoma, acute lymphoblastic leukemia including adult and pediatric ALL, non-Hodgkin lymphoma, indolent non-Hodgkin lymphoma, or chronic lymphocytic leukemia. For example, BCMA may be targeted in multiple myeloma or plasma cell leukemia (see, e.g., 2018 American Association for Cancer Research (AACR) Annual meeting Poster: Allogeneic Chimeric Antigen Receptor T Cells Targeting B Cell Maturation Antigen). For example, CLL1 may be targeted in acute myeloid leukemia. For example, MAGE A3, MAGE A6, SSX2, and/or KRAS may be targeted in solid tumors. For example, HPV E6 and/or HPV E7 may be targeted in cervical cancer or head and neck cancer. For example, WT1 may be targeted in acute myeloid leukemia (AML), myelodysplastic syndromes (MDS), chronic myeloid leukemia (CML), non-small cell lung cancer, breast, pancreatic, ovarian or colorectal cancers, or mesothelioma. For example, CD22 may be targeted in B cell malignancies, including non-Hodgkin lymphoma, diffuse large B-cell lymphoma, or acute lymphoblastic leukemia. For

example, CD171 may be targeted in neuroblastoma, glioblastoma, or lung, pancreatic, or ovarian cancers. For example, ROR1 may be targeted in ROR1+ malignancies, including non-small cell lung cancer, triple negative breast cancer, pancreatic cancer, prostate cancer, ALL, chronic lymphocytic leukemia, or mantle cell lymphoma. For example, MUC16 may be targeted in MUC16ecto+ epithelial ovarian, fallopian tube or primary peritoneal cancer. For example, CD70 may be targeted in both hematologic malignancies as well as in solid cancers such as renal cell carcinoma (RCC), gliomas (e.g., GBM), and head and neck cancers (HNSCC). CD70 is expressed in both hematologic malignancies as well as in solid cancers, while its expression in normal tissues is restricted to a subset of lymphoid cell types (see, e.g., 2018 American Association for Cancer Research (AACR) Annual meeting Poster: Allogeneic CRISPR Engineered Anti-CD70 CAR-T Cells Demonstrate Potent Preclinical Activity Against Both Solid and Hematological Cancer Cells).

[00833] Various strategies may for example be employed to genetically modify T cells by altering the specificity of the T cell receptor (TCR) for example by introducing new TCR α and β chains with selected peptide specificity (see U.S. Patent No. 8,697,854; PCT Patent Publications: WO2003020763, WO2004033685, WO2004044004, WO2005114215, WO2006000830, WO2008038002, WO2008039818, WO2004074322, WO2005113595, WO2006125962, WO2013166321, WO2013039889, WO2014018863, WO2014083173; U.S. Patent No. 8,088,379).

[00834] As an alternative to, or addition to, TCR modifications, chimeric antigen receptors (CARs) may be used in order to generate immunoresponsive cells, such as T cells, specific for selected targets, such as malignant cells, with a wide variety of receptor chimera constructs having been described (see U.S. Patent Nos. 5,843,728; 5,851,828; 5,912,170; 6,004,811; 6,284,240; 6,392,013; 6,410,014; 6,753,162; 8,211,422; and, PCT Publication WO9215322).

[00835] In general, CARs are comprised of an extracellular domain, a transmembrane domain, and an intracellular domain, wherein the extracellular domain comprises an antigen-binding domain that is specific for a predetermined target. While the antigen-binding domain of a CAR is often an antibody or antibody fragment (e.g., a single chain variable fragment, scFv), the binding domain is not particularly limited so long as it results in specific recognition of a target. For example, in some embodiments, the antigen-binding domain may comprise a receptor, such that the CAR is capable of binding to the ligand of the receptor. Alternatively, the antigen-binding domain may comprise a ligand, such that the CAR is capable of binding the endogenous receptor of that ligand.

[00836] The antigen-binding domain of a CAR is generally separated from the transmembrane domain by a hinge or spacer. The spacer is also not particularly limited, and it is designed to provide the CAR with flexibility. For example, a spacer domain may comprise a portion of a human Fc domain, including a portion of the CH3 domain, or the hinge region of any immunoglobulin, such as IgA, IgD, IgE, IgG, or IgM, or variants thereof. Furthermore, the hinge region may be modified so as to prevent off-target binding by FcRs or other potential interfering objects. For example, the hinge may comprise an IgG4 Fc domain with or without a S228P, L235E, and/or N297Q mutation (according to Kabat numbering) in order to decrease binding to FcRs. Additional spacers/hinges include, but are not limited to, CD4, CD8, and CD28 hinge regions.

[00837] The transmembrane domain of a CAR may be derived either from a natural or from a synthetic source. Where the source is natural, the domain may be derived from any membrane bound or transmembrane protein. Transmembrane regions of particular use in this disclosure may be derived from CD8, CD28, CD3, CD45, CD4, CD5, CDS, CD9, CD 16, CD22, CD33, CD37, CD64, CD80, CD86, CD 134, CD137, CD 154, TCR. Alternatively, the transmembrane domain may be synthetic, in which case it will comprise predominantly hydrophobic residues such as leucine and valine. Preferably a triplet of phenylalanine, tryptophan and valine will be found at each end of a synthetic transmembrane domain. Optionally, a short oligo- or polypeptide linker, preferably between 2 and 10 amino acids in length may form the linkage between the transmembrane domain and the cytoplasmic signaling domain of the CAR. A glycine-serine doublet provides a particularly suitable linker.

[00838] Alternative CAR constructs may be characterized as belonging to successive generations. First-generation CARs typically consist of a single-chain variable fragment of an antibody specific for an antigen, for example comprising a VL linked to a VH of a specific antibody, linked by a flexible linker, for example by a CD8α hinge domain and a CD8α transmembrane domain, to the transmembrane and intracellular signaling domains of either CD3ζ or FcRγ (scFv-CD3ζ or scFv-FcRγ; see U.S. Patent No. 7,741,465; U.S. Patent No. 5,912,172; U.S. Patent No. 5,906,936). Second-generation CARs incorporate the intracellular domains of one or more costimulatory molecules, such as CD28, OX40 (CD134), or 4-1BB (CD137) within the endodomain (for example scFv-CD28/OX40/4-1BB-CD3ζ; see U.S. Patent Nos. 8,911,993; 8,916,381; 8,975,071; 9,101,584; 9,102,760; 9,102,761). Third-generation CARs include a combination of costimulatory endodomains, such a CD3ζ-chain, CD97, GDI la-CD18, CD2, ICOS, CD27, CD154, CDS, OX40, 4-1BB, CD2, CD7, LIGHT, LFA-1, NKG2C, B7-H3, CD30, CD40, PD-1, or CD28 signaling domains (for example scFv-

CD28-4-1BB-CD3ζ or scFv-CD28-OX40-CD3ζ; see U.S. Patent No. 8,906,682; U.S. Patent No. 8,399,645; U.S. Pat. No. 5,686,281; PCT Publication No. WO2014134165; PCT Publication No. WO2012079000). In certain embodiments, the primary signaling domain comprises a functional signaling domain of a protein selected from the group consisting of CD3 zeta, CD3 gamma, CD3 delta, CD3 epsilon, common FcR gamma (FCERIG), FcR beta (Fc Epsilon R1b), CD79a, CD79b, Fc gamma RIIa, DAP10, and DAP12. In certain preferred embodiments, the primary signaling domain comprises a functional signaling domain of CD3ζ or FcRγ. In certain embodiments, the one or more costimulatory signaling domains comprise a functional signaling domain of a protein selected, each independently, from the group consisting of: CD27, CD28, 4-1BB (CD137), OX40, CD30, CD40, PD-1, ICOS, lymphocyte function-associated antigen-1 (LFA-1), CD2, CD7, LIGHT, NKG2C, B7-H3, a ligand that specifically binds with CD83, CDS, ICAM-1, GITR, BAFFR, HVEM (LIGHTR), SLAMF7, NKp80 (KLRF1), CD160, CD19, CD4, CD8 alpha, CD8 beta, IL2R beta, IL2R gamma, IL7R alpha, ITGA4, VLA1, CD49a, ITGA4, IA4, CD49D, ITGA6, VLA-6, CD49f, ITGAD, CD11d, ITGAE, CD103, ITGAL, CD11a, LFA-1, ITGAM, CD11b, ITGAX, CD11c, ITGB1, CD29, ITGB2, CD18, ITGB7, TNFR2, TRANCE/RANKL, DNAM1 (CD226), SLAMF4 (CD244, 2B4), CD84, CD96 (Tactile), CEACAM1, CRTAM, Ly9 (CD229), CD160 (BY55), PSGL1, CD100 (SEMA4D), CD69, SLAMF6 (NTB-A, Lyl08), SLAM (SLAMF1, CD150, IPO-3), BLAME (SLAMF8), SELPLG (CD162), LTBR, LAT, GADS, SLP-76, PAG/Cbp, NKp44, NKp30, NKp46, and NKG2D. In certain embodiments, the one or more costimulatory signaling domains comprise a functional signaling domain of a protein selected, each independently, from the group consisting of: 4-1BB, CD27, and CD28. In certain embodiments, a chimeric antigen receptor may have the design as described in U.S. Patent No. 7,446,190, comprising an intracellular domain of CD3ζ chain (such as amino acid residues 52-163 of the human CD3 zeta chain, as shown in SEQ ID NO: 14 of US 7,446,190), a signaling region from CD28 and an antigen-binding element (or portion or domain; such as scFv). The CD28 portion, when between the zeta chain portion and the antigen-binding element, may suitably include the transmembrane and signaling domains of CD28 (such as amino acid residues 114-220 of SEQ ID NO: 10, full sequence shown in SEQ ID NO: 6 of US 7,446,190; these can include the following portion of CD28 as set forth in Genbank identifier NM_006139 (sequence version 1, 2 or 3): IEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWVLVVVGGVLACYSLLVT VAFIIFWVRSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRS)) (SEQ ID NO:453). Alternatively, when the zeta sequence lies between the CD28 sequence and the

antigen-binding element, intracellular domain of CD28 can be used alone (such as amino sequence set forth in SEQ ID NO: 9 of US 7,446,190). Hence, certain embodiments employ a CAR comprising (a) a zeta chain portion comprising the intracellular domain of human CD3ζ chain, (b) a costimulatory signaling region, and (c) an antigen-binding element (or portion or domain), wherein the costimulatory signaling region comprises the amino acid sequence encoded by SEQ ID NO: 6 of US 7,446,190.

[00839] Alternatively, costimulation may be orchestrated by expressing CARs in antigen-specific T cells, chosen so as to be activated and expanded following engagement of their native αβTCR, for example by antigen on professional antigen-presenting cells, with attendant costimulation. In addition, additional engineered receptors may be provided on the immunoresponsive cells, for example to improve targeting of a T-cell attack and/or minimize side effects

[00840] By means of an example and without limitation, Kochenderfer et al., (2009) J Immunother. 32 (7): 689-702 described anti-CD19 chimeric antigen receptors (CAR). FMC63-28Z CAR contained a single chain variable region moiety (scFv) recognizing CD19 derived from the FMC63 mouse hybridoma (described in Nicholson et al., (1997) Molecular Immunology 34: 1157–1165), a portion of the human CD28 molecule, and the intracellular component of the human TCR-ζ molecule. FMC63-CD828BBZ CAR contained the FMC63 scFv, the hinge and transmembrane regions of the CD8 molecule, the cytoplasmic portions of CD28 and 4-1BB, and the cytoplasmic component of the TCR-ζ molecule. The exact sequence of the CD28 molecule included in the FMC63-28Z CAR corresponded to Genbank identifier NM_006139; the sequence included all amino acids starting with the amino acid sequence IEVMYPPPY (SEQ ID NO:454) and continuing all the way to the carboxy-terminus of the protein. To encode the anti-CD19 scFv component of the vector, the authors designed a DNA sequence which was based on a portion of a previously published CAR (Cooper et al., (2003) Blood 101: 1637–1644). This sequence encoded the following components in frame from the 5' end to the 3' end: an XhoI site, the human granulocyte-macrophage colony-stimulating factor (GM-CSF) receptor α-chain signal sequence, the FMC63 light chain variable region (as in Nicholson et al., *supra*), a linker peptide (as in Cooper et al., *supra*), the FMC63 heavy chain variable region (as in Nicholson et al., *supra*), and a NotI site. A plasmid encoding this sequence was digested with XhoI and NotI. To form the MSGV-FMC63-28Z retroviral vector, the XhoI and NotI-digested fragment encoding the FMC63 scFv was ligated into a second XhoI and NotI-digested fragment that encoded the MSGV retroviral backbone (as in Hughes et al., (2005) Human Gene Therapy 16: 457–472) as well as part of the extracellular portion of human

CD28, the entire transmembrane and cytoplasmic portion of human CD28, and the cytoplasmic portion of the human TCR-ζ molecule (as in Maher et al., 2002) Nature Biotechnology 20: 70–75). The FMC63-28Z CAR is included in the KTE-C19 (axicabtagene ciloleucel) anti-CD19 CAR-T therapy product in development by Kite Pharma, Inc. for the treatment of *inter alia* patients with relapsed/refractory aggressive B-cell non-Hodgkin lymphoma (NHL). Accordingly, in certain embodiments, cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may express the FMC63-28Z CAR as described by Kochenderfer et al. (*supra*). Hence, in certain embodiments, cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may comprise a CAR comprising an extracellular antigen-binding element (or portion or domain; such as scFv) that specifically binds to an antigen, an intracellular signaling domain comprising an intracellular domain of a CD3ζ chain, and a costimulatory signaling region comprising a signaling domain of CD28. Preferably, the CD28 amino acid sequence is as set forth in Genbank identifier NM_006139 (sequence version 1, 2 or 3) starting with the amino acid sequence IEVMYPPPY and continuing all the way to the carboxy-terminus of the protein. The sequence is reproduced herein: IEVMYPPPYLDNEKSNGTIIHVKGKHLCPSPLFPGPSKPFWVLVVVGGVLACYSLLVT VAFIIFWVRSKRSRLLHSDYMNMTPRRPGPTRKHYQPYAPPRDFAAYRS. Preferably, the antigen is CD19, more preferably the antigen-binding element is an anti-CD19 scFv, even more preferably the anti-CD19 scFv as described by Kochenderfer et al. (*supra*).

[00841] Additional anti-CD19 CARs are further described in WO2015187528. More particularly Example 1 and Table 1 of WO2015187528, incorporated by reference herein, demonstrate the generation of anti-CD19 CARs based on a fully human anti-CD19 monoclonal antibody (47G4, as described in US20100104509) and murine anti-CD19 monoclonal antibody (as described in Nicholson et al. and explained above). Various combinations of a signal sequence (human CD8-alpha or GM-CSF receptor), extracellular and transmembrane regions (human CD8-alpha) and intracellular T-cell signalling domains (CD28-CD3ζ; 4-1BB-CD3ζ; CD27-CD3ζ; CD28-CD27-CD3ζ, 4-1BB-CD27-CD3ζ; CD27-4-1BB-CD3ζ; CD28-CD27-FcεRI gamma chain; or CD28-FcεRI gamma chain) were disclosed. Hence, in certain embodiments, cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may comprise a CAR comprising an extracellular antigen-binding element that specifically binds to an antigen, an extracellular and transmembrane region as set forth in Table 1 of WO2015187528 and an intracellular T-cell signalling domain as set forth in Table

1 of WO2015187528. Preferably, the antigen is CD19, more preferably the antigen-binding element is an anti-CD19 scFv, even more preferably the mouse or human anti-CD19 scFv as described in Example 1 of WO2015187528. In certain embodiments, the CAR comprises, consists essentially of or consists of an amino acid sequence of SEQ ID NO: 1, SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4, SEQ ID NO: 5, SEQ ID NO: 6, SEQ ID NO: 7, SEQ ID NO: 8, SEQ ID NO: 9, SEQ ID NO: 10, SEQ ID NO: 11, SEQ ID NO: 12, or SEQ ID NO: 13 as set forth in Table 1 of WO2015187528.

[00842] By means of an example and without limitation, chimeric antigen receptor that recognizes the CD70 antigen is described in WO2012058460A2 (see also, Park et al., CD70 as a target for chimeric antigen receptor T cells in head and neck squamous cell carcinoma, Oral Oncol. 2018 Mar;78:145-150; and Jin et al., CD70, a novel target of CAR T-cell therapy for gliomas, Neuro Oncol. 2018 Jan 10;20(1):55-65). CD70 is expressed by diffuse large B-cell and follicular lymphoma and also by the malignant cells of Hodgkins lymphoma, Waldenstrom's macroglobulinemia and multiple myeloma, and by HTLV-1- and EBV-associated malignancies. (Agathanggelou et al. Am.J.Pathol. 1995;147: 1152-1160; Hunter et al., Blood 2004; 104:4881. 26; Lens et al., J Immunol. 2005;174:6212-6219; Baba et al., J Virol. 2008;82:3843-3852.) In addition, CD70 is expressed by non-hematological malignancies such as renal cell carcinoma and glioblastoma. (Junker et al., J Urol. 2005;173:2150-2153; Chahlavi et al., Cancer Res 2005;65:5428-5438) Physiologically, CD70 expression is transient and restricted to a subset of highly activated T, B, and dendritic cells.

[00843] By means of an example and without limitation, chimeric antigen receptor that recognizes BCMA has been described (see, e.g., US20160046724A1; WO2016014789A2; WO2017211900A1; WO2015158671A1; US20180085444A1; WO2018028647A1; US20170283504A1; and WO2013154760A1).

[00844] In certain embodiments, the immune cell may, in addition to a CAR or exogenous TCR as described herein, further comprise a chimeric inhibitory receptor (inhibitory CAR) that specifically binds to a second target antigen and is capable of inducing an inhibitory or immunosuppressive or repressive signal to the cell upon recognition of the second target antigen. In certain embodiments, the chimeric inhibitory receptor comprises an extracellular antigen-binding element (or portion or domain) configured to specifically bind to a target antigen, a transmembrane domain, and an intracellular immunosuppressive or repressive signaling domain. In certain embodiments, the second target antigen is an antigen that is not expressed on the surface of a cancer cell or infected cell or the expression of which is downregulated on a cancer cell or an infected cell. In certain embodiments, the second target

antigen is an MHC-class I molecule. In certain embodiments, the intracellular signaling domain comprises a functional signaling portion of an immune checkpoint molecule, such as for example PD-1 or CTLA4. Advantageously, the inclusion of such inhibitory CAR reduces the chance of the engineered immune cells attacking non-target (e.g., non-cancer) tissues.

[00845] Alternatively, T-cells expressing CARs may be further modified to reduce or eliminate expression of endogenous TCRs in order to reduce off-target effects. Reduction or elimination of endogenous TCRs can reduce off-target effects and increase the effectiveness of the T cells (U.S. 9,181,527). T cells stably lacking expression of a functional TCR may be produced using a variety of approaches. T cells internalize, sort, and degrade the entire T cell receptor as a complex, with a half-life of about 10 hours in resting T cells and 3 hours in stimulated T cells (von Essen, M. et al. 2004. J. Immunol. 173:384-393). Proper functioning of the TCR complex requires the proper stoichiometric ratio of the proteins that compose the TCR complex. TCR function also requires two functioning TCR zeta proteins with ITAM motifs. The activation of the TCR upon engagement of its MHC-peptide ligand requires the engagement of several TCRs on the same T cell, which all must signal properly. Thus, if a TCR complex is destabilized with proteins that do not associate properly or cannot signal optimally, the T cell will not become activated sufficiently to begin a cellular response.

[00846] Accordingly, in some embodiments, TCR expression may eliminated using RNA interference (e.g., shRNA, siRNA, miRNA, etc.), CRISPR, or other methods that target the nucleic acids encoding specific TCRs (e.g., TCR-α and TCR-β) and/or CD3 chains in primary T cells. By blocking expression of one or more of these proteins, the T cell will no longer produce one or more of the key components of the TCR complex, thereby destabilizing the TCR complex and preventing cell surface expression of a functional TCR.

[00847] In some instances, CAR may also comprise a switch mechanism for controlling expression and/or activation of the CAR. For example, a CAR may comprise an extracellular, transmembrane, and intracellular domain, in which the extracellular domain comprises a target-specific binding element that comprises a label, binding domain, or tag that is specific for a molecule other than the target antigen that is expressed on or by a target cell. In such embodiments, the specificity of the CAR is provided by a second construct that comprises a target antigen binding domain (e.g., an scFv or a bispecific antibody that is specific for both the target antigen and the label or tag on the CAR) and a domain that is recognized by or binds to the label, binding domain, or tag on the CAR. See, e.g., WO 2013/044225, WO 2016/000304, WO 2015/057834, WO 2015/057852, WO 2016/070061, US 9,233,125, US 2016/0129109. In this way, a T-cell that expresses the CAR can be administered to a subject,

but the CAR cannot bind its target antigen until the second composition comprising an antigen-specific binding domain is administered.

[00848] Alternative switch mechanisms include CARs that require multimerization in order to activate their signaling function (see, e.g., US 2015/0368342, US 2016/0175359, US 2015/0368360) and/or an exogenous signal, such as a small molecule drug (US 2016/0166613, Yung et al., Science, 2015), in order to elicit a T-cell response. Some CARs may also comprise a "suicide switch" to induce cell death of the CAR T-cells following treatment (Buddee et al., PLoS One, 2013) or to downregulate expression of the CAR following binding to the target antigen (WO 2016/011210).

[00849] Alternative techniques may be used to transform target immunoresponsive cells, such as protoplast fusion, lipofection, transfection or electroporation. A wide variety of vectors may be used, such as retroviral vectors, lentiviral vectors, adenoviral vectors, adeno-associated viral vectors, plasmids or transposons, such as a Sleeping Beauty transposon (see U.S. Patent Nos. 6,489,458; 7,148,203; 7,160,682; 7,985,739; 8,227,432), may be used to introduce CARs, for example using 2nd generation antigen-specific CARs signaling through CD3ζ and either CD28 or CD137. Viral vectors may for example include vectors based on HIV, SV40, EBV, HSV or BPV.

[00850] Cells that are targeted for transformation may for example include T cells, Natural Killer (NK) cells, cytotoxic T lymphocytes (CTL), regulatory T cells, human embryonic stem cells, tumor-infiltrating lymphocytes (TIL) or a pluripotent stem cell from which lymphoid cells may be differentiated. T cells expressing a desired CAR may for example be selected through co-culture with γ-irradiated activating and propagating cells (AaPC), which co-express the cancer antigen and co-stimulatory molecules. The engineered CAR T-cells may be expanded, for example by co-culture on AaPC in presence of soluble factors, such as IL-2 and IL-21. This expansion may for example be carried out so as to provide memory CAR+ T cells (which may for example be assayed by non-enzymatic digital array and/or multi-panel flow cytometry). In this way, CAR T cells may be provided that have specific cytotoxic activity against antigen-bearing tumors (optionally in conjunction with production of desired chemokines such as interferon-γ). CAR T cells of this kind may for example be used in animal models, for example to treat tumor xenografts.

[00851] In certain embodiments, ACT includes co-transferring CD4+ Th1 cells and CD8+ CTLs to induce a synergistic antitumor response (see, e.g., Li et al., Adoptive cell therapy with CD4+ T helper 1 cells and CD8+ cytotoxic T cells enhances complete rejection of an

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

[00852] In certain embodiments, Th17 cells are transferred to a subject in need thereof. Th17 cells have been reported to directly eradicate melanoma tumors in mice to a greater extent than Th1 cells (Muranski P, et al., Tumor-specific Th17-polarized cells eradicate large established melanoma. Blood. 2008 Jul 15; 112(2):362-73; and Martin-Orozco N, et al., T helper 17 cells promote cytotoxic T cell activation in tumor immunity. Immunity. 2009 Nov 20; 31(5):787-98). Those studies involved an adoptive T cell transfer (ACT) therapy approach, which takes advantage of $CD4^+$ T cells that express a TCR recognizing tyrosinase tumor antigen. Exploitation of the TCR leads to rapid expansion of Th17 populations to large numbers *ex vivo* for reinfusion into the autologous tumor-bearing hosts.

[00853] In certain embodiments, ACT may include autologous iPSC-based vaccines, such as irradiated iPSCs in autologous anti-tumor vaccines (see e.g., Kooreman, Nigel G. et al., Autologous iPSC-Based Vaccines Elicit Anti-tumor Responses In Vivo, Cell Stem Cell 22, 1–13, 2018, doi.org/10.1016/j.stem.2018.01.016).

[00854] Unlike T-cell receptors (TCRs) that are MHC restricted, CARs can potentially bind any cell surface-expressed antigen and can thus be more universally used to treat patients (see Irving et al., Engineering Chimeric Antigen Receptor T-Cells for Racing in Solid Tumors: Don't Forget the Fuel, Front. Immunol., 03 April 2017, doi.org/10.3389/fimmu.2017.00267). In certain embodiments, in the absence of endogenous T-cell infiltrate (e.g., due to aberrant antigen processing and presentation), which precludes the use of TIL therapy and immune checkpoint blockade, the transfer of CAR T-cells may be used to treat patients (see, e.g., Hinrichs CS, Rosenberg SA. Exploiting the curative potential of adoptive T-cell therapy for cancer. Immunol Rev (2014) 257(1):56–71. doi:10.1111/ imr.12132).

[00855] Approaches such as the foregoing may be adapted to provide methods of treating and/or increasing survival of a subject having a disease, such as a neoplasia, for example by administering an effective amount of an immunoresponsive cell comprising an antigen recognizing receptor that binds a selected antigen, wherein the binding activates the immunoresponsive cell, thereby treating or preventing the disease (such as a neoplasia, a pathogen infection, an autoimmune disorder, or an allogeneic transplant reaction).

[00856] In certain embodiments, the treatment can be administered after lymphodepleting pretreatment in the form of chemotherapy (typically a combination of cyclophosphamide and fludarabine) or radiation therapy. Initial studies in ACT had short lived responses and the transferred cells did not persist in vivo for very long (Houot et al., T-cell-based

immunotherapy: adoptive cell transfer and checkpoint inhibition. Cancer Immunol Res (2015) 3(10):1115–22; and Kamta et al., Advancing Cancer Therapy with Present and Emerging Immuno-Oncology Approaches. Front. Oncol. (2017) 7:64). Immune suppressor cells like Tregs and MDSCs may attenuate the activity of transferred cells by outcompeting them for the necessary cytokines. Not being bound by a theory lymphodepleting pretreatment may eliminate the suppressor cells allowing the TILs to persist.

[00857] In one embodiment, the treatment can be administrated into patients undergoing an immunosuppressive treatment (e.g., glucocorticoid treatment). The cells or population of cells, may be made resistant to at least one immunosuppressive agent due to the inactivation of a gene encoding a receptor for such immunosuppressive agent. In certain embodiments, the immunosuppressive treatment provides for the selection and expansion of the immunoresponsive T cells within the patient.

[00858] In certain embodiments, the treatment can be administered before primary treatment (e.g., surgery or radiation therapy) to shrink a tumor before the primary treatment. In another embodiment, the treatment can be administered after primary treatment to remove any remaining cancer cells.

[00859] In certain embodiments, immunometabolic barriers can be targeted therapeutically prior to and/or during ACT to enhance responses to ACT or CAR T-cell therapy and to support endogenous immunity (see, e.g., Irving et al., Engineering Chimeric Antigen Receptor T-Cells for Racing in Solid Tumors: Don't Forget the Fuel, Front. Immunol., 03 April 2017, doi.org/10.3389/fimmu.2017.00267).

[00860] The administration of cells or population of cells, such as immune system cells or cell populations, such as more particularly immunoresponsive cells or cell populations, as disclosed herein may be carried out in any convenient manner, including by aerosol inhalation, injection, ingestion, transfusion, implantation or transplantation. The cells or population of cells may be administered to a patient subcutaneously, intradermally, intratumorally, intranodally, intramedullary, intramuscularly, intrathecally, by intravenous or intralymphatic injection, or intraperitoneally. In some embodiments, the disclosed CARs may be delivered or administered into a cavity formed by the resection of tumor tissue (i.e. intracavity delivery) or directly into a tumor prior to resection (intratumoral delivery). In one embodiment, the cell compositions of the present invention are preferably administered by intravenous injection.

[00861] The administration of the cells or population of cells can consist of the administration of 10^4- 10^9 cells per kg body weight, preferably 10^5 to 10^6 cells/kg body weight including all integer values of cell numbers within those ranges. Dosing in CAR T cell therapies

may for example involve administration of from 10^6 to 10^9 cells/kg, with or without a course of lymphodepletion, for example with cyclophosphamide. The cells or population of cells can be administrated in one or more doses. In another embodiment, the effective amount of cells are administrated as a single dose. In another embodiment, the effective amount of cells are administrated as more than one dose over a period time. Timing of administration is within the judgment of managing physician and depends on the clinical condition of the patient. The cells or population of cells may be obtained from any source, such as a blood bank or a donor. While individual needs vary, determination of optimal ranges of effective amounts of a given cell type for a particular disease or conditions are within the skill of one in the art. An effective amount means an amount which provides a therapeutic or prophylactic benefit. The dosage administrated will be dependent upon the age, health and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment and the nature of the effect desired.

[00862] In another embodiment, the effective amount of cells or composition comprising those cells are administrated parenterally. The administration can be an intravenous administration. The administration can be directly done by injection within a tumor.

[00863] To guard against possible adverse reactions, engineered immunoresponsive cells may be equipped with a transgenic safety switch, in the form of a transgene that renders the cells vulnerable to exposure to a specific signal. For example, the herpes simplex viral thymidine kinase (TK) gene may be used in this way, for example by introduction into allogeneic T lymphocytes used as donor lymphocyte infusions following stem cell transplantation (Greco, et al., Improving the safety of cell therapy with the TK-suicide gene. Front. Pharmacol. 2015; 6: 95). In such cells, administration of a nucleoside prodrug such as ganciclovir or acyclovir causes cell death. Alternative safety switch constructs include inducible caspase 9, for example triggered by administration of a small-molecule dimerizer that brings together two nonfunctional icasp9 molecules to form the active enzyme. A wide variety of alternative approaches to implementing cellular proliferation controls have been described (see U.S. Patent Publication No. 20130071414; PCT Patent Publication WO2011146862; PCT Patent Publication WO2014011987; PCT Patent Publication WO2013040371; Zhou et al. BLOOD, 2014, 123/25:3895 – 3905; Di Stasi et al., The New England Journal of Medicine 2011; 365:1673-1683; Sadelain M, The New England Journal of Medicine 2011; 365:1735-173; Ramos et al., Stem Cells 28(6):1107-15 (2010)).

[00864] In a further refinement of adoptive therapies, genome editing may be used to tailor immunoresponsive cells to alternative implementations, for example providing edited CAR T cells (see Poirot et al., 2015, Multiplex genome edited T-cell manufacturing platform for "off-

the-shelf" adoptive T-cell immunotherapies, Cancer Res 75 (18): 3853; Ren et al., 2017, Multiplex genome editing to generate universal CAR T cells resistant to PD1 inhibition, Clin Cancer Res. 2017 May 1;23(9):2255-2266. doi: 10.1158/1078-0432.CCR-16-1300. Epub 2016 Nov 4; Qasim et al., 2017, Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells, Sci Transl Med. 2017 Jan 25;9(374); Legut, et al., 2018, CRISPR-mediated TCR replacement generates superior anticancer transgenic T cells. Blood, 131(3), 311-322; and Georgiadis et al., Long Terminal Repeat CRISPR-CAR-Coupled "Universal" T Cells Mediate Potent Anti-leukemic Effects, Molecular Therapy, In Press, Corrected Proof, Available online 6 March 2018). Cells may be edited using any CRISPR system and method of use thereof as described herein. CRISPR systems may be delivered to an immune cell by any method described herein. In preferred embodiments, cells are edited ex vivo and transferred to a subject in need thereof. Immunoresponsive cells, CAR T cells or any cells used for adoptive cell transfer may be edited. Editing may be performed for example to insert or knock-in an exogenous gene, such as an exogenous gene encoding a CAR or a TCR, at a preselected locus in a cell (e.g. TRAC locus); to eliminate potential alloreactive T-cell receptors (TCR) or to prevent inappropriate pairing between endogenous and exogenous TCR chains, such as to knock-out or knock-down expression of an endogenous TCR in a cell; to disrupt the target of a chemotherapeutic agent in a cell; to block an immune checkpoint, such as to knock-out or knock-down expression of an immune checkpoint protein or receptor in a cell; to knock-out or knock-down expression of other gene or genes in a cell, the reduced expression or lack of expression of which can enhance the efficacy of adoptive therapies using the cell; to knock-out or knock-down expression of an endogenous gene in a cell, said endogenous gene encoding an antigen targeted by an exogenous CAR or TCR; to knock-out or knock-down expression of one or more MHC constituent proteins in a cell; to activate a T cell; to modulate cells such that the cells are resistant to exhaustion or dysfunction; and/or increase the differentiation and/or proliferation of functionally exhausted or dysfunctional CD8+ T-cells (see PCT Patent Publications: WO2013176915, WO2014059173, WO2014172606, WO2014184744, and WO2014191128).

[00865] In certain embodiments, editing may result in inactivation of a gene. By inactivating a gene, it is intended that the gene of interest is not expressed in a functional protein form. In a particular embodiment, the CRISPR system specifically catalyzes cleavage in one targeted gene thereby inactivating said targeted gene. The nucleic acid strand breaks caused are commonly repaired through the distinct mechanisms of homologous recombination or non-homologous end joining (NHEJ). However, NHEJ is an imperfect repair process that often

results in changes to the DNA sequence at the site of the cleavage. Repair via non-homologous end joining (NHEJ) often results in small insertions or deletions (Indel) and can be used for the creation of specific gene knockouts. Cells in which a cleavage induced mutagenesis event has occurred can be identified and/or selected by well-known methods in the art. In certain embodiments, homology directed repair (HDR) is used to concurrently inactivate a gene (e.g., TRAC) and insert an endogenous TCR or CAR into the inactivated locus.

[00866] Hence, in certain embodiments, editing of cells (such as by CRISPR/Cas), particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to insert or knock-in an exogenous gene, such as an exogenous gene encoding a CAR or a TCR, at a preselected locus in a cell. Conventionally, nucleic acid molecules encoding CARs or TCRs are transfected or transduced to cells using randomly integrating vectors, which, depending on the site of integration, may lead to clonal expansion, oncogenic transformation, variegated transgene expression and/or transcriptional silencing of the transgene. Directing of transgene(s) to a specific locus in a cell can minimize or avoid such risks and advantageously provide for uniform expression of the transgene(s) by the cells. Without limitation, suitable 'safe harbor' loci for directed transgene integration include CCR5 or AAVS1. Homology-directed repair (HDR) strategies are known and described elsewhere in this specification allowing to insert transgenes into desired loci (e.g., TRAC locus).

[00867] Further suitable loci for insertion of transgenes, in particular CAR or exogenous TCR transgenes, include without limitation loci comprising genes coding for constituents of endogenous T-cell receptor, such as T-cell receptor alpha locus (TRA) or T-cell receptor beta locus (TRB), for example T-cell receptor alpha constant (TRAC) locus, T-cell receptor beta constant 1 (TRBC1) locus or T-cell receptor beta constant 2 (TRBC1) locus. Advantageously, insertion of a transgene into such locus can simultaneously achieve expression of the transgene, potentially controlled by the endogenous promoter, and knock-out expression of the endogenous TCR. This approach has been exemplified in Eyquem et al., (2017) Nature 543: 113-117, wherein the authors used CRISPR/Cas9 gene editing to knock-in a DNA molecule encoding a CD19-specific CAR into the TRAC locus downstream of the endogenous promoter; the CAR-T cells obtained by CRISPR were significantly superior in terms of reduced tonic CAR signaling and exhaustion.

[00868] T cell receptors (TCR) are cell surface receptors that participate in the activation of T cells in response to the presentation of antigen. The TCR is generally made from two chains, α and β, which assemble to form a heterodimer and associates with the CD3-transducing

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

[00869] Hence, in certain embodiments, editing of cells (such as by CRISPR/Cas), particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to knock-out or knock-down expression of an endogenous TCR in a cell. For example, NHEJ-based or HDR-based gene editing approaches can be employed to disrupt the endogenous TCR alpha and/or beta chain genes. For example, gene editing system or systems, such as CRISPR/Cas system or systems, can be designed to target a sequence found within the TCR beta chain conserved between the beta 1 and beta 2 constant region genes (TRBC1 and TRBC2) and/or to target the constant region of the TCR alpha chain (TRAC) gene.

[00870] Allogeneic cells are rapidly rejected by the host immune system. It has been demonstrated that, allogeneic leukocytes present in non-irradiated blood products will persist for no more than 5 to 6 days (Boni, Muranski et al. 2008 Blood 1;112(12):4746-54). Thus, to prevent rejection of allogeneic cells, the host's immune system usually has to be suppressed to some extent. However, in the case of adoptive cell transfer the use of immunosuppressive drugs also have a detrimental effect on the introduced therapeutic T cells. Therefore, to effectively use an adoptive immunotherapy approach in these conditions, the introduced cells would need to be resistant to the immunosuppressive treatment. Thus, in a particular embodiment, the present invention further comprises a step of modifying T cells to make them resistant to an immunosuppressive agent, preferably by inactivating at least one gene encoding a target for an immunosuppressive agent. An immunosuppressive agent is an agent that suppresses immune function by one of several mechanisms of action. An immunosuppressive agent can be, but is

not limited to a calcineurin inhibitor, a target of rapamycin, an interleukin-2 receptor α-chain blocker, an inhibitor of inosine monophosphate dehydrogenase, an inhibitor of dihydrofolic acid reductase, a corticosteroid or an immunosuppressive antimetabolite. The present invention allows conferring immunosuppressive resistance to T cells for immunotherapy by inactivating the target of the immunosuppressive agent in T cells. As non-limiting examples, targets for an immunosuppressive agent can be a receptor for an immunosuppressive agent such as: CD52, glucocorticoid receptor (GR), a FKBP family gene member and a cyclophilin family gene member.

[00871] In certain embodiments, editing of cells (such as by CRISPR/Cas), particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to block an immune checkpoint, such as to knock-out or knock-down expression of an immune checkpoint protein or receptor in a cell. Immune checkpoints are inhibitory pathways that slow down or stop immune reactions and prevent excessive tissue damage from uncontrolled activity of immune cells. In certain embodiments, the immune checkpoint targeted is the programmed death-1 (PD-1 or CD279) gene (PDCD1). In other embodiments, the immune checkpoint targeted is cytotoxic T-lymphocyte-associated antigen (CTLA-4). In additional embodiments, the immune checkpoint targeted is another member of the CD28 and CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. In further additional embodiments, the immune checkpoint targeted is a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3.

[00872] Additional immune checkpoints include Src homology 2 domain-containing protein tyrosine phosphatase 1 (SHP-1) (Watson HA, et al., SHP-1: the next checkpoint target for cancer immunotherapy? Biochem Soc Trans. 2016 Apr 15;44(2):356-62). SHP-1 is a widely expressed inhibitory protein tyrosine phosphatase (PTP). In T-cells, it is a negative regulator of antigen-dependent activation and proliferation. It is a cytosolic protein, and therefore not amenable to antibody-mediated therapies, but its role in activation and proliferation makes it an attractive target for genetic manipulation in adoptive transfer strategies, such as chimeric antigen receptor (CAR) T cells. Immune checkpoints may also include T cell immunoreceptor with Ig and ITIM domains (TIGIT/Vstm3/WUCAM/VSIG9) and VISTA (Le Mercier I, et al., (2015) Beyond CTLA-4 and PD-1, the generation Z of negative checkpoint regulators. Front. Immunol. 6:418).

[00873] WO2014172606 relates to the use of MT1 and/or MT2 inhibitors to increase proliferation and/or activity of exhausted CD8+ T-cells and to decrease CD8+ T-cell exhaustion (e.g., decrease functionally exhausted or unresponsive CD8+ immune cells). In

certain embodiments, metallothioneins are targeted by gene editing in adoptively transferred T cells.

[00874] In certain embodiments, targets of gene editing may be at least one targeted locus involved in the expression of an immune checkpoint protein. Such targets may include, but are not limited to CTLA4, PPP2CA, PPP2CB, PTPN6, PTPN22, PDCD1, ICOS (CD278), PDL1, KIR, LAG3, HAVCR2, BTLA, CD160, TIGIT, CD96, CRTAM, LAIR1, SIGLEC7, SIGLEC9, CD244 (2B4), TNFRSF10B, TNFRSF10A, CASP8, CASP10, CASP3, CASP6, CASP7, FADD, FAS, TGFBRII, TGFRBRI, SMAD2, SMAD3, SMAD4, SMAD10, SKI, SKIL, TGIF1, IL10RA, IL10RB, HMOX2, IL6R, IL6ST, EIF2AK4, CSK, PAG1, SIT1, FOXP3, PRDM1, BATF, VISTA, GUCY1A2, GUCY1A3, GUCY1B2, GUCY1B3, MT1, MT2, CD40, OX40, CD137, GITR, CD27, SHP-1, TIM-3, CEACAM-1, CEACAM-3, or CEACAM-5. In preferred embodiments, the gene locus involved in the expression of PD-1 or CTLA-4 genes is targeted. In other preferred embodiments, combinations of genes are targeted, such as but not limited to PD-1 and TIGIT.

[00875] By means of an example and without limitation, WO2016196388 concerns an engineered T cell comprising (a) a genetically engineered antigen receptor that specifically binds to an antigen, which receptor may be a CAR; and (b) a disrupted gene encoding a PD-L1, an agent for disruption of a gene encoding a PD- L1, and/or disruption of a gene encoding PD-L1, wherein the disruption of the gene may be mediated by a gene editing nuclease, a zinc finger nuclease (ZFN), CRISPR/Cas9 and/or TALEN. WO2015142675 relates to immune effector cells comprising a CAR in combination with an agent (such as CRISPR, TALEN or ZFN) that increases the efficacy of the immune effector cells in the treatment of cancer, wherein the agent may inhibit an immune inhibitory molecule, such as PD1, PD-L1, CTLA-4, TIM-3, LAG-3, VISTA, BTLA, TIGIT, LAIR1, CD160, 2B4, TGFR beta, CEACAM-1, CEACAM-3, or CEACAM-5. Ren et al., (2017) Clin Cancer Res 23 (9) 2255-2266 performed lentiviral delivery of CAR and electro-transfer of Cas mRNA and gRNAs targeting endogenous TCR, β-2 microglobulin (B2M) and PD1 simultaneously, to generate gene-disrupted allogeneic CAR T cells deficient of TCR, HLA class I molecule and PD1.

[00876] In certain embodiments, cells may be engineered to express a CAR, wherein expression and/or function of methylcytosine dioxygenase genes (TET1, TET2 and/or TET3) in the cells has been reduced or eliminated, such as by CRISPR, ZNF or TALEN (for example, as described in WO201704916).

[00877] In certain embodiments, editing of cells (such as by CRISPR/Cas), particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells,

may be performed to knock-out or knock-down expression of an endogenous gene in a cell, said endogenous gene encoding an antigen targeted by an exogenous CAR or TCR, thereby reducing the likelihood of targeting of the engineered cells. In certain embodiments, the targeted antigen may be one or more antigen selected from the group consisting of CD38, CD138, CS-1, CD33, CD26, CD30, CD53, CD92, CD100, CD148, CD150, CD200, CD261, CD262, CD362, human telomerase reverse transcriptase (hTERT), survivin, mouse double minute 2 homolog (MDM2), cytochrome P450 1B1 (CYP1B), HER2/neu, Wilms' tumor gene 1 (WT1), livin, alphafetoprotein (AFP), carcinoembryonic antigen (CEA), mucin 16 (MUC16), MUC1, prostate-specific membrane antigen (PSMA), p53, cyclin (D1), B cell maturation antigen (BCMA), transmembrane activator and CAML Interactor (TACI), and B-cell activating factor receptor (BAFF-R) (for example, as described in WO2016011210 and WO2017011804).

[00878] In certain embodiments, editing of cells (such as by CRISPR/Cas), particularly cells intended for adoptive cell therapies, more particularly immunoresponsive cells such as T cells, may be performed to knock-out or knock-down expression of one or more MHC constituent proteins, such as one or more HLA proteins and/or beta-2 microglobulin (B2M), in a cell, whereby rejection of non-autologous (e.g., allogeneic) cells by the recipient's immune system can be reduced or avoided. In preferred embodiments, one or more HLA class I proteins, such as HLA-A, B and/or C, and/or B2M may be knocked-out or knocked-down. Preferably, B2M may be knocked-out or knocked-down. By means of an example, Ren et al., (2017) Clin Cancer Res 23 (9) 2255-2266 performed lentiviral delivery of CAR and electro-transfer of Cas9 mRNA and gRNAs targeting endogenous TCR, β-2 microglobulin (B2M) and PD1 simultaneously, to generate gene-disrupted allogeneic CAR T cells deficient of TCR, HLA class I molecule and PD1.

[00879] In other embodiments, at least two genes are edited. Pairs of genes may include, but are not limited to PD1 and TCRα, PD1 and TCRβ, CTLA-4 and TCRα, CTLA-4 and TCRβ, LAG3 and TCRα, LAG3 and TCRβ, Tim3 and TCRα, Tim3 and TCRβ, BTLA and TCRα, BTLA and TCRβ, BY55 and TCRα, BY55 and TCRβ, TIGIT and TCRα, TIGIT and TCRβ, B7H5 and TCRα, B7H5 and TCRβ, LAIR1 and TCRα, LAIR1 and TCRβ, SIGLEC10 and TCRα, SIGLEC10 and TCRβ, 2B4 and TCRα, 2B4 and TCRβ, B2M and TCRα, B2M and TCRβ.

[00880] In certain embodiments, a cell may be multiplied edited (multiplex genome editing) as taught herein to (1) knock-out or knock-down expression of an endogenous TCR (for example, TRBC1, TRBC2 and/or TRAC), (2) knock-out or knock-down expression of an

immune checkpoint protein or receptor (for example PD1, PD-L1 and/or CTLA4); and (3) knock-out or knock-down expression of one or more MHC constituent proteins (for example, HLA-A, B and/or C, and/or B2M, preferably B2M).

[00881] Whether prior to or after genetic modification of the T cells, the T cells can be activated and expanded generally using methods as described, for example, in U.S. Patents 6,352,694; 6,534,055; 6,905,680; 5,858,358; 6,887,466; 6,905,681; 7,144,575; 7,232,566; 7,175,843; 5,883,223; 6,905,874; 6,797,514; 6,867,041; and 7,572,631. T cells can be expanded in vitro or in vivo.

[00882] Immune cells may be obtained using any method known in the art. In one embodiment, allogenic T cells may be obtained from healthy subjects. In one embodiment T cells that have infiltrated a tumor are isolated. T cells may be removed during surgery. T cells may be isolated after removal of tumor tissue by biopsy. T cells may be isolated by any means known in the art. In one embodiment, T cells are obtained by apheresis. In one embodiment, the method may comprise obtaining a bulk population of T cells from a tumor sample by any suitable method known in the art. For example, a bulk population of T cells can be obtained from a tumor sample by dissociating the tumor sample into a cell suspension from which specific cell populations can be selected. Suitable methods of obtaining a bulk population of T cells may include, but are not limited to, any one or more of mechanically dissociating (e.g., mincing) the tumor, enzymatically dissociating (e.g., digesting) the tumor, and aspiration (e.g., as with a needle).

[00883] The bulk population of T cells obtained from a tumor sample may comprise any suitable type of T cell. Preferably, the bulk population of T cells obtained from a tumor sample comprises tumor infiltrating lymphocytes (TILs).

[00884] The tumor sample may be obtained from any mammal. Unless stated otherwise, as used herein, the term "mammal" refers to any mammal including, but not limited to, mammals of the order Logomorpha, such as rabbits; the order Carnivora, including Felines (cats) and Canines (dogs); the order Artiodactyla, including Bovines (cows) and Swines (pigs); or of the order Perssodactyla, including Equines (horses). The mammals may be non-human primates, e.g., of the order Primates, Ceboids, or Simoids (monkeys) or of the order Anthropoids (humans and apes). In some embodiments, the mammal may be a mammal of the order Rodentia, such as mice and hamsters. Preferably, the mammal is a non-human primate or a human. An especially preferred mammal is the human.

[00885] T cells can be obtained from a number of sources, including peripheral blood mononuclear cells (PBMC), bone marrow, lymph node tissue, spleen tissue, and tumors. In

certain embodiments of the present invention, T cells can be obtained from a unit of blood collected from a subject using any number of techniques known to the skilled artisan, such as Ficoll separation. In one preferred embodiment, cells from the circulating blood of an individual are obtained by apheresis or leukapheresis. The apheresis product typically contains lymphocytes, including T cells, monocytes, granulocytes, B cells, other nucleated white blood cells, red blood cells, and platelets. In one embodiment, the cells collected by apheresis may be washed to remove the plasma fraction and to place the cells in an appropriate buffer or media for subsequent processing steps. In one embodiment of the invention, the cells are washed with phosphate buffered saline (PBS). In an alternative embodiment, the wash solution lacks calcium and may lack magnesium or may lack many if not all divalent cations. Initial activation steps in the absence of calcium lead to magnified activation. As those of ordinary skill in the art would readily appreciate a washing step may be accomplished by methods known to those in the art, such as by using a semi-automated "flow-through" centrifuge (for example, the Cobe 2991 cell processor) according to the manufacturer's instructions. After washing, the cells may be resuspended in a variety of biocompatible buffers, such as, for example, Ca-free, Mg-free PBS. Alternatively, the undesirable components of the apheresis sample may be removed and the cells directly resuspended in culture media.

[00886] In another embodiment, T cells are isolated from peripheral blood lymphocytes by lysing the red blood cells and depleting the monocytes, for example, by centrifugation through a PERCOLL™ gradient. A specific subpopulation of T cells, such as CD28+, CD4+, CDC, CD45RA+, and CD45RO+ T cells, can be further isolated by positive or negative selection techniques. For example, in one preferred embodiment, T cells are isolated by incubation with anti-CD3/anti-CD28 (i.e., 3×28)-conjugated beads, such as DYNABEADS® M-450 CD3/CD28 T, or XCYTE DYNABEADS™ for a time period sufficient for positive selection of the desired T cells. In one embodiment, the time period is about 30 minutes. In a further embodiment, the time period ranges from 30 minutes to 36 hours or longer and all integer values there between. In a further embodiment, the time period is at least 1, 2, 3, 4, 5, or 6 hours. In yet another preferred embodiment, the time period is 10 to 24 hours. In one preferred embodiment, the incubation time period is 24 hours. For isolation of T cells from patients with leukemia, use of longer incubation times, such as 24 hours, can increase cell yield. Longer incubation times may be used to isolate T cells in any situation where there are few T cells as compared to other cell types, such in isolating tumor infiltrating lymphocytes (TIL) from tumor tissue or from immunocompromised individuals. Further, use of longer incubation times can increase the efficiency of capture of CD8+ T cells.

[00887] Enrichment of a T cell population by negative selection can be accomplished with a combination of antibodies directed to surface markers unique to the negatively selected cells. A preferred method is cell sorting and/or selection via negative magnetic immunoadherence or flow cytometry that uses a cocktail of monoclonal antibodies directed to cell surface markers present on the cells negatively selected. For example, to enrich for CD4+ cells by negative selection, a monoclonal antibody cocktail typically includes antibodies to CD14, CD20, CD11b, CD16, HLA-DR, and CD8.

[00888] Further, monocyte populations (CD14+ cells) may be depleted from blood preparations by a variety of methodologies, including anti-CD14 coated beads or columns, or utilization of the phagocytotic activity of these cells to facilitate removal. Accordingly, in one embodiment, the invention uses paramagnetic particles of a size sufficient to be engulfed by phagocytotic monocytes. In certain embodiments, the paramagnetic particles are commercially available beads, for example, those produced by Life Technologies under the trade name Dynabeads™. In one embodiment, other non-specific cells are removed by coating the paramagnetic particles with "irrelevant" proteins (e.g., serum proteins or antibodies). Irrelevant proteins and antibodies include those proteins and antibodies or fragments thereof that do not specifically target the T cells to be isolated. In certain embodiments, the irrelevant beads include beads coated with sheep anti-mouse antibodies, goat anti-mouse antibodies, and human serum albumin.

[00889] In brief, such depletion of monocytes is performed by preincubating T cells isolated from whole blood, apheresed peripheral blood, or tumors with one or more varieties of irrelevant or non-antibody coupled paramagnetic particles at any amount that allows for removal of monocytes (approximately a 20:1 bead:cell ratio) for about 30 minutes to 2 hours at 22 to 37 degrees C., followed by magnetic removal of cells which have attached to or engulfed the paramagnetic particles. Such separation can be performed using standard methods available in the art. For example, any magnetic separation methodology may be used including a variety of which are commercially available, (e.g., DYNAL® Magnetic Particle Concentrator (DYNAL MPC®)). Assurance of requisite depletion can be monitored by a variety of methodologies known to those of ordinary skill in the art, including flow cytometric analysis of CD14 positive cells, before and after depletion.

[00890] For isolation of a desired population of cells by positive or negative selection, the concentration of cells and surface (e.g., particles such as beads) can be varied. In certain embodiments, it may be desirable to significantly decrease the volume in which beads and cells are mixed together (i.e., increase the concentration of cells), to ensure maximum contact of

cells and beads. For example, in one embodiment, a concentration of 2 billion cells/ml is used. In one embodiment, a concentration of 1 billion cells/ml is used. In a further embodiment, greater than 100 million cells/ml is used. In a further embodiment, a concentration of cells of 10, 15, 20, 25, 30, 35, 40, 45, or 50 million cells/ml is used. In yet another embodiment, a concentration of cells from 75, 80, 85, 90, 95, or 100 million cells/ml is used. In further embodiments, concentrations of 125 or 150 million cells/ml can be used. Using high concentrations can result in increased cell yield, cell activation, and cell expansion. Further, use of high cell concentrations allows more efficient capture of cells that may weakly express target antigens of interest, such as CD28-negative T cells, or from samples where there are many tumor cells present (i.e., leukemic blood, tumor tissue, etc.). Such populations of cells may have therapeutic value and would be desirable to obtain. For example, using high concentration of cells allows more efficient selection of CD8+ T cells that normally have weaker CD28 expression.

[00891] In a related embodiment, it may be desirable to use lower concentrations of cells. By significantly diluting the mixture of T cells and surface (e.g., particles such as beads), interactions between the particles and cells is minimized. This selects for cells that express high amounts of desired antigens to be bound to the particles. For example, CD4+ T cells express higher levels of CD28 and are more efficiently captured than CD8+ T cells in dilute concentrations. In one embodiment, the concentration of cells used is 5×10^6/ml. In other embodiments, the concentration used can be from about 1×10^5/ml to 1×10^6/ml, and any integer value in between.

[00892] T cells can also be frozen. Wishing not to be bound by theory, the freeze and subsequent thaw step provides a more uniform product by removing granulocytes and to some extent monocytes in the cell population. After a washing step to remove plasma and platelets, the cells may be suspended in a freezing solution. While many freezing solutions and parameters are known in the art and will be useful in this context, one method involves using PBS containing 20% DMSO and 8% human serum albumin, or other suitable cell freezing media, the cells then are frozen to −80° C at a rate of 1° per minute and stored in the vapor phase of a liquid nitrogen storage tank. Other methods of controlled freezing may be used as well as uncontrolled freezing immediately at −20° C. or in liquid nitrogen.

[00893] T cells for use in the present invention may also be antigen-specific T cells. For example, tumor-specific T cells can be used. In certain embodiments, antigen-specific T cells can be isolated from a patient of interest, such as a patient afflicted with a cancer or an infectious disease. In one embodiment, neoepitopes are determined for a subject and T cells

specific to these antigens are isolated. Antigen-specific cells for use in expansion may also be generated in vitro using any number of methods known in the art, for example, as described in U.S. Patent Publication No. US 20040224402 entitled, Generation and Isolation of Antigen-Specific T Cells, or in U.S. Pat. Nos. 6,040,177. Antigen-specific cells for use in the present invention may also be generated using any number of methods known in the art, for example, as described in Current Protocols in Immunology, or Current Protocols in Cell Biology, both published by John Wiley & Sons, Inc., Boston, Mass.

[00894] In a related embodiment, it may be desirable to sort or otherwise positively select (e.g. via magnetic selection) the antigen specific cells prior to or following one or two rounds of expansion. Sorting or positively selecting antigen-specific cells can be carried out using peptide-MHC tetramers (Altman, et al., Science. 1996 Oct. 4; 274(5284):94-6). In another embodiment, the adaptable tetramer technology approach is used (Andersen et al., 2012 Nat Protoc. 7:891-902). Tetramers are limited by the need to utilize predicted binding peptides based on prior hypotheses, and the restriction to specific HLAs. Peptide-MHC tetramers can be generated using techniques known in the art and can be made with any MHC molecule of interest and any antigen of interest as described herein. Specific epitopes to be used in this context can be identified using numerous assays known in the art. For example, the ability of a polypeptide to bind to MHC class I may be evaluated indirectly by monitoring the ability to promote incorporation of ^{125}I labeled β2-microglobulin (β2m) into MHC class I/β2m/peptide heterotrimeric complexes (see Parker et al., J. Immunol. 152:163, 1994).

[00895] In one embodiment cells are directly labeled with an epitope-specific reagent for isolation by flow cytometry followed by characterization of phenotype and TCRs. In one embodiment, T cells are isolated by contacting with T cell specific antibodies. Sorting of antigen-specific T cells, or generally any cells of the present invention, can be carried out using any of a variety of commercially available cell sorters, including, but not limited to, MoFlo sorter (DakoCytomation, Fort Collins, Colo.), FACSAria™, FACSArray™, FACSVantage™, BD™ LSR II, and FACSCalibur™ (BD Biosciences, San Jose, Calif.).

[00896] In a preferred embodiment, the method comprises selecting cells that also express CD3. The method may comprise specifically selecting the cells in any suitable manner. Preferably, the selecting is carried out using flow cytometry. The flow cytometry may be carried out using any suitable method known in the art. The flow cytometry may employ any suitable antibodies and stains. Preferably, the antibody is chosen such that it specifically recognizes and binds to the particular biomarker being selected. For example, the specific selection of CD3, CD8, TIM-3, LAG-3, 4-1BB, or PD-1 may be carried out using anti-CD3,

anti-CD8, anti-TIM-3, anti-LAG-3, anti-4-lBB, or anti-PD-1 antibodies, respectively. The antibody or antibodies may be conjugated to a bead (e.g., a magnetic bead) or to a fluorochrome. Preferably, the flow cytometry is fluorescence-activated cell sorting (FACS). TCRs expressed on T cells can be selected based on reactivity to autologous tumors. Additionally, T cells that are reactive to tumors can be selected for based on markers using the methods described in patent publication Nos. WO2014133567 and WO2014133568, herein incorporated by reference in their entirety. Additionally, activated T cells can be selected for based on surface expression of CD107a.

[00897] In one embodiment of the invention, the method further comprises expanding the numbers of T cells in the enriched cell population. Such methods are described in U.S. Patent No. 8,637,307 and is herein incorporated by reference in its entirety. The numbers of T cells may be increased at least about 3-fold (or 4-, 5-, 6-, 7-, 8-, or 9-fold), more preferably at least about 10-fold (or 20-, 30-, 40-, 50-, 60-, 70-, 80-, or 90-fold), more preferably at least about 100-fold, more preferably at least about 1,000 fold, or most preferably at least about 100,000-fold. The numbers of T cells may be expanded using any suitable method known in the art. Exemplary methods of expanding the numbers of cells are described in patent publication No. WO 2003057171, U.S. Patent No. 8,034,334, and U.S. Patent Application Publication No. 2012/0244133, each of which is incorporated herein by reference.

[00898] In one embodiment, *ex vivo* T cell expansion can be performed by isolation of T cells and subsequent stimulation or activation followed by further expansion. In one embodiment of the invention, the T cells may be stimulated or activated by a single agent. In another embodiment, T cells are stimulated or activated with two agents, one that induces a primary signal and a second that is a co-stimulatory signal. Ligands useful for stimulating a single signal or stimulating a primary signal and an accessory molecule that stimulates a second signal may be used in soluble form. Ligands may be attached to the surface of a cell, to an Engineered Multivalent Signaling Platform (EMSP), or immobilized on a surface. In a preferred embodiment both primary and secondary agents are co-immobilized on a surface, for example a bead or a cell. In one embodiment, the molecule providing the primary activation signal may be a CD3 ligand, and the co-stimulatory molecule may be a CD28 ligand or 4-1BB ligand.

[00899] In certain embodiments, T cells comprising a CAR or an exogenous TCR, may be manufactured as described in WO2015120096, by a method comprising: enriching a population of lymphocytes obtained from a donor subject; stimulating the population of lymphocytes with one or more T-cell stimulating agents to produce a population of activated

T cells, wherein the stimulation is performed in a closed system using serum-free culture medium; transducing the population of activated T cells with a viral vector comprising a nucleic acid molecule which encodes the CAR or TCR, using a single cycle transduction to produce a population of transduced T cells, wherein the transduction is performed in a closed system using serum-free culture medium; and expanding the population of transduced T cells for a predetermined time to produce a population of engineered T cells, wherein the expansion is performed in a closed system using serum-free culture medium. In certain embodiments, T cells comprising a CAR or an exogenous TCR, may be manufactured as described in WO2015120096, by a method comprising: obtaining a population of lymphocytes; stimulating the population of lymphocytes with one or more stimulating agents to produce a population of activated T cells, wherein the stimulation is performed in a closed system using serum-free culture medium; transducing the population of activated T cells with a viral vector comprising a nucleic acid molecule which encodes the CAR or TCR, using at least one cycle transduction to produce a population of transduced T cells, wherein the transduction is performed in a closed system using serum-free culture medium; and expanding the population of transduced T cells to produce a population of engineered T cells, wherein the expansion is performed in a closed system using serum-free culture medium. The predetermined time for expanding the population of transduced T cells may be 3 days. The time from enriching the population of lymphocytes to producing the engineered T cells may be 6 days. The closed system may be a closed bag system. Further provided is population of T cells comprising a CAR or an exogenous TCR obtainable or obtained by said method, and a pharmaceutical composition comprising such cells.

[00900] In certain embodiments, T cell maturation or differentiation in vitro may be delayed or inhibited by the method as described in WO2017070395, comprising contacting one or more T cells from a subject in need of a T cell therapy with an AKT inhibitor (such as, e.g., one or a combination of two or more AKT inhibitors disclosed in claim 8 of WO2017070395) and at least one of exogenous Interleukin-7 (IL-7) and exogenous Interleukin-15 (IL-15), wherein the resulting T cells exhibit delayed maturation or differentiation, and/or wherein the resulting T cells exhibit improved T cell function (such as, e.g., increased T cell proliferation; increased cytokine production; and/or increased cytolytic activity) relative to a T cell function of a T cell cultured in the absence of an AKT inhibitor.

[00901] In certain embodiments, a patient in need of a T cell therapy may be conditioned by a method as described in WO2016191756 comprising administering to the patient a dose of

cyclophosphamide between 200 mg/m2/day and 2000 mg/m2/day and a dose of fludarabine between 20 mg/m2/day and 900 mg/m^2/day.

KITS

[00902] In one aspect, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. In some embodiments, the kit comprises a vector system as taught herein or one or more of the components of the CRISPR-Cas system or complex as taught herein, such as crRNAs and/or Cas protein or Cas protein encoding mRNA, and instructions for using the kit. Elements may be provided individually or in combinations, and may be provided in any suitable container, such as a vial, a bottle, or a tube. In some embodiments, the kit includes instructions in one or more languages, for example in more than one language. The instructions may be specific to the applications and methods described herein. In some embodiments, a kit comprises one or more reagents for use in a process utilizing one or more of the elements described herein. Reagents may be provided in any suitable container. For example, a kit may provide one or more reaction or storage buffers. Reagents may be provided in a form that is usable in a particular assay, or in a form that requires addition of one or more other components before use (e.g., in concentrate or lyophilized form). A buffer can be any buffer, including but not limited to a sodium carbonate buffer, a sodium bicarbonate buffer, a borate buffer, a Tris buffer, a MOPS buffer, a HEPES buffer, and combinations thereof. In some embodiments, the buffer is alkaline. In some embodiments, the buffer has a pH from about 7 to about 10. In some embodiments, the kit comprises one or more oligonucleotides corresponding to a guide sequence for insertion into a vector so as to operably link the guide or crRNA sequence and a regulatory element. In some embodiments, the kit comprises a homologous recombination template polynucleotide. In some embodiments, the kit comprises one or more of the vectors and/or one or more of the polynucleotides described herein. The kit may advantageously allow to provide all elements of the systems of the invention.

[00903] The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1 – Exemplary Type IV Cas loci

[00904] Sequences of exemplary Type IV Cas loci are shown and direct repeat sequences in Table 1 .

[00905] Nucleic acids sequences of exemplary Type IV Cas proteins are shown in SEQ ID NOs: 1-8 in the Sequence Listing herein. All loci sequences of exemplary Type IV Cas proteins

are shown in SEQ ID NOs. 9-317 in the Sequence Listing herein. Complete loci sequences of exemplary Type IV Cas proteins are shown in SEQ ID NOs: 318-405 in the Sequence Listing herein.

Determining PAM sequences of Exemplary Type IV

[00906] PAM sequences of exemplary Type IV Cas were determined and the results are shown in FIG. 1.

EXAMPLE 2

[00907] Alignment of sequences of exemplary Type IV Cas proteins comprising HNH and DinG domains is shown in FIG. 2. FIG. 3 shows examples of loci comprising DinG and HNH sequences. The loci can have different configurations. Spacer BLAST was performed to identify PAM sequences (FIG. 4). (C/T)CN was determined as a PAM for the Type IV Cas. An exemplary Type IV Cas comprising DinG and HNH domains performed plasmid interference with 5' (C/T)CN PAM (FIG. 5).

EXAMPLE 3 – Exemplary Class 1, Type I Cas Loci

[00908] Sequences of exemplary class 1, Type I Cas loci are shown and direct repeat sequences in Table 2.

[00909] PAM sequences of exemplary Type I Cas were determined and the results are shown in FIG. 6.

* * *

[00910] Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention. This application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure come within known customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

CLAIMS

What is claimed is:

1. A non-naturally occurring, engineered composition comprising a Class 1, Type IV Cas system comprising a helicase comprising a DinG domain and a HNH domain that is less than 600 amino acids in size, and at least one guide sequence capable of complexing with the Cas system and directing binding of the guide-Cas system complex to a target polynucleotide.

2. 1The composition of any one of the preceding claims, further comprising one or more Cas proteins.

3. The composition of claim 2, wherein the one or more Cas proteins comprise Cas7 or a Cas7-like Cas protein, Cas5 or a Cas5-like Cas protein, Cas6 or a Cas6-like protein, and a Cse3 family protein.

4. The composition of claim 2, wherein the one or more Cas proteins comprise Csf2, Csf3, Cas6, and Pfam08798.

5. The composition of any one of the preceding claims, wherein the Cas protein is a protein with nucleotide sequence listed in Table 1 and SEQ ID NOs 1-405.

6. The composition of any one of the preceding claims, wherein the target sequence comprises a protospacer adjacent motif (PAM) at the 5' side of the target sequence.

7. The composition of claim 6, wherein the PAM sequence comprises CC.

8. The composition of claim 6, wherein the PAM sequence comprises (C/T)CN.

9. The composition of any one of claims 1-7, wherein the target sequence comprises a PAM at the 3' side of the target sequence.

10. The composition of claim 9, wherein the PAM sequence is GG.

11. The composition of any one of the preceding claims, wherein the composition comprises cascade/helicase activity.

12. The composition of any one of the preceding claims, wherein the composition does not comprise Cas3 DNA shredding activity.

13. The composition of any one of the preceding claims, further comprising a donor polynucleotide.

14. The composition of claim 13, wherein the donor polynucleotide:

a. introduces one or more mutations to the target polynucleotide;

b. corrects a premature stop codon in the target polynucleotide;

c. disrupts a splicing site;

d. restores a splicing site; or

e. a combination thereof.

15. The composition of claim 14, wherein the one or more mutations introduced by the donor polynucleotide comprises substitutions, deletions, insertions, or a combination thereof.

16. The composition of claim 14, wherein the one or more mutations causes a shift in an open reading frame on the target polynucleotide.

17. The composition of any one of the preceding claims, further comprising a plurality of guide molecules capable of complexing with the Cas protein and directing binding of the guide-Cas protein complex to one or more target polynucleotides.

18. A composition comprising one or more polynucleotides encoding:

a. one or more class 1, Type IV Cas proteins or functional fragments thereof, wherein the one or more class 1, Type IV Cas proteins comprises DinG protein with a length less than 600 amino acids; and

b. one or more guide molecules capable of complexing with the class 1, Type IV Cas protein and directing binding of the guide-Cas protein complex to one or more target polynucleotides.

19. The composition of claim 18, further comprising a donor polynucleotide.

20. The composition of claim 19, wherein the donor polynucleotide comprises a polynucleotide insert.

21. The composition of claim 18, wherein the one or more target polynucleotides encode a part of the components in claim 1.

22. The composition of claim 18, wherein the one or more class 1, Type IV Cas proteins comprise Csf2 (Cas7-like), Csf3 (Cas5-like), Cas6, and Pfam08798 (Cse3 family).

23. A vector comprising the one or more polynucleotides of any one of claims 18-22.

24. An engineered cell comprising the composition of any one of claims 1-23.

25. A method of modifying a target polynucleotide sequence in a cell, comprising introducing to the cell:

a. one or more class 1, Type IV Cas proteins or functional fragments thereof, wherein the one or more class 1, Type IV Cas proteins comprises DinG protein with a length less than 600 amino acids; and

b. one or more guide molecules capable of complexing with the class 1, Type IV Cas protein and directing binding of the guide-Cas protein complex to one or more target polynucleotides.

26. The method of claim 25, further comprising introducing a donor polynucleotide.

27. The method of either claim 25 or claim 26, wherein the donor polynucleotide:

a. introduces one or more mutations to the target polynucleotide;

b. corrects a premature stop codon in the target polynucleotide;

c. disrupts a splicing site;

d. restores a splicing site; or

e. a combination thereof.

28. The method of claim 27, wherein the one or more mutations introduced by the donor polynucleotide comprises substitutions, deletions, insertions, or a combination thereof.

29. The method of claim 27, wherein the one or more mutations causes a shift in an open reading frame on the target polynucleotide.

30. The method of any one of claims 27-29, wherein the cell is a prokaryotic cell.

31. The method of any one of claims 25-29, wherein the cell is a eukaryotic cell.

32. The method of any one of claims 25-29, wherein the cell is a mammalian cell, a cell of a non-human primate, or a human cell.

33. The method of any one of claims 25-29, wherein the cell is a plant cell.

34. A non-naturally occurring, engineered composition comprising a Cas protein that comprises an HNH domain and a helicase domain, and at least one guide sequence capable of complexing with the Cas protein and directing binding of the guide-Cas protein complex to a target polynucleotide.

35. The composition of claim 34, wherein the helicase domain is a DinG domain.

36. The composition of claim 34, wherein the Cas protein is a class 1 Type IV Cas protein.

37. A composition comprising one or more polynucleotides encoding one or more components of the composition of claim 34.

38. A method of modifying a target polynucleotide sequence in a cell, comprising introducing to the cell one or more components of composition of claim 34.

39. A non-naturally occurring, engineered composition comprising a class 1, Type I Cas system that comprises an HNH domain that is less than 400 amino acids in length, and at least one guide sequence capable of complexing with the Cas protein and directing binding of the guide-Cas protein complex to a target polynucleotide.

40. The composition of claim 39, wherein the class 1, Type I Cas system HNH domain is comprised in an McRA protein.

41. The composition of claim 40, wherein the class 1, Type I Cas system HNH domain is comprised in a Cas5 protein.

42. The composition of claims 39-41, further comprising one or more Cas proteins.

43. The composition of claim 42, wherein the one or more Cas proteins comprise Cas1, Cas2, Cas6, Cas7, Cas8e-Cse fusion, CasE, Cse2, and McrA.

44. The composition of claims 39-43, wherein the Cas protein is a protein with nucleotide sequence listed in Table 2 and SEQ ID NOs. 495-1212.

45. The composition of claims 39-45, wherein the target sequence comprises a protospacer adjacent motif (PAM) at the 5' side of the target sequence.

46. The composition of claim 45, wherein the PAM sequence is AAG.

47. The composition of any one of claims 39-45, wherein the target sequence comprises a PAM at the 3' side of the target sequence.

48. The composition of claim 47, wherein the PAM sequence is CTT.

49. The composition of claims 39-48, wherein the composition comprises cascade/helicase activity.

50. The composition of claims 39-49, wherein the composition does not comprise Cas3 and/or Cas3 DNA shredding activity.

51. The composition of claims 39-50, further comprising a donor polynucleotide.

52. The composition of claim 51, wherein the donor polynucleotide:

a. introduces one or more mutations to the target polynucleotide;

b. corrects a premature stop codon in the target polynucleotide;

c. disrupts a splicing site;

d. restores a splicing site; or

e. a combination thereof.

53. The composition of claim 52, wherein the one or more mutations introduced by the donor polynucleotide comprises substitutions, deletions, insertions, or a combination thereof.

54. The composition of claim 52, wherein the one or more mutations causes a shift in an open reading frame on the target polynucleotide.

55. The composition of claims 39-54, further comprising a plurality of guide molecules capable of complexing with the Cas protein and directing binding of the guide-Cas protein complex to one or more target polynucleotides.

56. A composition comprising one or more polynucleotides encoding:

a. one or more class 1, Type I Cas proteins or functional fragments thereof, wherein the one or more class 1, Type I Cas proteins comprises Cas5 protein with a length less than 400 amino acids; and

b. one or more guide molecules capable of complexing with the class 1, Type I Cas protein and directing binding of the guide-Cas protein complex to one or more target polynucleotides.

57. The composition of claim 56, further comprising a donor polynucleotide.

58. The composition of claim 57, wherein the donor polynucleotide comprises a polynucleotide insert.

59. The composition of claim 56, wherein the one or more target polynucleotides encode a part of the components in claim 1.

60. The composition of claim 56, wherein the one or more Type I Cas proteins comprise Cas1, Cas2, Cas6, Cas7, Cas8e-Cse fusion, CasE, Cse2, and McrA.

61. A vector comprising the one or more polynucleotides of any one of claims 56-60.

62. An engineered cell comprising the composition of any one of claims 39-61.

63. A method of modifying a target polynucleotide sequence in a cell, comprising introducing to the cell:

a. one or more class 1, Type I Cas proteins or functional fragments thereof, wherein the one or more class 1, Type I Cas proteins comprises Cas5 protein with a length less than 400 amino acids; and

b. one or more guide molecules capable of complexing with the class 1, Type I Cas protein and directing binding of the guide-Cas protein complex to one or more target polynucleotides.

64. The method of claim 63, further comprising introducing a donor polynucleotide.

65. The method of either claim 63 or claim 64, wherein the donor polynucleotide:

a. introduces one or more mutations to the target polynucleotide,

b. corrects a premature stop codon in the target polynucleotide,

c. disrupts a splicing site,

d. restores a splicing site, or

e. a combination thereof.

66. The method of claim 65, wherein the one or more mutations introduced by the donor polynucleotide comprises substitutions, deletions, insertions, or a combination thereof.

67. The method of claim 65, wherein the one or more mutations causes a shift in an open reading frame on the target polynucleotide.

68. The method of any one of claims 63-67, wherein the cell is a prokaryotic cell.

69. The method of any one of claims 63-67, wherein the cell is a eukaryotic cell.

70. The method of any one of claims 63-67, wherein the cell is a mammalian cell, a cell of a non-human primate, or a human cell.

71. The method of any one of claims 63-67, wherein the cell is a plant cell.

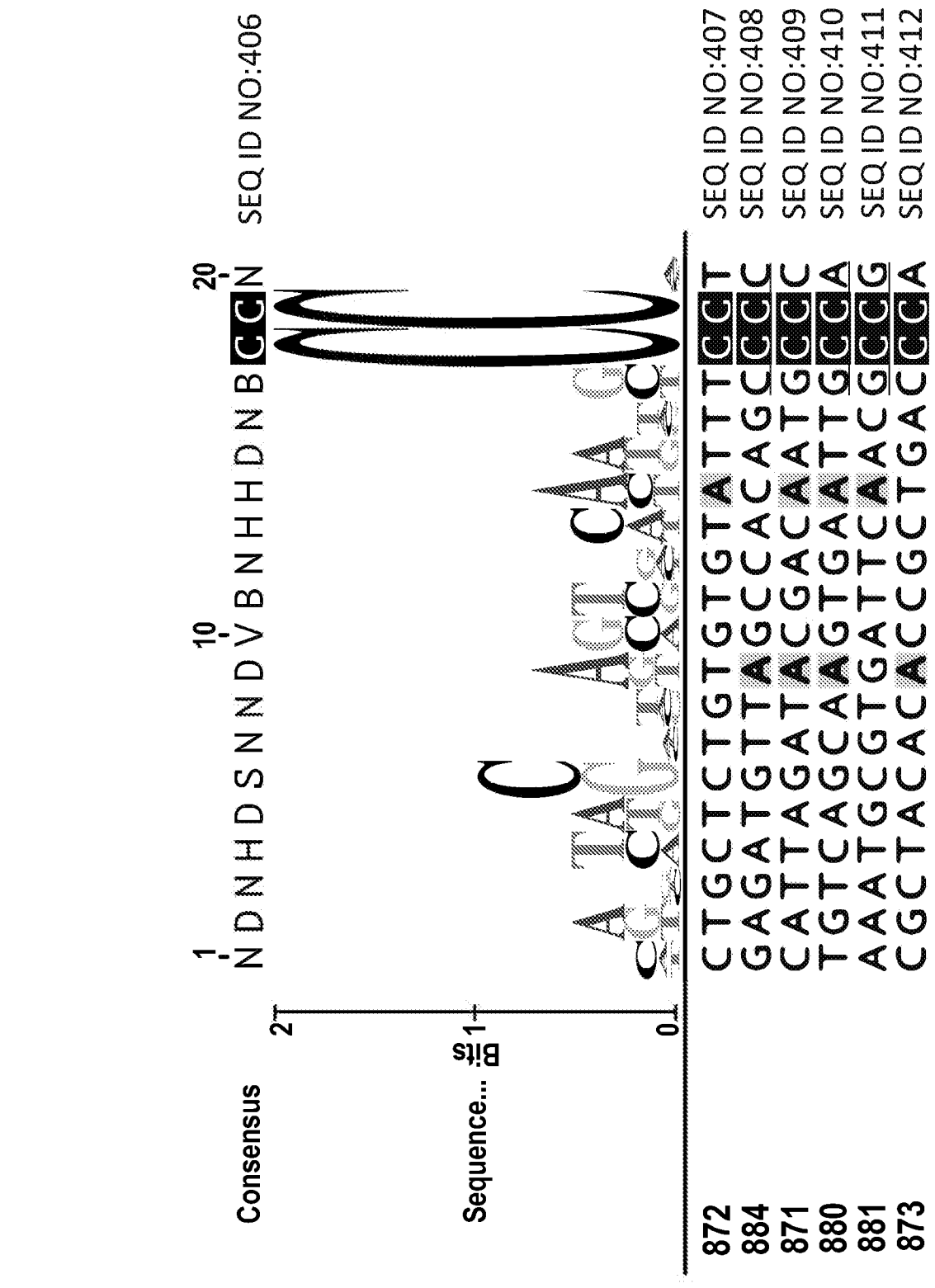

Consensus
Sequence...
Bits
2
1
0
1
10
20
N D N H D S N N D V B N H H D N B C G N
SEQ ID NO:406
872 CTGCTTCTTGTGTGTATTTTCGT SEQ ID NO:407
884 GAGATGTTTAGCCACACAGCCCC SEQ ID NO:408
871 CATTAGATACGACAAATGGCCC SEQ ID NO:409
880 TGTTCAGGCAAGTGAAATTGCCA SEQ ID NO:410
881 AAATGCGTGATTCAACGGCCG SEQ ID NO:411
873 CGCTACACACCGCTGACCCA SEQ ID NO:412

FIG. 1

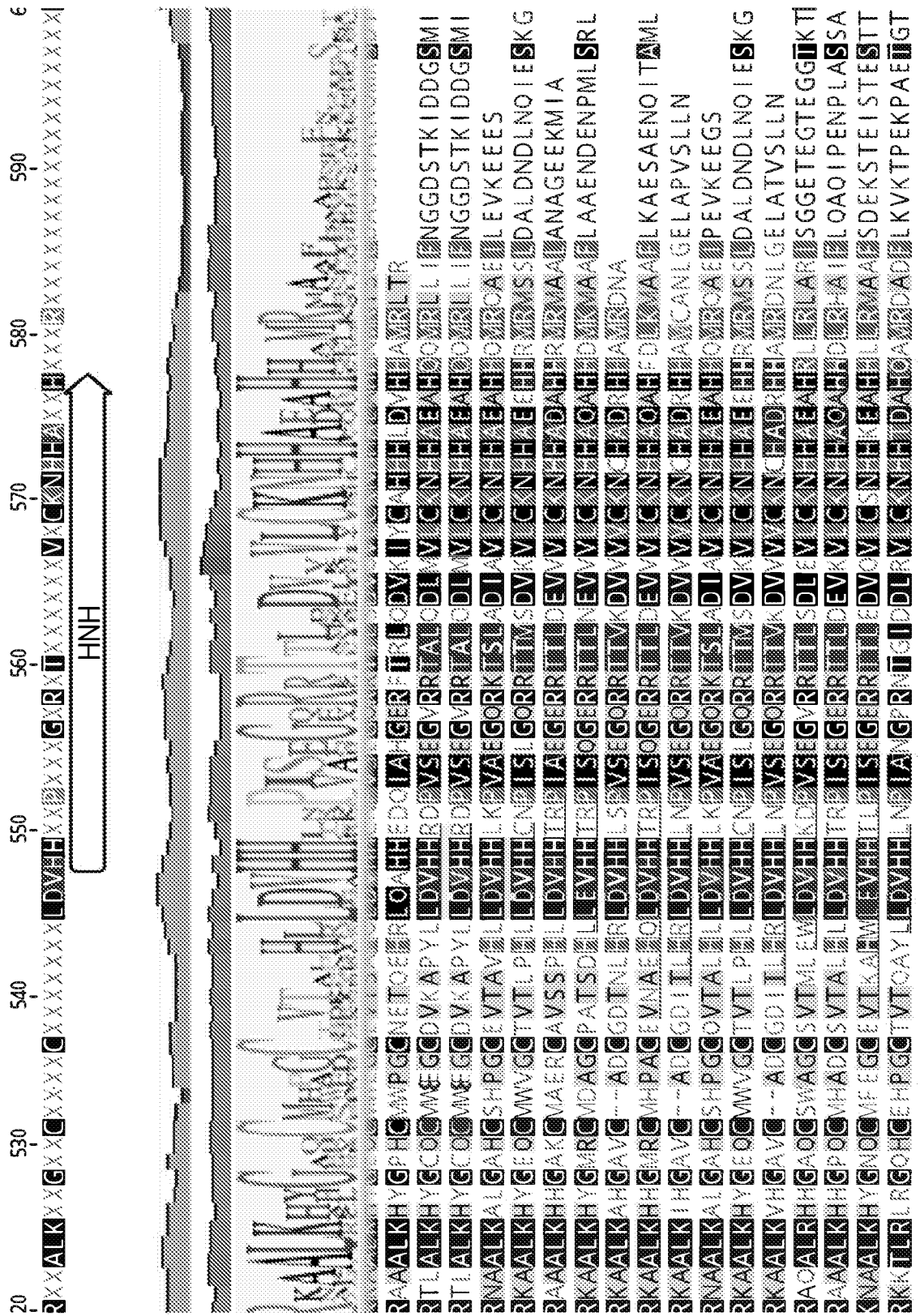


FIG. 2

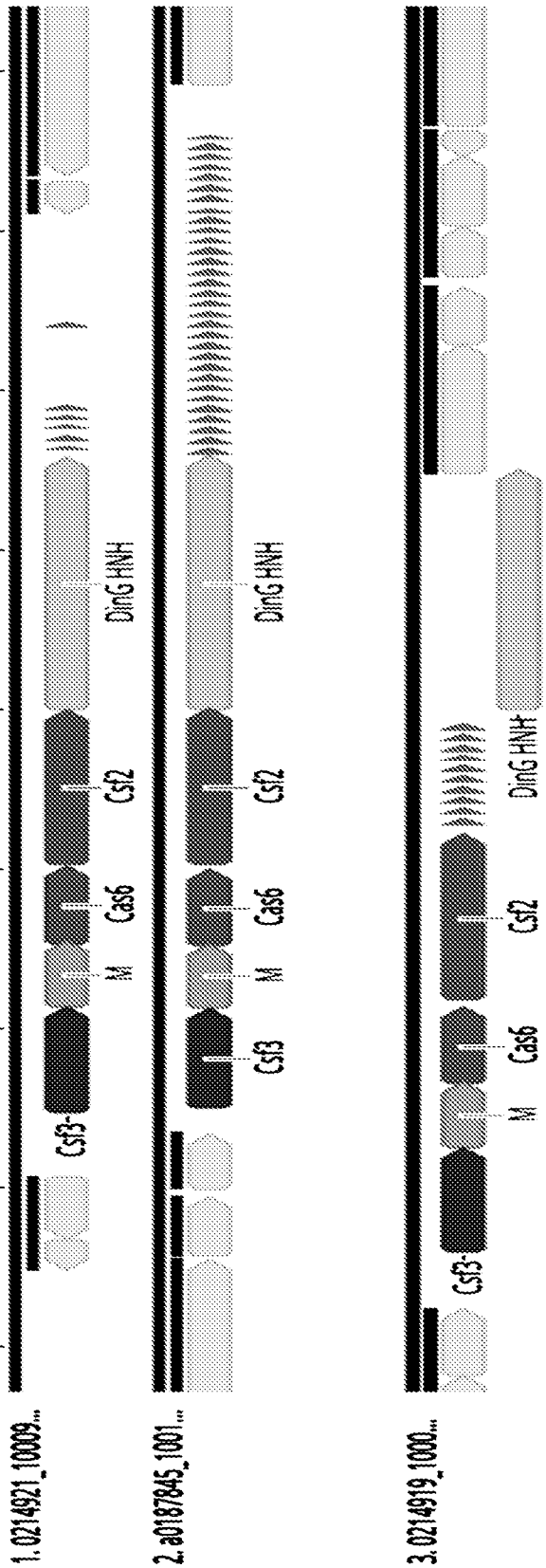

21,000
22,000
23,000
24,000
25,000
26,000
27,000
28,000
29,000
1. 0214921_10009...
Csf3-
M
Cas6
Csf2
DinG HNH
2. a0187845_1001...
Csf3
M
Cas6
Csf2
DinG HNH
3. 0214919_1000...
Csf3-
M
Cas6
Csf2
DinG HNH

FIG. 3

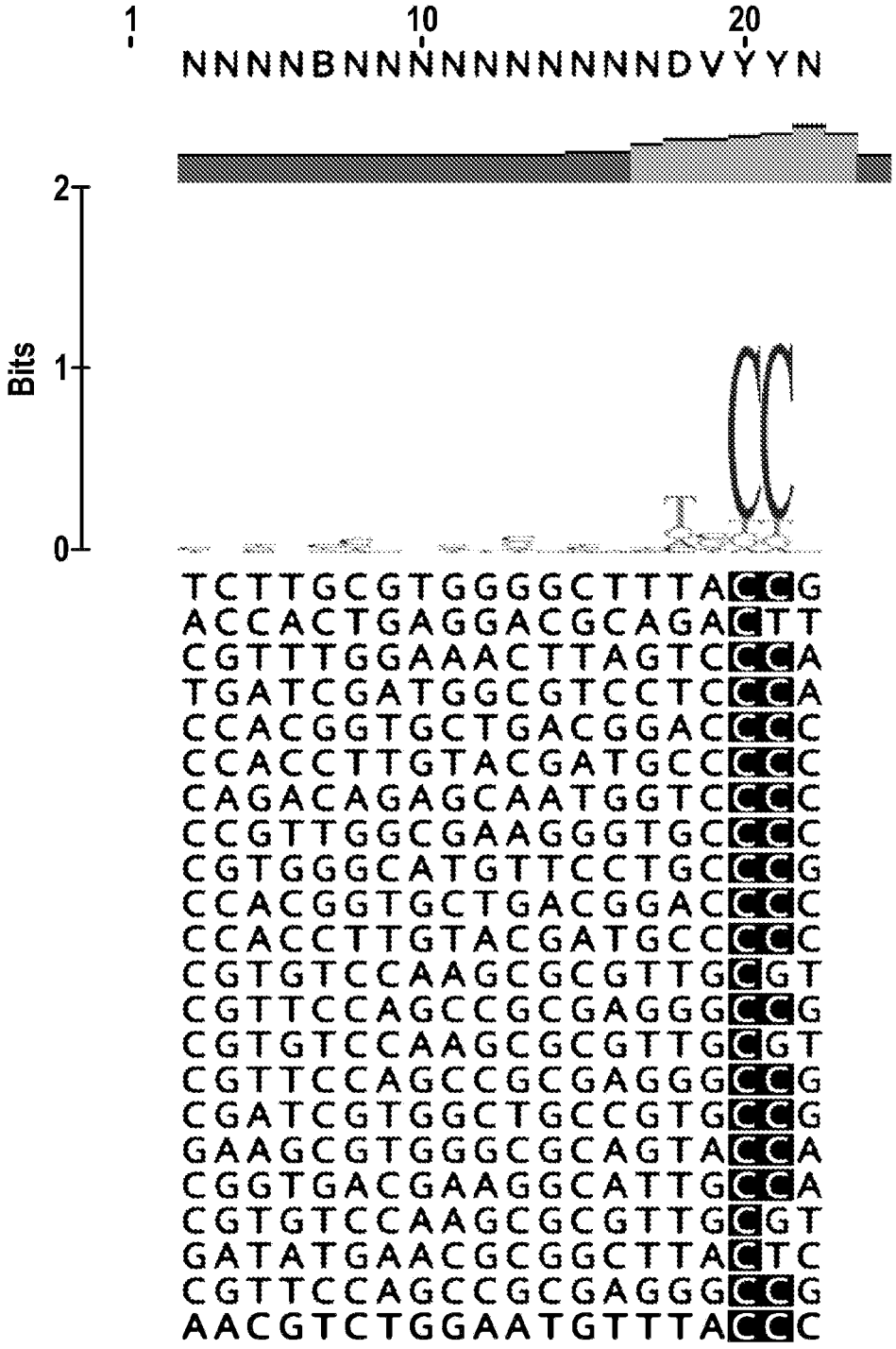

1
10
20
NNNNBNNNNNNNNNDVYYN
Bits
2
1
0
TCTTGCGTGGGGCTTTACCG
ACCACTGAGGACGCAGACTT
CGTTTGGAAACTTAGTCCCA
TGATCGATGGCGTCCTCCCA
CCACGGTGCTGACGGACCCC
CCACCTTGTACGATGCCCCC
CAGACAGAGCAATGGTCCCC
CCGTTGGCGAAGGGTGCCCC
CGTGGGCATGTTCCTGCCCG
CCACGGTGCTGACGGACCCC
CCACCTTGTACGATGCCCCC
CGTGTCCAAGCGCGTTGCGT
CGTTCCAGCCGCGAGGGCCG
CGTGTCCAAGCGCGTTGCGT
CGTTCCAGCCGCGAGGGCCG
CGATCGTGGCTGCCGTGCCG
GAAGCGTGGGCGCAGTACCA
CGGTGACGAAGGCATTGCCA
CGTGTCCAAGCGCGTTGCGT
GATATGAACGCGGCTTACTC
CGTTCCAGCCGCGAGGGCCG
AACGTCTGGAATGTTTACCC

FIG. 4

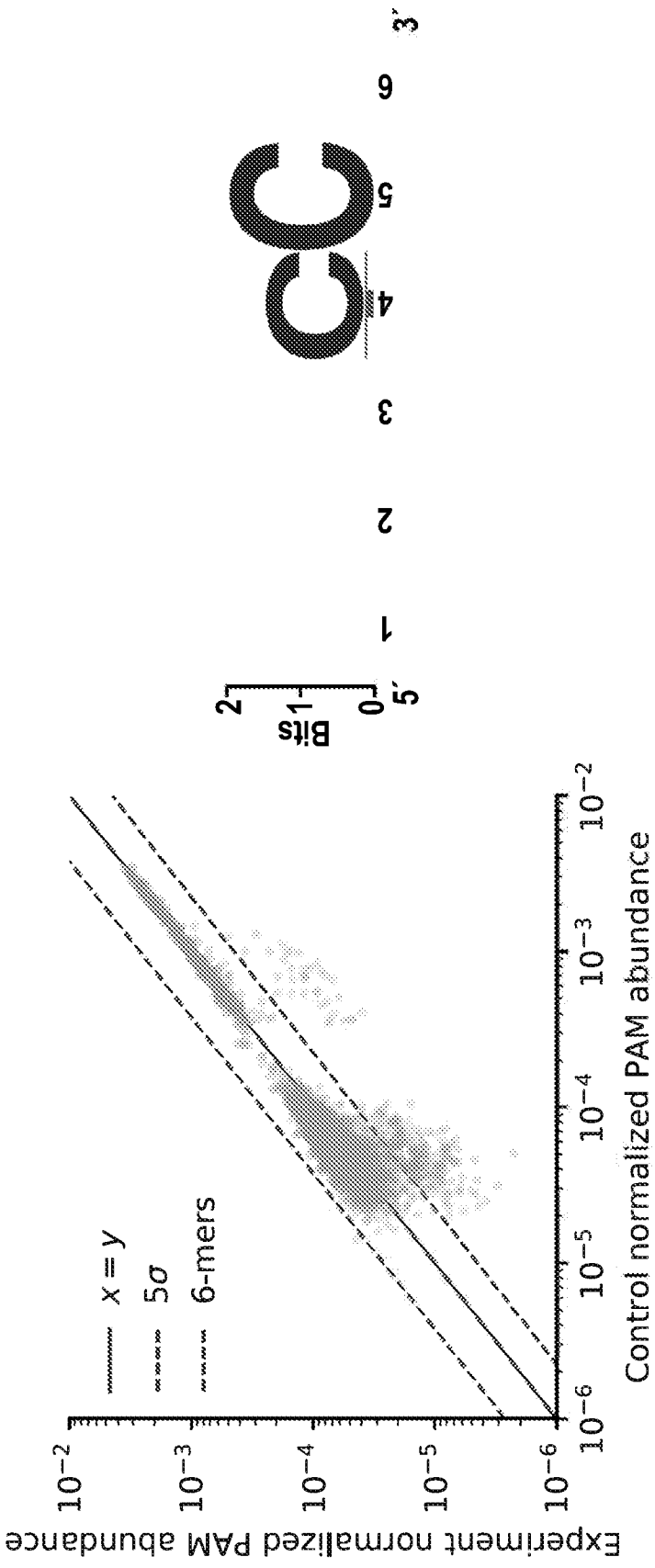
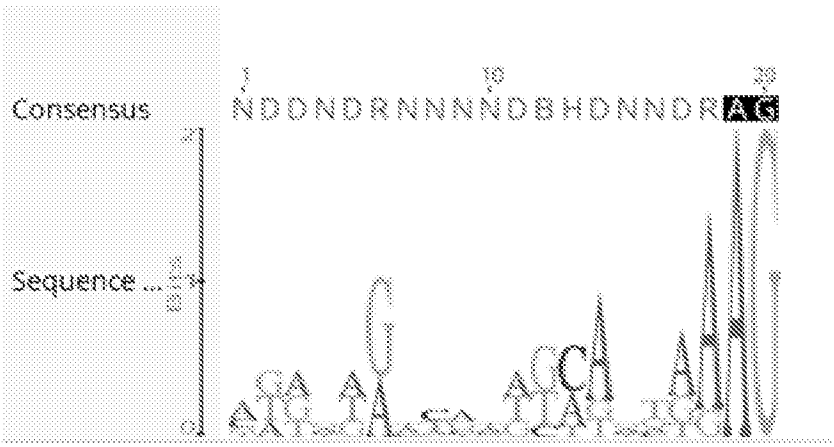

Experiment normalized PAM abundance
Control normalized PAM abundance
10^{-2}
10^{-3}
10^{-4}
10^{-5}
10^{-6}
$x = y$
5σ
6-mers
Bits
2
1
0
5′
1
2
3
4
5
6
3′
C
C

FIG. 5



2965	T T A T T G G G A T T G A A C T T T A G	SEQ ID NO:1123
2971	A G G A T A C C C G A T C A A C A T A G	SEQ ID NO:1124
2970	A T G G A G T C G T T G A A G T A T A G	SEQ ID NO:1125
2968	A G T C A G A T A A G T C T T G A T A G	SEQ ID NO:1126
2973	C G A G T A T T T C A G C A A A G T A G	SEQ ID NO:1127
2972	A A A C G A G T A T T T C A G T A C A G	SEQ ID NO:1128
2976	G T G A A G A A G A A C A A G G A T A G	SEQ ID NO:1129
2977	C A T G G G T C G C G G T G C T G T A G	SEQ ID NO:1130

FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 21/19494

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 15/90 (2021.01)

CPC - C12N 15/907, C12N 9/22, A61K 48/00, C12N 15/113, C12N 15/63, C12N 15/86

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2019/0021343 A1 (NORTH CAROLINA STATE UNIVERSITY) 24 January 2019 (24.01.2019); para [0011], [0014], [0066], [0072], [0075], [0139], [0166], [0174]; claim 1	1-4, 18-23, 25-29, 34-39, 56-58, 60, 63-67 ---- 40-41
Y	US 2016/0215300 A1 (PIONEER HI-BRED INTERNATIONAL, INC.) 28 July 2016 (28.07.2016); para [0049]	40-41

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"D" document cited by the applicant in the international application
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
29 April 2021	MAY 20 2021
Name and mailing address of the ISA/US Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Lee Young Telephone No. PCT Helpdesk: 571-272-4300

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 21/19494

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:

      ☒ in the form of an Annex C/ST.25 text file.

      ☐ on paper or in the form of an image file.

   b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

   c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

      ☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

      ☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US 21/19494

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. ☒ Claims Nos.: 5-17, 24, 30-33, 42-55, 59, 61-62, 68-71
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

☐ No protest accompanied the payment of additional search fees.