COMPOSITIONS AND METHODS OF USE OF CRISPR-CAS SYSTEMS IN NUCLEOTIDE REPEAT DISORDERS

RELATED APPLICATIONS AND INCORPORATION BY REFERENCE

[0001] This application claims priority from US provisional patent applications Serial Nos. 61/915,150, filed December 12, 2013; and 62/010,888 and 62/010,879, both filed June 11, 2014.

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

FIELD OF THE INVENTION

[0003] The present invention generally relates to the delivery, engineering, optimization and therapeutic applications of systems, methods, and compositions used for the control of gene expression involving sequence targeting, such as genome perturbation or gene-editing, that relate to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and components thereof. The invention relates to delivery, use, control and therapeutic applications of CRISPR-Cas systems and compositions, for brain and central nervous system (CNS) disorders and diseases. The invention relates to delivery, use, control and therapeutic applications of CRISPR-Cas systems and compositions, for nucleotide repeat elements (e.g., trinucleotide repeat, tetranucleotide repeat, nucleotide expansion elements) disorders and diseases.

STATEMENT AS TO FEDERALLY SPONSORED RESEARCH

[0004] This invention was made with government support under the NIH Pioneer Award (1DP1MH100706) awarded by the National Institutes of Health. The government has certain rights in the invention.

BACKGROUND OF THE INVENTION

[0005] Recent advances in genome sequencing techniques and analysis methods have significantly accelerated the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. Precise genome targeting technologies are needed to

enable systematic reverse engineering of causal genetic variations by allowing selective perturbation of individual genetic elements, as well as to advance synthetic biology, biotechnological, and medical applications. Although genome-editing techniques such as designer zinc fingers, transcription activator-like effectors (TALEs), or homing meganucleases are available for producing targeted genome perturbations, there remains a need for new genome engineering technologies that are affordable, easy to set up, scalable, and amenable to targeting multiple positions within the eukaryotic genome.

SUMMARY OF THE INVENTION

[0006] The invention provides in an aspect, a non-naturally occurring or engineered self-inactivating CRISPR-Cas composition comprising:

I. a first regulatory element operably linked to a CRISPR-Cas system RNA polynucleotide sequence, wherein the polynucleotide sequence comprises:

(a) at least one first guide sequence capable of hybridizing to a target DNA,

(b) at least one tracr mate sequence, and

(c) at least one tracr sequence, and

wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

II. a second regulatory element operably linked to a polynucleotide sequence encoding a CRISPR enzyme,

wherein parts I and II comprise a first CRISPR-Cas system, and wherein,

III. the composition further comprises

(a) at least one second guide sequence capable of hybridizing to a sequence in or of the CRISPR-Cas system,

(b) at least one tracr mate sequence, and

(c) at least one tracr sequence, and

wherein (a), (b) and (c) are arranged in a 5' to 3' orientation, wherein parts I and III comprise a second CRISPR-Cas system,

and said composition when transcribed comprises

a first CRISPR complex comprising the CRISPR enzyme complexed with (1) the first guide sequence that can be hybridized or can be hybridizable to the target sequence, and (2) the tracr mate sequence that can be hybridized to the tracr sequence,

a second CRISPR complex comprising a CRISPR enzyme complexed with (1) the second guide sequence that can be hybridized or hybridizable to a sequence of a polynucleotide comprising or

encoding the CRISPR-Cas system, and (2) the tracr mate sequence that can be hybridized to the tracr sequence,

wherein the first guide sequence directs sequence-specific binding of a first CRISPR complex to the target DNA, and

wherein the second guide sequence directs sequence-specific binding of a second CRISPR complex to a sequence comprising a polynucleotide comprising or encoding a component of the CRISPR-Cas system and whereby there can be diminished activity of the first CRISPR-Cas system over a period of time, and the CRISPR-Cas composition can be self-inactivating ("SIN CRISPR-Cas composition").

[0007] The target DNA sequence can be within a cell. The cell can be a eukaryotic cell, or a prokaryotic cell. The composition the first CRISPR-Cas system and/or the second CRISPR-Cas system can be codon optimized, e.g., for a eukaryotic cell. Part II can include coding for one or more nuclear localization signals (NLSs). Part I can be encoded by a first viral vector and part II can be encoded by a second viral vector. The first and second viral vectors can be lentiviral vectors or recombinant AAV. The recombinant AAV genome can comprise inverted terminal repeats (iTRs). Expression of the CRISPR enzyme can be driven by the inverted terminal repeat (iTR) in the AAV genome. The first regulatory element can be a RNA polymerase type III promoter and the second regulatory element can be a RNA polymerase type III promoter. The first regulatory element can be a U6 promoter or a H1 promoter. The second regulatory element can be a ubiquitous expression promoter or a cell-type specific promoter. There can be a selection marker comprising a FLAG-tag. The CRISPR enzyme can comprise a C-terminal NLS and an N-terminal NLS. The composition can be delivered via injection. The composition or a part thereof can be delivered via a liposome, a nanoparticle, an exosome, a microvesicles. 17. The composition can have the first guide sequence directing sequence-specific binding of the first CRISPR complex to the target DNA sequence and alters expression of a genomic locus in the cell. The composition can have wherein the first CRISPR complex mediating binding to or a double or single stranded DNA break, thereby editing a genomic locus in the cell. 19. The composition of any of the preceding claims, wherein the first and/or second CRISPR-Cas system can be a multiplexed CRISPR enzyme system further comprising multiple chimeras and/or multiple multiguide sequences and a single tracr sequence. In the composition the first CRISPR-Cas system can be a multiplexed CRISPR enzyme system to minimize off-target activity. The composition according any of the preceding claims, wherein the CRISPR enzyme can be a nickase. The CRISPR enzyme can comprise one or more mutations. The one or more mutations

can be selected from D10A, E762A, H840A, N854A, N863A or D986A. The one or more mutations can be in a RuvC1 domain of the CRISPR enzyme. The CRISPR complex mediates genome engineering that includes: modifying a target polynucleotide or expression thereof, knocking out a gene, amplifying or increasing or decreasing expression of a polynucleotide or gene, or repairing a mutation, or editing by inserting a polynucleotide. The CRISPR enzyme further comprises a functional domain. The CRISPR enzyme can be a Cas9. The second complex can binds to a sequence for CRISPR enzyme expression. The second guide sequence can be capable of hybridizing to (a) a sequence encoding the RNA or (b) a sequence encoding the CRISPR enzyme, or (c) a non-coding sequence comprising i) a sequence within a regeulatory element driving expression of non-coding RNA elements, ii) a sequence within a regulatory element driving expression of the CRISPR enzyme, iii) a sequence within 100bp of the ATG translational start codon of the CRISPR enzyme coding sequence, and iv) a sequence within an inverted terminal repeat of a viral vector. The second guide sequence can be expressed singularly to achieve inactivation of the first CRISPR-Cas system. The second CRISPR complex induces a frame shift in CRISPR enzyme coding sequence causing a loss of protein expression. The second guide sequence targets an iTR, wherein expression will result in the excision of an entire CRISPR-Cas cassette. The second guide sequence can be expressed in an array format to achieve inactivation of the first CRISPR-Cas9 system. The second guide sequences can be expressed in array format and targets both regulatory elements, thereby excising intervening nucleotides from within the first CRISPR-Cas system, effectively leading to its inactivation. The expression of the second guide sequences can be driven by a U6 promoter. The self-inactivation of the first CRISPR-Cas system limits duration of its activity and/or expression in targeted cells. Transient expression of the CRISPR enzyme can be normally lost within 48 hours. The invention also comprehends a non-naturally occurring or engineered composition comprising the first and second CRISPR complexes.

[0008] With respect to mutations of the CRISPR enzyme, when the enzyme is not SpCas9, mutations may be made at any or all residues corresponding to positions 10, 762, 840, 854, 863 and/or 986 of SpCas9 (which may be ascertained for instance by standard sequence comparison tools). In particular, any or all of the following mutations are preferred in SpCas9: D10A, E762A, H840A, N854A, N863A and/or D986A; as well as conservative substitution for any of the replacement amino acids is also envisaged. In an aspect the invention provides as to any or each or all embodiments herein-discussed wherein the CRISPR enzyme comprises at least one or more, or at least two or more mutations, wherein the at least one or more mutation or the at least

two or more mutations is as to D10, E762, H840, N854, N863, or D986 according to SpCas9 protein, e.g., D10A, E762A, H840A, N854A, N863A and/or D986A as to SpCas9, or N580 according to SaCas9, e.g., N580A as to SaCas9, or any corresponding mutation(s) in a Cas9 of an ortholog to Sp or Sa, or the CRISPR enzyme comprises at least one mutation wherein at least H840 or N863A as to Sp Cas9 or N580A as to Sa Cas9 is mutated; e.g., wherein the CRISPR enzyme comprises H840A, or D10A and H840A, or D10A and N863A, according to SpCas9 protein, or any corresponding mutation(s) in a Cas9 of an ortholog to Sp protein or Sa protein.

[0009] The invention in an aspect provides a method of treating or inhibiting a condition in a cell or tissue having a nucleotide element or trinucleotide repeat or other nucleic acid repeat element that gives rise to an adverse or disease condition caused by a defect in a genomic locus of interest in a cell in a subject or a non-human subject in need thereof comprising modifying the subject or a non-human subject by editing the genomic locus and wherein the condition can be susceptible to treatment or inhibition by editing the genomic locus comprising providing treatment comprising: delivering the non-naturally occurring or engineered composition of the invention.

[0010] The invention in an aspect provides use of a composition of the invention in the manufacture of a medicament for *ex vivo* gene or genome editing or for use in a method of modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest or in a method of treating or inhibiting a condition caused by a defect in a target sequence in a genomic locus of interest.

[0011] In a method or use of the invention, part III can be introduced into the cell sequentially or at a time point after the introduction of parts I and II.

[0012] In a use, composition or method of the invention, the RNA can be chimeric RNA (chiRNA).

[0013] The invention provides for use of a SIN CRISPR-Cas composition of any of the preceding claims or as disclosed herein for genome engineering or for a treatment of a condition or for preparing a medicament or pharmaceutical composition.

[0014] The invention also provides a non-naturally occurring or engineered RNA that can be a first CRISPR-Cas system or first CRISPR-Cas complex guide sequence capable of hybridizing to an RNA sequence of a second CRISPR-Cas system or a nucleic acid molecule for expression of a component of the second CRISPR-Cas complex, to diminish or eliminate functional expression of the second system or complex, whereby the first and/or second system or complex can be Self-Inactivating.

[0015] In an aspect the invention provides use of a SIN CRISPR-Cas composition or first and second CRISPR-Cas complexes of any of the preceding claims or as disclosed herein for genome engineering or for a treatment of a condition or for preparing a medicament or pharmaceutical composition. The genome engineering can include: modifying a target polynucleotide or expression thereof, knocking out a gene, amplifying or increasing or decreasing expression of a polynucleotide or gene, or repairing a mutation, or editing by inserting a polynucleotide.

[0016] In an aspect the invention provides a non-naturally occurring or engineered composition for use in a cell having a defective nucleotide element or trinucleotide repeat or other nucleotide repeat element or nucleotide expansion, the comprising:

A.

I. a first regulatory element operably linked to a CRISPR-Cas system RNA polynucleotide sequence, wherein the polynucleotide sequence comprises:

(a) at least one first guide sequence capable of hybridizing to a target DNA within the cell,

(b) at least one tracr mate sequence, and

(c) at least one tracr sequence, and

wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

II. a second regulatory element operably linked to a polynucleotide sequence encoding a CRISPR enzyme,

wherein parts A.I and A.II comprise a CRISPR-Cas system, and wherein, said composition when transcribed comprises

a CRISPR complex comprising the CRISPR enzyme complexed with (1) the guide sequence that can be hybridized or can be hybridizable to the target sequence, and (2) the tracr mate sequence that can be hybridized to the tracr sequence,

wherein the guide sequence directs sequence-specific binding of a CRISPR complex to the target DNA, and mediates impact or repair of the defect;

or,

B.

I. a CRISPR-Cas system RNA polynucleotide sequence, wherein the polynucleotide sequence comprises:

(a) at least one guide sequence capable of hybridizing to a target sequence in a eukaryotic cell,

(b) at least one tracr mate sequence, and

(c) at least one tracr sequence, and

wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

II. a CRISPR enzyme,

wherein parts B.I and B.II comprise the CRISPR complex.

[0017] The cell can be a eukaryotic cell or a prokaryotic cell. The CRISPR-Cas system can be codon optimized. Part A.II can include coding for one or more nuclear localization signals (NLSs); or part B.II can include one or more NLSs. Part A.I can be encoded by a first viral vector and/or part A.II can be encoded by a second viral vector. The first and second viral vectors can be lentiviral vectors or recombinant AAV. The recombinant AAV genome can comprise inverted terminal repeats (iTRs). The expression of the CRISPR enzyme can be driven by the inverted terminal repeat (iTR) in the AAV genome. The first regulatory element can be a RNA polymerase type III promoter and the second regulatory element can be a RNA polymerase type III promoter. The first regulatory element can be a U6 promoter or a H1 promoter. The second regulatory element can be a ubiquitous expression promoter or a cell-type specific promoter. There can be a selection marker comprising a FLAG-tag. The CRISPR enzyme can comprise a C-terminal NLS and an N-terminal NLS. The composition can be delivered via injection. The composition or a part thereof can be delivered via a liposome, a nanoparticle, an exosome, or a microvesicle. The guide sequence can direct sequence-specific binding of the CRISPR complex to the target DNA sequence and alters expression of a genomic locus in the cell. The CRISPR complex can mediate binding to or a double or single stranded DNA break, and there can optionally be insertion of DNA, whereby there can be editing of a genomic locus in the cell. The CRISPR-Cas system can be a multiplexed CRISPR enzyme system further comprising multiple chimeras and/or multiple multiguide sequences and a single tracr sequence. The CRISPR-Cas system can be a multiplexed CRISPR enzyme system to minimize off-target activity. The CRISPR enzyme can be a nickase. The CRISPR enzyme can comprise one or more mutations. The CRISPR enzyme comprises one or more mutations selected from D10A, E762A, H840A, N854A, N863A or D986A. The one or more mutations can be in a RuvC1 domain of the CRISPR enzyme. The CRISPR enzyme further comprises a functional domain. The composition of the CRISPR complex can mediate genome engineering that includes: modifying a target polynucleotide or expression thereof, knocking out a gene, amplifying or increasing or decreasing expression of a polynucleotide or gene, or repairing a mutation, or editing by inserting a polynucleotide. The CRISPR enzyme can be a Cas9. The CRISPR complex can mediate at least one double stranded DNA break thereby causing editing of the target DNA. The cell can be a mammalian brain or central nervous tissue cell. The nucleotide repeat element can

be selected from one or more of: a trinucleotide repeat comprising CTG, CAG, CGG, CCG, GAA, or TTC; a tetranucleotide repeat comprising CCTG, a pentanucleotide repeat comprising ATTCT or AGAAT; a hexanucleotide repeat comprising GGGGCC; and a dodecanucleotide repeat comprising CCCCGCCCCGCG or CGCGGGGCGGGG. The defect gives rise to a condition selected from one or more of: a Fragile X (FXS); Fragile X Tremor Ataxia (FXTAS); Unverricht-Lundborg disease (EPM1); Spinocerebellar ataxia type-12 (SCA12); Amyotrophic Lateral Scleroscan be (ALS); Fronto Temporal Dementia (FTD); Friedreich Ataxia; Myotonic Dystrophy type-1 (DM1); Myotonic Dystrophy type-2 (DM2); Spinocerebellar ataxia type-8 (SCA8); Spinocerebellar ataxia type-10 (SCA10); Spinocerebellar ataxia type-31 (SCA31); Oculopharyngeal muscular dystrophy (OPMD); Spinocerebellar ataxia type-1 (SCA1); Spinocerebellar ataxia type-2 (SCA2); Spinocerebellar ataxia type-3 (SCA3); Spinocerebellar ataxia type-6 (SCA6); Spinocerebellar ataxia type-7 (SCA7); Spinocerebellar ataxia type-17 (SCA17); Dentatorubral-pallidoluysian atrophy (DRPLA); Spinobulbar muscular atrophy (SBMA); Huntington's disease like type-2 (HDL2) and Huntington's Disease (HD).

[0018] The invention comprehends in an aspect a method of treating or inhibiting a condition in a cell having a defective nucleotide element or trinucleotide repeat or other nucleotide repeat element or nucleotide expansion, comprising delivering the non-naturally occurring or engineered composition of the invention. The invention also comprehends use of a composition of the invention to treat a disease or disorder. The invention additionally comprehends use of a composition of the invention to treat disease or disorder wherein the disease or disorder comprises a brain disease or disorder or a central nervous system disease or disorder. The invention further comprehends use of a composition of the invention in the manufacture of a medicament for *ex vivo* gene or genome editing or for use in a method of modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest or in a method of treating or inhibiting a condition. The condition can comprise a brain disease or disorder or a central nervous system disease or disorder. In any method, use or composition of any of the invention, the CRISPR-Cas system RNA can be a chimeric RNA (chiRNA). Also, in any method, use or composition of the invention, there can be at least one second guide sequence capable of hybridizing to an RNA sequence of the CRISPR-Cas system or a nucleic acid molecule for expression of a component of the CRISPR-Cas complex, to diminish or eliminate functional expression of the system or complex, whereby the system or complex can be Self-Inactivating; and, the second guide sequence can be capable of hybridizing to a nucleic acid molecule for expression of the CRISPR enzyme.

[0019] The invention involves the development and application of the CRISPR-Cas9 system as a tool for editing disease-causing nucleotide repeat expansions in the human genome. Applicants provide evidence that the sequences, plasmids and/or viral vectors that Applicants have designed and tested facilitate genomic editing of nucleotide repeat sequences at a number of disease-linked genomic loci including those associated with CAG triplet repeat disorders (i.e. Polyglutamine diseases), Fragile X and Fragile X-associated tremor/ataxia syndrome (FXTAS) and to other nucleotide repeat disorders or nucleotide expansion disorders as provided herein. Moreover, Applicants describe the design and application of CRISPR-Cas9 to the mammalian brain (and other tissues or organs of the central nervous system) using Adeno Associated Virus (AAV) as a vector. Finally, the invention also discloses a method for the self-inactivation of the Cas9 nuclease as means to limit the duration of its expression in targeted cells.

[0020] The CRISPR-Cas system does not require the generation of customized proteins to target specific sequences but rather a single Cas enzyme can be programmed by a short RNA molecule to recognize a specific DNA target. Adding the CRISPR-Cas system to the repertoire of genome sequencing techniques and analysis methods may significantly simplify the methodology and accelerate the ability to catalog and map genetic factors associated with a diverse range of biological functions and diseases. To utilize the CRISPR-Cas system effectively for genome editing without deleterious effects, it is critical to understand aspects of engineering, optimization and cell-type/tissue/organ specific delivery of these genome engineering tools, which are aspects of the claimed invention.

[0021] There exists a pressing need for alternative and robust systems and techniques for nucleic sequence targeting with a wide array of applications. Aspects of this invention address this need and provide related advantages. An exemplary CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within the target polynucleotide. The guide sequence is linked to a tracr mate sequence, which in turn hybridizes to a tracr sequence.

[0022] In one aspect, the invention provides methods for using one or more elements of a CRISPR-Cas system. The CRISPR complex of the invention provides an effective means for modifying a target polynucleotide. The CRISPR complex of the invention has a wide variety of utilities including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target polynucleotide in a multiplicity of cell types in various tissues and organs. As such the CRISPR complex of the invention has a broad spectrum of applications in, e.g., gene or genome editing, gene therapy, drug discovery, drug screening, disease diagnosis, and prognosis.

[0023] Aspects of the invention relate to Cas9 enzymes having improved targeting specificity in a CRISPR-Cas9 system having guide RNAs having optimal activity, smaller in length than wild-type Cas9 enzymes and nucleic acid molecules coding therefor, and chimeric Cas9 enzymes, as well as methods of improving the target specificity of a Cas9 enzyme or of designing a CRISPR-Cas9 system comprising designing or preparing guide RNAs having optimal activity and/or selecting or preparing a Cas9 enzyme having a smaller size or length than wild-type Cas9 whereby packaging a nucleic acid coding therefor into a delivery vector is more advanced as there is less coding therefor in the delivery vector than for wild-type Cas9, and/or generating chimeric Cas9 enzymes.

[0024] Also provided are uses of the present sequences, vectors, enzymes or systems, in medicine. Also provided are uses of the same in gene or genome editing.

[0025] In an additional aspect of the invention, a Cas9 enzyme may comprise one or more mutations and may be used as a generic DNA binding protein with or without fusion to a functional domain. The mutations may be artificially introduced mutations or gain- or loss-of-function mutations. The mutations may include but are not limited to mutations in one of the catalytic domains (D10 and H840) in the RuvC and HNH catalytic domains, respectively. Further mutations have been characterized and may be used in one or more compositions of the invention. In one aspect of the invention, the mutated Cas9 enzyme may be fused to a protein domain, e.g., such as a transcriptional activation domain. In one aspect, the transcriptional activation domain may be VP64. In other aspects of the invention, the transcriptional repressor domain may be KRAB or SID4X. Other aspects of the invention relate to the mutated Cas9 enzyme being fused to domains which include but are not limited to a transcriptional activator, repressor, a recombinase, a transposase, a histone remodeler, a demethylase, a DNA methyltransferase, a cryptochrome, a light inducible/controllable domain or a chemically inducible/controllable domain.

[0026] In a further embodiment, the invention provides for methods to generate mutant tracrRNA and direct repeat sequences or mutant chimeric guide sequences that allow for enhancing performance of these RNAs in cells. Aspects of the invention also provide for selection of said sequences.

[0027] Aspects of the invention also provide for methods of simplifying the cloning and delivery of components of the CRISPR complex. In the preferred embodiment of the invention, a suitable promoter, such as a Pol III promoter such as a U6 promoter, is amplified with a DNA oligo and added onto the guide RNA. The promoter can thus be positioned upstream, e.g.,

contiguous to and upstream, of a sequence encoding the guide RNA The resulting PCR product can then be transfected into cells to drive expression of the guide RNA. Aspects of the invention also relate to the guide RNA being transcribed *in vitro* or ordered from a synthesis company and directly transfected.

[0028] In one aspect, the invention provides for methods to improve activity by using a more active polymerase. In one aspect, a T7 promoter may be inserted upstream, e.g., contiguous to and upstream, of a sequence encoding a guide RNA. In a preferred embodiment, the expression of guide RNAs under the control of the T7 promoter is driven by the expression of the T7 polymerase in the cell. In an advantageous embodiment, the cell is a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a human cell. In a more preferred embodiment the human cell is a patient specific cell, e.g., a cell removed from a patient that may be modified and/or expanded into a cell population or a modified cell population, for instance, for re-administration to the patient.

[0029] In one aspect, the invention provides for methods of reducing the toxicity of Cas enzymes. In certain aspects, the Cas enzyme is any Cas9 as described herein, for instance any naturally-occurring bacterial Cas9 as well as any chimaeras, mutants, homologs or orthologs. In one aspect, the Cas enzyme is a nickase. In an embodiment, the Cas9 is delivered into the cell in the form of a nucleic acid molecule, e.g., DNA, RNA, mRNA. This allows for the transient expression of the enzyme thereby reducing toxicity. In another embodiment, the Cas9 is delivered into the cell in the nucleotide construct that encodes and expresses the Cas9 enzyme. In another embodiment, the invention also provides for methods of expressing Cas9 under the control of an inducible promoter, and the constructs used therein.

[0030] In another aspect, the invention provides for methods of improving the *in vivo* applications of the CRISPR-Cas system. In the preferred embodiment, the Cas enzyme is wildtype Cas9 or any of the modified versions described herein, including any naturally-occurring bacterial Cas9 as well as any chimaeras, mutants, homologs or orthologs. In one aspect, the Cas enzyme is a nickase. An advantageous aspect of the invention provides for the selection of Cas9 homologs that are easily packaged into viral vectors for delivery. Cas9 orthologs typically share the general organization of 3-4 RuvC domains and a HNH domain. The 5' most RuvC domain cleaves the non-complementary strand, and the HNH domain cleaves the complementary strand. All notations are in reference to the guide sequence.

[0031] The catalytic residue in the 5' RuvC domain is identified through homology comparison of the Cas9 of interest with other Cas9 orthologs (from S. pyogenes type II CRISPR

locus, S. thermophilus CRISPR locus 1, S. thermophilus CRISPR locus 3, and Franciscilla novicida type II CRISPR locus), and the conserved Asp residue (D10) is mutated to alanine to convert Cas9 into a complementary-strand nicking enzyme. Similarly, the conserved His and Asn residues in the HNH domains are mutated to Alanine to convert Cas9 into a non-complementary-strand nicking enzyme. In some embodiments, both sets of mutations may be made, to convert Cas9 into a non-cutting enzyme.

[0032] In some embodiments, the CRISPR enzyme is a type I or III CRISPR enzyme, preferably a type II CRISPR enzyme. This type II CRISPR enzyme may be any Cas enzyme. A preferred Cas enzyme may be identified as Cas9 as this can refer to the general class of enzymes that share homology to the biggest nuclease with multiple nuclease domains from the type II CRISPR system. Most preferably, the Cas9 enzyme is from, or is derived from, spCas9 or saCas9. By derived, Applicants mean that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as described herein

[0033] It will be appreciated that the terms Cas and CRISPR enzyme are generally used herein interchangeably, unless otherwise apparent. As mentioned above, many of the residue numberings used herein refer to the Cas9 enzyme from the type II CRISPR locus in *Streptococcus pyogenes* (annotated alternatively as SpCas9 or spCas9). However, it will be appreciated that this invention includes many more Cas9s from other species of microbes, such as SpCas9 or or from or derived from S. pyogenes, SaCas9 or or from or derived from S. aureus, St1Cas9 or or from or derived from S. thermophilus and so forth. However, it will be appreciated that this invention includes many more Cas9s from other species of microbes, such as SpCas9, SaCas9, St1Cas9 and so forth. Further examples are provided herein. The skilled person will be able to determine appropriate corresponding residues in Cas9 enzymes other than SpCas9 by comparison of the relevant amino acid sequences. Thus, where a specific amino acid replacement is referred to using the SpCas9 numbering, then, unless the context makes it apparent this is not intended to refer to other Cas9 enzymes, the disclosure is intended to encompass corresponding modifications in other Cas9 enzymes.

[0034] An example of a codon optimized sequence, in this instance optimized for humans (i.e. being optimized for expression in humans) is provided herein, see the SaCas9 human codon optimized sequence. 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 such as the brain, can be practiced from this disclosure and the knowledge in the art.

[0035] In further embodiments, the invention provides for methods of enhancing the function of Cas9 by generating chimeric Cas9 proteins. Chimeric Cas9 proteins chimeric Cas9s may be new Cas9 containing fragments from more than one naturally occurring Cas9. These methods may comprise fusing N-terminal fragments of one Cas9 homolog with C-terminal fragments of another Cas9 homolog. These methods also allow for the selection of new properties displayed by the chimeric Cas9 proteins.

[0036] It will be appreciated that in the present methods the modification may occur *ex vivo* or *in vitro*, for instance in a cell culture and in some instances not *in vivo*. In other embodiments, it may occur *in vivo*.

[0037] In one aspect, the invention provides a method of modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest comprising: delivering a non-naturally occurring or engineered composition comprising:

A) - I. a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises:

(a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell,

(b) a tracr mate sequence, and

(c) a tracr sequence, and

wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

II. a polynucleotide sequence encoding a CRISPR enzyme comprising one or more nuclear localization sequences,

wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and

wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or is hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence and the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA,

or

(B) I. a polynucleotide comprising:

(a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, and

(b) at least one or more tracr mate sequences,

II. a polynucleotide sequence encoding a CRISPR enzyme, and

III. a polynucleotide sequence comprising a tracr sequence,

wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and

wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence, and the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA.

[0038] In one aspect, the invention provides a non-naturally occurring or engineered composition for delivery to a cell or to one or more tissues containing cells having a nucleotide element or trinucleotide repeat or other nucleotide repeat element that gives rise to an adverse or disease condition, the composition comprising:

(A) I. a first regulatory element operably linked to a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises:

(a) at least one guide sequence capable of hybridizing to a target sequence in a eukaryotic cell,

(b) at least one tracr mate sequence, and

(c) at least one tracr sequence, and

wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

II. a second regulatory element operably linked to a polynucleotide sequence encoding a CRISPR enzyme comprising one or more nuclear localization sequences,

wherein the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and

wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or is hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence,

or

(B) I. a first regulatory element operably linked to a polynucleotide comprising:

(a) at least one guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, and

(b) at least one or more tracr mate sequences,

II. a second regulatory element operably linked to a polynucleotide sequence encoding a CRISPR enzyme, and

III. a third regulatory element operably linked to a polynucleotide sequence comprising a tracr sequence,

wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and

wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence, and the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA;

wherein the CRISPR complex mediates at least one double stranded DNA break thereby editing the targeted genomic locus in the cell.

[0039] In an embodiment for use in a eukaryotic cell, the vector system comprises a viral vector system, e.g., an AAV vector or AAV vector system or a lentivirus-derived vector system or a tobacco mosaic virus-derived system or an Agrobacterium Ti or Ri plasmid

[0040] Any or all of the polynucleotide sequence encoding a CRISPR enzyme, guide sequence, tracr mate sequence or tracr sequence, may be RNA, DNA or a combination of RNA and DNA. In one aspect, the polynucleotides comprising the sequence encoding a CRISPR enzyme, the guide sequence, tracr mate sequence or tracr sequence are RNA. In one aspect, the polynucleotides comprising the sequence encoding a CRISPR enzyme, the guide sequence, tracr mate sequence or tracr sequence are DNA. In one aspect, the polynucleotides are a mixture of DNA and RNA, wherein some of the polynucleotides comprising the sequence encoding one or more of the CRISPR enzyme, the guide sequence, tracr mate sequence or tracr sequence are DNA and some of the polynucleotides are RNA. In one aspect, the polynucleotide comprising the sequence encoding the CRISPR enzyme is a DNA and the guide sequence, tracr mate sequence or tracr sequence are RNA. The one or more polynucleotides comprising the sequence encoding a CRISPR enzyme, the guide sequence, tracr mate sequence or tracr sequence may be delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun.

[0041] It will be appreciated that where reference is made to a polynucleotide, where that polynucleotide is RNA and is said to 'comprise' a feature such as a tracr mate sequence, the RNA sequence includes the feature. Where the polynucleotide is DNA and is said to comprise a feature such as a tracr mate sequence, the DNA sequence is or can be transcribed into the RNA that comprises the feature at issue. Where the feature is a protein, such as the CRISPR enzyme, the DNA or RNA sequence referred to is, or can be, translated (and in the case of DNA

transcribed first). Furthermore, in cases where an RNA encoding the CRISPR enzyme is provided to a cell, it is understood that the RNA is capable of being translated by the cell into which it is delivered.

[0042] Accordingly, in certain embodiments the invention provides a method of modifying an organism, e.g., mammal including human or a non-human mammal or organism by manipulation of a target sequence in a genomic locus of interest comprising delivering a non-naturally occurring or engineered composition comprising a viral or plasmid vector system comprising one or more viral or plasmid vectors operably encoding a composition for expression thereof, wherein the composition comprises: (A) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, (b) a tracr mate sequence, and (c) a tracr sequence, and II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences (or optionally at least one or more nuclear localization sequences as some embodiments can involve no NLS), wherein (a), (b) and (c) are arranged in a 5' to 3' orientation, wherein components I and II are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence, or (B) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, and (b) at least one or more tracr mate sequences, II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and III. a third regulatory element operably linked to a tracr sequence, wherein components I, II and III are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr

mate sequence that is hybridized or hybridizable to the tracr sequence. In some embodiments, components I, II and III are located on the same vector. In other embodiments, components I and II are located on the same vector, while component III is located on another vector. In other embodiments, components I and III are located on the same vector, while component II is located on another vector. In other embodiments, components II and III are located on the same vector, while component I is located on another vector. In other embodiments, each of components I, II and III is located on different vectors. The invention also provides a viral or plasmid vector system as described herein.

[0043] Preferably, the vector is a viral vector, such as a lenti- or baculo- or preferably adeno-viral/adeno-associated viral vectors, but other means of delivery are known (such as yeast systems, microvesicles, gene guns/means of attaching vectors to gold nanoparticles) and are provided. In some embodiments, one or more of the viral or plasmid vectors may be delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun.

[0044] By manipulation of a target sequence, Applicants mean alteration of the target sequence, which may include the epigenetic manipulation of a target sequence. This epigenetic manipulation may be of the chromatin state of a target sequence, such as by modification of the methylation state of the target sequence (i.e. addition or removal of methylation or methylation patterns or CpG islands), histone modification, increasing or reducing accessibility to the target sequence, or by promoting 3D folding. In relation to nucleotide repeats, however, excision of the sequence repeats is the manipulation of primary interest.

[0045] It will be appreciated that where reference is made to a method of modifying an organism or mammal including human or a non-human mammal or organism by manipulation of a target sequence in a genomic locus of interest, this may apply to the organism (or mammal) as a whole or just a single cell or population of cells from that organism. In the case of humans, for instance, Applicants envisage, *inter alia*, a single cell or a population of cells and these may preferably be modified *ex vivo* and then re-introduced. In this case, a biopsy or other tissue or biological fluid sample may be necessary. Stem cells are also particularly preferred in this regard. But, of course, *in vivo* embodiments are also envisaged.

[0046] In certain embodiments the invention provides a method of treating or inhibiting a condition caused by a defect in a target sequence in a genomic locus of interest in a subject (e.g., mammal or human) or a non-human subject (e.g., mammal) in need thereof comprising modifying the subject or a non-human subject by manipulation of the target sequence and wherein the condition is susceptible to treatment or inhibition by manipulation of the target

sequence comprising providing treatment comprising: delivering a non-naturally occurring or engineered composition comprising an AAV or lentivirus vector system comprising one or more AAV or lentivirus vectors operably encoding a composition for expression thereof, wherein the target sequence is manipulated by the composition when expressed, wherein the composition comprises: (A) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, (b) a tracr mate sequence, and (c) a tracr sequence, and II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences (or optionally at least one or more nuclear localization sequences as some embodiments can involve no NLS, i.e., there can be zero NLSs but advantageously there is greater than zero NLSs, such as one or more or advantageously two or more NLSs, and thus the invention comprehends embodiments wherein there is 0, 1, 2, 3, or more NLSs) wherein (a), (b) and (c) are arranged in a 5' to 3' orientation, wherein components I and II are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence, or (B) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, and (b) at least one or more tracr mate sequences, II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and III. a third regulatory element operably linked to a tracr sequence, wherein components I, II and III are located on the same or different vectors of the system, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence. In some embodiments, components I, II and III are located on the same vector. In other embodiments, components I and

II are located on the same vector, while component III is located on another vector. In other embodiments, components I and III are located on the same vector, while component II is located on another vector. In other embodiments, components II and III are located on the same vector, while component I is located on another vector. In other embodiments, each of components I, II and III is located on different vectors. The invention also provides a viral (e.g. AAV or lentivirus) vector system as described herein, although other vector systems are known in the art and can be part of a vector system as described herein.

[0047] Some methods of the invention can include inducing expression. In some methods of the invention the organism or subject is a eukaryote, including e.g. a plant or an animal (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 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 viral vector is an AAV or a lentivirus, and can be part of a vector system as described herein. Delivery therefore can be via a vector, such as a viral vector, e.g., a recombinant viral vector delivery system; and, this system can be an AAV or lentivirus or derived from an AAV or a lentivirus (e.g., a recombinant AAV or lentivirus that expresses that which is foreign, heterologous or that which is not homologous or native to the virus may make some consider the virus "derived from" is parent virus). . In some methods of the invention the viral vector is a lentivirus-derived vector. In some methods of the invention the viral vector is an Agrobacterium Ti or Ri plasmid for use in plants. In some methods of the invention the CRISPR enzyme is a Cas9. In some methods of the invention the CRISPR enzyme comprises one or more mutations in one of the catalytic domains. In some methods of the invention the CRISPR enzyme is a Cas9 nickase. In some methods of the invention the expression of the guide sequence is under the control of the T7 promoter and that is driven by the expression of T7 polymerase. In some methods of the invention the expression of the guide sequence is under the control of a U6 promoter. In some methods of the invention the CRISPR enzyme comprises one or more mutations in one of the catalytic domains. In some methods of the invention the CRISPR enzyme is a Cas9 nickase.

[0048] The invention in some embodiments comprehends a method of delivering a CRISPR enzyme comprising delivering to a cell a nucleic acid molecule, e.g., a plasmid or RNA or mRNA encoding the CRISPR enzyme. In some of these methods the CRISPR enzyme is a Cas9.

[0049] The invention also provides methods of preparing the vector systems of the invention, in particular the viral vector systems as described herein. The invention in some embodiments comprehends a method of preparing the vector, e.g., AAV or lentivirus, of the invention comprising transfecting one or more plasmid(s) containing or consisting essentially of nucleic acid molecule(s) coding for the AAV into AAV-infectable cells, and supplying AAV rep and/or cap obligatory for replication and packaging of the AAV. In some embodiments the AAV rep and/or cap obligatory for replication and packaging of the AAV are supplied by transfecting the cells with helper plasmid(s) or helper virus(es). In some embodiments the helper virus is a poxvirus, adenovirus, herpesvirus or baculovirus. In some embodiments the poxvirus is a vaccinia virus. In some embodiments the cells are mammalian cells. And in some embodiments the cells are insect cells and the helper virus is baculovirus. In other embodiments, the virus is a lentivirus.

[0050] The invention further comprehends a composition of the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme) for use in medicine or in therapy. In some embodiments the invention comprehends a composition according to the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme) for use in a method according to the invention. In some embodiments the invention provides for the use of a composition of the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme) in *ex vivo* gene or genome editing. In certain embodiments the invention comprehends use of a composition of the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme) in the manufacture of a medicament for *ex vivo* gene or genome editing or for use in a method according of the invention. In some methods of the invention the CRISPR enzyme comprises one or more mutations in one of the catalytic domains. In some methods of the invention the CRISPR enzyme is a Cas9 nickase.

[0051] The invention comprehends in some embodiments a composition of the invention or a CRISPR enzyme thereof (including or alternatively mRNA encoding the CRISPR enzyme), wherein the target sequence is flanked at its 3' end by a 5' motif termed a proto-spacer adjacent motif or PAM, especially where the Cas9 is (or is derived from) *S. pyogenes* or *S. aureus* Cas9. For example, a suitable PAM is 5'-NRG or 5'-NNGRR (where N is any Nucleotide) for SpCas9 or SaCas9 enzymes (or derived enzymes), respectively, as mentioned below.

[0052] It will be appreciated that SpCas9 or SaCas9 are those from or derived from *S. pyogenes* or *S. aureus* Cas9.

[0053] Apects of the invention comprehend improving the specificity of a CRISPR enzyme, e.g. Cas9, mediated gene targeting and reducing the likelihood of off-target modification by the CRISPR enzyme, e.g. Cas9. The invention in some embodiments comprehends a method of modifying an organism or a non-human organism by minimizing off-target modifications by manipulation of a first and a second target sequence on opposite strands of a DNA duplex in a genomic locus of interest in a cell comprising delivering a non-naturally occurring or engineered composition comprising :

I. a first CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the first polynucleotide sequence comprises:

(a) a first guide sequence capable of hybridizing to the first target sequence,

(b) a first tracr mate sequence, and

(c) a first tracr sequence,

II. a second CRISPR-Cas system chiRNA polynucleotide sequence, wherein the second polynucleotide sequence comprises:

(a) a second guide sequence capable of hybridizing to the second target sequence,

(b) a second tracr mate sequence, and

(c) a second tracr sequence, and

III. a polynucleotide sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences and comprising one or more mutations, wherein (a), (b) and (c) are arranged in a 5' to 3' orientation, wherein when transcribed, the first and the second tracr mate sequence hybridize to the first and second tracr sequence respectively and the first and the second guide sequence directs sequence-specific binding of a first and a second CRISPR complex to the first and second target sequences respectively, wherein the first CRISPR complex comprises the CRISPR enzyme complexed with (1) the first guide sequence that is hybridized or hybridizable to the first target sequence, and (2) the first tracr mate sequence that is hybridized or hybridizable to the first tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized or hybridizable to the second target sequence, and (2) the second tracr mate sequence that is hybridized or hybridizable to the second tracr sequence, wherein the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA, and wherein the first guide sequence directs cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directs cleavage of the other or opposite strand of the DNA duplex near the second target sequence inducing an offset or double strand break, thereby modifying the organism or the

non-human organism by minimizing off-target modifications. In one aspect, the first nick and the second nick in the DNA is offset relative to each other by at least one base pair of the duplex. In one aspect, the first nick and the second nick are offset relative to each other so that the resulting DNA break has a 3' overhang. In one aspct, the first nick and the second nick are offset relative to each other so that the resulting DNA break has a 5' overhang. In one aspect, the first nick and the second nick are positioned relative to each other such that the overhang is at least 1 nucleotide (nt), at least 10 nt, at least 15 nt, at least 26 nt, at least 30 nt, at least 50 nt or more that at least 50 nt. Additional aspects of the invention comprising the resulting offset double nicked DNA strand can be appreciated by one skilled in the art, and exemplary uses of the double nick system are provided herein.

[0054] In some methods of the invention any or all of the polynucleotide sequence encoding the CRISPR enzyme, the first and the second guide sequence, the first and the second tracr mate sequence or the first and the second tracr sequence, is/are RNA. In further embodiments of the invention the polynucleotides comprising the sequence encoding the CRISPR enzyme, the first and the second guide sequence, the first and the second tracr mate sequence or the first and the second tracr sequence, is/are RNA and are delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun. In certain embodiments of the invention, the first and second tracr mate sequence share 100% identity and/or the first and second tracr sequence share 100% identity. In some embodiments, the polynucleotides may be comprised within a vector system comprising one or more vectors. In preferred embodiments of the invention the CRISPR enzyme is a Cas9 enzyme, e.g. SpCas9. In an aspect of the invention the CRISPR enzyme comprises one or more mutations in a catalytic domain, wherein the one or more mutations are selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the CRISPR enzyme has the D10A mutation. In preferred embodiments, the first CRISPR enzyme has one or more mutations such that the enzyme is a complementary strand nicking enzyme, and the second CRISPR enzyme has one or more mutations such that the enzyme is a non-complementary strand nicking enzyme. Alternatively the first enzyme may be a non-complementary strand nicking enzyme, and the second enzyme may be a complementary strand nicking enzyme.

[0055] In preferred methods of the invention the first guide sequence directing cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directing cleavage of the other strand near the second target sequence results in a 5' overhang. In embodiments of the invention the 5' overhang is at most 200 base pairs, preferably at most 100

base pairs, or more preferably at most 50 base pairs. In embodiments of the invention the 5' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs.

[0056] The invention in some embodiments comprehends a method of modifying an organism or a non-human organism by minimizing off-target modifications by manipulation of a first and a second target sequence on opposite strands of a DNA duplex in a genomic locus of interest in a cell comprising delivering a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising

I. a first regulatory element operably linked to

(a) a first guide sequence capable of hybridizing to the first target sequence, and

(b) at least one or more tracr mate sequences,

II. a second regulatory element operably linked to

(a) a second guide sequence capable of hybridizing to the second target sequence, and

(b) at least one or more tracr mate sequences,

III. a third regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and

IV. a fourth regulatory element operably linked to a tracr sequence,

wherein components I, II, III and IV are located on the same or different vectors of the system, when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the first and the second guide sequence direct sequence-specific binding of a first and a second CRISPR complex to the first and second target sequences respectively, wherein the first CRISPR complex comprises the CRISPR enzyme complexed with (1) the first guide sequence that is hybridized or hybridizable to the first target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized or hybridizable to the second target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence, wherein the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA, and wherein the first guide sequence directs cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directs cleavage of the other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism by minimizing off-target modifications.

[0057] The invention also provides a vector system as described herein. The system may comprise one, two, three or four different vectors. Components I, II, III and IV may thus be

located on one, two, three or four different vectors, and all combinations for possible locations of the components are herein envisaged, for example: components I, II, III and IV can be located on the same vector; components I, II, III and IV can each be located on different vectors; components I, II, II I and IV may be located on a total of two or three different vectors, with all combinations of locations envisaged, etc.

[0058] In some methods of the invention any or all of the polynucleotide sequence encoding the CRISPR enzyme, the first and the second guide sequence, the first and the second tracr mate sequence or the first and the second tracr sequence, is/are RNA. In further embodiments of the invention the first and second tracr mate sequence share 100% identity and/or the first and second tracr sequence share 100% identity. In preferred embodiments of the invention the CRISPR enzyme is a Cas9 enzyme, e.g. SpCas9. In an aspect of the invention the CRISPR enzyme comprises one or more mutations in a catalytic domain, wherein the one or more mutations are selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the CRISPR enzyme has the D10A mutation. In preferred embodiments, the first CRISPR enzyme has one or more mutations such that the enzyme is a complementary strand nicking enzyme, and the second CRISPR enzyme has one or more mutations such that the enzyme is a non-complementary strand nicking enzyme. Alternatively the first enzyme may be a non-complementary strand nicking enzyme, and the second enzyme may be a complementary strand nicking enzyme. In a further embodiment of the invention, one or more of the viral vectors are delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun.

[0059] In preferred methods of the invention the first guide sequence directing cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directing cleavage of the other or opposite strand near the second target sequence results in a 5' overhang. In embodiments of the invention the 5' overhang is at most 200 base pairs, preferably at most 100 base pairs, or more preferably at most 50 base pairs. In embodiments of the invention the 5' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs.

[0060] The invention in some embodiments comprehends a method of modifying a genomic locus of interest by minimizing off-target modifications by introducing into a cell containing and expressing a double stranded DNA molecule encoding a gene product of interest an engineered, non-naturally occurring CRISPR-Cas system comprising a Cas protein having one or more mutations and two guide RNAs that target a first strand and a second strand of the DNA

molecule respectively, whereby the guide RNAs target the DNA molecule encoding the gene product and the Cas protein nicks each of the first strand and the second strand of the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas protein and the two guide RNAs do not naturally occur together.

[0061] In preferred methods of the invention the Cas protein nicking each of the first strand and the second strand of the DNA molecule encoding the gene product results in a 5' overhang. In embodiments of the invention the 5' overhang is at most 200 base pairs, preferably at most 100 base pairs, or more preferably at most 50 base pairs. In embodiments of the invention the 5' overhang is at least 26 base pairs, preferably at least 30 base pairs or more preferably 34-50 base pairs.

[0062] Embodiments of the invention also comprehend the guide RNAs comprising a guide sequence fused to a tracr mate sequence and a tracr sequence. In an aspect of the invention the Cas protein is codon optimized for expression in a eukaryotic cell, preferably a mammalian cell or a human cell. As explained in more detail below, codon usage can even be optimized for expression in particular cell types *e.g.* for brain cells. In further embodiments of the invention the Cas protein is a type II CRISPR-Cas protein, e.g. a Cas9 protein. In a highly preferred embodiment the Cas protein is a Cas9 protein, e.g. SpCas9. In aspects of the invention the Cas protein has one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the Cas protein has the D10A mutation.

[0063] Aspects of the invention relate to the expression of the gene product being decreased or a template polynucleotide being further introduced into the DNA molecule encoding the gene product or an intervening sequence being excised precisely by allowing the two 5' overhangs to reanneal and ligate or the activity or function of the gene product being altered or the expression of the gene product being increased. In an embodiment of the invention, the gene product is a protein.

[0064] The invention also comprehends an engineered, non-naturally occurring CRISPR-Cas system comprising a Cas protein having one or more mutations and two guide RNAs that target a first strand and a second strand respectively of a double stranded DNA molecule encoding a gene product in a cell, whereby the guide RNAs target the DNA molecule encoding the gene product and the Cas protein nicks each of the first strand and the second strand of the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas protein and the two guide RNAs do not naturally occur together.

[0065] In aspects of the invention the guide RNAs may comprise a guide sequence fused to a tracr mate sequence and a tracr sequence. In an embodiment of the invention the Cas protein is a type II CRISPR-Cas protein. In an aspect of the invention the Cas protein is codon optimized for expression in a eukaryotic cell, preferably a mammalian cell or a human cell. In further embodiments of the invention the Cas protein is a type II CRISPR-Cas protein, e.g. a Cas9 protein. In a highly preferred embodiment the Cas protein is a Cas9 protein, e.g. SpCas9. In aspects of the invention the Cas protein has one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a highly preferred embodiment the Cas protein has the D10A mutation.

[0066] Aspects of the invention relate to the expression of the gene product being decreased or a template polynucleotide being further introduced into the DNA molecule encoding the gene product or an intervening sequence (such as a trinucleotide repeat or other nucleotide expansion element) being excised precisely by allowing the two 5' overhangs to reanneal and ligate or the activity or function of the gene product being altered or the expression of the gene product being increased. In an embodiment of the invention, the gene product is a protein.

[0067] The invention also comprehends an engineered, non-naturally occurring vector system comprising one or more vectors comprising:

a) a first regulatory element operably linked to each of two CRISPR-Cas system guide RNAs that target a first strand and a second strand respectively of a double stranded DNA molecule encoding a gene product,

b) a second regulatory element operably linked to a Cas protein,

wherein components (a) and (b) are located on same or different vectors of the system, whereby the guide RNAs target the DNA molecule encoding the gene product and the Cas protein nicks each of the first strand and the second strand of the DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the Cas protein and the two guide RNAs do not naturally occur together.

[0068] In preferred embodiments of the invention the vectors of the system are viral vectors. In a further embodiment, the vectors of the system are delivered via liposomes, nanoparticles, exosomes, microvesicles, or a gene-gun.

[0069] In one aspect, the invention provides a method of modifying a target polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR

enzyme complexed with a guide sequence hybridized or hybridizable to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by homologous recombination with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid changes in a protein expressed from a gene comprising the target sequence. In some embodiments, the method further comprises delivering one or more vectors to said eukaryotic cell, wherein the one or more vectors drive expression of one or more of: the CRISPR enzyme, the guide sequence linked to the tracr mate sequence, and the tracr sequence. In some embodiments, said vectors are delivered to the eukaryotic cell in a subject. In some embodiments, said modifying takes place in said eukaryotic cell in a cell culture. In some embodiments, the method further comprises isolating said eukaryotic cell from a subject prior to said modifying. In some embodiments, the method further comprises returning said eukaryotic cell and/or cells derived therefrom to said subject.

[0070] In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized or hybridizable to a target sequence within said polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. The nature of the Complex and the target can determine whether binding results in increased or decreased expression. For example, the target may be a gene product whose expression leads to the down-regulation or decreased expression of another gene product. Decreasing the expression of that first gene product can lead to expression being increased as to the second gene product (and of course expression of the first product is decreased). The complex can bind to a target and result in altered expression of a protein, e.g., a modified version being expressed. In that instance, the expression of the modified form of the protein is increased. These are but some of the ways that expression may be increased or decreased. In some embodiments, the method further comprises delivering one or more vectors

to said eukaryotic cells, wherein the one or more vectors drive expression of one or more of: the CRISPR enzyme, the guide sequence linked to the tracr mate sequence, and the tracr sequence.

[0071] In one aspect, the invention provides a method of generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, a disease gene is any gene associated with an increase in the risk of having or developing a disease. In some embodiments, the method comprises (a) introducing one or more vectors into a eukaryotic cell, wherein the one or more vectors drive expression of one or more of: a CRISPR enzyme, a guide sequence linked to a tracr mate sequence, and a tracr sequence; and (b) allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said disease gene, wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence within the target polynucleotide, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence, thereby generating a model eukaryotic cell comprising a mutated disease gene. In some embodiments, said cleavage comprises cleaving one or two strands at the location of the target sequence by said CRISPR enzyme. In some embodiments, said cleavage results in decreased transcription of a target gene. In some embodiments, the method further comprises repairing said cleaved target polynucleotide by homologous recombination with an exogenous template polynucleotide, wherein said repair results in a mutation comprising an insertion, deletion, or substitution of one or more nucleotides of said target polynucleotide. In some embodiments, said mutation results in one or more amino acid changes in a protein expression from a gene comprising the target sequence.

[0072] In one aspect, the invention provides for methods of modifying a target polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized or hybridizable to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence.

[0073] 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.

[0074] Where desired, to effect the modification of the expression in a cell, one or more vectors comprising a tracr sequence, a guide sequence linked to the tracr mate sequence, a

sequence encoding a CRISPR enzyme is delivered to a cell. In some methods, the one or more vectors comprises a regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR enzyme comprising a nuclear localization sequence; and a regulatory element operably linked to a tracr mate sequence and one or more insertion sites for inserting a guide sequence upstream of the tracr mate sequence. When expressed, the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a cell. Typically, the CRISPR complex comprises a CRISPR enzyme complexed with (1) the guide sequence that is hybridized or hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized or hybridizable to the tracr sequence.

[0075] 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 is not produced.

[0076] In certain embodiments, the CRISPR enzyme comprises one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A or D986A and/or the one or more mutations is in a RuvC1 or HNH domain of the CRISPR enzyme or is a mutation as otherwise as discussed herein. In some embodiments, the CRISPR enzyme has one or more mutations in a catalytic domain, wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and the guide sequence directs sequence-specific binding of a CRISPR complex to the target sequence, and wherein the enzyme further comprises a functional domain. Thus, in some embodiments a mutated Cas9 enzyme may be fused to a protein domain or functional domain. In one aspect, the functional domain is a transcriptional activation domain, preferably VP64. 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 (e.g., 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.

[0077] In some embodiments, the CRISPR enzyme is a type I or III CRISPR enzyme, but is preferably a type II CRISPR enzyme. This type II CRISPR enzyme may be any Cas enzyme. A Cas enzyme may be identified as Cas9 as this can refer to the general class of enzymes that share

homology to the biggest nuclease with multiple nuclease domains from the type II CRISPR system. Most preferably, the Cas9 enzyme is from, or is derived from, spCas9 or saCas9. By derived, Applicants mean that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as described herein.

[0078] It will be appreciated that the terms Cas and CRISPR enzyme are generally used herein interchangeably, unless otherwise apparent. As mentioned above, many of the residue numberings used herein refer to the Cas9 enzyme from the type II CRISPR locus in *Streptococcus pyogenes*. However, it will be appreciated that this invention includes many more Cas9s from other species of microbes, such as SpCas9, SaCa9, St1Cas9 and so forth.

[0079] An example of a codon optimized sequence, in this instance optimized for humans (i.e. being optimized for expression in humans) is provided herein, see the SaCas9 human codon optimized sequence. 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 such as the brain, can be employed in the practice of the invention, from the teachings herein in conjunction with the knowledge in the art.

[0080] Preferably, delivery is in the form of a vector which may be a viral vector, such as a lenti- or baculo- or preferably adeno-viral/adeno-associated viral vectors, but other means of delivery are known (such as yeast systems, microvesicles, gene guns/means of attaching vectors to gold nanoparticles) and are provided. A vector may mean not only a viral or yeast system (for instance, where the nucleic acids of interest may be operably linked to and under the control of (in terms of expression, such as to ultimately provide a processed RNA) a promoter), but also direct delivery of nucleic acids into a host cell. While in herein methods the vector may be a viral vector and this is advantageously an AAV, other viral vectors as herein discussed can be employed, such as lentivirus. For example, baculoviruses may be used for expression in insect cells. These insect cells may, in turn be useful for producing large quantities of further vectors, such as AAV or lentivirus vectors adapted for delivery of the present invention. Also envisaged is a method of delivering the present CRISPR enzyme comprising delivering to a cell mRNA encoding the CRISPR enzyme. It will be appreciated that in certain embodiments the CRISPR enzyme is truncated, and/or comprised of less than one thousand amino acids or less than four thousand amino acids, and/or is a nuclease or nickase, and/or is codon-optimized, and/or comprises one or more mutations, and/or comprises a chimeric CRISPR enzyme, and/or the other options as herein discussed. AAV and lentiviral vectors are preferred.

[0081] In certain embodiments, the target sequence is flanked or followed, at its 3' end, by a PAM suitable for the CRISPR enzyme, typically a Cas and in particular a Cas9.

[0082] For example, a suitable PAM is 5'-NRG or 5'-NNGRR for SpCas9 or SaCas9 enzymes (or derived enzymes), respectively. For S. pyogenes Cas9 or derived enzymes, a suitable PAM is 5'-NRG.

[0083] Expression of the components of a CRISPR system preferably does not take place systemically in a subject, but rather occurs only in desired cells, tissues or organs of interest. The invention utilises three principle ways of controlling expression in this way, which can be used singly or in combination. Firstly, expression can be under the control of regulatory elements which are specific to the desired cells, tissues or organs. Secondly, a delivery vehicle can be used which is specific to the desired cells, tissues or organs *e.g.* based on suitably-specific cell surface molecules. Thirdly, local delivery can be used *e.g.* by delivery into the desired cells, tissues or organs, such as by injection.

[0084] In an aspect the invention provides a non-naturally occurring or engineered composition Self Inactivating CRISPR-Cas system comprising

I. a first regulatory element operably linked to

(a) at least one first guide sequence capable of hybridizing to at least one first target sequence in the genome of a eukaryotic cell, and

(b) at least one or more tracr mate sequences, and

(c) at least one or more tracr sequences, and

II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme comprising, one, two or more nuclear localization signals (NLSs) and optionally a selection marker,

wherein the system further comprises

(a) at least a second guide sequence capable of hybridizing to a second target sequence selected from one or more of:

a sequence encoding the CRISPR type II enzyme, and

a sequence within a non-coding CRISPR-Cas construct selected from

i) within the promoter driving expression of the non-coding RNA elements,

ii) within the promoter driving expression of the Cas9 gene,

iii) within 100bp of the ATG translational start codon in the Cas9 coding sequence, and

iv) within the inverted terminal repeat of the AAV genome; and

(b) at least one or more tracr mate sequences for the at least one second guide sequence; wherein when transcribed, the tracr mate sequence hybridizes to the tracr sequence and

the first guide sequence is hybridized or hybridizable to the first target sequence and directs sequence-specific binding of a CRISPR complex to the first target sequence,

wherein the first CRISPR complex comprises the CRISPR enzyme complexed with (1) the first guide sequence that is hybridized or hybridizable to the first target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence, and

wherein the first CRISPR complex mediates binding to or a double or single stranded DNA break, thereby editing the genomic locus in the cell; and

the second guide sequence is hybridized or hybridizable to the second target sequence that inactivates one or more components of the CRISPR-Cas system, whereby all CRISPR complexes become self-inactivating.

[0085] Furthermore, the second guide sequence is optionally introduced into the system simultaneously with the CRISPR-Cas system comprising the first guide sequence, or sequentially at a time point after the introduction of the elements encoding the first CRISPR-Cas complex.

[0086] Accordingly, it is an object of the invention to not encompass within the invention any previously known product, process of making the product, or method of using the product such that Applicants reserve the right and hereby disclose a disclaimer of any previously known product, process, or method. It is further noted that the invention does not intend to encompass within the scope of the invention any product, process, or making of the product or method of using the product, which does not meet the written description and enablement requirements of the USPTO (35 U.S.C. §112, first paragraph) or the EPO (Article 83 of the EPC), such that Applicants reserve the right and hereby disclose a disclaimer of any such subject matter.

[0087] It is noted that in this disclosure and particularly in the claims and/or paragraphs, terms such as "comprises", "comprised", "comprising" and the like can have the meaning attributed to it in U.S. Patent law; e.g., they can mean "includes", "included", "including", and the like; and that terms such as "consisting essentially of" and "consists essentially of" have the meaning ascribed to them in U.S. Patent law, e.g., they allow for elements not explicitly recited, but exclude elements that are found in the prior art or that affect a basic or novel characteristic of the invention. It may be advantageous in the practice of the invention to be in compliance with Art. 53(c) EPC and Rule 28(b) and (c) EPC. Nothing herein is to be construed as a promise.

[0088] These and other embodiments are disclosed or are obvious from and encompassed by, the following Detailed Description.

BRIEF DESCRIPTION OF THE DRAWINGS

[0089] The novel features of the invention are set forth with particularity in the appended claims. A better 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 are utilized, and the accompanying drawings of which:

[0090] Figure 1 shows currently known diseases (highlighted) caused by an abnormal expansion of a nucleotide repeat sequence. Nucleotide repeats vary in size (triangles) and can reside in coding or non-coding regions of the disease-associated genes. Based on empirical and bioinformatics analyses Applicants have determined that each loci can be targeted using the CRISPR-Cas9 approach described in this application.

[0091] Figure 2A-B shows targeting design and editing of the human ATXN1 CAG repeat. a) Review of the CRISPR-Cas9 system. b) Guide sequences flanking the CAG nucleotide repeat in ATXN1 are removed using the CRISPR-Cas9 system. Shown are PCR products (primers depicted in gray) that correspond to an unedited (top dark band) and edited (lower band, arrow) ATXN1 after transient expression of CRISPR-Cas9 plasmids in a human HT1080 cells. As expected, successful editing of the repeat is only observed when both flanking guide non-coding RNAs are simultaneously expressed (compare lanes 1-5 versus lanes 6-11).

[0092] Figure 3A-C shows CAG repeats are excised by the CRISPR-Cas9 system. a) Edited genomic DNA (Figure 3b) was purified, cloned and sequenced to confirm editing of the CAG nucleotide repeat. Over 150 clones were sequenced. The analysis showed that although different sequences were present (Isoforms A-I) all but 1 (Isoform F) lacked the endogenous human ATXN1 CAG repeat. b) Most clones belonged to Isoform A. c) Isoform A produced a new genomic locus with an in-frame deletion lacking the CAG repeat.

[0093] Figure 4A-B shows targeting design and editing of the human FMR1 CGG repeat; a second example. a) Guide sequences that flank the CGG nucleotide repeat region in FMR1. Unlike the CAG repeats in ATXN1, the CGG repeats in FMR1 are in the 5' un-translated region (5'UTR). b) Successful editing of FMR1 as evidenced by the PCR products (primers depicted in gray) corresponding to an unedited (top dark band) and edited (lower band, arrow) human FMR1 5'UTR/exon-1 region. For this, CRISPR-Cas9 plasmids were transfected into HT1080 cells for transient.

[0094] Figure 5 shows CGG repeats are excised by the CRISPR-Cas9 system as confirmed by direct sequencing of the FMR1 5'UTR/exon-1 region. a) Edited genomic DNA (shown in Figure 4b) was sequenced to confirm appropriate editing of FMR1 CGG repeats. Sequence alignment analysis showed the absence of the CGG repeat sequence in over 100 sequenced clones when compared to wild-type sequence.

[0095] Figure 6A-C shows design of Adeno Associated Viral vectors for the delivery of CRISPR-Cas9 system into the mammalian brain. a) An AAV vector was engineered for in vivo CRISPR-Cas9-mediated genome editing. SpCas9 containing an N-terminal and C-terminal nuclear localization domain as well as an N-terminal Flag was cloned into an AAV shuttle plasmid. Because of the large size of the SpCas9 cDNA and the desire to obtain low levels of SpCas9 nuclease expression in vivo, Applicants omitted the use of a promoter. Instead, expression of the SpCas9 is driven by the basal transcriptional activity of the AAV inverted terminal repeat (iTR) sequences. The guide RNA (gcRNA) and the transactivating RNA (tracrRNA), were cloned into a different AAV shuttle plasmid and placed under the regulation of two different RNA polymerase type-III promoters: the U6 and H1 promoters respectively. A reporter gene (EGFP shown as an example), or any other sequence, can be cloned downstream of the non-coding expression cassettes. b) In this system, the non-coding CRISPR components are expressed as an array of chimeras (sgRNAs) driven by the U6 promoter. c) AAV plasmids described in 6a were used to target ATXN1 plasmids that carry either 30 CAG nucleotide repeats (normal range) or 80 CAG repeats (disease range). Excision of the normal or the expanded CAG repeat was observed only in the presence of both AAV plasmids (iTR-SpCas9 and AAVPS2/AAVPS5). This result confirms that the AAV-iTR-SpCas9 vector can produce sufficient SpCas9 to mediate effective gene editing.

[0096] Figure 7A-B shows a self-inactivating AAV-CRISPR-Cas9 system. a) Diagram of the SIN-CC9 concept. Applicants designed plasmids that co-express sgRNAs targeting genomic sequences of interest (shown in grays/black) with "self-inactivating" sgRNAs that target an SpCas9 sequence near the engineered ATG start site (shown in red/black). A regulatory sequence in the U6 promoter region can also be targeted with an sgRNA (shown in blue/black). The U6-driven sgRNAs as shown are designed in an array format such that multiple sgRNA sequences can be simultaneously released. When first delivered into target tissue/cells (left cell) sgRNAs begin to accumulate while Cas9 levels rise in the nucleus. Cas9 will complex with all of the sgRNAs to mediate genome editing and self-inactivation of the CRISPR-Cas9 plasmids. b) Left panel: Western blot analysis following transient co-expression of SpCas9 with sgRNAs

targeting the ATXN1 CAG repeat region or sgRNAs targeting the ATXN1 CAG repeat region (SpCas9+gATXN1) and a sequence just downstream of the SpCas9 start codon (SpCas9+gATXN1/gCas9). Transient expression of SpCas9 normally persists to at least 48hrs post-transfection. In contrast, in the presence of the anti-SpCas9 sgRNAs, transient expression of SpCas9 is lost by 48hrs post-transfection. Right panel: PCR products using primers that span the ATXN1 CAG repeat region showing successfull excision of the CAG repeat prior to the loss of SpCas9 expression.

[0097] Figure 8A-B shows the concept of CRISPR-Cas9-mediated allele-specific targeting.

[0098] Figure 9A-B shows loss of Ataxin-1 expression after excision of CAG repeats from within the ATXN1 locus. Figure 9a shows PCR across the human ATXN1 locus following a 5-day selection of HT1080 cells transfected the indicated plasmids. After selection with Puromycin, >90% of the cell population contained an edited ATXN1 locus lacking the CAG repeat, and the arrow shows the shorter locus. In Figure 9b, quantitative PCR analysis of ATXN1 expression reveals a significant reduction in the steady-state levels of ATXN1 mRNA. Two different cell lines were analyzed and similar results were observed. The y-axis shows transcript levels relative to the control (level 1.0).

[0099] Figure 10A-C shows targeting of expanded CTG repeats in the DMPK locus. Guide sequences were designed to flank the CTG nucleotide repeat region in the 3' un-translated region (3'UTR) of DMPK. The PCR products in Figure 10a indicate successful editing of the DMPK locus in HT1080 cells, comparing unedited (top arrow) and edited (bottom arrow). Figure 10b shows primary skin fibroblasts biopsied from a DM1 patient. After the CRISPR-Cas9 plasmids targeting the DMPK locus were introduced, Figure 10c shows that the CTG expansion is effectively excised.

[00100] Figure 11A-C shows AAV vectors for delivery of CRISPR-Cas9 system into mammalian tissue.The vectors are illustrated in Figure 11a, using the SaCas9 nuclease containing an N-terminal nuclear localization domain as well as an N-terminal HA-tag, cloned into an AAV shuttle plasmid and placed under the control of a CMV promoter. The non-coding RNA elements required for Cas9-mediated gene editing are also contained within the same AAV packaging genome. This allows for the co-delivery of a second AAV vector (example provided CMV-EGFP) that could serve as a transduction marker or a template donor whenever HR is desired. Figure 11b shows results from successful vector delivery to mice as indicated by expression of a EGFP marker. Figure 11c shows that delivery of the anti-CTG SaCas9 led to efficient excision of the CTG repeats from the HSA^{LR} transgenic locus, but no editing was

observed in mice receiving the AAV9-EGFP virus alone or an AAV9 expressing SaCas9 with a control guide (scrambled sequence).

[0100] Figure 12 shows synthetic expression constructs that contain the first 342 nucleotides (from the ATG start site) of the ATXN2 mRNA fused in-frame to the N-terminus of EGFP (G allele) or mCherry (C allele). Experimental results described in Example 37 indicate that allele-specific targeting of the C allele using CRISPR-Cas9 results in a loss of mCherry (C allele) but not EGFP (G allele) expression in cultured cells.

[0101] Figure 13 depicts one aspect of a Self-Inactivating CRISPR-Cas9 system; see Examples 3, 4.

[0102] Figure 14 depicts an exemplary self-inactivating CRISPR-Cas9 system for a chimeric tandem array transcript specific to the ATXN1 locus. The ATXN1aPS9 guide edits the ATXN1 locus while the U6aPS1 and CMVaPS1 guides inactivate the CRISPR-Cas9 system; see Examples 3, 4.

[0103] Figure 15A-C shows tandem guide RNAs are efficiently processed especially in the first position. Tandem guide RNAs are efficiently processed especially in the first position ((A) Schematic showing tandem guide RNA scaffolds encoding for either EMX1.3 or EMX63 in the first or second position with position of Emx1.3 Northern probe shown in red. (B) Northern blot analysis examining processing of tandem sgRNA in cells. (C) SURVEYOR assay examining independent sgRNA activity targeting two genomic loci, DYRK1A and GRIN2B. The three left lanes in both panels are tsgRNAs targeting DYRK1A in the first position and GRIN2B in the second position. Conversely, three right lanes target GRIN2B first and then DYRK1A second).

[0104] Figure 16A-C shows optimization of tsgRNA scoffold pairings. Optimization of tsgRNA scaffold pairings ((A) Schematic of tandem scaffold design with first spacer targeting Grin2B using Scaffold A and second spacer targeting Cas9 itself using Scaffold B in a Cas9-T2A-GFP expressing plasmid. (B) Single U6-guide controls show both an increase in the percentage of GFP-negative cells as well as a decrease in mean fluorescence intensity of the positive fraction. (C) 12x12 matrix of tandem scaffold pairings and results of subsequent analyses by flow cytometry).

[0105] Figure 17 depicts tandem pairs between divergent scaffolds improves second spacer activity. Tandem pairs between divergent scaffolds improve second spacer activity (Sequence alignment of the sgRNA scaffolds used in the previous study to the sp85 scaffold).

[0106] Figure 18A-E provides evidence of in vivo CRISPR/Cas9 genome editing efficacy and therapeutic benefit in a polyglutamine disease mouse model.

[0107] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE INVENTION

[0108] With respect to general information on CRISPR-Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, viral vectors, adenovirus, AAV, lentivirus, and making and using thereof, including as to amounts and formulations, all useful in the practice of the instant invention, reference is made to: US Patents Nos. 8,697,359, 8,771,945, 8,795,965, 8,865,406, 8,871,445, 8,889,356, 8,889,418 and 8,895,308; 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 Patent Applications EP 2 771 468 (EP13818570.7), EP 2 764 103 (EP13824232.6), and EP 2 784 162 (EP14170383.5); and PCT Patent Publications WO 2014/093661 (PCT/US2013/074743), WO 2014/093694 (PCT/US2013/074790), WO 2014/093595 (PCT/US2013/074611), WO 2014/093718 (PCT/US2013/074825), WO 2014/093709 (PCT/US2013/074812), WO 2014/093622 (PCT/US2013/074667), WO 2014/093635 (PCT/US2013/074691), WO 2014/093655 (PCT/US2013/074736), WO 2014/093712 (PCT/US2013/074819), WO2014/093701 (PCT/US2013/074800), and WO2014/018423 (PCT/US2013/051418). Reference is also made to 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.: 61/915,251, 61/915,301 and 61/915,260, each filed December 12, 2013; 61/930,214, filed January 22, 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. Each of these patents, patent publications, and applications, and all documents cited therein or during their prosecution ("appln cited documents") and all documents cited or referenced in the appln cited documents, together with any instructions, descriptions, product specifications, and product sheets for any products mentioned therein or in any document therein and incorporated by reference herein, are hereby incorporated herein by reference, and may be employed in the practice of the invention. All documents (e.g., these patents, patent publications and applications and the appln cited documents) are incorporated herein by reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

[0109] Also with respect to general information on CRISPR-Cas Systems, mention is made of:

- *Multiplex genome engineering using CRISPR/Cas systems*. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, F. *Science* Feb 15;339(6121):819-23 (2013);
- *RNA-guided editing of bacterial genomes using CRISPR-Cas systems*. Jiang W., Bikard D., Cox D., Zhang F, Marraffini LA. Nat Biotechnol Mar;31(3):233-9 (2013);
- *One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering*. Wang H., Yang H., Shivalila CS., Dawlaty MM., Cheng AW., Zhang F., Jaenisch R. Cell May 9;153(4):910-8 (2013);
- *Optical control of mammalian endogenous transcription and epigenetic states*. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. Nature. 2013 Aug 22;500(7463):472-6. doi: 10.1038/Nature12466. Epub 2013 Aug 23;

- *Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity*. Ran, FA., Hsu, PD., Lin, CY., Gootenberg, JS., Konermann, S., Trevino, AE., Scott, DA., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. *Cell* Aug 28. pii: S0092-8674(13)01015-5. (2013*);*
- *DNA targeting specificity of RNA-guided Cas9 nucleases*. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. *Nat Biotechnol* doi:10.1038/nbt.2647 (2013);
- *Genome engineering using the CRISPR-Cas9 system*. Ran, FA., Hsu, PD., Wright, J., Agarwala, V., Scott, DA., Zhang, F. Nature Protocols Nov;8(11):2281-308. (2013);
- *Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells*. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelson, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. *Science* Dec 12. (2013). [Epub ahead of print];
- *Crystal structure of cas9 in complex with guide RNA and target DNA*. Nishimasu, H., Ran, FA., Hsu, PD., Konermann, S., Shehata, SI., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. *Cell* Feb 27. (2014). 156(5):935-49;
- *Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells*. Wu X., Scott DA., Kriz AJ., Chiu AC., Hsu PD., Dadon DB., Cheng AW., Trevino AE., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp PA. *Nat Biotechnol*. (2014) Apr 20. doi: 10.1038/nbt.2889,
- *CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling*, Platt et al., Cell 159(2): 440-455 (2014) DOI: 10.1016/j.cell.2014.09.014,
- *Development and Applications of CRISPR-Cas9 for Genome Engineering*, Hsu et al, Cell 157, 1262-1278 (June 5, 2014) (Hsu 2014),
- *Genetic screens in human cells using the CRISPR/Cas9 system*, Wang et al., Science. 2014 January 3; 343(6166): 80–84. doi:10.1126/science.1246981,
- *Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation*, Doench et al., Nature Biotechnology published online 3 September 2014; doi:10.1038/nbt.3026, and
- *In vivo interrogation of gene function in the mammalian brain using CRISPR-Cas9*, Swiech et al, Nature Biotechnology ; published online 19 October 2014; doi:10.1038/nbt.3055.

each of which is incorporated herein by reference, and discussed briefly below:

- Cong *et al.* engineered type II CRISPR-Cas systems for use in eukaryotic cells based on both *Streptococcus thermophilus* Cas9 and also *Streptoccocus pyogenes* Cas9 and demonstrated that Cas9 nucleases can be directed by short RNAs to induce precise cleavage of DNA in human and mouse cells. Their study further showed that Cas9 as converted into a nicking enzyme can be used to facilitate homology-directed repair in eukaryotic cells with minimal mutagenic activity. Additionally, their study demonstrated that multiple guide sequences can be encoded into a single CRISPR array to enable simultaneous editing of several at endogenous genomic loci sites within the mammalian genome, demonstrating easy programmability and wide applicability of the RNA-guided nuclease technology. This ability to use RNA to program sequence specific DNA cleavage in cells defined a new class of genome engineering tools. These studies further showed that other CRISPR loci are likely to be transplantable into mammalian cells and can also mediate mammalian genome cleavage. Importantly, it can be envisaged that several aspects of the CRISPR-Cas system can be further improved to increase its efficiency and versatility.
- Jiang *et al.* used the clustered, regularly interspaced, short palindromic repeats (CRISPR)–associated Cas9 endonuclease complexed with dual-RNAs to introduce precise mutations in the genomes of *Streptococcus pneumoniae* and *Escherichia coli.* The approach relied on dual-RNA:Cas9-directed cleavage at the targeted genomic site to kill unmutated cells and circumvents the need for selectable markers or counter-selection systems. The study reported reprogramming dual-RNA:Cas9 specificity by changing the sequence of short CRISPR RNA (crRNA) to make single- and multinucleotide changes carried on editing templates. The study showed that simultaneous use of two crRNAs enabled multiplex mutagenesis. Furthermore, when the approach was used in combination with recombineering, in *S. pneumoniae*, nearly 100% of cells that were recovered using the described approach contained the desired mutation, and in *E. coli,* 65% that were recovered contained the mutation.
- Konermann *et al.* addressed the need in the art for versatile and robust technologies that enable optical and chemical modulation of DNA-binding domains based CRISPR Cas9 enzyme and also Transcriptional Activator Like Effectors
- As discussed in the present specification, the Cas9 nuclease from the microbial CRISPR-Cas system is targeted to specific genomic loci by a 20 nt guide sequence, which can tolerate certain mismatches to the DNA target and thereby promote undesired off-target

mutagenesis. To address this, Ran *et al.* described an approach that combined a Cas9 nickase mutant with paired guide RNAs to introduce targeted double-strand breaks. Because individual nicks in the genome are repaired with high fidelity, simultaneous nicking *via* appropriately offset guide RNAs is required for double-stranded breaks and extends the number of specifically recognized bases for target cleavage. The authors demonstrated that using paired nicking can reduce off-target activity by 50- to 1,500-fold in cell lines and to facilitate gene knockout in mouse zygotes without sacrificing on-target cleavage efficiency. This versatile strategy enables a wide variety of genome editing applications that require high specificity.

- Hsu *et al.* characterized SpCas9 targeting specificity in human cells to inform the selection of target sites and avoid off-target effects. The study evaluated >700 guide RNA variants and SpCas9-induced indel mutation levels at >100 predicted genomic off-target loci in 293T and 293FT cells. The authors that SpCas9 tolerates mismatches between guide RNA and target DNA at different positions in a sequence-dependent manner, sensitive to the number, position and distribution of mismatches. The authors further showed that SpCas9-mediated cleavage is unaffected by DNA methylation and that the dosage of SpCas9 and sgRNA can be titrated to minimize off-target modification. Additionally, to facilitate mammalian genome engineering applications, the authors reported providing a web-based software tool to guide the selection and validation of target sequences as well as off-target analyses.
- Ran *et al.* described a set of tools for Cas9-mediated genome editing *via* non-homologous end joining (NHEJ) or homology-directed repair (HDR) in mammalian cells, as well as generation of modified cell lines for downstream functional studies. To minimize off-target cleavage, the authors further described a double-nicking strategy using the Cas9 nickase mutant with paired guide RNAs. The protocol provided by the authors experimentally derived guidelines for the selection of target sites, evaluation of cleavage efficiency and analysis of off-target activity. The studies showed that beginning with target design, gene modifications can be achieved within as little as 1–2 weeks, and modified clonal cell lines can be derived within 2–3 weeks.
- Shalem *et al.* described a new way to interrogate gene function on a genome-wide scale. Their studies showed that delivery of a genome-scale CRISPR-Cas9 knockout (GeCKO) library targeted 18,080 genes with 64,751 unique guide sequences enabled both negative and positive selection screening in human cells. First, the authors showed use of the

GeCKO library to identify genes essential for cell viability in cancer and pluripotent stem cells. Next, in a melanoma model, the authors screened for genes whose loss is involved in resistance to vemurafenib, a therapeutic that inhibits mutant protein kinase BRAF. Their studies showed that the highest-ranking candidates included previously validated genes NF1 and MED12 as well as novel hits NF2, CUL3, TADA2B, and TADA1. The authors observed a high level of consistency between independent guide RNAs targeting the same gene and a high rate of hit confirmation, and thus demonstrated the promise of genome-scale screening with Cas9.

- Nishimasu *et al.* reported the crystal structure of *Streptococcus pyogenes* Cas9 in complex with sgRNA and its target DNA at 2.5 A° resolution. The structure revealed a bilobed architecture composed of target recognition and nuclease lobes, accommodating the sgRNA:DNA heteroduplex in a positively charged groove at their interface. Whereas the recognition lobe is essential for binding sgRNA and DNA, the nuclease lobe contains the HNH and RuvC nuclease domains, which are properly positioned for cleavage of the complementary and non-complementary strands of the target DNA, respectively. The nuclease lobe also contains a carboxyl-terminal domain responsible for the interaction with the protospacer adjacent motif (PAM). This high-resolution structure and accompanying functional analyses have revealed the molecular mechanism of RNA-guided DNA targeting by Cas9, thus paving the way for the rational design of new, versatile genome-editing technologies.
- Wu *et al.* mapped genome-wide binding sites of a catalytically inactive Cas9 (dCas9) from *Streptococcus pyogenes* loaded with single guide RNAs (sgRNAs) in mouse embryonic stem cells (mESCs). The authors showed that each of the four sgRNAs tested targets dCas9 to between tens and thousands of genomic sites, frequently characterized by a 5-nucleotide seed region in the sgRNA and an NGG protospacer adjacent motif (PAM). Chromatin inaccessibility decreases dCas9 binding to other sites with matching seed sequences; thus 70% of off-target sites are associated with genes. The authors showed that targeted sequencing of 295 dCas9 binding sites in mESCs transfected with catalytically active Cas9 identified only one site mutated above background levels. The authors proposed a two-state model for Cas9 binding and cleavage, in which a seed match triggers binding but extensive pairing with target DNA is required for cleavage.
- Hsu 2014 is a review article that discusses generally CRISPR-Cas9 history from yogurt to genome editing, including genetic screening of cells, that is in the information, data

and findings of the applications in the lineage of this specification filed prior to June 5, 2014. The general teachings of Hsu 2014 do not involve the specific models, animals of the instant specification.

[0110] In general, the CRISPR-Cas or CRISPR system is as used in the foregoing documents, such as WO 2014/093622 (PCT/US2013/074667) and refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or "RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas9, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell. 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. In some embodiments it may be preferred in a CRISPR complex that the tracr sequence has one or more hairpins and is 30 or more nucleotides in length, 40 or more nucleotides in length, or 50 or more nucleotides in length; the guide sequence is between 10 to 30 nucleotides in length, the CRISPR/Cas enzyme is a Type II Cas9 enzyme. In embodiments of the invention the terms guide sequence and guide RNA 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. 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. For example, for the S. pyogenes Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXGG where NNNNNNNNNNNNXGG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. A unique target sequence in a genome may include an S. pyogenes Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNXGG where NNNNNNNNNNNXGG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. For the S. thermophilus CRISPR1 Cas9, a unique target sequence in a genome may include a Cas9 target

site of the form MMMMMMMMNNNNNNNNNNNNXXAGAAW where NNNNNNNNNNNNXXAGAAW (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. A unique target sequence in a genome may include an S. thermophilus CRISPR1 Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNNXXAGAAW where NNNNNNNNNNNXXAGAAW (N is A, G, T, or C; X can be anything; and W is A or T) has a single occurrence in the genome. For the S. pyogenes Cas9, a unique target sequence in a genome may include a Cas9 target site of the form MMMMMMMMNNNNNNNNNNNNXGGXG where NNNNNNNNNNNNXGGXG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. A unique target sequence in a genome may include an S. pyogenes Cas9 target site of the form MMMMMMMMMNNNNNNNNNNNXGGXG where NNNNNNNNNNNXGGXG (N is A, G, T, or C; and X can be anything) has a single occurrence in the genome. In each of these sequences "M" may be A, G, T, or C, and need not be considered in identifying a sequence as unique. In some embodiments, a guide sequence is selected to reduce the degree secondary structure within the guide sequence. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the guide sequence participate in self-complementary base pairing when optimally folded. Optimal folding may be determined by any suitable polynucleotide folding algorithm. Some programs are based on calculating the minimal Gibbs free energy. An example of one such algorithm is mFold, as described by Zuker and Stiegler (Nucleic Acids Res. 9 (1981), 133-148). Another example folding algorithm is the online webserver RNAfold, developed at Institute for Theoretical Chemistry at the University of Vienna, using the centroid structure prediction algorithm (see e.g. A.R. Gruber et al., 2008, Cell 106(1): 23-24; and PA Carr and GM Church, 2009, Nature Biotechnology 27(12): 1151-62).

[0111] In general, a tracr mate sequence includes any sequence that has sufficient complementarity with a tracr sequence to promote one or more of: (1) excision of a guide sequence flanked by tracr mate sequences in a cell containing the corresponding tracr sequence; and (2) formation of a CRISPR complex at a target sequence, wherein the CRISPR complex comprises the tracr mate sequence hybridized to the tracr sequence. In general, degree of complementarity is with reference to the optimal alignment of the tracr mate sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the tracr sequence or tracr mate sequence.

In some embodiments, the degree of complementarity between the tracr sequence and tracr mate sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher. In some embodiments, the tracr sequence is about or more than about 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 40, 50, or more nucleotides in length. In some embodiments, the tracr sequence and tracr mate sequence are contained within a single transcript, such that hybridization between the two produces a transcript having a secondary structure, such as a hairpin. In an embodiment of the invention, the transcript or transcribed polynucleotide sequence has at least two or more hairpins. In preferred embodiments, the transcript has two, three, four or five hairpins. In a further embodiment of the invention, the transcript has at most five hairpins. In a hairpin structure the portion of the sequence 5' of the final "N" and upstream of the loop corresponds to the tracr mate sequence, and the portion of the sequence 3' of the loop corresponds to the tracr sequence Further non-limiting examples of single polynucleotides comprising a guide sequence, a tracr mate sequence, and a tracr sequence are as follows (listed 5' to 3'), where "N" represents a base of a guide sequence, the first block of lower case letters represent the tracr mate sequence, and the second block of lower case letters represent the tracr sequence, and the final poly-T sequence represents the transcription terminator: (1) NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagataa ggcttcatgccgaaatcaacaccctgtcatttatggcagggtgtttcgttatttaaTTTTTT; (2) NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccg aaatcaacaccctgtcatttatggcagggtgtttcgttatttaaTTTTTT; (3) NNNNNNNNNNNNNNNNNNNNgtttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccg aaatcaacaccctgtcatttatggcagggtgtTTTTTT; (4) NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaactt gaaaaagtggcaccgagtcggtgcTTTTTT; (5) NNNNNNNNNNNNNNNNNNNNgttttagagctaGAAATAGcaagttaaaataaggctagtccgttatcaac ttgaaaaagtgTTTTTTT; and (6) NNNNNNNNNNNNNNNNNNNNgttttagagctagAAATAGcaagttaaaataaggctagtccgttatcaTT TTTTTT. In some embodiments, sequences (1) to (3) are used in combination with Cas9 from *S. thermophilus* CRISPR1. In some embodiments, sequences (4) to (6) are used in combination with Cas9 from *S. pyogenes*. In some embodiments, the tracr sequence is a separate transcript from a transcript comprising the tracr mate sequence.

[0112] In some embodiments, candidate tracrRNA may be subsequently predicted by sequences that fulfill any or all of the following criteria: 1. sequence homology to direct repeats (motif search in Geneious with up to 18-bp mismatches); 2. presence of a predicted Rho-independent transcriptional terminator in direction of transcription; and 3. stable hairpin secondary structure between tracrRNA and direct repeat. In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[0113] In some embodiments, chimeric synthetic guide RNAs (sgRNAs) designs may incorporate at least 12 bp of duplex structure between the direct repeat and tracrRNA.

[0114] For minimization of toxicity and off-target effect, it will be important to control the concentration of CRISPR enzyme mRNA and guide RNA delivered. Optimal concentrations of CRISPR enzyme mRNA and guide RNA can be determined by testing different concentrations in a cellular or non-human eukaryote animal model and using deep sequencing the analyze the extent of modification at potential off-target genomic loci. For example, for the guide sequence targeting 5'-GAGTCCGAGCAGAAGAAGAA-3' in the EMX1 gene of the human genome, deep sequencing can be used to assess the level of modification at the following two off-target loci, 1: 5'-GAGTCCTAGCAGGAGAAGAA-3' and 2: 5'-GAGTCTAAGCAGAAGAAGAA-3'. 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. Alternatively, to minimize the level of toxicity and off-target effect, CRISPR enzyme nickase mRNA (for example S. pyogenes Cas9 with the D10A mutation) can be delivered with a pair of guide RNAs targeting a site of interest. The two guide RNAs need to be spaced as follows. Guide sequences and strategies to mimize toxicity and off-target effects can be as in WO 2014/093622 (PCT/US2013/074667).

[0115] The CRISPR system is derived advantageously from a type II CRISPR system. In some embodiments, one or more elements of a CRISPR system is derived from a particular organism comprising an endogenous CRISPR system, such as *Streptococcus pyogenes*. In preferred embodiments of the invention, the CRISPR system is a type II CRISPR system and the Cas enzyme is Cas9, which catalyzes DNA cleavage. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10,

Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologues thereof, or modified versions thereof.

[0116] In some embodiments, the unmodified CRISPR enzyme has DNA cleavage activity, such as Cas9. In some embodiments, the CRISPR enzyme directs cleavage of one or both strands at the location of a target sequence, such as within the target sequence and/or within the complement of the target sequence. In some embodiments, the CRISPR enzyme directs cleavage of one or both strands within about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, 25, 50, 100, 200, 500, or more base pairs from the first or last nucleotide of a target sequence. In some embodiments, a vector encodes a CRISPR enzyme that is mutated to with respect to a corresponding wild-type enzyme such that the mutated CRISPR enzyme lacks the ability to cleave one or both strands of a target polynucleotide containing a target sequence. For example, an aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of Cas9 from S. pyogenes converts Cas9 from a nuclease that cleaves both strands to a nickase (cleaves a single strand). Other examples of mutations that render Cas9 a nickase include, without limitation, H840A, N854A, and N863A. As a further example, two or more catalytic domains of Cas9 (RuvC I, RuvC II, and RuvC III or the HNH domain) may be mutated to produce a mutated Cas9 substantially lacking all DNA cleavage activity. In some embodiments, a D10A mutation is combined with one or more of H840A, N854A, or N863A mutations to produce a Cas9 enzyme substantially lacking all DNA cleavage activity. In some embodiments, a CRISPR enzyme is considered to substantially lack all DNA cleavage activity when the DNA cleavage activity of the mutated enzyme is about no more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the DNA cleavage activity of the non-mutated form of the enzyme; an example can be when the DNA cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form. Where the enzyme is not SpCas9, mutations may be made at any or all residues corresponding to positions 10, 762, 840, 854, 863 and/or 986 of SpCas9 (which may be ascertained for instance by standard sequence comparison tools). In particular, any or all of the following mutations are preferred in SpCas9: D10A, E762A, H840A, N854A, N863A and/or D986A; as well as conservative substitution for any of the replacement amino acids is also envisaged. The same (or conservative substitutions of these mutations) at corresponding positions in other Cas9s are also preferred. Particularly preferred are D10 and H840 in SpCas9. However, in other Cas9s, residues corresponding to SpCas9 D10 and H840 are also preferred. Orthologs of SpCas9 can be used in the practice of the invention. A Cas enzyme may be identified Cas9 as this can refer to the general class of enzymes that share homology to the biggest nuclease with multiple nuclease domains from the type II

CRISPR system. Most preferably, the Cas9 enzyme is from, or is derived from, spCas9 (*S. pyogenes* Cas9) or saCas9 (*S. aureus* Cas9). StCas9" refers to wild type Cas9 from S. thermophilus, the protein sequence of which is given in the SwissProt database under accession number G3ECR1. Similarly, S pyogenes Cas9 or spCas9 is included in SwissProt under accession number Q99ZW2. By derived, Applicants mean that the derived enzyme is largely based, in the sense of having a high degree of sequence homology with, a wildtype enzyme, but that it has been mutated (modified) in some way as described herein. It will be appreciated that the terms Cas and CRISPR enzyme are generally used herein interchangeably, unless otherwise apparent. As mentioned above, many of the residue numberings used herein refer to the Cas9 enzyme from the type II CRISPR locus in *Streptococcus pyogenes*. However, it will be appreciated that this invention includes many more Cas9s from other species of microbes, such as SpCas9, SaCa9, St1Cas9 and so forth. Enzymatic action by Cas9 derived from *Streptococcus pyogenes* or any closely related Cas9 generates double stranded breaks at target site sequences which hybridize to 20 nucleotides of the guide sequence and that have a protospacer-adjacent motif (PAM) sequence (examples include NGG/NRG or a PAM that can be determined as described herein) following the 20 nucleotides of the target sequence. CRISPR activity through Cas9 for site-specific DNA recognition and cleavage is defined by the guide sequence, the tracr sequence that hybridizes in part to the guide sequence and the PAM sequence. More aspects of the CRISPR system are described in Karginov and Hannon, The CRISPR system: small RNA-guided defence in bacteria and archaea, Mole Cell 2010, January 15; 37(1): 7. The type II CRISPR locus from *Streptococcus pyogenes* SF370, which contains a cluster of four genes Cas9, Cas1, Cas2, and Csn1, as well as two non-coding RNA elements, tracrRNA and a characteristic array of repetitive sequences (direct repeats) interspaced by short stretches of non-repetitive sequences (spacers, about 30bp each). In this system, targeted DNA double-strand break (DSB) is generated in four sequential steps. First, two non-coding RNAs, the pre-crRNA array and tracrRNA, are transcribed from the CRISPR locus. Second, tracrRNA hybridizes to the direct repeats of pre-crRNA, which is then processed into mature crRNAs containing individual spacer sequences. Third, the mature crRNA:tracrRNA complex directs Cas9 to the DNA target consisting of the protospacer and the corresponding PAM via heteroduplex formation between the spacer region of the crRNA and the protospacer DNA. Finally, Cas9 mediates cleavage of target DNA upstream of PAM to create a DSB within the protospacer. A pre-crRNA array consisting of a single spacer flanked by two direct repeats (DRs) is also encompassed by the term "tracr-mate sequences"). In certain embodiments, Cas9 may be constitutively present or

inducibly present or conditionally present or administered or delivered. Cas9 optimization may be used to enhance function or to develop new functions, one can generate chimeric Cas9 proteins. And Cas9 may be used as a generic DNA binding protein.

[0117] Typically, in the context of an endogenous CRISPR system, formation of a CRISPR complex (comprising a guide sequence hybridized to a target sequence and complexed with one or more Cas proteins) results in cleavage of one or both strands in or near (e.g. within 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 50, or more base pairs from) the target sequence. Without wishing to be bound by theory, the tracr sequence, which may comprise or consist of all or a portion of a wild-type tracr sequence (e.g. about or more than about 20, 26, 32, 45, 48, 54, 63, 67, 85, or more nucleotides of a wild-type tracr sequence), may also form part of a CRISPR complex, such as by hybridization along at least a portion of the tracr sequence to all or a portion of a tracr mate sequence that is operably linked to the guide sequence.

[0118] 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 CRISPR enzyme 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 www.kazusa.orjp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. "Codon usage tabulated from the international DNA sequence databases: status for the year 2000" Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. In some embodiments, one or more codons (e.g. 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a CRISPR enzyme correspond to the most frequently used codon for a particular amino acid.

[0119] In some embodiments, a vector encodes a CRISPR enzyme comprising 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 CRISPR enzyme 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 CRISPR enzyme comprises at most 6 NLSs. In some embodiments, an NLS is considered near the N- or C-terminus when the nearest amino acid of the NLS is within about 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 40, 50, or more amino acids along the polypeptide chain from the N- or C-terminus. Non-limiting examples of NLSs include an NLS sequence derived from: the NLS of the SV40 virus large T-antigen, having the amino acid sequence PKKKRKV; the NLS from nucleoplasmin (e.g. the nucleoplasmin bipartite NLS with the sequence KRPAATKKAGQAKKKK); the c-myc NLS having the amino acid sequence PAAKRVKLD or RQRRNELKRSP; the hRNPA1 M9 NLS having the sequence NQSSNFGPMKGGNFGGRSSGPYGGGGQYFAKPRNQGGY; the sequence RMRIZFKNKGKDTAELRRRRVEVSVELRKAKKDEQILKRRNV of the IBB domain from importin-alpha; the sequences VSRKRPRP and PPKKARED of the myoma T protein; the sequence POPKKKPL of human p53; the sequence SALIKKKKKMAP of mouse c-abl IV; the

sequences DRLRR and PKQKKRK of the influenza virus NS1; the sequence RKLKKKIKKL of the Hepatitis virus delta antigen; the sequence REKKKFLKRR of the mouse Mx1 protein; the sequence KRKGDEVDGVDEVAKKKSKK of the human poly(ADP-ribose) polymerase; and the sequence RKCLQAGMNLEARKTKK of the steroid hormone receptors (human) glucocorticoid. In general, the one or more NLSs are of sufficient strength to drive accumulation of the CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In general, strength of nuclear localization activity may derive from the number of NLSs in the CRISPR enzyme, 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 CRISPR enzyme, 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 CRISPR enzyme activity), as compared to a control no exposed to the CRISPR enzyme or complex, or exposed to a CRISPR enzyme lacking the one or more NLSs.

[0120] Aspects of the invention relate to the expression of the gene product being decreased or a template polynucleotide being further introduced into the DNA molecule encoding the gene product or an intervening sequence being excised precisely by allowing the two 5' overhangs to reanneal and ligate or the activity or function of the gene product being altered or the expression of the gene product being increased. In an embodiment of the invention, the gene product is a protein. Only sgRNA pairs creating 5' overhangs with less than 8bp overlap between the guide sequences (offset greater than -8 bp) were able to mediate detectable indel formation. Importantly, each guide used in these assays is able to efficiently induce indels when paired with wildtype Cas9, indicating that the relative positions of the guide pairs are the most important parameters in predicting double nicking activity. Since Cas9n and Cas9H840A nick opposite strands of DNA, substitution of Cas9n with Cas9H840A with a given sgRNA pair should have resulted in the inversion of the overhang type; but no indel formation is observed as with Cas9H840A indicating that Cas9H840A is a CRISPR enzyme substantially lacking all DNA cleavage activity (which is when the DNA cleavage activity of the mutated enzyme is about no

more than 25%, 10%, 5%, 1%, 0.1%, 0.01%, or less of the DNA cleavage activity of the non-mutated form of the enzyme; whereby an example can be when the DNA cleavage activity of the mutated form is nil or negligible as compared with the non-mutated form, e.g., when no indel formation is observed as with Cas9H840A in the eukaryotic system in contrast to the biochemical or prokaryotic systems). Nonetheless, a pair of sgRNAs that will generate a 5' overhang with Cas9n should in principle generate the corresponding 3' overhang instead, and double nicking. Therefore, sgRNA pairs that lead to the generation of a 3' overhang with Cas9n can be used with another mutated Cas9 to generate a 5' overhang, and double nicking. Accordingly, in some embodiments, a recombination template is also provided. A recombination template may be a component of another vector as described herein, contained in a separate vector, or provided as a separate polynucleotide. In some embodiments, a recombination template is designed to serve as a template in homologous recombination, such as within or near a target sequence nicked or cleaved by a CRISPR enzyme as a part of a CRISPR complex. A template polynucleotide may be of any suitable length, such as about or more than about 10, 15, 20, 25, 50, 75, 100, 150, 200, 500, 1000, or more nucleotides in length. In some embodiments, the template polynucleotide is complementary to a portion of a polynucleotide comprising the target sequence. When optimally aligned, a template polynucleotide might overlap with one or more nucleotides of a target sequences (e.g. about or more than about 1, 5, 10, 15, 20, or more nucleotides). In some embodiments, when a template sequence and a polynucleotide comprising a target sequence are optimally aligned, the nearest nucleotide of the template polynucleotide is within about 1, 5, 10, 15, 20, 25, 50, 75, 100, 200, 300, 400, 500, 1000, 5000, 10000, or more nucleotides from the target sequence.

[0121] In some embodiments, one or more vectors driving expression of one or more elements of a CRISPR system are introduced into a host cell such that expression of the elements of the CRISPR system direct formation of a CRISPR complex at one or more target sites. For example, a Cas enzyme, a guide sequence linked to a tracr-mate sequence, and a tracr sequence could each be operably linked to separate regulatory elements on separate vectors. Or, RNA(s) of the CRISPR System can be delivered to a transgenic Cas9 animal or mammal, e.g., an animal or mammal that constitutively or inducibly or conditionally expresses Cas9; or an animal or mammal that is otherwise expressing Cas9 or has cells containing Cas9, such as by way of prior administration thereto of a vector or vectors that code for and express in vivo Cas9. Alternatively, two or more of the elements expressed from the same or different regulatory elements, may be combined in a single vector, with one or more additional vectors providing any

components of the CRISPR system not included in the first vector. CRISPR system elements that are combined in a single vector may be arranged in any suitable orientation, such as one element located 5' with respect to ("upstream" of) or 3' with respect to ("downstream" of) a second element. The coding sequence of one element may be located on the same or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In some embodiments, a single promoter drives expression of a transcript encoding a CRISPR enzyme and one or more of the guide sequence, tracr mate sequence (optionally operably linked to the guide sequence), and a tracr sequence embedded within one or more intron sequences (e.g. each in a different intron, two or more in at least one intron, or all in a single intron). In some embodiments, the CRISPR enzyme, guide sequence, tracr mate sequence, and tracr sequence are operably linked to and expressed from the same promoter. Delivery vehicles, vectors, particles, nanoparticles, formulations and components thereof for expression of one or more elements of a CRISPR system are as used in the foregoing documents, such as WO 2014/093622 (PCT/US2013/074667). In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a "cloning site"). In some embodiments, one or more insertion sites (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors. In some embodiments, a vector comprises an insertion site upstream of a tracr mate sequence, and optionally downstream of a regulatory element operably linked to the tracr mate sequence, such that following insertion of a guide sequence into the insertion site and upon expression the guide sequence directs sequence-specific binding of a CRISPR complex to a target sequence in a eukaryotic cell. In some embodiments, a vector comprises two or more insertion sites, each insertion site being located between two tracr mate sequences so as to allow insertion of a guide sequence at each site. In such an arrangement, the two or more guide sequences may comprise two or more copies of a single guide sequence, two or more different guide sequences, or combinations of these. When multiple different guide sequences are used, a single expression construct may be used to target CRISPR activity to multiple different, corresponding target sequences within a cell. For example, a single vector may comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide sequences. In some embodiments, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-sequence-containing vectors may be provided, and optionally delivered to a cell. In some embodiments, a vector comprises a regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, such as a Cas protein.

CRISPR enzyme or CRISPR enzyme mRNA or CRISPR guide RNA or RNA(s) can be delivered separately; and advantageously at least one of these is delivered via a nanoparticle complex. CRISPR enzyme mRNA can be delivered prior to the guide RNA to give time for CRISPR enzyme to be expressed. CRISPR enzyme mRNA might be administered 1-12 hours (preferably around 2-6 hours) prior to the administration of guide RNA. Alternatively, CRISPR enzyme mRNA and guide RNA can be administered together. Advantageously, a second booster dose of guide RNA can be administered 1-12 hours (preferably around 2-6 hours) after the initial administration of CRISPR enzyme mRNA + guide RNA. Additional administrations of CRISPR enzyme mRNA and/or guide RNA might be useful to achieve the most efficient levels of genome modification.

[0122] In one aspect, the invention provides methods for using one or more elements of a CRISPR system. The CRISPR complex of the invention provides an effective means for modifying a target polynucleotide. The CRISPR complex of the invention has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating) a target polynucleotide in a multiplicity of cell types. As such the CRISPR complex of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis. An exemplary CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within the target polynucleotide. The guide sequence is linked to a tracr mate sequence, which in turn hybridizes to a tracr sequence. In one embodiment, this invention provides a method of cleaving a target polynucleotide. The method comprises modifying a target polynucleotide using a CRISPR complex that binds to the target polynucleotide and effect cleavage of said target polynucleotide. Typically, the CRISPR complex of the invention, when introduced into a cell, creates a break (e.g., a single or a double strand break) in the genome sequence. For example, the method can be used to cleave a disease gene in a cell. The break created by the CRISPR complex can be repaired by a repair processes such as the error prone non-homologous end joining (NHEJ) pathway or the high fidelity homology-directed repair (HDR). During these repair process, an exogenous polynucleotide 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). Without wishing to be bound by theory, it is believed that the target sequence should be associated with a PAM (protospacer adjacent motif); that is, a short sequence recognized by the CRISPR complex. The precise sequence and length requirements for the PAM differ depending on the CRISPR enzyme used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence) Examples of PAM sequences are given in the examples section below, and the skilled person will be able to identify further PAM sequences for use with a given CRISPR enzyme. In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence within said polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. Similar considerations and conditions apply as above for methods of modifying a target polynucleotide. In fact, these sampling, culturing and re-introduction options apply across the aspects of the present invention. In one aspect, the invention provides for methods of modifying a target polynucleotide in a eukaryotic cell, which may be *in vivo*, *ex vivo* or *in vitro*. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal, and modifying the cell or cells. Culturing may occur at any stage *ex vivo*. The cell or cells may even be re-introduced into the non-human animal or plant. For re-introduced cells it is particularly preferred that the cells are stem cells.

[0123] Indeed, in any aspect of the invention, the CRISPR complex may comprise a CRISPR enzyme complexed with a guide sequence hybridized to a target sequence, wherein said guide sequence may be linked to a tracr mate sequence which in turn may hybridize to a tracr sequence.

[0124] The invention relates to the engineering and optimization of systems, methods and compositions used for the control of gene expression involving sequence targeting, such as genome perturbation or gene-editing, that relate to the CRISPR-Cas system and components

thereof. In advantageous embodiments, the Cas enzyme is Cas9. An advantage of the present methods is that the CRISPR system minimizes or avoids off-target binding and its resulting side effects. This is achieved using systems arranged to have a high degree of sequence specificity for the target DNA.

[0125] The CRISPR system is particularly suitable for editing nucleotide repeats, such a trinucleotide repeats or other nucleotide expansion elements. These repeats are a DNA mutation responsible for causing many disorders. Many of these display neurological symptoms, so the use of CRISPR in brain and other central nervous system (CNS) tissues is of particular interest, but non-neurological symptoms are also seen and so delivery to and expression in other tissues is also useful. For example, myotonic dystrophy is caused by expansion of nucleotide repeats (a trinucleotide for DM1, but a tetranucleotide for DM2) but causes muscular dystrophy, cataracts, heart conduction defects, and myotonia, and so diverse target tissues are involved.

Nucleotide repeat expansion disorders

[0126] The CRISPR-Cas9 system is a powerful tool for editing nucleotide repeat expansions which can occur in the genome. These repeats are a mutation in genomic DNA and they are responsible for causing many disorders *e.g.* see 'Human Nucleotide Expansion Disorders' (eds. Fry & Usdin, 2006) Nucleic Acids and Molecular Biology, Vol. 19 (ISBN 978-3-540-33336-4). Most of these disorders are neurodegenerative, but they can affect a variety of tissues.

[0127] The nucleotides that comprise the nucleotide expansion elements involved in diseases and disorders vary, but they are commonly trinucleotide repeats, usually involving CTG, CAG, CGG, CCG, GAA, or TTC. Longer repeats are also seen, such as a CCTG tetranucleotide, ATTCT and AGAAT pentanucleotides, GGGGCC hexanucleotides and CCCCGCCCCGCG and CGCGGGGCGGGG dodecanucleotides. The nature of the CRISPR system means that it is useful for editing all such nucleotide repeats. For example, the use of CRISPR editing of nucleotide repeats includes the excession of the repeat. It is preferred that the excession of the repeat results in the repair to the wildtype. In the presence of multiple repeats, multiple guides may be employed to target the multiple repeats.

[0128] The repeats can occur within coding or within non-coding regions *e.g.* within an exon, a 5'UTR, a 3'UTR, a promoter element or an intron. The invention can be used regardless of the location of the repeat.

[0129] Nucleotide repeat disorders, and in particular trinucleotide repeat disorders and nucleotide expansion disorders, are thus preferred conditions to be treated. These are also exemplified herein.

[0130] For example, 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.

[0131] As mentioned above, nucleotide repeat expansion proteins are a diverse set of proteins associated with susceptibility for developing a nucleotide repeat expansion disorder, the presence of a nucleotide repeat expansion disorder, the severity of a nucleotide 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 X-associated tremor/ataxia syndrome (FXTAS); Fragile XE Mental Retardation (FRAXE); FRAXF; Friedreich Ataxia (FRDA); Myotonic Dystrophy (DM), in particular type 1 (DM1) or the tetranucleotide variant for DM2; and Spinocerebellar Ataxias (SCA types 8, and 12). Other nucleotide expansion disorders include progressive myoclonus epilepsy (12-mer repeat), DM2 myotonic dystrophy (4-mer repeat element), C9orf72 (6-mer repeat element) and SCA type 10 (5-mer repeat element).

[0132] The proteins associated with nucleotide repeat expansion disorders are typically selected based on an experimental association of the protein associated with a nucleotide repeat expansion disorder to a nucleotide repeat expansion disorder. For example, the production rate or circulating concentration of a protein associated with a nucleotide repeat expansion disorder may be elevated or depressed in a population having a nucleotide repeat expansion disorder relative to a population lacking the nucleotide 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 nucleotide 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). Knowing the nucleotide repeat, CRISPR repair involving excision of the repeat, preferably to wildtype, may be excercised. In the same manner, the nucleotide repeat is considered a mutation. The mutation may be repaired by reintroduction of the missing wildtype sequence into the mutant. In such a case a repair template may be used which allows for reintroduction of the missing wildtye sequence into the mutant. Such CRISPR repair may be used to repair any mutation.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.

[0133] Nucleotide repeats vary in size and can reside in coding or non-coding regions of the disease-associated genes. The skilled person will be able to recognize such repeats and whether such repeats are normal or aberrant. Each loci can be targeted using the CRISPR-Cas9 approach described in this application.

[0134] An exemplary, abnormal expansion of a nucleotide repeat sequence to be targeted for CRISPR repair in EPM1 is CCCCGCCCCGCG. Using the described CRISPR repair the repeat is excised from the affected sequence.

[0135] An exemplary, abnormal expansion of a nucleotide repeat sequence to be targeted for CRISPR repair in C9ORF72 is GGGGCC. Using the described CRISPR repair the repeat is excised from the affected sequence.

[0136] An exemplary, abnormal expansion of a nucleotide repeat sequence to be targeted for CRISPR repair in DM2 is CCTG. Using the described CRISPR repair the repeat is excised from the affected sequence.

[0137] An exemplary, abnormal expansion of a nucleotide repeat sequence to be targeted for CRISPR repair in OPMD is GCG/Ala. Using the described CRISPR repair the repeat is excised from the affected sequence.

[0138] An exemplary, abnormal expansion of a nucleotide repeat sequence to be targeted for CRISPR repair in SCA10 is ATTCT. Using the described CRISPR repair the repeat is excised from the affected sequence.

[0139] An exemplary, abnormal expansion of a nucleotide repeat sequence to be targeted for CRISPR repair in Fragile X, FXTAS is CGG. Using the described CRISPR repair the repeat is excised from the affected sequence.

[0140] An exemplary, abnormal expansion of a nucleotide repeat sequence to be targeted for CRISPR repair in SCA12 is CAG. Using the described CRISPR repair the repeat is excised from the affected sequence.

[0141] An exemplary, abnormal expansion of a nucleotide repeat sequence to be targeted for CRISPR repair in Friedreich Ataxia is GAA. Using the described CRISPR repair the repeat is excised from the affected sequence.

[0142] An exemplary, abnormal expansion of a nucleotide repeat sequence to be targeted for CRISPR repair in SCAs 1-3, 6, 7, 17, DRPLA, HD, SBMA, HDL2 (SCAs 8, 12) is CAG/Polyglutamine. Using the described CRISPR repair the repeat is excised from the affected sequence.

[0143] An exemplary, abnormal expansion of a nucleotide repeat sequence to be targeted for CRISPR repair in SCA31 is TGGAA. Using the described CRISPR repair the repeat is excised from the affected sequence.

[0144] An exemplary, abnormal expansion of a nucleotide repeat sequence to be targeted for CRISPR repair in SCA8, DM1 is CTG. Using the described CRISPR repair the repeat is excised from the affected sequence.

[0145] 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.

Huntington's Disease (HD):

[0146] 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), and therefore Applicant postulates that it may be 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 (Htt) 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.

[0147] 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.

[0148] 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 10^{12} viral genomes/ml) into the straiatum. A similar dosage of CRISPR Cas targeted to Htt may be contemplated for humans in the present invention, for example, about 10-20 ml of 4 x 10^{12} viral genomes/ml) CRISPR Cas targeted to Htt may be injected intrastriatally.

[0149] In another example, a CRISPR Cas targetd 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.

[0150] 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.

[0151] In another example, the methods of US Patent Publication No. 20130253040 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.

[0152] Possible target genes of CRISPR complex in regard to Huntington's Disease: PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; and TGM2.

C9ORF72

[0153] C9orf72 (chromosome 9 open reading frame 72) is a protein which in humans encodes a protein found in many regions of the brain, in the cytoplasm of neurons and in presynaptic terminals. Mutation(s) of the C9orf72 gene have been identified that contain a hexanucleotide repeat expansion element of the six letter string of nucleotides GGGGCC. The

mutations in C9orf72 are significant because it is the first pathogenic mechanism identified to be a genetic link between familial frontotemporal dementia (FTD) and amyotrophic lateral sclerosis (ALS).

CFTR

[0154] According to another aspect, a method of gene therapy for the treatment of a subject having a mutation in the CFTR gene is provided and comprises administering a therapeutically effective amount of a CRISPR-Cas gene therapy particle, optionally via a biocompatible pharmaceutical carrier, to the cells of a subject. Preferably, the target DNA comprises the mutation deltaF508. In general, it is of preferred that the mutation is repaired to the wildtype. In this case, the mutation is a deletion of the three nucleotides that comprise the codon for phenylalanine (F) at position 508. Accordingly, repair in this instance requires reintroduction of the missing codon into the mutant.

[0155] To implement this Gene Repair Strategy, it is preferred that an adenovirus/AAV vector system is introduced into the host cell, cells or patient. Preferably, the system comprises a Cas9 (or Cas9 nickase) and the guide RNA along with a adenovirus/AAV vector system comprising the homology repair template containing the F508 residue. This may be introduced into the subject via one of the methods of delivery discussed earlier. The CRISPR-Cas system may be guided by the CFTRdelta 508 chimeric guide RNA. It targets a specific site of the CFTR genomic locus to be nicked or cleaved. After cleavage, the repair template is inserted into the cleavage site via homologous recombination correcting the deletion that results in cystic fibrosis or causes cystic fibrosis related symptoms. This strategy to direct delivery and provide systemic introduction of CRISPR systems with appropriate guide RNAs can be employed to target genetic mutations to edit or otherwise manipulate genes that cause metabolic, liver, kidney and protein diseases and disorders.

[0156] One aspect is an AAV vector engineered for in vivo CRISPR-Cas9-mediated genome editing. Cas9 (e.g. SpCas9) containing an N-terminal and C-terminal nuclear localization domain as well as an N-terminal Flag may be cloned into an AAV shuttle plasmid. Because of possible size constraints of the Cas9 cDNA and the desire to obtain low levels of Cas9 nuclease expression in vivo, the use of a promoter may be omitted. Instead, expression of the Cas9 may be driven by the basal transcriptional activity of the AAV inverted terminal repeat (iTR) sequences. The guide RNA (gcRNA) and the transactivating RNA (tracrRNA) may be cloned into a different AAV shuttle plasmid and placed under the regulation of two different RNA polymerase type-III promoters: the U6 and H1 promoters respectively. A reporter gene (e.g. EGFP), or any

other sequence, can be cloned downstream of the non-coding expression cassettes. In one system, the non-coding CRISPR components may be expressed as an array of chimeras (sgRNAs) driven by the U6 promoter. Such AAV plasmids may be used, e.g., to target ATXN1 plasmids that carry either 30 CAG nucleotide repeats (normal range) or 80 CAG repeats (disease range).

[0157] Vectors may use the Cas9 nuclease containing an N-terminal nuclear localization domain as well as an N-terminal HA-tag, cloned into an AAV shuttle plasmid and placed under the control of a CMV promoter. The non-coding RNA elements required for Cas9-mediated gene editing are also contained within the same AAV packaging genome. This allows for the co-delivery of a second AAV vector that may serve as a transduction marker or a template donor whenever HR is desired. Successful vector delivery may be indicated by expression of a marker (e.g. EGFP). AAV vectors may be used for delivery of CRISPR-Cas9 system into mammalian tissue.

[0158] Guide sequences flanking the repeat may be removed using the CRISPR-Cas9 system. Guide sequences may be designed to flank the nucleotide repeat region in the 3' un-translated region (3'UTR). Successful editing of the repeat may be confirmed when both flanking guide non-coding RNAs are simultaneously expressed. Sequencing of affected sequence may also be used for confirmation of successful repair. Upon proper excision of the abnormal expansion, preferably to wildtype, the sequence is repaired.

Self-inactivating systems

[0159] Once all copies of a gene in the genome of a cell have been edited, continued CRISRP/Cas9 expression in that cell is no longer necessary. Indeed, sustained expression would be undesirable in case of off-target effects at unintended genomic sites, *etc.* Thus time-limited expression would be useful. Inducible expression offers one approach, but in addition Applicants have engineered a Self-Inactivating CRISPR-Cas9 system that relies on the use of a non-coding guide target sequence within the CRISPR vector itself. Thus, after expression begins, the CRISPR system will lead to its own destruction, but before destruction is complete it will have time to edit the genomic copies of the target gene (which, with a normal point mutation in a diploid cell, requires at most two edits). Simply, the self inactivating CRISPR-Cas system includes additional RNA (i.e., guide RNA) that targets the coding sequence for the CRISPR enzyme itself or that targets one or more non-coding guide target sequences complementary to unique sequences present in one or more of the following:

(a) within the promoter driving expression of the non-coding RNA elements,

(b) within the promoter driving expression of the Cas9 gene,

(c) within 100bp of the ATG translational start codon in the Cas9 coding sequence,

(d) within the inverted terminal repeat (iTR) of a viral delivery vector, e.g., in the AAV genome.

[0160] Furthermore, that RNA can be delivered via a vector, e.g., a separate vector or the same vector that is encoding the CRISPR complex. When provided by a separate vector, the CRISPR RNA that targets Cas expression can be administered sequentially or simultaneously. When administered sequentially, the CRISPR RNA that targets Cas expression is to be delivered after the CRISPR RNA that is intended for e.g. gene editing or gene engineering. This period may be a period of minutes (e.g. 5 minutes, 10 minutes, 20 minutes, 30 minutes, 45 minutes, 60 minutes). This period may be a period of hours (e.g. 2 hours, 4 hours, 6 hours, 8 hours, 12 hours, 24 hours). This period may be a period of days (e.g. 2 days, 3 days, 4 days, 7 days). This period may be a period of weeks (e.g. 2 weeks, 3 weeks, 4 weeks). This period may be a period of months (e.g. 2 months, 4 months, 8 months, 12 months). This period may be a period of years (2 years, 3 years, 4 years). In this fashion, the Cas enzyme associates with a first gRNA/chiRNA capable of hybridizing to a first target, such as a genomic locus or loci of interest and undertakes the function(s) desired of the CRISPR-Cas system (e.g., gene engineering); and subsequently the Cas enzyme may then associate with the second gRNA/chiRNA capable of hybridizing to the sequence comprising at least part of the Cas or CRISPR cassette. Where the gRNA/chiRNA targets the sequences encoding expression of the Cas protein, the enzyme becomes impeded and the system becomes self inactivating. In the same manner, CRISPR RNA that targets Cas expression applied via, for example liposome, lipofection, nanoparticles, microvesicles as explained herein, may be administered sequentially or simultaneously. Similarly, self-inactivation may be used for inactivation of one or more guide RNA used to target one or more targets.

[0161] In some aspects, a single gRNA is provided that is capable of hybridization to a sequence downstream of a CRISPR enzyme start codon, whereby after a period of time there is a loss of the CRISPR enzyme expression. In some aspects, one or more gRNA(s) are provided that are capable of hybridization to one or more coding or non-coding regions of the polynucleotide encoding the CRISPR-Cas system, whereby after a period of time there is a inactivation of one or more, or in some cases all, of the CRISPR-Cas system. In some aspects of the system, and not to be limited by theory, the cell may comprise a plurality of CRISPR-Cas complexes, wherein a first subset of CRISPR complexes comprise a first chiRNA capable of targeting a genomic locus or loci to be edited, and a second subset of CRISPR complexes

comprise at least one second chiRNA capable of targeting the polynucleotide encoding the CRISPR-Cas system, wherein the first subset of CRISPR-Cas complexes mediate editing of the targeted genomic locus or loci and the second subset of CRISPR complexes eventually inactivate the CRISPR-Cas system, thereby inactivating further CRISPR-Cas expression in the cell.

[0162] Thus the invention provides a CRISPR-Cas system comprising one or more vectors for delivery to a eukaryotic cell, wherein the vector(s) encode(s): (i) a CRISPR enzyme; (ii) a first guide RNA capable of hybridizing to a target sequence in the cell; (iii) a second guide RNA capable of hybridizing to one or more target sequence(s) in the vector which encodes the CRISPR enzyme; (iv) at least one tracr mate sequence; and (v) at least one tracr sequence, The first and second complexes can use the same tracr and tracr mate, thus differeing only by the guide sequence, wherein, when expressed within the cell: the first guide RNA directs sequence-specific binding of a first CRISPR complex to the target sequence in the cell; the second guide RNA directs sequence-specific binding of a second CRISPR complex to the target sequence in the vector which encodes the CRISPR enzyme; the CRISPR complexes comprise (a) a tracr mate sequence hybridised to a tracr sequence and (b) a CRISPR enzyme bound to a guide RNA, such that a guide RNA can hybridize to its target sequence; and the second CRISPR complex inactivates the CRISPR-Cas system to prevent continued expression of the CRISPR enzyme by the cell.

[0163] Further characteristics of the vector(s), the encoded enzyme, the guide sequences, *etc.* are disclosed elsewhere herein. For instance, one or both of the guide sequence(s) can be part of a chiRNA sequence which provides the guide, tracr mate and tracr sequences within a single RNA, such that the system can encode (i) a CRISPR enzyme; (ii) a first chiRNA comprising a sequence capable of hybridizing to a first target sequence in the cell, a first tracr mate sequence, and a first tracr sequence; (iii) a second guide RNA capable of hybridizing to the vector which encodes the CRISPR enzyme, a second tracr mate sequence, and a second tracr sequence. Similarly, the enzyme can include one or more NLS, *etc.*

[0164] The various coding sequences (CRISPR enzyme, guide RNAs, tracr and tracr mate) can be included on a single vector or on multiple vectors. For instance, it is possible to encode the enzyme on one vector and the various RNA sequences on another vector, or to encode the enzyme and one chiRNA on one vector, and the remaining chiRNA on another vector, or any other permutation. In general, a system using a total of one or two different vectors is preferred.

[0165] Where multiple vectors are used, it is possible to deliver them in unequal numbers, and ideally with an excess of a vector which encodes the first guide RNA relative to the second

guide RNA, thereby assisting in delaying final inactivation of the CRISPR system until genome editing has had a chance to occur.

[0166] The first guide RNA can target any target sequence of interest within a genome, as described elsewhere herein. The second guide RNA targets a sequence within the vector which encodes the CRISPR Cas9 enzyme, and thereby inactivates the enzyme's expression from that vector. Thus the target sequence in the vector must be capable of inactivating expression. Suitable target sequences can be, for instance, near to or within the translational start codon for the Cas9 coding sequence, in a non-coding sequence in the promoter driving expression of the non-coding RNA elements, within the promoter driving expression of the Cas9 gene, within 100bp of the ATG translational start codon in the Cas9 coding sequence, and/or within the inverted terminal repeat (iTR) of a viral delivery vector, e.g., in the AAV genome. A double stranded break near this region can induce a frame shift in the Cas9 coding sequence, causing a loss of protein expression. An alternative target sequence for the "self-inactivating" guide RNA would aim to edit/inactivate regulatory regions/sequences needed for the expression of the CRISPR-Cas9 system or for the stability of the vector. For instance, if the promoter for the Cas9 coding sequence is disrupted then transcription can be inhibited or prevented. Similarly, if a vector includes sequences for replication, maintenance or stability then it is possible to target these. For instance, in a AAV vector a useful target sequence is within the iTR. Other useful sequences to target can be promoter sequences, polyadenlyation sites, etc.

[0167] Furthermore, if the guide RNAs are expressed in array format, the "self-inactivating" guide RNAs that target both promoters simultaneously will result in the excision of the intervening nucleotides from within the CRISPR-Cas expression construct, effectively leading to its complete inactivation. Similarly, excision of the intervening nucleotides will result where the guide RNAs target both ITRs, or targets two or more other CRISPR-Cas components simultaneously. Self-inactivation as explained herein is applicable, in general, with CRISPR-Cas9 systems in order to provide regulation of the CRISPR-Cas9. For example, self-inactivation as explained herein may be applied to the CRISPR repair of mutations, for example expansion disorders, as explained herein. As a result of this self-inactivation, CRISPR repair is only transiently active.

[0168] Addition of non-targeting nucleotides to the 5' end (e.g. 1 – 10 nucleotides, preferably 1 – 5 nucleotides) of the "self-inactivating" guide RNA can be used to delay its processing and/or modify its efficiency as a means of ensuring editing at the targeted genomic locus prior to CRISPR-Cas9 shutdown.

[0169] In one aspect of the self-inactivating AAV-CRISPR-Cas9 system, plasmids that co-express one or more sgRNA targeting genomic sequences of interest (e.g. 1-2, 1-5, 1-10, 1 -15, 1-20, 1-30) may be established with "self-inactivating" sgRNAs that target an SpCas9 sequence at or near the engineered ATG start site (e.g. within 5 nucleotides, within 15 nucleotides, within 30 nucleotides, within 50 nucleotides, within 100 nucleotides). A regulatory sequence in the U6 promoter region can also be targeted with an sgRNA. The U6-driven sgRNAs may be designed in an array format such that multiple sgRNA sequences can be simultaneously released. When first delivered into target tissue/cells (left cell) sgRNAs begin to accumulate while Cas9 levels rise in the nucleus. Cas9 complexes with all of the sgRNAs to mediate genome editing and self-inactivation of the CRISPR-Cas9 plasmids.

[0170] One aspect of a self-inactivating CRISPR-Cas9 system is expression of singly or in tandam array format from 1 up to 4 or more different guide sequences; e.g. up to about 20 or about 30 guides sequences. Each individual self inactivating guide sequence may target a different target. Such may be processed from, e.g. one chimeric pol3 transcript. Pol3 promoters such as U6 or H1 promoters may be used. Pol2 promoters such as those mentioned throughout herein. Inverted terminal repeat (iTR) sequences may flank the Pol3 promoter - sgRNA(s)-Pol2 promoter- Cas9.

[0171] One aspect of a chimeric, tandem array transcript is that one or more guide(s) edit the one or more target(s) while one or more self inactivating guides inactivate the CRISPR/Cas9 system. Thus, for example, the described CRISPR-Cas9 system for repairing expansion disorders may be directly combined with the self-inactivating CRISPR-Cas9 system described herein. Such a system may, for example, have two guides directed to the target region for repair as well as at least a third guide directed to self-inactivation of the CRISPR-Cas9.

Delivery generally

[0172] Vector delivery, e.g., plasmid, viral delivery: The CRISPR enzyme, for instance a Cas9, and/or any of the present RNAs, for instance a guide RNA, can be delivered using any suitable vector, e.g., plasmid or viral vectors, such as adeno associated virus (AAV), lentivirus, adenovirus or other viral vector types, or combinations thereof. Cas9 and one or more guide RNAs can be packaged into one or more vectors, e.g., plasmid or viral vectors. In some embodiments, the vector, e.g., plasmid or viral vector is delivered to the tissue of interest by, for example, an intramuscular injection, while other times the delivery is via intravenous, transdermal, intranasal, oral, mucosal, or other delivery methods. Such delivery may be either via a single dose, or multiple doses. One skilled in the art understands that the actual dosage to

be delivered herein may vary greatly depending upon a variety of factors, such as the vector choice, the target cell, organism, or tissue, the general condition of the subject to be treated, the degree of transformation/modification sought, the administration route, the administration mode, the type of transformation/modification sought, etc.

[0173] Such a dosage may further contain, for example, a carrier (water, saline, ethanol, glycerol, lactose, sucrose, calcium phosphate, gelatin, dextran, agar, pectin, peanut oil, sesame oil, etc.), a diluent, a pharmaceutically-acceptable carrier (e.g., phosphate-buffered saline), a pharmaceutically-acceptable excipient, and/or other compounds known in the art. The dosage may further contain one or more pharmaceutically acceptable salts such as, for example, a mineral acid salt such as a hydrochloride, a hydrobromide, a phosphate, a sulfate, etc.; and the salts of organic acids such as acetates, propionates, malonates, benzoates, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, gels or gelling materials, flavorings, colorants, microspheres, polymers, suspension agents, etc. may also be present herein. In addition, one or more other conventional pharmaceutical ingredients, such as preservatives, humectants, suspending agents, surfactants, antioxidants, anticaking agents, fillers, chelating agents, coating agents, chemical stabilizers, etc. may also be present, especially if the dosage form is a reconstitutable form. Suitable exemplary ingredients include microcrystalline cellulose, carboxymethylcellulose sodium, polysorbate 80, phenylethyl alcohol, chlorobutanol, potassium sorbate, sorbic acid, sulfur dioxide, propyl gallate, the parabens, ethyl vanillin, glycerin, phenol, parachlorophenol, gelatin, albumin and a combination thereof. A thorough discussion of pharmaceutically acceptable excipients is available in REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991) which is incorporated by reference herein.

[0174] In an embodiment herein the delivery is via an adenovirus, which may be at a single booster dose containing at least 1 x 10^5 particles (also referred to as particle units, pu) of adenoviral vector. In an embodiment herein, the dose preferably is at least about 1 x 10^6 particles (for example, about 1 x 10^6-1 x 10^{12} particles), more preferably at least about 1 x 10^7 particles, more preferably at least about 1 x 10^8 particles (e.g., about 1 x 10^8-1 x 10^{11} particles or about 1 x 10^8-1 x 10^{12} particles), and most preferably at least about 1 x 10^9 particles (e.g., about 1 x 10^9-1 x 10^{10} particles or about 1 x 10^9-1 x 10^{12} particles), or even at least about 1 x 10^{10} particles (e.g., about 1 x 10^{10}-1 x 10^{12} particles) of the adenoviral vector. Alternatively, the dose comprises no more than about 1 x 10^{14} particles, preferably no more than about 1 x 10^{13} particles, even more preferably no more than about 1 x 10^{12} particles, even more preferably no

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

[0175] In an embodiment herein, the delivery is via an AAV. A therapeutically effective dosage for in vivo delivery of the AAV to a human is believed to be in the range of from about 20 to about 50 ml of saline solution containing from about 1 x 10^{10} to about 1 x 10^{10} functional AAV/ml solution. The dosage may be adjusted to balance the therapeutic benefit against any side effects. In an embodiment herein, the AAV dose is generally in the range of concentrations of from about 1 x 10^{5} to 1 x 10^{50} genomes AAV, from about 1 x 10^{8} to 1 x 10^{20} genomes AAV, from about 1 x 10^{10} to about 1 x 10^{16} genomes, or about 1 x 10^{11} to about 1 x 10^{16} genomes AAV. A human dosage may be about 1 x 10^{13} genomes AAV. Such concentrations may be delivered in from about 0.001 ml to about 100 ml, about 0.05 to about 50 ml, or about 10 to about 25 ml of a carrier solution. Other effective dosages can be readily established by one of ordinary skill in the art through routine trials establishing dose response curves. See, for example, U.S. Patent No. 8,404,658 B2 to Hajjar, et al., granted on March 26, 2013, at col. 27, lines 45-60.

[0176] In an embodiment herein the delivery is via a plasmid. In such plasmid compositions, the dosage should be a sufficient amount of plasmid to elicit a response. For instance, suitable quantities of plasmid DNA in plasmid compositions can be from about 0.1 to about 2 mg, or from about 1 μg to about 10 μg per 70 kg individual. Plasmids of the invention will generally comprise (i) a promoter; (ii) a sequence encoding a CRISPR enzyme, operably linked to said promoter; (iii) a selectable marker; (iv) an origin of replication; and (v) a transcription terminator downstream of and operably linked to (ii). The plasmid can also encode the RNA components of a CRISPR complex, but one or more of these may instead be encoded on a different vector.

[0177] The doses herein are based on an average 70 kg individual. The frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician,

veterinarian), or scientist skilled in the art. It is also noted that mice used in experiments are typically about 20g and from mice experiments one can scale up to a 70 kg individual.

[0178] In some embodiments the RNA molecules of the invention are delivered in liposome or lipofectin formulations and the like and can be prepared by methods well known to those skilled in the art. Such methods are described, for example, in U.S. Pat. Nos. 5,593,972, 5,589,466, and 5,580,859, which are herein incorporated by reference. Delivery systems aimed specifically at the enhanced and improved delivery of siRNA into mammalian cells have been developed, (see, for example, Shen et al FEBS Let. 2003, 539:111-114; Xia et al., Nat. Biotech. 2002, 20:1006-1010; Reich et al., Mol. Vision. 2003, 9: 210-216; Sorensen et al., J. Mol. Biol. 2003, 327: 761-766; Lewis et al., Nat. Gen. 2002, 32: 107-108 and Simeoni et al., NAR 2003, 31, 11: 2717-2724) and may be applied to the present invention. siRNA has recently been successfully used for inhibition of gene expression in primates (see for example. Tolentino et al., Retina 24(4):660 which may also be applied to the present invention.

[0179] Indeed, RNA delivery is a useful method of *in vivo* delivery. It is possible to deliver Cas9 and gRNA (and, for instance, HR repair template) into cells using liposomes or nanoparticles. Thus delivery of the CRISPR enzyme, such as a Cas9 and/or delivery of the RNAs of the invention may be in RNA form and via microvesicles, liposomes or nanoparticles. For example, Cas9 mRNA and gRNA can be packaged into liposomal particles for delivery *in vivo*. Liposomal transfection reagents such as lipofectamine from Life Technologies and other reagents on the market can effectively deliver RNA molecules into the liver.

[0180] Means of delivery of RNA also preferred include delivery of RNA via nanoparticles (Cho, S., Goldberg, M., Son, S., Xu, Q., Yang, F., Mei, Y., Bogatyrev, S., Langer, R. and Anderson, D., Lipid-like nanoparticles for small interfering RNA delivery to endothelial cells, Advanced Functional Materials, 19: 3112-3118, 2010) or exosomes (Schroeder, A., Levins, C., Cortez, C., Langer, R., and Anderson, D., Lipid-based nanotherapeutics for siRNA delivery, Journal of Internal Medicine, 267: 9-21, 2010, PMID: 20059641). Indeed, exosomes have been shown to be particularly useful in delivery siRNA, a system with some parallels to the CRISPR system. For instance, El-Andaloussi S, et al. ("Exosome-mediated delivery of siRNA in vitro and in vivo." Nat Protoc. 2012 Dec;7(12):2112-26. doi: 10.1038/nprot.2012.131. Epub 2012 Nov 15.) describe how exosomes are promising tools for drug delivery across different biological barriers and can be harnessed for delivery of siRNA in vitro and in vivo. Their approach is to generate targeted exosomes through transfection of an expression vector, comprising an exosomal protein fused with a peptide ligand. The exosomes are then purify and characterized

from transfected cell supernatant, then RNA is loaded into the exosomes. Delivery or administration according to the invention can be performed with exosomes, in particular but not limited to the brain. Vitamin E (α-tocopherol) may be conjugated with CRISPR Cas and delivered to the brain along with high density lipoprotein (HDL), for example in a similar manner as was done by Uno et al. (HUMAN GENE THERAPY 22:711–719 (June 2011)) for delivering short-interfering RNA (siRNA) to the brain. Mice were infused via Osmotic minipumps (model 1007D; Alzet, Cupertino, CA) filled with phosphate-buffered saline (PBS) or free TocsiBACE or Toc-siBACE/HDL and connected with Brain Infusion Kit 3 (Alzet). A brain-infusion cannula was placed about 0.5mm posterior to the bregma at midline for infusion into the dorsal third ventricle. Uno et al. found that as little as 3 nmol of Toc-siRNA with HDL could induce a target reduction in comparable degree by the same ICV infusion method. A similar dosage of CRISPR Cas conjugated to α-tocopherol and co-administered with HDL targeted to the brain may be contemplated for humans in the present invention, for example, about 3 nmol to about 3 μmol of CRISPR Cas targeted to the brain may be contemplated. Zou et al. ((HUMAN GENE THERAPY 22:465-475 (April 2011)) describes a method of lentiviral-mediated delivery of short-hairpin RNAs targeting PKCγ for in vivo gene silencing in the spinal cord of rats. Zou et al. administered about 10 μl of a recombinant lentivirus having a titer of 1×10^9 transducing units (TU)/ml by an intrathecal catheter. A similar dosage of CRISPR Cas expressed in a lentiviral vector targeted to the brain may be contemplated for humans in the present invention, for example, about 10-50 ml of CRISPR Cas targeted to the brain in a lentivirus having a titer of 1×10^9 transducing units (TU)/ml may be contemplated.

[0181] In terms of local delivery to the brain, this can be achieved in various ways. For instance, material can be delivered intrastriatally *e.g.* by injection. Injection can be performed stereotactically via a craniotomy.

[0182] Enhancing NHEJ or HR efficiency is also helpful for delivery. It is preferred that NHEJ efficiency is enhanced by co-expressing end-processing enzymes such as Trex2 (Dumitrache et al. Genetics. 2011 August; 188(4): 787–797). It is preferred that HR efficiency is increased by transiently inhibiting NHEJ machineries such as Ku70 and Ku86. HR efficiency can also be increased by co-expressing prokaryotic or eukaryotic homologous recombination enzymes such as RecBCD, RecA.

Packaging and Promoters generally

[0183] Ways to package Cas9 coding nucleic acid molecules, e.g., DNA, into vectors, e.g., viral vectors, to mediate genome modification *in vivo* include:

To achieve NHEJ-mediated gene knockout:

Single virus vector:

Vector containing two or more expression cassettes:

Promoter-Cas9 coding nucleic acid molecule -terminator

Promoter-gRNA1-terminator

Promoter-gRNA2-terminator

Promoter-gRNA(N)-terminator (up to size limit of vector)

Double virus vector:

Vector 1 containing one expression cassette for driving the expression of Cas9

Promoter-Cas9 coding nucleic acid molecule-terminator

Vector 2 containing one more expression cassettes for driving the expression of one or more guideRNAs

Promoter-gRNA1-terminator

Promoter-gRNA(N)-terminator (up to size limit of vector)

To mediate homology-directed repair.

In addition to the single and double virus vector approaches described above, an additional vector is used to deliver a homology-direct repair template.

[0184] The promoter used to drive Cas9 coding nucleic acid molecule expression can include:

AAV ITR can serve as a promoter: this is advantageous for eliminating the need for an additional promoter element (which can take up space in the vector). The additional space freed up can be used to drive the expression of additional elements (gRNA, etc.). Also, ITR activity is relatively weaker, so can be used to reduce potential toxicity due to over expression of Cas9.

For ubiquitous expression, can use promoters: CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains, etc.

For brain or other CNS expression, can use promoters: SynapsinI for all neurons, CaMKIIalpha for excitatory neurons, GAD67 or GAD65 or VGAT for GABAergic neurons, etc.

For liver expression, can use Albumin promoter.

For lung expression, can use SP-B.

For endothelial cells, can use ICAM.

For hematopoietic cells can use IFNbeta or CD45.

For Osteoblasts can use OG-2.

[0185] The promoter used to drive guide RNA can include:

Pol III promoters such as U6 or H1

Use of Pol II promoter and intronic cassettes to express gRNA

Adeno associated virus (AAV)

[0186] Cas9 and one or more guide RNA can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other plasmid or viral vector types, in particular, using formulations and doses from, for example, US Patents Nos. 8,454,972 (formulations, doses for adenovirus), 8,404,658 (formulations, doses for AAV) and 5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For examples, for AAV, the route of administration, formulation and dose can be as in US Patent No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in US Patent No. 8,404,658 and as in clinical trials involving adenovirus. For plasmid delivery, the route of administration, formulation and dose can be as in US Patent No 5,846,946 and as in clinical studies involving plasmids. Doses may be based on or extrapolated to an average 70 kg individual (*e.g.* a male adult human), and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed. The viral vectors can be injected into the tissue of interest. For cell-type specific genome modification, the expression of Cas9 can be driven by a cell-type specific promoter. For example, liver-specific expression might use the Albumin promoter and neuron-specific expression (*e.g.* for targeting CNS disorders) might use the Synapsin I promoter.

[0187] In terms of *in vivo* delivery, AAV is advantageous over other viral vectors for a couple of reasons:

Low toxicity (this may be due to the purification method not requiring ultra centrifugation of cell particles that can activate the immune response)

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

[0188] AAV has a packaging limit of 4.5 or 4.75 Kb. This means that Cas9 as well as a promoter and transcription terminator have to be all fit into the same viral vector. Constructs larger than 4.5 or 4.75 Kb will lead to significantly reduced virus production. SpCas9 is quite

large, the gene itself is over 4.1 Kb, which makes it difficult for packing into AAV. Therefore embodiments of the invention include utilizing homologs of Cas9 that are shorter. For example:

Species	Cas9 Size
Corynebacter diphtheriae	3252
Eubacterium ventriosum	3321
Streptococcus pasteurianus	3390
Lactobacillus farciminis	3378
Sphaerochaeta globus	3537
Azospirillum B510	3504
Gluconacetobacter diazotrophicus	3150
Neisseria cinerea	3246
Roseburia intestinalis	3420
Parvibaculum lavamentivorans	3111
Staphylococcus aureus	3159
Nitratifractor salsuginis DSM 16511	3396
Campylobacter lari CF89-12	3009
Streptococcus thermophilus LMD-9	3396

[0189] These species are therefore, in general, preferred Cas9 species.

[0190] As to AAV, the AAV can be AAV1, AAV2, AAV5 or any combination thereof. One can select the AAV of the AAV with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof for targeting brain or neuronal cells; and one can select AAV4 for targeting cardiac tissue. AAV8 is useful for delivery to the liver. The herein promoters and vectors are preferred individually. A tabulation of certain AAV serotypes as to these cells (see Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)) is as follows:

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

Lentivirus

[0191] Lentiviruses are complex retroviruses that have the ability to infect and express their genes in both mitotic and post-mitotic cells. The most commonly known lentivirus is the human immunodeficiency virus (HIV), which uses the envelope glycoproteins of other viruses to target a broad range of cell types.

[0192] Lentiviruses may be prepared as follows. After cloning pCasES10 (which contains a lentiviral transfer plasmid backbone), HEK293FT at low passage (p=5) were seeded in a T-75 flask to 50% confluence the day before transfection in DMEM with 10% fetal bovine serum and without antibiotics. After 20 hours, media was changed to OptiMEM (serum-free) media and transfection was done 4 hours later. Cells were transfected with 10 μg of lentiviral transfer plasmid (pCasES10) and the following packaging plasmids: 5 μg of pMD2.G (VSV-g pseudotype), and 7.5ug of psPAX2 (gag/pol/rev/tat). Transfection was done in 4mL OptiMEM with a cationic lipid delivery agent (50uL Lipofectamine 2000 and 100ul Plus reagent). After 6 hours, the media was changed to antibiotic-free DMEM with 10% fetal bovine serum. These methods use serum during cell culture, but serum-free methods are preferred.

[0193] Lentivirus may be purified as follows. Viral supernatants were harvested after 48 hours. Supernatants were first cleared of debris and filtered through a 0.45um low protein binding (PVDF) filter. They were then spun in a ultracentrifuge for 2 hours at 24,000 rpm. Viral pellets were resuspended in 50ul of DMEM overnight at 4C. They were then aliquotted and immediately frozen at -80°C.

[0194] In another embodiment, minimal non-primate lentiviral vectors based on the equine infectious anemia virus (EIAV) are also contemplated, especially for ocular gene therapy (see, e.g., Balagaan, J Gene Med 2006; 8: 275 – 285). In another embodiment, RetinoStat®, an equine infectious anemia virus-based lentiviral gene therapy vector that expresses angiostatic proteins endostatin and angiostatin that is delivered via a subretinal injection for the treatment of the web form of age-related macular degeneration is also contemplated (see, e.g., Binley et al., HUMAN GENE THERAPY 23:980–991 (September 2012)) and this vector may be modified for the CRISPR-Cas system of the present invention.

[0195] In another embodiment, self-inactivating lentiviral vectors with an siRNA targeting a common exon shared by HIV tat/rev, a nucleolar-localizing TAR decoy, and an anti–CCR5-specific hammerhead ribozyme (see, e.g., DiGiusto et al. (2010) Sci Transl Med 2:36ra43) may be used/and or adapted to the CRISPR-Cas system of the present invention. A minimum of 2.5 ×

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

[0196] Lentiviral vectors have been disclosed as in the treatment for Parkinson's Disease, see, e.g., US Patent Publication No. 20120295960 and US Patent Nos. 7303910 and 7351585. Lentiviral vectors have also been disclosed for the treatment of ocular diseases, see e.g., US Patent Publication Nos. 20060281180, 20090007284, US20110117189; US20090017543; US20070054961, US20100317109. Lentiviral vectors have also been disclosed for delivery to the brain, see, e.g., US Patent Publication Nos. US20110293571; US20110293571, US20040013648, US20070025970, US20090111106 and US Patent No. US7259015.

RNA delivery

[0197] RNA delivery: The CRISPR enzyme, for instance a Cas9, and/or any of the present RNAs, for instance a guide RNA, can also be delivered in the form of RNA. Cas9 mRNA can be generated using *in vitro* transcription. For example, Cas9 mRNA can be synthesized using a PCR cassette containing the following elements: T7_promoter-kozak sequence (GCCACC)-Cas9-3' UTR from beta globin-polyA tail (a string of 120 or more adenines). The cassette can be used for transcription by T7 polymerase. Guide RNAs can also be transcribed using in vitro transcription from a cassette containing T7_promoter-GG-guide RNA sequence.

[0198] To enhance expression and reduce possible toxicity, the CRISPR enzyme-coding sequence and/or the guide RNA can be modified to include one or more modified nucleoside *e.g.* using pseudo-U or 5-Methyl-C.

[0199] mRNA delivery methods are especially promising for liver delivery currently.

[0200] Much clinical work on RNA delivery has focused on RNAi or antisense, but these systems can be adapted for delivery of RNA for implementing the present invention. References below to RNAi *etc.* should be read accordingly.

Nanoparticles

[0201] CRISPR enzyme mRNA and guide RNA may be delivered simultaneously using nanoparticles or lipid envelopes.

[0202] For example, Su X, Fricke J, Kavanagh DG, Irvine DJ ("In vitro and in vivo mRNA delivery using lipid-enveloped pH-responsive polymer nanoparticles" Mol Pharm. 2011 Jun

6;8(3):774-87. doi: 10.1021/mp100390w. Epub 2011 Apr 1) describes biodegradable core-shell structured nanoparticles with a poly(β-amino ester) (PBAE) core enveloped by a phospholipid bilayer shell. These were developed for in vivo mRNA delivery. The pH-responsive PBAE component was chosen to promote endosome disruption, while the lipid surface layer was selected to minimize toxicity of the polycation core. Such are, therefore, preferred for delivering RNA of the present invention.

[0203] In one embodiment, nanoparticles based on self assembling bioadhesive polymers are contemplated, which may be applied to oral delivery of peptides, intravenous delivery of peptides and nasal delivery of peptides, all to the brain. Other embodiments, such as oral absorption and ocular delivery of hydrophobic drugs are also contemplated. The molecular envelope technology involves an engineered polymer envelope which is protected and delivered to the site of the disease (see, e.g., Mazza, M. et al. ACSNano, 2013. 7(2): 1016-1026; Siew, A., et al. Mol Pharm, 2012. 9(1):14-28; Lalatsa, A., et al. J Contr Rel, 2012. 161(2):523-36; Lalatsa, A., et al., Mol Pharm, 2012. 9(6):1665-80; Lalatsa, A., et al. Mol Pharm, 2012. 9(6):1764-74; Garrett, N.L., et al. J Biophotonics, 2012. 5(5-6):458-68; Garrett, N.L., et al. J Raman Spect, 2012. 43(5):681-688; Ahmad, S., et al. J Royal Soc Interface 2010. 7:S423-33; Uchegbu, I.F. Expert Opin Drug Deliv, 2006. 3(5):629-40; Qu, X.,et al. Biomacromolecules, 2006. 7(12):3452-9 and Uchegbu, I.F., et al. Int J Pharm, 2001. 224:185-199). Doses of about 5 mg/kg are contemplated, with single or multiple doses, depending on the target tissue.

[0204] In one embodiment, nanoparticles that can deliver RNA to a cancer cell to stop tumor growth developed by Dan Anderson's lab at MIT may be used/and or adapted to the CRISPR Cas system of the present invention. In particular, the Anderson lab developed fully automated, combinatorial systems for the synthesis, purification, characterization, and formulation of new biomaterials and nanoformulations. See, e.g., Alabi et al., Proc Natl Acad Sci U S A. 2013 Aug 6;110(32):12881-6; Zhang et al., Adv Mater. 2013 Sep 6;25(33):4641-5; Jiang et al., Nano Lett. 2013 Mar 13;13(3):1059-64; Karagiannis et al., ACS Nano. 2012 Oct 23;6(10):8484-7; Whitehead et al., ACS Nano. 2012 Aug 28;6(8):6922-9 and Lee et al., Nat Nanotechnol. 2012 Jun 3;7(6):389-93.

[0205] US patent application 20110293703 relates to lipidoid compounds are also particularly useful in the administration of polynucleotides, which may be applied to deliver the CRISPR Cas system of the present invention. In one aspect, the aminoalcohol lipidoid compounds are combined with an agent to be delivered to a cell or a subject to form microparticles, nanoparticles, liposomes, or micelles. The agent to be delivered by the particles,

liposomes, or micelles may be in the form of a gas, liquid, or solid, and the agent may be a polynucleotide, protein, peptide, or small molecule. The minoalcohol lipidoid compounds may be combined with other aminoalcohol lipidoid compounds, polymers (synthetic or natural), surfactants, cholesterol, carbohydrates, proteins, lipids, etc. to form the particles. These particles may then optionally be combined with a pharmaceutical excipient to form a pharmaceutical composition.

[0206] US Patent Publication No. 20110293703 also provides methods of preparing the aminoalcohol lipidoid compounds. One or more equivalents of an amine are allowed to react with one or more equivalents of an epoxide-terminated compound under suitable conditions to form an aminoalcohol lipidoid compound of the present invention. In certain embodiments, all the amino groups of the amine are fully reacted with the epoxide-terminated compound to form tertiary amines. In other embodiments, all the amino groups of the amine are not fully reacted with the epoxide-terminated compound to form tertiary amines thereby resulting in primary or secondary amines in the aminoalcohol lipidoid compound. These primary or secondary amines are left as is or may be reacted with another electrophile such as a different epoxide-terminated compound. As will be appreciated by one skilled in the art, reacting an amine with less than excess of epoxide-terminated compound will result in a plurality of different aminoalcohol lipidoid compounds with various numbers of tails. Certain amines may be fully functionalized with two epoxide-derived compound tails while other molecules will not be completely functionalized with epoxide-derived compound tails. For example, a diamine or polyamine may include one, two, three, or four epoxide-derived compound tails off the various amino moieties of the molecule resulting in primary, secondary, and tertiary amines. In certain embodiments, all the amino groups are not fully functionalized. In certain embodiments, two of the same types of epoxide-terminated compounds are used. In other embodiments, two or more different epoxide-terminated compounds are used. The synthesis of the aminoalcohol lipidoid compounds is performed with or without solvent, and the synthesis may be performed at higher temperatures ranging from 30-100 °C., preferably at approximately 50-90 °C. The prepared aminoalcohol lipidoid compounds may be optionally purified. For example, the mixture of aminoalcohol lipidoid compounds may be purified to yield an aminoalcohol lipidoid compound with a particular number of epoxide-derived compound tails. Or the mixture may be purified to yield a particular stereo- or regioisomer. The aminoalcohol lipidoid compounds may also be alkylated using an alkyl halide (e.g., methyl iodide) or other alkylating agent, and/or they may be acylated.

[0207] US Patent Publication No. 20110293703 also provides libraries of aminoalcohol lipidoid compounds prepared by the inventive methods. These aminoalcohol lipidoid compounds may be prepared and/or screened using high-throughput techniques involving liquid handlers, robots, microtiter plates, computers, etc. In certain embodiments, the aminoalcohol lipidoid compounds are screened for their ability to transfect polynucleotides or other agents (e.g., proteins, peptides, small molecules) into the cell.

[0208] US Patent Publication No. 20130302401 relates to a class of poly(beta-amino alcohols) (PBAAs) has been prepared using combinatorial polymerization. The inventive PBAAs may be used in biotechnology and biomedical applications as coatings (such as coatings of films or multilayer films for medical devices or implants), additives, materials, excipients, non-biofouling agents, micropatterning agents, and cellular encapsulation agents. When used as surface coatings, these PBAAs elicited different levels of inflammation, both in vitro and in vivo, depending on their chemical structures. The large chemical diversity of this class of materials allowed us to identify polymer coatings that inhibit macrophage activation in vitro. Furthermore, these coatings reduce the recruitment of inflammatory cells, and reduce fibrosis, following the subcutaneous implantation of carboxylated polystyrene microparticles. These polymers may be used to form polyelectrolyte complex capsules for cell encapsulation. The invention may also have many other biological applications such as antimicrobial coatings, DNA or siRNA delivery, and stem cell tissue engineering. The teachings of US Patent Publication No. 20130302401 may be applied to the CRISPR Cas system of the present invention.

[0209] In another embodiment, lipid nanoparticles (LNPs) are contemplated. An antitransthyretin small interfering RNA has been encapsulated in lipid nanoparticles and delivered to humans (see, e.g., Coelho et al., N Engl J Med 2013;369:819-29), and such a ssystem may be adapted and applied to the CRISPR Cas system of the present invention. Doses of about 0.01 to about 1 mg per kg of body weight administered intravenously are contemplated. Medications to reduce the risk of infusion-related reactions are contemplated, such as dexamethasone, acetampinophen, diphenhydramine or cetirizine, and ranitidine are contemplated. Multiple doses of about 0.3 mg per kilogram every 4 weeks for five doses are also contemplated.

[0210] LNPs have been shown to be highly effective in delivering siRNAs to the liver (see, e.g., Tabernero et al., Cancer Discovery, April 2013, Vol. 3, No. 4, pages 363-470) and are therefore contemplated for delivering RNA encoding CRISPR Cas to the liver. A dosage of about four doses of 6 mg/kg of the LNP every two weeks may be contemplated. Tabernero et al.

demonstrated that tumor regression was observed after the first 2 cycles of LNPs dosed at 0.7 mg/kg, and by the end of 6 cycles the patient had achieved a partial response with complete regression of the lymph node metastasis and substantial shrinkage of the liver tumors. A complete response was obtained after 40 doses in this patient, who has remained in remission and completed treatment after receiving doses over 26 months. Two patients with RCC and extrahepatic sites of disease including kidney, lung, and lymph nodes that were progressing following prior therapy with VEGF pathway inhibitors had stable disease at all sites for approximately 8 to 12 months, and a patient with PNET and liver metastases continued on the extension study for 18 months (36 doses) with stable disease.

[0211] However, the charge of the LNP must be taken into consideration. As cationic lipids combined with negatively charged lipids to induce nonbilayer structures that facilitate intracellular delivery. Because charged LNPs are rapidly cleared from circulation following intravenous injection, ionizable cationic lipids with pKa values below 7 were developed (see, e.g., Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). Negatively charged polymers such as RNA may be loaded into LNPs at low pH values (e.g., pH 4) where the ionizable lipids display a positive charge. However, at physiological pH values, the LNPs exhibit a low surface charge compatible with longer circulation times. Four species of ionizable cationic lipids have been focused upon, namely 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxy-keto-N,N-dimethyl-3-aminopropane (DLinKDMA), and 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA). It has been shown that LNP siRNA systems containing these lipids exhibit remarkably different gene silencing properties in hepatocytes in vivo, with potencies varying according to the series DLinKC2-DMA>DLinKDMA>DLinDMA>>DLinDAP employing a Factor VII gene silencing model (see, e.g., Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). A dosage of 1 μg/ml of LNP or CRISPR-Cas RNA in or associated with the LNP may be contemplated, especially for a formulation containing DLinKC2-DMA.

[0212] Preparation of LNPs and CRISPR Cas encapsulation may be used/and or adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011). The cationic lipids 1,2-dilineoyl-3-dimethylammonium-propane (DLinDAP), 1,2-dilinoleyloxy-3-N,N-dimethylaminopropane (DLinDMA), 1,2-dilinoleyloxyketo-N,N-dimethyl-3-aminopropane (DLinK-DMA), 1,2-dilinoleyl-4-(2-dimethylaminoethyl)-[1,3]-dioxolane (DLinKC2-DMA), (3-o-[2″-(methoxypolyethyleneglycol 2000) succinoyl]-1,2-dimyristoyl-sn-glycol (PEG-S-DMG),

and R-3-[(ω-methoxy-poly(ethylene glycol)2000) carbamoyl]-1,2-dimyristyloxlpropyl-3-amine (PEG-C-DOMG) may be provided by Tekmira Pharmaceuticals (Vancouver, Canada) or synthesized. Cholesterol may be purchased from Sigma (St Louis, MO). The specific CRISPR Cas RNA may be encapsulated in LNPs containing DLinDAP, DLinDMA, DLinK-DMA, and DLinKC2-DMA (cationic lipid:DSPC:CHOL: PEGS-DMG or PEG-C-DOMG at 40:10:40:10 molar ratios). When required, 0.2% SP-DiOC18 (Invitrogen, Burlington, Canada) may be incorporated to assess cellular uptake, intracellular delivery, and biodistribution. Encapsulation may be performed by dissolving lipid mixtures comprised of cationic lipid:DSPC:cholesterol:PEG-c-DOMG (40:10:40:10 molar ratio) in ethanol to a final lipid concentration of 10 mmol/l. This ethanol solution of lipid may be added drop-wise to 50 mmol/l citrate, pH 4.0 to form multilamellar vesicles to produce a final concentration of 30% ethanol vol/vol. Large unilamellar vesicles may be formed following extrusion of multilamellar vesicles through two stacked 80 nm Nuclepore polycarbonate filters using the Extruder (Northern Lipids, Vancouver, Canada). Encapsulation may be achieved by adding RNA dissolved at 2 mg/ml in 50 mmol/l citrate, pH 4.0 containing 30% ethanol vol/vol drop-wise to extruded preformed large unilamellar vesicles and incubation at 31 °C for 30 minutes with constant mixing to a final RNA/lipid weight ratio of 0.06/1 wt/wt. Removal of ethanol and neutralization of formulation buffer were performed by dialysis against phosphate-buffered saline (PBS), pH 7.4 for 16 hours using Spectra/Por 2 regenerated cellulose dialysis membranes. Nanoparticle size distribution may be determined by dynamic light scattering using a NICOMP 370 particle sizer, the vesicle/intensity modes, and Gaussian fitting (Nicomp Particle Sizing, Santa Barbara, CA). The particle size for all three LNP systems may be ~70 nm in diameter. RNA encapsulation efficiency may be determined by removal of free RNA using VivaPureD MiniH columns (Sartorius Stedim Biotech) from samples collected before and after dialysis. The encapsulated RNA may be extracted from the eluted nanoparticles and quantified at 260 nm. RNA to lipid ratio was determined by measurement of cholesterol content in vesicles using the Cholesterol E enzymatic assay from Wako Chemicals USA (Richmond, VA). In conjunction with the herein discussion of LNPs and PEG lipids, PEGylated liposomes or LNPs are likewise suitable for delivery of a CRISPR-Cas system or components thereof.

[0213] Preparation of large LNPs may be used/and or adapted from Rosin et al, Molecular Therapy, vol. 19, no. 12, pages 1286-2200, Dec. 2011. A lipid premix solution (20.4 mg/ml total lipid concentration) may be prepared in ethanol containing DLinKC2-DMA, DSPC, and cholesterol at 50:10:38.5 molar ratios. Sodium acetate may be added to the lipid premix at a

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

[0214] Spherical Nucleic Acid (SNA™) constructs and other nanoparticles (particularly gold nanoparticles) are also contemplated as a means to delivery CRISPR-Cas system to intended targets. Significant data show that AuraSense Therapeutics' Spherical Nucleic Acid (SNA™) constructs, based upon nucleic acid-functionalized gold nanoparticles, are useful.

[0215] Literature that may be employed in conjunction with herein teachings include: Cutler et al., J. Am. Chem. Soc. 2011 133:9254-9257, Hao et al., Small. 2011 7:3158-3162, Zhang et al., ACS Nano. 2011 5:6962-6970, Cutler et al., J. Am. Chem. Soc. 2012 134:1376-1391, Young et al., Nano Lett. 2012 12:3867-71, Zheng et al., Proc. Natl. Acad. Sci. USA. 2012 109:11975-80, Mirkin, Nanomedicine 2012 7:635-638 Zhang et al., J. Am. Chem. Soc. 2012 134:16488-1691, Weintraub, Nature 2013 495:S14-S16, Choi et al., Proc. Natl. Acad. Sci. USA. 2013 110(19):7625-7630, Jensen et al., Sci. Transl. Med. 5, 209ra152 (2013) and Mirkin, et al., Small, 10:186-192.

[0216] Self-assembling nanoparticles with RNA may be constructed with polyethyleneimine (PEI) that is PEGylated with an Arg-Gly-Asp (RGD) peptide ligand attached at the distal end of the polyethylene glycol (PEG). This system has been used, for example, as a means to target tumor neovasculature expressing integrins and deliver siRNA inhibiting vascular endothelial growth factor receptor-2 (VEGF R2) expression and thereby achieve tumor angiogenesis (see, e.g., Schiffelers et al., Nucleic Acids Research, 2004, Vol. 32, No. 19). Nanoplexes may be prepared by mixing equal volumes of aqueous solutions of cationic polymer and nucleic acid to give a net molar excess of ionizable nitrogen (polymer) to phosphate (nucleic acid) over the

range of 2 to 6. The electrostatic interactions between cationic polymers and nucleic acid resulted in the formation of polyplexes with average particle size distribution of about 100 nm, hence referred to here as nanoplexes. A dosage of about 100 to 200 mg of CRISPR Cas is envisioned for delivery in the self-assembling nanoparticles of Schiffelers et al.

[0217] The nanoplexes of Bartlett et al. (PNAS, September 25, 2007,vol. 104, no. 39) may also be applied to the present invention. The nanoplexes of Bartlett et al. are prepared by mixing equal volumes of aqueous solutions of cationic polymer and nucleic acid to give a net molar excess of ionizable nitrogen (polymer) to phosphate (nucleic acid) over the range of 2 to 6. The electrostatic interactions between cationic polymers and nucleic acid resulted in the formation of polyplexes with average particle size distribution of about 100 nm, hence referred to here as nanoplexes. The DOTA-siRNA of Bartlett et al. was synthesized as follows: 1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid mono(N-hydroxysuccinimide ester) (DOTA-NHSester) was ordered from Macrocyclics (Dallas, TX). The amine modified RNA sense strand with a 100-fold molar excess of DOTA-NHS-ester in carbonate buffer (pH 9) was added to a microcentrifuge tube. The contents were reacted by stirring for 4 h at room temperature. The DOTA-RNAsense conjugate was ethanol-precipitated, resuspended in water, and annealed to the unmodified antisense strand to yield DOTA-siRNA. All liquids were pretreated with Chelex-100 (Bio-Rad, Hercules, CA) to remove trace metal contaminants. Tf-targeted and nontargeted siRNA nanoparticles may be formed by using cyclodextrin-containing polycations. Typically, nanoparticles were formed in water at a charge ratio of 3 (+/-) and an siRNA concentration of 0.5 g/liter. One percent of the adamantane-PEG molecules on the surface of the targeted nanoparticles were modified with Tf (adamantane-PEG-Tf). The nanoparticles were suspended in a 5% (wt/vol) glucose carrier solution for injection.

[0218] Davis et al. (Nature, Vol 464, 15 April 2010) conducts a RNA clinical trial that uses a targeted nanoparticle-delivery system (clinical trial registration number NCT00689065). Patients with solid cancers refractory to standard-of-care therapies are administered doses of targeted nanoparticles on days 1, 3, 8 and 10 of a 21-day cycle by a 30-min intravenous infusion. The nanoparticles consist of a synthetic delivery system containing: (1) a linear, cyclodextrin-based polymer (CDP), (2) a human transferrin protein (TF) targeting ligand displayed on the exterior of the nanoparticle to engage TF receptors (TFR) on the surface of the cancer cells, (3) a hydrophilic polymer (polyethylene glycol (PEG) used to promote nanoparticle stability in biological fluids), and (4) siRNA designed to reduce the expression of the RRM2 (sequence used in the clinic was previously denoted siR2B+5). The TFR has long been known to be upregulated

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

Particle delivery systems and/or formulations:

[0219] Several types of particle delivery systems and/or formulations are known to be useful in a diverse spectrum of biomedical applications. In general, a particle is defined as a small object that behaves as a whole unit with respect to its transport and properties. Particles are further classified according to diameter Coarse particles cover a range between 2,500 and 10,000 nanometers. Fine particles are sized between 100 and 2,500 nanometers. Ultrafine particles, or nanoparticles, are generally between 1 and 100 nanometers in size. The basis of the 100-nm limit is the fact that novel properties that differentiate particles from the bulk material typically develop at a critical length scale of under 100 nm.

[0220] As used herein, a particle delivery system/formulation is defined as any biological delivery system/formulation which includes a particle in accordance with the present invention. A particle in accordance with the present invention is any entity having a greatest dimension (e.g. diameter) of less than 100 microns (μm). In some embodiments, inventive particles have a greatest dimension of less than 10 μm. In some embodiments, inventive particles have a greatest dimension of less than 2000 nanometers (nm). In some embodiments, inventive particles have a greatest dimension of less than 1000 nanometers (nm). In some embodiments, inventive particles have a greatest dimension of less than 900 nm, 800 nm, 700 nm, 600 nm, 500 nm, 400 nm, 300 nm, 200 nm, or 100 nm. Typically, inventive particles have a greatest dimension (e.g., diameter)

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

[0221] Particle characterization (including e.g., characterizing morphology, dimension, etc.) is done using a variety of different techniques. Common techniques are electron microscopy (TEM, SEM), atomic force microscopy (AFM), dynamic light scattering (DLS), X-ray photoelectron spectroscopy (XPS), powder X-ray diffraction (XRD), Fourier transform infrared spectroscopy (FTIR), matrix-assisted laser desorption/ionization time-of-flight mass spectrometry(MALDI-TOF), ultraviolet-visible spectroscopy, dual polarisation interferometry and nuclear magnetic resonance (NMR). Characterization (dimension measurements) may be made as to native particles (i.e., preloading) or after loading of the cargo (herein cargo refers to e.g., one or more components of CRISPR-Cas system e.g., CRISPR enzyme or mRNA or guide RNA, or any combination thereof, and may include additional carriers and/or excipients) to provide particles of an optimal size for delivery for any *in vitro*, *ex vivo* and/or *in vivo* application of the present invention. In certain preferred embodiments, particle dimension (e.g., diameter) characterization is based on measurements using dynamic laser scattering (DLS). Mention is made of US Patent No. 8,709,843; US Patent No. 6,007,845; US Patent No. 5,855,913; US Patent No. 5,985,309; US. Patent No. 5,543,158; and the publication by James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84, concerning particles, methods of making and using them and measurements thereof.

[0222] Particles delivery systems within the scope of the present invention may be provided in any form, including but not limited to solid, semi-solid, emulsion, or colloidal particles. As such any of the delivery systems described herein, including but not limited to, e.g., lipid-based systems, liposomes, micelles, microvesicles, exosomes, or gene gun may be provided as particle delivery systems within the scope of the present invention.

Nanoparticles

[0223] In terms of this invention, it is preferred to have one or more components of CRISPR complex, e.g., CRISPR enzyme or mRNA or guide RNA delivered using nanoparticles or lipid

envelopes. Other delivery systems or vectors are may be used in conjunction with the nanoparticle aspects of the invention.

[0224] In general, a "nanoparticle" refers to any particle having a diameter of less than 1000 nm. In certain preferred embodiments, nanoparticles of the invention have a greatest dimension (e.g., diameter) of 500 nm or less. In other preferred embodiments, nanoparticles of the invention have a greatest dimension ranging between 25 nm and 200 nm. In other preferred embodiments, nanoparticles of the invention have a greatest dimension of 100 nm or less. In other preferred embodiments, nanoparticles of the invention have a greatest dimension ranging between 35 nm and 60 nm.

[0225] Nanoparticles encompassed in the present invention may be provided in different forms, e.g., as solid nanoparticles (e.g., metal such as silver, gold, iron, titanium), non-metal, lipid-based solids, polymers), suspensions of nanoparticles, or combinations thereof. Metal, dielectric, and semiconductor nanoparticles may be prepared, as well as hybrid structures (e.g., core–shell nanoparticles). Nanoparticles made of semiconducting material may also be labeled quantum dots if they are small enough (typically sub 10 nm) that quantization of electronic energy levels occurs. Such nanoscale particles are used in biomedical applications as drug carriers or imaging agents and may be adapted for similar purposes in the present invention.

[0226] Semi-solid and soft nanoparticles have been manufactured, and are within the scope of the present invention. A prototype nanoparticle of semi-solid nature is the liposome. Various types of liposome nanoparticles are currently used clinically as delivery systems for anticancer drugs and vaccines. Nanoparticles with one half hydrophilic and the other half hydrophobic are termed Janus particles and are particularly effective for stabilizing emulsions. They can self-assemble at water/oil interfaces and act as solid surfactants.

[0227] US Patent No. 8,709,843, incorporated herein by reference, provides a drug delivery system for targeted delivery of therapeutic agent-containing particles to tissues, cells, and intracellular compartments. The invention provides targeted particles comprising comprising polymer conjugated to a surfactant, hydrophilic polymer or lipid.

[0228] US Patent No. 6,007,845, incorporated herein by reference, provides particles which have a core of a multiblock copolymer formed by covalently linking a multifunctional compound with one or more hydrophobic polymers and one or more hydrophilic polymers, and conatin a biologically active material.

[0229] US Patent No. 5,855,913, incorporated herein by reference, provides a particulate composition having aerodynamically light particles having a tap density of less than 0.4 g/cm3

with a mean diameter of between 5 μm and 30 μm, incorporating a surfactant on the surface thereof for drug delivery to the pulmonary system.

[0230] US Patent No. 5,985,309, incorporated herein by reference, provides particles incorporating a surfactant and/or a hydrophilic or hydrophobic complex of a positively or negatively charged therapeutic or diagnostic agent and a charged molecule of opposite charge for delivery to the pulmonary system.

[0231] US. Patent No. 5,543,158, incorporated herein by reference, provides biodegradable injectable nanoparticles having a biodegradable solid core containing a biologically active material and poly(alkylene glycol) moieties on the surface.

[0232] WO2012135025 (also published as US20120251560), incorporated herein by reference, describes conjugated polyethyleneimine (PEI) polymers and conjugated aza-macrocycles (collectively referred to as "conjugated lipomer" or "lipomers"). In certain embodiments, it can envisioned that such conjugated lipomers can be used in the context of the CRISPR-Cas system to achieve in vitro, ex vivo and in vivo genomic perturbations to modify gene expression, including modulation of protein expression.

[0233] In one embodiment, the nanoparticle may be epoxide-modified lipid–polymer, advantageously 7C1 (see, e.g., James E. Dahlman and Carmen Barnes et al. Nature Nanotechnology (2014) published online 11 May 2014, doi:10.1038/nnano.2014.84). C71 was synthesized by reacting C15 epoxide-terminated lipids with PEI600 at a 14:1 molar ratio, and was formulated with C14PEG2000 to produce nanoparticles (diameter between 35 and 60 nm) that were stable in PBS solution for at least 40 days.

[0234] An epoxide-modified lipid-polymer may be utilized to deliver the CRISPR-Cas system of the present invention to pulmonary, cardiovascular or renal cells, however, one of skill in the art may adapt the system to deliver to other target organs. Dosage ranging from about 0.05 to about 0.6 mg/kg are envisioned. Dosages over several days or weeks are also envisioned, with a total dosage of about 2 mg/kg.

Exosomes

[0235] Exosomes are endogenous nano-vesicles that transport RNAs and proteins, and which can deliver RNA to the brain and other target organs. To reduce immunogenicity, Alvarez-Erviti et al. (2011, Nat Biotechnol 29: 341) used self-derived dendritic cells for exosome production. Targeting to the brain was achieved by engineering the dendritic cells to express Lamp2b, an exosomal membrane protein, fused to the neuron-specific RVG peptide. Purified exosomes were loaded with exogenous RNA by electroporation. Intravenously injected RVG-targeted exosomes

delivered GAPDH siRNA specifically to neurons, microglia, oligodendrocytes in the brain, resulting in a specific gene knockdown. Pre-exposure to RVG exosomes did not attenuate knockdown, and non-specific uptake in other tissues was not observed. The therapeutic potential of exosome-mediated siRNA delivery was demonstrated by the strong mRNA (60%) and protein (62%) knockdown of BACE1, a therapeutic target in Alzheimer's disease.

[0236] To obtain a pool of immunologically inert exosomes, Alvarez-Erviti et al. harvested bone marrow from inbred C57BL/6 mice with a homogenous major histocompatibility complex (MHC) haplotype. As immature dendritic cells produce large quantities of exosomes devoid of T-cell activators such as MHC-II and CD86, Alvarez-Erviti et al. selected for dendritic cells with granulocyte/macrophage-colony stimulating factor (GM-CSF) for 7 d. Exosomes were purified from the culture supernatant the following day using well-established ultracentrifugation protocols. The exosomes produced were physically homogenous, with a size distribution peaking at 80 nm in diameter as determined by nanoparticle tracking analysis (NTA) and electron microscopy. Alvarez-Erviti et al. obtained 6-12 μg of exosomes (measured based on protein concentration) per 10^6 cells.

[0237] Next, Alvarez-Erviti et al. investigated the possibility of loading modified exosomes with exogenous cargoes using electroporation protocols adapted for nanoscale applications. As electroporation for membrane particles at the nanometer scale is not well-characterized, nonspecific Cy5-labeled RNA was used for the empirical optimization of the electroporation protocol. The amount of encapsulated RNA was assayed after ultracentrifugation and lysis of exosomes. Electroporation at 400 V and 125 μF resulted in the greatest retention of RNA and was used for all subsequent experiments.

[0238] Alvarez-Erviti et al. administered 150 μg of each BACE1 siRNA encapsulated in 150 μg of RVG exosomes to normal C57BL/6 mice and compared the knockdown efficiency to four controls: untreated mice, mice injected with RVG exosomes only, mice injected with BACE1 siRNA complexed to an in vivo cationic liposome reagent and mice injected with BACE1 siRNA complexed to RVG-9R, the RVG peptide conjugated to 9 D-arginines that electrostatically binds to the siRNA. Cortical tissue samples were analyzed 3 d after administration and a significant protein knockdown (45%, $P < 0.05$, versus 62%, $P < 0.01$) in both siRNA-RVG-9R-treated and siRNARVG exosome-treated mice was observed, resulting from a significant decrease in BACE1 mRNA levels (66% [+ or -] 15%, $P < 0.001$ and 61% [+ or -] 13% respectively, $P < 0.01$). Moreover, Applicants demonstrated a significant decrease (55%, $P < 0.05$) in the total [beta]-amyloid 1-42 levels, a main component of the amyloid

plaques in Alzheimer's pathology, in the RVG-exosome-treated animals. The decrease observed was greater than the β-amyloid 1-40 decrease demonstrated in normal mice after intraventricular injection of BACE1 inhibitors. Alvarez-Erviti et al. carried out 5'-rapid amplification of cDNA ends (RACE) on BACE1 cleavage product, which provided evidence of RNAi-mediated knockdown by the siRNA.

[0239] Finally, Alvarez-Erviti et al. investigated whether RNA-RVG exosomes induced immune responses in vivo by assessing IL-6, IP-10, TNFα and IFN-α serum concentrations. Following exosome treatment, nonsignificant changes in all cytokines were registered similar to siRNA-transfection reagent treatment in contrast to siRNA-RVG-9R, which potently stimulated IL-6 secretion, confirming the immunologically inert profile of the exosome treatment. Given that exosomes encapsulate only 20% of siRNA, delivery with RVG-exosome appears to be more efficient than RVG-9R delivery as comparable mRNA knockdown and greater protein knockdown was achieved with fivefold less siRNA without the corresponding level of immune stimulation. This experiment demonstrated the therapeutic potential of RVG-exosome technology, which is potentially suited for long-term silencing of genes related to neurodegenerative diseases. The exosome delivery system of Alvarez-Erviti et al. may be applied to deliver the CRISPR-Cas system of the present invention to therapeutic targets, especially neurodegenerative diseases. A dosage of about 100 to 1000 mg of CRISPR Cas encapsulated in about 100 to 1000 mg of RVG exosomes may be contemplated for the present invention.

[0240] El-Andaloussi et al. (Nature Protocols 7,2112–2126(2012)) discloses how exosomes derived from cultured cells can be harnessed for delivery of RNA in vitro and in vivo. This protocol first describes the generation of targeted exosomes through transfection of an expression vector, comprising an exosomal protein fused with a peptide ligand. Next, El-Andaloussi et al. explain how to purify and characterize exosomes from transfected cell supernatant. Next, El-Andaloussi et al. detail crucial steps for loading RNA into exosomes. Finally, El-Andaloussi et al. outline how to use exosomes to efficiently deliver RNA in vitro and in vivo in mouse brain. Examples of anticipated results in which exosome-mediated RNA delivery is evaluated by functional assays and imaging are also provided. The entire protocol takes ~3 weeks. Delivery or administration according to the invention may be performed using exosomes produced from self-derived dendritic cells. From the herein teachings, this can be employed in the practice of the invention

[0241] In another embodiment, the plasma exosomes of Wahlgren et al. (Nucleic Acids Research, 2012, Vol. 40, No. 17 e130) are contemplated. Exosomes are nano-sized vesicles (30–90nm in size) produced by many cell types, including dendritic cells (DC), B cells, T cells, mast cells, epithelial cells and tumor cells. These vesicles are formed by inward budding of late endosomes and are then released to the extracellular environment upon fusion with the plasma membrane. Because exosomes naturally carry RNA between cells, this property may be useful in gene therapy, and from this disclosure can be employed in the practice of the instant invention.

[0242] Exosomes from plasma can be prepared by centrifugation of buffy coat at 900g for 20 min to isolate the plasma followed by harvesting cell supernatants, centrifuging at 300g for 10 min to eliminate cells and at 16 500g for 30 min followed by filtration through a 0.22 mm filter. Exosomes are pelleted by ultracentrifugation at 120 000g for70 min. Chemical transfection of siRNA into exosomes is carried out according to the manufacturer's instructions in RNAi Human/Mouse Starter Kit (Quiagen, Hilden, Germany). siRNA is added to 100 ml PBS at a final concentration of 2 mmol/ml. After adding HiPerFect transfection reagent, the mixture is incubated for 10 min at RT. In order to remove the excess of micelles, the exosomes are re-isolated using aldehyde/sulfate latex beads. The chemical transfection of CRISPR Cas into exosomes may be conducted similarly to siRNA. The exosomes may be co-cultured with monocytes and lymphocytes isolated from the peripheral blood of healthy donors. Therefore, it may be contemplated that exosomes containing CRISPR Cas may be introduced to monocytes and lymphocytes of and autologously reintroduced into a human. Accordingly, delivery or administration according to the invention may beperformed using plasma exosomes.

Liposomes

[0243] Delivery or administration according to the invention can be performed with liposomes. Liposomes are spherical vesicle structures composed of a uni- or multilamellar lipid bilayer surrounding internal aqueous compartments and a relatively impermeable outer lipophilic phospholipid bilayer. Liposomes have gained considerable attention as drug delivery carriers because they are biocompatible, nontoxic, can deliver both hydrophilic and lipophilic drug molecules, protect their cargo from degradation by plasma enzymes, and transport their load across biological membranes and the blood brain barrier (BBB) (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0244] Liposomes can be made from several different types of lipids; however, phospholipids are most commonly used to generate liposomes as drug carriers. Although

liposome formation is spontaneous when a lipid film is mixed with an aqueous solution, it can also be expedited by applying force in the form of shaking by using a homogenizer, sonicator, or an extrusion apparatus (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0245] Several other additives may be added to liposomes in order to modify their structure and properties. For instance, either cholesterol or sphingomyelin may be added to the liposomal mixture in order to help stabilize the liposomal structure and to prevent the leakage of the liposomal inner cargo. Further, liposomes are prepared from hydrogenated egg phosphatidylcholine or egg phosphatidylcholine, cholesterol, and dicetyl phosphate, and their mean vesicle sizes were adjusted to about 50 and 100 nm. (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0246] A liposome formulation may be mainly comprised of natural phospholipids and lipids such as 1,2-distearoryl-sn-glycero-3-phosphatidyl choline (DSPC), sphingomyelin, egg phosphatidylcholines and monosialoganglioside. Since this formulation is made up of phospholipids only, liposomal formulations have encountered many challenges, one of the ones being the instability in plasma. Several attempts to overcome these challenges have been made, specifically in the manipulation of the lipid membrane. One of these attempts focused on the manipulation of cholesterol. Addition of cholesterol to conventional formulations reduces rapid release of the encapsulated bioactive compound into the plasma or 1,2-dioleoyl-sn-glycero-3-phosphoethanolamine (DOPE) increases the stability (see, e.g., Spuch and Navarro, Journal of Drug Delivery, vol. 2011, Article ID 469679, 12 pages, 2011. doi:10.1155/2011/469679 for review).

[0247] In a particularly advantageous embodiment, Trojan Horse liposomes (also known as Molecular Trojan Horses) are desirable and protocols may be found at http://cshprotocols.cshlp.org/content/2010/4/pdb.prot5407.long. These particles allow delivery of a transgene to the entire brain after an intravascular injection. Without being bound by limitation, it is believed that neutral lipid particles with specific antibodies conjugated to surface allow crossing of the blood brain barrier via endocytosis. Applicant postulates utilizing Trojan Horse Liposomes to deliver the CRISPR family of nucleases to the brain via an intravascular injection, which would allow whole brain transgenic animals without the need for embryonic manipulation. About 1-5 g of DNA or RNA may be contemplated for in vivo administration in liposomes.

[0248] In another embodiment, the CRISPR Cas system may be administered in liposomes, such as a stable nucleic-acid-lipid particle (SNALP) (see, e.g., Morrissey et al., Nature Biotechnology, Vol. 23, No. 8, August 2005). Daily intravenous injections of about 1, 3 or 5 mg/kg/day of a specific CRISPR Cas targeted in a SNALP are contemplated. The daily treatment may be over about three days and then weekly for about five weeks. In another embodiment, a specific CRISPR Cas encapsulated SNALP) administered by intravenous injection to at doses of about 1 or 2.5 mg/kg are also contemplated (see, e.g., Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006). The SNALP formulation may contain the lipids 3-N-[(wmethoxypoly(ethylene glycol) 2000) carbamoyl] -1,2-dimyristyloxy-propylamine (PEG-C-DMA), 1,2-dilinoleyloxy-N,N-dimethyl-3-aminopropane (DLinDMA), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC) and cholesterol, in a 2:40:10:48 molar per cent ratio (see, e.g., Zimmerman et al., Nature Letters, Vol. 441, 4 May 2006).

[0249] In another embodiment, stable nucleic-acid-lipid particles (SNALPs) have proven to be effective delivery molecules to highly vascularized HepG2-derived liver tumors but not in poorly vascularized HCT-116 derived liver tumors (see, e.g., Li, Gene Therapy (2012) 19, 775–780). The SNALP liposomes may be prepared by formulating D-Lin-DMA and PEG-C-DMA with distearoylphosphatidylcholine (DSPC), Cholesterol and siRNA using a 25:1 lipid/siRNA ratio and a 48/40/10/2 molar ratio of Cholesterol/D-Lin-DMA/DSPC/PEG-C-DMA. The resulted SNALP liposomes are about 80–100 nm in size.

[0250] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich, St Louis, MO, USA), dipalmitoylphosphatidylcholine (Avanti Polar Lipids, Alabaster, AL, USA), 3-N-[(w-methoxy poly(ethylene glycol)2000)carbamoyl]-1,2-dimyrestyloxypropylamine, and cationic 1,2-dilinoleyloxy-3-N,Ndimethylaminopropane (see, e.g., Geisbert et al., Lancet 2010; 375: 1896-905). A dosage of about 2 mg/kg total CRISPR Cas per dose administered as, for example, a bolus intravenous infusion may be contemplated.

[0251] In yet another embodiment, a SNALP may comprise synthetic cholesterol (Sigma-Aldrich), 1,2-distearoyl-sn-glycero-3-phosphocholine (DSPC; Avanti Polar Lipids Inc.), PEG-cDMA, and 1,2-dilinoleyloxy-3-(N;N-dimethyl)aminopropane (DLinDMA) (see, e.g., Judge, J. Clin. Invest. 119:661-673 (2009)). Formulations used for in vivo studies may comprise a final lipid/RNA mass ratio of about 9:1.

[0252] The safety profile of RNAi nanomedicines has been reviewed by Barros and Gollob of Alnylam Pharmaceuticals (see, e.g., Advanced Drug Delivery Reviews 64 (2012) 1730–1737). The stable nucleic acid lipid particle (SNALP) is comprised of four different lipids — an

ionizable lipid (DLinDMA) that is cationic at low pH, a neutral helper lipid, cholesterol, and a diffusible polyethylene glycol (PEG)-lipid. The particle is approximately 80 nm in diameter and is charge-neutral at physiologic pH. During formulation, the ionizable lipid serves to condense lipid with the anionic RNA during particle formation. When positively charged under increasingly acidic endosomal conditions, the ionizable lipid also mediates the fusion of SNALP with the endosomal membrane enabling release of RNA into the cytoplasm. The PEG-lipid stabilizes the particle and reduces aggregation during formulation, and subsequently provides a neutral hydrophilic exterior that improves pharmacokinetic properties.

[0253] To date, two clinical programs have been initiated using SNALP formulations with RNA. Tekmira Pharmaceuticals recently completed a phase I single-dose study of SNALP-ApoB in adult volunteers with elevated LDL cholesterol. ApoB is predominantly expressed in the liver and jejunum and is essential for the assembly and secretion of VLDL and LDL. Seventeen subjects received a single dose of SNALP-ApoB (dose escalation across 7 dose levels). There was no evidence of liver toxicity (anticipated as the potential dose-limiting toxicity based on preclinical studies). One (of two) subjects at the highest dose experienced flu-like symptoms consistent with immune system stimulation, and the decision was made to conclude the trial.

[0254] Alnylam Pharmaceuticals has similarly advanced ALN-TTR01, which employs the SNALP technology described above and targets hepatocyte production of both mutant and wild-type TTR to treat TTR amyloidosis (ATTR). Three ATTR syndromes have been described: familial amyloidotic polyneuropathy (FAP) and familial amyloidotic cardiomyopathy (FAC) — both caused by autosomal dominant mutations in TTR; and senile systemic amyloidosis (SSA) cause by wildtype TTR. A placebo-controlled, single dose-escalation phase I trial of ALN-TTR01 was recently completed in patients with ATTR. ALN-TTR01 was administered as a 15-minute IV infusion to 31 patients (23 with study drug and 8 with placebo) within a dose range of 0.01 to 1.0 mg/kg (based on siRNA). Treatment was well tolerated with no significant increases in liver function tests. Infusion-related reactions were noted in 3 of 23 patients at≥0.4 mg/kg; all responded to slowing of the infusion rate and all continued on study. Minimal and transient elevations of serum cytokines IL-6, IP-10 and IL-1ra were noted in two patients at the highest dose of 1 mg/kg (as anticipated from preclinical and NHP studies). Lowering of serum TTR, the expected pharmacodynamics effect of ALN-TTR01, was observed at 1 mg/kg.

[0255] In yet another embodiment, a SNALP may be made by solubilizing a cationic lipid, DSPC, cholesterol and PEG-lipid e.g., in ethanol, e.g., at a molar ratio of 40:10:40:10, respectively (see, Semple et al., Nature Niotechnology, Volume 28 Number 2 February 2010, pp.

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

Other Lipids

[0256] Other cationic lipids, such as amino lipid 2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]-dioxolane (DLin-KC2-DMA) may be utilized to encapsulate CRISPR Cas or components thereof or nucleic acid molecule(s) coding therefor e.g., similar to SiRNA (see, e.g., Jayaraman, Angew. Chem. Int. Ed. 2012, 51, 8529 –8533), and hence may be employed in the practice of the invention. A preformed vesicle with the following lipid composition may be contemplated: amino lipid, distearoylphosphatidylcholine (DSPC), cholesterol and (R)-2,3-bis(octadecyloxy) propyl-1-(methoxy poly(ethylene glycol)2000)propylcarbamate (PEG-lipid) in the molar ratio 40/10/40/10, respectively, and a FVII siRNA/total lipid ratio of approximately 0.05 (w/w). To ensure a narrow particle size distribution in the range of 70–90 nm and a low polydispersity

index of 0.11±0.04 (n=56), the particles may be extruded up to three times through 80 nm membranes prior to adding the CRISPR Cas RNA. Particles containing the highly potent amino lipid 16 may be used, in which the molar ratio of the four lipid components 16, DSPC, cholesterol and PEG-lipid (50/10/38.5/1.5) which may be further optimized to enhance in vivo activity.

[0257] Michael S D Kormann et al. ("Expression of therapeutic proteins after delivery of chemically modified mRNA in mice: Nature Biotechnology, Volume:29, Pages: 154–157 (2011)) describes the use of lipid envelopes to deliver RNA. Use of lipid envelopes is also preferred in the present invention.

[0258] In another embodiment, lipids may be formulated with the CRISPR Cas system of the present invention to form lipid nanoparticles (LNPs). Lipids include, but are not limited to, DLin-KC2-DMA4, C12-200 and colipids disteroylphosphatidyl choline, cholesterol, and PEG-DMG may be formulated with CRISPR Cas instead of siRNA (see, e.g., Novobrantseva, Molecular Therapy–Nucleic Acids (2012) 1, e4; doi:10.1038/mtna.2011.3) using a spontaneous vesicle formation procedure. The component molar ratio may be about 50/10/38.5/1.5 (DLin-KC2-DMA or C12-200/disteroylphosphatidyl choline/cholesterol/PEG-DMG). The final lipid:siRNA weight ratio may be ~12:1 and 9:1 in the case of DLin-KC2-DMA and C12-200 lipid nanoparticles (LNPs), respectively. The formulations may have mean particle diameters of ~80 nm with >90% entrapment efficiency. A 3 mg/kg dose may be contemplated.

[0259] Tekmira has a portfolio of approximately 95 patent families, in the U.S. and abroad, that are directed to various aspects of LNPs and LNP formulations (see, e.g., U.S. Pat. Nos. 7,982,027; 7,799,565; 8,058,069; 8,283,333; 7,901,708; 7,745,651; 7,803,397; 8,101,741; 8,188,263; 7,915,399; 8,236,943 and 7,838,658 and European Pat. Nos 1766035; 1519714; 1781593 and 1664316), all of which may be used and/or adapted to the present invention.

[0260] The CRISPR Cas system or components thereof or nucleic acid molecule(s) coding therefor may be delivered encapsulated in PLGA Microspheres such as that further described in US published applications 20130252281 and 20130245107 and 20130244279 (assigned to Moderna Therapeutics) which relate to aspects of formulation of compositions comprising modified nucleic acid molecules which may encode a protein, a protein precursor, or a partially or fully processed form of the protein or a protein precursor. The formulation may have a molar ratio 50:10:38.5:1.5-3.0 (cationic lipid:fusogenic lipid:cholesterol:PEG lipid). The PEG lipid may be selected from, but is not limited to PEG-c-DOMG, PEG-DMG. The fusogenic lipid may

be DSPC. See also, Schrum et al., Delivery and Formulation of Engineered Nucleic Acids, US published application 20120251618.

[0261] Nanomerics' technology addresses bioavailability challenges for a broad range of therapeutics, including low molecular weight hydrophobic drugs, peptides, and nucleic acid based therapeutics (plasmid, siRNA, miRNA). Specific administration routes for which the technology has demonstrated clear advantages include the oral route, transport across the blood-brain-barrier, delivery to solid tumours, as well as to the eye. See, e.g., Mazza et al., 2013, ACS Nano. 2013 Feb 26;7(2):1016-26; Uchegbu and Siew, 2013, J Pharm Sci. 102(2):305-10 and Lalatsa et al., 2012, J Control Release. 2012 Jul 20; 161(2):523-36.

[0262] US Patent Publication No. 20050019923 describes cationic dendrimers for delivering bioactive molecules, such as polynucleotide molecules, peptides and polypeptides and/or pharmaceutical agents, to a mammalian body. The dendrimers are suitable for targeting the delivery of the bioactive molecules to, for example, the liver, spleen, lung, kidney or heart (or even the brain). Dendrimers are synthetic 3-dimensional macromolecules that are prepared in a step-wise fashion from simple branched monomer units, the nature and functionality of which can be easily controlled and varied. Dendrimers are synthesised from the repeated addition of building blocks to a multifunctional core (divergent approach to synthesis), or towards a multifunctional core (convergent approach to synthesis) and each addition of a 3-dimensional shell of building blocks leads to the formation of a higher generation of the dendrimers. Polypropylenimine dendrimers start from a diaminobutane core to which is added twice the number of amino groups by a double Michael addition of acrylonitrile to the primary amines followed by the hydrogenation of the nitriles. This results in a doubling of the amino groups. Polypropylenimine dendrimers contain 100% protonable nitrogens and up to 64 terminal amino groups (generation 5, DAB 64). Protonable groups are usually amine groups which are able to accept protons at neutral pH. The use of dendrimers as gene delivery agents has largely focused on the use of the polyamidoamine. and phosphorous containing compounds with a mixture of amine/amide or N--P(O_2)S as the conjugating units respectively with no work being reported on the use of the lower generation polypropylenimine dendrimers for gene delivery. Polypropylenimine dendrimers have also been studied as pH sensitive controlled release systems for drug delivery and for their encapsulation of guest molecules when chemically modified by peripheral amino acid groups. The cytotoxicity and interaction of polypropylenimine dendrimers with DNA as well as the transfection efficacy of DAB 64 has also been studied.

[0263] US Patent Publication No. 20050019923 is based upon the observation that, contrary to earlier reports, cationic dendrimers, such as polypropylenimine dendrimers, display suitable properties, such as specific targeting and low toxicity, for use in the targeted delivery of bioactive molecules, such as genetic material. In addition, derivatives of the cationic dendrimer also display suitable properties for the targeted delivery of bioactive molecules. See also, Bioactive Polymers, US published application 20080267903, which discloses "Various polymers, including cationic polyamine polymers and dendrimeric polymers, are shown to possess anti-proliferative activity, and may therefore be useful for treatment of disorders characterised by undesirable cellular proliferation such as neoplasms and tumours, inflammatory disorders (including autoimmune disorders), psoriasis and atherosclerosis. The polymers may be used alone as active agents, or as delivery vehicles for other therapeutic agents, such as drug molecules or nucleic acids for gene therapy. In such cases, the polymers' own intrinsic anti-tumour activity may complement the activity of the agent to be delivered." The disclosures of these patent publications may be employed in conjunction with herein teachings for delivery of CRISPR Cas system(s) or component(s) thereof or nucleic acid molecule(s) coding therefor.

Supercharged proteins

[0264] Supercharged proteins are a class of engineered or naturally occurring proteins with unusually high positive or negative net theoretical charge and may be employed in delivery of CRISPR Cas system(s) or component(s) thereof or nucleic acid molecule(s) coding therefor. Both supernegatively and superpositively charged proteins exhibit a remarkable ability to withstand thermally or chemically induced aggregation. Superpositively charged proteins are also able to penetrate mammalian cells. Associating cargo with these proteins, such as plasmid DNA, RNA, or other proteins, can enable the functional delivery of these macromolecules into mammalian cells both in vitro and in vivo. David Liu's lab reported the creation and characterization of supercharged proteins in 2007 (Lawrence et al., 2007, Journal of the American Chemical Society 129, 10110–10112).

[0265] The nonviral delivery of RNA and plasmid DNA into mammalian cells are valuable both for research and therapeutic applications (Akinc et al., 2010, Nat. Biotech. 26, 561–569). Purified +36 GFP protein (or other superpositively charged protein) is mixed with RNAs in the appropriate serum-free media and allowed to complex prior addition to cells. Inclusion of serum at this stage inhibits formation of the supercharged protein-RNA complexes and reduces the effectiveness of the treatment. The following protocol has been found to be effective for a variety of cell lines (McNaughton et al., 2009, Proc. Natl. Acad. Sci. USA 106, 6111–6116).

However, pilot experiments varying the dose of protein and RNA should be performed to optimize the procedure for specific cell lines.

(1) One day before treatment, plate 1 x 10^5 cells per well in a 48-well plate.

(2) On the day of treatment, dilute purified +36 GFP protein in serumfree media to a final concentration 200nM. Add RNA to a final concentration of 50nM. Vortex to mix and incubate at room temperature for 10min.

(3) During incubation, aspirate media from cells and wash once with PBS.

(4) Following incubation of +36 GFP and RNA, add the protein-RNA complexes to cells.

(5) Incubate cells with complexes at 37 °C for 4h.

(6) Following incubation, aspirate the media and wash three times with 20 U/mL heparin PBS. Incubate cells with serum-containing media for a further 48h or longer depending upon the assay for activity.

(7) Analyze cells by immunoblot, qPCR, phenotypic assay, or other appropriate method.

[0266] David Liu's lab has further found +36 GFP to be an effective plasmid delivery reagent in a range of cells. As plasmid DNA is a larger cargo than siRNA, proportionately more +36 GFP protein is required to effectively complex plasmids. For effective plasmid delivery Applicants have developed a variant of +36 GFP bearing a C-terminal HA2 peptide tag, a known endosome-disrupting peptide derived from the influenza virus hemagglutinin protein. The following protocol has been effective in a variety of cells, but as above it is advised that plasmid DNA and supercharged protein doses be optimized for specific cell lines and delivery applications.

(1) One day before treatment, plate 1 x 10^5 per well in a 48-well plate.

(2) On the day of treatment, dilute purified þ36 GFP protein in serumfree media to a final concentration 2 mM. Add 1mg of plasmid DNA. Vortex to mix and incubate at room temperature for 10min.

(3) During incubation, aspirate media from cells and wash once with PBS.

(4) Following incubation of þ36 GFP and plasmid DNA, gently add the protein-DNA complexes to cells.

(5) Incubate cells with complexes at 37 C for 4h.

(6) Following incubation, aspirate the media and wash with PBS. Incubate cells in serum-containing media and incubate for a further 24--48h.

(7) Analyze plasmid delivery (e.g., by plasmid-driven gene expression) as appropriate.

[0267] See also, e.g., McNaughton et al., Proc. Natl. Acad. Sci. USA 106, 6111-6116 (2009); Cronican et al., ACS Chemical Biology 5, 747-752 (2010); Cronican et al., Chemistry & Biology 18, 833-838 (2011); Thompson et al., Methods in Enzymology 503, 293-319 (2012); Thompson, D.B., et al., Chemistry & Biology 19 (7), 831-843 (2012). The methods of the super charged proteins may be used and/or adapted for delivery of the CRISPR Cas system of the present invention. These systems of Dr. Lui and documents herein in inconjunction with herein teachints can be employed in the delivery of CRISPR Cas system(s) or component(s) thereof or nucleic acid molecule(s) coding therefor.

Implantable devices

[0268] In another embodiment, implantable devices are also contemplated for delivery of the CRISPR Cas system or component(s) thereof or nucleic acid molecule(s) coding therefor. For example, US Patent Publication 20110195123 discloses an implantable medical device which elutes a drug locally and in prolonged period is provided, including several types of such a device, the treatment modes of implementation and methods of implantation. The device comprising of polymeric substrate, such as a matrix for example, that is used as the device body, and drugs, and in some cases additional scaffolding materials, such as metals or additional polymers, and materials to enhance visibility and imaging. An implantable delivery device can be advantageous in providing release locally and over a prolonged period, where drug is released directly to the extracellular matrix (ECM) of the diseased area such as tumor, inflammation, degeneration or for symptomatic objectives, or to injured smooth muscle cells, or for prevention. One kind of drug is RNA, as disclosed above, and this system may be used/and or adapted to the CRISPR Cas system of the present invention. The modes of implantation in some embodiments are existing implantation procedures that are developed and used today for other treatments, including brachytherapy and needle biopsy. In such cases the dimensions of the new implant described in this invention are similar to the original implant. Typically a few devices are implanted during the same treatment procedure.

[0269] As described in US Patent Publication 20110195123, there is provided a drug delivery implantable or insertable system, including systems applicable to a cavity such as the abdominal cavity and/or any other type of administration in which the drug delivery system is not anchored or attached, comprising a biostable and/or degradable and/or bioabsorbable polymeric substrate, which may for example optionally be a matrix. It should be noted that the

term "insertion" also includes implantation. The drug delivery system is preferably implemented as a "Loder" as described in US Patent Publication 20110195123.

[0270] The polymer or plurality of polymers are biocompatible, incorporating an agent and/or plurality of agents, enabling the release of agent at a controlled rate, wherein the total volume of the polymeric substrate, such as a matrix for example, in some embodiments is optionally and preferably no greater than a maximum volume that permits a therapeutic level of the agent to be reached. As a non-limiting example, such a volume is preferably within the range of 0.1 m^3 to 1000 mm^3, as required by the volume for the agent load. The Loder may optionally be larger, for example when incorporated with a device whose size is determined by functionality, for example and without limitation, a knee joint, an intra-uterine or cervical ring and the like.

[0271] The drug delivery system (for delivering the composition) is designed in some embodiments to preferably employ degradable polymers, wherein the main release mechanism is bulk erosion; or in some embodiments, non degradable, or slowly degraded polymers are used, wherein the main release mechanism is diffusion rather than bulk erosion, so that the outer part functions as membrane, and its internal part functions as a drug reservoir, which practically is not affected by the surroundings for an extended period (for example from about a week to about a few months). Combinations of different polymers with different release mechanisms may also optionally be used. The concentration gradient at the surface is preferably maintained effectively constant during a significant period of the total drug releasing period, and therefore the diffusion rate is effectively constant (termed "zero mode" diffusion). By the term "constant" it is meant a diffusion rate that is preferably maintained above the lower threshold of therapeutic effectiveness, but which may still optionally feature an initial burst and/or may fluctuate, for example increasing and decreasing to a certain degree. The diffusion rate is preferably so maintained for a prolonged period, and it can be considered constant to a certain level to optimize the therapeutically effective period, for example the effective silencing period.

[0272] The drug delivery system optionally and preferably is designed to shield the nucleotide based therapeutic agent from degradation, whether chemical in nature or due to attack from enzymes and other factors in the body of the subject.

[0273] The drug delivery system as described in US Patent Publication 20110195123 is optionally associated with sensing and/or activation appliances that are operated at and/or after implantation of the device, by non and/or minimally invasive methods of activation and/or acceleration/deceleration, for example optionally including but not limited to thermal heating

and cooling, laser beams, and ultrasonic, including focused ultrasound and/or RF (radiofrequency) methods or devices.

[0274] According to some embodiments of US Patent Publication 20110195123, the site for local delivery may optionally include target sites characterized by high abnormal proliferation of cells, and suppressed apoptosis, including tumors, active and or chronic inflammation and infection including autoimmune diseases states, degenerating tissue including muscle and nervous tissue, chronic pain, degenerative sites, and location of bone fractures and other wound locations for enhancement of regeneration of tissue, and injured cardiac, smooth and striated muscle.

[0275] The site for implantation of the composition, or target site, preferably features a radius, area and/or volume that is sufficiently small for targeted local delivery. For example, the target site optionally has a diameter in a range of from about 0.1 mm to about 5 cm.

[0276] The location of the target site is preferably selected for maximum therapeutic efficacy. For example, the composition of the drug delivery system (optionally with a device for implantation as described above) is optionally and preferably implanted within or in the proximity of a tumor environment, or the blood supply associated thereof.

[0277] For example the composition (optionally with the device) is optionally implanted within or in the proximity to pancreas, prostate, breast, liver, via the nipple, within the vascular system and so forth.

[0278] The target location is optionally selected from the group consisting of (as non-limiting examples only, as optionally any site within the body may be suitable for implanting a Loder): 1. brain at degenerative sites like in Parkinson or Alzheimer disease at the basal ganglia, white and gray matter; 2. spine as in the case of amyotrophic lateral sclerosis (ALS); 3. uterine cervix to prevent HPV infection; 4. active and chronic inflammatory joints; 5. dermis as in the case of psoriasis; 6. sympathetic and sensoric nervous sites for analgesic effect; 7. Intra osseous implantation; 8. acute and chronic infection sites; 9. Intra vaginal; 10. Inner ear--auditory system, labyrinth of the inner ear, vestibular system; 11. Intra tracheal; 12. Intra-cardiac; coronary, epicardiac; 13. urinary bladder; 14. biliary system; 15. parenchymal tissue including and not limited to the kidney, liver, spleen; 16. lymph nodes; 17. salivary glands; 18. dental gums; 19. Intra-articular (into joints); 20. Intra-ocular; 21. Brain tissue; 22. Brain ventricles; 23. Cavities, including abdominal cavity (for example but without limitation, for ovary cancer); 24. Intra esophageal and 25. Intra rectal.

[0279] Optionally insertion of the system (for example a device containing the composition) is associated with injection of material to the ECM at the target site and the vicinity of that site to affect local pH and/or temperature and/or other biological factors affecting the diffusion of the drug and/or drug kinetics in the ECM, of the target site and the vicinity of such a site.

[0280] Optionally, according to some embodiments, the release of said agent could be associated with sensing and/or activation appliances that are operated prior and/or at and/or after insertion, by non and/or minimally invasive and/or else methods of activation and/or acceleration/deceleration, including laser beam, radiation, thermal heating and cooling, and ultrasonic, including focused ultrasound and/or RF (radiofrequency) methods or devices, and chemical activators.

[0281] According to other embodiments of US Patent Publication 20110195123, the drug preferably comprises a RNA, for example for localized cancer cases in breast, pancreas, brain, kidney, bladder, lung, and prostate as described below. Although exemplified with RNAi, many drugs are applicable to be encapsulated in Loder, and can be used in association with this invention, as long as such drugs can be encapsulated with the Loder substrate, such as a matrix for example, and this system may be used and/or adapted to deliver the CRISPR Cas system of the present invention.

[0282] As another example of a specific application, neuro and muscular degenerative diseases develop due to abnormal gene expression. Local delivery of RNAs may have therapeutic properties for interfering with such abnormal gene expression. Local delivery of anti apoptotic, anti inflammatory and anti degenerative drugs including small drugs and macromolecules may also optionally be therapeutic. In such cases the Loder is applied for prolonged release at constant rate and/or through a dedicated device that is implanted separately. All of this may be used and/or adapted to the CRISPR Cas system of the present invention.

[0283] As yet another example of a specific application, psychiatric and cognitive disorders are treated with gene modifiers. Gene knockdown is a treatment option. Loders locally delivering agents to central nervous system sites are therapeutic options for psychiatric and cognitive disorders including but not limited to psychosis, bi-polar diseases, neurotic disorders and behavioral maladies. The Loders could also deliver locally drugs including small drugs and macromolecules upon implantation at specific brain sites. All of this may be used and/or adapted to the CRISPR Cas system of the present invention.

[0284] As another example of a specific application, silencing of innate and/or adaptive immune mediators at local sites enables the prevention of organ transplant rejection. Local

delivery of RNAs and immunomodulating reagents with the Loder implanted into the transplanted organ and/or the implanted site renders local immune suppression by repelling immune cells such as CD8 activated against the transplanted organ. All of this may be used/and or adapted to the CRISPR Cas system of the present invention.

[0285] As another example of a specific application, vascular growth factors including VEGFs and angiogenin and others are essential for neovascularization. Local delivery of the factors, peptides, peptidomimetics, or suppressing their repressors is an important therapeutic modality; silencing the repressors and local delivery of the factors, peptides, macromolecules and small drugs stimulating angiogenesis with the Loder is therapeutic for peripheral, systemic and cardiac vascular disease.

[0286] The method of insertion, such as implantation, may optionally already be used for other types of tissue implantation and/or for insertions and/or for sampling tissues, optionally without modifications, or alternatively optionally only with non-major modifications in such methods. Such methods optionally include but are not limited to brachytherapy methods, biopsy, endoscopy with and/or without ultrasound, such as ERCP, stereotactic methods into the brain tissue, Laparoscopy, including implantation with a laparoscope into joints, abdominal organs, the bladder wall and body cavities.

Patient-specific screening methods

[0287] A CRISPR-Cas system that targets nucleotide, e.g., trinucleotide repeats can be used to screen patients or patent samples for the presence of such repeats. The repeats can be the target of the RNA of the CRISPR-Cas system, and if there is binding thereto by the CRISPR-Cas system, that binding can be detected, to thereby indicate that such a repeat is present. Thus, a CRISPR-Cas system can be used to screen patients or patient samples for the presence of the repeat. The patient can then be administered suitable compound(s) to address the condition; or, can be administed a CRISPR-Cas system to bind to and cause insertion, deletion or mutation and alleviate the condition.

Nucleic acids, amino acids and proteins, Regulatory sequences, Vectors, etc

[0288] Nucleic acids, amino acids and proteins: The invention uses nucleic acids to bind target DNA sequences. This is advantageous as nucleic acids are much easier and cheaper to produce than proteins, and the specificity can be varied according to the length of the stretch where homology is sought. Complex 3-D positioning of multiple fingers, for example is not required. The terms "polynucleotide", "nucleotide", "nucleotide sequence", "nucleic acid" and "oligonucleotide" are used interchangeably. They refer to a polymeric form of nucleotides of

any length, either deoxyribonucleotides or ribonucleotides, or analogs thereof. Polynucleotides may have any three dimensional structure, and may perform any function, known or unknown. The following are non-limiting examples of polynucleotides: coding or non-coding regions of a gene or gene fragment, loci (locus) defined from linkage analysis, exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, short interfering RNA (siRNA), short-hairpin RNA (shRNA), micro-RNA (miRNA), ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes, and primers. The term also encompasses nucleic-acid-like structures with synthetic backbones, see, e.g., Eckstein, 1991; Baserga et al., 1992; Milligan, 1993; WO 97/03211; WO 96/39154; Mata, 1997; Strauss-Soukup, 1997; and Samstag, 1996. A polynucleotide may comprise one or more modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure may be imparted before or after assembly of the polymer. The sequence of nucleotides may be interrupted by non-nucleotide components. A polynucleotide may be further modified after polymerization, such as by conjugation with a labeling component. As used herein the term "wild type" is a term of the art understood by skilled persons and means the typical form of an organism, strain, gene or characteristic as it occurs in nature as distinguished from mutant or variant forms. A "wild type" can be a base line. As used herein the term "variant" should be taken to mean the exhibition of qualities that have a pattern that deviates from what occurs in nature. The terms "non-naturally occurring" or "engineered" are used interchangeably and indicate the involvement of the hand of man. The terms, when referring to nucleic acid molecules or polypeptides mean that the nucleic acid molecule or the polypeptide is at least substantially free from at least one other component with which they are naturally associated in nature and as found in nature. "Complementarity" refers to the ability of a nucleic acid to form hydrogen bond(s) with another nucleic acid sequence by either traditional Watson-Crick base pairing or other non-traditional types. A percent complementarity indicates the percentage of residues in a nucleic acid molecule which can form hydrogen bonds (e.g., Watson-Crick base pairing) with a second nucleic acid sequence (e.g., 5, 6, 7, 8, 9, 10 out of 10 being 50%, 60%, 70%, 80%, 90%, and 100% complementary). "Perfectly complementary" means that all the contiguous residues of a nucleic acid sequence will hydrogen bond with the same number of contiguous residues in a second nucleic acid sequence. "Substantially complementary" as used herein refers to a degree of complementarity that is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 97%, 98%, 99%, or 100% over a region of 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25,

30, 35, 40, 45, 50, or more nucleotides, or refers to two nucleic acids that hybridize under stringent conditions. As used herein, "stringent conditions" for hybridization refer to conditions under which a nucleic acid having complementarity to a target sequence predominantly hybridizes with the target sequence, and substantially does not hybridize to non-target sequences. Stringent conditions are generally sequence-dependent, and vary depending on a number of factors. In general, the longer the sequence, the higher the temperature at which the sequence specifically hybridizes to its target sequence. Non-limiting examples of stringent conditions are described in detail in Tijssen (1993), Laboratory Techniques In Biochemistry And Molecular Biology-Hybridization With Nucleic Acid Probes Part I, Second Chapter "Overview of principles of hybridization and the strategy of nucleic acid probe assay", Elsevier, N.Y. Where reference is made to a polynucleotide sequence, then complementary or partially complementary sequences are also envisaged. These are preferably capable of hybridising to the reference sequence under highly stringent conditions. Generally, in order to maximize the hybridization rate, relatively low-stringency hybridization conditions are selected: about 20 to 25° C lower than the thermal melting point (T_m). The T_m is the temperature at which 50% of specific target sequence hybridizes to a perfectly complementary probe in solution at a defined ionic strength and pH. Generally, in order to require at least about 85% nucleotide complementarity of hybridized sequences, highly stringent washing conditions are selected to be about 5 to 15° C lower than the T_m. In order to require at least about 70% nucleotide complementarity of hybridized sequences, moderately-stringent washing conditions are selected to be about 15 to 30° C lower than the T_m. Highly permissive (very low stringency) washing conditions may be as low as 50° C below the T_m, allowing a high level of mis-matching between hybridized sequences. Those skilled in the art will recognize that other physical and chemical parameters in the hybridization and wash stages can also be altered to affect the outcome of a detectable hybridization signal from a specific level of homology between target and probe sequences. Preferred highly stringent conditions comprise incubation in 50% formamide, 5×SSC, and 1% SDS at 42° C, or incubation in 5×SSC and 1% SDS at 65° C, with wash in 0.2×SSC and 0.1% SDS at 65° C. "Hybridization" refers to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding may occur by Watson Crick base pairing, Hoogstein binding, or in any other sequence specific manner. The complex may comprise two strands forming a duplex structure, three or more strands forming a multi stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction may constitute a step

in a more extensive process, such as the initiation of PCR, or the cleavage of a polynucleotide by an enzyme. A sequence capable of hybridizing with a given sequence is referred to as the "complement" of the given sequence. As used herein, the term "genomic locus" or "locus" (plural loci) is the specific location of a gene or DNA sequence on a chromosome. A "gene" refers to stretches of DNA or RNA that encode a polypeptide or an RNA chain that has functional role to play in an organism and hence is the molecular unit of heredity in living organisms. For the purpose of this invention it may be considered that genes include regions which regulate the production of the gene product, whether or not such regulatory sequences are adjacent to coding and/or transcribed sequences. Accordingly, a gene includes, but is not necessarily limited to, promoter sequences, terminators, translational regulatory sequences such as ribosome binding sites and internal ribosome entry sites, enhancers, silencers, insulators, boundary elements, replication origins, matrix attachment sites and locus control regions. As used herein, "expression of a genomic locus" or "gene expression" is the process by which information from a gene is used in the synthesis of a functional gene product. The products of gene expression are often proteins, but in non-protein coding genes such as rRNA genes or tRNA genes, the product is functional RNA. The process of gene expression is used by all known life - eukaryotes (including multicellular organisms), prokaryotes (bacteria and archaea) and viruses to generate functional products to survive. As used herein "expression" of a gene or nucleic acid encompasses not only cellular gene expression, but also the transcription and translation of nucleic acid(s) in cloning systems and in any other context. As used herein, "expression" also refers to the process by which a polynucleotide is transcribed from a DNA template (such as into and mRNA or other RNA transcript) and/or the process by which a transcribed mRNA is subsequently translated into peptides, polypeptides, or proteins. Transcripts and encoded polypeptides may be collectively referred to as "gene product." If the polynucleotide is derived from genomic DNA, expression may include splicing of the mRNA in a eukaryotic cell. The terms "polypeptide", "peptide" and "protein" are used interchangeably herein to refer to polymers of amino acids of any length. The polymer may be linear or branched, it may comprise modified amino acids, and it may be interrupted by non amino acids. The terms also encompass an amino acid polymer that has been modified; for example, disulfide bond formation, glycosylation, lipidation, acetylation, phosphorylation, or any other manipulation, such as conjugation with a labeling component. As used herein the term "amino acid" includes natural and/or unnatural or synthetic amino acids, including glycine and both the D or L optical isomers, and amino acid analogs and peptidomimetics. As used herein, the term

"domain" or "protein domain" refers to a part of a protein sequence that may exist and function independently of the rest of the protein chain. As described in aspects of the invention, sequence identity is related to sequence homology. Homology comparisons may be conducted by eye, or more usually, with the aid of readily available sequence comparison programs. These commercially available computer programs may calculate percent (%) homology between two or more sequences and may also calculate the sequence identity shared by two or more amino acid or nucleic acid sequences. In some preferred embodiments, the capping region of the dTALEs described herein have sequences that are at least 95% identical or share identity to the capping region amino acid sequences provided herein. Sequence homologies may be generated by any of a number of computer programs known in the art, for example BLAST or FASTA, etc. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (University of Wisconsin, U.S.A; Devereux et al., 1984, Nucleic Acids Research 12:387). Examples of other software than may perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel et al., 1999 ibid – Chapter 18), FASTA (Atschul et al., 1990, J. Mol. Biol., 403-410) and the GENEWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel et al., 1999 ibid, pages 7-58 to 7-60). However it is preferred to use the GCG Bestfit program. Percentage (%) sequence homology may be calculated over contiguous sequences, i.e., one sequence is aligned with the other sequence and each amino acid or nucleotide in one sequence is directly compared with the corresponding amino acid or nucleotide in the other sequence, one residue at a time. This is called an "ungapped" alignment. Typically, such ungapped alignments are performed only over a relatively short number of residues. Although this is a very simple and consistent method, it fails to take into consideration that, for example, in an otherwise identical pair of sequences, one insertion or deletion may cause the following amino acid residues to be put out of alignment, thus potentially resulting in a large reduction in % homology when a global alignment is performed. Consequently, most sequence comparison methods are designed to produce optimal alignments that take into consideration possible insertions and deletions without unduly penalizing the overall homology or identity score. This is achieved by inserting "gaps" in the sequence alignment to try to maximize local homology or identity. However, these more complex methods assign "gap penalties" to each gap that occurs in the alignment so that, for the same number of identical amino acids, a sequence alignment with as few gaps as possible - reflecting higher relatedness between the two compared sequences - may achieve a higher score than one with many gaps. "Affinity gap costs" are typically used that charge a relatively high

cost for the existence of a gap and a smaller penalty for each subsequent residue in the gap. This is the most commonly used gap scoring system. High gap penalties may, of course, produce optimized alignments with fewer gaps. Most alignment programs allow the gap penalties to be modified. However, it is preferred to use the default values when using such software for sequence comparisons. For example, when using the GCG Wisconsin Bestfit package the default gap penalty for amino acid sequences is -12 for a gap and -4 for each extension. Calculation of maximum % homology therefore first requires the production of an optimal alignment, taking into consideration gap penalties. A suitable computer program for carrying out such an alignment is the GCG Wisconsin Bestfit package (Devereux et al., 1984 *Nuc. Acids Research* 12 p387). Examples of other software than may perform sequence comparisons include, but are not limited to, the BLAST package (see Ausubel et al., 1999 *Short Protocols in Molecular Biology*, 4th Ed. – Chapter 18), FASTA (Altschul et al., 1990 *J. Mol. Biol.* 403-410) and the GENEWORKS suite of comparison tools. Both BLAST and FASTA are available for offline and online searching (see Ausubel et al., 1999, *Short Protocols in Molecular Biology*, pages 7-58 to 7-60). However, for some applications, it is preferred to use the GCG Bestfit program. A new tool, called BLAST 2 Sequences is also available for comparing protein and nucleotide sequences (see *FEMS Microbiol Lett.* 1999 174(2): 247-50; *FEMS Microbiol Lett.* 1999 177(1): 187-8 and the website of the National Center for Biotechnology information at the website of the National Institutes for Health). Although the final % homology may be measured in terms of identity, the alignment process itself is typically not based on an all-or-nothing pair comparison. Instead, a scaled similarity score matrix is generally used that assigns scores to each pair-wise comparison based on chemical similarity or evolutionary distance. An example of such a matrix commonly used is the BLOSUM62 matrix - the default matrix for the BLAST suite of programs. GCG Wisconsin programs generally use either the public default values or a custom symbol comparison table, if supplied (see user manual for further details). For some applications, it is preferred to use the public default values for the GCG package, or in the case of other software, the default matrix, such as BLOSUM62. Alternatively, percentage homologies may be calculated using the multiple alignment feature in DNASIS^{TM} (Hitachi Software), based on an algorithm, analogous to CLUSTAL (Higgins DG & Sharp PM (1988), *Gene* 73(1), 237-244). Once the software has produced an optimal alignment, it is possible to calculate % homology, preferably % sequence identity. The software typically does this as part of the sequence comparison and generates a numerical result. The sequences may also have deletions, insertions or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent

substance. Deliberate amino acid substitutions may be made on the basis of similarity 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. However, it is more useful to include mutation data as well. The sets of amino acids thus derived are likely to be conserved for structural 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 below which describes a 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

[0289] Embodiments of the invention 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.

[0290] For purpose of this invention, amplification means any method employing a primer and a polymerase capable of replicating a target sequence with reasonable fidelity. Amplification may be carried out by natural or recombinant DNA polymerases such as TaqGold™, T7 DNA polymerase, Klenow fragment of E.coli DNA polymerase, and reverse transcriptase. A preferred amplification method is PCR.

[0291] In certain aspects the invention involves vectors. A used herein, a "vector" is a tool that allows or facilitates the transfer of an entity from one environment to another. It is a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements. In general, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a "plasmid," which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses (AAVs)). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-

linked. Such vectors are referred to herein as "expression vectors." Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0292] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell). With regards to recombination and cloning methods, mention is made of U.S. patent application 10/815,730, published September 2, 2004 as US 2004-0171156 A1, the contents of which are herein incorporated by reference in their entirety.

[0293] Aspects of the invention relate to bicistronic vectors for chimeric RNA and Cas9. Bicistronic expression vectors for chimeric RNA and Cas9 are preferred. In general and particularly in this embodiment Cas9 is preferably driven by the CBh promoter. The chimeric RNA may preferably be driven by a Pol III promoter, such as a U6 promoter. Ideally the two are combined. The chimeric guide RNA typically consists of a 20bp guide sequence (Ns) and this may be joined to the tracr sequence (running from the first "U" of the lower strand to the end of the transcript). The tracr sequence may be truncated at various positions as indicated. The guide and tracr sequences are separated by the tracr-mate sequence, which may be GUUUUAGAGCUA. This may be followed by the loop sequence GAAA as shown. Both of these are preferred examples. Applicants have demonstrated Cas9-mediated indels at the human *EMX1* and *PVALB* loci by SURVEYOR assays. ChiRNAs are indicated by their "+n" designation, and crRNA refers to a hybrid RNA where guide and tracr sequences are expressed as separate transcripts. Throughout this application, chimeric RNA may also be called single guide, or synthetic guide RNA (sgRNA). The loop is preferably GAAA, but it is not limited to this sequence or indeed to being only 4bp in length. Indeed, preferred loop forming sequences for use in hairpin structures are four nucleotides in length, and most preferably have the sequence GAAA. However, longer or shorter loop sequences may be used, as may alternative sequences. The sequences preferably include a nucleotide triplet (for example, AAA), and an additional nucleotide (for example C or G). Examples of loop forming sequences include CAAA and AAAG. In practicing any of the methods disclosed herein, a suitable vector can be

introduced to a cell or an embryo via one or more methods known in the art, including without limitation, microinjection, electroporation, sonoporation, biolistics, calcium phosphate-mediated transfection, cationic transfection, liposome transfection, dendrimer transfection, heat shock transfection, nucleofection transfection, magnetofection, lipofection, impalefection, optical transfection, proprietary agent-enhanced uptake of nucleic acids, and delivery via liposomes, immunoliposomes, virosomes, or artificial virions. In some methods, the vector is introduced into an embryo by microinjection. The vector or vectors may be microinjected into the nucleus or the cytoplasm of the embryo. In some methods, the vector or vectors may be introduced into a cell by nucleofection.

[0294] The term "regulatory element" is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g. transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g. liver, pancreas), or particular cell types (e.g. lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector comprises one or more pol III promoter (e.g. 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g. 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g. 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) [see, e.g., Boshart et al, Cell, 41:521-530 (1985)], the SV40 promoter, the dihydrofolate reductase promoter, the β-actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1α promoter. Also encompassed by the term "regulatory element" are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β-globin (Proc. Natl. Acad. Sci. USA., Vol.

78(3), p. 1527-31, 1981). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc. A vector can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (e.g., clustered regularly interspersed short palindromic repeats (CRISPR) transcripts, proteins, enzymes, mutant forms thereof, fusion proteins thereof, etc.). With regards to regulatory sequences, mention is made of U.S. patent application 10/491,026, the contents of which are incorporated by reference herein in their entirety. With regards to promoters, mention is made of PCT publication WO 2011/028929 and U.S. application 12/511,940, the contents of which are incorporated by reference herein in their entirety.

[0295] Vectors can be designed for expression of CRISPR transcripts (e.g. nucleic acid transcripts, proteins, or enzymes) in prokaryotic or eukaryotic cells. For example, CRISPR transcripts can be expressed in bacterial cells such as Escherichia coli, insect cells (using baculovirus expression vectors), yeast cells, or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be transcribed and translated in vitro, for example using T7 promoter regulatory sequences and T7 polymerase.

[0296] Vectors may be introduced and propagated in a prokaryote or prokaryotic cell. In some embodiments, a prokaryote is used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a eukaryotic cell (e.g. amplifying a plasmid as part of a viral vector packaging system). In some embodiments, a prokaryote is used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism. Expression of proteins in prokaryotes is most often carried out in Escherichia coli with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, such as to the amino terminus of the recombinant protein. Such fusion vectors may serve one or more purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety

subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Example fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion E. coli expression vectors include pTrc (Amrann et al., (1988) Gene 69:301-315) and pET 11d (Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89). In some embodiments, a vector is a yeast expression vector. Examples of vectors for expression in yeast Saccharomyces cerivisae include pYepSec1 (Baldari, et al., 1987. EMBO J. 6: 229-234), pMFa (Kuijan and Herskowitz, 1982. Cell 30: 933-943), pJRY88 (Schultz et al., 1987. Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (InVitrogen Corp, San Diego, Calif.). In some embodiments, a vector drives protein expression in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, et al., 1983. Mol. Cell. Biol. 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. Virology 170: 31-39).

[0297] In some embodiments, a vector is capable of driving expression of one or more sequences in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. Nature 329: 840) and pMT2PC (Kaufman, et al., 1987. EMBO J. 6: 187-195). When used in mammalian cells, the expression vector's control functions are typically provided by one or more regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and known in the art. For other suitable expression systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of Sambrook, et al., MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

[0298] In some embodiments, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, et al., 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in

particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Baneiji, et al., 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, et al., 1985. *Science* 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α-fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546). With regards to these prokaryotic and eukaryotic vectors, mention is made of U.S. Patent 6,750,059, the contents of which are incorporated by reference herein in their entirety. Other embodiments of the invention may relate to the use of viral vectors, with regards to which mention is made of U.S. Patent application 13/092,085, the contents of which are incorporated by reference herein in their entirety. Tissue-specific regulatory elements are known in the art and in this regard, mention is made of U.S. Patent 7,776,321, the contents of which are incorporated by reference herein in their entirety. In some embodiments, a regulatory element is operably linked to one or more elements of a CRISPR system so as to drive expression of the one or more elements of the CRISPR system. In general, CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats), also known as SPIDRs (SPacer Interspersed Direct Repeats), constitute a family of DNA loci that are usually specific to a particular bacterial species. The CRISPR locus comprises a distinct class of interspersed short sequence repeats (SSRs) that were recognized in E. coli (Ishino et al., J. Bacteriol., 169:5429-5433 [1987]; and Nakata et al., J. Bacteriol., 171:3553-3556 [1989]), and associated genes. Similar interspersed SSRs have been identified in Haloferax mediterranei, Streptococcus pyogenes, Anabaena, and Mycobacterium tuberculosis (See, Groenen et al., Mol. Microbiol., 10:1057-1065 [1993]; Hoe et al., Emerg. Infect. Dis., 5:254-263 [1999]; Masepohl et al., Biochim. Biophys. Acta 1307:26-30 [1996]; and Mojica et al., Mol. Microbiol., 17:85-93 [1995]). The CRISPR loci typically differ from other SSRs by the structure of the repeats, which have been termed short regularly spaced repeats (SRSRs) (Janssen et al., OMICS J. Integ. Biol., 6:23-33 [2002]; and Mojica et al., Mol. Microbiol., 36:244-246 [2000]). In general, the repeats are short elements that occur in clusters that are regularly spaced by unique intervening sequences with a substantially constant length (Mojica et al., [2000], supra). Although the repeat sequences are highly conserved between strains, the number of interspersed repeats and the sequences of the spacer regions typically differ from strain to strain (van Embden et al., J. Bacteriol., 182:2393-2401 [2000]). CRISPR

loci have been identified in more than 40 prokaryotes (See e.g., Jansen et al., Mol. Microbiol., 43:1565-1575 [2002]; and Mojica et al., [2005]) including, but not limited to Aeropyrum, Pyrobaculum, Sulfolobus, Archaeoglobus, Halocarcula, Methanobacterium, Methanococcus, Methanosarcina, Methanopyrus, Pyrococcus, Picrophilus, Thermoplasma, Corynebacterium, Mycobacterium, Streptomyces, Aquifex, Porphyromonas, Chlorobium, Thermus, Bacillus, Listeria, Staphylococcus, Clostridium, Thermoanaerobacter, Mycoplasma, Fusobacterium, Azarcus, Chromobacterium, Neisseria, Nitrosomonas, Desulfovibrio, Geobacter, Myxococcus, Campylobacter, Wolinella, Acinetobacter, Erwinia, Escherichia, Legionella, Methylococcus, Pasteurella, Photobacterium, Salmonella, Xanthomonas, Yersinia, Treponema, and Thermotoga.

[0299] In some embodiments, the CRISPR enzyme is part of a fusion protein comprising one or more heterologous protein domains (e.g. about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more domains in addition to the CRISPR enzyme). A CRISPR enzyme fusion protein may comprise any additional protein sequence, and optionally a linker sequence between any two domains. Examples of protein domains that may be fused to a CRISPR enzyme include, without limitation, epitope tags, reporter gene sequences, and protein domains having one or more of the following activities: methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity and nucleic acid binding activity. Non-limiting examples of epitope tags include histidine (His) tags, V5 tags, FLAG tags, influenza hemagglutinin (HA) tags, Myc tags, VSV-G tags, and thioredoxin (Trx) tags. Examples of reporter genes include, but are not limited to, glutathione-S-transferase (GST), horseradish peroxidase (HRP), chloramphenicol acetyltransferase (CAT) beta-galactosidase, beta-glucuronidase, luciferase, green fluorescent protein (GFP), HcRed, DsRed, cyan fluorescent protein (CFP), yellow fluorescent protein (YFP), and autofluorescent proteins including blue fluorescent protein (BFP). A CRISPR enzyme may be fused to a gene sequence encoding a protein or a fragment of a protein that bind DNA molecules or bind other cellular molecules, including but not limited to maltose binding protein (MBP), S-tag, Lex A DNA binding domain (DBD) fusions, GAL4 DNA binding domain fusions, and herpes simplex virus (HSV) BP16 protein fusions. Additional domains that may form part of a fusion protein comprising a CRISPR enzyme are described in US20110059502, incorporated herein by reference. In some embodiments, a tagged CRISPR enzyme is used to identify the location of a target sequence.

[0300] In some embodiments, a CRISPR enzyme 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 enzyme 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 enzyme, 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.

[0301] 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 Sambrook, Fritsch and Maniatis, MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (F. M. Ausubel, et al. eds., (1987)); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.): PCR 2: A PRACTICAL APPROACH (M.J. MacPherson, B.D. Hames and G.R. Taylor eds. (1995)), Harlow and Lane, eds. (1988) ANTIBODIES, A LABORATORY MANUAL, and ANIMAL CELL CULTURE (R.I. Freshney, ed. (1987)).

Modifying a target

[0302] In one aspect, the invention provides for methods of modifying a target polynucleotide in a eukaryotic cell, which may be *in vivo, ex vivo* or *in vitro*. In some embodiments, the method comprises sampling a cell or population of cells from a human or non-human animal, and modifying the cell or cells. Culturing may occur at any stage *ex vivo*. The cell or cells may even be re-introduced into the non-human animal or plant. For re-introduced cells it is particularly preferred that the cells are stem cells.

[0303] In some embodiments, the method comprises allowing a CRISPR complex to bind to the target polynucleotide to effect cleavage of said target polynucleotide thereby modifying the target polynucleotide, wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized or hybridizable to a target sequence within said target polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence.

[0304] In one aspect, the invention provides a method of modifying expression of a polynucleotide in a eukaryotic cell. In some embodiments, the method comprises allowing a CRISPR complex to bind to the polynucleotide such that said binding results in increased or decreased expression of said polynucleotide; wherein the CRISPR complex comprises a CRISPR enzyme complexed with a guide sequence hybridized or hybridizable to a target sequence within said polynucleotide, wherein said guide sequence is linked to a tracr mate sequence which in turn hybridizes to a tracr sequence. Similar considerations and conditions apply as above for methods of modifying a target polynucleotide. In fact, these sampling, culturing and re-introduction options apply across the aspects of the present invention.

[0305] Indeed, in any aspect of the invention, the CRISPR complex may comprise a CRISPR enzyme complexed with a guide sequence hybridized or hybridizable to a target sequence, wherein said guide sequence may be linked to a tracr mate sequence which in turn may hybridize to a tracr sequence. Similar considerations and conditions apply as above for methods of modifying a target polynucleotide.

Kits

[0306] In one aspect, the invention provides kits containing any one or more of the elements disclosed in the above methods and compositions. 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.

[0307] 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 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 allows to provide all elements of the systems of the invention.

EXAMPLES

[0308] The following examples are given for the purpose of illustrating various embodiments of the invention and are not meant to limit the present invention in any fashion. The present examples, along with the methods described herein are presently representative of preferred embodiments, are exemplary, and are not intended as limitations on the scope of the invention. Changes therein and other uses which are encompassed within the spirit of the invention as defined by the scope of the claims will occur to those skilled in the art.

Example 1: CRISPR-Cas9 system as a tool for editing disease-causing nucleotide repeat expansions in the human genome

[0309] The invention involves the development and application of the CRISPR-Cas9 system as a tool for editing disease-causing nucleotide repeat expansions in the human genome (Figure 1). Applicants provide evidence that the sequences, plasmids and/or viral vectors that Applicants have designed and tested facilitate genomic editing of nucleotide repeat sequences at a number of disease-linked genomic loci including those associated with CAG triplet repeat disorders (i.e. Polyglutamine diseases), Fragile X and Fragile X-associated tremor/ataxia syndrome (FXTAS). Moreover, Applicants describe the design and application of CRISPR-Cas9 to the mammalian brain using Adeno Associated Virus (AAV) as a vector.

[0310] Selection of target sequences. The goal of this approach is to use the CRISPR-Cas9 system to generate DNA double stranded breaks in genomic sequences that flank (i.e. immediately upstream and immediately downstream) the disease-associated nucleotide repeat sequences. Through the process of non-homologous recombination this should result in the CRISPR-Cas9-mediated genomic excision of the disease-causing nucleotide repeat sequence. A list of genes that can be targeted using this approach is provided in the below table. Target sequences are chosen based on three primary criteria: 1) the presence of a protospacer adjacent motif (PAM) sequence upstream and downstream from the nucleotide repeat, 2) a low predicted off-target potential of the 20-nucleotide target sequence associated with the identified PAM motifs (bioinformatics analysis based on algorithms developed in the Zhang laboratory) the proximity of the target sequence (within 100 nucleotides) to the nucleotide repeat expansion sequence.

Disease	Gene ID	Coding or Noncoding
Fragile X (FXS)	*FMR1*	non coding

Fragile X Tremor Ataxia (FXTAS)	*FMR1*	non coding
Unverricht-Lundborg disease (EPM1)	*CSTB*	non coding
Spinocerebellar ataxia type-12 (SCA12)	*PPP2R2B*	non coding
Amyotrophic Lateral Sclerosis (ALS)	*C9ORF72*	non coding
Fronto Temporal Dementia (FTD)	*C9ORF72*	non coding
Friedreich Ataxia	*FXN*	non coding
Myotonic Dystrophy type-1 (DM1)	*DMPK*	non coding
Myotonic Dystrophy type-2 (DM2)	*CNBP*	non coding
Spinocerebellar ataxia type-8 (SCA8)	*ATXN8OS*	non coding
Spinocerebellar ataxia type-10 (SCA10)	*ATXN10*	non coding
Spinocerebellar ataxia type-31 (SCA31)	*BEAN and TK2*	non coding
Oculopharyngeal muscular dystrophy (OPMD)	*PABPN1*	coding
Spinocerebellar ataxia type-1 (SCA1)	*ATXN1*	coding
Spinocerebellar ataxia type-2 (SCA2)	*ATXN2*	coding
Spinocerebellar ataxia type-3 (SCA3)	*ATXN3*	coding
Spinocerebellar ataxia type-6 (SCA6)	*CACNA1A*	coding
Spinocerebellar ataxia type-7 (SCA7)	*ATXN7*	coding
Spinocerebellar ataxia type-17 (SCA17)	*TBP*	coding
Dentatorubral-pallidoluysian atrophy (DRPLA)	*ATN1*	coding
Spinobulbar muscular atrophy (SBMA)	*AR*	coding
Huntington's disease like type-2 (HDL2)	*JPH3*	coding
Huntington's Disease (HD)	*HTT*	coding

[0311] Preliminary screen of CRISPR-Cas9 guide sequences in cultured human cell lines. To assess efficacy, target sequences that flank the nucleotide repeat region are individually cloned into the pX260 CRISPR-Cas9 vector system see protocols Zhang laboratory publications and described at www.genome-engineering.org. The pX260 vector system promotes the expression of the CRISPR target guide RNA under the control of a U6 RNA polymerase III promoter, a CRISPR transactivating RNA (tracrRNA) under the control of an H1 RNA polymerase III promoter and a nuclear-targeted, codon optimized *S. pyogenes Cas9* gene under the control of a Chicken beta-actin RNA polymerase II promoter. The screening process involves the transient expression of target guide sequences in human cell lines (i.e. HEK293, HeLa or HT1080) followed by genomic DNA extraction and PCR amplification to assess the CRISPR-Cas9-mediated excision of the nucleotide repeat sequence (Figure 2a). To confirm the excision and non-homologous repair of the CRISRP/Cas9 targeted site, the region around the intended target site is amplified by PCR, and the PCR amplicon cloned and sequenced. The results from these analyses are shown in Figures 2b-5.

[0312] ATXN1 editing: HT1080 cells were transfected with a plasmid encoding SpCas9 plus/minus plasmids encoding guide RNAs for editing trinucleotide CAG repeats in the *ATXN1*

coding sequence. The plasmids also permitted selection of transfected cells using puromycin. After five days of selection and growth, >90% of the cell population contained an edited *ATXN1* locus lacking the CAG repeat. Figure 9a shows that the genomic locus was shortened by approximately 150bp in cells which received the PS2 and PS5 plasmids (see Figure 2b). Furthermore, Figure 9b shows that steady-state levels of *ATXN1* transcripts were significantly reduced in cells receiving PS2+PS5 or PS1+PS5. Expression of SpCas9 alone did not affect *ATXN1* transcript levels.

[0313] Similar results for *ATXN1* have been achieved in neuroprogenitor cells, and also for editing the coding sequence of the *HTT* gene.

[0314] DMPK editing: A similar approach was used for editing CTG repeats in the *DMPK* non-coding 3'-UTR. Guide sequences upstream and downstream of the repeats were designed and expressed in HT1080 cells. Figure 10a shows that the genomic locus was shortened when the cells were co-transfected with plasmids encoding Cas9 and the guide sequences. The same result was also seen in primary skin fibroblasts obtained from a DM1 myotonic dystrophy patient who had around 550 CTG repeats. The fibroblasts were transfected with the plasmids via electroporation, and Figure 10c shows that the CTG expansion is effectively excised using the Cas9 CRISPR system. At least 1500nt of sequence were excised from the *DMPK* locus in the genome. DMPK protein expression was unaffected.

[0315] Design and use of AAVs to deliver the CRISPR-Cas9 system into mammalian tissues including but not limited to brain, lung, liver and muscle: To achieve efficient expression of the CRISPR-Cas9 system in target disease tissues Applicants engineered the required expression cassettes into AAVs. In one, the nuclear-targeted, codon-optimized *S. pyogenes Cas9* nuclease found in the pX260 plasmid was subcloned between the inverted terminal repeats (iTR) of an AAV serotype-2 shuttle plasmid using standard PCR cloning techniques. In the absence of a minimal promoter, the basal transcriptional activity present in the AAV serotype-2 iTRs is sufficient to drive the production of Cas9 nuclease (Figure 6). This system is also amenable to the expression of other members of the *Cas* family of bacterial nucleases. In addition, the incorporation of short, minimal synthetic or naturally occurring promoter sequences or RNA stabilizing elements could be considered if higher levels, tissue or cell specific Cas9 expression is desired.

[0316] The non-coding RNA components of the CRISR/Cas9 system are subcloned into a second AAV shuttle vector. CRISPR-Cas9 guide target sequences can be expressed individually or as part of an array format where multiple guide target sequences (gcRNA) are produced

intracellularly under the control of an RNA polymerase type III promoter (i.e. U6 or H1 RNA). The non-coding tracrRNA can be expressed from the same AAV shuttle vector using a pol III promoter. The gcRNA and the tracrRNA can also be expressed as a chimeric molecule, with the gcRNA and tracrRNA sequences fused as a single transcript (i.e. sgRNA) under the control of a RNA polymerase type III promoter. Multiple sgRNAs can also be expressed in array format using an RNA polymerase type III promoter (i.e. U6 or H1 RNA). The non-coding RNA CRISPR-Cas9 components described above are small enough that when cloned into AAV shuttle vectors sufficient space remains to include other elements such as reporter genes, antibiotic resistance genes or other sequences, which are cloned into the AAV shuttle plasmid using standard methods.

[0317] For expression of AAV CRISPR-Cas9 components in target tissues the iTR-SpCas9 AAV and the AAV.guide.taCRISPR are produced following previously described AAV purification protocols and simultaneously delivered in vivo. This can be achieved, for example, by combining both viruses at specified ratios in the same buffer prior to infusion. This results in the transduction of targeted tissue with both the Cas9 nuclease and the non-coding RNAs required to guide the nuclease to the targeted genomic locus. The diverse tissue tropism of AAV capsids, or synthetically modified AAV capsids, provides an opportunity to deliver AAV CRISPR-Cas9 components effectively to different tissues. Although a two-virus system is described, a three-virus system can also be used to deliver these components into target tissues. This could be desirable, for example, when several non-coding target guide RNAs need to be delivered at different times.

[0318] In an attempt to increase efficiency, alternative AAV vectors were designed. Figure 11a shows a system where an AAV9 virus encodes SaCas9 (with a N-terminal NLS, under the control of a CMV promoter) and synthetic guide RNA(s). If desired, this vector can be delivered with a second vector encoding, for example a transduction marker (*e.g.* EGFP, as shown in Figure 11a) or a template donor if nomologous recombination is desired. This experiment used the transgenic mouse HSA^{LR} model of DM1 which has an expanded CTG repeat in a human skeletal actin (hACTA1) transgene. Six-weeks old mice received an intrajugular infusion (systemic delivery, targeting primarily muscle and liver) of AAV9 coding either for an EGFP marker or for SaCas9 and guide sequences targeting the CTG expansion. Fluorescence in muscle biopsy (Figure 11b) confirms that the vector effectively targets muscle tissue. The PCR results for muscle tissue in Figure 11c show that CTG repeat region is excised in mice receiving SaCas9 and the sgRNA, but not in mice receiving the EGFP-coding vector or in mice receiving SaCas9

in combination with a control sgRNA whose sequence had been scrambled. The HSA^{LR} model shows nuclear foci due to retention of transcripts, but FISH analysis showed a reduction in nuclear foci in treated mice.

Example 2: Allele Specific CRISPR-Cas9-mediated inactivation of mutant alleles in dominantly inherited diseases

[0319] A number of single nucleotide polymorphisms (SNPs) exist in linkage disequilibrium with genetic mutations that underlie dominantly inherited diseases. The presence of these SNPs in mutant alleles can lead to the creation of de novo protospacer adjacent motifs (PAM) that can be targeted using the CRISPR-Cas9 system described in this application. Since Cas9 nucleases require the presence of a PAM sequence in order to mediate a double stranded DNA break, SNPs in linkage disequilibrium can be exploited to achieve allele-specific inactivation of mutant alleles in dominantly inherited diseases.

[0320] In Figure 8 Applicants describe the concept of CRISPR-Cas9-mediated allele-specific targeting. As shown in Figure 8a, the presence of a SNP (X to a G, where X is any nucleotide but G) in the beta allele results in the formation of a de novo 5'-NGG-3' PAM that is missing in the alpha allele. This one nucleotide difference can be used to design a guide sequence (gray Ns) to target the inactivation of the beta allele using the SpCas9. Applicants provide one example of this strategy in Figure 8b. However, this strategy can be done for any other PAM motif of any other candidate Cas9. A CAG repeat nucleotide expansion in the ATXN2 gene underlies the dominantly inherited neurodegenerative disease Spinocebellar ataxia type-2 (SCA2). As shown by Choudry et al, the rs695871 SNP (G to a C) exists in linkage disequilibrium with alleles containing an expanded CAG repeat. The SNP results in the formation of a 5'-NGG-3' PAM (underlined) that can be targeted using SpCas9 to preferentially inactivate the mutant ATXN2 allele carrying the CAG expansion, while maintaining normal activity from the second allele lacking the CAG expansion.

[0321] This concept was validated *in vitro* using synthetic expression constructs that contain the first 342 nucleotides (from the ATG start site) of the ATXN2 mRNA fused in-frame to the N-terminus of EGFP (G allele) or mCherry (C allele). Experimental results indicate that allele-specific targeting of the C allele using CRISPR-Cas9 results in a loss of mCherry (C allele) but not EGFP (G allele) expression in cultured cells (See Figures).

[0322] More in particular, synthetic fusion expression constructs containing the first 342bp (from the ATG start site) of the ATXN2 mRNA were fused to the N-terminus of EGFP (G allele) or mCherry (C allele). These constructs were transiently co-transfected into human HT1080

(fibrosarcoma cell line) cells with control guides or guides that specifically target the C allele (Allele specific guides=AS-sgRNA). EGFP and mCherry expression was analyzed at 72hrs post-transfection under a microscope using a UV lamp and filters for EGFP and mCherry excitation. As shown in the figure, mCherry expression was only inhibited in cells that received the AS-sgRNA constructs and not control. In contrast, EGFP expression was similar in cells that received control or C-allele specific (AS-sgRNA) guide RNA CRISPR/Cas9.

Example 3: Design of a Self-INactivating CRISPR-Cas9 system, e.g., to limit and/or prevent unnecessary long-term, chronic expression of the Cas nuclease gene—to control Cas nuclease expression.

[0323] The invention also provides a method for the self-inactivation of the CRISPR-Cas9 system as a means to limit the duration of its activity and/or expression in targeted cells. Figure 13 depicts one aspect of a Self-Inactivating CRISPR-Cas9 system, and Figure 14 depicts an exemplary self-inactivating CRISPR-Cas9 system for a chimeric tandem array transcript specific to the ATXN1 locus.

[0324] In principle, human genome editing via CRISPR-Cas9 requires, at most, two Cas9 molecules (targeting genomic sites on two different alleles). Thus, delivery of CRISRP/Cas9 systems that result in sustained cellular expression of the *Cas9* gene and/or its non-coding RNA components is unnecessary to successfully achieve editing of disease-causing mutations. Moreover, sustained *CRISPR-Cas9* activity could lead to undesirable off-target effects at unintended genomic sites, which over time, could be deleterious for the host cell, tissue or organism.

[0325] Applicants have engineered a Self-Inactivating CRISPR-Cas9 system (SIN-CC9) that relies on the use of a non-coding guide target sequences complementary to unique sequences present: i) within the promoter driving expression of the non-coding RNA elements, ii) within the promoter driving expression of the Cas9 gene, iii) within 100bp of the ATG translational start codon in the Cas9 coding sequence, iv) within the inverted terminal repeat of the AAV genome. The "self-inactivating" guide RNAs can be expressed singly or in array format to achieve inactivation of the CRISPR-Cas9 system. For example, a double stranded break near the ATG translational start codon in the Cas9 coding sequence will induce a frame shift in the Cas9 coding sequence causing a loss of protein expression. If expressed in array format, "self-inactivating" guide RNAs that target both promoters simultaneously will result in the excision of 400-800 nucleotides from within the AAV transgene/genome, effectively leading to its complete inactivation. These strategies are diagrammed in Figures 7 and 12.

[0326] Figure 12 provides an exemplary self-inactivating construct that comprises a cloning cassette optionally flanked by an inverted terminal repeat (iTR) on each of the 5' and 3' ends of the cassette, with the cassette further comprising a Pol III promoter driving the sgRNAs and a Pol II promoter driving Cas9. A "self-inactivating" sgRNA (whose target sites are shown as red lines) can be expressed singly or in tandem array format from 1 up to 4 different guide sequences processed from one chimeric Pol III transcript. The black brackets depict the expected DNA excision resulting from targeted double strand breaks when at least two different "self-inactivating" sgRNAs are expressed in tandem arrays. A wide selection of self-inactivating target sequences are available for use in a SaCas9 system, including, but not limited to, inactivating target sequences in Cas9, CMV, U6, modU6, ITR and the like. Pol III promoters may be selected from, but are not limited to, the U6 or H1 promoters. Pol II promoters may be selected from, but are not limited to, those proomoters provided throughout the description, including the ubiquitous and cell-type specific promoters.

[0327] To accomplish CRISPR-Cas9 inactivation, while also achieving editing of the targeted genomic site, the CRISPR-Cas9 guide RNAs that target a desired genomic locus and the "self-inactivating" guide RNA(s) are co-delivered/co-expressed in the same cell/target tissue following procedures described above. The basic concept is presented in diagram form (Figure 7a). This approach results in editing of the intended genomic site followed by the inactivation of the *Cas9* nuclease gene within 48 hours (Figure 7b). If necessary, the addition of non-targeting nucleotides to the 5' end of the "self-inactivating" guide RNA can be used to delay its processing and/or modify its efficiency as a means of ensuring editing at the targeted genomic locus prior to CRISPR-Cas9 shutdown.

[0328] Applicants have put the SIN-CC9 concept into effect with both SpCas9 and SaCas9, and targeting sequences both at/near the Cas9 ATG and within the ITR of an AAV vector (See, e.g., Figure 12). Furthermore, this system is believed to be sufficiently rigorous to "inactivate" any other recombinant AAV-based gene therapy system. Designing and adding this option for self-inactivation of such systems is of particular interest for use in gene therapy systems that might lead to tumor formation due to unanticipated integration of the AAV sequences in the human genome.

Example 4: U6-driven tandem guide RNAs deliver two functional sgRNAs; Optimization of tandem sgRNA scaffold architecture; Processing of tandem sgRNAS into individual subunits occurs; Targeting Cas9 against itself (Self-INactivating; SIN).

[0329] Using pooled delivery of independently transcribed sgRNAs is stochastic in nature and may be less reproducible than a single vector system; for instance, in applications where target saturation may not be desired or achievable. Many endogenous microbial CRISPR systems naturally occur as a single-promoter driven array of direct repeats interspaced by protospacers, which are transcribed as a single transcript prior to their processing into individual mature crRNAs. However, given that the chimeric sgRNA system works much more efficiently than the native crRNA:tracrRNA duplex, Applicants sought to develop a system by which a single promoter may drive the expression of multiple sgRNAs arranged in tandem, similar to the native microbial CRISPR loci. Without wishing to be bound by any particular theory, structurally stable sgRNA scaffolds may be more likely to fold into independent, functionally active units when multiple units are transcribed together in the same transcript. Applicants inserted an 8-nt linker between tandem adjacent sgRNAs; for each the invariant sgRNA scaffold (non-guide region), Applicants used either pairs of original sp85 sgRNA or scaffolds with stabilized distal hairpins. Strikingly, Applicants observed that when the tandem synthetic guide RNAs (tsgRNAs) targeted closely approximated genomic loci previously shown to induce indels with Cas9 nickase, the stabilized scaffolds were able to induce indels at frequencies similar to those induced by co-transfected individual sgRNAs. Moreover, when paired with wild-type Cas9 nuclease, tsgRNAs were similarly able to induce genomic microdeletions in the human EMX1 locus at levels comparable to multiplexed, individual sgRNAs. Having shown that sgRNAs transcribed in tandem are able to simultaneously target two genomic loci, Applicants next sought to determine the optimal linker for connecting the adjacent guide-scaffolds. Applicants designed tsgRNAs using linker sequences of varying lengths in a genomic microdeletion assay with two sgRNAs. Given that endogenous individual protospacers are separated by 36-nt long direct repeat sequences, and also tested linkers that encoded for either half of a direct repeat or a full-length direct repeat. Interestingly, there was not a strong correlation between linker sequence length and the efficiency of genome modification, even in cases where there was no linker separating the distal end of the sgRNA from the guide sequence of the second. However, it appeared that inclusion of direct repeat sequences may adversely affect activity while there is a modest preference towards an 8-16, e.g., 10-14, for instance 12-nt linker length for cleavage efficiency. To address whether co-transcribed tandem sgRNAs (transcribing multiple sgRNAs under the same promoter) are processed to individual guide-scaffold units, Applicants designed tandem sgRNAs that carried the same guide in either the first or second position (Figure 15A). Subsequent Northern blot analyses of transfected cells showed three distinct RNA species,

corresponding to a 200+ nt (likely unprocessed tandem RNA transcript), a ~140 nt transcript (consistent with premature transcriptional termination signaled by the poly-U tract in the second scaffold), and a ~100 nt fully processed sgRNA (Figure 15B). When the target spacer is in the first position in the tsgRNA, Applicants observed abundant fully processed sgRNA of the same size as individually U6-transcribed sgRNAs. When placed in the second position, there were only trace amounts of fully processed sgRNA present. Consistent with this, reversing spacer order in microdeletion assays could significantly alter the efficiency of genomic modification. When testing other pairs of sgRNAs targeting different genomic loci, the same guide sequence typically had better activity when placed in the first rather than the second position (Figure 15C). These observations suggest that while most spacers are compatible with a single guide transcript, the sequence of the second spacer may be more likely to influence activity of the second sgRNA in the context of a tandem sgRNA. Pairing of sequence-divergent scaffolds results in better second spacer activity: To optimize the activity of the second spacer, Applicants devised an assay for assessing its activity by fluorescence cytometry. By targeting the second guide against Cas9 itself in a plasmid expressing Cas9-2A-GFP, Applicants assessed indel activity by measuring the fluorescence fraction and intensity of transfected cells (Figure 16A). Transfecting cells with single sgRNAs targeting Cas9 or co-delivering Cas9-targeting sgRNA with another sgRNA significantly reduced the mean fluorescence intensity (MFI) of the Cas9-2A-GFP-transfected GFP-positive fraction, whereas cells transfected with Cas9-2A-EGFP and a non-Cas9-targeting sgRNA maintained high MFI (Figure 16B). Given that each sgRNA scaffold needs to fold into a stable secondary structure, without wishing to be bound by any particular theory, a potential reason for the decreased activity of the second spacer may be due to secondary structure interactions not within a single but between the two sgRNA scaffolds. Without wishing to be bound by any particular theory, the use of divergent, minimally homologous sgRNA scaffolds that are less likely to base-pair with each other could reduce interactions between the pair and aid individual folding. Applicants designed a set of twelve distinct sgRNA scaffolds, each with the first guide targeting GRIN2B and the second targeting Cas9, and performed a pair-wise comparison of all scaffold combinations. Subsequent flow-cytometric analyses identified five potential candidate sgRNA scaffolds that significantly reduced both the MFI of the GFP-positive fraction as well as the overall percentage of GFP-positive cells; the levels of reductions are similar to those obtained by transfecting singly transcribed Cas9-targeting sgRNA (Figure 16C). Consistent with the notion that inter-scaffold interactions may be disrupting proper sgRNA folding and processing, most of the five scaffolds

showed relatively poor activity when transcribed in tandem with highly homologous sgRNAs. Indeed, sequence alignment analysis of the twelve scaffolds showed that the pairs of tandem scaffolds that showed the highest activity had the greatest sequence divergence between the two sgRNAs (Figure 17). Tandem-arrayed sgRNAs represents a potentially useful approach for co-delivery of two sgRNAs in a single RNA transcript. While some guide sequences appear to function well in the second position, optimization of the sgRNA architecture to maximize inter-scaffold sequence divergence and improve structural stability can aid processing and activity of tandem sgRNAs. And, sgRNAs can be designed to so that the system is SIN.

Example 5: CRISPR/Cas9 in vivo genome editing efficacy and therapeutic benefit in a polyglutamine disease mouse model.

[0330] This Example is to be read in conjunction with Figure 18, wherein Applicants provide evidence of *in vivo* CRISPR/Cas9 genome editing efficacy and therapeutic benefit in a polyglutamine disease mouse model. The Example demonstrates gene editing of an expanded CAG trinucleotide repeat using AAV-delivered CRISPR/Cas9 (AAV-CC9) in a mouse model of Spinocerebellar ataxia type-1 (B05 transgenic mouse line) using the small Cas9 nuclease from *Staphylococcus aureus* (SaCas9). To achieve this, Applicants developed a single AAV vector that expresses a CMV promoter-driven, HA-tagged SaCas9 and contains a U6 promoter cassette that drives expression of a multimeric sgRNA transcript (see Figure 18A). When expressed in cells, this non-coding RNA transcript is processed to release two individual sgRNAs that complex with SaCas9 to mediate gene editing. CTRL-AAV-CC9 and ATXN1-AAV-CC9 vectors were generated carrying control (targeting the *Renilla reniformis* Luciferase gene) and anti-*ATXN1* sgRNAs and delivered into the cerebellum of adult SCA1/BO5 mice using a stereotactic apparatus. SCA1/BO5 transgenic mice carry numerous copies (>20) of a mutant human *ATXN1* transgene that contains 84 CAG nucleotide repeats and is almost exclusively expressed in cerebellar Purkinje cells (Burright EN, Clark HB, Servadio A, Matilla T, Feddersen RM, Yunis WS, Duvick LA, Zoghbi HY, Orr HT. SCA1 transgenic mice: a model for neurodegeneration caused by an expanded CAG trinucleotide repeat. Cell. 1995 Sep 22;82(6):937-48. PubMed PMID: 7553854). As shown in Figure 18B, expression of HA-SaCas9 persists in the nucleus of SCA1/BO5 Purkinje cells (red signal) at16 weeks post *ATXN1*-AAV-CC9 delivery. A faster migrating band can be detected following PCR amplification across the *ATXN1* CAG expansion, confirming excision/editing of the mutant *ATXN1* CAG expanded transgene in SCA1/BO5 cerebellar tissue (Figure 18C). A semi-quantitative analysis of the ratio between edited and unedited PCR products suggests a 41% gene editing efficiency. However, a

more significant loss in *ATXN1* transgene expression was observed (~80% loss in mRNA by qPCR and ~70% loss in protein as detected by Western blot) in SCA1/BO5 mice expressing the *ATXN1*-AAV-CC9 vector (Figures 18D, 18E). The fact that the *ATXN1* transgene is primarily expressed in Purkinje cells (highly transduced by AAV) while the *ATXN1* transgene genomic sequence is present in the genomes of all cells (including Purkinje cells) explains the apparent discrepancy in gene editing efficiency between the genome-based PCR and mRNA-based qPCR assays. Finally, mice injected with *ATXN1*-AAV-CC9 performed better in the rotarod apparatus when compared to non-injected or control-injected SCA1/BO5 mice. This impact in phenotypic progression provides the extent of mutant *ATXN1* transgene inactivation observed in injected SCA1/BO5 mice; demonstrates the efficacy of the invention; and provides a surprising and superior result as such a result was not previously achievable through other genome editing techniques.

[0331] References

Banker G, Goslin K. Developments in neuronal cell culture. Nature. 1988 Nov 10;336(6195):185-6.

Bedell, V.M. et al. In vivo genome editing using a high-efficiency TALEN system. Nature 491, 114-U133 (2012).

Bhaya, D., Davison, M. & Barrangou, R. CRISPR-Cas systems in bacteria and archaea: versatile small RNAs for adaptive defense and regulation. Annu Rev Genet 45, 273-297 (2011).

Bobis-Wozowicz, S., Osiak, A., Rahman, S.H. & Cathomen, T. Targeted genome editing in pluripotent stem cells using zinc-finger nucleases. Methods 53, 339-346 (2011).

Boch, J. et al. Breaking the code of DNA binding specificity of TAL-type III effectors. Science 326, 1509-1512 (2009).

Bogenhagen, D.F. & Brown, D.D. Nucleotide sequences in Xenopus 5S DNA required for transcription termination. Cell 24, 261-270 (1981).

Bultmann, S. et al. Targeted transcriptional activation of silent oct4 pluripotency gene by combining designer TALEs and inhibition of epigenetic modifiers. Nucleic Acids Res 40, 5368-5377 (2012).

Carlson, D.F. et al. Efficient TALEN-mediated gene knockout in livestock. Proc Natl Acad Sci U S A 109, 17382-17387 (2012).

Chen, F.Q. et al. High-frequency genome editing using ssDNA oligonucleotides with zinc-finger nucleases. Nat Methods 8, 753-U796 (2011).

Chen, Y.H., M. Chang, and B.L. Davidson, Molecular signatures of disease brain endothelia provide new sites for CNS-directed enzyme therapy. Nat Med, 2009. 15(10): p. 1215-8.

Chen, Y.H., et al., Sialic acid deposition impairs the utility of AAV9, but not peptide-modified AAVs for brain gene therapy in a mouse model of lysosomal storage disease. Mol Ther, 2012. 20(7): p. 1393-9.

Cho, S.W., Kim, S., Kim, J.M. & Kim, J.S. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. Nat Biotechnol 31, 230-232 (2013).

Christian, M. et al. Targeting DNA double-strand breaks with TAL effector nucleases. Genetics 186, 757-761 (2010).

Cong, L. et al. Multiplex genome engineering using CRISPR-Cas systems. Science 339, 819-823 (2013).

Davidson, B.L., et al., Recombinant adeno-associated virus type 2, 4, and 5 vectors: transduction of variant cell types and regions in the mammalian central nervous system. Proc Natl Acad Sci U S A, 2000. 97(7): p. 3428-32.

Davidson, B.L. and J.A. Chiorini, Recombinant adeno-associated viral vector types 4 and 5. Preparation and application for CNS gene transfer. Methods Mol Med, 2003. 76: p. 269-85.

Deltcheva, E. et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602-607 (2011).

Deveau, H., Garneau, J.E. & Moineau, S. CRISPR-Cas system and its role in phage-bacteria interactions. Annu Rev Microbiol 64, 475-493 (2010).

Ding, Q. et al. A TALEN genome-editing system for generating human stem cell-based disease models. Cell Stem Cell 12, 238-251 (2013).

Garneau, J.E. et al. The CRISPR-Cas bacterial immune system cleaves bacteriophage and plasmid DNA. Nature 468, 67-71 (2010).

Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc Natl Acad Sci U S A 109, E2579-2586 (2012).

Geurts, A.M. et al. Knockout Rats via Embryo Microinjection of Zinc-Finger Nucleases. Science 325, 433-433 (2009).

Gray SJ, Foti SB, Schwartz JW, Bachaboina L, Taylor-Blake B, Coleman J, Ehlers MD, Zylka MJ, McCown TJ, Samulski RJ. Optimizing promoters for recombinant adeno-associated

virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors. Hum Gene Ther. 2011 Sep;22(9):1143-53. doi: 10.1089/hum.2010.245.

Guschin, D.Y. et al. A rapid and general assay for monitoring endogenous gene modification. Methods Mol Biol 649, 247-256 (2010).

Hasty, P., Rivera-Perez, J. & Bradley, A. The length of homology required for gene targeting in embryonic stem cells. Mol Cell Biol 11, 5586-5591 (1991).

Horvath, P. & Barrangou, R. CRISPR-Cas, the immune system of bacteria and archaea. Science 327, 167-170 (2010).

Hsu, P.D. & Zhang, F. Dissecting neural function using targeted genome engineering technologies. ACS Chem Neurosci 3, 603-610 (2012).

Hughes, S.M., et al., Viral-mediated gene transfer to mouse primary neural progenitor cells. Mol Ther, 2002. 5(1): p. 16-24.

Hwang, W.Y. et al. Efficient genome editing in zebrafish using a CRISPR-Cas system. Nat Biotechnol 31, 227-229 (2013).

Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L.A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Nat Biotechnol 31, 233-239 (2013).

Jinek, M. et al. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. Science 337, 816-821 (2012).

Jinek, M. et al. RNA-programmed genome editing in human cells. eLife 2, e00471 (2013).

Kaplitt, M.G., et al., Safety and tolerability of gene therapy with an adeno-associated virus (AAV) borne GAD gene for Parkinson's disease: an open label, phase I trial. Lancet. 2007 Jun 23;369(9579):2097-105.

Levitt N. Briggs D. Gil A. Proudfoot N.J. Definition of an efficient synthetic poly(A) site. Genes Dev. 1989;3:1019–1025.

Liu D, Fischer I. Two alternative promoters direct neuron-specific expression of the rat microtubule-associated protein 1B gene. J Neurosci. 1996 Aug 15;16(16):5026-36.

Liu, G., et al., Adeno-associated virus type 4 (AAV4) targets ependyma and astrocytes in the subventricular zone and RMS. Gene Ther, 2005. 12(20): p. 1503-8.

Lopes, V.S., etc al., Retinal gene therapy with a large MYO7A cDNA using adeno-assocaited virus. Gene Ther, 2013 Jan 24. doi: 10.1038/gt 2013.3.[Epub ahead of print]

Lotery, A.J., et al., Adeno-associated virus type 5: transduction efficiency and cell-type specificity in the primate retina. Hum Gene Ther, 2003. 14(17): p. 1663-71.

Mahfouz, M.M. et al. De novo-engineered transcription activator-like effector (TALE) hybrid nuclease with novel DNA binding specificity creates double-strand breaks. Proc Natl Acad Sci U S A 108, 2623-2628 (2011).

Makarova, K.S. et al. Evolution and classification of the CRISPR-Cas systems. Nat Rev Microbiol 9, 467-477 (2011).

Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823-826 (2013).

McClure C, Cole KL, Wulff P, Klugmann M, Murray AJ. Production and titering of recombinant adeno-associated viral vectors. J Vis Exp. 2011 Nov 27;(57):e3348. doi: 10.3791/3348.

Michaelis, L.M., Maud "Die kinetik der invertinwirkung.". Biochem. z (1913).

Miller, J.C. et al. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat Biotechnol 25, 778-785 (2007).

Miller, J.C. et al. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol 29, 143-148 (2011).

Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. Science 326, 1501 (2009).Porteus, M.H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. Science 300, 763 (2003).

Mussolino, C. et al. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic acids research 39, 9283-9293 (2011).

Nathwani, A.C., et al., Adenovirus-associated virus vector-mediated gene transfer in hemophilia B. N Engl J Med. 2011 Dec 22;365(25):2357-65. doi: 10.1056/NEJMoa1108046. Epub 2011 Dec 10.

Oliveira, T.Y. et al. Translocation capture sequencing: a method for high throughput mapping of chromosomal rearrangements. J Immunol Methods 375, 176-181 (2012).

Perez, E.E. et al. Establishment of HIV-1 resistance in CD4(+) T cells by genome editing using zinc-finger nucleases. Nat Biotechnol 26, 808-816 (2008).

Qi, L.S. et al. Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell 152, 1173-1183 (2013).

Ran, F.A., et al., Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. Cell, 2013. 154(6): p. 1380-9.

Ran, F.A., et al., Genome engineering using the CRISPR-Cas9 system. Nat Protoc, 2013. 8(11): p. 2281-308.

REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991)

Reyon, D. et al. FLASH assembly of TALENs for high-throughput genome editing. Nat Biotechnol 30, 460-465 (2012).

Rodriguez-Lebron, E., et al., Allele-specific RNAi mitigates phenotypic progression in a transgenic model of Alzheimer's disease. Mol Ther, 2009. 17(9): p. 1563-73.

Saleh-Gohari, N. & Helleday, T. Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells. Nucleic Acids Res 32, 3683-3688 (2004).

Sander, J.D. et al. Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). Nat Methods 8, 67-69 (2011).

Sanjana, N.E. et al. A transcription activator-like effector toolbox for genome engineering. Nat Protoc 7, 171-192 (2012).

Sapranauskas, R. et al. The Streptococcus thermophilus CRISPR-Cas system provides immunity in Escherichia coli. Nucleic Acids Res 39, 9275-9282 (2011).

Shen, B. et al. Generation of gene-modified mice via Cas9/RNA-mediated gene targeting. Cell Res 23, 720-723 (2013).

Smithies, O., Gregg, R.G., Boggs, S.S., Koralewski, M.A. & Kucherlapati, R.S. Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. Nature 317, 230-234 (1985).

Soldner, F. et al. Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. Cell 146, 318-331 (2011).

Takasu, Y. et al. Targeted mutagenesis in the silkworm Bombyx mori using zinc finger nuclease mRNA injection. Insect Biochem Molec 40, 759-765 (2010).

Tangri S, et al., Rationally engineered therapeutic proteins with reduced immunogenicity, J Immunol. 2005 Mar 15;174(6):3187-96.

Thomas, K.R., Folger, K.R. & Capecchi, M.R. High frequency targeting of genes to specific sites in the mammalian genome. Cell 44, 419-428 (1986).

Tuschl, T. Expanding small RNA interference. Nat Biotechnol 20, 446-448 (2002).

Urabe, M., C. Ding, and R.M. Kotin, Insect cells as a factory to produce adeno-associated virus type 2 vectors. Hum Gene Ther, 2002. 13(16): p. 1935-43.

Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. & Gregory, P.D. Genome editing with engineered zinc finger nucleases. Nat Rev Genet 11, 636-646 (2010).

Valton, J. et al. Overcoming transcription activator-like effector (TALE) DNA binding domain sensitivity to cytosine methylation. J Biol Chem 287, 38427-38432 (2012).

Wang, H. et al. One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR-Cas-Mediated Genome Engineering. Cell 153, 910-918 (2013).

Watanabe, T. et al. Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. Nat Commun 3 (2012).

Wilson, E.B. Probable inference, the law of succession, and statistical inference. J Am Stat Assoc 22, 209-212 (1927).

Wood, A.J. et al. Targeted genome editing across species using ZFNs and TALENs. Science 333, 307 (2011).

Wu, S., Ying, G.X., Wu, Q. & Capecchi, M.R. A protocol for constructing gene targeting vectors: generating knockout mice for the cadherin family and beyond. Nat Protoc 3, 1056-1076 (2008).

Yang, G.S., et al., Virus-mediated transduction of murine retina with adeno-associated virus: effects of viral capsid and genome size. J Virol, 2002. 76(15): p. 7651-60.

Zabner, J., et al., Adeno-associated virus type 5 (AAV5) but not AAV2 binds to the apical surfaces of airway epithelia and facilitates gene transfer. J Virol, 2000. 74(8): p. 3852-8.

Zhang, F. et al. Efficient construction of sequence-specific TAL effectors for modulating mammalian transcription. Nat Biotechnol 29, 149-153 (2011).

* * *

[0332] While preferred embodiments of the present invention have been shown and described herein, it will be obvious to those skilled in the art that such embodiments are provided by way of example only. Numerous variations, changes, and substitutions will now occur to those skilled in the art without departing from the invention. It should be understood that various alternatives to the embodiments of the invention described herein may be employed in practicing the invention.

WHAT IS CLAIMED IS:

1. A non-naturally occurring or engineered composition for use in a cell having a defective nucleotide element or trinucleotide repeat or other nucleotide repeat element or nucleotide expansion, the comprising:

A.

I. a first regulatory element operably linked to a CRISPR-Cas system RNA polynucleotide sequence, wherein the polynucleotide sequence comprises:

(a) at least one first guide sequence capable of hybridizing to a target DNA within the cell,

(b) at least one tracr mate sequence, and

(c) at least one tracr sequence, and

wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

II. a second regulatory element operably linked to a polynucleotide sequence encoding a CRISPR enzyme,

wherein parts A.I and A.II comprise a CRISPR-Cas system, and wherein, said composition when transcribed comprises

a CRISPR complex comprising the CRISPR enzyme complexed with (1) the guide sequence that is hybridized or is hybridizable to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence,

wherein the guide sequence directs sequence-specific binding of a CRISPR complex to the target DNA, and mediates impact or repair of the defect;

or,

B.

I. a CRISPR-Cas system RNA polynucleotide sequence, wherein the polynucleotide sequence comprises:

(a) at least one guide sequence capable of hybridizing to a target sequence in a eukaryotic cell,

(b) at least one tracr mate sequence, and

(c) at least one tracr sequence, and

wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

II. a CRISPR enzyme,

wherein parts B.I and B.II comprise the CRISPR complex.

2. The composition of claim 1, wherein the cell is a eukaryotic cell.

3. The composition of claim 2, wherein the CRISPR-Cas system is codon optimized.

4. The composition of claim 2, wherein part A.II includes coding for one or more nuclear localization signals (NLSs); or part B.II includes one or more NLSs.

5. The composition according to any one of the preceding claims, wherein part A.I is encoded by a first viral vector and part A.II is encoded by a second viral vector.

6. The composition of claim 5, wherein the first and second viral vectors are lentiviral vectors or recombinant AAV.

7. The composition of claim 6, wherein the recombinant AAV genome comprises inverted terminal repeats (iTRs).

8. The composition of claim 7, wherein expression of the CRISPR enzyme is driven by the inverted terminal repeat (iTR) in the AAV genome.

9. The composition of any one of the preceding claims, wherein the first regulatory element is a RNA polymerase type III promoter and the second regulatory element is a RNA polymerase type III promoter.

10. The composition of any one of the preceding claims, wherein the first regulatory element is a U6 promoter or a H1 promoter.

11. The composition of any one of the preceding claims, wherein the second regulatory element is a ubiquitous expression promoter or a cell-type specific promoter.

12. The composition of any one of the preceding claims 1, wherein there is a selection marker comprising a FLAG-tag.

13. The composition of claim 4, wherein the CRISPR enzyme comprises a C-terminal NLS and an N-terminal NLS.

14. The composition of any one of the preceding claims is delivered via injection.

15. The composition of any one of the preceding claims wherein the composition or a part thereof is delivered via a liposome, a nanoparticle, an exosome or a microvesicle.

16. The composition of any of the preceding claims, wherein the guide sequence directs sequence-specific binding of the CRISPR complex to the target DNA sequence and alters expression of a genomic locus in the cell.

17. The composition of any of the preceding claims, wherein the CRISPR complex mediates binding to or a double or single stranded DNA break, and there can optionally be insertion of DNA, whereby there is editing of a genomic locus in the cell.

18. The composition of any of the preceding claims, wherein the CRISPR-Cas system is a multiplexed CRISPR enzyme system further comprising multiple chimeras and/or multiple multiguide sequences and a single tracr sequence.

19. The composition according any of the preceding claims, wherein the CRISPR enzyme is a nickase.

20. The composition of any of the preceding claims, wherein the CRISPR-Cas system is a multiplexed CRISPR enzyme system to minimize off-target activity.

21. The composition according to any of the preceding claims, wherein the CRISPR enzyme comprises one or more mutations.

22. The composition according to claim 21, wherein the CRISPR enzyme comprises one or more mutations selected from D10A, E762A, H840A, N854A, N863A or D986A.

23. The composition according to claim 21 wherein the one or more mutations is in a RuvC1 domain of the CRISPR enzyme.

24. The composition according to any preceding claim, wherein the CRISPR enzyme further comprises a functional domain.

25. The composition according to any of the preceding claims, wherein the CRISPR complex mediates genome engineering that includes: modifying a target polynucleotide or expression thereof, knocking out a gene, amplifying or increasing or decreasing expression of a polynucleotide or gene, or repairing a mutation, or editing by inserting a polynucleotide.

26. The composition of any one of the preceding claims, wherein the CRISPR enzyme is a Cas9

27. The composition of any of the preceding claims wherein the CRISPR complex mediates at least one double stranded DNA break thereby causing editing of the target DNA.

28. The composition of any one of the preceding claims wherein the cell is a mammalian brain or central nervous tissue cell.

29. The composition of any of the preceding claims wherein the nucleotide repeat element is selected from one or more of: a trinucleotide repeat comprising CTG, CAG, CGG, CCG, GAA, or TTC; a tetranucleotide repeat comprising CCTG, a pentanucleotide repeat comprising ATTCT or AGAAT; a hexanucleotide repeat comprising GGGGCC; and a dodecanucleotide repeat comprising CCCCGCCCCGCG or CGCGGGGCGGGG.

30. The composition of any one of the preceding claims, wherein the defect gives rise to a condition selected from one or more of: a Fragile X (FXS); Fragile X Tremor Ataxia (FXTAS); Unverricht-Lundborg disease (EPM1); Spinocerebellar ataxia type-12 (SCA12); Amyotrophic

Lateral Sclerosis (ALS); Fronto Temporal Dementia (FTD); Friedreich Ataxia; Myotonic Dystrophy type-1 (DM1); Myotonic Dystrophy type-2 (DM2); Spinocerebellar ataxia type-8 (SCA8); Spinocerebellar ataxia type-10 (SCA10); Spinocerebellar ataxia type-31 (SCA31); Oculopharyngeal muscular dystrophy (OPMD); Spinocerebellar ataxia type-1 (SCA1); Spinocerebellar ataxia type-2 (SCA2); Spinocerebellar ataxia type-3 (SCA3); Spinocerebellar ataxia type-6 (SCA6); Spinocerebellar ataxia type-7 (SCA7); Spinocerebellar ataxia type-17 (SCA17); Dentatorubral-pallidoluysian atrophy (DRPLA); Spinobulbar muscular atrophy (SBMA); Huntington's disease like type-2 (HDL2) and Huntington's Disease (HD).

31. A method of treating or inhibiting a condition in a cell having a defective nucleotide element or trinucleotide repeat or other nucleotide repeat element or nucleotide expansion, comprising delivering the non-naturally occurring or engineered composition of any one of claims 1-30.

32. Use of the composition of any one of claims 1-30 to treat a disease or disorder.

33. Use of claim 32 wherein the disease or disorder comprises a brain disease or disorder or a central nervous system disease or disorder.

34. Use of a composition as defined in any of claims 1-30 in the manufacture of a medicament for *ex vivo* gene or genome editing or for use in a method of modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest or in a method of treating or inhibiting a condition.

35. Use of claim 34 wherein the condition comprises a brain disease or disorder or a central nervous system disease or disorder.

36. The method, use or composition of any of the preceding claims, wherein the CRISPR-Cas system RNA is a chimeric RNA (chiRNA).

37. The method, use or composition of any preceding claim, further comprising at least one second guide sequence capable of hybridizing to an RNA sequence of the CRISPR-Cas system or a nucleic acid molecule for expression of a component of the CRISPR-Cas complex, to diminish or eliminate functional expression of the system or complex, whereby the system or complex is Self-Inactivating.

38. The method, use or composition of claim 38 wherein the second guide sequence is capable of hybridizing to a nucleic acid molecule for expression of the CRISPR enzyme.