(12) STANDARD PATENT (11) Application No. **AU 2014281026 B2**
(19) AUSTRALIAN PATENT OFFICE

(54) Title
Delivery, engineering and optimization of tandem guide systems, methods and compositions for sequence manipulation

(51) International Patent Classification(s)
C12N 9/22 (2006.01) ***C12N 15/63*** (2006.01)
C12N 15/10 (2006.01)

(21) Application No: **2014281026** (22) Date of Filing: **2014.06.10**

(87) WIPO No: **WO14/204724**

(30) Priority Data

(31) Number	(32) Date	(33) Country
61/979,942	**2014.04.15**	**US**
61/862,468	**2013.08.05**	**US**
61/915,407	**2013.12.12**	**US**
61/847,537	**2013.07.17**	**US**
61/871,301	**2013.08.28**	**US**
61/862,355	**2013.08.05**	**US**
61/836,123	**2013.06.17**	**US**

(43) Publication Date: **2014.12.24**
(44) Accepted Journal Date: **2020.05.28**

(71) Applicant(s)
President and Fellows of Harvard College;Massachusetts Institute of Technology;The Broad Institute, Inc.

(72) Inventor(s)
Zhang, Feng;Lin, Chie-yu;Ran, Fei

(74) Agent / Attorney
Phillips Ormonde Fitzpatrick, PO Box 323, Collins Street West, VIC, 8007, AU

(56) Related Art
WO 2013/176772 A1

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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
24 December 2014 (24.12.2014)

WIPO | PCT

(10) International Publication Number
WO 2014/204724 A1

(51) International Patent Classification:
C12N 15/63 (2006.01) *C12N 15/10* (2006.01)
C12N 9/22 (2006.01)

(21) International Application Number:
PCT/US2014/041800

(22) International Filing Date:
10 June 2014 (10.06.2014)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

61/836,123	17 June 2013 (17.06.2013)	US
61/847,537	17 July 2013 (17.07.2013)	US
61/862,355	5 August 2013 (05.08.2013)	US
61/862,468	5 August 2013 (05.08.2013)	US
61/871,301	28 August 2013 (28.08.2013)	US
61/915,407	12 December 2013 (12.12.2013)	US
61/979,942	15 April 2014 (15.04.2014)	US

(71) Applicants: **THE BROAD INSTITUTE INC.** [US/US]; 7 Cambridge Center, Cambridge, MA 02142 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Ave., Cambridge, MA 02142 (US).

(72) Inventors; and

(71) Applicants : **ZHANG, Feng** [US/US]; 100 Pacific Street, Apt. 11, Cambridge, MA 02139 (US). **LIN, Chie-yu** [US/US]; C/o Massachusetts Institute Of Technology, 77 Massachusetts Ave., Cambridge, MA 02142 (US). **RAN, Fei** [US/US]; 129 St. Botolph Street, Apt. 2, Boston, MA 02115 (US).

(74) Agents: **RAJAN, Lakshmi** et al.; Vedder Price P.C., 1633 Broadway, New York, NY 10019 (US).

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

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

Published:

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

(54) Title: DELIVERY, ENGINEERING AND OPTIMIZATION OF TANDEM GUIDE SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION

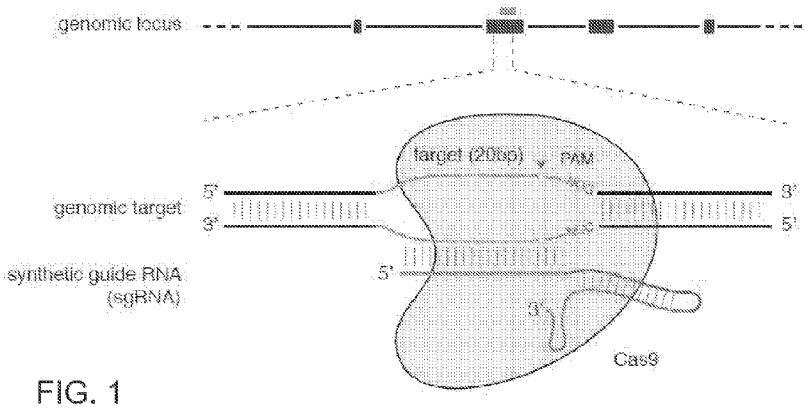


FIG. 1

(57) Abstract: The invention provides for delivery, engineering and optimization of systems, methods, and compositions for manipulation of sequences and/or activities of target sequences. Provided are vectors and vector systems, some of which encode one or more components of a CRISPR complex, as well as methods for the design and use of such vectors. Also provided are methods of directing CRISPR complex formation in prokaryotic and eukaryotic cells to ensure enhanced specificity for target recognition and avoidance of toxicity.

WO 2014/204724 A1

DELIVERY, ENGINEERING AND OPTIMIZATION OF TANDEM GUIDE SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION

RELATED APPLICATIONS AND INCORPORATION BY REFERENCE

[0001] Priority is claimed from US Provisional applications: 61/836,123, filed June 17, 2013; 61/847,537, filed July 17, 2013; 61/862,355, filed August 5, 2013; 61/862,468, filed August 5, 2013; 61/871,301, filed August 28, 2013; 61/915,407, filed December 12, 2013; and 61/979,942, filed April 15, 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 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 Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) and components thereof.

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 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 and optimization of these genome engineering tools, which are aspects of the claimed invention.

[0007] There exists a pressing need for alternative and robust systems and techniques for sequence targeting with a wide array of applications. Aspects of this invention address this need and provides 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.

[0008] 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 utilities 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 or genome editing, gene therapy, drug discovery, 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.

[0009] Aspects of the invention relate to Cas9 enzymes having improved target 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 advanced as there is less coding therefor in the delivery vector than for wild-type Cas9, and/or generating chimeric Cas9 enzymes.

[0010] 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.

[0011] 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). Further mutations have been characterized. In one aspect of the invention, the transcriptional activation domain may be VP64. Other aspects of the invention relate to the mutated Cas 9 enzyme being fused to domains which include but are not limited to a transcriptional repressor, a recombinase, a transposase, a histone remodeler, a DNA methyltransferase, a cryptochrome, a light inducible/controllable domain or a chemically inducible/controllable domain.

[0012] 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.

[0013] 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 the U6 promoter, is amplified with a DNA oligo and added onto 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.

[0014] In one aspect, the invention provides for methods to improve activity by using a more active polymerase. 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.

[0015] 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 a preferred embodiment, the Cas9 is delivered into the cell in the form of mRNA. This allows for the transient expression of the enzyme thereby reducing toxicity. In another preferred embodiment, the invention also provides for methods of expressing Cas9 under the control of an inducible promoter the constructs used therein.

[0016] 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. 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.

[0017] 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 Francisella novicida type II CRISPR locus), and the conserved Asp residue 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.

[0018] 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 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 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.

[0019] 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, SaCas9, StlCas9 and so forth. Further examples are provided herein.

[0020] 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 is known.

[0021] In further embodiments, the invention provides for methods of enhancing the function of Cas9 by generating chimeric Cas9 proteins. 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 proteins.

[0022] It will be appreciated that in the present methods, where the organism is an animal or a plant, 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*.

[0023] In one aspect the invention provides for a method of modifying a genomic locus of interest in a cell comprising delivering a non-naturally occurring or engineered composition comprising : I. a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence may comprise: (a) a first guide sequence capable of hybridizing to the first target sequence, (b) a first tracr mate

sequence, (c) a first tracr sequence, (d) a second guide sequence capable of hybridizing to the second target sequence, (e) a second tracr mate sequence, and (f) a second tracr sequence, and II. a polynucleotide sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences, wherein (a), (b), (c), (d), (e) and (f) may be arranged in a 5' to 3' orientation, wherein the polynucleotide sequence may comprise a linker sequence between the first tracr sequence and the second guide sequence, whereby the first guide sequence and the second guide sequence are in tandem, wherein when transcribed, the first and the second tracr mate sequence may 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 may comprise the CRISPR enzyme complexed with (1) the first guide sequence that is hybridized to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence, wherein the second CRISPR complex may comprise the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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 other strand near the second target sequence inducing a double strand break, thereby modifying the genomic locus.

[0024] The invention also comprehends a method of modifying an organism or a non-human organism 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, (b) a first tracr mate sequence, (c) a first tracr sequence, (d) a second guide sequence capable of hybridizing to the second target sequence, (e) a second tracr mate sequence, and (f) a second tracr sequence, and wherein a linker sequence is present between the first tracr sequence and the second guide sequence, whereby the first guide sequence and the second guide sequence are in tandem, and II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and wherein components I and II are

located on the same or different vectors of the system, when transcribed, a first tracr mate sequence hybridizes to a first tracr sequence 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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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 other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism with a reduction in likelihood of off-target modifications. In some instances, the second tracr may hybridize with the first tracr-mate.

[0025] The invention also comprehends a non-naturally occurring or engineered composition for modifying an organism or a non-human organism 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 : I. a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises: (a) a first guide sequence capable of hybridizing to the first target sequence, (b) a first tracr mate sequence, (c) a first tracr sequence, (d) a second guide sequence capable of hybridizing to the second target sequence, (e) a second tracr mate sequence, and (f) a second tracr sequence, and II. a polynucleotide sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences, wherein (a), (b), (c), (d), (e) and (f) may be arranged in a 5' to 3' orientation, wherein the polynucleotide sequence comprises a linker sequence between the first tracr sequence and the second guide sequence, whereby the first guide sequence and the second guide sequence are in tandem, 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 to the first target sequence, and (2) the first tracr

mate sequence that is hybridized to the first tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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 other strand near the second target sequence inducing a double strand break to modify the organism or the non-human organism with a reduction in likelihood of off-target modifications.

[0026] In embodiments of the methods and compositions of the invention, the modifying of the organism or the non-human organism with a reduction in likelihood of off-target modifications may comprise inducing a microdeletion, wherein the microdeletion may comprise deletion of a sequence between the first and second target sequences. In embodiments of the invention, the linker sequence may comprise at least 5 nucleotides, or at least 10 nucleotides or at least 20 nucleotides. In an advantageous embodiment, the linker sequence has 8 or 12 nucleotides. In other embodiments, the first or second guide sequence, the first or second tracr sequence and/or the first or second tracr mate sequences are modified, wherein the modification comprises optimized first or second tracr sequence and/or optimized first or second guide sequence and/or, co-fold structure of first or second tracr sequence and/or first or second tracr mate sequence(s) respectively and/or stabilizing secondary structures (hair pins) of first or second tracr sequence and/or first or second tracr sequence with a reduced region of base-pairing and/or first or second tracr sequence fused RNA elements.

[0027] In a further embodiments of the methods and compositions 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. The invention also comprehends the polynucleotides encoding 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 nanoparticles, exosomes, microvesicles, or a gene-gun. In a further embodiment, the first and second tracr mate sequence share 100% identity and/or the first and second tracr sequence share 100% identity. In a preferred embodiment, the CRISPR enzyme is a Cas9 enzyme, e.g. SpCas9.

It is not necessarily the case for 100% identity between the first tracr and second tracr sequence: Many of the altered scaffolds have changes to the DR/TracR sequences as well as the distal hairpins. It is also possible to have tandems involving alternate scaffolds that have distinct DR/TRACR sequences.

[0028] In another aspect, the invention comprehends a method of modifying an organism or a non-human organism 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 CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises: (a) a first guide sequence capable of hybridizing to the first target sequence, (b) a first tracr mate sequence, (c) a first tracr sequence, (d) a second guide sequence capable of hybridizing to the second target sequence, (e) a second tracr mate sequence, and (f) a second tracr sequence, and II. 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), (c), (d), (e) and (f) are may be in a 5' to 3' orientation, wherein the polynucleotide sequence comprises a linker sequence between the first tracr sequence and the second guide sequence, whereby the first guide sequence and the second guide sequence are in tandem, 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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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 other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism.

[0029] The invention comprehends a method of modifying an organism or a non-human organism 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, (b) a first tracr mate sequence, (c) a first tracr sequence, (d) a second guide sequence capable of hybridizing to the second target sequence, (e) a second tracr mate sequence, and (f) a second tracr sequence, and wherein a linker sequence is present between the first tracr sequence and the second guide sequence, whereby the first guide sequence and the second guide sequence are in tandem, and II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and wherein components I and II are located on the same or different vectors of the system, when transcribed, a first tracr mate sequence hybridizes to a first tracr sequence 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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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 other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism. In an embodiment of the invention, the one or more of the viral vectors are delivered via nanoparticles, exosomes, microvesicles, or a gene-gun. In some instances, the second tracr may hybridize with the first tracr-mate.

[0030] The invention also comprehends a non-naturally occurring or engineered composition modifying an organism or a non-human organism 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: I. a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises: (a) a first guide sequence capable of hybridizing to the first target sequence, (b) a first tracr mate sequence, (c) a first tracr sequence, (d) a second guide sequence capable of hybridizing to the second target sequence, (e) a second tracr

mate sequence, and (f) a second tracr sequence, and II. 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), (c), (d), (e) and (f) may be arranged in a 5' to 3' orientation, wherein the polynucleotide sequence comprises a linker sequence between the first tracr sequence and the second guide sequence, whereby the first guide sequence and the second guide sequence are in tandem, 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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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 other strand near the second target sequence inducing a double strand break to modify the organism or the non-human organism.

[0031] In embodiments of the methods and compositions 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 other strand near the second target sequence results in a 5' overhang. In some embodiments, the 5' overhang is at most 200 base pairs, or at most 100 base pairs, or at most 50 base pairs or at least 26 base pairs or at least 30 base pairs. In a preferred embodiment, the 5' overhang is 34-50 base pairs. The invention comprehends 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 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. It is not necessarily the case for 100% identity between the first tracr and second tracr sequence: Many of the altered scaffolds have changes to the DR/TracR sequences as well as the distal hairpins. It is also possible to have tandems involving alternate scaffolds that have distinct DR/TRACR sequences.

In a preferred embodiment, the CRISPR enzyme is a Cas9 enzyme, e.g. SpCas9. In a further embodiment, the CRISPR enzyme comprises one or more mutations in a catalytic domain, wherein the one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A. In a preferred embodiment, the CRISPR enzyme has the D10A mutation. In yet another embodiment, the modification comprises optimized first or second tracr sequence and/or optimized first or second guide sequence and/or, co-fold structure of first or second tracr sequence and/or first or second tracr mate sequence(s) respectively and/or stabilizing secondary structures (hair pins) of first or second tracr sequence and/or first or second tracr sequence with a reduced region of base-pairing and/or first or second tracr sequence fused RNA elements. In embodiments of the invention, the linker sequence may comprise at least 5 nucleotides, or at least 10 nucleotides or at least 20 nucleotides. In an advantageous embodiment, the linker sequence has 8 or 12 nucleotides.

[0031a] In a further aspect the invention provides for a method of modifying a DNA duplex at a locus of interest in a cell, the method comprising delivering to the cell:

I. a polynucleotide comprising:

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

(b) a first tracr mate sequence,

(c) a first tracr sequence,

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

(e) a second tracr mate sequence, and

(f) a second tracr sequence,

wherein (a), (b), (c), (d), (e) and (f) are arranged in a 5' to 3' orientation and a linker sequence is present between (c) and (d), whereby the first guide sequence and the second guide sequence are in tandem,

and

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

wherein the first target sequence is on a first strand of the DNA duplex and the second target sequence is on the opposite strand of the DNA duplex, and when the first and second guide sequences are hybridized to said target sequences in the duplex, the 5' ends of the first guide sequence and the second guide sequence are offset relative to each other by at least one base pair of the duplex;

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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence,

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

and wherein said first strand of the DNA duplex is cleaved near said first target sequence, and said opposite strand of the DNA duplex is cleaved near said second target sequence, resulting in a double strand break with a 5' overhang or a 3' overhang.

[0031b] In a further aspect the invention provides for a method of modifying a DNA duplex at a locus of interest in a cell, the method comprising delivering to the cell a vector system comprising one or more vectors comprising:

I. a first polynucleotide comprising a regulatory element operably linked to

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

(b) a first tracr mate sequence,

(c) a first tracr sequence,

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

(e) a second tracr mate sequence, and

(f) a second tracr sequence, and

wherein (a), (b), (c), (d), (e) and (f) are arranged in a 5' to 3' orientation and a linker sequence is present between (c) and (d), whereby the first guide sequence and the second guide sequence are in tandem,

and

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

wherein components I and II are located on the same or different vectors of the system, wherein the first target sequence is on a first strand of the DNA duplex and the second target sequence is on the opposite strand of the DNA duplex, and when the first and second guide sequences are hybridized to said target sequences in the duplex, the 5' ends of the first guide

sequence and the second guide sequence are offset relative to each other by at least one base pair of the duplex;

wherein when transcribed, the first and the second tracr mate sequences hybridize to a tracr sequence, and the first and the second guide sequences 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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to a tracr sequence,

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

and wherein said first strand of the DNA duplex is cleaved near said first target sequence, and said opposite strand of the DNA duplex is cleaved near said second target sequence, resulting in a double strand break with a 5' overhang or a 3' overhang.

[0031c] In a further aspect the invention provides for a kit or composition comprising:

I. a polynucleotide comprising:

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

(b) a first tracr mate sequence,

(c) a first tracr sequence.

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

(e) a second tracr mate sequence, and

(f) a second tracr sequence,

wherein (a), (b), (c), (d), (e) and (f) are arranged in a 5' to 3' orientation and a linker sequence is present between (c) and (d), whereby the first guide sequence and the second guide sequence are in tandem,

and

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

wherein the first target sequence is on a first strand of the DNA duplex and the second target sequence is on the opposite strand of the DNA duplex, and when the first and second guide sequences are hybridized to said target sequences in the duplex, the 5' ends of the first guide sequence and the second guide sequence are offset relative to each other by at least one base pair of the duplex;

2014281026 05 May 2020

and optionally wherein each of I and II is provided in the same or a different vector.

[0032] In other aspects of the methods and compositions of the invention, more than two guide sequences may be in tandem. The term "tandem single guide RNA" or "tsgRNA" is used to refer to one or more single guide RNAs (sgRNAs) connected by one or more linker sequences. In a preferred embodiment, two guide sequences are in tandem in the polynucleotide sequence. A key aspect of the invention is the utilization of a single promoter or operable element to drive the expression of more than one guide RNA.

[0033] 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:

[0034] delivering a non-naturally occurring or engineered composition comprising:

[0035] A) - I. a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises:

[0036] (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell,

[0037] (b) a tracr mate sequence, and

[0038] (c) a tracr sequence, and

[0039] II. a polynucleotide sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences,

[0040] wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

[0041] 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

[0042] wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized 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,

[0043] or

[0044] (B) I. polynucleotides comprising:

[0045] (a) a guide sequence capable of hybridizing to a target sequence in a prokaryotic cell, and

[0046] (b) at least one or more tracr mate sequences,

[0047] II. a polynucleotide sequence encoding a CRISPR enzyme, and

[0048] III. a polynucleotide sequence comprising a tracr sequence,

[0049] 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

[0050] wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized 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.

[0051] Any or all of the polynucleotide sequence encoding a CRISPR enzyme, guide sequence, tracr mate sequence or tracr sequence, may be RNA. The polynucleotides encoding the sequence encoding a CRISPR enzyme, the guide sequence, tracr mate sequence or tracr sequence may be RNA and are delivered via nanoparticles, exosomes, microvesicles, or a gene-gun.

[0052] It will be appreciated that where reference is made to a polynucleotide, which is RNA and is said to 'comprise' a feature such a tracr mate sequence, the RNA sequence includes the feature. Where the polynucleotide is DNA and is said to comprise a feature such a tracr mate sequence, the DNA sequence is or can be transcribed into the RNA 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).

[0053] Accordingly, in certain embodiments the invention provides 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 to the target sequence, and (2) the tracr mate sequence that is hybridized 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 prokaryotic 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 to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence.

[0054] 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 nanoparticles, exosomes, microvesicles, or a gene-gun.

[0055] By manipulation of a target sequence, Applicants also mean the epigenetic manipulation of a target sequence. This 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 or reducing 3D folding.

[0056] 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 (if the organism is multicellular). 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.

[0057] 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 vector system comprising one or more AAV vectors comprising 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) 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 to the target sequence, and (2) the tracr mate sequence that is hybridized 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 prokaryotic 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 to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence.

[0058] Some methods of the invention can include inducing expression. In some methods of the invention the organism or subject is a eukaryote (including mammal including human) or a non-human eukaryote or a non-human animal or a non-human mammal. In some methods of the invention the organism or subject is a plant. In some methods of the invention the organism or subject is a mammal or a non-human mammal. In some methods of the invention the organism or subject is algae. In some methods of the invention the viral vector is an AAV. In some methods of the invention the CRISPR enzyme is a Cas9. In some methods of the invention the

expression of the guide sequence is under the control of the T7 promoter is driven by the expression of T7 polymerase.

[0059] By manipulation of a target sequence, Applicants also mean the epigenetic manipulation of a target sequence. This 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 or reducing 3D folding.

[0060] It will be appreciated that where reference is made to a method of modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest, this may apply to the organism as a whole or just a single cell or population of cells from that organism (if the organism is multicellular). 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.

[0061] 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 or a non-human subject 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 vector system comprising one or more AAV vectors comprising 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) 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 to the target sequence, and (2) the tracr mate sequence that is hybridized 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 prokaryotic 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 to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence.

[0062] Some methods of the invention can include inducing expression. In some methods of the invention the organism or subject is a eukaryote or a non-human eukaryote or a non-human animal. In some methods of the invention the organism or subject is a plant. In some methods of the invention the organism or subject is a mammal or a non-human mammal. In some methods of the invention the organism or subject is algae. In some methods of the invention the viral vector is an AAV. In some methods of the invention the CRISPR enzyme is a Cas9. In some methods of the invention the expression of the guide sequence is under the control of the T7 promoter is driven by the expression of T7 polymerase.

[0063] The invention in some embodiments comprehends a method of delivering a CRISPR enzyme comprising delivering to a cell mRNA encoding the CRISPR enzyme. In some of these methods the CRISPR enzyme is a Cas9.

[0064] The invention in some embodiments comprehends a method of preparing the AAV of the invention comprising transfecting plasmid(s) containing or consisting essentially of nucleic acid molecule(s) coding for the AAV into AAV-infected 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.

[0065] In plants, pathogens are often host-specific. For example, *Fusarium oxysporum* f. sp. *lycopersici* causes tomato wilt but attacks only tomato, and *F. oxysporum f. dianthii Puccinia graminis* f. sp. *tritici* attacks only wheat. Plants have existing and induced defenses to resist most pathogens. Mutations and recombination events across plant generations lead to genetic variability that gives rise to susceptibility, especially as pathogens reproduce with more frequency than plants. In plants there can be non-host resistance, e.g., the host and pathogen are incompatible. There can also be Horizontal Resistance, e.g., partial resistance against all races of a pathogen, typically controlled by many genes and Vertical Resistance, e.g., complete resistance to some races of a pathogen but not to other races, typically controlled by a few genes. In a Gene-for-Gene level, plants and pathogens evolve together, and the genetic changes in one balance changes in other. Accordingly, using Natural Variability, breeders combine most useful genes for Yield, Quality, Uniformity, Hardiness, Resistance. The sources of resistance genes include native or foreign Varieties, Heirloom Varieties, Wild Plant Relatives, and Induced Mutations, e.g., treating plant material with mutagenic agents. Using the present invention, plant breeders are provided with a new tool to induce mutations. Accordingly, one skilled in the art can analyze the genome of sources of resistance genes, and in Varieties having desired characteristics or traits employ the present invention to induce the rise of resistance genes, with more precision than previous mutagenic agents and hence accelerate and improve plant breeding programs.

[0066] 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. 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. 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 5'-motif (where N is any Nucleotide), 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 for SpCas9 or SaCas9 enzymes (or derived enzymes), respectively, as mentioned below.

[0067] It will be appreciated that SpCas9 or SaCas9 are those from or derived from *S. pyogenes* or *S. aureus* Cas9.

[0068] 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 with a reduction in likelihood of 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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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 other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism with a reduction in likelihood of off-target modifications.

[0069] 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 encoding 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 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. It is not necessarily the case for 100% identity between the first tracr and second tracr sequence: Many of the altered scaffolds have changes to the DR/TracR sequences as well as the distal hairpins. It is also possible to have tandems involving alternate scaffolds that have distinct DR/TRACR sequences. 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.

[0070] 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 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.

[0071] The invention in some embodiments comprehends a method of modifying an organism or a non-human organism with a reduction in likelihood of 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 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 to the first target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence, wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the tracr mate sequence that is hybridized 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 other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism with a reduction in likelihood of off-target modifications.

[0072] 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. It is not necessarily the case for 100% identity between the first tracr and second tracr sequence: Many of the altered scaffolds have changes to the DR/TracR sequences as well as the distal hairpins. It is also possible to have tandems involving alternate scaffolds that have distinct DR/TRACR sequences. 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 a further embodiment of the invention, one or more of the viral vectors are delivered via nanoparticles, exosomes, microvesicles, or a gene-gun.

[0073] 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 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.

[0074] The invention in some embodiments comprehends a method of modifying a genomic locus of interest with a reduction in likelihood of off-target modifications by introducing into a cell containing and expressing a double stranded DNA molecule encoding the gene product 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.

[0075] 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.

[0076] 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, wherein it may be 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 Cas 9 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.

[0077] 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.

[0078] 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.

[0079] 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, wherein it may be 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 Cas 9 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.

[0080] 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.

[0081] 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.

[0082] 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, wherein it may be 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 Cas 9

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.

[0083] 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. 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 nanoparticles, exosomes, microvesicles, or a gene-gun.

[0084] 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 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.

[0085] 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. 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.

[0086] 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 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 to the target sequence within the target polynucleotide, and (2) the tracr mate sequence that is hybridized 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.

[0087] In one aspect the invention provides for a method of selecting one or more prokaryotic cell(s) by introducing one or more mutations in a gene in the one or more prokaryotic cell (s), the method comprising: introducing one or more vectors into the prokaryotic cell (s), 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, a tracr sequence, and an editing template; wherein the editing template comprises the one or more mutations that abolish CRISPR enzyme cleavage; allowing homologous recombination of the editing template with the target polynucleotide in the cell(s) to be selected; allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said gene, wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence within the target polynucleotide, and (2) the tracr mate sequence that is hybridized to the tracr sequence, wherein binding of the CRISPR complex to the target polynucleotide induces cell death, thereby allowing one or more prokaryotic cell(s) in which one or more mutations have been introduced to be selected. In a preferred embodiment, the CRISPR enzyme is Cas9. In another aspect of the invention the cell to be selected may be a eukaryotic cell. Aspects of the invention allow for selection of specific cells without requiring a selection marker or a two-step process that may include a counter-selection system.

[0088] 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 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.

[0089] 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.

[0090] 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 to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence.

[0091] In some methods, a target polynucleotide can be inactivated to effect the modification of the expression in a cell. For example, upon the binding of a CRISPR complex to a target sequence in a cell, the target polynucleotide is inactivated such that the sequence is not transcribed, the coded protein is not produced, or the sequence does not function as the wild-type sequence does. For example, a protein or microRNA coding sequence may be inactivated such that the protein is not produced.

[0092] In certain embodiments, the CRISPR enzyme comprises one or more mutations 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. In some embodiments, a transcriptional activation domain is VP64. In some embodiments, a transcription repression domains is KRAB. In some embodiments, a transcription repression domain is SID, or concatemers of SID (i.e. SID4X). In some embodiments, an epigenetic modifying enzyme is provided. In some embodiments, an activation domain is provided, which may be the P65 activation domain.

[0093] 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 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 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.

[0094] 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.

[0095] 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 is known.

[0096] 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 (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. 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 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 the CRISPR enzyme is truncated, comprised of less than one thousand amino acids or less than four thousand amino acids, is a nuclease or nickase, is codon-optimized comprises one or more mutations, and/or comprises a chimeric CRISPR enzyme, or the other options as herein discussed. AAV viral vectors are preferred.

[0097] 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.

[0098] For example, a suitable PAM is 5'-NRG or 5'-NNGRR for SpCas9 or SaCas9 enzymes (or derived enzymes), respectively.

[0099] It will be appreciated that SpCas9 or SaCas9 are those from or derived from *S. pyogenes* or *S. aureus* Cas9.

[0100] The invention does not intend to encompass within the scope of the invention any previously known product, process of making the product, or method of using the product such that Applicants reserve the right and hereby disclose a disclaimer of any previously known product, process, or method. It is further noted that the invention does not intend to encompass within the scope of the invention any product, process, or making of the product or method of using the product, which does not meet the written description and enablement requirements of the USPTO (35 U.S.C. § 112, first paragraph) or the EPO (Article 83 of the EPC), such that Applicants reserve the right and hereby disclose a disclaimer of any previously described product, process of making the product, or method of using the product.

[0101] 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.

[0101a] The discussion of documents, acts, materials, devices, articles and the like is included in this specification solely for the purpose of providing a context for the present invention. It is not suggested or represented that any or all of these matters formed part of the prior art base or were common general knowledge in the field relevant to the present invention as it existed before the priority date of each claim of this application.

[0101b] Throughout the description and claims of this specification, the word “comprise” and variations of the word, such as “comprising” and “comprises”, is not intended to exclude other additives, components, integers or steps.

[0102] These and other embodiments are disclosed or are obvious from and encompassed by, the following Detailed Description.

2014281026 05 May 2020

BRIEF DESCRIPTION OF THE DRAWINGS

[0103] 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:

[0104] **Figure 1** shows a schematic model of the CRISPR system. The Cas9 nuclease from Streptococcus pyogenes (yellow) is targeted to genomic DNA by a synthetic guide RNA (sgRNA) consisting of a 20-nt guide sequence (blue) and a scaffold (red). The guide sequence base-pairs with the DNA target (blue), directly upstream of a requisite 5'-NGG protospacer

adjacent motif (PAM; magenta), and Cas9 mediates a double-stranded break (DSB) ~3 bp upstream of the PAM (red triangle).

[0105] **Figure 2A-F** shows a schematic representation assay carried out to evaluate the cleavage specificity of Cas9 form *Streptococcus pyogenes*. Single base pair mismatches between the guide RNA sequence and the target DNA are mapped against cleavage efficiency in %.

[0106] **Figure 3A-F** is a circular phylogenetic tree of Cas genes

[0107] **Figure 4A-F** shows the linear phylogenetic analysis revealing five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids).

[0108] **Figure 5** shows a graph representing the length distribution of Cas9 orthologs.

[0109] **Figure 6A-M** shows sequences where the mutation points are located within the SpCas9 gene.

[0110] **Figure 7A** shows the Conditional Cas9, Rosa26 targeting vector map.

[0111] **Figure 7B** shows the Constitutive Cas9, Rosa26 targeting vector map.

[0112] **Figure 8** shows a schematic of the important elements in the Constitutive and Conditional Cas9 constructs.

[0113] **Figure 9** shows delivery and *in vivo* mouse brain Cas9 expression data.

[0114] **Figure 10** shows RNA delivery of Cas9 and chimeric RNA into cells (A) Delivery of a GFP reporter as either DNA or mRNA into Neuro-2A cells. (B) Delivery of Cas9 and chimeric RNA against the Icam2 gene as RNA results in cutting for one of two spacers tested. (C) Delivery of Cas9 and chimeric RNA against the F7 gene as RNA results in cutting for one of two spacers tested.

[0115] **Figure 11** shows how DNA double-strand break (DSB) repair promotes gene editing. In the error-prone non-homologous end joining (NHEJ) pathway, the ends of a DSB are processed by endogenous DNA repair machineries and rejoined together, which can result in random insertion/deletion (indel) mutations at the site of junction. Indel mutations occurring within the coding region of a gene can result in frame-shift and a premature stop codon, leading to gene knockout. Alternatively, a repair template in the form of a plasmid or single-stranded oligodeoxynucleotides (ssODN) can be supplied to leverage the homology-directed repair (HDR) pathway, which allows high fidelity and precise editing.

[0116] **Figure 12A-C** shows shows anticipated results for HDR in HEK and HUES9 cells. (a) Either a targeting plasmid or an ssODN (sense or antisense) with homology arms can be used to edit the sequence at a target genomic locus cleaved by Cas9 (red triangle). To assay the efficiency of HDR, we introduced a HindIII site (red bar) into the target locus, which was PCR-amplified with primers that anneal outside of the region of homology. Digestion of the PCR product with HindIII reveals the occurrence of HDR events. (b) ssODNs, oriented in either the sense or the antisense (s or a) direction relative to the locus of interest, can be used in combination with Cas9 to achieve efficient HDR-mediated editing at the target locus. A minimal homology region of 40 bp, and preferably 90 bp, is recommended on either side of the modification (red bar). (c) Example of the effect of ssODNs on HDR in the EMX1 locus is shown using both wild-type Cas9 and Cas9 nickase (D10A). Each ssODN contains homology arms of 90 bp flanking a 12-bp insertion of two restriction sites.

[0117] **Figure 13A-C** shows the repair strategy for Cystic Fibrosis delta F508 mutation.

[0118] **Figure 14A-B** shows (a) a schematic of the GAA repeat expansion in FXN intron 1 and (b) a schematic of the strategy adopted to excise the GAA expansion region using the CRISPR/Cas system.

[0119] **Figure 15** shows a screen for efficient SpCas9 mediated targeting of Tet1-3 and Dnmt1, 3a and 3b gene loci. Surveyor assay on DNA from transfected N2A cells demonstrates efficient DNA cleavage by using different gRNAs.

[0120] **Figure 16** shows a strategy of multiplex genome targeting using a 2-vector system in an AAV1/2 delivery system. Tet1-3 and Dnmt1, 3a and 3b gRNA under the control of the U6 promoter. GFP-KASH under the control of the human synapsin promoter. Restriction sides shows simple gRNA replacement strategy by subcloning. HA-tagged SpCas9 flanked by two nuclear localization signals (NLS) is shown. Both vectors are delivered into the brain by AAV1/2 virus in a 1:1 ratio.

[0121] **Figure 17** shows verification of multiplex DNMT targeting vector #1 functionality using Surveyor assay. N2A cells were co-transfected with the DNMT targeting vector #1 (+) and the SpCas9 encoding vector for testing SpCas9 mediated cleavage of DNMTs genes family loci. gRNA only (-) is negative control. Cells were harvested for DNA purification and downstream processing 48 h after transfection.

[0122] **Figure 18** shows verification of multiplex DNMT targeting vector #2 functionality using Surveyor assay. N2A cells were co-transfected with the DNMT targeting vector #1 (+) and the SpCas9 encoding vector for testing SpCas9 mediated cleavage of DNMTs genes family loci. gRNA only (-) is negative control. Cells were harvested for DNA purification and downstream processing 48 h after transfection.

[0123] **Figure 19** shows schematic overview of short promoters and short polyA versions used for HA-SpCas9 expression in vivo. Sizes of the encoding region from L-ITR to R-ITR are shown on the right.

[0124] **Figure 20** shows schematic overview of short promoters and short polyA versions used for HA-SaCas9 expression in vivo. Sizes of the encoding region from L-ITR to R-ITR are shown on the right.

[0125] **Figure 21** shows expression of SpCas9 and SaCas9 in N2A cells. Representative Western blot of HA-tagged SpCas9 and SaCas9 versions under the control of different short promoters and with or short polyA (spA) sequences. Tubulin is loading control. mCherry (mCh) is a transfection control. Cells were harvested and further processed for Western blotting 48 h after transfection.

[0126] **Figure 22** shows screen for efficient SaCas9 mediated targeting of Tet3 gene locus. Surveyor assay on DNA from transfected N2A cells demonstrates efficient DNA cleavage by using different gRNAs with NNGGGT PUM sequence. GFP transfected cells and cells expressing only SaCas9 are controls.

[0127] **Figure 23** shows expression of HA-SaCas9 in the mouse brain. Animals were injected into dentate gyri with virus driving expression of HA-SaCas9 under the control of human Synapsin promoter. Animals were sacrificed 2 weeks after surgery. HA tag was detected using rabbit monoclonal antibody C29F4 (Cell Signaling). Cell nuclei stained in blue with DAPI stain.

[0128] **Figure 24** shows expression of SpCas9 and SaCas9 in cortical primary neurons in culture 7 days after transduction. Representative Western blot of HA-tagged SpCas9 and SaCas9 versions under the control of different promoters and with bgh or short polyA (spA) sequences. Tubulin is loading control.

[0129] **Figure 25** shows LIVE/DEAD stain of primary cortical neurons 7 days after transduction with AAV1 particles carrying SpCas9 with different promoters and multiplex

gRNAs constructs (example shown on the last panel for DNMTs). Neurons after AAV transduction were compared with control untransduced neurons. Red nuclei indicate permeabilized, dead cells (second line of panels). Live cells are marked in green color (third line of panels).

[0130] **Figure 26** shows LIVE/DEAD stain of primary cortical neurons 7 days after transduction with AAV1 particles carrying SaCas9 with different promoters. Red nuclei indicate permeabilized, dead cells (second line of panels). Live cells are marked in green color (third line of panels).

[0131] **Figure 27** shows comparison of morphology of neurons after transduction with AAV1 virus carrying SpCas9 and gRNA multiplexes for TETs and DNMTs genes loci. Neurons without transduction are shown as a control.

[0132] **Figure 28** shows verification of multiplex DNMT targeting vector #1 functionality using Surveyor assay in primary cortical neurons. Cells were co-transduced with the DNMT targeting vector #1 and the SpCas9 viruses with different promoters for testing SpCas9 mediated cleavage of DNMTs genes family loci.

[0133] **Figure 29** shows in vivo efficiency of SpCas9 cleavage in the brain. Mice were injected with AAV1/2 virus carrying gRNA multiplex targeting DNMT family genes loci together with SpCas9 viruses under control of 2 different promoters: mouse Mecp2 and rat Map1b. Two weeks after injection brain tissue was extracted and nuclei were prepped and sorted using FACS, based on the GFP expression driven by Synapsin promoter from gRNA multiplex construct. After gDNA extraction Surveyor assay was run. + indicates GFP positive nuclei and – control, GFP-negative nuclei from the same animal. Numbers on the gel indicate assessed SpCas9 efficiency.

[0134] **Figure 30** shows purification of GFP-KASH labeled cell nuclei from hippocampal neurons. The outer nuclear membrane (ONM) of the cell nuclear membrane is tagged with a fusion of GFP and the KASH protein transmembrane domain. Strong GFP expression in the brain after one week of stereotactic surgery and AAV1/2 injection. Density gradient centrifugation step to purify cell nuclei from intact brain. Purified nuclei are shown. Chromatin stain by Vybrant® DyeCycle™ Ruby Stain is shown in red, GFP labeled nuclei are green. Representative FACS profile of GFP+ and GFP- cell nuclei (Magenta: Vybrant® DyeCycle™ Ruby Stain, Green: GFP).

[0135] **Figure 31** shows efficiency of SpCas9 cleavage in the mouse brain. Mice were injected with AAV1/2 virus carrying gRNA multiplex targeting TET family genes loci together with SpCas9 viruses under control of 2 different promoters: mouse Mecp2 and rat Map1b. Three weeks after injection brain tissue was extracted, nuclei were prepped and sorted using FACS, based on the GFP expression driven by Synapsin promoter from gRNA multiplex construct. After gDNA extraction Surveyor assay was run. + indicates GFP positive nuclei and – control, GFP-negative nuclei from the same animal. Numbers on the gel indicate assessed SpCas9 efficiency.

[0136] **Figure 32** shows GFP-KASH expression in cortical neurons in culture. Neurons were transduced with AAV1 virus carrying gRNA multiplex constructs targeting TET genes loci. The strongest signal localize around cells nuclei due to KASH domain localization.

[0137] **Figure 33** shows (top) a list of spacing (as indicated by the pattern of arrangement for two PAM sequences) between pairs of guide RNAs. Only guide RNA pairs satisfying patterns 1, 2, 3, 4 exhibited indels when used with SpCas9(D10A) nickase. (bottom) Gel images showing that combination of SpCas9(D10A) with pairs of guide RNA satisfying patterns 1, 2, 3, 4 led to the formation of indels in the target site.

[0138] **Figure 34** shows a list of U6 reverse primer sequences used to generate U6-guide RNA expression casssettes. Each primer needs to be paired with the U6 forward primer "gcactgagggcctatttcccatgattc" to generate amplicons containing U6 and the desired guide RNA.

[0139] **Figure 35** shows a Genomic sequence map from the human Emx1 locus showing the locations of the 24 patterns listed in Figure 33.

[0140] **Figure 36** shows on (right) a gel image indicating the formation of indels at the target site when variable 5' overhangs are present after cleavage by the Cas9 nickase targeted by different pairs of guide RNAs. on (left) a table indicating the lane numbers of the gel on the right and various parameters including identifying the guide RNA pairs used and the length of the 5' overhang present following cleavage by the Cas9 nickase.

[0141] **Figure 37** shows a Genomic sequence map from the human Emx1 locus showing the locations of the different pairs of guide RNAs that result in the gel patterns of Figure 36 (right) and which are further described in Example 30.

[0142] **Figure 38A-D** shows (A) Optimization of sgRNA scaffold for additional stability; Architectures A and B are indicated. (B) a schematic illustrating a tandem sgRNA (2 sgRNAs

connected by a linker) driven by a single U6 promoter (C) a gel figure indicating tsgRNAs are designed to target a region of EMX1 so that, when co-transfected into HEK cells with wt Cas9, mediates a genomic microdeletion. Here, the targets flank the last exon of EMX1. tsgRNAs are synthesized with either custom scaffold A or B that differ from the wt sgRNA scaffold (as indicated above). A single tsgRNA encoding 2 spacers, with either architecture A or B, when co-delivered with wt Cas9, mediates a similar level of genomic microdeletion as two wt sgRNAs co-delivered with wt Cas9. (D) a gel figure where, alternatively, tsgRNAs are designed to target a region of EMX1 so that, when co-transfected into HEK cells with D10A Cas9, mediates an indel as previously described in double nickase experiments. Spacers are chosen so that the left and right targets are offset with at least 0, or 4-bp of non-overlap in between the 20-bp targets, and resulting cuts from D10A would generate 5'-overhangs. The gel image shows SURVEYOR assay measuring indels from either D10A Cas9 co-transfected with either a single tsgRNA encoding 2 spacers or two separate sgRNA, where the former can mediate a similar level of indels as the latter.

[0143] **Figure 39A-C** shows (A) tandem lentiviral vector design. (B) the scaffold sequences for LentiCRISPR. (C) a schematic and gel image indicating that the tandem guide RNAs constructd of modified scaffolds are more efficiently processed to individual units.

[0144] **Figure 40A-B** shows paired sgRNAs induce indels via double-nicking with the nickase Cas9n (A) Schematic illustrating DNA double strand break using a pair of Cas9 D10A nickases (Cas9n). Two sgRNA target Cas9n to nick both strands of DNA. The D10A mutation renders Cas9 capable of only cleaving the DNA strand that is complementary to the sgRNA. The offset distance refers to the length of DNA between the closest ends of the paired sgRNAs, in this case 4bp. (B) Representative gel image showing Cas9n mediated indel in the EMX1 locus of the human genome, as detected using the SURVEYOR nuclease assay with 650bp band representing the unmodified genomic target and bands around 300bp indicating the presence of indels.

[0145] **Figure 41A-B.** shows double-nicking is able to induce indels (A) Graphs showing indel frequency corresponding to indicated sgRNA offset distances across three different human genes: EMX1, DYRK1A, GRIN2B. (B) As an example, sequence of the human EMX1 locus targeted by Cas9n. sgRNA target sites and PAMs are indicated by blue and magenta bars respectively. Below, selected sequences showing representative indels.

[0146] **Figure 42A-C** shows double nicking strategy is able to facilitate homologous recombination (A) Schematic illustrating HDR targeted via a single stranded oligodeoxynucleotide (ssODN) template at a DSB created by a pair of Cas9n. Successful recombination at the DSB site introduces a HindIII restriction site. (B) Restriction digest assay gel showing successful insertion of HindIII cleavage sites by double nicking-facilitated HDR in HEK 293FT cells. Upper bands are unmodified template; lower bands are HindIII cleavage product. (C) Double nicking enhances HDR in HUES62 cells. HDR frequencies determined using deep sequencing. (n = 3; error bars show mean ± s.e.)

[0147] **Figure 43** shows characterization of double nicking spacing for homologous recombination. Schematic illustrating HDR with ssODN template (shown in blue, introduced HindIII site in red). Red arrowheads indicate binding site of respective sgRNA with black bars corresponding putative overhangs resulting from paired nicking activity. Panels at right show efficiency of recombination with the indicated sgRNA pairs, overhang length and type, and offset distances between paired sgRNAs.

[0148] **Figure 44A-D** shows double-nicking reduces non-specific activity at known off-target sites. (A) Schematic showing the target human EMX1 locus and sgRNA target sites. Genomic off-target sites for the right sgRNA are listed below. Off-target sites were identified previously as described in Hsu et al48. (B) SURVEYOR gels showing modification at the on-target site by Cas9n with two sgRNAs as well as by wild-type Cas9 with individual sgRNA. Indels at off-target 5 were only observed for wild-type Cas9 with sgRNA 1. (C) The levels of off-target modification are quantified using deep sequencing at all five off-target loci. (D) Specificity comparison of Cas9n and wild-type Cas9. The specificity ratio is calculated by taking the ratio of on-target and off-target modification rates. (n = 3; error bars show mean ± s.e.)

[0149] **Figure 45A-B** shows rational mutagenesis of sgRNA architecture. (A) Schematic of the chimeric sgRNA architecture with 20-bp guide sequence encoding for the target specificity. The different regions of the sgRNA interrogated by mutagenesis are named and highlighted above. (B) Description of mutations made and corresponding indel activity at human EMX1 locus.

[0150] **Figure 46** shows distal hairpin and DDR stabilization retains comparable indel activity. Schematic showing sequence of three alternative scaffold architectures aimed at

stabilization of the hairpins with changes base-pairs denoted in black. Indel-inducing activity of corresponding scaffolds compared to the wild-type sp85 architecture is shown at right.

[0151] **Figure 47A-C** shows U6-driven tandem guide RNAs are able to efficiently target two genomic loci. (A) Schematic illustrating a tandem sgRNA (2 sgRNAs connected by a linker) driven by a single U6 promoter. Images of PAGE gels of SURVEYOR assays demonstrate that tsgRNAs using modified RNA scaffolds delivered with the nickase Cas9n or wild-type Cas9 are able to induce genomic indels or microdeletions, (B and C, respectively) with frequencies comparable to co-delivery of two independent sgRNAs.

[0152] **Figure 48** shows optimization of tandem sgRNA linker length and structure. Gel quantification of band intensities from PCR amplification of the human EMX1 target loci comparing relative abundance of wild-type and modified DNA with varying tandem linkers of 0, 4, 8, 12, 16 base pairs, half direct repeat or full direct repeat (listed in bottom panel).

[0153] **Figure 49A-C** shows tandem guide RNAs are efficiently processed in only 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.

[0154] **Figure 50A-C** shows 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.

[0155] **Figure 51** shows tandem pairs between divergent scaffolds improve second spacer activity. Sequence alignment of the sgRNA scaffolds used in the previous study to the sp85 scaffold.

[0156] **Figure 52** shows different truncations of the Bi-directional U6 promoter and associated microdeletion activity and quantification of indel-inducing activity coming off of either side of the bi-directional promoter.

[0157] **Figure 53** shows a sequence of the smallest U6-8/U6-1 bi-directional reporter.

DETAILED DESCRIPTION OF THE INVENTION

[00100] Reference is made to US provisional patent application 61/862,468, filed on August 8, 2013. Reference is 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 applications 61/836,123 and 61/847,537, filed on June 17, 2013 and July 17, 2013 respectively. Reference is also made to US provisional patent applications 61/736,527 and 61/748,427, filed on December 12, 2012 and January 2, 2013, respectively. Reference is also made to US provisional patent application 61/791,409, filed on March 15, 2013. Reference is also made to US provisional patent application 61/799,800, filed March 15, 2013. Reference is also made to US provisional patent applications 61/835,936, 61/836,127, 61/836,101, 61/836,123, 61/836,080 and 61/835,973, each filed June 17, 2013. Reference is also made to US provisional patent application 61/842,322 and US patent application 14/054,414, filed on July 2, 2013 and October 15, 2013 respectively. Reference is also made to US provisional patent application 61/915,407, filed December 12, 2013. Each of these 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 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.

[00101] 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, and

- *Development and Applications of CRISPR-Cas9 for Genome Engineering*, Hsu et al, Cell 157, 1262-1278 (June 5, 2014) (Hsu 2014),

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.

[0100] 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.

[0101] An advantage of the present methods is that the CRISPR system avoids off-target binding and its resulting side effects. This is achieved using systems arranged to have a high degree of sequence specificity for the target DNA.

[0102] Cas9 optimization may be used to enhance function or to develop new functions, one can generate chimeric Cas9 proteins. Examples that the Applicants have generated are provided in Example 6. Chimeric Cas9 proteins can be made by combining fragments from different Cas9 homologs. For example, two example chimeric Cas9 proteins from the Cas9s described herein. For example, Applicants fused the N-term of St1Cas9 (fragment from this protein is in bold) with C-term of SpCas9. The benefit of making chimeric Cas9s include any or all of:

[0103] reduced toxicity;

[0104] improved expression in eukaryotic cells;

[0105] enhanced specificity;

[0106] reduced molecular weight of protein, make protein smaller by combining the smallest domains from different Cas9 homologs; and/or

[0107] altering the PAM sequence requirement.

[0108] The Cas9 may be used as a generic DNA binding protein. For example, and as shown in Example 7, Applicants used Cas9 as a generic DNA binding protein by mutating the two catalytic domains (D10 and H840) responsible for cleaving both strands of the DNA target. In order to upregulate gene transcription at a target locus Applicants fused the transcriptional activation domain (VP64) to Cas9. Other transcriptional activation domains are known. As shown in Example 11, transcriptional activation is possible. As also shown in Example 11, gene repression (in this case of the beta-catenin gene) is possible using a Cas9 repressor (DNA-binding domain) that binds to the target gene sequence, thus repressing its activity.

[0109] Transgenic animals are also provided. Preferred examples include animals comprising Cas9, in terms of polynucleotides encoding Cas9 or the protein itself. Mice, rats and rabbits are preferred. To generate transgenic mice with the constructs, as exemplified herein one may inject pure, linear DNA into the pronucleus of a zygote from a pseudo pregnant female, e.g. a CB56 female. Founders may then be identified are identified, genotyped, and backcrossed to CB57 mice. The constructs may then be cloned and optionally verified, for instance by Sanger sequencing. Knock outs are envisaged where for instance one or more genes are knocked out in a model. However, are knockins are also envisaged (alone or in combination). An example Knock in Cas9 mouse was generated and this is exemplified, but Cas9 knockins are preferred. To generate a Cas9 knock in mice one may target the same constitutive and conditional constructs to the Rosa26 locus, as described herein (Figures 7A-B and 8).

[0110] Utility of the conditional Cas9 mouse: Applicants have shown in 293 cells that the Cas9 conditional expression construct can be activated by co-expression with Cre. Applicants also show that the correctly targeted R1 mESCs can have active Cas9 when Cre is expressed. Because Cas9 is followed by the P2A peptide cleavage sequence and then EGFP Applicants identify successful expression by observing EGFP. Applicants have shown Cas9 activation in mESCs. This same concept is what makes the conditional Cas9 mouse so useful. Applicants may cross their conditional Cas9 mouse with a mouse that ubiquitously expresses Cre (ACTB-Cre line) and may arrive at a mouse that expresses Cas9 in every cell. It should only take the delivery of chimeric RNA to induce genome editing in embryonic or adult mice. Interestingly, if the conditional Cas9 mouse is crossed with a mouse expressing Cre under a tissue specific promoter,

there should only be Cas9 in the tissues that also express Cre. This approach may be used to edit the genome in only precise tissues by delivering chimeric RNA to the same tissue.

[0111] As mentioned above, transgenic animals are also provided, as are transgenic plants, especially crops and algae. The transgenic may be useful in applications outside of providing a disease model. These may include food of feed production through expression of, for instance, higher protein, carbohydrate, nutrient or vitamins levels than would normally be seen in the wildtype. In this regard, transgenic plants, especially pulses and tubers, and animals, especially mammals such as livestock (cows, sheep, goats and pigs), but also poultry and edible insects, are preferred.

[0112] Transgenic algae or other plants such as rape may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol), for instance. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[0113] In terms of i*n 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.

[0114] 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

[0115] These species are therefore, in general, preferred Cas9 species. Applicants have shown delivery and *in vivo* mouse brain Cas9 expression data.

[0116] Two ways to package Cas9 coding nucleic acid molecules, e.g., DNA, into viral vectors to mediate genome modification *in vivo* are preferred:

[0117] To achieve NHEJ-mediated gene knockout:

[0118] Single virus vector:

[0119] Vector containing two or more expression cassettes:

[0120] Promoter-Cas9 coding nucleic acid molecule -terminator

[0121] Promoter-gRNA1-terminator

[0122] Promoter-gRNA2-terminator

[0123] Promoter-gRNA(N)-terminator (up to size limit of vector)

[0124] Double virus vector:

[0125] Vector 1 containing one expression cassette for driving the expression of Cas9

[0126] Promoter-Cas9 coding nucleic acid molecule-terminator

[0127] Vector 2 containing one more expression cassettes for driving the expression of one or more guideRNAs

[0128] Promoter-gRNA1-terminator

[0129] Promoter-gRNA(N)-terminator (up to size limit of vector)

[0130] 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.

[0131] Promoter used to drive Cas9 coding nucleic acid molecule expression can include:

[0132] 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 toxicity due to over expression of Cas9.

[0133] For ubiquitous expression, can use promoters: CMV, CAG, CBh, PGK, SV40, Ferritin heavy or light chains, etc.

[0134] For brain expression, can use promoters: SynapsinI for all neurons, CaMKIIalpha for excitatory neurons, GAD67 or GAD65 or VGAT for GABAergic neurons, etc.

[0135] For liver expression, can use Albumin promoter

[0136] For lung expression, can use SP-B

[0137] For endothelial cells, can use ICAM

[0138] For hematopoietic cells can use IFNbeta or CD45

[0139] For Osteoblasts can use OG-2

[0140] Promoter used to drive guide RNA can include:

[0141] Pol III promoters such as U6 or H1

[0142] Use of Pol II promoter and intronic cassettes to express gRNA

[0143] 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 above promoters and vectors are preferred individually.

[0144] RNA delivery is also a useful method of i*n vivo* delivery. Figure 9 shows delivery and *in vivo* mouse brain Cas9 expression data. 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 Invivofectamine from Life Technologies and other reagents on the market can effectively deliver RNA molecules into the liver.

[0145] 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.

[0146] Various means of delivery are described herein, and further discussed in this section.

[0147] 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 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 viral vectors. In some embodiments, the viral vector is delivered to the tissue of interest by, for example, an intramuscular injection, while other times the viral delivery is via intravenous, transdermal, intranasal, oral, mucosal, or other delivery methods. Such delivery may be either via a single dose, or multiple doses. One skilled in the art understands that the actual dosage to be delivered herein may vary greatly depending upon a variety of factors, such as the vector chose, the target cell, organism, or tissue, the general condition of the subject to be treated, the degree of transformation/modification sought, the administration route, the administration mode, the type of transformation/modification sought, etc.

[0148] Such a dosage may further contain, for example, a carrier (water, saline, ethanol, glycerol, lactose, sucrose, calcium phosphate, gelatin, dextran, agar, pectin, peanut oil, sesame oil, etc.), a diluent, a pharmaceutically-acceptable carrier (e.g., phosphate-buffered saline), a pharmaceutically-acceptable excipient, and/or other compounds known in the art. Such a dosage formulation is readily ascertainable by one skilled in the art. The dosage may further contain one or more pharmaceutically acceptable salts such as, for example, a mineral acid salt such as a hydrochloride, a hydrobromide, a phosphate, a sulfate, etc.; and the salts of organic acids such as acetates, propionates, malonates, benzoates, etc. Additionally, auxiliary substances, such as wetting or emulsifying agents, pH buffering substances, gels or gelling materials, flavorings, colorants, , microspheres, polymers, suspension agents, etc. may also be present herein. In addition, one or more other conventional pharmaceutical ingredients, such as preservatives, humectants, suspending agents, surfactants, antioxidants, anticaking agents, fillers, chelating agents, coating agents, chemical stabilizers, etc. may also be present, especially if the dosage form is a reconstitutable form. Suitable exemplary ingredients include microcrystalline cellulose, carboxymethylcellulose sodium, polysorbate 80, phenylethyl alcohol, chlorobutanol, potassium sorbate, sorbic acid, sulfur dioxide, propyl gallate, the parabens, ethyl vanillin, glycerin, phenol, parachlorophenol, gelatin, albumin and a combination thereof. A thorough discussion of

pharmaceutically acceptable excipients is available in REMINGTON'S PHARMACEUTICAL SCIENCES (Mack Pub. Co., N.J. 1991) which is incorporated by reference herein.

[0149] 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 10^0 particles (e.g., about 1 x 10^9-1 x 10^{10} particles or about 1 x 10^9-1 x 10^{12} particles), or even at least about 1 x 10^{10} particles (e.g., about 1 x 10^{10}-1 x 10^{12} particles) of the adenoviral vector. Alternatively, the dose comprises no more than about 1 x 10^{14} particles, preferably no more than about 1 x 10^{13} particles, even more preferably no more than about 1 x 10^{12} particles, even more preferably no more than about 1 x 10^{11} particles, and most preferably no more than about 1 x 10^{10} particles (e.g., no more than about 1 x 10^9 articles). Thus, the dose may contain a single dose of adenoviral vector with, for example, about 1 x 10^6 particle units (pu), about 2 x 10^6 pu, about 4 x 10^6 pu, about 1 x 10^7 pu, about 2 x 10^7 pu, about 4 x 10^7 pu, about 1 x 10^8 pu, about 2 x 10^8 pu, about 4 x 10^8 pu, about 1 x 10^9 pu, about 2 x 10^9 pu, about 4 x 10^9 pu, about 1 x 10^{10} pu, about 2 x 10^{10} pu, about 4 x 10^{10} pu, about 1 x 10^{11} pu, about 2 x 10^{11} pu, about 4 x 10^{11} pu, about 1 x 10^{12} pu, about 2 x 10^{12} pu, or about 4 x 10^{12} pu of adenoviral vector. See, for example, the adenoviral vectors in U.S. Patent No. 8,454,972 B2 to Nabel, et. al., granted on June 4, 2013; incorporated by reference herein, and the dosages at col 29, lines 36-58 thereof. In an embodiment herein, the adenovirus is delivered via multiple doses.

[0150] 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.

[0151] 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.

[0152] 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. A mouse used in experiments is typically considered 20g, and one skilled in the art can extrapolate dosages from a 20g mouse to a 70 kg individual.

[0153] Cas9 and one or more guide RNA can be delivered using adeno associated virus (AAV), lentivirus, adenovirus or other plasmid or viral vector types, 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, 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.

[0154] 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 might use the Synapsin I promoter.

[0155] 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.

[0156] To enhance expression and reduce toxicity, the CRISPR enzyme and/or guide RNA can be modified using pseudo-U or 5-Methyl-C.

[0157] CRISPR enzyme mRNA and guide RNA may be delivered simultaneously using nanoparticles or lipid envelopes.

[0158] 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.

[0159] Furthermore, 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) Published online 09 January 2011) describes the use of lipid envelopes to deliver RNA. Use of lipid envelopes is also preferred in the present invention.

[0160] mRNA delivery methods are especially promising for liver delivery currently.

[0161] CRISPR enzyme mRNA and guide RNA might also be delivered separately. CRISPR enzyme mRNA can be delivered prior to the guide RNA to give time for CRISPR enzyme to be expressed. CRISPR enzyme mRNA might be administered 1-12 hours (preferably around 2-6 hours) prior to the administration of guide RNA.

[0162] 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

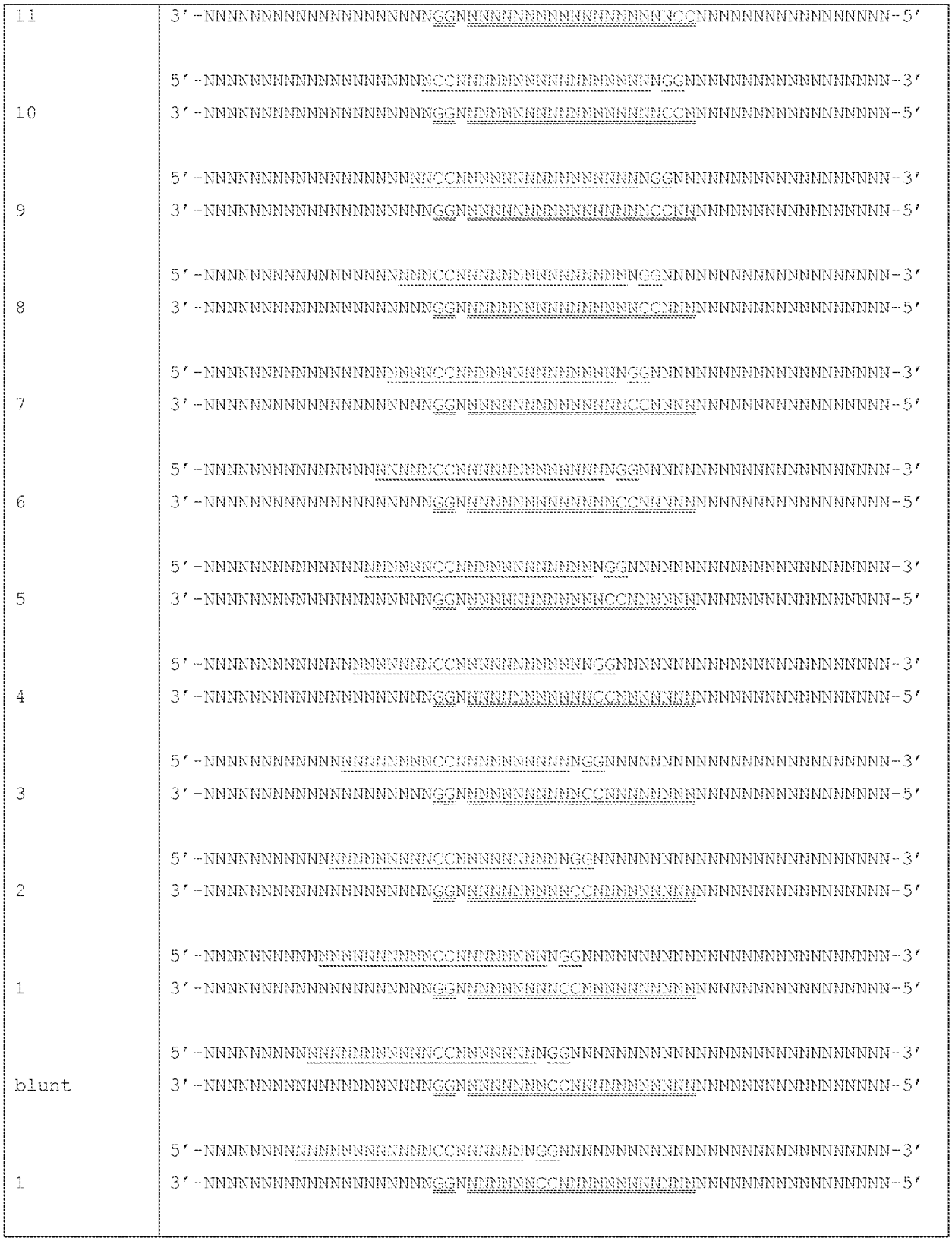
(preferably around 2-6 hours) after the initial administration of CRISPR enzyme mRNA + guide RNA.

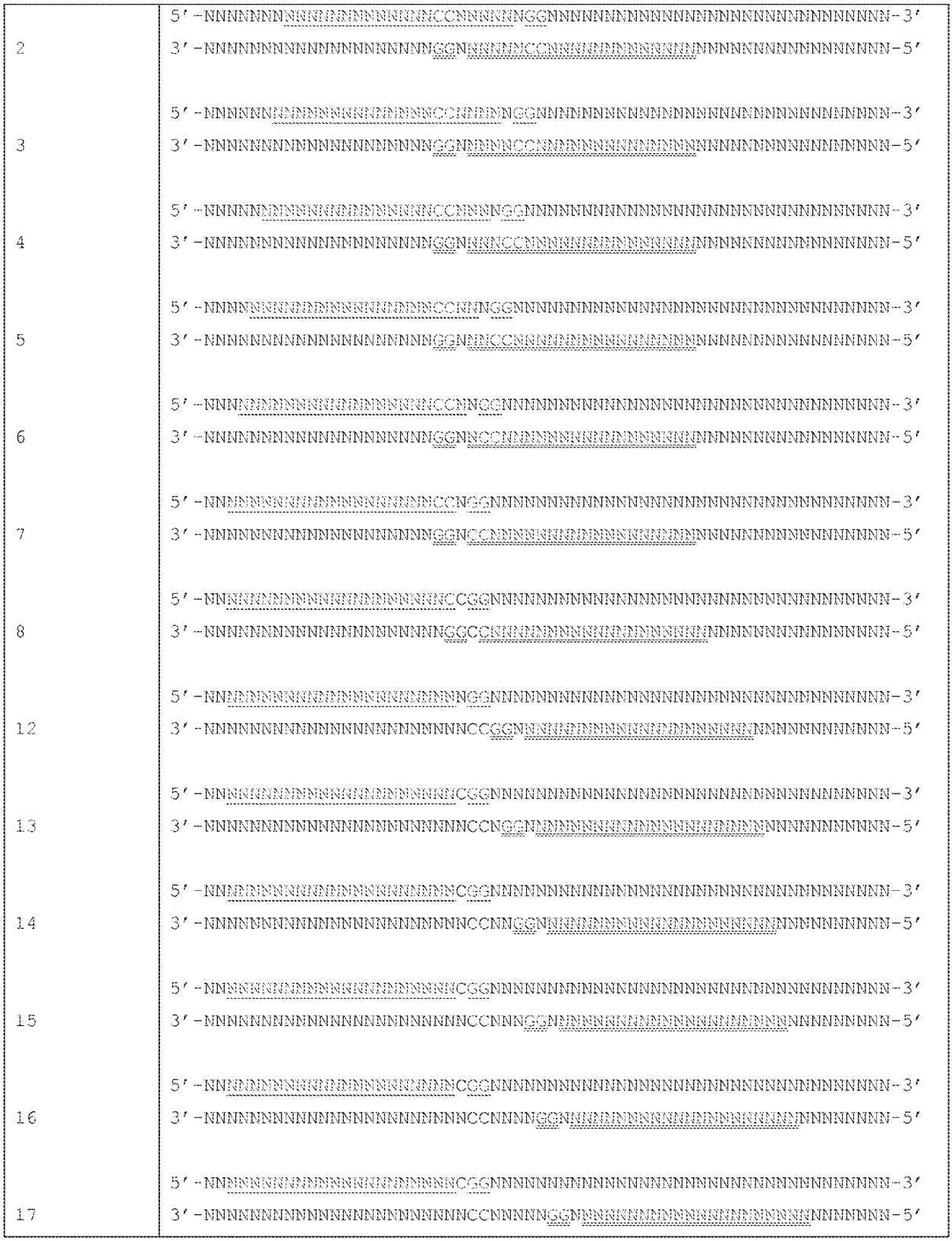
[0163] Additional administrations of CRISPR enzyme mRNA and/or guide RNA might be useful to achieve the most efficient levels of genome modification.

[0164] 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 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 or with a reduction in likelihood of the level of off-target modification should be chosen for in vivo delivery.

[0165] 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 in red (single underline) and blue (double underline) respectively (these examples are based on the PAM requirement for *Streptococcus pyogenes* Cas9.

Overhang length (bp)	Guide RNA design (guide sequence and PAM color coded)
	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNN-3'
14	3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNN-5'
	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNN-3'
13	3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNN-5'
	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNN-3'
12	3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNN-5'
	5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNN-3'


11
3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3'
10
3'-NNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3'
9
3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3'
8
3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3'
7
3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3'
6
3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3'
5
3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3'
4
3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3'
3
3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3'
2
3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3'
1
3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3'
blunt
3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'
5'-NNNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNN-3'
1
3'-NNNNNNNNNNNNNNNNNNNNGGNNNNNNNNNNNNNNNNNNCCNNNNNNNNNNNNNNNN-5'



[0166] Further interrogation of the system have given Applicants evidence of the 5' overhang (see, e.g., Ran et al., Cell. 2013 Sep 12;154(6):1380-9 and US Provisional Patent Application Serial No. 61/871,301 filed August 28, 2013). Applicants have further identified parameters that relate to efficient cleavage by the Cas9 nickase mutant when combined with two guide RNAs and these parameters include but are not limited to the length of the 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 or 1-34 base pairs. In other 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 other strand near the second target sequence results in a blunt cut or a 3' overhang. In embodiments of the invention the 3' overhang is at most 150, 100 or 25 base pairs or at least 15, 10 or 1 base pairs. In preferred embodiments the 3' overhang is 1-100 base pairs.

[0167] 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.

[0168] 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.

[0169] Since Cas9n and Cas9H840A nick opposite strands of DNA, substitution of Cas9n with Cas9H840A with a given sgRNA pair should result in the inversion of the overhang type. For example, a pair of sgRNAs that will generate a 5' overhang with Cas9n should in principle generate the corresponding 3' overhang instead. Therefore, sgRNA pairs that lead to the

generation of a 3' overhang with Cas9n might be used with Cas9H840A to generate a 5' overhang. Unexpectedly, Applicants tested Cas9H840A with a set of sgRNA pairs designed to generate both 5' and 3' overhangs (offset range from –278 to +58 bp), but were unable to observe indel formation. Further work may be needed to identify the necessary design rules for sgRNA pairing to allow double nicking by Cas9H840A.

[0170] Additional delivery options for the brain include encapsulation of CRISPR enzyme and guide RNA in the form of either DNA or RNA into liposomes and conjugating to molecular Trojan horses for trans-blood brain barrier (BBB) delivery. Molecular Trojan horses have been shown to be effective for delivery of B-gal expression vectors into the brain of non-human primates. The same approach can be used to delivery vectors containing CRISPR enzyme and guide RNA. For instance, Xia CF and Boado RJ, Pardridge WM ("Antibody-mediated targeting of siRNA via the human insulin receptor using avidin-biotin technology." Mol Pharm. 2009 May-Jun;6(3):747-51. doi: 10.1021/mp800194) describes how delivery of short interfering RNA (siRNA) to cells in culture, and in vivo, is possible with combined use of a receptor-specific monoclonal antibody (mAb) and avidin-biotin technology. The authors also report that because the bond between the targeting mAb and the siRNA is stable with avidin-biotin technology, and RNAi effects at distant sites such as brain are observed in vivo following an intravenous administration of the targeted siRNA.

[0171] Zhang Y, Schlachetzki F, Pardridge WM. ("Global non-viral gene transfer to the primate brain following intravenous administration." Mol Ther. 2003 Jan;7(1):11-8.) describe how expression plasmids encoding reporters such as luciferase were encapsulated in the interior of an "artificial virus" comprised of an 85 nm pegylated immunoliposome, which was targeted to the rhesus monkey brain in vivo with a monoclonal antibody (MAb) to the human insulin receptor (HIR). The HIRMAb enables the liposome carrying the exogenous gene to undergo transcytosis across the blood-brain barrier and endocytosis across the neuronal plasma membrane following intravenous injection. The level of luciferase gene expression in the brain was 50-fold higher in the rhesus monkey as compared to the rat. Widespread neuronal expression of the beta-galactosidase gene in primate brain was demonstrated by both histochemistry and confocal microscopy. The authors indicate that this approach makes feasible reversible adult transgenics in 24 hours. Accordingly, the use of immunoliposome is preferred. These may be used in conjunction with antibodies to target specific tissues or cell surface proteins.

[0172] Other means of delivery or RNA are also preferred, such as 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, exozomes 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 siRNA is loaded into the exosomes.

[0173] Targeted deletion of genes is preferred. Examples are exemplified in Example 12. Preferred are, therefore, genes involved in cholesterol biosynthesis, fatty acid biosynthesis, and other metabolic disorders, genes encoding mis-folded proteins involved in amyloid and other diseases, oncogenes leading to cellular transformation, latent viral genes, and genes leading to dominant-negative disorders, amongst other disorders. As exemplified here, Applicants prefer gene delivery of a CRISPR-Cas system to the liver, brain, ocular, epithelial, hematopoetic, or another tissue of a subject or a patient in need thereof, suffering from metabolic disorders, amyloidosis and protein-aggregation related diseases, cellular transformation arising from genetic mutations and translocations, dominant negative effects of gene mutations, latent viral infections, and other related symptoms, using either viral or nanoparticle delivery system.

[0174] Therapeutic applications of the CRISPR-Cas system include Glaucoma, Amyloidosis, and Huntington's disease. These are exemplified in Example 14 and the features described therein are preferred alone or in combination.

[0175] As an example, chronic infection by HIV-1 may be treated or prevented. In order to accomplish this, one may generate CRISPR-Cas guide RNAs that target the vast majority of the HIV-1 genome while taking into account HIV-1 strain variants for maximal coverage and effectiveness. One may accomplish delivery of the CRISPR-Cas system by conventional

adenoviral or lentiviral-mediated infection of the host immune system. Depending on approach, host immune cells could be a) isolated, transduced with CRISPR-Cas, selected, and re-introduced in to the host or b) transduced in vivo by systemic delivery of the CRISPR-Cas system. The first approach allows for generation of a resistant immune population whereas the second is more likely to target latent viral reservoirs within the host. This is discussed in more detail in the Examples section.

[0176] It is also envisaged that the present invention generates a gene knockout cell library. Each cell may have a single gene knocked out. This is exemplified in Example 17.

[0177] One may make a library of ES cells where each cell has a single gene knocked out, and the entire library of ES cells will have every single gene knocked out. This library is useful for the screening of gene function in cellular processes as well as diseases. To make this cell library, one may integrate Cas9 driven by an inducible promoter (e.g. doxycycline inducible promoter) into the ES cell. In addition, one may integrate a single guide RNA targeting a specific gene in the ES cell. To make the ES cell library, one may simply mix ES cells with a library of genes encoding guide RNAs targeting each gene in the human genome. One may first introduce a single BxB1 attB site into the AAVS1 locus of the human ES cell. Then one may use the BxB1 integrase to facilitate the integration of individual guide RNA genes into the BxB1 attB site in AAVS1 locus. To facilitate integration, each guide RNA gene may be contained on a plasmid that carries of a single attP site. This way BxB1 will recombine the attB site in the genome with the attP site on the guide RNA containing plasmid. To generate the cell library, one may take the library of cells that have single guide RNAs integrated and induce Cas9 expression. After induction, Cas9 mediates double strand break at sites specified by the guide RNA.

[0178] Chronic administration of protein therapeutics may elicit unacceptable immune responses to the specific protein. The immunogenicity of protein drugs can be ascribed to a few immunodominant helper T lymphocyte (HTL) epitopes. Reducing the MHC binding affinity of these HTL epitopes contained within these proteins can generate drugs with lower immunogenicity (Tangri S, et al. ("Rationally engineered therapeutic proteins with reduced immunogenicity" J Immunol. 2005 Mar 15;174(6):3187-96.) In the present invention, the immunogenicity of the CRISPR enzyme in particular may be reduced following the approach first set out in Tangri et al with respect to erythropoietin and subsequently developed.

Accordingly, directed evolution or rational design may be used to reduce the immunogenicity of the CRISPR enzyme (for instance a Cas9) in the host species (human or other species).

[0179] In Example 28, Applicants used 3 guideRNAs of interest and able to visualize efficient DNA cleavage *in vivo* occurring only in a small subset of cells. Essentially, what Applicants have shown here is targeted *in vivo* cleavage. In particular, this provides proof of concept that specific targeting in higher organisms such as mammals can also be achieved. It also highlights multiplex aspect in that multiple guide sequences (i.e. separate targets) can be used simultaneously (in the sense of co-delivery). In other words, Applicants used a multiple approach, with several different sequences targeted at the same time, but independently.

[0180] A suitable example of a protocol for producing AAV, a preferred vector of the invention is provided in Example 29.

[0181] Trinucleotide repeat disorders are preferred conditions to be treated. These are also exemplified herein.

[0182] 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.

[0183] 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 such as those in Table B.

[0184] For an example of CFTRdelta508 chimeric guide RNA, see Example which demonstrates gene transfer or gene delivery of a CRISPR-Cas system in airways of subject or a patient in need thereof, suffering from cystic fibrosis or from cystic fibrosis (CF) related symptoms, using adeno-associated virus (AAV) particles. In particular, they exemplify a repair strategy for Cystic Fibrosis delta F508 mutation. This type of strategy should apply across all organisms. With particular reference to CF, suitable patients may include: Human, non-primate human, canine, feline, bovine, equine and other domestic animals. In this instance, Applicants utilized a CRISPR-Cas system comprising a Cas9 enzyme to target deltaF508 or other CFTR-inducing mutations.

[0185] The treated subjects in this instance receive pharmaceutically effective amount of aerosolized AAV vector system per lung endobronchially delivered while spontaneously breathing. As such, aerosolized delivery is preferred for AAV delivery in general. An adenovirus or an AAV particle may be used for delivery. Suitable gene constructs, each operably linked to one or more regulatory sequences, may be cloned into the delivery vector. In this instance, the following constructs are provided as examples: Cbh or EF1a promoter for Cas9, U6 or H1 promoter for chimeric guide RNA),: A preferred arrangement is to use a CFTRdelta508 targeting chimeric guide, a repair template for deltaF508 mutation and a codon optimized Cas9 enzyme (preferred Cas9s are those with nuclease or nickase activity) with optionally one or more nuclear localization signal or sequence(s) (NLS(s)), e.g., two (2) NLSs. Constructs without NLS are also envisaged.

[0186] In order to identify the Cas9 target site, Applicants analyzed the human CFTR genomic locus and identified the Cas9 target site. Preferably, in general and in this CF case, the PAM may contain a NGG or a NNAGAAW motif.

[0187] Accordingly, in the case of CF, the present method comprises manipulation of a target sequence in a genomic locus of interest comprising

[0188] delivering a non-naturally occurring or engineered composition comprising a viral vector system comprising one or more viral vectors operably encoding a composition for expression thereof, wherein the composition comprises:

[0189] a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising

[0190] I. a first regulatory element operably linked to a CRISPR-Cas system chimeric RNA (chiRNA) polynucleotide sequence, wherein the polynucleotide sequence comprises

[0191] (a) a guide sequence capable of hybridizing to the CF target sequence in a suitable mammalian cell,

[0192] (b) a tracr mate sequence, and

[0193] (c) a tracr sequence, and

[0194] 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,

[0195] wherein (a), (b) and (c) are arranged in a 5' to 3' orientation,

[0196] wherein components I and II are located on the same or different vectors of the system,

[0197] 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

[0198] wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence, and (2) the tracr mate sequence that is hybridized to the tracr sequence. In respect of CF, preferred target DNA sequences comprise the CFTRdelta508 mutation. A preferred PAM is described above. A preferred CRISPR enzyme is any Cas (described herein, but particularly that described in Example 16).

[0199] Alternatives to CF include any genetic disorder and examples of these are well known. Another preferred method or use of the invention is for correcting defects in the EMP2A and EMP2B genes that have been identified to be associated with Lafora disease.

[0200] In some embodiments, a "guide sequence" may be distinct from "guide RNA". A guide sequence may refer to an approx. 20bp sequence, within the guide RNA, that specifies the target site.

[0201] In some embodiments, the Cas9 is (or is derived from) the SpCas9. In such embodiments, preferred mutations are at any or all or positions 10, 762, 840, 854, 863 and/or 986 of SpCas9 or corresponding positions in other Cas9s (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. These are advantageous as they provide nickase activity.

[0202] It will be readily apparent that a host of other diseases can be treated in a similar fashion. Some examples of genetic diseases caused by mutations are provided herein, but many more are known. The above strategy can be applied to these diseases.

[0203] The invention uses nucleic acids to bind target DNA sequences. This is advantageous as nucleic acids are much easier and cheaper to produce 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.

[0204] 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., WO 97/03211; WO 96/39154. 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.

[0205] 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.

[0206] 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.

[0207] 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.

[0208] "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.

[0209] 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.

[0210] "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.

[0211] 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.

[0212] 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.

[0213] 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.

[0214] 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.

[0215] 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.

[0216] 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.

[0217] 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.

[0218] 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.

[0219] 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.

[0220] 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).

[0221] 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.

[0222] Alternatively, percentage homologies may be calculated using the multiple alignment feature in DNASIS™ (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.

[0223] 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

[0224] 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.

[0225] 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.

[0226] 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)).

[0227] In one aspect, the invention provides for vectors that are used in the engineering and optimization of CRISPR-Cas systems.

[0228] 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.

[0229] 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.

[0230] Aspects of the invention relate to 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 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.

[0231] In one embodiment, the U6 promoter may be a bi-directional promoter. Figure 52 depicts panels of data: the first showing different truncations of the Bi-directional U6 promoter and associated microdeletion activity and the second quantifying the indel-inducing activity coming off of either side of the bi-directional promoter. Figure 52 represents yet another means by which it is possible to deliver two independent sgRNAs off of a single U6 promoter that is independent of and complementary to the tandem sgRNA approach. The increase in size of the U6 promoter is minimal, which makes this amenable and attractive to in vivo/AAV delivery applications where vector size represents a significant constraint. The sequence of the smallest U6-8/U6-1 bi-directional reporter that retains good function in both directions which consists of joining end to end of the U6-1 promoter in the forward direction and a truncated version of the

U6-8 promoter in the reverse direction (see Figure 53). Additionally, this may address the issue that the first spacer tends to work significantly better than the second in the current iteration of the tandem sgRNA architecture as processing of independent sgRNAS does not have to occur with the BiU6 approach. See, e.g., WO2005035718 for additional disclosure on bi-directional promoters.

[0232] 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 I 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.

[0233] 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.

[0234] 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.

[0235] 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).

[0236] 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.).

[0237] 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).

[0238] 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.

[0239] 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.

[0240] 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.

[0241] In general, "CRISPR system" refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or other sequences and transcripts from a CRISPR locus. In embodiments of the invention the terms guide sequence and guide RNA are used interchangeably. In some embodiments, one or more elements of a CRISPR system is derived from a type I, type II, or type III 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 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.

[0242] In some embodiments, direct repeats may be identified *in silico* by searching for repetitive motifs that fulfill any or all of the following criteria:

[0243] 1. found in a 2Kb window of genomic sequence flanking the type II CRISPR locus;

[0244] 2. span from 20 to 50 bp; and

[0245] 3. interspaced by 20 to 50 bp.

[0246] 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.

[0247] In some embodiments, candidate tracrRNA may be subsequently predicted by sequences that fulfill any or all of the following criteria:

[0248] 1. sequence homology to direct repeats (motif search in Geneious with up to 18-bp mismatches);

[0249] 2. presence of a predicted Rho-independent transcriptional terminator in direction of transcription; and

[0250] 3. stable hairpin secondary structure between tracrRNA and direct repeat.

[0251] 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.

[0252] In some embodiments, chimeric synthetic guide RNAs (sgRNAs) designs may incorporate at least 12 bp of duplex structure between the direct repeat and tracrRNA.

[0253] 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. 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.

[0254] 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 (Figure 2A). 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 (Figure 2A). Figure 2B demonstrates the nuclear localization of the codon optimized Cas9. To promote precise transcriptional initiation, the RNA polymerase III-based U6 promoter was selected to drive the expression of tracrRNA (Figure 2C). Similarly, a U6 promoter-based construct was developed to express a pre-crRNA array consisting of a single spacer flanked by two direct repeats (DRs, also encompassed by the term "tracr-mate sequences"; Figure 2C). The initial spacer was designed to target a 33-base-pair (bp) target site (30-bp protospacer plus a 3-bp CRISPR motif (PAM) sequence satisfying the NGG recognition motif of Cas9) in the human EMX1 locus (Figure 2C), a key gene in the development of the cerebral cortex.

[0255] 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. 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. 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.

[0256] 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.

[0257] 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. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologues thereof, or modified versions thereof. In 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 less than about 25%, 10%, 5%, 1%, 0.1%, 0.01%, or lower with respect to its non-mutated form.

[0258] An aspartate-to-alanine substitution (D10A) in the RuvC I catalytic domain of SpCas9 was engineered to convert the nuclease into a nickase (SpCas9n) (see e.g. Sapranauskas et al., 2011, Nucleic Acis Research, 39: 9275; Gasiunas et al., 2012, Proc. Natl. Acad. Sci. USA, 109:E2579), such that nicked genomic DNA undergoes the high-fidelity homology-directed repair (HDR). Surveyor assay confirmed that SpCas9n does not generate indels at the EMX1 protospacer target. Co-expression of EMX1-targeting chimeric crRNA (having the tracrRNA component as well) with SpCas9 produced indels in the target site, whereas co-expression with SpCas9n did not (n=3). Moreover, sequencing of 327 amplicons did not detect any indels induced by SpCas9n. The same locus was selected to test CRISPR-mediated HR by co-transfecting HEK 293FT cells with the chimeric RNA targeting EMX1, hSpCas9 or hSpCas9n, as well as a HR template to introduce a pair of restriction sites (HindIII and NheI) near the protospacer.

[0259] Preferred orthologs are described herein. 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 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.

[0260] 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.

[0261] 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 is known.

[0262] 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, mouse, rat, rabbit, dog, 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.

[0263] 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/ (visited Jul. 9, 2002), 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.

[0264] 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. one or more NLS at the amino-terminus and 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.

[0265] 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.

[0266] 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. 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.

[0267] 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.

[0268] 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).

[0269] 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) NNNNNNNNNNNNNNNNNNNNgttttgtactctcaagatttaGAAAtaaatcttgcagaagctacaaagataa ggcttcatgccgaaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT; (2) NNNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccg

aaatcaacaccctgtcattttatggcagggtgttttcgttatttaaTTTTTT; (3)

NNNNNNNNNNNNNNNNNNNNgttttgtactctcaGAAAtgcagaagctacaaagataaggcttcatgccg

aaatcaacaccctgtcattttatggcagggtgtTTTTTT; (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.

[0270] 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.

[0271] 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.

[0272] 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 cryptochrome 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.

[0273] In some aspects, the invention provides methods comprising delivering one or more polynucleotides, such as or one or more vectors as described herein, one or more transcripts thereof, and/or one or proteins transcribed therefrom, to a host cell. In some aspects, the invention further provides cells produced by such methods, and animals comprising or produced from such cells. In some embodiments, a CRISPR enzyme in combination with (and optionally complexed with) a guide sequence is delivered to a cell. Conventional viral and non-viral based gene transfer methods can be used to introduce nucleic acids in mammalian cells or target tissues. Such methods can be used to administer nucleic acids encoding components of a CRISPR system to cells in culture, or in a host organism. Non-viral vector delivery systems include DNA plasmids, RNA (e.g. a transcript of a vector described herein), naked nucleic acid, and nucleic acid complexed with a delivery vehicle, such as a liposome. Viral vector delivery systems include DNA and RNA viruses, which have either episomal or integrated genomes after delivery to the cell. For a review of gene therapy procedures, see Anderson, Science 256:808-813 (1992); Nabel & Felgner, TIBTECH 11:211-217 (1993); Mitani & Caskey, TIBTECH 11:162-166 (1993); Dillon, TIBTECH 11:167-175 (1993); Miller, Nature 357:455-460 (1992); Van Brunt, Biotechnology 6(10):1149-1154 (1988); Vigne, Restorative Neurology and Neuroscience 8:35-36 (1995); Kremer & Perricaudet, British Medical Bulletin 51(1):31-44 (1995); Haddada et al., in Current Topics in Microbiology and Immunology Doerfler and Böhm (eds) (1995); and Yu et al., Gene Therapy 1:13-26 (1994).

[0274] Methods of non-viral delivery of nucleic acids include lipofection, microinjection, biolistics, virosomes, liposomes, immunoliposomes, polycation or lipid:nucleic acid conjugates, naked DNA, artificial virions, and agent-enhanced uptake of DNA. Lipofection is described in e.g., U.S. Pat. Nos. 5,049,386, 4,946,787; and 4,897,355) and lipofection reagents are sold commercially (e.g., Transfectam™ and Lipofectin™). Cationic and neutral lipids that are suitable for efficient receptor-recognition lipofection of polynucleotides include those of Felgner, WO 91/17424; WO 91/16024. Delivery can be to cells (e.g. in vitro or ex vivo administration) or target tissues (e.g. in vivo administration).

[0275] The preparation of lipid:nucleic acid complexes, including targeted liposomes such as immunolipid complexes, is well known to one of skill in the art (see, e.g., Crystal, Science 270:404-410 (1995); Blaese et al., Cancer Gene Ther. 2:291-297 (1995); Behr et al., Bioconjugate Chem. 5:382-389 (1994); Remy et al., Bioconjugate Chem. 5:647-654 (1994); Gao

et al., Gene Therapy 2:710-722 (1995); Ahmad et al., Cancer Res. 52:4817-4820 (1992); U.S. Pat. Nos. 4,186,183, 4,217,344, 4,235,871, 4,261,975, 4,485,054, 4,501,728, 4,774,085, 4,837,028, and 4,946,787).

[0276] The use of RNA or DNA viral based systems for the delivery of nucleic acids take advantage of highly evolved processes for targeting a virus to specific cells in the body and trafficking the viral payload to the nucleus. Viral vectors can be administered directly to patients (in vivo) or they can be used to treat cells in vitro, and the modified cells may optionally be administered to patients (ex vivo). Conventional viral based systems could include retroviral, lentivirus, adenoviral, adeno-associated and herpes simplex virus vectors for gene transfer. Integration in the host genome is possible with the retrovirus, lentivirus, and adeno-associated virus gene transfer methods, often resulting in long term expression of the inserted transgene. Additionally, high transduction efficiencies have been observed in many different cell types and target tissues.

[0277] The tropism of a retrovirus can be altered by incorporating foreign envelope proteins, expanding the potential target population of target cells. Lentiviral vectors are retroviral vectors that are able to transduce or infect non-dividing cells and typically produce high viral titers. Selection of a retroviral gene transfer system would therefore depend on the target tissue. Retroviral vectors are comprised of cis-acting long terminal repeats with packaging capacity for up to 6-10 kb of foreign sequence. The minimum cis-acting LTRs are sufficient for replication and packaging of the vectors, which are then used to integrate the therapeutic gene into the target cell to provide permanent transgene expression. Widely used retroviral vectors include those based upon murine leukemia virus (MuLV), gibbon ape leukemia virus (GaLV), Simian Immuno deficiency virus (SIV), human immuno deficiency virus (HIV), and combinations thereof (see, e.g., Buchscher et al., J. Virol. 66:2731-2739 (1992); Johann et al., J. Virol. 66:1635-1640 (1992); Sommnerfelt et al., Virol. 176:58-59 (1990); Wilson et al., J. Virol. 63:2374-2378 (1989); Miller et al., J. Virol. 65:2220-2224 (1991); PCT/US94/05700).

[0278] In applications where transient expression is preferred, adenoviral based systems may be used. Adenoviral based vectors are capable of very high transduction efficiency in many cell types and do not require cell division. With such vectors, high titer and levels of expression have been obtained. This vector can be produced in large quantities in a relatively simple system.

[0279] Adeno-associated virus ("AAV") vectors may also be used to transduce cells with target nucleic acids, e.g., in the in vitro production of nucleic acids and peptides, and for in vivo and ex vivo gene therapy procedures (see, e.g., West et al., Virology 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); Muzyczka, J. Clin. Invest. 94:1351 (1994). Construction of recombinant AAV vectors are described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., Mol. Cell. Biol. 5:3251-3260 (1985); Tratschin, et al., Mol. Cell. Biol. 4:2072-2081 (1984); Hermonat & Muzyczka, PNAS 81:6466-6470 (1984); and Samulski et al., J. Virol. 63:03822-3828 (1989).

[0280] Packaging cells are typically used to form virus particles that are capable of infecting a host cell. Such cells include 293 cells, which package adenovirus, and ψ2 cells or PA317 cells, which package retrovirus. Viral vectors used in gene therapy are usually generated by producer a cell line that packages a nucleic acid vector into a viral particle. The vectors typically contain the minimal viral sequences required for packaging and subsequent integration into a host, other viral sequences being replaced by an expression cassette for the polynucleotide(s) to be expressed. The missing viral functions are typically supplied in trans by the packaging cell line. For example, AAV vectors used in gene therapy typically only possess ITR sequences from the AAV genome which are required for packaging and integration into the host genome. Viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely rep and cap, but lacking ITR sequences. The cell line may also infected with adenovirus as a helper. The helper virus promotes replication of the AAV vector and expression of AAV genes from the helper plasmid. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, e.g., heat treatment to which adenovirus is more sensitive than AAV.

[0281] Accordingly, AAV is considered an ideal candidate for use as a transducing vector. Such AAV transducing vectors can comprise sufficient cis-acting functions to replicate in the presence of adenovirus or herpesvirus or poxvirus (e.g., vaccinia virus) helper functions provided in trans. Recombinant AAV (rAAV) can be used to carry exogenous genes into cells of a variety of lineages. In these vectors, the AAV cap and/or rep genes are deleted from the viral genome and replaced with a DNA segment of choice. Current AAV vectors may accommodate up to 4300 bases of inserted DNA.

[0282] There are a number of ways to produce rAAV, and the invention provides rAAV and methods for preparing rAAV. For example, plasmid(s) containing or consisting essentially of the desired construct are transfected into AAV-infected cells. In addition, a second or additional helper plasmid is cotransfected into these cells to provide the AAV rep and/or cap genes which are obligatory for replication and packaging of the recombinant viral construct. Under these conditions, the rep and/or cap proteins of AAV act in trans to stimulate replication and packaging of the rAAV construct. Two to Three days after transfection, rAAV is harvested. Traditionally rAAV is harvested from the cells along with adenovirus. The contaminating adenovirus is then inactivated by heat treatment. In the instant invention, rAAV is advantageously harvested not from the cells themselves, but from cell supernatant. Accordingly, in an initial aspect the invention provides for preparing rAAV, and in addition to the foregoing, rAAV can be prepared by a method that comprises or consists essentially of: infecting susceptible cells with a rAAV containing exogenous DNA including DNA for expression, and helper virus (e.g., adenovirus, herpesvirus, poxvirus such as vaccinia virus) wherein the rAAV lacks functioning cap and/or rep (and the helper virus (e.g., adenovirus, herpesvirus, poxvirus such as vaccinia virus) provides the cap and/or rev function that the rAAV lacks); or infecting susceptible cells with a rAAV containing exogenous DNA including DNA for expression, wherein the recombinant lacks functioning cap and/or rep, and transfecting said cells with a plasmid supplying cap and/or rep function that the rAAV lacks; or infecting susceptible cells with a rAAV containing exogenous DNA including DNA for expression, wherein the recombinant lacks functioning cap and/or rep, wherein said cells supply cap and/or rep function that the recombinant lacks; or transfecting the susceptible cells with an AAV lacking functioning cap and/or rep and plasmids for inserting exogenous DNA into the recombinant so that the exogenous DNA is expressed by the recombinant and for supplying rep and/or cap functions whereby transfection results in an rAAV containing the exogenous DNA including DNA for expression that lacks functioning cap and/or rep.

[0283] The rAAV can be from an AAV as herein described, and advantageously can be an rAAV1, rAAV2, AAV5 or rAAV having hybrid capsid which may comprise AAV1, AAV2, AAV5 or any combination thereof. One can select the AAV of the rAAV with regard to the cells to be targeted by the rAAV; 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.

[0284] In addition to 293 cells, other cells that can be used in the practice of the invention and the relative infectivity of certain AAV serotypes *in vitro* as to these cells (see Grimm, D. et al, J. Virol. 82: 5887-5911 (2008)) are 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

[0285] The invention provides rAAV that contains or consists essentially of an exogenous nucleic acid molecule encoding a CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) system, e.g., a plurality of cassettes comprising or consisting a first cassette comprising or consisting essentially of a promoter, a nucleic acid molecule encoding a CRISPR-associated (Cas) protein (putative nuclease or helicase proteins), e.g., Cas9 and a terminator, and a two, or more, advantageously up to the packaging size limit of the vector, e.g., in total (including the first cassette) five, cassettes comprising or consisting essentially of a promoter, nucleic acid molecule encoding guide RNA (gRNA) and a terminator (e.g., each cassette schematically represented as Promoter-gRNA1-terminator, Promoter-gRNA2-terminator ... Promoter-gRNA(N)-terminator (where N is a number that can be inserted that is at an upper limit of the packaging size limit of the vector), or two or more individual rAAVs, each containing one or more than one cassette of a CRISPR system, e.g., a first rAAV containing the first cassette comprising or consisting essentially of a promoter, a nucleic acid molecule encoding Cas, e.g.,

Cas9 and a terminator, and a second rAAV containing a plurality, four, cassettes comprising or consisting essentially of a promoter, nucleic acid molecule encoding guide RNA (gRNA) and a terminator (e.g., each cassette schematically represented as Promoter-gRNA1-terminator, Promoter-gRNA2-terminator ... Promoter-gRNA(N)-terminator (where N is a number that can be inserted that is at an upper limit of the packaging size limit of the vector). As rAAV is a DNA virus, the nucleic acid molecules in the herein discussion concerning AAV or rAAV are advantageously DNA. The promoter is in some embodiments advantageously human Synapsin I promoter (hSyn).

[0286] Additional methods for the delivery of nucleic acids to cells are known to those skilled in the art. See, for example, US20030087817, incorporated herein by reference.

[0287] In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In some embodiments, a cell is transfected as it naturally occurs in a subject. In some embodiments, a cell that is transfected is taken from a subject. In some embodiments, the cell is derived from cells taken from a subject, such as a cell line. A wide variety of cell lines for tissue culture are known in the art. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEKn, HEKa, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2, HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/ 3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr -/-, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN / OPCT

cell lines, Peer, PNT-1A / PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassus, Va.)). In some embodiments, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences. In some embodiments, a cell transiently transfected with the components of a CRISPR system as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a CRISPR complex, is used to establish a new cell line comprising cells containing the modification but lacking any other exogenous sequence. In some embodiments, cells transiently or non-transiently transfected with one or more vectors described herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[0288] In some embodiments, one or more vectors described herein are used to produce a non-human transgenic animal or transgenic plant. In some embodiments, the transgenic animal is a mammal, such as a mouse, rat, or rabbit. Methods for producing transgenic plants and animals are known in the art, and generally begin with a method of cell transfection, such as described herein.

[0289] With recent advances in crop genomics, the ability to use CRISPR-Cas systems to perform efficient and cost effective gene editing and manipulation will allow the rapid selection and comparison of single and and multiplexed genetic manipulations to transform such genomes for improved production and enhanced traits. In this regard reference is made to US patents and publications: US Patent No. 6,603,061 - Agrobacterium-Mediated Plant Transformation Method; US Patent No. 7,868,149 - Plant Genome Sequences and Uses Thereof and US 2009/0100536 - Transgenic Plants with Enhanced Agronomic Traits, all the contents and disclosure of each of which are herein incorporated by reference in their entirety. In the practice of the invention, the contents and disclosure of Morrell et al "Crop genomics: advances and applications" Nat Rev Genet. 2011 Dec 29;13(2):85-96 are also herein incorporated by reference in their entirety. In an advantageous embodiment of the invention, the CRISPR/Cas9 system is used to engineer microalgae. That the CRISPR-Cas system is able to be employed in plant systems is also provided in the manuscript "Efficient Genome Editing in Plants using a

CRISPR/Cas System", by Feng et al. submitted for publication to Nature Biotechnology in July 2013, incorporated by reference in its entirety, wherein it is demonstrated that engineered CRISPR/Cas complexes may be used to create double strand breaks at specific sites of the plant genome to achieve targeted genome modifications in both dicot and monocot plants. Accordingly, reference herein to animal cells may also apply, *mutatis mutandis,* to plant cells unless otherwise apparent.

[0290] 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 or plant (including micro-algae), 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 (including micro-algae). For re-introduced cells it is particularly preferred that the cells are stem cells.

[0291] 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.

[0292] 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.

[0293] 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. Similar considerations and conditions apply as above for methods of modifying a target polynucleotide.

[0294] 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.

[0295] 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.

[0296] 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.

[0297] 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.

[0298] 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) (Figure 11). 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.

[0299] 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.

[0300] 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.

[0301] 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.

[0302] 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.

[0303] 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).

[0304] In an exemplary method for modifying a target polynucleotide by integrating an exogenous polynucleotide template, a double stranded break is introduced into the genome sequence by the CRISPR complex, the break is repaired via homologous recombination by 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.

[0305] 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.

[0306] In some methods, a target polynucleotide can be inactivated to effect the modification of the expression in a cell. For example, upon the binding of a CRISPR complex to a target sequence in a cell, the target polynucleotide is inactivated such that the sequence is not transcribed, the coded protein is not produced, or the sequence does not function as the wild-type sequence does. For example, a protein or microRNA coding sequence may be inactivated such that the protein is not produced.

[0307] 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.

[0308] The inactivated target sequence may include a deletion mutation (i.e., deletion of one or more nucleotides), an insertion mutation (i.e., insertion of one or more nucleotides), or a nonsense mutation (i.e., substitution of a single nucleotide for another nucleotide such that a stop codon is introduced). In some methods, the inactivation of a target sequence results in "knock-out" of the target sequence.

[0309] A method of the invention may be used to create a plant, an animal or cell that may be used as a disease model. As used herein, "disease" refers to a disease, disorder, or indication in a subject. For example, a method of the invention may be used to create an animal or cell that comprises a modification in one or more nucleic acid sequences associated with a disease, or a plant, animal or cell in which the expression of one or more nucleic acid sequences associated with a disease are altered. Such a nucleic acid sequence may encode a disease associated protein sequence or may be a disease associated control sequence. Accordingly, it is understood that in embodiments of the invention, a plant, subject, patient, organism or cell can be a non-human subject, patient, organism or cell. Thus, the invention provides a plant, animal or cell, produced by the present methods, or a progeny thereof. The progeny may be a clone of the produced plant or animal, or may result from sexual reproduction by crossing with other individuals of the same species to introgress further desirable traits into their offspring. The cell may be *in vivo* or *ex vivo* in the cases of multicellular organisms, particularly animals or plants. In the instance where the cell is in cultured, a cell line may be established if appropriate culturing conditions are met and preferably if the cell is suitably adapted for this purpose (for instance a stem cell). Bacterial cell lines produced by the invention are also envisaged. Hence, cell lines are also envisaged.

[0310] In some methods, the disease model can be used to study the effects of mutations on the animal or cell and development and/or progression of the disease using measures commonly used in the study of the disease. Alternatively, such a disease model is useful for studying the effect of a pharmaceutically active compound on the disease.

[0311] In some methods, the disease model can be used to assess the efficacy of a potential gene therapy strategy. That is, a disease-associated gene or polynucleotide can be modified such that the disease development and/or progression is inhibited or reduced. In particular, the method comprises modifying a disease-associated gene or polynucleotide such that an altered

protein is produced and, as a result, the animal or cell has an altered response. Accordingly, in some methods, a genetically modified animal may be compared with an animal predisposed to development of the disease such that the effect of the gene therapy event may be assessed.

[0312] In another embodiment, this invention provides a method of developing a biologically active agent that modulates a cell signaling event associated with a disease gene. The method comprises contacting a test compound with a cell comprising one or more vectors that drive expression of one or more of a CRISPR enzyme, a guide sequence linked to a tracr mate sequence, and a tracr sequence; and detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with, e.g., a mutation in a disease gene contained in the cell.

[0313] A cell model or animal model can be constructed in combination with the method of the invention for screening a cellular function change. Such a model may be used to study the effects of a genome sequence modified by the CRISPR complex of the invention on a cellular function of interest. For example, a cellular function model may be used to study the effect of a modified genome sequence on intracellular signaling or extracellular signaling. Alternatively, a cellular function model may be used to study the effects of a modified genome sequence on sensory perception. In some such models, one or more genome sequences associated with a signaling biochemical pathway in the model are modified.

[0314] Several disease models have been specifically investigated. These include *de novo* autism risk genes CHD8, KATNAL2, and SCN2A; and the syndromic autism (Angelman Syndrome) gene UBE3A. These genes and resulting autism models are of course preferred, but serve to show the broad applicability of the invention across genes and corresponding models.

[0315] An altered expression of one or more genome sequences associated with a signaling biochemical pathway can be determined by assaying for a difference in the mRNA levels of the corresponding genes between the test model cell and a control cell, when they are contacted with a candidate agent. Alternatively, the differential expression of the sequences associated with a signaling biochemical pathway is determined by detecting a difference in the level of the encoded polypeptide or gene product.

[0316] To assay for an agent-induced alteration in the level of mRNA transcripts or corresponding polynucleotides, nucleic acid contained in a sample is first extracted according to standard methods in the art. For instance, mRNA can be isolated using various lytic enzymes or

chemical solutions according to the procedures set forth in Sambrook et al. (1989), or extracted by nucleic-acid-binding resins following the accompanying instructions provided by the manufacturers. The mRNA contained in the extracted nucleic acid sample is then detected by amplification procedures or conventional hybridization assays (e.g. Northern blot analysis) according to methods widely known in the art or based on the methods exemplified herein.

[0317] For purpose of this invention, amplification means any method employing a primer and a polymerase capable of replicating a target sequence with reasonable fidelity. Amplification may be carried out by natural or recombinant DNA polymerases such as TaqGold™, T7 DNA polymerase, Klenow fragment of E.coli DNA polymerase, and reverse transcriptase. A preferred amplification method is PCR. In particular, the isolated RNA can be subjected to a reverse transcription assay that is coupled with a quantitative polymerase chain reaction (RT-PCR) in order to quantify the expression level of a sequence associated with a signaling biochemical pathway.

[0318] Detection of the gene expression level can be conducted in real time in an amplification assay. In one aspect, the amplified products can be directly visualized with fluorescent DNA-binding agents including but not limited to DNA intercalators and DNA groove binders. Because the amount of the intercalators incorporated into the double-stranded DNA molecules is typically proportional to the amount of the amplified DNA products, one can conveniently determine the amount of the amplified products by quantifying the fluorescence of the intercalated dye using conventional optical systems in the art. DNA-binding dye suitable for this application include SYBR green, SYBR blue, DAPI, propidium iodine, Hoeste, SYBR gold, ethidium bromide, acridines, proflavine, acridine orange, acriflavine, fluorcoumanin, ellipticine, daunomycin, chloroquine, distamycin D, chromomycin, homidium, mithramycin, ruthenium polypyridyls, anthramycin, and the like.

[0319] In another aspect, other fluorescent labels such as sequence specific probes can be employed in the amplification reaction to facilitate the detection and quantification of the amplified products. Probe-based quantitative amplification relies on the sequence-specific detection of a desired amplified product. It utilizes fluorescent, target-specific probes (e.g., TaqMan® probes) resulting in increased specificity and sensitivity. Methods for performing probe-based quantitative amplification are well established in the art and are taught in U.S. Patent No. 5,210,015.

[0320] In yet another aspect, conventional hybridization assays using hybridization probes that share sequence homology with sequences associated with a signaling biochemical pathway can be performed. Typically, probes are allowed to form stable complexes with the sequences associated with a signaling biochemical pathway contained within the biological sample derived from the test subject in a hybridization reaction. It will be appreciated by one of skill in the art that where antisense is used as the probe nucleic acid, the target polynucleotides provided in the sample are chosen to be complementary to sequences of the antisense nucleic acids. Conversely, where the nucleotide probe is a sense nucleic acid, the target polynucleotide is selected to be complementary to sequences of the sense nucleic acid.

[0321] Hybridization can be performed under conditions of various stringency. Suitable hybridization conditions for the practice of the present invention are such that the recognition interaction between the probe and sequences associated with a signaling biochemical pathway is both sufficiently specific and sufficiently stable. Conditions that increase the stringency of a hybridization reaction are widely known and published in the art. See, for example, (Sambrook, et al., (1989); Nonradioactive In Situ Hybridization Application Manual, Boehringer Mannheim, second edition). The hybridization assay can be formed using probes immobilized on any solid support, including but are not limited to nitrocellulose, glass, silicon, and a variety of gene arrays. A preferred hybridization assay is conducted on high-density gene chips as described in U.S. Patent No. 5,445,934.

[0322] For a convenient detection of the probe-target complexes formed during the hybridization assay, the nucleotide probes are conjugated to a detectable label. Detectable labels suitable for use in the present invention include any composition detectable by photochemical, biochemical, spectroscopic, immunochemical, electrical, optical or chemical means. A wide variety of appropriate detectable labels are known in the art, which include fluorescent or chemiluminescent labels, radioactive isotope labels, enzymatic or other ligands. In preferred embodiments, one will likely desire to employ a fluorescent label or an enzyme tag, such as digoxigenin, ß-galactosidase, urease, alkaline phosphatase or peroxidase, avidin/biotin complex.

[0323] The detection methods used to detect or quantify the hybridization intensity will typically depend upon the label selected above. For example, radiolabels may be detected using photographic film or a phosphoimager. Fluorescent markers may be detected and quantified using a photodetector to detect emitted light. Enzymatic labels are typically detected by

providing the enzyme with a substrate and measuring the reaction product produced by the action of the enzyme on the substrate; and finally colorimetric labels are detected by simply visualizing the colored label.

[0324] An agent-induced change in expression of sequences associated with a signaling biochemical pathway can also be determined by examining the corresponding gene products. Determining the protein level typically involves a) contacting the protein contained in a biological sample with an agent that specifically bind to a protein associated with a signaling biochemical pathway; and (b) identifying any agent:protein complex so formed. In one aspect of this embodiment, the agent that specifically binds a protein associated with a signaling biochemical pathway is an antibody, preferably a monoclonal antibody.

[0325] The reaction is performed by contacting the agent with a sample of the proteins associated with a signaling biochemical pathway derived from the test samples under conditions that will allow a complex to form between the agent and the proteins associated with a signaling biochemical pathway. The formation of the complex can be detected directly or indirectly according to standard procedures in the art. In the direct detection method, the agents are supplied with a detectable label and unreacted agents may be removed from the complex; the amount of remaining label thereby indicating the amount of complex formed. For such method, it is preferable to select labels that remain attached to the agents even during stringent washing conditions. It is preferable that the label does not interfere with the binding reaction. In the alternative, an indirect detection procedure may use an agent that contains a label introduced either chemically or enzymatically. A desirable label generally does not interfere with binding or the stability of the resulting agent:polypeptide complex. However, the label is typically designed to be accessible to an antibody for an effective binding and hence generating a detectable signal.

[0326] A wide variety of labels suitable for detecting protein levels are known in the art. Non-limiting examples include radioisotopes, enzymes, colloidal metals, fluorescent compounds, bioluminescent compounds, and chemiluminescent compounds.

[0327] The amount of agent:polypeptide complexes formed during the binding reaction can be quantified by standard quantitative assays. As illustrated above, the formation of agent:polypeptide complex can be measured directly by the amount of label remained at the site of binding. In an alternative, the protein associated with a signaling biochemical pathway is

tested for its ability to compete with a labeled analog for binding sites on the specific agent. In this competitive assay, the amount of label captured is inversely proportional to the amount of protein sequences associated with a signaling biochemical pathway present in a test sample.

[0328] A number of techniques for protein analysis based on the general principles outlined above are available in the art. They include but are not limited to radioimmunoassays, ELISA (enzyme linked immunoradiometric assays), "sandwich" immunoassays, immunoradiometric assays, in situ immunoassays (using e.g., colloidal gold, enzyme or radioisotope labels), western blot analysis, immunoprecipitation assays, immunofluorescent assays, and SDS-PAGE.

[0329] Antibodies that specifically recognize or bind to proteins associated with a signaling biochemical pathway are preferable for conducting the aforementioned protein analyses. Where desired, antibodies that recognize a specific type of post-translational modifications (e.g., signaling biochemical pathway inducible modifications) can be used. Post-translational modifications include but are not limited to glycosylation, lipidation, acetylation, and phosphorylation. These antibodies may be purchased from commercial vendors. For example, anti-phosphotyrosine antibodies that specifically recognize tyrosine-phosphorylated proteins are available from a number of vendors including Invitrogen and Perkin Elmer. Anti-phosphotyrosine antibodies are particularly useful in detecting proteins that are differentially phosphorylated on their tyrosine residues in response to an ER stress. Such proteins include but are not limited to eukaryotic translation initiation factor 2 alpha (eIF-2α). Alternatively, these antibodies can be generated using conventional polyclonal or monoclonal antibody technologies by immunizing a host animal or an antibody-producing cell with a target protein that exhibits the desired post-translational modification.

[0330] In practicing the subject method, it may be desirable to discern the expression pattern of an protein associated with a signaling biochemical pathway in different bodily tissue, in different cell types, and/or in different subcellular structures. These studies can be performed with the use of tissue-specific, cell-specific or subcellular structure specific antibodies capable of binding to protein markers that are preferentially expressed in certain tissues, cell types, or subcellular structures.

[0331] An altered expression of a gene associated with a signaling biochemical pathway can also be determined by examining a change in activity of the gene product relative to a control cell. The assay for an agent-induced change in the activity of a protein associated with a

signaling biochemical pathway will dependent on the biological activity and/or the signal transduction pathway that is under investigation. For example, where the protein is a kinase, a change in its ability to phosphorylate the downstream substrate(s) can be determined by a variety of assays known in the art. Representative assays include but are not limited to immunoblotting and immunoprecipitation with antibodies such as anti-phosphotyrosine antibodies that recognize phosphorylated proteins. In addition, kinase activity can be detected by high throughput chemiluminescent assays such as AlphaScreen™ (available from Perkin Elmer) and eTag™ assay (Chan-Hui, et al. (2003) Clinical Immunology 111: 162-174).

[0332] Where the protein associated with a signaling biochemical pathway is part of a signaling cascade leading to a fluctuation of intracellular pH condition, pH sensitive molecules such as fluorescent pH dyes can be used as the reporter molecules. In another example where the protein associated with a signaling biochemical pathway is an ion channel, fluctuations in membrane potential and/or intracellular ion concentration can be monitored. A number of commercial kits and high-throughput devices are particularly suited for a rapid and robust screening for modulators of ion channels. Representative instruments include FLIPRTM (Molecular Devices, Inc.) and VIPR (Aurora Biosciences). These instruments are capable of detecting reactions in over 1000 sample wells of a microplate simultaneously, and providing real-time measurement and functional data within a second or even a minisecond.

[0333] 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.

[0334] 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).

[0335] 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.

[0336] 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).

[0337] The target polynucleotide of a CRISPR complex may include a number of disease-associated genes and polynucleotides as well as signaling biochemical pathway-associated genes and polynucleotides as listed in US provisional patent applications 61/736,527 and 61/748,427 having Broad reference BI-2011/008/WSGR Docket No. 44063-701.101 and BI-2011/008/WSGR Docket No. 44063-701.102 respectively, both entitled SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION filed on December 12, 2012 and January 2, 2013, respectively, the contents of all of which are herein incorporated by reference in their entirety.

[0338] 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.

[0339] Examples of disease-associated genes and polynucleotides are listed in Tables A and B. Disease specific information is available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web. Examples of signaling biochemical pathway-associated genes and polynucleotides are listed in Table C.

[0340] Mutations in these genes and pathways can result in production of improper proteins or proteins in improper amounts which affect function. Further examples of genes, diseases and proteins are hereby incorporated by reference from US Provisional applications 61/736,527 and 61/748,427. Such genes, proteins and pathways may be the target polynucleotide of a CRISPR complex.

Table A

DISEASE/DISORDERS	GENE(S)
Neoplasia	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4;
	Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF;
	HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alpha; PPAR
	gamma; WT1 (Wilms Tumor); FGF Receptor Family
	members (5 members: 1, 2, 3, 4, 5); CDKN2a; APC; RB
	(retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR
	(Androgen Receptor); TSG101; IGF; IGF Receptor; Igf1 (4
	variants); Igf2 (3 variants); Igf 1 Receptor; Igf 2 Receptor;
	Bax; Bcl2; caspases family (9 members:
	1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
Age-related Macular	Abcr; Ccl2; Cc2; cp (ceruloplasmin); Timp3; cathepsinD;
Degeneration	Vldlr; Ccr2
Schizophrenia	Neuregulin1 (Nrg1); Erb4 (receptor for Neuregulin);

	Complexin1 (Cplx1); Tph1 Tryptophan hydroxylase; Tph2
	Tryptophan hydroxylase 2; Neurexin 1; GSK3; GSK3a;
	GSK3b
Disorders	5-HTT (Slc6a4); COMT; DRD (Drd1a); SLC6A3; DAOA;
	DTNBP1; Dao (Dao1)
Trinucleotide Repeat	HTT (Huntington's Dx); SBMA/SMAX1/AR (Kennedy's
Disorders	Dx); FXN/X25 (Friedrich's Ataxia); ATX3 (Machado-
	Joseph's Dx); ATXN1 and ATXN2 (spinocerebellar
	ataxias); DMPK (myotonic dystrophy); Atrophin-1 and Atn1
	(DRPLA Dx); CBP (Creb-BP - global instability); VLDLR
	(Alzheimer's); Atxn7; Atxn10
Fragile X Syndrome	FMR2; FXR1; FXR2; mGLUR5
Secretase Related	APH-1 (alpha and beta); Presenilin (Psen1); nicastrin
Disorders	(Ncstn); PEN-2
Others	Nos1; Parp1; Nat1; Nat2
Prion - related disorders	Prp
ALS	SOD1; ALS2; STEX; FUS; TARDBP; VEGF (VEGF-a;
	VEGF-b; VEGF-c)
Drug addiction	Prkce (alcohol); Drd2; Drd4; ABAT (alcohol); GRIA2;
	Grm5; Grin1; Htr1b; Grin2a; Drd3; Pdyn; Gria1 (alcohol)
Autism	Mecp2; BZRAP1; MDGA2; Sema5A; Neurexin 1; Fragile X
	(FMR2 (AFF2); FXR1; FXR2; Mglur5)
Alzheimer's Disease	E1; CHIP; UCH; UBB; Tau; LRP; PICALM; Clusterin; PS1;
	SORL1; CR1; Vldlr; Uba1; Uba3; CHIP28 (Aqp1,
	Aquaporin 1); Uchl1; Uchl3; APP
Inflammation	IL-10; IL-1 (IL-1a; IL-1b); IL-13; IL-17 (IL-17a (CTLA8); IL-
	17b; IL-17c; IL-17d; IL-17f); Il-23; Cx3cr1; ptpn22; TNFa;
	NOD2/CARD15 for IBD; IL-6; IL-12 (IL-12a; IL-12b);
	CTLA4; Cx3cl1
Parkinson's Disease	x-Synuclein; DJ-1; LRRK2; Parkin; PINK1

Table B:

Blood and coagulation diseases and disorders	Anemia (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); Bare lymphocyte syndrome (TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5), Bleeding disorders (TBXA2R, P2RX1, P2X1); Factor H and factor H-like 1 (HF1, CFH, HUS); Factor V and factor VIII (MCFD2); Factor VII deficiency (F7); Factor X deficiency (F10); Factor XI deficiency (F11); Factor XII deficiency (F12, HAF); Factor XIIIA deficiency (F13A1, F13A); Factor XIIIB deficiency (F13B); Fanconi anemia (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1,

	BACH1, FANCJ, PHF9, FANCL, FANCM, KIAA1596); Hemophagocytic lymphohistiocytosis disorders (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemophilia A (F8, F8C, HEMA); Hemophilia B (F9, HEMB), Hemorrhagic disorders (PI, ATT, F5); Leukocyde deficiencies and disorders (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2, EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); Sickle cell anemia (HBB); Thalassemia (HBA2, HBB, HBD, LCRB, HBA1).
Cell dysregulation and oncology diseases and disorders	B-cell non-Hodgkin lymphoma (BCL7A, BCL7); Leukemia (TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL11, ARLTS1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN).
Inflammation and immune related diseases and disorders	AIDS (KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1); Autoimmune lymphoproliferative syndrome (TNFRSF6, APT1, FAS, CD95, ALPS1A); Combined immunodeficiency, (IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228), HIV susceptibility or infection (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCKR5 (CCR5)); Immunodeficiencies (CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI); Inflammation (IL-10, IL-1 (IL-1a, IL-1b), IL-13, IL-17 (IL-17a (CTLA8), IL-17b, IL-17c, IL-17d, IL-17f), Il-23, Cx3cr1, ptpn22, TNFa, NOD2/CARD15 for IBD, IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3cl1); Severe combined immunodeficiencies (SCIDs)(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).
Metabolic, liver, kidney and protein diseases and disorders	Amyloid neuropathy (TTR, PALB); Amyloidosis (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); Cirrhosis (KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); Cystic fibrosis (CFTR, ABCC7, CF, MRP7); Glycogen storage diseases (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Hepatic adenoma, 142330 (TCF1, HNF1A, MODY3), Hepatic failure, early onset, and neurologic disorder (SCOD1, SCO1), Hepatic lipase deficiency (LIPC), Hepatoblastoma, cancer and carcinomas (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5;

	Medullary cystic kidney disease (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); Phenylketonuria (PAH, PKU1, QDPR, DHPR, PTS); Polycystic kidney and hepatic disease (FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63).
Muscular / Skeletal diseases and disorders	Becker muscular dystrophy (DMD, BMD, MYF6), Duchenne Muscular Dystrophy (DMD, BMD); Emery-Dreifuss muscular dystrophy (LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); Facioscapulohumeral muscular dystrophy (FSHMD1A, FSHD1A); Muscular dystrophy (FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3, LGMD1C, SEPN1, SELN, RSMD1, PLEC1, PLTN, EBS1); Osteopetrosis (LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); Muscular atrophy (VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1).
Neurological and neuronal diseases and disorders	ALS (SOD1, ALS2, STEX, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c); Alzheimer disease (APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLAU, URK, ACE, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); Autism (Mecp2, BZRAP1, MDGA2, Sema5A, Neurexin 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); Fragile X Syndrome (FMR2, FXR1, FXR2, mGLUR5); Huntington's disease and disease like disorders (HD, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); Parkinson disease (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK5, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); Rett syndrome (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Synuclein, DJ-1); Schizophrenia (Neuregulin1 (Nrg1), Erb4 (receptor for Neuregulin), Complexin1 (Cplx1), Tph1 Tryptophan hydroxylase, Tph2, Tryptophan hydroxylase 2, Neurexin 1, GSK3, GSK3a, GSK3b, 5-HTT (Slc6a4), COMT, DRD (Drd1a), SLC6A3, DAOA, DTNBP1, Dao (Dao1)); Secretase Related Disorders (APH-1 (alpha and beta), Presenilin (Psen1), nicastrin, (Ncstn), PEN-2, Nos1, Parp1, Nat1, Nat2); Trinucleotide Repeat Disorders (HTT (Huntington's Dx), SBMA/SMAX1/AR (Kennedy's Dx), FXN/X25 (Friedrich's Ataxia), ATX3 (Machado- Joseph's Dx), ATXN1 and ATXN2 (spinocerebellar ataxias), DMPK (myotonic dystrophy), Atrophin-1 and

	Atn1 (DRPLA Dx), CBP (Creb-BP - global instability), VLDLR (Alzheimer's), Atxn7, Atxn10).
Occular diseases and disorders	Age-related macular degeneration (Aber, Ccl2, Cc2, cp (ceruloplasmin), Timp3, cathepsinD, Vldlr, Ccr2); Cataract (CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2, CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); Corneal clouding and dystrophy (APOA1, TGFBI, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD); Cornea plana congenital (KERA, CNA2); Glaucoma (MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A); Leber congenital amaurosis (CRB1, RP12, CRX, CORD2, CRD, RPGRIP1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3); Macular dystrophy (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2).

Table C:

CELLULAR FUNCTION	GENES
PI3K/AKT Signaling	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2;
	PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1;
	AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2;
	PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2;
	ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3;
	PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7;
	YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A;
	CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1;
	CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1;
	PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2;
	TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK;
	HSP90AA1; RPS6KB1
ERK/MAPK Signaling	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2;
	EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6;
	MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1;
	PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A;
	PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN;
	EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC;
	CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ;
	PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1;

	MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1;
	PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1;
	CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
Glucocorticoid Receptor	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1;
Signaling	MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I;
	PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2;
	MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1;
	MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13;
	RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1;
	PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3;
	MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP;
	CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2;
	PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1;
	ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1;
	STAT1; IL6; HSP90AA1
Axonal Guidance Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12;
	IGF1; RAC1; RAP1A; EIF4E; PRKCZ; NRP1; NTRK2;
	ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2;
	PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2;
	CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11;
	PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1;
	FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1;
	GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3;
	CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B;
	AKT3; PRKCA
Ephrin Receptor Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1;
	PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2;
	MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2;
	DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14;
	CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1;
	KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2;
	PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10;
	MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2;
	EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4;
	AKT3; SGK
Actin Cytoskeleton	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1;
Signaling	PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6;
	ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8;
	PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8;
	F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD;
	PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7;

	PPP1CC; PXN; VIL2; RAF1; GSN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3;
	ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL;
	BRAF; VAV3; SGK
Huntington's Disease	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2;
Signaling	MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2;
	PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST;
	GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1;
	GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2;
	HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A;
	HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1;
	PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX;
	ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3
Apoptosis Signaling	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1;
	BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB;
	CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8;
	BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF;
	RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2;
	CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2;
	BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK;
	CASP3; BIRC3; PARP1
B Cell Receptor Signaling	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11;
	AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1;
	MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9;
	EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB;
	MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1;
	NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN;
	GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Leukocyte Extravasation	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA;
Signaling	RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11;
	MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12;
	PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB;
	MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK;
	MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2;
	CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK;
	CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
Integrin Signaling	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A;
	TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2;
	CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8;
	CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP;
	RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1;

	TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2;
	CRKL; BRAF; GSK3B; AKT3
Acute Phase Response	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11;
Signaling	AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14;
	PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS;
	MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1;
	TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1;
	IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1;
	CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN;
	AKT3; IL1R1; IL6
PTEN Signaling	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11;
	MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA;
	CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1;
	MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR;
	RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2;
	AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1;
	NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2;
	GSK3B; AKT3; FOXO1; CASP3; RPS6KB1
p53 Signaling	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A;
	BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2;
	PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1;
	PMAIP1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9;
	CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A;
	HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1;
	SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN;
	SNAI2; GSK3B; BAX; AKT3
Aryl Hydrocarbon Receptor	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1;
Signaling	NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1;
	SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1;
	MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1;
	SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF;
	CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1;
	CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1;
	HSP90AA1
Xenobiotic Metabolism	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1;
Signaling	NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A;
	PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1;
	ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD;
	GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL;
	NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1;
	CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1;
	NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1;
	HSP90AA1
SAPK/JNK Signaling	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1;

	GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA;
	FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1;
	GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A;
	TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2;
	PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1;
	CRKL; BRAF; SGK
PPAr/RXR Signaling	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN;
	RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2;
	ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8;
	IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A;
	NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7;
	CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1;
	TGFBR1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1;
	ADIPOQ
NF-KB Signaling	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAF6;
	TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2;
	KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF;
	INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1;
	PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10;
	GSK3B; AKT3; TNFAIP3; IL1R1
Neuregulin Signaling	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1;
	MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI;
	CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS;
	PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2;
	ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3;
	EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL;
	AKT3; PRKCA; HSP90AA1; RPS6KB1
Wnt & Beta catenin	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO;
Signaling	AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A;
	WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK;
	LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1;
	PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1; CCND1;
	GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B;
	AKT3; SOX2
Insulin Receptor Signaling	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1;
	PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3;
	MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1;
	SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN;
	MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1;
	GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK;
	RPS6KB1
IL-6 Signaling	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11;

	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3;
	MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1;
	MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG;
	RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3;
	MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6
Hepatic Cholestasis	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA;
	RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8;
	PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1;
	TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8;
	CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4;
	JUN; IL1R1; PRKCA; IL6
IGF-1 Signaling	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2;
	PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8;
	IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A;
	YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1;
	PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3;
	FOXO1; SRF; CTGF; RPS6KB1
NRF2-mediated Oxidative	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1;
Stress Response	NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL;
	NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP;
	MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1;
	GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
Hepatic Fibrosis/Hepatic	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF;
Stellate Cell Activation	SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9;
	IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8;
	PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX;
	IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
PPAR Signaling	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB;
	NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3;
	NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2;
	PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG;
	RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA;
	MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Fc Epsilon RI Signaling	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11;
	AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD;
	MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN;
	MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3;
	VAV3; PRKCA
G-Protein Coupled	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB;
Receptor Signaling	PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB;
	PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1;
	IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK;

	PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3;
	PRKCA
Inositol Phosphate	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6;
Metabolism	MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2;
	PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1;
	MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
PDGF Signaling	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB;
	PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC;
	PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2;
	PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC;
	JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
VEGF Signaling	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF;
	AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3;
	BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN;
	RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN;
	VEGFA; AKT3; FOXO1; PRKCA
Natural Killer Cell Signaling	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11;
	KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB;
	PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6;
	PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1;
	PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
Cell Cycle: G1/S	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC;
Checkpoint Regulation	ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11;
	HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1;
	E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1;
	GSK3B; RBL1; HDAC6
T Cell Receptor Signaling	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS;
	NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN;
	MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10;
	JUN; VAV3
Death Receptor Signaling	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD;
	FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8;
	DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB;
	CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3;
	BIRC3
FGF Signaling	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11;
	AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8;
	MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1;
	AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4;
	AKT3; PRKCA; HGF
GM-CSF Signaling	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A;

	STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3;
	ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3;
	STAT1
Amyotrophic Lateral	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2;
Sclerosis Signaling	PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1;
	PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1;
	APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
JAK/Stat Signaling	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B;
	PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A;
	PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3;
	STAT1
Nicotinate and Nicotinamide	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1;
Metabolism	PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1;
	PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2;
	MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
Chemokine Signaling	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ;
	CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13;
	RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1;
	MAP2K2; MAP2K1; JUN; CCL2; PRKCA
IL-2 Signaling	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS;
	STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2;
	JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
Synaptic Long Term	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS;
Depression	PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3;
	KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA;
	YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Estrogen Receptor	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2;
Signaling	SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1;
	HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP;
	MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Protein Ubiquitination	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4;
Pathway	CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7;
	USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8;
	USP1; VHL; HSP90AA1; BIRC3
IL-10 Signaling	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2;
	MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF;
	IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1;
	JUN; IL1R1; IL6
VDR/RXR Activation	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1;
	NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD;
	RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1;

	LRP5; CEBPB; FOXO1; PRKCA
TGF-beta Signaling	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1;
	FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2;
	SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2;
	MAP2K1; TGFBR1; SMAD4; JUN; SMAD5
Toll-like Receptor Signaling	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13;
	RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK;
	NFKB1; TLR2; JUN
p38 MAPK Signaling	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS;
	CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2;
	MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1;
	SRF; STAT1
Neurotrophin/TRK Signaling	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS;
	PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A;
	RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1;
	CDC42; JUN; ATF4
FXR/RXR Activation	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8;
	APOB; MAPK10; PPARG; MTTP; MAPK9; PPARGC1A;
	TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
Synaptic Long Term Potentiation	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1;
	PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS;
	PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1;
	ATF4; PRKCA
Calcium Signaling	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1;
	CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11;
	HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4;
	HDAC6
EGF Signaling	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1;
	STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Hypoxia Signaling in the Cardiovascular System	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT;
	HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM;
	VEGFA; JUN; ATF4; VHL; HSP90AA1
LPS/IL-1 Mediated Inhibition of RXR Function	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1;
	MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2;
	TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
LXR/RXR Activation	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA;
	NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1;
	SREBF1; IL1R1; CCL2; IL6; MMP9
Amyloid Processing	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2;
	CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1;

	PSEN1; CSNK1A1; GSK3B; AKT3; APP
IL-4 Signaling	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1;
	PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1;
	FRAP1; AKT3; RPS6KB1
Cell Cycle: G2/M DNA	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC;
Damage Checkpoint	CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A;
Regulation	PRKDC; ATM; SFN; CDKN2A
Nitric Oxide Signaling in the	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3;
Cardiovascular System	CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1;
	VEGFA; AKT3; HSP90AA1
Purine Metabolism	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4;
	PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C;
	NT5E; POLD1; NME1
cAMP-mediated Signaling	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3;
	SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Mitochondrial Dysfunction	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9;
	PARK7; PSEN1; PARK2; APP; CASP3
Notch Signaling	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2;
	PSEN1; NOTCH3; NOTCH1; DLL4
Endoplasmic Reticulum	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4;
Stress Pathway	EIF2AK3; CASP3
Pyrimidine Metabolism	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B;
	NT5E; POLD1; NME1
Parkinson's Signaling	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7;
	PARK2; CASP3
Cardiac & Beta Adrenergic	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC;
Signaling	PPP2R5C
Glycolysis/Gluconeogenesis	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Interferon Signaling	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Sonic Hedgehog Signaling	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B
Glycerophospholipid	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Metabolism	
Phospholipid Degradation	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Tryptophan Metabolism	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Lysine Degradation	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Nucleotide Excision Repair	ERCC5; ERCC4; XPA; XPC; ERCC1
Pathway	

Starch and Sucrose	UCHL1; HK2; GCK; GPI; HK1
Metabolism	
Aminosugars Metabolism	NQO1; HK2; GCK; HK1
Arachidonic Acid	PRDX6; GRN; YWHAZ; CYP1B1
Metabolism	
Circadian Rhythm Signaling	CSNK1E; CREB1; ATF4; NR1D1
Coagulation System	BDKRB1; F2R; SERPINE1; F3
Dopamine Receptor	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Signaling	
Glutathione Metabolism	IDH2; GSTP1; ANPEP; IDH1
Glycerolipid Metabolism	ALDH1A1; GPAM; SPHK1; SPHK2
Linoleic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Methionine Metabolism	DNMT1; DNMT3B; AHCY; DNMT3A
Pyruvate Metabolism	GLO1; ALDH1A1; PKM2; LDHA
Arginine and Proline	ALDH1A1; NOS3; NOS2A
Metabolism	
Eicosanoid Signaling	PRDX6; GRN; YWHAZ
Fructose and Mannose	HK2; GCK; HK1
Metabolism	
Galactose Metabolism	HK2; GCK; HK1
Stilbene, Coumarine and	PRDX6; PRDX1; TYR
Lignin Biosynthesis	
Antigen Presentation	CALR; B2M
Pathway	
Biosynthesis of Steroids	NQO1; DHCR7
Butanoate Metabolism	ALDH1A1; NLGN1
Citrate Cycle	IDH2; IDH1
Fatty Acid Metabolism	ALDH1A1; CYP1B1
Glycerophospholipid	PRDX6; CHKA
Metabolism	
Histidine Metabolism	PRMT5; ALDH1A1
Inositol Metabolism	ERO1L; APEX1
Metabolism of Xenobiotics	GSTP1; CYP1B1
by Cytochrome p450	
Methane Metabolism	PRDX6; PRDX1
Phenylalanine Metabolism	PRDX6; PRDX1
Propanoate Metabolism	ALDH1A1; LDHA
Selenoamino Acid	PRMT5; AHCY
Metabolism	
Sphingolipid Metabolism	SPHK1; SPHK2
Aminophosphonate	PRMT5

Metabolism	
Androgen and Estrogen	PRMT5
Metabolism	
Ascorbate and Aldarate	ALDH1A1
Metabolism	
Bile Acid Biosynthesis	ALDH1A1
Cysteine Metabolism	LDHA
Fatty Acid Biosynthesis	FASN
Glutamate Receptor	GNB2L1
Signaling	
NRF2-mediated Oxidative	PRDX1
Stress Response	
Pentose Phosphate	GPI
Pathway	
Pentose and Glucuronate	UCHL1
Interconversions	
Retinol Metabolism	ALDH1A1
Riboflavin Metabolism	TYR
Tyrosine Metabolism	PRMT5, TYR
Ubiquinone Biosynthesis	PRMT5
Valine, Leucine and	ALDH1A1
Isoleucine Degradation	
Glycine, Serine and	CHKA
Threonine Metabolism	
Lysine Degradation	ALDH1A1
Pain/Taste	TRPM5; TRPA1
Pain	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2;
	Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca;
	Prkacb; Prkar1a; Prkar2a
Mitochondrial Function	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Developmental Neurology	BMP-4; Chordin (Chrd); Noggin (Nog); WNT (Wnt2;
	Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b;
	Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenin;
	Dkk-1; Frizzled related proteins; Otx-2; Gbx2; FGF-8;
	Reelin; Dab1; unc-86 (Pou4f1 or Brn3a); Numb; Reln

[0341] Embodiments of the invention also relate to methods and compositions related to knocking out genes, amplifying genes and repairing particular mutations associated with DNA repeat instability and neurological disorders (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Second Edition, Academic Press, Oct 13, 2011 – Medical). Specific aspects of tandem repeat sequences have been found to be responsible for

more than twenty human diseases (New insights into repeat instability: role of RNA•DNA hybrids. McIvor EI, Polak U, Napierala M. RNA Biol. 2010 Sep-Oct;7(5):551-8). The CRISPR-Cas system may be harnessed to correct these defects of genomic instability.

[0342] A further aspect of the invention relates to utilizing the CRISPR-Cas system for correcting defects in the EMP2A and EMP2B genes that have been identified to be associated with Lafora disease. Lafora disease is an autosomal recessive condition which is characterized by progressive myoclonus epilepsy which may start as epileptic seizures in adolescence. A few cases of the disease may be caused by mutations in genes yet to be identified. The disease causes seizures, muscle spasms, difficulty walking, dementia, and eventually death. There is currently no therapy that has proven effective against disease progression. Other genetic abnormalities associated with epilepsy may also be targeted by the CRISPR-Cas system and the underlying genetics is further described in Genetics of Epilepsy and Genetic Epilepsies, edited by Giuliano Avanzini, Jeffrey L. Noebels, Mariani Foundation Paediatric Neurology:20; 2009).

[0343] In yet another aspect of the invention, the CRISPR-Cas system may be used to correct ocular defects that arise from several genetic mutations further described in Genetic Diseases of the Eye, Second Edition, edited by Elias I. Traboulsi, Oxford University Press, 2012.

[0344] Several further aspects of the invention relate to correcting defects associated with a wide range of genetic diseases which are further described on the website of the National Institutes of Health under the topic subsection Genetic Disorders (website at health.nih.gov/topic/GeneticDisorders). The genetic brain diseases may include but are not limited to Adrenoleukodystrophy, Agenesis of the Corpus Callosum, Aicardi Syndrome, Alpers' Disease, Alzheimer's Disease, Barth Syndrome, Batten Disease, CADASIL, Cerebellar Degeneration, Fabry's Disease, Gerstmann-Straussler-Scheinker Disease, Huntington's Disease and other Triplet Repeat Disorders, Leigh's Disease, Lesch-Nyhan Syndrome, Menkes Disease, Mitochondrial Myopathies and NINDS Colpocephaly. These diseases are further described on the website of the National Institutes of Health under the subsection Genetic Brain Disorders.

[0345] In some embodiments, the condition may be neoplasia. In some embodiments, where the condition is neoplasia, the genes to be targeted are any of those listed in Table A (in this case PTEN and so forth). In some embodiments, the condition may be Age-related Macular Degeneration. In some embodiments, the condition may be a Schizophrenic Disorder. In some embodiments, the condition may be a Trinucleotide Repeat Disorder. In some embodiments, the

condition may be Fragile X Syndrome. In some embodiments, the condition may be a Secretase Related Disorder. In some embodiments, the condition may be a Prion - related disorder. In some embodiments, the condition may be ALS. In some embodiments, the condition may be a drug addiction. In some embodiments, the condition may be Autism. In some embodiments, the condition may be Alzheimer's Disease. In some embodiments, the condition may be inflammation. In some embodiments, the condition may be Parkinson's Disease.

[0346] Examples of proteins associated with Parkinson's disease include but are not limited to α-synuclein, DJ-1, LRRK2, PINK1, Parkin, UCHL1, Synphilin-1, and NURR1.

[0347] Examples of addiction-related proteins may include ABAT for example.

[0348] Examples of inflammation-related proteins may include the monocyte chemoattractant protein-1 (MCP1) encoded by the Ccr2 gene, the C-C chemokine receptor type 5 (CCR5) encoded by the Ccr5 gene, the IgG receptor IIB (FCGR2b, also termed CD32) encoded by the Fcgr2b gene, or the Fc epsilon R1g (FCER1g) protein encoded by the Fcer1g gene, for example.

[0349] Examples of cardiovascular diseases associated proteins may include IL1B (interleukin 1, beta), XDH (xanthine dehydrogenase), TP53 (tumor protein p53), PTGIS (prostaglandin I2 (prostacyclin) synthase), MB (myoglobin), IL4 (interleukin 4), ANGPT1 (angiopoietin 1), ABCG8 (ATP-binding cassette, sub-family G (WHITE), member 8), or CTSK (cathepsin K), for example.

[0350] Examples of Alzheimer's disease associated proteins may include the very low density lipoprotein receptor protein (VLDLR) encoded by the VLDLR gene, the ubiquitin-like modifier activating enzyme 1 (UBA1) encoded by the UBA1 gene, or the NEDD8-activating enzyme E1 catalytic subunit protein (UBE1C) encoded by the UBA3 gene, for example.

[0351] Examples of proteins associated Autism Spectrum Disorder may include the benzodiazapine receptor (peripheral) associated protein 1 (BZRAP1) encoded by the BZRAP1 gene, the AF4/FMR2 family member 2 protein (AFF2) encoded by the AFF2 gene (also termed MFR2), the fragile X mental retardation autosomal homolog 1 protein (FXR1) encoded by the FXR1 gene, or the fragile X mental retardation autosomal homolog 2 protein (FXR2) encoded by the FXR2 gene, for example.

[0352] Examples of proteins associated Macular Degeneration may include the ATP-binding cassette, sub-family A (ABC1) member 4 protein (ABCA4) encoded by the ABCR gene, the

apolipoprotein E protein (APOE) encoded by the APOE gene, or the chemokine (C-C motif) Ligand 2 protein (CCL2) encoded by the CCL2 gene, for example.

[0353] Examples of proteins associated Schizophrenia may include NRG1, ErbB4, CPLX1, TPH1, TPH2, NRXN1, GSK3A, BDNF, DISC1, GSK3B, and combinations thereof.

[0354] Examples of proteins involved in tumor suppression may include ATM (ataxia telangiectasia mutated), ATR (ataxia telangiectasia and Rad3 related), EGFR (epidermal growth factor receptor), ERBB2 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 2), ERBB3 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 3), ERBB4 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 4), Notch 1, Notch2, Notch 3, or Notch 4, for example.

[0355] Examples of proteins associated with a secretase disorder may include PSENEN (presenilin enhancer 2 homolog (C. elegans)), CTSB (cathepsin B), PSEN1 (presenilin 1), APP (amyloid beta (A4) precursor protein), APH1B (anterior pharynx defective 1 homolog B (C. elegans)), PSEN2 (presenilin 2 (Alzheimer disease 4)), or BACE1 (beta-site APP-cleaving enzyme 1), for example.

[0356] Examples of proteins associated with Amyotrophic Lateral Sclerosis may include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

[0357] Examples of proteins associated with prion diseases may include SOD1 (superoxide dismutase 1), ALS2 (amyotrophic lateral sclerosis 2), FUS (fused in sarcoma), TARDBP (TAR DNA binding protein), VAGFA (vascular endothelial growth factor A), VAGFB (vascular endothelial growth factor B), and VAGFC (vascular endothelial growth factor C), and any combination thereof.

[0358] Examples of proteins related to neurodegenerative conditions in prion disorders may include A2M (Alpha-2-Macroglobulin), AATF (Apoptosis antagonizing transcription factor), ACPP (Acid phosphatase prostate), ACTA2 (Actin alpha 2 smooth muscle aorta), ADAM22 (ADAM metallopeptidase domain), ADORA3 (Adenosine A3 receptor), or ADRA1D (Alpha-1D adrenergic receptor for Alpha-1D adrenoreceptor), for example.

[0359] Examples of proteins associated with Immunodeficiency may include A2M [alpha-2-macroglobulin]; AANAT [arylalkylamine N-acetyltransferase]; ABCA1 [ATP-binding cassette,

sub-family A (ABC1), member 1]; ABCA2 [ATP-binding cassette, sub-family A (ABC1), member 2]; or ABCA3 [ATP-binding cassette, sub-family A (ABC1), member 3]; for example.

[0360] Examples of proteins associated with Trinucleotide Repeat Disorders include AR (androgen receptor), FMR1 (fragile X mental retardation 1), HTT (huntingtin), or DMPK (dystrophia myotonica-protein kinase), FXN (frataxin), ATXN2 (ataxin 2), for example.

[0361] Examples of proteins associated with Neurotransmission Disorders include SST (somatostatin), NOS1 (nitric oxide synthase 1 (neuronal)), ADRA2A (adrenergic, alpha-2A-, receptor), ADRA2C (adrenergic, alpha-2C-, receptor), TACR1 (tachykinin receptor 1), or HTR2c (5-hydroxytryptamine (serotonin) receptor 2C), for example.

[0362] Examples of neurodevelopmental-associated sequences include A2BP1 [ataxin 2-binding protein 1], AADAT [aminoadipate aminotransferase], AANAT [arylalkylamine N-acetyltransferase], ABAT [4-aminobutyrate aminotransferase], ABCA1 [ATP-binding cassette, sub-family A (ABC1), member 1], or ABCA13 [ATP-binding cassette, sub-family A (ABC1), member 13], for example.

[0363] Further examples of preferred conditions treatable with the present system include may be selected from: Aicardi-Goutières Syndrome; Alexander Disease; Allan-Herndon-Dudley Syndrome; POLG-Related Disorders; Alpha-Mannosidosis (Type II and III); Alström Syndrome; Angelman; Syndrome; Ataxia-Telangiectasia; Neuronal Ceroid-Lipofuscinoses; Beta-Thalassemia; Bilateral Optic Atrophy and (Infantile) Optic Atrophy Type 1; Retinoblastoma (bilateral); Canavan Disease; Cerebrooculofacioskeletal Syndrome 1 [COFS1]; Cerebrotendinous Xanthomatosis; Cornelia de Lange Syndrome; MAPT-Related Disorders; Genetic Prion Diseases; Dravet Syndrome; Early-Onset Familial Alzheimer Disease; Friedreich Ataxia [FRDA]; Fryns Syndrome; Fucosidosis; Fukuyama Congenital Muscular Dystrophy; Galactosialidosis; Gaucher Disease; Organic Acidemias; Hemophagocytic Lymphohistiocytosis; Hutchinson-Gilford Progeria Syndrome; Mucolipidosis II; Infantile Free Sialic Acid Storage Disease; PLA2G6-Associated Neurodegeneration; Jervell and Lange-Nielsen Syndrome; Junctional Epidermolysis Bullosa; Huntington Disease; Krabbe Disease (Infantile); Mitochondrial DNA-Associated Leigh Syndrome and NARP; Lesch-Nyhan Syndrome; LIS1-Associated Lissencephaly; Lowe Syndrome; Maple Syrup Urine Disease; MECP2 Duplication Syndrome; ATP7A-Related Copper Transport Disorders; LAMA2-Related Muscular Dystrophy; Arylsulfatase A Deficiency; Mucopolysaccharidosis Types I, II or III; Peroxisome Biogenesis

Disorders, Zellweger Syndrome Spectrum; Neurodegeneration with Brain Iron Accumulation Disorders; Acid Sphingomyelinase Deficiency; Niemann-Pick Disease Type C; Glycine Encephalopathy; ARX-Related Disorders; Urea Cycle Disorders; COL1A1/2-Related Osteogenesis Imperfecta; Mitochondrial DNA Deletion Syndromes; PLP1-Related Disorders; Perry Syndrome; Phelan-McDermid Syndrome; Glycogen Storage Disease Type II (Pompe Disease) (Infantile); MAPT-Related Disorders; MECP2-Related Disorders; Rhizomelic Chondrodysplasia Punctata Type 1; Roberts Syndrome; Sandhoff Disease; Schindler Disease - Type 1; Adenosine Deaminase Deficiency; Smith-Lemli-Opitz Syndrome; Spinal Muscular Atrophy; Infantile-Onset Spinocerebellar Ataxia; Hexosaminidase A Deficiency; Thanatophoric Dysplasia Type 1; Collagen Type VI-Related Disorders; Usher Syndrome Type I; Congenital Muscular Dystrophy; Wolf-Hirschhorn Syndrome; Lysosomal Acid Lipase Deficiency; and Xeroderma Pigmentosum.

[0364] As will be apparent, it is envisaged that the present system can be used to target any polynucleotide sequence of interest. Some examples of conditions or diseases that might be usefully treated using the present system are included in the Tables above and examples of genes currently associated with those conditions are also provided there. However, the genes exemplified are not exhaustive.

[0365] For example, "wild type 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 is included in SwissProt under accession number Q99ZW2.

[0366] The ability to use CRISPR-Cas systems to perform efficient and cost effective gene editing and manipulation will allow the rapid selection and comparison of single and and multiplexed genetic manipulations to transform such genomes for improved production and enhanced traits. In this regard reference is made to US patents and publications: US Patent No. 6,603,061 - Agrobacterium-Mediated Plant Transformation Method; US Patent No. 7,868,149 - Plant Genome Sequences and Uses Thereof and US 2009/0100536 - Transgenic Plants with Enhanced Agronomic Traits, all the contents and disclosure of each of which are herein incorporated by reference in their entirety. In the practice of the invention, the contents and disclosure of Morrell et al "Crop genomics:advances and applications" Nat Rev Genet. 2011 Dec 29;13(2):85-96 are also herein incorporated by reference in their entirety.

EXAMPLES

[0367] 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: Methodological improvement to simplify cloning and delivery.

[0368] Rather than encoding the U6-promoter and guide RNA on a plasmid, Applicants amplified the U6 promoter with a DNA oligo to add on the guide RNA. The resulting PCR product may be transfected into cells to drive expression of the guide RNA.

[0369] Example primer pair that allows the generation a PCR product consisting of U6-promoter::guideRNA targeting human Emx1 locus:

[0370] Forward Primer: AAACTCTAGAgagggcctatttcccatgattc

[0371] Reverse Primer (carrying the guide RNA, which is underlined): acctctagAAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGC CTTATTTTAACTTGCTATGCTGTTTTGTTTCCAAAACAGCATAGCTCTAAAACCCC TAGTCATTGGAGGTGACGGTGTTTCGTCCTTTCCACaag

Example 2: Methodological improvement to improve activity:

[0372] Rather than use pol3 promoters, in particular RNA polymerase III (e.g. U6 or H1 promoters), to express guide RNAs in eukaryotic cells, Applicants express the T7 polymerase in eukaryotic cells to drive expression of guide RNAs using the T7 promoter.

[0373] One example of this system may involve introduction of three pieces of DNA:

[0374] 1. expression vector for Cas9

[0375] 2. expression vector for T7 polymerase

[0376] 3. expression vector containing guideRNA fused to the T7 promoter

Example 3: Methodological improvement to reduce toxicity of Cas9: Delivery of Cas9 in the form of mRNA.

[0377] Delivery of Cas9 in the form of mRNA enables transient expression of Cas9 in cells, to reduce toxicity. For example, humanized SpCas9 may be amplified using the following primer pair:

[0378] Forward Primer (to add on T7 promoter for in vitro transcription): TAATACGACTCACTATAGGAAGTGCGCCACCATGGCCCCAAAGAAGAAGCGG

[0379] Reverse Primer (to add on polyA tail): GGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTttcttaCTTTTTCTTTTTTGCCTGGCCG

[0380] Applicants transfect the Cas9 mRNA into cells with either guide RNA in the form of RNA or DNA cassettes to drive guide RNA expression in eukaryotic cells.

Example 4: Methodological improvement to reduce toxicity of Cas9: Use of an inducible promoter

[0381] Applicants transiently turn on Cas9 expression only when it is needed for carrying out genome modification. 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).

Example 5: Improvement of the Cas9 system for in vivo application

[0382] Applicants conducted a Metagenomic search for a Cas9 with small molecular weight. Most Cas9 homologs are fairly large. For example the SpCas9 is around 1368aa long, which is too large to be easily packaged into viral vectors for delivery. A graph representing the length distribution of Cas9 homologs is generated from sequences deposited in GenBank (Figure 5). Some of the sequences may have been mis-annotated and therefore the exact frequency for each length may not necessarily be accurate. Nevertheless it provides a glimpse at distribution of Cas9 proteins and suggest that there are shorter Cas9 homologs.

[0383] Through computational analysis, Applicants found that in the bacterial strain *Campylobacter*, there are two Cas9 proteins with less than 1000 amino acids. The sequence for one Cas9 from *Campylobacter jejuni* is presented below. At this length, CjCas9 can be easily packaged into AAV, lentiviruses, Adenoviruses, and other viral vectors for robust delivery into primary cells and *in vivo* in animal models. In a preferred embodiment of the invention, the Cas9 protein from *S. aureus* is used.

[0384] >Campylobacter jejuni Cas9 (CjCas9)

[0385] MARILAFDIGISSIGWAFSENDELKDCGVRIFTKVENPKTGESLALPRRLAR
SARKRLARRKARLNHLKHLIANEFKLNYEDYQSFDESLAKAYKGSLISPYELRFRALN
ELLSKQDFARVILHIAKRRGYDDIKNSDDKEKGAILKAIKQNEEKLANYQSVGEYLYK
EYFQKFKENSKEFTNVRNKKESYERCIAQSFLKDELKLIFKKQREFGFSFSKKFEEEVL
SVAFYKRALKDFSHLVGNCSFFTDEKRAPKNSPLAFMFVALTRIINLLNNLKNTEGIL
YTKDDLNALLNEVLKNGTLTYKQTKKLLGLSDDYEFKGEKGTYFIEFKKYKEFIKAL
GEHNLSQDDLNEIAKDITLIKDEIKLKKALAKYDLNQNQIDSLSKLEFKDHLNISFKAL
KLVTPLMLEGKKYDEACNELNLKVAINEDKKDFLPAFNETYYKDEVTNPVVLRAIKE
YRKVLNALLKKYGKVHKINIELAREVGKNHSQRAKIEKEQNENYKAKKDAELECEK
LGLKINSKNILKLRLFKEQKEFCAYSGEKIKISDLQDEKMLEIDHIYPYSRSFDDSYMN
KVLVFTKQNQEKLNQTPFEAFGNDSAKWQKIEVLAKNLPTKKQKRILDKNYKDKEQ
KNFKDRNLNDTRYIARLVLNYTKDYLDFLPLSDDENTKLNDTQKGSKVHVEAKSGM
LTSALRHTWGFSAKDRNNHLHHAIDAVIIAYANNSIVKAFSDFKKEQESNSAELYAK
KISELDYKNKRKFFEPFSGFRQKVLDKIDEIFVSKPERKKPSGALHEETFRKEEEFYQS
YGGKEGVLKALELGKIRKVNGKIVKNGDMFRVDIFKHKKTNKFYAVPIYTMDFALK
VLPNKAVARSKKGEIKDWILMDENYEFCFSLYKDSLILIQTKDMQEPEFVYYNAFTSS
TVSLIVSKHDNKFETLSKNQKILFKNANEKEVIAKSIGIQNLKVFEKYIVSALGEVTKA
EFRQREDFKK.

[0386] The putative tracrRNA element for this CjCas9 is:

[0387] TATAATCTCATAAGAAATTTAAAAAGGGACTAAAATAAAGAGTTTGCG
GGACTCTGCGGGGTTACAATCCCCTAAAACCGCTTTTAAAATT

[0388] The Direct Repeat sequence is:

[0389] ATTTTACCATAAAGAAATTTAAAAAGGGACTAAAAC

[0390] An example of a chimeric guideRNA for CjCas9 is:

[0391] NNNNNNNNNNNNNNNNNNNGUUUUAGUCCCGAAAGGGACUAAAAU
AAAGAGUUUGCGGGACUCUGCGGGGUUACAAUCCCCUAAAACCGCUUUU

Example 6: Cas9 optimization

[0392] For enhanced function or to develop new functions, Applicants generate chimeric Cas9 proteins by combining fragments from different Cas9 homologs. For example, two example chimeric Cas9 proteins:

[0393] For example, Applicants fused the N-term of St1Cas9 (fragment from this protein is in bold) with C-term of SpCas9 (fragment from this protein is underlined).

[0394] >St1(N)Sp(C)Cas9

[0395] **MSDLVLGLDIGIGSVGVGILNKVTGEIIHKNSRIFPAAQAENNLVRRTN RQGRRLARRKKHRRVRLNRLFEESGLITDFTKISINLNPYQLRVKGLTDELSNEE LFIALKNMVKHRGISYLDDASDDGNSSVGDYAQIVKENSKQLETKTPGQIQLERY QTYGQLRGDFTVEKDGKKHRLINVFPTSAYRSEALRILQTQQEFNPQITDEFINR YLEILTGKRKYYHGPGNEKSRTDYGRYRTSGETLDNIFGILIGKCTFYPDEFRAAK ASYTAQEFNLLNDLNNLTVPTETKKLSKEQKNQIINYVKNEKAMGPAKLFKYIAK LLSCDVADIKGYRIDKSGKAEIHTFEAYRKMKTLETLDIEQMDRETLDKLAYVLT LNTEREGIQEALEHEFADGSFSQKQVDELVQFRKANSSIFGKGWHNFSVKLMME LIPELYETSEEQMTILTRLGKQKTTSSSNKTKYIDEKLLTEEIYNPVVAKSVRQAIK IVNAAIKEYGDFDNIVIEMARENQTTQKGQKNSRERMKRIEEGIKELGSQILKEHPVE NTQLQNEKLYLYYLQNGRDMYVDQELDINRLSDYDVDHIVPQSFLKDDSIDNKVLT RSDKNRGKSDNVPSEEVVKKMKNYWRQLLNAKLITQRKFDNLTKAERGGLSELDK AGFIKRQLVETRQITKHVAQILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQ FYKVREINNYHHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQ EIGKATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRDFATVRKV LSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDPKKYGGFDSPTVAYSV LVVAKVEKGKSKKLKSVKELLGITIMERSSFEKNPIDFLEAKGYKEVKKDLIIKLPKYS LFELENGRKRMLASAGELQKGNELALPSKYVNFLYLASHYEKLKGSPEDNEQKQLF VEQHKHYLDEIIEQISEFSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNL GAPAAFKYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD

[0396] >Sp(N)St1(C)Cas9

[0397] MDKKYSIGLDIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFHRLEESFLVE EDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDKADLRLIYLALAHMIKFRG HFLIEGDLNPDNSDVDKLFIQLVQTYNQLFEENPINASGVDAKAILSARLSKSRRLENL IAQLPGEKKNGLFGNLIALSLGLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIG DQYADLFLAAKNLSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQ QLPEKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKLNREDLL

RKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEKILTFRIPYYVGPLARGN
SRFAWMTRKSEETITPWNFEEVVDKGASAQSFIERMTNFDKNLPNEKVLPKHSLLYE
YFTVYNELTKVKYVTEGMRKPAFLSGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIE
CFDSVEISGVEDRFNASLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIE
ERLKTYAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDGFAN
RNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKGILQTVKVVDELV
KVMGRHKPENIVIEMARETNEDDEKKAIQKIQKANKDEKDAAMLKAANQYNGKA
ELPHSVFHGHKQLATKIRLWHQQGERCLYTGKTISIHDLINNSNQFEVDHILPLSI
TFDDSLANKVLVYATANQEKGQRTPYQALDSMDDAWSFRELKAFVRESKTLSNK
KKEYLLTEEDISKFDVRKKFIERNLVDTRYASRVVLNALQEHFRAHKIDTKVSVVR
GQFTSQLRRHWGIEKTRDTYHHHAVDALIIAASSQLNLWKKQKNTLVSYSEDQL
LDIETGELISDDEYKESVFKAPYQHFVDTLKSKEFEDSILFSYQVDSKFNRKISDATI
YATRQAKVGKDKADETYVLGKIKDIYTQDGYDAFMKIYKKDKSKFLMYRHDPQT
FEKVIEPILENYPNKQINEKGKEVPCNPFLKYKEEHGYIRKYSKKGNGPEIKSLKYY
DSKLGNHIDITPKDSNNKVVLQSVSPWRADVYFNKTTGKYEILGLKYADLQFEKG
TGTYKISQEKYNDIKKKEGVDSDSEFKFTLYKNDLLLVKDTETKEQQLFRFLSRT
MPKQKHYVELKPYDKQKFEGGEALIKVLGNVANSGQCKKGLGKSNISIYKVRTD
VLGNQHIIKNEGDKPKLDF

[0398] The benefit of making chimeric Cas9 include:

[0399] reduce toxicity

[0400] improve expression in eukaryotic cells

[0401] enhance specificity

[0402] reduce molecular weight of protein, make protein smaller by combining the smallest domains from different Cas9 homologs.

[0403] Altering the PAM sequence requirement

Example 7: Utilization of Cas9 as a generic DNA binding protein

[0404] Applicants used Cas9 as a generic DNA binding protein by mutating the two catalytic domains (D10 and H840) responsible for cleaving both strands of the DNA target. In order to upregulate gene transcription at a target locus Applicants fused the transcriptional activation domain (VP64) to Cas9. Applicants hypothesized that it would be important to see strong nuclear

localization of the Cas9-VP64 fusion protein because transcription factor activation strength is a function of time spent at the target. Therefore, Applicants cloned a set of Cas9-VP64-GFP constructs, transfected them into 293 cells and assessed their localization under a fluorescent microscope 12 hours post-transfection.

[0405] The same constructs were cloned as a 2A-GFP rather than a direct fusion in order to functionally test the constructs without a bulky GFP present to interfere. Applicants elected to target the Sox2 locus with the Cas9 transactivator because it could be useful for cellular reprogram and the locus has already been validated as a target for TALE-TF mediated transcriptional activation. For the Sox2 locus Applicants chose eight targets near the transcriptional start site (TSS). Each target was 20bp long with a neighboring NGG protospacer adjacent motif (PAM). Each Cas9-VP64 construct was co-transfected with each PCR generated chimeric crispr RNA (chiRNA) in 293 cells. 72 hours post transfection the transcriptional activation was assessed using RT-qPCR.

[0406] To further optimize the transcriptional activator, Applicants titrated the ratio of chiRNA (Sox2.1 and Sox2.5) to Cas9 (NLS-VP64-NLS-hSpCas9-NLS-VP64-NLS), transfected into 293 cells, and quantified using RT-qPCR. These results indicate that Cas9 can be used as a generic DNA binding domain to upregulate gene transcription at a target locus.

[0407] Applicants designed a second generation of constructs. (Table below).

pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-hSpCsn1(D10A, H840A)-NLS
pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-hSpCsn1(D10A, H840A)
pLenti-EF1a-GFP-2A-6xHis-NLS-VP64-NLS-NLS-hSpCsn1(D10A, H840A)
pLenti-EF1a-GFP-2A-6xHis-NLS-hSpCsn1(D10A, H840A)-NLS
pLenti-EF1a-GFP-2A-6xHis-NLS-hSpCsn1(D10A, H840A)
pLenti-EF1a-GFP-2A-6xHis-NLS-NLS-hSpCsn1(D10A, H840A)

[0408] Applicants use these constructs to assess transcriptional activation (VP64 fused constructs) and repression (Cas9 only) by RT-qPCR. Applicants assess the cellular localization of each construct using anti-His antibody, nuclease activity using a Surveyor nuclease assay, and DNA binding affinity using a gel shift assay. In a preferred embodiment of the invention, the gel shift assay is an EMSA gel shift assay.

Example 8: Cas9 transgenic and knock in mice

[0409] To generate a mouse that expresses the Cas9 nuclease Applicants submit two general strategies, transgenic and knock in. These strategies may be applied to generate any other model organism of interest, for e.g. Rat. For each of the general strategies Applicants made a constitutively active Cas9 and a Cas9 that is conditionally expressed (Cre recombinase dependent). The constitutively active Cas9 nuclease is expressed in the following context: pCAG-NLS-Cas9-NLS-P2A-EGFP-WPRE-bGHpolyA. pCAG is the promoter, NLS is a nuclear localization signal, P2A is the peptide cleavage sequence, EGFP is enhanced green fluorescent protein, WPRE is the woodchuck hepatitis virus posttranscriptional regulatory element, and bGHpolyA is the bovine growth hormone poly-A signal sequence (Figures 7A-B). The conditional version has one additional stop cassette element, loxP-SV40 polyA x3-loxP, after the promoter and before NLS-Cas9-NLS (i.e. pCAG-loxP-SV40polyAx3-loxP-NLS-Cas9-NLS-P2A-EGFP-WPRE-bGHpolyA). The important expression elements can be visualized as in Figure 8. The constitutive construct should be expressed in all cell types throughout development, whereas, the conditional construct will only allow Cas9 expression when the same cell is expressing the Cre recombinase. This latter version will allow for tissue specific expression of Cas9 when Cre is under the expression of a tissue specific promoter. Moreover, Cas9 expression could be induced in adult mice by putting Cre under the expression of an inducible promoter such as the TET on or off system.

[0410] Validation of Cas9 constructs: Each plasmid was functionally validated in three ways: 1) transient transfection in 293 cells followed by confirmation of GFP expression; 2) transient transfection in 293 cells followed by immunofluorescence using an antibody recognizing the P2A sequence; and 3) transient transfection followed by Surveyor nuclease assay. The 293 cells may be 293FT or 293 T cells depending on the cells that are of interest. In a preferred embodiment the cells are 293FT cells. The results of the Surveyor were run out on the top and bottom row of the gel for the conditional and constitutive constructs, respectively. Each was tested in the presence and absence of chimeric RNA targeted to the hEMX1 locus (chimeric RNA hEMX1.1). The results indicate that the construct can successfully target the hEMX1 locus only in the presence of chimeric RNA (and Cre in the conditional case). The gel was quantified and the results are presented as average cutting efficiency and standard deviation for three samples.

[0411] Transgenic Cas9 mouse: To generate transgenic mice with constructs, Applicants inject pure, linear DNA into the pronucleus of a zygote from a pseudo pregnant CB56 female. Founders are identified, genotyped, and backcrossed to CB57 mice. The constructs were successfully cloned and verified by Sanger sequencing.

[0412] Knock in Cas9 mouse: To generate Cas9 knock in mice Applicants target the same constitutive and conditional constructs to the Rosa26 locus. Applicants did this by cloning each into a Rosa26 targeting vector with the following elements: Rosa26 short homology arm – constitutive/conditional Cas9 expression cassette – pPGK-Neo-Rosa26 long homology arm – pPGK-DTA. pPGK is the promoter for the positive selection marker Neo, which confers resistance to neomycin, a 1 kb short arm, a 4.3 kb long arm, and a negative selection diphtheria toxin (DTA) driven by PGK.

[0413] The two constructs were electroporated into R1 mESCs and allowed to grow for 2 days before neomycin selection was applied. Individual colonies that had survived by days 5-7 were picked and grown in individual wells. 5-7 days later the colonies were harvested, half were frozen and the other half were used for genotyping. Genotyping was done by genomic PCR, where one primer annealed within the donor plasmid (AttpF) and the other outside of the short homology arm (Rosa26-R) Of the 22 colonies harvested for the conditional case, 7 were positive (Left). Of the 27 colonies harvested for the constitutive case, zero were positive (Right). It is likely that Cas9 causes some level of toxicity in the mESC and for this reason there were no positive clones. To test this Applicants introduced a Cre expression plasmid into correctly targeted conditional Cas9 cells and found very low toxicity after many days in culture. The reduced copy number of Cas9 in correctly targeted conditional Cas9 cells (1-2 copies per cell) is enough to allow stable expression and relatively no cytotoxicity. Moreover, this data indicates that the Cas9 copy number determines toxicity. After electroporation each cell should get several copies of Cas9 and this is likely why no positive colonies were found in the case of the constitutive Cas9 construct. This provides strong evidence that utilizing a conditional, Cre-dependent strategy should show reduced toxicity. Applicants inject correctly targeted cells into a blastocyst and implant into a female mouse. Chimerics are identified and backcrossed. Founders are identified and genotyped.

[0414] Utility of the conditional Cas9 mouse: Applicants have shown in 293 cells that the Cas9 conditional expression construct can be activated by co-expression with Cre. Applicants

also show that the correctly targeted R1 mESCs can have active Cas9 when Cre is expressed. Because Cas9 is followed by the P2A peptide cleavage sequence and then EGFP Applicants identify successful expression by observing EGFP. This same concept is what makes the conditional Cas9 mouse so useful. Applicants may cross their conditional Cas9 mouse with a mouse that ubiquitously expresses Cre (ACTB-Cre line) and may arrive at a mouse that expresses Cas9 in every cell. It should only take the delivery of chimeric RNA to induce genome editing in embryonic or adult mice. Interestingly, if the conditional Cas9 mouse is crossed with a mouse expressing Cre under a tissue specific promoter, there should only be Cas9 in the tissues that also express Cre. This approach may be used to edit the genome in only precise tissues by delivering chimeric RNA to the same tissue.

Example 9: Cas9 diversity and chimeric RNAs

[0415] The CRISPR-Cas system is an adaptive immune mechanism against invading exogenous DNA employed by diverse species across bacteria and archaea. The type II CRISPR-Cas system consists of a set of genes encoding proteins responsible for the "acquisition" of foreign DNA into the CRISPR locus, as well as a set of genes encoding the "execution" of the DNA cleavage mechanism; these include the DNA nuclease (Cas9), a non-coding transactivating cr-RNA (tracrRNA), and an array of foreign DNA-derived spacers flanked by direct repeats (crRNAs). Upon maturation by Cas9, the tracrRNA and crRNA duplex guide the Cas9 nuclease to a target DNA sequence specified by the spacer guide sequences, and mediates double-stranded breaks in the DNA near a short sequence motif in the target DNA that is required for cleavage and specific to each CRISPR-Cas system. The type II CRISPR-Cas systems are found throughout the bacterial kingdom and highly diverse in in Cas9 protein sequence and size, tracrRNA and crRNA direct repeat sequence, genome organization of these elements, and the motif requirement for target cleavage. One species may have multiple distinct CRISPR-Cas systems.

[0416] Applicants evaluated 207 putative Cas9s from bacterial species identified based on sequence homology to known Cas9s and structures orthologous to known subdomains, including the HNH endonuclease domain and the RuvC endonuclease domains [information from the Eugene Koonin and Kira Makarova]. Phylogenetic analysis based on the protein sequence

conservation of this set revealed five families of Cas9s, including three groups of large Cas9s (~1400 amino acids) and two of small Cas9s (~1100 amino acids) (Figures 3 and 4A-F).

[0417] Applicants have also optimized Cas9 guide RNA using *in vitro* methods.

Example 10: Cas9 mutations

[0418] In this example, Applicants show that the following mutations can convert SpCas9 into a nicking enzyme: D10A, E762A, H840A, N854A, N863A, D986A.

[0419] Applicants provide sequences showing where the mutation points are located within the SpCas9 gene (Figure 6A-M). Applicants also show that the nickases are still able to mediate homologous recombination. Furthermore, Applicants show that SpCas9 with these mutations (individually) do not induce double strand break.

[0420] Cas9 orthologs all 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.

[0421] 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 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.

Example 11: Cas9 Transcriptional Activation and Cas9 Repressor

[0422] Cas9 Transcriptional Activation

[0423] A second generation of constructs have been designed and are in the pipeline to be tested (Table 1). These constructs will be used to assess transcriptional activation (VP64 fused constructs) and repression (Cas9 only) by RT-qPCR. Applicants will also assess the cellular localization of each construct using anti-His antibody,, nuclease activity using a Surveyor nuclease assay, and DNA binding affinity using a gel shift assay.

[0424] Cas Repressor

[0425] It has been shown previously that dCas9 can be used as a generic DNA binding domain to repress gene expression. Applicants report an improved dCas9 design as well as dCas9 fusions to the repressor domains KRAB and SID4x. From the plasmid library created for modulating transcription using Cas9 in Table 1, the following repressor plasmids were functionally characterized by qPCR: pXRP27, pXRP28, pXRP29, pXRP48, pXRP49, pXRP50, pXRP51, pXRP52, pXRP53, pXRP56, pXRP58, pXRP59, pXRP61, and pXRP62.

[0426] Each dCas9 repressor plasmid was co-transfected with two guide RNAs targeted to the coding strand of the beta-catenin gene. RNA was isolated 72 hours after transfection and gene expression was quantified by RT-qPCR. The endogenous control gene was GAPDH. Two validated shRNAs were used as positive controls. Negative controls were certain plasmids transfected without gRNA, these are denoted as "pXRP## control". The plasmids pXRP28, pXRP29, pXRP48, and pXRP49 could repress the beta-catenin gene when using the specified targeting strategy. These plasmids correspond to dCas9 without a functional domain (pXRP28 and pXRP28) and dCas9 fused to SID4x (pXRP48 and pXRP49).

[0427] Further work investigates: repeating the above experiment, targeting different genes, utilizing other gRNAs to determine the optimal targeting position, and multiplexed repression.

Table 1

pXRP024-pLenti2-EF1a-VP64-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP025-pLenti2-EF1a-VP64-NLS-$GGGGS_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP026-pLenti2-EF1a-VP64-NLS-$EAAAK_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP027-pLenti2-EF1a-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP028-pLenti2-EF1a-NLS-$GGGGS_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP029-pLenti2-EF1a-NLS-$EAAAK_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP030-pLenti2-pSV40-VP64-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP031-pLenti2-pPGK-VP64-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP032-pLenti2-LTR-VP64-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP033-pLenti2-pSV40-VP64-NLS-$GGGGS_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP034-pLenti2-pPGK-VP64-NLS-$GGGGS_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP035-pLenti2-LTR-VP64-NLS-$GGGGS_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP036-pLenti2-pSV40-VP64-NLS-$EAAAK_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP037-pLenti2-pPGK-VP64-NLS-$EAAAK_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP038-pLenti2-LTR-VP64-NLS-$EAAAK_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP048-pLenti2-EF1a-SID4x-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP049-pLenti2-EF1a-SID4X-NLS-$GGGGS_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE

pXRP050-pLenti2-EF1a-SID4X-NLS-EAAAK$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP051-pLenti2-EF1a-KRAB-NLS-FLAG-Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP052-pLenti2-EF1a-KRAB-NLS-GGGGS$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP053-pLenti2-EF1a-KRAB-NLS-EAAAK$_3$Linker-dCas9-NLS-gLuc-2A-GFP-WPRE
pXRP054-pLenti2-EF1a-dCas9-Linker-FLAG-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP055-pLenti2-EF1a-dCas9-Linker-FLAG-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP056-pLenti2-EF1a-dCas9-Linker-FLAG-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP057-pLenti2-EF1a-dCas9-GGGGGS$_3$-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP058-pLenti2-EF1a-dCas9-GGGGGS$_3$-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP059-pLenti2-EF1a-dCas9-GGGGGS$_3$-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP060-pLenti2-EF1a-dCas9-EAAAK$_3$-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP061-pLenti2-EF1a-dCas9-EAAAK$_3$-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP062-pLenti2-EF1a-dCas9-EAAAK$_3$-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP024-pLenti2-EF1a-VP64-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP025-pLenti2-EF1a-VP64-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP026-pLenti2-EF1a-VP64-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP027-pLenti2-EF1a-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP028-pLenti2-EF1a-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP029-pLenti2-EF1a-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP030-pLenti2-pSV40-VP64-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP031-pLenti2-pPGK-VP64-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP032-pLenti2-LTR-VP64-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP033-pLenti2-pSV40-VP64-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP034-pLenti2-pPGK-VP64-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP035-pLenti2-LTR-VP64-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP036-pLenti2-pSV40-VP64-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP037-pLenti2-pPGK-VP64-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP038-pLenti2-LTR-VP64-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP048-pLenti2-EF1a-SID4x-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP049-pLenti2-EF1a-SID4X-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP050-pLenti2-EF1a-SID4X-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP051-pLenti2-EF1a-KRAB-NLS-FLAG-Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP052-pLenti2-EF1a-KRAB-NLS-GGGGS$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP053-pLenti2-EF1a-KRAB-NLS-EAAAK$_3$Linker-Cas9-NLS-gLuc-2A-GFP-WPRE
pXRP054-pLenti2-EF1a-Cas9-Linker-FLAG-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP055-pLenti2-EF1a-Cas9-Linker-FLAG-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP056-pLenti2-EF1a-Cas9-Linker-FLAG-NLS-KRAB-gLuc-2A-GFP-WPRE
pXRP057-pLenti2-EF1a-Cas9-GGGGGS$_3$-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP058-pLenti2-EF1a-Cas9-GGGGGS$_3$-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP059-pLenti2-EF1a-Cas9-GGGGGS$_3$-NLS-KRAB-gLuc-2A-GFP-WPRE

pXRP060-pLenti2-EF1a-Cas9-EAAAK$_3$-NLS-VP64-gLuc-2A-GFP-WPRE
pXRP061-pLenti2-EF1a-Cas9-EAAAK$_3$-NLS-SID4X-gLuc-2A-GFP-WPRE
pXRP062-pLenti2-EF1a-Cas9-EAAAK$_3$-NLS-KRAB-gLuc-2A-GFP-WPRE

Example 12: Targeted deletion of genes involved in cholesterol biosynthesis, fatty acid biosynthesis, and other metabolic disorders, genes encoding mis-folded proteins involved in amyloid and other diseases, oncogenes leading to cellular transformation, latent viral genes, and genes leading to dominant-negative disorders, amongst other disorders.

[0428] Applicants demonstrate gene delivery of a CRISPR-Cas system in the liver, brain, ocular, epithelial, hematopoetic, or another tissue of a subject or a patient in need thereof, suffering from metabolic disorders, amyloidosis and protein-aggregation related diseases, cellular transformation arising from genetic mutations and translocations, dominant negative effects of gene mutations, latent viral infections, and other related symptoms, using either viral or nanoparticle delivery system.

[0429] **Study Design:** Subjects or patients in need thereof suffering from metabolic disorders, amyloidosis and protein aggregation related disease which include but are not limited to human, non-primate human, canine, feline, bovine, equine, other domestic animals and related mammals. The CRISPR-Cas system is guided by a chimeric guide RNA and targets a specific site of the human genomic loci to be cleaved. After cleavage and non-homologous end-joining mediated repair, frame-shift mutation results in knock out of genes.

[0430] Applicants select guide-RNAs targeting genes involved in above-mentioned disorders to be specific to endogenous loci with minimal off-target activity. Two or more guide RNAs may be encoded into a single CRISPR array to induce simultaneous double-stranded breaks in DNA leading to micro-deletions of affected genes or chromosomal regions.

[0431] Identification and design of gene targets

[0432] For each candidate disease gene, Applicants select DNA sequences of interest include protein-coding exons, sequences including and flanking known dominant negative mutation sites, sequences including and flanking pathological repetitive sequences. For gene-knockout approaches, early coding exons closest to the start codon offer best options for achieving complete knockout and minimize possibility of truncated protein products retaining partial function.

[0433] Applicants analyze sequences of interest for all possible targetable 20-bp sequences immediately 5' to a NGG motif (for SpCas9 system) or a NNAGAAW (for St1Cas9 system).

Applicants choose sequences for unique, single RNA-guided Cas9 recognition in the genome to minimize off-target effects based on computational algorithm to determine specificity.

[0434] Cloning of guide sequences into a delivery system

[0435] Guide sequences are synthesized as double-stranded 20-24 bp oligonucleotides. After 5'-phosphorylation treatment of oligos and annealing to form duplexes, oligos are ligated into suitable vector depending on the delivery method:

[0436] Virus-based delivery methods

[0437] AAV-based vectors (PX260, 330, 334, 335) have been described elsewhere

[0438] Lentiviral-based vectors use a similar cloning strategy of directly ligating guide sequences into a single vector carrying a U6 promoter-driven chimeric RNA scaffold and a EF1a promoter-driven Cas9 or Cas9 nickase.

[0439] Virus production is described elsewhere.

[0440] Nanoparticle-based RNA delivery methods

[0441] 1. Guide sequences are synthesized as an oligonucleotide duplex encoding T7 promoter—guide sequence—chimeric RNA. A T7 promoter is added 5' of Cas9 by PCR method.

[0442] 2. T7-driven Cas9 and guide-chimeric RNAs are transcribed *in vitro*, and Cas9 mRNA is further capped and A-tailed using commercial kits. RNA products are purified per kit instructions.

[0443] Hydrodynamic tail vein delivery methods (for mouse)

[0444] Guide sequences are cloned into AAV plasmids as described above and elsewhere in this application.

[0445] *In vitro* validation on cell lines

[0446] Transfection

[0447] 1. DNA plasmid transfection

[0448] Plasmids carrying guide sequences are transfected into human embryonic kidney (HEK293T) or human embryonic stem (hES) cells, other relevant cell types using lipid-, chemical-, or electroporation-based methods. For a 24-well transfection of HEK293T cells (~260,000 cells), 500ng of total DNA is transfected into each single well using Lipofectamine 2000. For a 12-well transfection of hES cells, 1ug of total DNA is transfected into a single well using Fugene HD.

[0449] 2. RNA transfection

[0450] Purified RNA described above is used for transfection into HEK293T cells. 1-2ug of RNA may be transfected into ~260,000 using Lipofectamine 2000 per manufacturer's instruction. RNA delivery of Cas9 and chimeric RNA is shown in Figure 10.

[0451] Assay of indel formation *in vitro*

[0452] Cells are harvested 72-hours post-transfection and assayed for indel formation as an indication of double-stranded breaks.

[0453] Briefly, genomic region around target sequence is PCR amplified (~400-600 bp amplicon size) using high-fidelity polymerase. Products are purified, normalized to equal concentration, and slowly annealed from 95°C to 4°C to allow formation of DNA heteroduplexes. Post annealing, the Cel-I enzyme is used to cleave heteroduplexes, and resulting products are separated on a polyacrylamide gel and indel efficiency calculated.

[0454] In vivo proof of principle in animal

[0455] Delivery mechanisms

[0456] AAV or Lentivirus production is described elsewhere.

[0457] Nanoparticle formulation: RNA mixed into nanoparticle formulation

[0458] Hydrodynamic tail vein injections with DNA plasmids in mice are conducted using a commercial kit

[0459] Cas9 and guide sequences are delivered as virus, nanoparticle-coated RNA mixture, or DNA plasmids, and injected into subject animals. A parallel set of control animals is injected with sterile saline, Cas9 and GFP, or guide sequence and GFP alone.

[0460] Three weeks after injection, animals are tested for amelioration of symptoms and sacrificed. Relevant organ systems analyzed for indel formation. Phenotypic assays include blood levels of HDL, LDL, lipids,

[0461] Assay for indel formation

[0462] DNA is extracted from tissue using commercial kits; indel assay will be performed as described for *in vitro* demonstration.

[0463] Therapeutic applications of the CRISPR-Cas system are amenable for achieving tissue-specific and temporally controlled targeted deletion of candidate disease genes. Examples include genes involved in cholesterol and fatty acid metabolism, amyloid diseases, dominant negative diseases, latent viral infections, among other disorders.

[0464] Examples of a single guide-RNA to introduce targeted indels at a gene locus

Disease	GENE	SPACER	PAM	Mechanism	References
Hypercholesterolemia	HMG-CR	GCCAAATTGGACGACCCTCG	CGG	Knockout	Fluvastatin: a review of its pharmacology and use in the management of hypercholesterolaemia.(Plosker GL et al. Drugs 1996, 51(3):433-459)
Hypercholesterolemia	SQLE	CGAGGAGACCCCCGTTTCGG	TGG	Knockout	Potential role of nonstatin cholesterol lowering agents (Trapani et al. IUBMB Life, Volume 63, Issue 11, pages 964–971, November 2011)
Hyperlipidemia	DGAT1	CCCGCCGCCGCCGTGGCTCG	AGG	Knockout	DGAT1 inhibitors as anti-obesity and anti-diabetic agents. (Birch AM et al. Current Opinion in Drug Discovery & Development [2010, 13(4):489-496)
Leukemia	BCR-ABL	TGAGCTCTACGAGATCCACA	AGG	Knockout	Killing of leukemic cells with a BCR/ABL fusion gene by RNA interference (RNAi).(Fuchs et al. Oncogene 2002, 21(37):5716-5724)

[0465] Examples of a pair of guide-RNA to introduce chromosomal microdeletion at a gene locus

Disease	GENE	SPACER	PAM	Mechanism	References
Hyperlipidemia	PLIN2 guide1	CTCAAAATTCATACCGGTTG	TGG	Microdeletion	Perilipin-2 Null Mice are Protected Against Diet-Induced Obesity, Adipose Inflammation and Fatty Liver Disease (McManaman JL et al. The Journal of Lipid Research, jlr.M035063. First Published on February 12, 2013)
Hyperlipidemia	PLIN2 guide2	CGTTAAACAACAACCGGACT	TGG	Microdeletion	
Hyperlipidemia	SREBP guide1	TTCACCCCGCGGCGCTGAAT	ggg	Microdeletion	Inhibition of SREBP by a Small Molecule, Betulin, Improves Hyperlipidemia and Insulin Resistance and Reduces Atherosclerotic Plaques (Tang J et

				al. Cell Metabolism, Volume 13, Issue 1, 44-56, 5 January 2011)
Hyperlipidemia	SREBP guide2	ACCACTACC AGTCCGTCC AC	agg	Microdeletion

Example 13: Targeted integration of repair for genes carrying disease-causing mutations; reconstitution of enzyme deficiencies and other related diseases.

[0466] Study design

I. Identification and design of gene targets

- Described in Example 16

II. Cloning of guide sequences and repair templates into a delivery system

- Described above in Example 16
- Applicants clone DNA repair templates to include homology arms with diseased allele as well a wild-type repair template

III. *In vitro* validation on cell lines

a. Transfection is described above in Example 16; Cas9, guide RNAs, and repair template are co-transfected into relevant cell types.

b. Assay for repair *in vitro*

i. Applicants harvest cells 72-hours post-transfection and assay for repair

ii. Briefly, Applicants amplify genomic region around repair template PCR using high-fidelity polymerase. Applicants sequence products for decreased incidence of mutant allele.

IV. *In vivo* proof of principle in animal

a. Delivery mechanisms are described above Examples 16 and 29.

b. Assay for repair *in vivo*

i. Applicants perform the repair assay as described in the *in vitro* demonstration.

V. Therapeutic applications

The CRISPR-Cas system is amenable for achieving tissue-specific and temporally controlled targeted deletion of candidate disease genes. Examples include genes

involved in cholesterol and fatty acid metabolism, amyloid diseases, dominant negative diseases, latent viral infections, among other disorders.

[0467] Example of one single missense mutation with repair template:

Disease	**GENE**	**SPACER**	**PAM**
Familial amyloid polyneuropathy	TTR	AGCCTTTCTGAACACATGCA	CGG

Mechanism	**References**
V30M repair	Transthyretin mutations in health and disease (Joao et al. Human Mutation, Volume 5, Issue 3, pages 191–196, 1995)

V30M allele CCTGCCATCAATGTGGCCATGCATGTGTTCAGAAAGGCT

WT allele CCTGCCATCAATGTGGCCGTGCATGTGTTCAGAAAGGCT

Example 14: Therapeutic application of the CRISPR-Cas system in Glaucoma, Amyloidosis, and Huntington's disease

[1000] Glaucoma: Applicants design guide RNAs to target the first exon of the mycilin (MYOC) gene. Applicants use adenovirus vectors (Ad5) to package both Cas9 as well as a guide RNA targeting the MYOC gene. Applicants inject adenoviral vectors into the trabecular meshwork where cells have been implicated in the pathophysiology of glaucoma. Applicants initially test this out in mouse models carrying the mutated MYOC gene to see whether they improve visual acuity and decrease pressure in the eyes. Therapeutic application in humans employ a similar strategy.

[0468] Amyloidosis: Applicants design guide RNAs to target the first exon of the transthyretin (TTR) gene in the liver. Applicants use AAV8 to package Cas9 as well as guide RNA targeting the first exon of the TTR gene. AAV8 has been shown to have efficient targeting of the liver and will be administered intravenously. Cas9 can be driven either using liver specific promoters such as the albumin promoter, or using a constitutive promoter. A pol3 promoter drives the guide RNA.

[0469] Alternatively, Applicants utilize hydrodynamic delivery of plasmid DNA to knockout the TTR gene. Applicants deliver a plasmid encoding Cas9 and the guideRNA targeting Exon1 of TTR.

[0470] As a further alternative approach, Applicants administer a combination of RNA (mRNA for Cas9, and guide RNA). RNA can be packaged using liposomes such as Invivofectamine from Life Technologies and delivered intravenously. To reduce RNA-induced

immunogenicity, increase the level of Cas9 expression and guide RNA stability, Applicants modify the Cas9 mRNA using 5' capping. Applicants also incorporate modified RNA nucleotides into Cas9 mRNA and guide RNA to increase their stability and reduce immunogenicity (e.g. activation of TLR). To increase efficiency, Applicants administer multiple doses of the virus, DNA, or RNA.

[0471] Huntington's Disease: Applicants design guide RNA based on allele specific mutations in the HTT gene of patients. For example, in a patient who is heterozygous for HTT with expanded CAG repeat, Applicants identify nucleotide sequences unique to the mutant HTT allele and use it to design guideRNA. Applicants ensure that the mutant base is located within the last 9 bp of the guide RNA (which Applicants have ascertained has the ability to discriminate between single DNA base mismatches between the target size and the guide RNA).

[0472] Applicants package the mutant HTT allele specific guide RNA and Cas9 into AAV9 and deliver into the striatum of Huntington's patients. Virus is injected into the striatum stereotactically via a craniotomy. AAV9 is known to transduce neurons efficiently. Applicants drive Cas9 using a neuron specific promoter such as human Synapsin I.

Example 15: Therapeutic application of the CRISPR-Cas system in HIV

[0473] Chronic viral infection is a source of significant morbidity and mortality. While there exists for many of these viruses conventional antiviral therapies that effectively target various aspects of viral replication, current therapeutic modalities are usually non-curative in nature due to "viral latency." By its nature, viral latency is characterized by a dormant phase in the viral life cycle without active viral production. During this period, the virus is largely able to evade both immune surveillance and conventional therapeutics allowing for it to establish long-standing viral reservoirs within the host from which subsequent re-activation can permit continued propagation and transmission of virus. Key to viral latency is the ability to stably maintain the viral genome, accomplished either through episomal or proviral latency, which stores the viral genome in the cytoplasm or integrates it into the host genome, respectively. In the absence of effective vaccinations which would prevent primary infection, chronic viral infections characterized by latent reservoirs and episodes of lytic activity can have significant consequences: human papilloma virus (HPV) can result in cervical cancer, hepatitis C virus (HCV) predisposes to hepatocellular carcinoma, and human immunodeficiency virus eventually

destroys the host immune system resulting in susceptibility to opportunistic infections. As such, these infections require life-long use of currently available antiviral therapeutics. Further complicating matters is the high mutability of many of these viral genomes which lead to the evolution of resistant strains for which there exists no effective therapy.

[0474] The CRISPR-Cas system is a bacterial adaptive immune system able to induce double-stranded DNA breaks (DSB) in a multiplex-able, sequence-specific manner and has been recently re-constituted within mammalian cell systems. It has been shown that targeting DNA with one or numerous guide-RNAs can result in both indels and deletions of the intervening sequences, respectively. As such, this new technology represents a means by which targeted and multiplexed DNA mutagenesis can be accomplished within a single cell with high efficiency and specificity. Consequently, delivery of the CRISPR-Cas system directed against viral DNA sequences could allow for targeted disruption and deletion of latent viral genomes even in the absence of ongoing viral production.

[0475] As an example, chronic infection by HIV-1 represents a global health issue with 33 million individuals infected and an annual incidence of 2.6 million infections. The use of the multimodal highly active antiretroviral therapy (HAART), which simultaneously targets multiple aspects of viral replication, has allowed HIV infection to be largely managed as a chronic, not terminal, illness. Without treatment, progression of HIV to AIDS occurs usually within 9-10 years resulting in depletion of the host immune system and occurrence of opportunistic infections usually leading to death soon thereafter. Secondary to viral latency, discontinuation of HAART invariably leads to viral rebound. Moreover, even temporary disruptions in therapy can select for resistant strains of HIV uncontrollable by available means. Additionally, the costs of HAART therapy are significant: within the US $10,000-15,0000 per person per year. As such, treatment approaches directly targeting the HIV genome rather than the process of viral replication represents a means by which eradication of latent reservoirs could allow for a curative therapeutic option.

[0476] Development and delivery of an HIV-1 targeted CRISPR-Cas system represents a unique approach differentiable from existing means of targeted DNA mutagenesis, i.e. ZFN and TALENs, with numerous therapeutic implications. Targeted disruption and deletion of the HIV-1 genome by CRISPR-mediated DSB and indels in conjunction with HAART could allow for

simultaneous prevention of active viral production as well as depletion of latent viral reservoirs within the host.

[0477] Once integrated within the host immune system, the CRISPR-Cas system allows for generation of a HIV-1 resistant sub-population that, even in the absence of complete viral eradication, could allow for maintenance and re-constitution of host immune activity. This could potentially prevent primary infection by disruption of the viral genome preventing viral production and integration, representing a means to "vaccination". Multiplexed nature of the CRISPR-Cas system allows targeting of multiple aspects of the genome simultaneously within individual cells.

[0478] As in HAART, viral escape by mutagenesis is minimized by requiring acquisition of multiple adaptive mutations concurrently. Multiple strains of HIV-1 can be targeted simultaneously which minimizes the chance of super-infection and prevents subsequent creation of new recombinants strains. Nucleotide, rather than protein, mediated sequence-specificity of the CRISPR-Cas system allows for rapid generation of therapeutics without need for significantly altering delivery mechanism.

[0479] In order to accomplish this, Applicants generate CRISPR-Cas guide RNAs that target the vast majority of the HIV-1 genome while taking into account HIV-1 strain variants for maximal coverage and effectiveness. Sequence analyses of genomic conservation between HIV-1 subtypes and variants should allow for targeting of flanking conserved regions of the genome with the aims of deleting intervening viral sequences or induction of frame-shift mutations which would disrupt viral gene functions.

[0480] Applicants accomplish delivery of the CRISPR-Cas system by conventional adenoviral or lentiviral-mediated infection of the host immune system. Depending on approach, host immune cells could be a) isolated, transduced with CRISPR-Cas, selected, and re-introduced in to the host or b) transduced in vivo by systemic delivery of the CRISPR-Cas system. The first approach allows for generation of a resistant immune population whereas the second is more likely to target latent viral reservoirs within the host.

Examples of potential HIV-1 targeted spacers adapted from Mcintyre *et al*, which generated shRNAs against HIV-1 optimized for maximal coverage of HIV-1 variants.
CACTGCTTAAGCCTCGCTCGAGG
TCACCAGCAATATTCGCTCGAGG
CACCAGCAATATTCCGCTCGAGG
TAGCAACAGACATACGCTCGAGG

GGGCAGTAGTAATACGCTCGAGG CCAATTCCCATACATTATTGTAC

Example 16: Targeted correction of deltaF508 or other mutations in cystic fibrosis

[0481] An aspect of the invention provides for a pharmaceutical composition that may comprise an CRISPR-Cas gene therapy particle and a biocompatible pharmaceutical carrier. According to another aspect, a method of gene therapy for the treatment of a subject having a mutation in the CFTR gene comprises administering a therapeutically effective amount of a CRISPR-Cas gene therapy particle to the cells of a subject.

[0482] This Example demonstrates gene transfer or gene delivery of a CRISPR-Cas system in airways of subject or a patient in need thereof, suffering from cystic fibrosis or from cystic fibrosis related symptoms, using adeno-associated virus (AAV) particles.

[0483] Study Design: Subjects or patients in need there of: Human, non-primate human, canine, feline, bovine, equine and other domestic animals, related. This study tests efficacy of gene transfer of a CRISPR-Cas system by a AAV vector. Applicants determine transgene levels sufficient for gene expression and utilize a CRISPR-Cas system comprising a Cas9 enzyme to target deltaF508 or other CFTR-inducing mutations.

[0484] The treated subjects receive pharmaceutically effective amount of aerosolized AAV vector system per lung endobronchially delivered while spontaneously breathing. The control subjects receive equivalent amount of a pseudotyped AAV vector system with an internal control gene. The vector system may be delivered along with a pharmaceutically acceptable or biocompatible pharmaceutical carrier. Three weeks or an appropriate time interval following vector administration, treated subjects are tested for amelioration of cystic fibrosis related symptoms.

[0485] Applicants use an adenovirus or an AAV particle.

[0486] Applicants clone the following gene constructs, each operably linked to one or more regulatory sequences (Cbh or EF1a promoter for Cas9, U6 or H1 promoter for chimeric guide RNA), into one or more adenovirus or AAV vectors or any other compatible vector: A CFTRdelta508 targeting chimeric guide RNA (Figure 13B), a repair template for deltaF508 mutation (Figure 13C) and a codon optimized Cas9 enzyme with optionally one or more nuclear localization signal or sequence(s) (NLS(s)), e.g., two (2) NLSs.

[0487] Identification of Cas9 target site

[0488] Applicants analyzed the human CFTR genomic locus and identified the Cas9 target site (Figure 13A). (PAM may contain a NGG or a NNAGAAW motif).

[0489] Gene Repair Strategy

[0490] Applicants introduce an adenovirus/AAV vector system comprising 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 into the subject via one of the methods of delivery discussed earlier. The CRISPR-Cas system is guided by the CFTRdelta 508 chimeric guide RNA and 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 such as those in Table B.

Example 17: Generation of Gene Knockout Cell Library

[0491] This example demonstrates how to generate a library of cells where each cell has a single gene knocked out:

[0492] Applicants make a library of ES cells where each cell has a single gene knocked out, and the entire library of ES cells will have every single gene knocked out. This library is useful for the screening of gene function in cellular processes as well as diseases.

[0493] To make this cell library, Applicants integrate Cas9 driven by an inducible promoter (e.g. doxycycline inducible promoter) into the ES cell. In addition, Applicants integrate a single guide RNA targeting a specific gene in the ES cell. To make the ES cell library, Applicants simply mix ES cells with a library of genes encoding guide RNAs targeting each gene in the human genome. Applicants first introduce a single BxB1 attB site into the AAVS1 locus of the human ES cell. Then Applicants use the BxB1 integrase to facilitate the integration of individual guide RNA genes into the BxB1 attB site in AAVS1 locus. To facilitate integration, each guide RNA gene is contained on a plasmid that carries of a single attP site. This way BxB1 will recombine the attB site in the genome with the attP site on the guide RNA containing plasmid.

[0494] To generate the cell library, Applicants take the library of cells that have single guide RNAs integrated and induce Cas9 expression. After induction, Cas9 mediates double strand break at sites specified by the guide RNA. To verify the diversity of this cell library, Applicants carry out whole exome sequencing to ensure that Applicants are able to observe mutations in every single targeted gene. This cell library can be used for a variety of applications, including who library-based screens, or can be sorted into individual cell clones to facilitate rapid generation of clonal cell lines with individual human genes knocked out.

Example 18: Engineering of Microalgae using Cas9

[0495] Methods of delivering Cas9

[0496] Method 1: Applicants deliver Cas9 and guide RNA using a vector that expresses Cas9 under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2-tubulin.

[0497] Method 2: Applicants deliver Cas9 and T7 polymerase using vectors that expresses Cas9 and T7 polymerase under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2-tubulin. Guide RNA will be delivered using a vector containing T7 promoter driving the guide RNA.

[0498] Method 3: Applicants deliver Cas9 mRNA and *in vitro* transcribed guide RNA to algae cells. RNA can be *in vitro* transcribed. Cas9 mRNA will consist of the coding region for Cas9 as well as 3'UTR from Cop1 to ensure stabilization of the Cas9 mRNA.

[0499] For Homologous recombination, Applicants provide an additional homology directed repair template.

[0500] Sequence for a cassette driving the expression of Cas9 under the control of beta-2 tubulin promoter, followed by the 3' UTR of Cop1.

TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGA
GACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTC
CTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATA
GCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAAC
ACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGC
TAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACATGTACCCATA
CGATGTTCCAGATTACGCTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCGA
CAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACC

GACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGC
GAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAG
ACGGAAGAACCGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAA
GGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGAT
AAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTAC
CACGAGAAGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACC
GACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCACATGATCAAGTTCC
GGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACA
AGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCAT
CAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAG
CAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCT
GTTCGGCAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAAC
TTCGACCTGGCCGAGGATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGAC
GACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGG
CCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACAC
CGAGATCACCAAGGCCCCCCTGAGCGCCTCTATGATCAAGAGATACGACGAGCA
CCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAG
TACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACG
GCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGA
TGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGA
AGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGC
TGCACGCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCG
GGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTG
GCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATC
ACCCCCTGGAACTTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCA
TCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCA
AGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAA
ATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAA
GGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCT
GAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGC
GTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTA

TCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATA
TCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGA
AAACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGA
GATACACCGGCTGGGGCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACA
AGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAG
AAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAG
AAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTG
GCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGAC
GAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATG
GCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAAT
GAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACA
CCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAG
AATGGGCGGGATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGAC
TACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACA
ACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCT
CCGAAGAGGTCGTGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCA
AGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCC
TGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGC
AGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACG
ACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGC
TGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAA
CTACCACCACGCCCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATC
AAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTAC
GACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCC
AAGTACTTCTTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGG
CCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGG
AGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCA
TGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCA
AAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGG
ACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGT
GCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAA

AGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCAT
CGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAA
GCTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGC
CTCTGCCGGCGAACTGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGT
GAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGAT
AATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATC
ATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGG
ACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGG
CCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTT
CAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGT
GCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGAT
CGACCTGTCTCAGCTGGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGAGGC
CAGCTAAGGATCCGGCAAGACTGGCCCCGCTTGGCAACGCAACAGTGAGCCCCTC
CCTAGTGTGTTTGGGGATGTGACTATGTATTCGTGTGTTGGCCAACGGGTCAACC
CGAACAGATTGATACCCGCCTTGGCATTTCCTGTCAGAATGTAACGTCAGTTGAT
GGTACT

[0501] Sequence for a cassette driving the expression of T7 polymerase under the control of beta-2 tubulin promoter, followed by the 3' UTR of Cop1:

TCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCGGGCTGCGA
GACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCCCGAAGCTC
CTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTGTTTAAATA
GCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCATATTCAAAC
ACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCGCACTCCGC
TAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACatgcctaagaagaaga
ggaaggttaacacgattaacatcgctaagaacgacttctctgacatcgaactggctgctatcccgttcaacactctggctgaccatt
acggtgagcgtttagctcgcgaacagttggcccttgagcatgagtcttacgagatgggtgaagcacgcttccgcaagatgtttga
gcgtcaacttaaagctggtgaggttgcggataacgctgccgccaagcctctcatcactaccctactccctaagatgattgcacgc
atcaacgactggtttgaggaagtgaaagctaagcgcggcaagcgcccgacagccttccagttcctgcaagaaatcaagccgga
agccgtagcgtacatcaccattaagaccactctggcttgcctaaccagtgctgacaatacaaccgttcaggctgtagcaagcgca
atcggtcgggccattgaggacgaggctcgcttcggtcgtatccgtgaccttgaagctaagcacttcaagaaaaacgttgaggaa
caactcaacaagcgcgtagggcacgtctacaagaaagcatttatgcaagttgtcgaggctgacatgctctctaagggtctactcg

gtggcgaggcgtggtcttcgtggcataaggaagactctattcatgtaggagtacgctgcatcgagatgctcattgagtcaaccgg
aatggttagcttacaccgccaaaatgctggcgtagtaggtcaagactctgagactatcgaactcgcacctgaatacgctgaggct
atcgcaacccgtgcaggtgcgctggctggcatctctccgatgttccaaccttgcgtagttcctcctaagccgtggactggcattac
tggtggtggctattgggctaacggtcgtcgtcctctggcgctggtgcgtactcacagtaagaaagcactgatgcgctacgaagac
gtttacatgcctgaggtgtacaaagcgattaacattgcgcaaaacaccgcatggaaaatcaacaagaaagtcctagcggtcgcc
aacgtaatcaccaagtggaagcattgtccggtcgaggacatccctgcgattgagcgtgaagaactcccgatgaaaccggaaga
catcgacatgaatcctgaggctctcaccgcgtggaaacgtgctgccgctgctgtgtaccgcaaggacaaggctcgcaagtctcg
ccgtatcagccttgagttcatgcttgagcaagccaataagtttgctaaccataaggccatctggttcccttacaacatggactggcg
cggtcgtgtttacgctgtgtcaatgttcaacccgcaaggtaacgatatgaccaaaggactgcttacgctggcgaaaggtaaacca
atcggtaaggaaggttactactggctgaaaatccacggtgcaaactgtgcgggtgtcgacaaggttccgttccctgagcgcatca
agttcattgaggaaaaccacgagaacatcatggcttgcgctaagtctccactggagaacacttggtgggctgagcaagattctcc
gttctgcttccttgcgttctgctttgagtacgctggggtacagcaccacggcctgagctataactgctcccttccgctggcgtttgac
gggtcttgctctggcatccagcacttctccgcgatgctccgagatgaggtaggtggtcgcgcggttaacttgcttcctagtgaaac
cgttcaggacatctacgggattgttgctaagaaagtcaacgagattctacaagcagacgcaatcaatgggaccgataacgaagta
gttaccgtgaccgatgagaacactggtgaaatctctgagaaagtcaagctgggcactaaggcactggctggtcaatggctggct
tacggtgttactcgcagtgtgactaagcgttcagtcatgacgctggcttacgggtccaaagagttcggcttccgtcaacaagtgct
ggaagataccattcagccagctattgattccggcaagggtctgatgttcactcagccgaatcaggctgctggatacatggctaag
ctgatttgggaatctgtgagcgtgacggtggtagctgcggttgaagcaatgaactggcttaagtctgctgctaagctgctggctgc
tgaggtcaaagataagaagactggagagattcttcgcaagcgttgcgctgtgcattgggtaactcctgatggtttccctgtgtggc
aggaatacaagaagcctattcagacgcgcttgaacctgatgttcctcggtcagttccgcttacagcctaccattaacaccaacaa
agatagcgagattgatgcacacaaacaggagtctggtatcgctcctaactttgtacacagccaagacggtagccaccttcgtaag
actgtagtgtgggcacacgagaagtacggaatcgaatcttttgcactgattcacgactccttcggtacgattccggctgacgctgc
gaacctgttcaaagcagtgcgcgaaactatggttgacacatatgagtcttgtgatgtactggctgatttctacgaccagttcgctga
ccagttgcacgagtctcaattggacaaaatgccagcacttccggctaaaggtaacttgaacctccgtgacatcttagagtcggact
tcgcgttcgcgtaaGGATCCGGCAAGACTGGCCCCGCTTGGCAACGCAACAGTGAGCCC
CTCCCTAGTGTGTTTGGGGATGTGACTATGTATTCGTGTGTTGGCCAACGGGTCA
ACCCGAACAGATTGATACCCGCCTTGGCATTTCCTGTCAGAATGTAACGTCAGTT
GATGGTACT

[0502] Sequence of guide RNA driven by the T7 promoter (T7 promoter, Ns represent targeting sequence):

gaaat**TAATACGACTCACTATA**NNNNNNNNNNNNNNNNNNNgttttagagc

taGAAAtagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcttttttt

[0503] Gene delivery:

[0504] Chlamydomonas reinhardtii strain CC-124 and CC-125 from the Chlamydomonas Resource Center will be used for electroporation. Electroporation protocol follows standard recommended protocol from the GeneArt Chlamydomonas Engineering kit.

[0505] Also, Applicants generate a line of Chlamydomonas reinhardtii that expresses Cas9 constitutively. This can be done by using pChlamy1 (linearized using PvuI) and selecting for hygromycin resistant colonies. Sequence for pChlamy1 containing Cas9 is below. In this way to achieve gene knockout one simply needs to deliver RNA for the guideRNA. For homologous recombination Applicants deliver guideRNA as well as a linearized homologous recombination template.

[0506] pChlamy1-Cas9:

TGCGGTATTTCACACCGCATCAGGTGGCACTTTTCGGGGAAATGTGC
GCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGA
GATTATCAAAAAGGATCTTCACCTAGATCCTTTTAAATTAAAAATGAAGTTTTAA
ATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACCAATGCTTAATC
AGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACT
CCCCGTCGTGTAGATAACTACGATACGGGAGGGCTTACCATCTGGCCCCAGTGCT
GCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCAGCAATAAACC
AGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCAT
CCAGTCTATTAATTGTTGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGT
TTGCGCAACGTTGTTGCCATTGCTACAGGCATCGTGGTGTCACGCTCGTCGTTTGG
TATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCC
ATGTTGTGCAAAAAAGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTA
AGTTGGCCGCAGTGTTATCACTCATGGTTATGGCAGCACTGCATAATTCTCTTACT
GTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATT
CTGAGAATAGTGTATGCGGCGACCGAGTTGCTCTTGCCCGGCGTCAATACGGGAT
AATACCGCGCCACATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAACGTTCTT
CGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACC
CACTCGTGCACCCAACTGATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGT

GAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAAGGGCGACACGG
AAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGG
TTATTGTCTCATGACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAG
ACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCCTTTTTTTCTGCGCGTAATC
TGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATC
AAGAGCTACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACC
AAATACTGTTCTTCTAGTGTAGCCGTAGTTAGGCCACCACTTCAAGAACTCTGTA
GCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGTTGCCAGTGG
CGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCG
CAGCGGTCGGGCTGAACGGGGGGTTCGTGCACACAGCCCAGCTTGGAGCGAACG
ACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACGCTTC
CCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGA
GAGCGCACGAGGGAGCTTCCAGGGGGAAACGCCTGGTATCTTTATAGTCCTGTCG
GGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAGGGGGGCGG
AGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTG
GCCTTTTGCTCACATGTTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTA
TTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCGAACGACCGAGCGCAG
CGAGTCAGTGAGCGAGGAAGCGGTCGCTGAGGCTTGACATGATTGGTGCGTATGT
TTGTATGAAGCTACAGGACTGATTTGGCGGGCTATGAGGGCGGGGGAAGCTCTGG
AAGGGCCGCGATGGGGCGCGCGGCGTCCAGAAGGCGCCATACGGCCCGCTGGCG
GCACCCATCCGGTATAAAAGCCCGCGACCCCGAACGGTGACCTCCACTTTCAGCG
ACAAACGAGCACTTATACATACGCGACTATTCTGCCGCTATACATAACCACTCAG
CTAGCTTAAGATCCCATCAAGCTTGCATGCCGGGCGCGCCAGAAGGAGCGCAGC
CAAACCAGGATGATGTTTGATGGGGTATTTGAGCACTTGCAACCCTTATCCGGAA
GCCCCCTGGCCCACAAAGGCTAGGCGCCAATGCAAGCAGTTCGCATGCAGCCCCT
GGAGCGGTGCCCTCCTGATAAACCGGCCAGGGGGCCTATGTTCTTTACTTTTTTAC
AAGAGAAGTCACTCAACATCTTAAAATGGCCAGGTGAGTCGACGAGCAAGCCCG
GCGGATCAGGCAGCGTGCTTGCAGATTTGACTTGCAACGCCCGCATTGTGTCGAC
GAAGGCTTTTGGCTCCTCTGTCGCTGTCTCAAGCAGCATCTAACCCTGCGTCGCCG
TTTCCATTTGCAGGAGATTCGAGGTACCATGTACCCATACGATGTTCCAGATTACG
CTTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCGACAAGAAGTACAGCATCG

GCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGTACAA
GGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAA
GAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAGGCCAC
CCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTG
CTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTC
CACAGACTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCAC
CCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCA
TCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGC
TGATCTATCTGGCCCTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGA
GGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGTTCATCCAGCTGGT
GCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGA
CGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCT
GATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAACCTGATTGC
CCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGAT
GCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTG
GCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCG
ACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCC
CCTGAGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTG
CTGAAAGCTCTCGTGCGGCAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCG
ACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAG
AGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAAC
TGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGGAAGCAGCGGACCTTCGACA
ACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGCG
GCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGAT
CCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGCAGA
TTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAG
GAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAAC
TTCGATAAGAACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTAC
GAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGA
ATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCT
GCTGTTCAAGACCAACCGGAAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTT

CAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAGATCGGTTC
AACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACT
TCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGA
CACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCCCACC
TGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGG
GCAGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGA
CAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCT
GATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTC
CGGCCAGGGCGATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGC
CATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGGACGAGCTCGTGAAAGT
GATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCA
GACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAG
AGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCCGTGGAAAACA
CCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATA
TGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACC
ATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACAACAAGGTGCTGAC
CAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCG
TGAAGAAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCC
AGAGAAAGTTCGACAATCTGACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGG
ATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGC
ACGTGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACA
AGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCCAAGCTGGTGTCCGATTT
CCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCC
CACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTA
AGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAGA
TGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTA
CAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATC
CGGAAGCGGCCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGAT
AAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAAT
ATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTG
CCCAAGAGGAACAGCGATAAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAG

AAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGGTGGTGGCCA
AAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGG
ATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAG
CCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAGTACT
CCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAAC
TGCAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACC
TGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAAC
AGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCA
GCGAGTTCTCCAAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTC
CGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGCAGGCCGAGAATATCAT
CCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACA
CCACCATCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCC
TGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATCGACCTGTCTCAGCT
GGGAGGCGACAGCCCCAAGAAGAAGAGAAAGGTGGAGGCCAGCTAACATATGATT
CGAATGTCTTTCTTGCGCTATGACACTTCCAGCAAAAGGTAGGGCG
GGCTGCGAGACGGCTTCCCGGCGCTGCATGCAACACCGATGATGCTTCGACCCCC
CGAAGCTCCTTCGGGGCTGCATGGGCGCTCCGATGCCGCTCCAGGGCGAGCGCTG
TTTAAATAGCCAGGCCCCCGATTGCAAAGACATTATAGCGAGCTACCAAAGCCAT
ATTCAAACACCTAGATCACTACCACTTCTACACAGGCCACTCGAGCTTGTGATCG
CACTCCGCTAAGGGGGCGCCTCTTCCTCTTCGTTTCAGTCACAACCCGCAAACAT
GACACAAGAATCCCTGTTACTTCTCGACCGTATTGATTCGGATGATTCCTACGCG
AGCCTGCGGAACGACCAGGAATTCTGGGAGGTGAGTCGACGAGCAAGCCCGGCG
GATCAGGCAGCGTGCTTGCAGATTTGACTTGCAACGCCCGCATTGTGTCGACGAA
GGCTTTTGGCTCCTCTGTCGCTGTCTCAAGCAGCATCTAACCCTGCGTCGCCGTTT
CCATTTGCAGCCGCTGGCCCGCCGAGCCCTGGAGGAGCTCGGGCTGCCGGTGCCG
CCGGTGCTGCGGGTGCCCGGCGAGAGCACCAACCCCGTACTGGTCGGCGAGCCC
GGCCCGGTGATCAAGCTGTTCGGCGAGCACTGGTGCGGTCCGGAGAGCCTCGCG
TCGGAGTCGGAGGCGTACGCGGTCCTGGCGGACGCCCCGGTGCCGGTGCCCCGC
CTCCTCGGCCGCGGCGAGCTGCGGCCCGGCACCGGAGCCTGGCCGTGGCCCTACC
TGGTGATGAGCCGGATGACCGGCACCACCTGGCGGTCCGCGATGGACGGCACGA
CCGACCGGAACGCGCTGCTCGCCCTGGCCCGCGAACTCGGCCGGGTGCTCGGCCG

GCTGCACAGGGTGCCGCTGACCGGGAACACCGTGCTCACCCCCCATTCCGAGGTC
TTCCCGGAACTGCTGCGGGAACGCCGCGCGGCGACCGTCGAGGACCACCGCGGG
TGGGGCTACCTCTCGCCCCGGCTGCTGGACCGCCTGGAGGACTGGCTGCCGGACG
TGGACACGCTGCTGGCCGGCCGCGAACCCCGGTTCGTCCACGGCGACCTGCACGG
GACCAACATCTTCGTGGACCTGGCCGCGACCGAGGTCACCGGGATCGTCGACTTC
ACCGACGTCTATGCGGGAGACTCCCGCTACAGCCTGGTGCAACTGCATCTCAACG
CCTTCCGGGGCGACCGCGAGATCCTGGCCGCGCTGCTCGACGGGGCGCAGTGGA
AGCGGACCGAGGACTTCGCCCGCGAACTGCTCGCCTTCACCTTCCTGCACGACTT
CGAGGTGTTCGAGGAGACCCCGCTGGATCTCTCCGGCTTCACCGATCCGGAGGAA
CTGGCGCAGTTCCTCTGGGGGCCGCCGGACACCGCCCCCGGCGCCTGATAAGGAT
CCGGCAAGACTGGCCCCGCTTGGCAACGCAACAGTGAGCCCCTCCCTAGTGTGTT
TGGGGATGTGACTATGTATTCGTGTGTTGGCCAACGGGTCAACCCGAACAGATTG
ATACCCGCCTTGGCATTTCCTGTCAGAATGTAACGTCAGTTGATGGTACT

[0507] For all modified Chlamydomonas reinhardtii cells, Applicants use PCR, SURVEYOR nuclease assay, and DNA sequencing to verify successful modification.

Example 19: Use of Cas9 to target a variety of disease types

[0508] **Diseases that involve mutations in protein coding sequence:**

[0509] Dominant disorders may be targeted by inactivating the dominant negative allele. Applicants use Cas9 to target a unique sequence in the dominant negative allele and introduce a mutation via NHEJ. The NHEJ-induced indel may be able to introduce a frame-shift mutation in the dominant negative allele and eliminate the dominant negative protein. This may work if the gene is haplo-sufficient (e.g. MYOC mutation induced glaucoma and Huntington's disease).

[0510] Recessive disorders may be targeted by repairing the disease mutation in both alleles. For dividing cells, Applicants use Cas9 to introduce double strand breaks near the mutation site and increase the rate of homologous recombination using an exogenous recombination template. For dividing cells, this may be achieved using multiplexed nickase activity to catalyze the replacement of the mutant sequence in both alleles via NHEJ-mediated ligation of an exogenous DNA fragment carrying complementary overhangs.

[0511] Applicants also use Cas9 to introduce protective mutations (e.g. inactivation of CCR5 to prevent HIV infection, inactivation of PCSK9 for cholesterol reduction, or introduction of the A673T into APP to reduce the likelihood of Alzheimer's disease).

[0512] Diseases that involve non-coding sequences

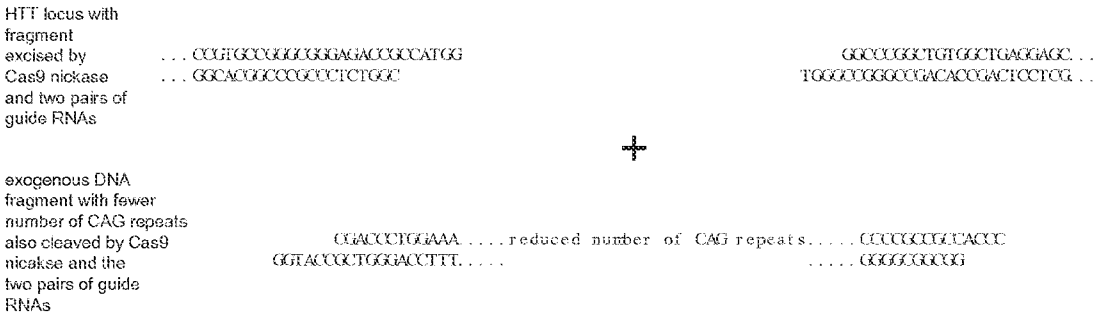
[0513] Applicants use Cas9 to disrupt non-coding sequences in the promoter region, to alter transcription factor binding sites and alter enhancer or repressor elements. For example, Cas9 may be used to excise out the Klf1 enhancer EHS1 in hematopoietic stem cells to reduce BCL11a levels and reactivate fetal globin gene expression in differentiated erythrocytes

[0514] Applicants also use Cas9 to disrupt functional motifs in the 5' or 3' untranslated regions. For example, for the treatment of myotonic dystrophy, Cas9 may be used to remove CTG repeat expansions in the DMPK gene.

Example 20: Multiplexed Nickase

[0515] Aspects of optimization and the teachings of Cas9 detailed in this application may also be used to generate Cas9 nickases. Applicants use Cas9 nickases in combination with pairs of guide RNAs to generate DNA double strand breaks with defined overhangs. When two pairs of guide RNAs are used, it is possible to excise an intervening DNA fragment. If an exogenous piece of DNA is cleaved by the two pairs of guide RNAs to generate compatible overhangs with the genomic DNA, then the exogenous DNA fragment may be ligated into the genomic DNA to replace the excised fragment. For example, this may be used to remove trinucleotide repeat expansion in the huntintin (HTT) gene to treat Huntington's Disease.

[0516] If an exogenous DNA that bears fewer number of CAG repeats is provided, then it may be able to generate a fragment of DNA that bears the same overhangs and can be ligated into the HTT genomic locus and replace the excised fragment.



[0517] The ligation of the exogenous DNA fragment into the genome does not require homologous recombination machineries and therefore this method may be used in post-mitotic cells such as neurons.

Example 21: Delivery of CRISPR System

[0518] Cas9 and its chimeric guide RNA, or combination of tracrRNA and crRNA, can be delivered either as DNA or RNA. Delivery of Cas9 and guide RNA both as RNA (normal or containing base or backbone modifications) molecules can be used to reduce the amount of time that Cas9 protein persist in the cell. This may reduce the level of off-target cleavage activity in the target cell. Since delivery of Cas9 as mRNA takes time to be translated into protein, it might be advantageous to deliver the guide RNA several hours following the delivery of Cas9 mRNA, to maximize the level of guide RNA available for interaction with Cas9 protein.

[0519] In situations where guide RNA amount is limiting, it may be desirable to introduce Cas9 as mRNA and guide RNA in the form of a DNA expression cassette with a promoter driving the expression of the guide RNA. This way the amount of guide RNA available will be amplified via transcription.

[0520] A variety of delivery systems can be introduced to introduce Cas9 (DNA or RNA) and guide RNA (DNA or RNA) into the host cell. These include the use of liposomes, viral vectors, electroporation, nanoparticles , nanowires (Shalek et al., Nano Letters, 2012), exosomes. Molecular trojan horses liposomes (Pardridge et al., Cold Spring Harb Protoc; 2010; doi:10.1101/pdb.prot5407) may be used to deliver Cas9 and guide RNA across the blood brain barrier.

Example 22: Cas9 orthologs

[0521] Applicants analyzed Cas9 orthologs (Figures 3 and 4A-F) to identify the relevant PAM sequences and the corresponding chimeric guide RNAs. This expanded set of PAMs may provide broader targeting across the genome and also significantly increases the number of unique target sites and provides potential for identifying novel Cas9s with increased levels of specificity in the genome. Applicants determined the PAM for *Staphylococcus aureus sp. Aureus* Cas9 to be NNGRR *Staphylococcus aureus sp. Aureus* Cas9 is also known as SaCas9.

[0522] The specificity of Cas9 orthologs can be evaluated by testing the ability of each Cas9 to tolerate mismatches between the guide RNA and its DNA target. For example, the specificity of SpCas9 has been characterized by testing the effect of mutations in the guide RNA on cleavage efficiency. Libraries of guide RNAs were made with single or multiple mismatches between the guide sequence and the target DNA. Based on these findings, target sites for SpCas9 can be selected based on the following guidelines:

[0523] To maximize SpCas9 specificity for editing a particular gene, one should choose a target site within the locus of interest such that potential 'off-target' genomic sequences abide by the following four constraints: First and foremost, they should not be followed by a PAM with either 5'-NGG or NAG sequences. Second, their global sequence similarity to the target sequence should be minimized. Third, a maximal number of mismatches should lie within the PAM-proximal region of the off-target site. Finally, a maximal number of mismatches should be consecutive or spaced less than four bases apart.

[0524] Similar methods can be used to evaluate the specificity of other Cas9 orthologs and to establish criteria for the selection of specific target sites within the genomes of target species.

Example 23: Therapeutic strategies for Trinucleotide repeat disorders

[0525] As previously mentioned in the application, the target polynucleotide of a CRISPR complex may include a number of disease-associated genes and polynucleotides and some of these disease associated gene may belong to a set of genetic disorders referred to as Trinucleotide repeat disorders (referred to as also trinucleotide repeat expansion disorders, triplet repeat expansion disorders or codon reiteration disorders).

[0526] These diseases are caused by mutations in which the trinucleotide repeats of certain genes exceed the normal, stable threshold which may usually differ in a gene. The discovery of more repeat expansion disorders has allowed for the classification of these disorders into a number of categories based on underlying similar characteristics. Huntington's disease (HD) and the spinocerebellar ataxias that are caused by a CAG repeat expansion in protein-coding portions of specific genes are included in Category I. Diseases or disorders with expansions that tend to make them phenotypically diverse and include expansions are usually small in magnitude and also found in exons of genes are included in Category II. Category III includes disorders or diseases which are characterized by much larger repeat expansions than either Category I or II

and are generally located outside protein coding regions. Examples of Category III diseases or disorders include but are not limited to Fragile X syndrome, myotonic dystrophy, two of the spinocerebellar ataxias, juvenile myoclonic epilepsy, and Friedreich's ataxia.

[0527] Similar therapeutic strategies, like the one mentioned for Friedreich's ataxia below may be adopted to address other trinucleotide repeat or expansion disorders as well. For example, another triple repeat disease that can be treated using almost identical strategy is dystrophia myotonica 1 (DM1), where there is an expanded CTG motif in the 3' UTR. In Friedreich's ataxia, the disease results from expansion of GAA trinucleotides in the first intron of frataxin (FXN). One therapeutic strategy using CRISPR is to excise the GAA repeat from the first intron. The expanded GAA repeat is thought to affect the DNA structure and leads to recruit the formation of heterochromatin which turn off the frataxin gene (Figure 14A).

[0528] Competitive Advantage over other therapeutic strategies are listed below:

[0529] siRNA knockdown is not applicable in this case, as disease is due to reduced expression of frataxin. Viral gene therapy is currently being explored. HSV-1 based vectors were used to deliver the frataxin gene in animal models and have shown therapeutic effect. However, long term efficacy of virus-based frataxin delivery suffer from several problems: First, it is difficult to regulate the expression of frataxin to match natural levels in health individuals, and second, long term over expression of frataxin leads to cell death.

[0530] Nucleases may be used to excise the GAA repeat to restore healthy genotype, but Zinc Finger Nuclease and TALEN strategies require delivery of two pairs of high efficacy nucleases, which is difficult for both delivery as well as nuclease engineering (efficient excision of genomic DNA by ZFN or TALEN is difficult to achieve).

[0531] In contrast to above strategies, the CRISPR-Cas system has clear advantages. The Cas9 enzyme is more efficient and more multiplexible, by which it is meant that one or more targets can be set at the same time. So far, efficient excision of genomic DNA > 30% by Cas9 in human cells and may be as high as 30%, and may be improved in the future. Furthermore, with regard to certain trinucleotide repeat disorders like Huntington's disease (HD), trinucleotide repeats in the coding region may be addressed if there are differences between the two alleles. Specifically, if a HD patient is heterozygous for mutant HTT and there are nucleotide differences such as SNPs between the wt and mutant HTT alleles, then Cas9 may be used to specifically

target the mutant HTT allele. ZFN or TALENs will not have the ability to distinguish two alleles based on single base differences.

[0532] In adopting a strategy using the CRISPR-Cas 9 enzyme to address Friedreich's ataxia, Applicants design a number of guide RNAs targeting sites flanking the GAA expansion and the most efficient and specific ones are chosen (Figure 14B).

[0533] Applicants deliver a combination of guide RNAs targeting the intron 1 of FXN along with Cas9 to mediate excision of the GAA expansion region. AAV9 may be used to mediate efficient delivery of Cas9 and in the spinal cord.

[0534] If the Alu element adjacent to the GAA expansion is considered important, there may be constraints to the number of sites that can be targeted but Applicants may adopt strategies to avoid disrupting it.

[0535] Alternative Strategies:

[0536] Rather than modifying the genome using Cas9, Applicants may also directly activate the FXN gene using Cas9 (nuclease activity deficient)-based DNA binding domain to target a transcription activation domain to the FXN gene.

Example 24: Strategies for minimizing off-target cleavage using Cas9 nickase

[0537] As previously mentioned in the application, Cas9 may be mutated to mediate single strand cleavage via one or more of the following mutations: D10A, E762A, and H840A.

[0538] To mediate gene knockout via NHEJ, Applicants use a nickase version of Cas9 along with two guide RNAs. Off-target nicking by each individual guide RNA may be primarily repaired without mutation, double strand breaks (which can lead to mutations via NHEJ) only occur when the target sites are adjacent to each other. Since double strand breaks introduced by double nicking are not blunt, co-expression of end-processing enzymes such as TREX1 will increase the level of NHEJ activity.

[0539] The following list of targets in tabular form are for genes involved in the following diseases:

[0540] Lafora's Disease – target GSY1 or PPP1R3C (PTG) to reduce glycogen in neurons.

[0541] Hypercholesterolemia – target PCSK9

[0542] Target sequences are listed in pairs (L and R) with different number of nucleotides in the spacer (0 to 3bp). Each spacer may also be used by itself with the wild type Cas9 to introduce double strand break at the target locus.

GYS1 (human)	GGCC-L	ACCCTTGTTAGCCACCTCCC
	GGCC-R	GAACGCAGTGCTCTTCGAAG
	GGNCC-L	CTCACGCCCTGCTCCGTGTA
	GGNCC-R	GGCGACAACTACTTCCTGGT
	GGNNCC-L	CTCACGCCCTGCTCCGTGTA
	GGNNCC-R	GGGCGACAACTACTTCCTGG
	GGNNNCC-L	CCTCTTCAGGGCCGGGGTGG
	GGNNNCC-R	GAGGACCCAGGTGGAACTGC
PCSK9 (human)	GGCC-L	TCAGCTCCAGGCGGTCCTGG
	GGCC-R	AGCAGCAGCAGCAGTGGCAG
	GGNCC-L	TGGGCACCGTCAGCTCCAGG
	GGNCC-R	CAGCAGTGGCAGCGGCCACC
	GGNNCC-L	ACCTCTCCCCTGGCCCTCAT
	GGNNCC-R	CCAGGACCGCCTGGAGCTGA
	GGNNNCC-L	CCGTCAGCTCCAGGCGGTCC
	GGNNNCC-R	AGCAGCAGCAGCAGTGGCAG
PPP1R3C (PTG) (human)	GGCC-L	ATGTGCCAAGCAAAGCCTCA
	GGCC-R	TTCGGTCATGCCCGTGGATG
	GGNCC-L	GTCGTTGAAATTCATCGTAC
	GGNCC-R	ACCACCTGTGAAGAGTTTCC
	GGNNCC-L	CGTCGTTGAAATTCATCGTA
	GGNNCC-R	ACCACCTGTGAAGAGTTTCC
Gys1 (mouse)	GGCC-L	GAACGCAGTGCTTTTCGAGG
	GGCC-R	ACCCTTGTTGGCCACCTCCC
	GGNCC-L	GGTGACAACTACTATCTGGT
	GGNCC-R	CTCACACCCTGCTCCGTGTA
	GGNNCC-L	GGGTGACAACTACTATCTGG
	GGNNCC-R	CTCACACCCTGCTCCGTGTA
	GGNNNCC-L	CGAGAACGCAGTGCTTTTCG
	GGNNNCC-R	ACCCTTGTTGGCCACCTCCC
PPP1R3C (PTG) (mouse)	GGCC-L	ATGAGCCAAGCAAATCCTCA
	GGCC-R	TTCCGTCATGCCCGTGGACA
	GGNCC-L	CTTCGTTGAAAACCATTGTA
	GGNCC-R	CCACCTCTGAAGAGTTTCCT
	GGNNCC-L	CTTCGTTGAAAACCATTGTA
	GGNNCC-R	ACCACCTCTGAAGAGTTTCC
	GGNNNCC-L	CTTCCACTCACTCTGCGATT

PCSK9 (mouse)	GGNNNCC-R	ACCATGTCTCAGTGTCAAGC
	GGCC-L	GGCGGCAACAGCGGCAACAG
	GGCC-R	ACTGCTCTGCGTGGCTGCGG
	GGNNCC-L	CCGCAGCCACGCAGAGCAGT
	GGNNCC-R	GCACCTCTCCTCGCCCCGAT

[0543] Alternative strategies for improving stability of guide RNA and increasing specificity

[0544] 1. Nucleotides in the 5' of the guide RNA may be linked via thiolester linkages rather than phosphoester linkage like in natural RNA. Thiolester linkage may prevent the guide RNA from being digested by endogenous RNA degradation machinery.

[0545] 2. Nucleotides in the guide sequence (5' 20bp) of the guide RNA can use bridged nucleic acids (BNA) as the bases to improve the binding specificity.

Example 25: CRISPR-Cas for rapid, multiplex genome editing

[0546] Aspects of the invention relate to protocols and methods by which efficiency and specificity of gene modification may be tested within 3-4 days after target design, and modified clonal cell lines may be derived within 2-3 weeks.

[0547] Programmable nucleases are powerful technologies for mediating genome alteration with high precision. The RNA-guided Cas9 nuclease from the microbial CRISPR adaptive immune system can be used to facilitate efficient genome editing in eukaryotic cells by simply specifying a 20-nt targeting sequence in its guide RNA. Applicants describe a set of protocols for applying Cas9 to facilitate efficient genome editing in mammalian cells and generate cell lines for downstream functional studies. Beginning with target design, efficient and specific gene modification can be achieved within 3-4 days, and modified clonal cell lines can be derived within 2-3 weeks.

[0548] The ability to engineer biological systems and organisms holds enormous potential for applications across basic science, medicine, and biotechnology. Programmable sequence-specific endonucleases that facilitate precise editing of endogenous genomic loci are now enabling systematic interrogation of genetic elements and causal genetic variations in a broad range of species, including those that have not been genetically tractable previously. A number of genome editing technologies have emerged in recent years, including zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and the RNA-guided CRISPR-

Cas nuclease system. The first two technologies use a common strategy of tethering endonuclease catalytic domains to modular DNA-binding proteins for inducing targeted DNA double stranded breaks (DSB) at specific genomic loci. By contrast, Cas9 is a nuclease guided by small RNAs through Watson-Crick base-pairing with target DNA, presenting a system that is easy to design, efficient, and well-suited for high-throughput and multiplexed gene editing for a variety of cell types and organisms. Here Applicants describe a set of protocols for applying the recently developed Cas9 nuclease to facilitate efficient genome editing in mammalian cells and generate cell lines for downstream functional studies.

[0549] Like ZFNs and TALENs, Cas9 promotes genome editing by stimulating DSB at the target genomic loci. Upon cleavage by Cas9, the target locus undergoes one of two major pathways for DNA damage repair, the error-prone non-homologous end joining (NHEJ) or the high-fidelity homology directed repair (HDR) pathway. Both pathways may be utilized to achieve the desired editing outcome.

[0550] NHEJ: In the absence of a repair template, the NHEJ process re-ligates DSBs, which may leave a scar in the form of indel mutations. This process can be harnessed to achieve gene knockouts, as indels occurring within a coding exon may lead to frameshift mutations and a premature stop codon. Multiple DSBs may also be exploited to mediate larger deletions in the genome.

[0551] HDR: Homology directed repair is an alternate major DNA repair pathway to NHEJ. Although HDR typically occurs at lower frequencies than NHEJ, it may be harnessed to generate precise, defined modifications at a target locus in the presence of an exogenously introduced repair template. The repair template may be either in the form of double stranded DNA, designed similarly to conventional DNA targeting constructs with homology arms flanking the insertion sequence, or single-stranded DNA oligonucleotides (ssODNs). The latter provides an effective and simple method for making small edits in the genome, such as the introduction of single nucleotide mutations for probing causal genetic variations. Unlike NHEJ, HDR is generally active only in dividing cells and its efficiency varies depending on the cell type and state.

[0552] Overview of CRISPR: The CRISPR-Cas system, by contrast, is at minimum a two-component system consisting of the Cas9 nuclease and a short guide RNA. Re-targeting of Cas9 to different loci or simultaneous editing of multiple genes simply requires cloning a different 20-bp oligonucleotide. Although specificity of the Cas9 nuclease has yet to be thoroughly

elucidated, the simple Watson-Crick base-pairing of the CRISPR-Cas system is likely more predictable than that of ZFN or TALEN domains.

[0553] The type II CRISPR-Cas (clustered regularly interspaced short palindromic repeats) is a bacterial adaptive immune system that uses Cas9, to cleave foreign genetic elements. Cas9 is guided by a pair of non-coding RNAs, a variable crRNA and a required auxiliary tracrRNA. The crRNA contains a 20-nt guide sequence determines specificity by locating the target DNA via Watson-Crick base-pairing. In the native bacterial system, multiple crRNAs are co-transcribed to direct Cas9 against various targets. In the CRISPR-Cas system derived from *Streptococcus pyogenes*, the target DNA must immediately precede a 5'-NGG/NRG protospacer adjacent motif (PAM), which can vary for other CRISPR systems.

[0554] CRISPR-Cas is reconstituted in mammalian cells through the heterologous expression of human codon-optimized Cas9 and the requisite RNA components. Furthermore, the crRNA and tracrRNA can be fused to create a chimeric, synthetic guide RNA (sgRNA). Cas9 can thus be re-directed toward any target of interest by altering the 20-nt guide sequence within the sgRNA.

[0555] Given its ease of implementation and multiplex capability, Cas9 has been used to generate engineered eukaryotic cells carrying specific mutations via both NHEJ and HDR. In addition, direct injection of sgRNA and mRNA encoding Cas9 into embryos has enabled the rapid generation of transgenic mice with multiple modified alleles; these results hold promise for editing organisms that are otherwise genetically intractable.

[0556] A mutant Cas9 carrying a disruption in one of its catalytic domains has been engineered to nick rather than cleave DNA, allowing for single-stranded breaks and preferential repair through HDR, potentially ameliorating unwanted indel mutations from off-target DSBs. Additionally, a Cas9 mutant with both DNA-cleaving catalytic residues mutated has been adapted to enable transcriptional regulation in *E. coli*, demonstrating the potential of functionalizing Cas9 for diverse applications. Certain aspects of the invention relate to the construction and application of Cas9 for multiplexed editing of human cells.

[0557] Applicants have provided a human codon-optimized, nuclear localization sequence-flanked Cas9 to facilitate eukaryotic gene editing. Applicants describe considerations for designing the 20-nt guide sequence, protocols for rapid construction and functional validation of sgRNAs, and finally use of the Cas9 nuclease to mediate both NHEJ- and HDR-based genome

modifications in human embryonic kidney (HEK-293FT) and human stem cell (HUES9) lines. This protocol can likewise be applied to other cell types and organisms.

[0558] Target selection for sgRNA: There are two main considerations in the selection of the 20-nt guide sequence for gene targeting: 1) the target sequence should precede the 5'-NGG PAM for S. pyogenes Cas9, and 2) guide sequences should be chosen to minimize off-target activity. Applicants provided an online Cas9 targeting design tool that takes an input sequence of interest and identifies suitable target sites. To experimentally assess off-target modifications for each sgRNA, Applicants also provide computationally predicted off-target sites for each intended target, ranked according to Applicants' quantitative specificity analysis on the effects of base-pairing mismatch identity, position, and distribution.

[0559] The detailed information on computationally predicted off-target sites is as follows:

[0560] Considerations for Off-target Cleavage Activities: Similar to other nucleases, Cas9 can cleave off-target DNA targets in the genome at reduced frequencies. The extent to which a given guide sequence exhibit off-target activity depends on a combination of factors including enzyme concentration, thermodynamics of the specific guide sequence employed, and the abundance of similar sequences in the target genome. For routine application of Cas9, it is important to consider ways to minimize the degree of off-target cleavage and also to be able to detect the presence of off-target cleavage.

[0561] Minimizing off-target activity: For application in cell lines, Applicants recommend following two steps to reduce the degree of off-target genome modification. First, using our online CRISPR target selection tool, it is possible to computationally assess the likelihood of a given guide sequence to have off-target sites. These analyses are performed through an exhaustive search in the genome for off-target sequences that are similar sequences as the guide sequence. Comprehensive experimental investigation of the effect of mismatching bases between the sgRNA and its target DNA revealed that mismatch tolerance is 1) position dependent – the 8-14 bp on the 3' end of the guide sequence are less tolerant of mismatches than the 5' bases, 2) quantity dependent – in general more than 3 mismatches are not tolerated, 3) guide sequence dependent – some guide sequences are less tolerant of mismatches than others, and 4) concentration dependent – off-target cleavage is highly sensitive to the amount of transfected DNA. The Applicants' target site analysis web tool (available at the website genome-engineering.org/tools) integrates these criteria to provide predictions for likely off-target sites in

the target genome. Second, Applicants recommend titrating the amount of Cas9 and sgRNA expression plasmid to minimize off-target activity.

[0562] Detection of off-target activities: Using Applicants' CRISPR targeting web tool, it is possible to generate a list of most likely off-target sites as well as primers performing SURVEYOR or sequencing analysis of those sites. For isogenic clones generated using Cas9, Applicants strongly recommend sequencing these candidate off-target sites to check for any undesired mutations. It is worth noting that there may be off target modifications in sites that are not included in the predicted candidate list and full genome sequence should be performed to completely verify the absence of off-target sites. Furthermore, in multiplex assays where several DSBs are induced within the same genome, there may be low rates of translocation events and can be evaluated using a variety of techniques such as deep sequencing.

[0563] The online tool provides the sequences for all oligos and primers necessary for 1) preparing the sgRNA constructs, 2) assaying target modification efficiency, and 3) assessing cleavage at potential off-target sites. It is worth noting that because the U6 RNA polymerase III promoter used to express the sgRNA prefers a guanine (G) nucleotide as the first base of its transcript, an extra G is appended at the 5' of the sgRNA where the 20-nt guide sequence does not begin with G.

[0564] Approaches for sgRNA construction and delivery: Depending on the desired application, sgRNAs may be delivered as either 1) PCR amplicons containing an expression cassette or 2) sgRNA-expressing plasmids. PCR-based sgRNA delivery appends the custom sgRNA sequence onto the reverse PCR primer used to amplify a U6 promoter template. The resulting amplicon may be co-transfected with a plasmid containing Cas9 (PX165). This method is optimal for rapid screening of multiple candidate sgRNAs, as cell transfections for functional testing can be performed mere hours after obtaining the sgRNA-encoding primers. Because this simple method obviates the need for plasmid-based cloning and sequence verification, it is well suited for testing or co-transfecting a large number of sgRNAs for generating large knockout libraries or other scale-sensitive applications. Note that the sgRNA-encoding primers are over 100-bp, compared to the ~20-bp oligos required for plasmid-based sgRNA delivery.

[0565] Construction of an expression plasmid for sgRNA is also simple and rapid, involving a single cloning step with a pair of partially complementary oligonucleotides. After annealing the oligo pairs, the resulting guide sequences may be inserted into a plasmid bearing both Cas9 and

an invariant scaffold bearing the remainder of the sgRNA sequence (PX330). The transfection plasmids may also be modified to enable virus production for *in vivo* delivery.

[0566] In addition to PCR and plasmid-based delivery methods, both Cas9 and sgRNA can be introduced into cells as RNA.

[0567] Design of repair template: Traditionally, targeted DNA modifications have required use of plasmid-based donor repair templates that contain homology arms flanking the site of alteration. The homology arms on each side can vary in length, but are typically longer than 500 bp. This method can be used to generate large modifications, including insertion of reporter genes such as fluorescent proteins or antibiotic resistance markers. The design and construction of targeting plasmids has been described elsewhere.

[0568] More recently, single-stranded DNA oligonucleotides (ssODNs) have been used in place of targeting plasmids for short modifications within a defined locus without cloning. To achieve high HDR efficiencies, ssODNs contain flanking sequences of at least 40 bp on each side that are homologous to the target region, and can be oriented in either the sense or antisense direction relative to the target locus.

[0569] Functional testing

[0570] SURVEYOR nuclease assay: Applicants detected indel mutations either by the SURVEYOR nuclease assay (or PCR amplicon sequencing. Applicants online CRISPR target design tool provides recommended primers for both approaches. However, SURVEYOR or sequencing primers may also be designed manually to amplify the region of interest from genomic DNA and to avoid non-specific amplicons using NCBI Primer-BLAST. SURVEYOR primers should be designed to amplify 300-400 bp (for a 600-800 bp total amplicon) on either side of the Cas9 target for allowing clear visualization of cleavage bands by gel electrophoresis. To prevent excessive primer dimer formation, SURVEYOR primers should be designed to be typically under 25-nt long with melting temperatures of ~60°C. Applicants recommend testing each pair of candidate primers for specific PCR amplicons as well as for the absence of non-specific cleavage during the SURVEYOR nuclease digestion process.

[0571] Plasmid- or ssODN-mediated HDR: HDR can be detected via PCR-amplification and sequencing of the modified region. PCR primers for this purpose should anneal outside the region spanned by the homology arms to avoid false detection of residual repair template (HDR Fwd and Rev, Figure 12). For ssODN-mediated HDR, SURVEYOR PCR primers can be used.

[0572] Detection of indels or HDR by sequencing: Applicants detected targeted genome modifications by either Sanger or next-generation deep sequencing (NGS). For the former, genomic DNA from modified region can be amplified using either SURVEYOR or HDR primers. Amplicons should be subcloned into a plasmid such as pUC19 for transformation; individual colonies can be sequenced to reveal clonal genotype.

[0573] Applicants designed next-generation sequencing (NGS) primers for shorter amplicons, typically in the 100-200 bp size range. For detecting NHEJ mutations, it is important to design primers with at least 10-20 bp between the priming regions and the Cas9 target site to allow detection of longer indels. Applicants provide guidelines for a two-step PCR method to attach barcoded adapters for multiplex deep sequencing. Applicants recommend the Illumina platform, due to its generally low levels of false positive indels. Off-target analysis (described previously) can then be performed through read alignment programs such as ClustalW, Geneious, or simple sequence analysis scripts.

[0574] Materials and Reagents

[0575] **sgRNA preparation:**

[0576] UltraPure DNaseRNase-free distilled water (Life Technologies, cat. no. 10977-023)

[0577] Herculase II fusion polymerase (Agilent Technologies, cat. no. 600679)

[0578] CRITICAL. Standard Taq polymerase, which lacks 3'-5' exonuclease proofreading activity, has lower fidelity and can lead to amplification errors. Herculase II is a high-fidelity polymerase (equivalent fidelity to Pfu) that produces high yields of PCR product with minimal optimization. Other high-fidelity polymerases may be substituted.

[0579] Herculase II reaction buffer (5x; Agilent Technologies, included with polymerase)

[0580] dNTP solution mix (25 mM each; Enzymatics, cat. no. N205L)

[0581] MgCl2 (25mM; ThermoScientific, cat. no. R0971)

[0582] QIAquick gel extraction kit (Qiagen, cat. no. 28704)

[0583] QIAprep spin miniprep kit (Qiagen, cat. no. 27106)

[0584] UltraPure TBE buffer (10X; Life Technologies, cat. no. 15581-028)

[0585] SeaKem LE agarose (Lonza, cat. no. 50004)

[0586] SYBR Safe DNA stain (10,000x; Life Technologies, cat. no. S33102)

[0587] 1-kb Plus DNA ladder (Life Technologies, cat. no. 10787-018)

[0588] TrackIt CyanOrange loading buffer (Life Technologies, cat. no. 10482-028)

[0589] FastDigest BbsI (BpiI) (Fermentas/ThermoScientific, cat. no. FD1014)

[0590] Fermentas Tango Buffer (Fermentas/ThermoScientific, cat. no. BY5)

[0591] DL-dithiothreitol (DTT; Fermentas/ThermoScientific, cat. no. R0862)

[0592] T7 DNA ligase (Enzymatics, cat. no. L602L)

[0593] Critical:Do not substitute the more commonly used T4 ligase. T7 ligase has 1,000-fold higher activity on the sticky ends than on the blunt ends and higher overall activity than commercially available concentrated T4 ligases.

[0594] T7 2X Rapid Ligation Buffer (included with T7 DNA ligase, Enzymatics, cat. no. L602L)

[0595] T4 Polynucleotide Kinase (New England Biolabs, cat. no M0201S)

[0596] T4 DNA Ligase Reaction Buffer (10X; New England Biolabs, cat. no B0202S)

[0597] Adenosine 5'-triphosphate (10 mM; New England Biolabs, cat. no. P0756S)

[0598] PlasmidSafe ATP-dependent DNase (Epicentre, cat. no. E3101K)

[0599] One Shot Stbl3 chemically competent Escherichia coli (E. coli) (Life Technologies, cat. no. C7373-03)

[0600] SOC medium (New England Biolabs, cat. no. B9020S)

[0601] LB medium (Sigma, cat. no. L3022)

[0602] LB agar medium (Sigma, cat. no. L2897)

[0603] Ampicillin, sterile filtered (100 mg ml−1; Sigma, cat. no. A5354)

[0604] **Mammalian cell culture:**

[0605] HEK293FT cells (Life Technologies, cat. no. R700-07)

[0606] Dulbecco's minimum Eagle's medium (DMEM, 1X, high glucose; Life Technologies, cat. no. 10313-039)

[0607] Dulbecco's minimum Eagle's medium (DMEM, 1X, high glucose, no phenol red; Life Technologies, cat. no. 31053-028)

[0608] Dulbecco's phosphate-buffered saline (DPBS, 1X; Life Technologies, cat. no. 14190-250)

[0609] Fetal bovine serum, qualified and heat inactivated (Life Technologies, cat. no. 10438-034)

[0610] Opti-MEM I reduced-serum medium (FBS; Life Technologies, cat. no. 11058-021)

[0611] Penicillin-streptomycin (100x; Life Technologies, cat. no. 15140-163)

[0612] TrypLE™ Express (1X, no Phenol Red; Life Technologies, cat. no. 12604-013)

[0613] Lipofectamine 2000 transfection reagent (Life Technologies, cat. no. 11668027)

[0614] Amaxa SF Cell Line 4D-Nucleofector® X Kit S (32 RCT; Lonza, cat. no V4XC-2032)

[0615] HUES 9 cell line (HARVARD STEM CELL SCIENCE)

[0616] Geltrex LDEV-Free Reduced Growth Factor Basement Membrane Matrix (Life Technologies, cat. no. A1413201)

[0617] mTeSR1 medium (Stemcell Technologies, cat. no. 05850)

[0618] Accutase cell detachment solution (Stemcell Technologies, cat. no. 07920)

[0619] ROCK Inhibitor (Y-27632; Millipore, cat. no. SCM075)

[0620] Amaxa P3 Primary Cell 4D-Nucleofector® X Kit S (32 RCT; Lonza cat. no. V4XP-3032)

[0621] **Genotyping analysis:**

[0622] QuickExtract DNA extraction solution (Epicentre, cat. no. QE09050)

[0623] PCR primers for SURVEYOR, RFLP analysis, or sequencing (see Primer table)

[0624] Herculase II fusion polymerase (Agilent Technologies, cat. no. 600679)

[0625] CRITICAL. As Surveyor assay is sensitive to single-base mismatches, it is particularly important to use a high-fidelity polymerase. Other high-fidelity polymerases may be substituted.

[0626] Herculase II reaction buffer (5x; Agilent Technologies, included with polymerase)

[0627] dNTP solution mix (25 mM each; Enzymatics, cat. no. N205L)

[0628] QIAquick gel extraction kit (Qiagen, cat. no. 28704)

[0629] Taq Buffer (10x; Genscript, cat. no. B0005)

[0630] SURVEYOR mutation detection kit for standard gel electrophoresis (Transgenomic, cat. no. 706025)

[0631] UltraPure TBE buffer (10x; Life Technologies, cat. no. 15581-028)

[0632] SeaKem LE agarose (Lonza, cat. no. 50004)

[0633] 4-20% TBE Gels 1.0 mm, 15 Well (Life Technologies, cat. no. EC62255BOX)

[0634] Novex® Hi-Density TBE Sample Buffer (5X; Life Technologies, cat. no. LC6678)

[0635] SYBR Gold Nucleic Acid Gel Stain (10,000X; Life Technologies, cat. no. S-11494)

[0636] 1-kb Plus DNA ladder (Life Technologies, cat. no. 10787-018)

[0637] TrackIt CyanOrange loading buffer (Life Technologies, cat. no. 10482-028)

[0638] FastDigest HindIII (Fermentas/ThermoScientific, cat. no. FD0504)

[0639] **Equipment**

[0640] Filtered sterile pipette tips (Corning)

[0641] Standard 1.5ml microcentrifuge tubes (Eppendorf, cat. no. 0030 125.150)

[0642] Axygen 96-well PCR plates (VWR, cat. no. PCR-96M2-HSC)

[0643] Axygen 8-Strip PCR tubes (Fischer Scientific, cat. no. 14-222-250)

[0644] Falcon tubes, polypropylene, 15 ml (BD Falcon, cat. no. 352097)

[0645] Falcon tubes, polypropylene, 50 ml (BD Falcon, cat. no. 352070)

[0646] Round-bottom Tube with cell strainer cap, 5ml (BD Falcon, cat. no. 352235)

[0647] Petri dishes (60 mm × 15 mm; BD Biosciences, cat. no. 351007)

[0648] Tissue culture plate (24 well; BD Falcon, cat. no. 353047)

[0649] Tissue culture plate (96 well, flat bottom; BD Falcon, cat. no. 353075)

[0650] Tissue culture dish (100 mm; BD Falcon, 353003)

[0651] 96-well thermocycler with programmable temperature stepping functionality (Applied Biosystems Veriti, cat. no. 4375786).

[0652] Desktop microcentrifuges 5424, 5804 (Eppendorf)

[0653] Gel electrophoresis system (PowerPac basic power supply, Bio-Rad, cat. no. 164-5050, and Sub-Cell GT System gel tray, Bio-Rad, cat. no. 170-4401)

[0654] Novex XCell SureLock Mini-Cell (Life Technologies, cat. no. EI0001)

[0655] Digital gel imaging system (GelDoc EZ, Bio-Rad, cat. no. 170-8270, and blue sample tray, Bio-Rad, cat. no. 170-8273)

[0656] Blue light transilluminator and orange filter goggles (SafeImager 2.0; Invitrogen, cat. no. G6600)

[0657] Gel quantification software (Bio-Rad, ImageLab, included with GelDoc EZ,or open-source ImageJ from the National Institutes of Health, available at the website rsbweb.nih.gov/ij/) UV spectrophotometer (NanoDrop 2000c, Thermo Scientific)

[0658] **Reagent Setup**

[0659] Tris-borate EDTA (TBE) electrophoresis solution Dilute TBE buffer in distilled water to 1X working solution for casting agarose gels and for use as a buffer for gel electrophoresis. Buffer may be stored at room temperature (18 – 22 °C) for at least 1 year.

[0660] • ATP, 10 mM Divide 10 mM ATP into 50-µl aliquots and store at − 20 °C for up to 1 year; avoid repeated freeze-thaw cycles.

[0661] • DTT, 10 mM Prepare 10 mM DTT solution in distilled water and store in 20-µl aliquots at − 70 °C for up to 2 years; for each reaction, use a new aliquot, as DTT is easily oxidized.

[0662] • D10 culture medium For culture of HEK293FT cells, prepare D10 culture medium by supplementing DMEM with 1X GlutaMAX and 10% (vol/vol) fetal bovine serum. As indicated in the protocol, this medium can also be supplemented with 1X penicillin-streptomycin . D10 medium can be made in advance and stored at 4 °C for up to 1 month.

[0663] • mTeSR1 culture medium For culture of human embryonic stem cells, prepare mTeSR1 medium by supplementing the 5X supplement (included with mTeSR1 basal medium), and 100 ug/ml Normocin.

[0664] **Procedure**

[0665] Design of targeting components and use of the online tool • Timing 1 d

[0666] 1| Input target genomic DNA sequence. Applicants provide an online Cas9 targeting design tool that takes an input sequence of interest, identifies and ranks suitable target sites, and computationally predicts off-target sites for each intended target. Alternatively, one can manually select guide sequence by identifying the 20-bp sequence directly upstream of any 5'-NGG.

[0667] 2| Order necessary oligos and primers as specified by the online tool. If the site is chosen manually, the oligos and primers should be designed.

[0668] Preparation of sgRNA expression construct

[0669] 3| To generate the sgRNA expression construct, either the PCR- or plasmid-based protocol can be used.

[0670] (A) via PCR amplification • Timing 2 h

[0671] (i) Applicants prepare diluted U6 PCR template. Applicants recommend using PX330 as a PCR template, but any U6-containing plasmid may likewise be used as the PCR template. Applicants diluted template with ddH_2O to a concentration of 10 ng/ul. Note that if a plasmid or cassette already containing an U6-driven sgRNA is used as a template, a gel extraction needs to be performed to ensure that the product contains only the intended sgRNA and no trace sgRNA carryover from template.

[0672] (ii) Applicants prepared diluted PCR oligos. U6-Fwd and U6-sgRNA-Rev primers are diluted to a final concentration of 10 uM in ddH_2O (add 10 ul of 100 uM primer to 90 ul ddH_2O).

[0673] (iii) U6-sgRNA PCR reaction. Applicants set up the following reaction for each U6-sgRNA-Rev primer and mastermix as needed:

Component:	Amount (ul)	Final concentration
Herculase II PCR buffer, 5X	10	1X
dNTP, 100mM (25mM each)	0.5	1 mM
U6 template (PX330)	1	0.2 ng/ul
U6-Fwd primer	1	0.2 uM
U6-sgRNA-Rev primer (variable)	1	0.2 uM
Herculase II Fusion polymerase	0.5	
Distilled water	36	
Total	50	

[0674] (iv) Applicants performed PCR reaction on the reactions from step (iii) using the following cycling conditions:

Cycle number	Denature	Anneal	Extend
1	95°C, 2 m		
2-31	95°C, 20 s	60°C, 20 s	72°C, 20 s
32			72°C, 3 m

[0675] (v) After the reaction is completed, Applicants ran the product on a gel to verify successful, single-band amplification. Cast a 2% (wt/vol) agarose gel in 1X TBE buffer with 1X SYBR Safe dye. Run 5 ul of the PCR product in the gel at 15 V cm-1 for 20-30 min. Successful amplicons should yield one single 370-bp product and the template should be invisible. It should not be necessary to gel extract the PCR amplicon.

[0676] (vi) Applicants purified the PCR product using the QIAquick PCR purification kit according to the manufacturer's directions. Elute the DNA in 35 ul of Buffer EB or water. Purified PCR products may be stored at 4°C or -20°C.

[0677] **(B) Cloning sgRNA into Cas9-containing bicistronic expression vector • Timing 3 d**

[0678] (i) *Prepare the sgRNA oligo inserts.* Applicants resuspended the top and bottom strands of oligos for each sgRNA design to a final concentration of 100 uM. Phosphorylate and anneal the oligo as follows:

Oligo 1 (100 uM)	1 ul
Oligo 2 (100 uM)	1 ul
T4 Ligation Buffer, 10X	1 ul
T4 PNK	1 ul
ddH_2O	6 ul
Total	10 ul

[0679] (ii) Anneal in a thermocycler using the following parameters:

[0680] 37°C for 30 m

[0681] 95°C for 5 m

[0682] Ramp down to 25°C at 5°C per m

[0683] (iii) Applicants diluted phosphorylated and annealed oligos 1:200 by add 1ul of oligo to 199 ul room temperature ddH_2O.

[0684] (iv) *Clone sgRNA oligo into PX330.* Applicants set up Golden Gate reaction for each sgRNA. Applicants recommend also setting up a no-insert, PX330 only negative control.

PX330 (100 ng)	x ul
Diluted oligo duplex from step (iii)	2 ul
Tango Buffer, 10X	2 ul
DTT, 10mM	1 ul
ATP, 10mM	1 ul
FastDigest BbsI	1 ul
T7 Ligase	0.5 ul
ddH_2O	x ul
Total	20 ul

[0685] (v) Incubate the Golden Gate reaction for a total of 1 h:

[0686] Cycle number	Condition
[0687] 1-6	37°C for 5 m, 21°C for 5 m

[0688] (vi) Applicants treated Golden Gate reaction with PlasmidSafe exonuclease to digest any residual linearized DNA. This step is optional but highly recommended.

Golden Gate reaction from step 4	11 ul
10X PlasmidSafe Buffer	1.5 ul
ATP, 10 mM	1.5 ul
PlasmidSafe exonuclease	1 ul

[0689] Total 15 ul

[0690] (vii) Applicants incubated the PlasmidSafe reaction at 37°C for 30 min, followed by inactivation at 70°C for 30 min. **Pause point:** after completion, the reaction may be frozen and continued later. The circular DNA should be stable for at least 1 week.

[0691] (viii) *Transformation.* Applicants transformed the PlasmidSafe-treated plasmid into a competent *E. coli* strain, according to the protocol supplied with the cells. Applicants recommend Stbl3 for quick transformation. Briefly, Applicants added 5ul of the product from step (vii) into 20ul of ice-cold chemically competent Stbl3 cells. This is then incubated on ice for 10 m, heat shocked at 42°C for 30 s, returned immediately to ice for 2 m, 100 ul of SOC medium is added, and this is plated onto an LB plate containing 100 ug/ml ampicillin with incubation overnight at 37°C.

[0692] (ix) Day 2: Applicants inspected plates for colony growth. Typically, there are no colonies on the negative control plates (ligation of BbsI-digested PX330 only, no annealed sgRNA oligo), and tens to hundreds of colonies on the PX330-sgRNA cloning plates.

[0693] (x) From each plate, Applicants picked 2-3 colonies to check correct insertion of sgRNA. Applicants used a sterile pipette tip to inoculate a single colony into a 3 ml culture of LB medium with 100 ug/ml ampicillin. Incubate and shake at 37°C overnight.

[0694] (xi) Day 3: Applicants isolated plasmid DNA from overnight cultures using a QiAprep Spin miniprep kit according to the manufacturer's instructions.

[0695] (xii) *Sequence validate CRISPR plasmid.* Applicants verified the sequence of each colony by sequencing from the U6 promoter using the U6-Fwd primer. Optional: sequence the Cas9 gene using primers listed in the following Primer table.

Primer	**Sequence (5' to 3')**	**Purpose**
U6-For	GAGGGCCTATTTCCCATGATTCC	Amplify U6-sgRNA
U6-Rev	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGA TAACGGACTAGCCTTATTTTAACTTGCTATTTCTAG CTCTAAAACNNNNNNNNNNNNNNNNNNNCCGGTGTTTC	Amplify U6-sgRNA; N is reverse

	GTCCTTTCCACAAG	complement of target
sgRNA-top	CACCGNNNNNNNNNNNNNNNNNN	Clone sgRNA into PX330
sgRNA-bottom	AAACNNNNNNNNNNNNNNNNNNC	Clone sgRNA into PX330
U6-*EMX1*-Rev	AAAAAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTCTAAAACCCCTAGTCATTGGAGGTGACCGGTGTTTCGTCCTTTCCACAAG	Amplify U6-EMX1 sgRNA
EMX1-top	CACCGTCACCTCCAATGACTAGGG	Clone EMX1 sgRNA into PX330
EMX1-bottom	AAACCCCTAGTCATTGGAGGTGAC	Clone EMX1 sgRNA into PX330
ssODN-sense	CAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGACATCGATGTCACCTCCAATGACAAGCTTGCTAGCGGTGGGCAACCACAAACCCACGAGGGCAGAGTGCTGCTTGCTGCTGGCCAGGCCCCTGCGTGGGCCCAAGCTGGACTCTGGCCACTCCCT	EMX1 HDR (sense; insertion underlined)
ssODN-antisense	AGGGAGTGGCCAGAGTCCAGCTTGGGCCCACGCAGGGGCCTGGCCAGCAGCAAGCAGCACTCTGCCCTCGTGGGTTTGTGGTTGCCCACCGCTAGCAAGCTTGTCATTGGAGGTGACATCGATGTCCTCCCCATTGGCCTGCTTCGTGGCAATGCGCCACCGGTTGATGTGATGGGAGCCCTTCTTCTTCTG	EMX1 HDR (antisense; insertion underlined)
EMX1-SURV-F	CCATCCCCTTCTGTGAATGT	EMX1 SURVEYOR assay PCR, sequencing
EMX1-SURV-R	GGAGATTGGAGACACGGAGA	EMX1 SURVEYOR assay PCR, sequencing
EMX1-HDR-F	GGCTCCCTGGGTTCAAAGTA	EMX1 RFLP analysis PCR, sequencing
EMX1-HDR-R	AGAGGGGTCTGGATGTCGTAA	EMX1 RFLP analysis PCR, sequencing
pUC19-F	CGCCAGGGTTTTCCCAGTCACGAC	pUC19 multiple cloning site F

		primer, for Sanger sequencing

[0696] Applicants referenced the sequencing results against the PX330 cloning vector sequence to check that the 20 bp guide sequence was inserted between the U6 promoter and the remainder of the sgRNA scaffold. Details and sequence of the PX330 map in GenBank vector map format (*.gb file) can be found at the website crispr.genome-engineering.org.

[0697] **(Optional) Design of ssODN template • Timing 3 d planning ahead**

[0698] 3| *Design and order ssODN.* Either the sense or antisense ssODN can be purchased directly from supplier. Applicants recommend designing homology arms of at least 40 bp on either side and 90 bp for optimal HDR efficiency. In Applicants' experience, antisense oligos have slightly higher modification efficiencies.

[0699] 4| Applicants resuspended and diluted ssODN ultramers to a final concentration of 10 uM. Do not combine or anneal the sense and antisense ssODNs. Store at -20°C.

[0700] 5| Note for HDR applications, Applicants recommend cloning sgRNA into the PX330 plasmid.

[0701] **Functional validation of sgRNAs: cell culture and transfections • Timing 3-4 d**

[0702] The CRISPR-Cas system has been used in a number of mammalian cell lines. Conditions may vary for each cell line. The protocols below details transfection conditions for HEK239FT cells. Note for ssODN-mediated HDR transfections, the Amaxa SF Cell Line Nucleofector Kit is used for optimal delivery of ssODNs. This is described in the next section.

[0703] 7| *HEK293FT maintenance.* Cells are maintained according to the manufacturer's recommendations. Briefly, Applicants cultured cells in D10 medium (GlutaMax DMEM supplemented with 10% Fetal Bovine Serum), at 37°C and 5% CO2.

[0704] 8| To passage, Applicants removed medium and rinsed once by gently adding DPBS to side of vessel, so as not to dislodge cells. Applicants added 2 ml of TrypLE to a T75 flask and incubated for 5 m at 37°C. 10 ml of warm D10 medium is added to inactivate and transferred to a 50 ml Falcon tube. Applicants dissociated cells by triturating gently, and re-seeded new flasks as necessary. Applicants typically passage cells every 2-3 d at a split ratio of 1:4 or 1:8, never allowing cells to reach more than 70% confluency. Cell lines are restarted upon reaching passage number 15.

[0705] 9| *Prepare cells for transfection.* Applicants plated well-dissociated cells onto 24-well plates in D10 medium without antibiotics 16-24 h before transfection at a seeding density of 1.3 x 10^5 cells per well and a seeding volume of 500 ul. Scale up or down according to the manufacturer's manual as needed. It is suggested to not plate more cells than recommended density as doing so may reduce transfection efficiency.

[0706] 10| On the day of transfection, cells are optimal at 70-90% confluency. Cells may be transfected with Lipofectamine 2000 or Amaxa SF Cell Line Nucleofector Kit according to the manufacturers' protocols.

[0707] (A) For sgRNAs cloned into PX330, Applicants transfected 500 ng of sequence-verified CRISPR plasmid; if transfecting more than one plasmid, mix at equimolar ratio and no more than 500 ng total.

[0708] (B) For sgRNA amplified by PCR, Applicants mixed the following:

PX165 (Cas9 only)	200 ng
sgRNA amplicon (each)	40 ng
pUC19	fill up total DNA to 500 ng

[0709] Applicants recommend transfecting in technical triplicates for reliable quantification and including transfection controls (e.g. GFP plasmid) to monitor transfection efficiency. In addition, PX330 cloning plasmid and/or sgRNA amplicon may be transfected alone as a negative control for downstream functional assays.

[0710] 11| Applicants added Lipofectamine complex to cells gently as HEK293FT cells may detach easily from plate easily and result in lower transfection efficiency.

[0711] 12| Applicants checked cells 24 h after transfection for efficiency by estimating the fraction of fluorescent cells in the control (e.g., GFP) transfection using a fluorescence microscope. Typically cells are more than 70% transfected.

[0712] 13| Applicants supplemented the culture medium with an additional 500 ul of warm D10 medium. Add D10 very slowly to the side of the well and do not use cold medium, as cells can detach easily.

[0713] 14| Cells are incubated for a total of 48-72 h post-transfection before harvested for indel analysis. Indel efficiency does not increase noticeably after 48 h.

[0714] (Optional) Co-transfection of CRISPR plasmids and ssODNs or targeting plasmids for HR • Timing 3-4 d

[0715] 15| *Linearize targeting plasmid.* Targeting vector is linearized if possible by cutting once at a restriction site in the vector backbone near one of the homology arms or at the distal end of either homology arm.

[0716] 16| Applicants ran a small amount of the linearized plasmid alongside uncut plasmid on a 0.8-1% agarose gel to check successful linearization. Linearized plasmid should run above the supercoiled plasmid.

[0717] 17| Applicants purified linearized plasmid with the QIAQuick PCR Purification kit.

[0718] 18| Prepare cells for transfection. Applicants cultured HEK293FT in T75 or T225 flasks. Sufficient cell count before day of transfection is planned for. For the Amaxa strip-cuvette format, 2 x 10^6 cells are used per transfection.

[0719] 19| Prepare plates for transfection. Applicants added 1 ml of warm D10 medium into each well of a 12 well plate. Plates are placed into the incubator to keep medium warm.

[0720] 20| Nucleofection. Applicants transfected HEK293FT cells according to the Amaxa SF Cell Line Nucleofector 4D Kit manufacturer's instructions, adapted in the steps below.

[0721] a. For ssODN and CRISPR cotransfection, pre-mix the following DNA in PCR tubes:

pCRISPR plasmid (Cas9 + sgRNA)	500 ng
ssODN template (10uM)	1 ul

[0722] b. For HDR targeting plasmid and CRISPR cotransfection, pre-mix the following DNA in PCR tubes:

CRISPR plasmid (Cas9 + sgRNA)	500 ng
Linearized targeting plasmid	500 ng

[0723] For transfection controls, see previous section. In addition, Applicants recommend transfecting ssODN or targeting plasmid alone as a negative control.

[0724] 21| Dissociate to single cells. Applicants removed medium and rinsed once gently with DPBS, taking care not to dislodge cells. 2 ml of TrypLE is added to a T75 flask and incubated for 5 m at 37°C. 10 ml of warm D10 medium is added to inactivate and triturated gently in a 50 ml Falcon tube. It is recommended that cells are triturated gently and dissociated to single cells. Large clumps will reduce transfection efficiency. Applicants took a 10 ul aliquot from the suspension and diluted into 90 ul of D10 medium for counting. Applicants counted cells and calculated the number of cells and volume of suspension needed for transfection. Applicants typically transfected 2 x 10^5 cells per condition using the Amaxa Nucleocuvette strips, and

recommend calculating for 20% more cells than required to adjust for volume loss in subsequent pipetting steps. The volume needed is transferred into a new Falcon tube.

[0725] 23| Applicants spun down the new tube at 200 x g for 5 m.

[0726] Applicants prepared the transfection solution by mixing the SF solution and S1 supplement as recommended by Amaxa. For Amaxa strip-cuvettes, a total of 20 ul of supplemented SF solution is needed per transfection. Likewise, Applicants recommend calculating for 20% more volume than required.

[0727] 25| Applicants removed medium completely from pelleted cells from step 23 and gently resuspended in appropriate volume (20 ul per 2 x 10^5 cells) of S1-supplemented SF solution. Do not leave cells in SF solution for extended period of time.

[0728] 26| 20 ul of resuspended cells is pipetted into each DNA pre-mix from step 20. Pipette gently to mix and transfer to Nucleocuvette strip chamber. This is repeated for each transfection condition.

[0729] Electroporate cells using the Nucleofector 4D program recommended by Amaxa, CM-130.

[0730] 28| Applicants gently and slowly pipetted 100 ul of warm D10 medium into each Nucleocuvette strip chamber, and transferred all volume into the pre-warmed plate from step 19. CRITICAL. Cells are very fragile at this stage and harsh pipetting can cause cell death. Incubate for 24 h. At this point, transfection efficiency can be estimated from fraction of fluorescent cells in positive transfection control. Nucleofection typically results in greater than 70-80% transfection efficiency. Applicants slowly added 1 ml warm D10 medium to each well without dislodging the cells. Incubate cells for a total of 72 h.

[0731] **Human embryonic stem cell (HUES 9) culture and transfection • Timing 3-4 d**

[0732] Maintaining hESC (HUES9) line. Applicants routinely maintain HUES9 cell line in feeder-free conditions with mTesR1 medium. Applicants prepared mTeSR1 medium by adding the 5X supplement included with basal medium and 100 ug/ml Normocin. Applicants prepared a 10 ml aliquot of mTeSR1 medium supplemented further with 10 uM Rock Inhibitor. Coat tissue culture plate. Dilute cold GelTrex 1:100 in cold DMEM and coat the entire surface of a 100 mm tissue culture plate.

[0733] Place plate in incubator for at least 30 m at 37°C. Thaw out a vial of cells at 37°C in a 15 ml Falcon tube, add 5 ml of mTeSR1 medium, and pellet at 200 x g for 5 m. Aspirate off

GelTrex coating and seed ~1 x 106 cells with 10 ml mTeSR1 medium containing Rock Inhibitor. Change to normal mTeSR1 medium 24 h after transfection and re-feed daily. Passaging cells. Re-feed cells with fresh mTeSR1 medium daily and passage before reaching 70% confluency. Aspirate off mTeSR1 medium and wash cells once with DPBS. Dissociate cells by adding 2 ml Accutase and incubating at 37°C for 3 – 5 m. Add 10 ml mTeSR1 medium to detached cells, transfer to 15 ml Falcon tube and resuspend gently. Re-plate onto GelTrex-coated plates in mTeSR1 medium with 10 uM Rock Inhibitor. Change to normal mTeSR1 medium 24 h after plating.

[0734] Transfection. Applicants recommend culturing cells for at least 1 week post-thaw before transfecting using the Amaxa P3 Primary Cell 4-D Nucleofector Kit (Lonza). Re-feed log-phase growing cells with fresh medium 2 h before transfection. Dissociate to single cells or small clusters of no more than 10 cells with accutase and gentle resuspension. Count the number of cells needed for nucleofection and spin down at 200 x g for 5 m. Remove medium completely and resuspend in recommended volume of S1-supplemented P3 nucleofection solution. Gently plate electroporated cells into coated plates in presence of 1X Rock Inhibitor.

[0735] Check transfection success and re-feed daily with regular mTeSR1 medium beginning 24 h after nucleofection. Typically, Applicants observe greater than 70% transfection efficiency with Amaxa Nucleofection. Harvest DNA. 48-72 h post transfection, dissociate cells using accutase and inactivate by adding 5 x volume of mTeSR1. Spin cells down at 200 x g for 5 m. Pelleted cells can be directed processed for DNA extraction with QuickExtract solution. It is recommended to not mechanically dissociate cells without accutase. It is recommended to not spin cells down without inactivating accutase or above the recommended speed; doing so may cause cells to lyse.

[0736] Isolation of clonal cell lines by FACS. Timing • 2-3 h hands-on; 2-3 weeks expansion

[0737] Clonal isolation may be performed 24 h post-transfection by FACS or by serial dilution.

[0738] 54| *Prepare FACS buffer.* Cells that do not need sorting using colored fluorescence may be sorted in regular D10 medium supplemented with 1X penicillin/streptinomycin. If colored fluorescence sorting is also required, a phenol-free DMEM or DPBS is substituted for

normal DMEM. Supplement with 1X penicillin/streptinomycin and filter through a .22 um Steriflip filter.

[0739] 55| *Prepare 96 well plates.* Applicants added 100 ul of D10 media supplemented with 1X penicillin/streptinomycin per well and prepared the number of plates as needed for the desired number of clones.

[0740] 56| *Prepare cells for FACS.* Applicants dissociated cells by aspirating the medium completely and adding 100 ul TrypLE per well of a 24-well plate. Incubate for 5 m and add 400 ul warm D10 media.

[0741] 57| Resuspended cells are transferred into a 15 ml Falcon tube and gently triturated 20 times. Recommended to check under the microscope to ensure dissociation to single cells.

[0742] 58| Spin down cells at 200 x g for 5 minutes.

[0743] 59| Applicants aspirated the media, and resuspended the cells in 200 ul of FACS media.

[0744] 60| Cells are filtered through a 35 um mesh filter into labeled FACS tubes. Applicants recommend using the BD Falcon 12 x 75 mm Tube with Cell Strainer cap. Place cells on ice until sorting.

[0745] 61| Applicants sorted single cells into 96-well plates prepared from step 55. Applicants recommend that in one single designated well on each plate, sort 100 cells as a positive control.

[0746] NOTE. The remainder of the cells may be kept and used for genotyping at the population level to gauge overall modification efficiency.

[0747] 62| Applicants returned cells into the incubator and allowed them to expand for 2-3 weeks. 100 ul of warm D10 medium is added 5 d post sorting. Change 100 ul of medium every 3-5 d as necessary.

[0748] 63| Colonies are inspected for "clonal" appearance 1 week post sorting: rounded colonies radiating from a central point. Mark off wells that are empty or may have been seeded with doublets or multiplets.

[0749] 64| When cells are more than 60% confluent, Applicants prepared a set of replica plates for passaging. 100 ul of D10 medium is added to each well in the replica plates. Applicants dissociated cells directly by pipetting up and down vigorously 20 times. 20% of the

resuspended volume was plated into the prepared replica plates to keep the clonal lines. Change the medium every 2-3 d thereafter and passage accordingly.

[0750] 65| Use the remainder 80% of cells for DNA isolation and genotyping.

[0751] Optional: Isolation of clonal cell lines by dilution. Timing • 2-3 h hands-on; 2-3 weeks expansion

[0752] 66| Applicants dissociated cells from 24-well plates as described above. Make sure to dissociate to single cells. A cell strainer can be used to prevent clumping of cells.

[0753] 67| The number of cells are counted in each condition. Serially dilute each condition in D10 medium to a final concentration of 0.5 cells per 100 ul. For each 96 well plate, Applicants recommend diluting to a final count of 60 cells in 12 ml of D10. Accurate count of cell number is recommended for appropriate clonal dilution. Cells may be recounted at an intermediate serial dilution stage to ensure accuracy.

[0754] 68| Multichannel pipette was used to pipette 100 ul of diluted cells to each well of a 96 well plate.

[0755] NOTE. The remainder of the cells may be kept and used for genotyping at the population level to gauge overall modification efficiency.

[0756] 69| Applicants inspected colonies for "clonal" appearance ~1 week post plating: rounded colonies radiating from a central point. Mark off wells that may have seeded with doublets or multiplets.

[0757] 70| Applicants returned cells to the incubator and allowed them to expand for 2-3 weeks. Re-feed cells as needed as detailed in previous section.

[0758] SURVEYOR assay for CRISPR cleavage efficiency. Timing • 5-6 h

[0759] Before assaying cleavage efficiency of transfected cells, Applicants recommend testing each new SURVEYOR primer on negative (untransfected) control samples through the step of SURVEYOR nuclease digestion using the protocol described below. Occasionally, even single-band clean SURVEYOR PCR products can yield non-specific SURVEYOR nuclease cleavage bands and potentially interfere with accurate indel analysis.

[0760] 71| *Harvest cells for DNA.* Dissociate cells and spin down at 200 x g for 5 m. NOTE. Replica plate at this stage as needed to keep transfected cell lines.

[0761] 72| Aspirate the supernatant completely.

[0762] 73| Applicants used QuickExtract DNA extraction solution according to the manufacturer's instructions. Applicants typically used 50 ul of the solution for each well of a 24 well plate and 10 ul for a 96 well plate.

[0763] 74| Applicants normalized extracted DNA to a final concentration of 100-200 ng/ul with ddH2O. **Pause point:** Extracted DNA may be stored at -20°C for several months.

[0764] 75| *Set up the SURVEYOR PCR.* Master mix the following using SURVEYOR primers provided by Applicants online/computer algorithm tool:

Component:	Amount (ul)	Final concentration
Herculase II PCR buffer, 5X	10	1X
dNTP, 100mM (25mM each)	1	1 mM
SURVEYOR Fwd primer (10uM)	1	0.2 uM
SURVEYOR Rev primer (10uM)	1	0.2 uM
Herculase II Fusion polymerase	1	
$MgCl_2$ (25mM)	2	1 mM
Distilled water	33	
Total	49 (for each reaction)	

[0765] 76| Applicants added 100-200 ng of normalized genomic DNA template from step 74 for each reaction.

[0766] 77| PCR reaction was performed using the following cycling conditions, for no more than 30 amplification cycles:

Cycle number	Denature	Anneal	Extend
1	95°C, 2 min		
2-31	95°C, 20 s	60°C, 20 s	72°C, 30 s
32			72°C, 3 min

[0767] 78| Applicants ran 2-5 ul of PCR product on a 1% gel to check for single-band product. Although these PCR conditions are designed to work with most pairs of SURVEYOR primers, some primers may need additional optimization by adjusting the template concentration, $MgCl_2$ concentration, and/or the annealing temperature.

[0768] 79| Applicants purified the PCR reactions using the QIAQuick PCR purification kit and normalized eluant to 20 ng/ul. **Pause point:** Purified PCR product may be stored at -20°C.

[0769] 80| *DNA heteroduplex formation.* The annealing reaction was set up as follows:

[0770] Taq PCR buffer, 10X 2 ul

[0771] Normalized DNA (20 ng/ul) 18 ul

[0772] Total volume 20 ul

[0773] 81| Anneal the reaction using the following conditions:

Cycle number	Condition
1	95°C, 10 mn
2	95°C-85°C, -2°C/s
3	85°C, 1 min
4	85°C-75°C, -0.3°C/s
5	75°C, 1 min
6	75°C-65°C, -0.3°C/s
7	65°C, 1 min
8	65°C-55°C, -0.3°C/s
9	55°C, 1 min
10	55°C-45°C, -0.3°C/s
11	45°C, 1 min
12	45°C-35°C, -0.3°C/s
13	35°C, 1 min
14	35°C-25°C, -0.3°C/s
15	25°C, 1 min

[0774] 82| *SURVEYOR nuclease S digestion.* Applicants prepared master-mix and added the following components on ice to annealed heteroduplexes from step 81 for a total final volume of 25 ul:

Component	Amount (ul)	Final Concentration
$MgCl_2$ solution, 0.15M	2.5	15mM
ddH_2O	0.5	
SURVEYOR nuclease S	1	1X
SURVEYOR enhancer S	1	1X
Total	5	

[0775] 83| Vortex well and spin down. Incubate the reaction at 42°C for 1 h.

[0776] 84| Optional: 2 ul of the Stop Solution from the SURVEYOR kit may be added. **Pause point.** The digested product may be stored at -20°C for analysis at a later time.

[0777] 85| *Visualize the SURVEYOR reaction.* SURVEYOR nuclease digestion products may be visualized on a 2% agarose gel. For better resolution, products may be run on a 4-20% gradient Polyacrylamide TBE gel. Applicants loaded 10 ul of product with the recommended loading buffer and ran the gel according to manufacturer's instructions. Typically, Applicants run until the bromophenol blue dye has migrated to the bottom of the gel. Include DNA ladder and negative controls on the same gel.

[0778] 86| Applicants stained the gel with 1X SYBR Gold dye diluted in TBE. The gel was gently rocked for 15 m.

[0779] 87| Applicants imaged the gel using a quantitative imaging system without overexposing the bands. The negative controls should have only one band corresponding to the size of the PCR product, but may have occasionally non-specific cleavage bands of other sizes. These will not interfere with analysis if they are different in size from target cleavage bands. The sum of target cleavage band sizes, provided by Applicants online/computer algorithm tool, should be equal to the size of the PCR product.

[0780] 88| *Estimate the cleavage intensity.* Applicants quantified the integrated intensity of each band using ImageJ or other gel quantification software.

[0781] 89| For each lane, Applicants calculated the fraction of the PCR product cleaved (f_{cut}) using the following formula: $f_{cut} = (b + c) / (a + b + c)$, where a is the integrated intensity of the undigested PCR product and b and c are the integrated intensities of each cleavage product. 90| Cleavage efficiency may be estimated using the following formula, based on the binomial probability distribution of duplex formation:

[0782] 91| $indel\ (\%) = 100 \times (1 - \sqrt{(1 - f_{cut})})$

[0783] **Sanger sequencing for assessing CRISPR cleavage efficiency. Timing • 3 d**

[0784] Initial steps are identical to Steps 71-79 of the SURVEYOR assay. Note: SURVEYOR primers may be used for Sanger sequencing if appropriate restriction sites are appended to the Forward and Reverse primers. For cloning into the recommended pUC19 backbone, EcoRI may be used for the Fwd primer and HindIII for the Rev primer.

[0785] 92| *Amplicon digestion.* Set up the digestion reaction as follows:

Component	Amount (ul)

Fast Digest buffer, 10X	3
FastDigest EcoRI	1
FastDigest HindIII	1
Normalized DNA (20 ng/ul)	10
ddH_2O	15
Total volume	30

[0786] 93| pUC19 backbone digestion. Set up the digestion reaction as follows:

[0787]

Component	Amount (ul)
Fast Digest buffer, 10X	3
FastDigest EcoRI	1
FastDigest HindIII	1
FastAP Alkaline Phosphatase	1
pUC19 vector (200 ng/ul)	5
ddH_2O	20
Total volume	30

[0788] 94| Applicants purified the digestion reactions using the QIAQuick PCR purification kit. **Pause point:** Purified PCR product may be stored at -20°C.

[0789] 95| Applicants ligated the digested pUC19 backbone and Sanger amplicons at a 1:3 vector:insert ratio as follows:

[0790]

Component	Amount (ul)
Digested pUC19	x (50 ng)
Digested insert	x (1:3 vector:insert molar ratio)
T7 ligase	1
2X Rapid Ligation Buffer	10
ddH_2O	x
Total volume	20

[0791] 96| *Transformation.* Applicants transformed the PlasmidSafe-treated plasmid into a competent *E. coli* strain, according to the protocol supplied with the cells. Applicants recommend Stbl3 for quick transformation. Briefly, 5ul of the product from step 95 is added into 20ul of ice-cold chemically competent Stbl3 cells, incubated on ice for 10 m, heat shocked at

42°C for 30 s, returned immediately to ice for 2 m, 100 ul of SOC medium is added, and plated onto an LB plate containing 100 ug/ml ampicillin. This is incubated overnight at 37°C.

[0792] 97| Day 2: Applicants inspected plates for colony growth. Typically, there are no colonies on the negative control plates (ligation of EcoRI-HindIII digested pUC19 only, no Sanger amplicon insert), and tens to hundreds of colonies on the pUC19-Sanger amplicon cloning plates.

[0793] 98| Day 3: Applicants isolated plasmid DNA from overnight cultures using a QIAprep Spin miniprep kit according to the manufacturer's instructions.

[0794] 99| *Sanger sequencing.* Applicants verified the sequence of each colony by sequencing from the pUC19 backbone using the pUC19-For primer. Applicants referenced the sequencing results against the expected genomic DNA sequence to check for the presence of Cas9-induced NHEJ mutations. % editing efficiency = (# modified clones)/(# total clones). It is important to pick a reasonable number of clones (>24) to generate accurate modification efficiencies.

[0795] **Genotyping for microdeletion. Timing • 2-3 d hands on; 2-3 weeks expansion**

[0796] 100| Cells were transfected as described above with a pair of sgRNAs targeting the region to be deleted.

[0797] 101| 24 h post-transfection, clonal lines are isolated by FACS or serial dilution as described above.

[0798] 102| Cells are expanded for 2-3 weeks.

[0799] 103| Applicants harvested DNA from clonal lines as described above using 10 ul QuickExtract solution and normalized genomic DNA with ddH_2O to a final concentration of 50-100 ng/ul.

[0800] 104| *PCR Amplify the modified region.* The PCR reaction is set up as follows:

Component:	Amount (ul)	Final concentration
Herculase II PCR buffer, 5X	10	1X
dNTP, 100mM (25mM each)	1	1 mM
Out Fwd primer (10uM)	1	0.2 uM
Out Rev primer (10uM)	1	0.2 uM
Herculase II Fusion polymerase	1	
MgCl2 (25mM)	2	1 mM

ddH_2O	32
Total	48 (for each reaction)

[0801] Note: if deletion size is more than 1 kb, set up a parallel set of PCR reactions with In-Fwd and In-Rev primers to screen for the presence of the wt allele.

[0802] 105| To screen for inversions, a PCR reaction is set up as follows:

Component:	Amount (ul)	Final concentration
Herculase II PCR buffer, 5X	10	1X
dNTP, 100mM (25mM each)	1	1 mM
Out Fwd *or* Out-Rev primer (10uM)	1	0.2 uM
In Fwd *or* In-Rev primer (10uM)	1	0.2 uM
Herculase II Fusion polymerase	1	
$MgCl_2$ (25mM)	2	1 mM
ddH_2O	32	
Total	48 (for each reaction)	

[0803] Note: primers are paired either as Out-Fwd + In Fwd, or Out-Rev + In-Rev.

[0804] 106| Applicants added 100-200 ng of normalized genomic DNA template from step 103 for each reaction.

[0805] 107| PCR reaction was performed using the following cycling conditions:

Cycle number	Denature	Anneal	Extend
1	95°C, 2 min		
2-31	95°C, 20 s	60°C, 20 s	72°C, 30 s
32			72°C, 3 m

[0806] 108| Applicants run 2-5 ul of PCR product on a 1-2% gel to check for product. Although these PCR conditions are designed to work with most primers, some primers may need additional optimization by adjusting the template concentration, $MgCl_2$ concentration, and/or the annealing temperature.

[0807] **Genotyping for targeted modifications via HDR. Timing • 2-3 d, 2-3 h hands on**

[0808] 109| Applicants harvested DNA as described above using QuickExtract solution and normalized genomic DNA with TE to a final concentration of 100-200 ng/ul.

[0809] 110| *PCR Amplify the modified region.* The PCR reaction is set up as follows:

Component:	Amount (ul)	Final concentration

Herculase II PCR buffer, 5X	10	1X
dNTP, 100 mM (25 mM each)	1	1 mM
HDR Fwd primer (10 uM)	1	0.2 uM
HDR Rev primer (10 uM)	1	0.2 uM
Herculase II Fusion polymerase	1	
$MgCl_2$ (25mM)	2	1 mM
ddH_2O	33	
Total	49 (for each reaction)	

[0810] 111| Applicants added 100-200 ng of genomic DNA template from step 109 for each reaction and run the following program.

Cycle number	Denature	Anneal	Extend
1	95°C, 2 min		
2-31	95°C, 20 s	60°C, 20 s	72°C, *30-60 s per kb*
32			72°C, 3 min

[0811] 112| Applicants ran 5 ul of PCR product on a 0.8-1% gel to check for single-band product. Primers may need additional optimization by adjusting the template concentration, $MgCl_2$ concentration, and/or the annealing temperature.

[0812] 113| Applicants purified the PCR reactions using the QIAQuick PCR purification kit.

[0813] 114| In the HDR example, a *Hind*III restriction site is inserted into the EMX1 gene. These are detected by a restriction digest of the PCR amplicon:

Component	Amount (ul)
Purified PCR amplicon (200-300ng)	x
F.D. buffer, Green	1
HindIII	0.5
ddH2O	x
Total	10

[0814] i. The DNA is digested for 10 m at 37°C:

[0815] ii. Applicants ran 10 ul of the digested product with loading dye on a 4-20% gradient polyacrylamide TBE gel until the xylene cyanol band had migrated to the bottom of the gel.

[0816] iii. Applicants stained the gel with 1X SYBR Gold dye while rocking for 15 m.

[0817] iv. The cleavage products are imaged and quantified as described above in the SURVEYOR assay section. HDR efficiency is estimated by the formula: (b + c)/(a + b + c), where a is the integrated intensity for the undigested HDR PCR product, and b and c are the integrated intensities for the HindIII-cut fragments.

[0818] 115| Alternatively, purified PCR amplicons from step 113 may be cloned and genotyped using Sanger sequencing or NGS.

[0819] **Deep sequencing and off-target analysis • Timing 1 - 2 d**

[0820] The online CRISPR target design tool generates candidate genomic off-target sites for each identified target site. Off-target analysis at these sites can be performed by SURVEYOR nuclease assay, Sanger sequencing, or next-generation deep sequencing. Given the likelihood of low or undetectable modification rates at many of these sites, Applicants recommend deep sequencing with the Illumina Miseq platform for high sensitivity and accuracy. Protocols will vary with sequencing platform; here, Applicants briefly describe a fusion PCR method for attaching sequencing adapters.

[0821] 116| *Design deep sequencing primers.* Next-generation sequencing (NGS) primers are designed for shorter amplicons, typically in the 100-200 bp size range. Primers may be manually designed using NCBI Primer-Blast or generated with online CRISPR target design tools (website at genome-engineering.org/tools).

[0822] 117| Harvest genomic DNA from Cas9-targeted cells. Normalize QuickExtract genomic DNA to 100-200 ng/ul with ddH2O.

[0823] 118| *Initial library preparation PCR.* Using the NGS primers from step 116, prepare the initial library preparation PCR

[0824]

Component:	Amount (ul)	Final concentration
Herculase II PCR buffer, 5X	10	1X
dNTP, 100mM (25mM each)	1	1 mM
NGS Fwd primer (10uM)	1	0.2 uM
NGS Rev primer (10uM)	1	0.2 uM
Herculase II Fusion polymerase	1	
MgCl2 (25mM)	2	1 mM
ddH2O	33	

Total	49 (for each reaction)

[0825] 119| Add 100-200 ng of normalized genomic DNA template for each reaction.

[0826] 120| Perform PCR reaction using the following cycling conditions, for no more than **20** amplification cycles:

Cycle number	Denature	Anneal	Extend
1	95°C, 2 min		
2-21	95°C, 20 s	60°C, 20 s	72°C, 15 s
22			72°C, 3 min

[0827] 121| Run 2-5 ul of PCR product on a 1% gel to check for single-band product. As with all genomic DNA PCRs, NGS primers may require additional optimization by adjusting the template concentration, $MgCl_2$ concentration, and/or the annealing temperature.

[0828] 122| Purify the PCR reactions using the QIAQuick PCR purification kit and normalize eluant to 20 ng/ul. **Pause point:** Purified PCR product may be stored at -20°C.

[0829] 123| *Nextera XT DNA Sample Preparation Kit.* Following the manufacturer's protocol, generate Miseq sequencing-ready libraries with unique barcodes for each sample.

[0830] 124| *Analyze sequencing data*. Off-target analysis may be performed through read alignment programs such as ClustalW, Geneious, or simple sequence analysis scripts.

[0831] **Timing**

[0832] Steps 1 – 2 Design and synthesis of sgRNA oligos and ssODNs: 1-5 d, variable depending on supplier

[0833] Steps 3 – 5 Construction of CRISPR plasmid or PCR expression cassette: 2 h to 3 d

[0834] Steps 6 – 53 Transfection into cell lines: 3 d (1 h hands-on time)

[0835] Steps 54 – 70 Optional derivation of clonal lines: 1-3 weeks, variable depending on cell type

[0836] Steps 71 – 91 Functional validation of NHEJ via SURVEYOR: 5-6 h

[0837] Steps 92 – 124 Genotyping via Sanger or next-gen deep sequencing: 2-3 d (3-4 h hands on time)

[0838] **Addressing Situations Concerning Herein Examples**

Situation	Solution
No amplification of sgRNA	Titrate U6-template concentration
SURVEYOR or HDR PCR dirty or no amplification	Titrate MgCl2; normalize and titrate template concentration; annealing temp gradient; redesign primers

Unequal amplification of alleles in microdeletion PCRs	Set up separate PCRs to detect wildtype and deletion alleles; Redesign primers with similar sized amplicons
Colonies on negative control plate	Increase *Bbs*I; increase Golden Gate reaction cycle number, cut PX330 separately with Antarctic Phosphate treatment
No sgRNA sequences or wrong sequences	Screen additional colonies
Low lipofectamine transfection efficiency	Check cell health and density; titrate DNA; add GFP transfection control
Low nucleofection transfection efficiency	Check cell health and density; titrate DNA; suspend to single cell
Clumps or no cells after FACS	Filter cells before FACS; dissociate to single cells; resuspend in appropriate density
Clumps or no cells in serial dilution	Recount cells; dissociate to single cells and filter through strainer; check serial dilution
High SURVEYOR background on negative sample	Redesign primers to prime from different locations
Dirty SURVEYOR result on gel	Purify PCR product; reduce input DNA; reduce 42°C incubation to 30 m
No SURVEYOR cleavage	Purify and normalize PCR product; re-anneal with TaqB buffer; Redesign sgRNAs; sequence verify Cas9 on px330 backbone
Samples do not sink in TBE acrylamide gel	Supplement with MgCl2 to a final concentration of 15mM or add loading buffer containing glycerol

[0839] Discussion

[0840] CRISPR-Cas may be easily multiplexed to facilitate simultaneous modification of several genes and mediate chromosomal microdeletions at high efficiencies. Applicants used two sgRNAs to demonstrate simultaneous targeting of the human *GRIN2B* and *DYRK1A* loci at efficiencies of up to 68% in HEK293FT cells. Likewise, a pair of sgRNAs may be used to mediate microdeletions, such as excision of an exon, which can be genotyped by PCR on a clonal level. Note that the precise location of exon junctions can vary. Applicants also demonstrated the use of ssODNs and targeting vector to mediate HDR with both wildtype and nickase mutant of Cas9 in HEK 293FT and HUES9 cells (Figure 12). Note that Applicants have not been able to detect HDR in HUES9 cells using the Cas9 nickase, which may be due to low efficiency or a potential difference in repair activities in HUES9 cells. Although these values are typical, there is some variability in the cleavage efficiency of a given sgRNA, and on rare occasions certain sgRNAs may not work for reasons yet unknown. Applicants recommend designing two sgRNAs for each locus, and testing their efficiencies in the intended cell type.

Example 26: NLSs

[0841] Cas9 Transcriptional Modulator: Applicants set out to turn the Cas9/gRNA CRISPR system into a generalized DNA binding system in which functions beyond DNA cleavage can be executed. For instance, by fusing functional domain(s) onto a catalytically inactive Cas9 Applicants have imparted novel functions, such as transcriptional activation/repression, methylation/demethylation, or chromatin modifications. To accomplish this goal Applicants made a catalytically inactive Cas9 mutant by changing two residues essential for nuclease activity, D10 and H840, to alanine. By mutating these two residues the nuclease activity of Cas9 is abolished while maintaining the ability to bind target DNA. The functional domains Applicants decided to focus on to test Applicants' hypothesis are the transcriptional activator VP64 and the transcriptional repressors SID and KRAB.

[0842] Cas9 Nuclear localization: Applicants hypothesized that the most effective Cas9 transcriptional modulator would be strongly localized to the nucleus where it would have its greatest influence on transcription. Moreover, any residual Cas9 in the cytoplasm could have unwanted effects. Applicants determined that wild-type Cas9 does not localize into the nucleus without including multiple nuclear localization signals (NLSs) (although a CRISPR system need not have one or more NLSs but advantageously has at least one or more NLS(s)). Because multiple NLS sequences were required it was reasoned that it is difficult to get Cas9 into the nucleus and any additional domain that is fused to Cas9 could disrupt the nuclear localization. Therefore, Applicants made four Cas9-VP64-GFP fusion constructs with different NLS sequences (pXRP02- pLenti2-EF1a-NLS-hSpCsn1(10A,840A)-NLS-VP64-EGFP, pXRP04-pLenti2-EF1a-NLS-hSpCsn1(10A,840A)-NLS-VP64-2A-EGFP-NLS, pXRP06- pLenti2-EF1a-NLS-EGFP-VP64-NLS-hSpCsn1(10A,840A)-NLS, pXRP08- pLenti2-EF1a-NLS-VP64-NLS-hSpCsn1(10A,840A)-NLS-VP64-EGFP-NLS). These constructs were cloned into a lenti backbone under the expression of the human EF1a promoter. The WPRE element was also added for more robust protein expression. Each construct was transfected into HEK 293FT cells using Lipofectame 2000 and imaged 24 hours post-transfection. The best nuclear localization is obtained when the fusion proteins have NLS sequences on both the N- and C-term of the fusion protein. The highest observed nuclear localization occurred in the construct with four NLS elements.

[0843] To more robustly understand the influence of NLS elements on Cas9 Applicants made 16 Cas9-GFP fusions by adding the same alpha importin NLS sequence on either the N- or C-term looking at zero to three tandem repeats. Each construct was transfected into HEK 293FT cells using Lipofectame 2000 and imaged 24 hours post-transfection. Notably, the number of NLS elements does not directly correlate with the extent of nuclear localization. Adding an NLS on the C-term has a greater influence on nuclear localization than adding on the N-term.

[0844] Cas9 Transcriptional Activator: Applicants functionally tested the Cas9-VP64 protein by targeting the Sox2 locus and quantifying transcriptional activation by RT-qPCR. Eight DNA target sites were chosen to span the promoter of Sox2. Each construct was transfected into HEK 293FT cells using Lipofectame 2000 and 72 hours post-transfection total RNA was extracted from the cells. 1 ug of RNA was reverse transcribed into cDNA (qScript Supermix) in a 40 ul reaction. 2 ul of reaction product was added into a single 20 ul TaqMan assay qPCR reaction. Each experiment was performed in biological and technical triplicates. No RT control and no template control reactions showed no amplification. Constructs that do not show strong nuclear localization, pXRP02 and pXRP04, result in no activation. For the construct that did show strong nuclear localization, pXRP08, moderate activation was observed. Statistically significant activation was observed in the case of guide RNAs Sox2.4 and Sox2.5.

Example 27: In Vivo Mouse Data

[0845] Material and reagents

Herculase II fusion polymerase (Agilent Technologies, cat. no. 600679)

10x NEBuffer 4 (NEB, cat. No. B7004S)

BsaI HF (NEB, cat. No. R3535S)

T7 DNA ligase (Enzymatics, cat. no. L602L)

Fast Digest buffer, 10X (ThermoScientific, cat. No. B64)

FastDigest NotI (ThermoScientific, cat. No. FD0594)

FastAP Alkaline Phosphatase (ThermoScientific, cat. No. EF0651)

Lipofectamine2000 (Life Technologies, cat. No. 11668-019)

Trypsin (Life Technologies, cat. No. 15400054)

Forceps #4 (Sigma, cat. No. Z168777-1EA)

Forceps #5 (Sigma, cat. No. F6521-1EA)

10x Hank's Balanced Salt Solution (Sigma, cat. No. H4641-500ML)

Penicillin/Streptomycin solution (Life Technologies, cat. No. P4333)

Neurobasal (Life Technologies, cat. No. 21103049)

B27 Supplement (Life Technologies, cat. No. 17504044)

L-glutamine (Life Technologies, cat. No. 25030081)

Glutamate (Sigma, cat. No. RES5063G-A7)

β-mercaptoethanol (Sigma, cat. No. M6250-100ML)

HA rabbit antibody (Cell Signaling, cat. No. 3724S)

LIVE/DEAD® Cell Imaging Kit (Life Technologies, cat. No. R37601)

30G World Precision Instrument syringe (World Precision Instruments, cat. No. NANOFIL)

Stereotaxic apparatus (Kopf Instruments)

UltraMicroPump3 (World Precision Instruments, cat. No. UMP3-4)

Sucrose (Sigma, cat. No. S7903)

Calcium chloride (Sigma, cat. No. C1016)

Magnesium acetate (Sigma, cat. No. M0631)

Tris-HCl (Sigma, cat. no T5941)

EDTA (Sigma, cat. No. E6758)

NP-40 (Sigma, cat. No. NP40)

Phenylmethanesulfonyl fluoride (Sigma, cat. No. 78830)

Magnesium chloride (Sigma, cat. No. M8266)

Potassium chloride (Sigma, cat. No. P9333)

β -glycerophosphate (Sigma, cat. No. G9422)

Glycerol (Sigma, cat. No. G9012)

Vybrant® DyeCycle™ Ruby Stain (Life technologies, cat. No. S4942)

FACS Aria Flu-act-cell sorter (Koch Institute of MIT, Cambridge US)

DNAeasy Blood & Tissue Kit (Qiagen, cat. No. 69504)

[0846] Procedure

[0847] ***Constructing gRNA multiplexes for using in vivo in the brain***

Applicants designed and PCR amplified single gRNAs targeting mouse TET and DNMT family members (as described herein) Targeting efficiency was assessed in N2a cell line (Figure 15). To obtain simultaneous modification of several genes in vivo, efficient gRNA was multiplexed in

AAV-packaging vector (Figure 16). To facilitate further analysis of system efficiency applicants added to the system expression cassette consistent of GFP-KASH domain fusion protein under control of human Synapsin I promoter (Figure 16). This modification allows for further analysis of system efficiency in neuronal population (more detail procedure in section ***Sorting nuclei and in vivo results***).

[0848] All 4 parts of the system were PCR amplified using Herculase II Fusion polymerase using following primers:

1st U6 Fw:

gagggtctcgtccttgcggccgcgctagcgagggcctatttcccatgattc

1st gRNA Rv:

ctcggtctcggtAAAAAAgcaccgactcggtgccacttttcaagttgataacggactagc

cttattttaacttgctaTTTCtagctctaaaacNNNNNNNNNNNNNNNNNNNNGGTGTTTC

GTCCTTTCCAC

2nd U6 Fw:

gagggtctcTTTaccggtgagggcctatttcccatgattcc

2nd gRNA Rv:

ctcggtctcctcAAAAAAgcaccgactcggtgccacttttcaagttgataacggacta

gc cttattttaacttgctaTTTCtagctctaaaacNNNNNNNNNNNNNNNNNNNNGGTGTTTC

GTCCTTTCCAC

3rd U6 Fw:

gagggtctcTTTgagctcgagggcctatttcccatgattc

3rd gRNA Rv:

ctcggtctcgcgtAAAAAAgcaccgactcggtgccacttttcaagttgataacggact

ag ccttattttaacttgctaTTTCtagctctaaaacNNNNNNNNNNNNNNNNNNNNGGTGTTT

CGTCCTTTCCA

hSyn_GFP-kash Fw: gagggtctcTTacgcgtgtgtctagac

hSyn_GFP-kash Rv:

ctcggtctcAaggaCAGGGAAGGGAGCAGTGGTTCACGCCTGTAATCCCAGCAATTTGG

GA GGCCAAGGTGGGTAGATCACCTGAGATTAGGAGTTGC

(NNNNNNNNNNNNNNNNNNN is a reverse compliment targeted genomic sequence)

[0849] Applicants used Golden Gate strategy to assemble all parts (1:1 molecular ratio) of the system in a single step reaction:

1st U6_gRNA	18 ng
2nd U6_gRNA	18 ng
3rd U6_gRNA	18 ng
Syn_GFP-kash	100 ng
10x NEBuffer 4	1.0 µl
10x BSA	1.0 µl
10 mM ATP	1.0 µl
BsaI HF	0.75 µl
T7 ligase	0.25 µl
ddH_2O	10 µl

Cycle number	Condition
1-50	37°C for 5 m, 21°C for 5 m

[0850] Golden Gate reaction product was PCR amplified using Herculase II fusion polymerase and following primers:

Fw 5' cctgtccttgcggccgcgctagcgagggcc

Rv 5' cacgcggccgcaaggacagggaagggagcag

[0851] PCR product was cloned into AAV backbone, between ITR sequences using NotI restriction sites:

PCR product digestion:

Fast Digest buffer, 10X	3 µl
FastDigest NotI	1 µl
DNA	1 µg
ddH_2O	up to 30 µl

[0852] AAV backbone digestion:

Fast Digest buffer, 10X	3 µl

FastDigest NotI	1 µl
FastAP Alkaline Phosphatase	1 µl
AAV backbone	1 µg
ddH_2O	up to 30 µl

[0853] After 20 min incubation in 37°C samples were purified using QIAQuick PCR purification kit. Standardized samples were ligated at a 1:3 vector:insert ratio as follows:

Digested pUC19	50 ng
Digested insert	1:3 vector:insert molar ratio
T7 ligase	1 µl
2X Rapid Ligation Buffer	5 µl
ddH_2O	up to 10 µl

[0854] After transformation of bacteria with ligation reaction product, applicants confirmed obtained clones with Sanger sequencing.

[0855] Positive DNA clones were tested in N2a cells after co-transfection with Cas9 construct (Figures 17 and 18).

[0856] ***Design of new Cas9 constructs for AAV delivery***

[0857] AAV delivery system despite its unique features has packing limitation – to successfully deliver expressing cassette in vivo it has to be in size < then 4.7 kb. To decrease the size of SpCas9 expressing cassette and facilitate delivery applicants tested several alteration: different promoters, shorter polyA signal and finally a smaller version of Cas9 from *Staphylococcus aureus* (SaCas9) (Figures 19 and 20). All tested promoters were previously tested and published to be active in neurons, including mouse *Mecp2* (Gray et al., 2011), rat *Map1b* and truncated rat *Map1b* (Liu and Fischer, 1996). Alternative synthetic polyA sequence was previously shown to be functional as well (Levitt et al., 1989; Gray et al., 2011). All cloned constructs were expressed in N2a cells after transfection with Lipofectamine 2000, and tested with Western blotting method (Figure 21).

[0858] ***Testing AAV multiplex system in primary neurons***

[0859] To confirm functionality of developed system in neurons, Applicants use primary neuronal cultures *in vitro*. Mouse cortical neurons was prepared according to the protocol published previously by Banker and Goslin (Banker and Goslin, 1988).

[0860] Neuronal cells are obtained from embryonic day 16. Embryos are extracted from the euthanized pregnant female and decapitated, and the heads are placed in ice-cold HBSS. The brains are then extracted from the skulls with forceps (#4 and #5) and transferred to another change of ice-cold HBSS. Further steps are performed with the aid of a stereoscopic microscope in a Petri dish filled with ice-cold HBSS and #5 forceps. The hemispheres are separated from each other and the brainstem and cleared of meninges. The hippocampi are then very carefully dissected and placed in a 15 ml conical tube filled with ice-cold HBSS. Cortices that remain after hippocampal dissection can be used for further cell isolation using an analogous protocol after removing the brain steam residuals and olfactory bulbs. Isolated hippocampi are washed three times with 10 ml ice-cold HBSS and dissociated by 15 min incubation with trypsin in HBSS (4 ml HBSS with the addition of 10 µl 2.5% trypsin per hippocampus) at 37°C. After trypsinization, the hippocampi are very carefully washed three times to remove any traces of trypsin with HBSS preheated to 37°C and dissociated in warm HBSS. Applicants usually dissociate cells obtained from 10-12 embryos in 1 ml HBSS using 1 ml pipette tips and dilute dissociated cells up to 4 ml. Cells are plated at a density of 250 cells/mm2 and cultured at 37°C and 5% CO2 for up to 3 week

[0861] **HBSS**

435 ml H2O

50 ml 10x Hank's Balanced Salt Solution

16.5 ml 0.3M HEPES pH 7.3

5 ml penicillin-streptomycin solution

Filter (0.2 µm) and store 4°C

[0862] **Neuron Plating Medium (100 ml)**

97 ml Neurobasal

2 ml B27 Supplement

1 ml penicillin-streptomycin solution

250 µl glutamine

125 µl glutamate

[0863] Neurons are transduced with concentrated AAV1/2 virus or AAV1 virus from filtered medium of HEK293FT cells, between 4-7 days in culture and keep for at least one week in culture after transduction to allow for delivered gene expression.

[0864] *AAV-driven expression of the system*

[0865] Applicants confirmed expression of SpCas9 and SaCas9 in neuronal cultures after AAV delivery using Western blot method (Figure 24). One week after transduction neurons were collected in NuPage SDS loading buffer with β-mercaptoethanol to denaturate proteins in 95°C for 5 min. Samples were separated on SDS PAGE gel and transferred on PVDF membrane for WB protein detection. Cas9 proteins were detected with HA antibody.

[0866] Expression of *Syn*-GFP-kash from gRNA multiplex AAV was confirmed with fluorescent microscopy (Figure 32).

[0867] *Toxicity*

[0868] To assess the toxicity of AAV with CRISPR system Applicants tested overall morphology of neurons one week after virus transduction (Figure 27). Additionally, Applicants tested potential toxicity of designed system with the LIVE/DEAD® Cell Imaging Kit, which allows to distinguish live and dead cells in culture. It is based on the presence of intracellular esterase activity (as determined by the enzymatic conversion of the non-fluorescent calcein AM to the intensely green fluorescent calcein). On the other hand, the red, cell-impermeant component of the Kit enters cells with damaged membranes only and bind to DNA generating fluorescence in dead cells. Both flourophores can be easily visualized in living cells with fluorescent microscopy. AAV-driven expression of Cas9 proteins and multiplex gRNA constructs in the primary cortical neurons was well tolerated and not toxic (Figures 25 and 26), what indicates that designed AAV system is suitable for in vivo tests.

[0869] ***Virus production***

[0870] Concentrated virus was produced according to the methods described in McClure et al., 2011. Supernatant virus production occurred in HEK293FT cells.

[0871] ***Brain surgeries***

[0872] For viral vector injections 10-15 week old male C57BL/6N mice were anesthetized with a Ketamine/Xylazine cocktail (Ketamine dose of 100 mg/kg and Xylazine dose of 10 mg/kg) by intraperitoneal injection. Intraperitonial administration of Buprenex was used as a pre-emptive analgesic (1 mg/kg). Animals were immobilized in a Kopf stereotaxic apparatus using intra-aural positioning studs and tooth bar to maintain an immobile skull. Using a hand-held drill, a hole (1-2mm) at -3.0 mm posterior to Bregma and 3.5 mm lateral for injection in the CA1 region of the hippocampus was made. Using 30G World Precision Instrument syringe at a depth

of 2.5 mm, the solution of AAV viral particles in a total volume of 1 ul was injected. The injection was monitored by a 'World Precision Instruments UltraMicroPump3' injection pump at a flow rate of 0.5 ul/min to prevent tissue damage. When the injection was complete, the injection needle was removed slowly, at a rate of 0.5 mm/min. After injection, the skin was sealed with 6-0 Ethilon sutures. Animals were postoperatively hydrated with 1 mL lactated Ringer's (subcutaneous) and housed in a temperature controlled (37°C) environment until achieving an ambulatory recovery. 3 weeks after surgery animals were euthanized by deep anesthesia followed by tissue removal for nuclei sorting or with 4% paraformaldehyde perfusion for immunochemistry.

[0873] ***Sorting nuclei and in vivo results***

[0874] Applicants designed a method to specifically genetically tag the gRNA targeted neuronal cell nuclei with GFP for Fluorescent Activated Cell Sorting (FACS) of the labeled cell nuclei and downstream processing of DNA, RNA and nuclear proteins. To that purpose the applicants' multiplex targeting vector was designed to express both a fusion protein between GFP and the mouse nuclear membrane protein domain KASH (Starr DA, 2011, Current biology) and the 3 gRNAs to target specific gene loci of interest (Figure 16). GFP-KASH was expressed under the control of the human Synapsin promoter to specifically label neurons. The amino acid of the fusion protein GFP-KASH was:

MVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICT
TGKLPVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGN
YKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIK
VNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMV
LLEFVTAAGITLGMDELYKSGLRSREEEEETDSRMPHLDSPGSSQPRRSFLSRVIRAAL
PLQLLLLLLLLLACLLPASEDDYSCTQANNFARSFYPMLRYTNGPPPT

[0875] One week after AAV1/2 mediated delivery into the brain a robust expression of GFP-KASH was observed. For FACS and downstream processing of labeled nuclei, the hippocampi were dissected 3 weeks after surgery and processed for cell nuclei purification using a gradient centrifugation step. For that purpose the tissue was homogenized in 320 mM Sucrose, 5 mM CaCl, 3 mM Mg(Ac)2, 10 mM Tris pH 7.8, 0.1 mM EDTA, 0.1% NP40, 0.1 mM Phenylmethanesulfonyl fluoride (PMSF), 1 mM β-mercaptoethanol using 2ml Dounce homogenizer (Sigma) The homogenisate was centrifuged on a 25% to 29% Optiprep® gradient

according to the manufacture's protocol for 30 min at 3.500 rpm at 4 °C. The nuclear pellet was resuspended in 340 mM Sucrose, 2 mM MgCl2, 25 mM KCl, 65 mM glycerophosphate, 5% glycerol, 0.1 mM PMSF, 1 mM β-mercaptoethanol and Vybrant® DyeCycle™ Ruby Stain (Life technologies) was added to label cell nuclei (offers near-infrared emission for DNA). The labeled and purified nuclei were sorted by FACS using an Aria Flu-act-cell sorter and BDFACS Diva software. The sorted GFP+ and GFP- nuclei were finally used to purify genomic DNA using DNAeasy Blood & Tissue Kit (Qiagen) for Surveyor assay analysis of the targeted genomic regions. The same approach can be easily used to purify nuclear RNA or protein from targeted cells for downstream processing. Due to the 2-vector system (Figure 16) the applicants using in this approach efficient Cas9 mediated DNA cleavage was expected to occur only in a small subset of cells in the brain (cells which were co-infected with both the multiplex targeting vector and the Cas9 encoding vector). The method described here enables the applicants to specifically purify DNA, RNA and nuclear proteins from the cell population expressing the 3 gRNAs of interest and therefore are supposed to undergo Cas9 mediated DNA cleavage. By using this method the applicants were able to visualize efficient DNA cleavage *in vivo* occurring only in a small subset of cells.

[0876] Essentially, what Applicants have shown here is targeted *in vivo* cleavage. Furthermore, Applicants used a multiple approach, with several different sequences targeted at the same time, but independently. Presented system can be applied for studying brain pathologic conditions (gene knock out, e.g. Parkinson disease) and also open a field for further development of genome editing tools in the brain. By replacing nuclease activity with gene transcription regulators or epigenetic regulators it will be possible to answer whole spectrum of scientific question about role of gene regulation and epigenetic changes in the brain in not only in the pathologic conditions but also in physiological process as learning and memory formation. Finally, presented technology can be applied in more complex mammalian system as primates, what allows to overcome current technology limitations.

Example 28: Model Data

[0877] Several disease models have been specifically investigated. These include *de novo* autism risk genes CHD8, KATNAL2, and SCN2A; and the syndromic autism (Angelman

Syndrome) gene UBE3A. These genes and resulting autism models are of course preferred, but show that the invention may be applied to any gene and therefore any model is possible.

[0878] Applicants have made these cells lines using Cas9 nuclease in human embryonic stem cells (hESCs). The lines were created by transient transfection of hESCs with Cbh-Cas9-2A-EGFP and pU6-sgRNA. Two sgRNAs are designed for each gene targeting most often the same exons in which patient nonsense (knock-out) mutations have been recently described from whole exome sequencing studies of autistic patients. The Cas9-2A-EGFP and pU6 plasmids were created specifically for this project.

Example 29: AAV production system or protocol

[0879] An AAV production system or protocol that was developed for, and works particularly well with, high through put screening uses is provided herein, but it has broader applicability in the present invention as well. Manipulating endogenous gene expression presents various challenges, as the rate of expression depends on many factors, including regulatory elements, mRNA processing, and transcript stability. To overcome this challenge, Applicants developed an adeno-associated virus (AAV)-based vector for the delivery. AAV has an ssDNA-based genome and is therefore less susceptible to recombination.

[0880] AAV1/2 (serotype AAV1/2, i.e., hybrid or mosaic AAV1 / AAV2 capsid AAV) heparin purified concentrated virus protocol

[0881] **Media: D10 + HEPES**

500ml bottle DMEM high glucose + Glutamax (GIBCO)

50ml Hyclone FBS (heat-inactivated) (Thermo Fischer)

5.5ml HEPES solution (1M, GIBCO)

Cells: low passage HEK293FT (passage <10 at time of virus production, thaw new cells of passage 2-4 for virus production, grow up for 3-5 passages)

[0882] **Transfection reagent: Polyethylenimine (PEI) "Max"**

Dissolve 50mg PEI "Max" in 50ml sterile Ultrapure H20

Adjust pH to 7.1

Filter with 0.22um fliptop filter

Seal tube and wrap with parafilm

Freeze aliquots at -20°C (for storage, can also be used immediately)

[0883] **Cell Culture**

Culture low passage HEK293FT in D10 + HEPES

Passage everyday between 1:2 and 1:2.5

Advantageously do not allow cells to reach more than 85% confluency

[0884] For T75

- Warm 10ml HBSS (-Mg2+, -Ca2+, GIBCO) + 1ml TrypLE Express (GIBCO) per flask to 37°C (Waterbath)

Aspirate media fully

- Add 10ml warm HBSS gently (to wash out media completely)

- Add 1ml TrypLE per Flask

- Place flask in incubator (37°C) for 1min

- Rock flask to detach cells

- Add 9ml D10 + HEPES media (37°C)

- Pipette up and down 5 times to generate single cell suspension

- Split at 1:2 – 1:2.5 (12ml media for T75) ratio (if cells are growing more slowly, discard and thaw a new batch, they are not in optimal growth)

- transfer to T225 as soon as enough cells are present (for ease of handling large amounts of cells)

[0885] **AAV production (5*15cm dish scale per construct):**

Plate 10 million cells in 21.5 ml media into a 15cm dish

Incubate for 18-22 hours at 37°C

Transfection is ideal at 80% confluence

[0886] **Per plate**

Prewarm 22ml media (D10 + HEPES)

[0887] Prepare tube with DNA mixture (use endofree maxiprep DNA):

5.2 ug vector of interest plasmid

4.35 ug AAV 1 serotype plasmid

4.35 ug AAV 2 serotype plasmid

10.4 ug pDF6 plasmid (adenovirus helper genes) □ Vortex to mix

Add 434 uL DMEM (no serum!)

Add 130 ul PEI solution

Vortex 5-10 seconds

Add DNA/DMEM/PEI mixture to prewarmed media

Vortex briefly to mix

Replace media in 15cm dish with DNA/DMEM/PEI mixture

Return to 37°C incubator

Incubate 48h before harvesting (make sure medium isn't turning too acidic)

[0888] Virus harvest:

1. aspirate media carefully from 15cm dish dishes (advantageously do not dislodge cells)

2. Add 25 ml RT DPBS (Invitrogen) to each plate and gently remove cells with a cell scraper. Collect suspension in 50 ml tubes.

3. Pellet cells at 800x g for 10 minutes.

4. Discard supernatant

[0889] pause point: freeze cell pellet at -80C if desired

5. resuspend pellet in 150 mM NaCl, 20 mM Tris pH 8.0, use 10 ml per tissue culture plate.

6. Prepare a fresh solution of 10% sodium deoxycholate in dH2O. Add 1.25 ml of this per tissue culture plate for a final concentration of 0.5%. Add benzonase nuclease to a final concentration of 50 units per ml. Mix tube thoroughly.

7. Incubate at 37°C for 1 hour (Waterbath).

8. Remove cellular debris by centrifuging at 3000 x g for 15 mins. Transfer to fresh 50 ml tube and ensure all cell debris has been removed to prevent blocking of heparin columns.

[0890] Heparin column purification of AAV1/2:

1. **Set up** HiTrap heparin columns using a peristaltic pump so that solutions flow through the column at 1 ml per minute. It is important to ensure no air bubbles are introduced into the heparin column.

2. **Equilibrate** the column with 10 ml 150 mM NaCl, 20 mM Tris, pH 8.0 using the peristaltic pump.

3. **Binding of virus:** Apply 50 ml virus solution to column and allow to flow through.

4. **Wash step 1:** column with 20 ml 100 mM NaCl, 20 mM Tris, pH 8.0. (using the peristaltic pump)

5. **Wash step 2:** Using a 3 ml or 5 ml syringe continue to wash the column with 1 ml 200 mM NaCl, 20 mM Tris, pH 8.0, followed by 1 ml 300 mM NaCl, 20 mM Tris, pH 8.0.

Discard the flow-through.

(prepare the syringes with different buffers during the 50min flow through of virus solution above)

6. **Elution** Using 5 ml syringes and gentle pressure (flow rate of < 1ml/min) elute the virus from the column by applying:

1.5 ml 400 mM NaCl, 20 mM Tris, pH 8.0

3.0 ml 450 mM NaCl, 20 mM Tris, pH 8.0

1.5 ml 500 mM NaCl, 20 mM Tris, pH 8.0

Collect these in a 15 ml centrifuge tube.

[0891] Concentration of AAV1/2:

1. Concentration step 1: Concentrate the eluted virus using Amicon ultra 15ml centrifugal filter units with a 100,000 molecular weight cutoff. Load column eluate into the concentrator and centrifuge at 2000x g for 2 minutes (at room temperature. Check concentrated volume – it should be approximately 500 µl. If necessary, centrifuge in 1min intervals until correct volume is reached.

2. **buffer exchange:** Add 1ml sterile DPBS to filter unit, centrifuge in 1min intervals until correct volume (500ul) is reached.

3. **Concentration step 2:** Add 500ul concentrate to an Amicon Ultra 0.5ml 100K filter unit. Centrifuge at 6000g for 2min. Check concentrated volume – it should be approximately 100 µl. If necessary, centrifuge in 1min intervals until correct volume is reached.

4. **Recovery**: Invert filter insert and insert into fresh collection tube. Centrifuge at 1000g for 2min.

Aliquot and freeze at -80°C

1ul is typically required per injection site, small aliquots (e.g. 5ul) are therefore recommended (avoid freeze-thaw of virus).

determine DNaseI-resistant GC particle titer using qPCR (see separate protocol)

[0892] Materials

Amicon Ultra, 0.5ml, 100K; MILLIPORE; UFC510024

Amicon Ultra, 15ml, 100K; MILLIPORE; UFC910024

Benzonase nuclease; Sigma-Aldrich, E1014

HiTrap Heparin cartridge; Sigma-Aldrich; 54836

Sodium deoxycholate; Sigma-Aldrich; D5670

[0893] AAV1 supernatant production protocol

Media: D10 + HEPES

500ml bottle DMEM high glucose + Glutamax (Invitrogen)

50ml Hyclone FBS (heat-inactivated) (Thermo Fischer)

5.5ml HEPES solution (1M, GIBCO)

Cells: low passage HEK293FT (passage <10 at time of virus production)

Thaw new cells of passage 2-4 for virus production, grow up for 2-5 passages

Transfection reagent: Polyethylenimine (PEI) "Max"

Dissolve 50mg PEI "Max" in 50ml sterile Ultrapure H20

Adjust pH to 7.1

Filter with 0.22um fliptop filter

Seal tube and wrap with parafilm

Freeze aliquots at -20°C (for storage, can also be used immediately)

Cell Culture

Culture low passage HEK293FT in D10 + HEPES Passage everyday between 1:2 and 1:2.5

Advantageously do let cells reach more than 85% confluency

For T75

- Warm 10ml HBSS (-Mg2+, -Ca2+, GIBCO) + 1ml TrypLE Express (GIBCO) per flask to 37°C (Waterbath)

- Aspirate media fully

- Add 10ml warm HBSS gently (to wash out media completely)

- Add 1ml TrypLE per Flask

- Place flask in incubator (37°C) for 1min

- Rock flask to detach cells

- Add 9ml D10 + HEPES media (37°C)

- Pipette up and down 5 times to generate single cell suspension

- Split at 1:2 – 1:2.5 (12ml media for T75) ratio (if cells are growing more slowly, discard and thaw a new batch, they are not in optimal growth)

- transfer to T225 as soon as enough cells are present (for ease of handling large amounts of cells)

AAV production (single 15cm dish scale)

Plate 10 million cells in 21.5 ml media into a 15cm dish

Incubate for 18-22 hours at 37°C

Transfection is ideal at 80% confluence per plate

Prewarm 22ml media (D10 + HEPES)

Prepare tube with DNA mixture (use endofree maxiprep DNA):

5.2 ug vector of interest plasmid

8.7 ug AAV 1 serotype plasmid

10.4 ug DF6 plasmid (adenovirus helper genes)

Vortex to mix

Add 434 uL DMEM (no serum!) Add 130 ul PEI solution

Vortex 5-10 seconds

Add DNA/DMEM/PEI mixture to prewarmed media

Vortex briefly to mix

Replace media in 15cm dish with DNA/DMEM/PEI mixture

Return to 37°C incubator

Incubate 48h before harvesting (advantageously monitor to ensure medium is not turning too acidic)

[0894] Virus harvest:

Remove supernatant from 15cm dish

Filter with 0.45 um filter (low protein binding) Aliquot and freeze at -80°C

Transduction (primary neuron cultures in 24-well format, 5DIV)

Replace complete neurobasal media in each well of neurons to be transduced with fresh neurobasal (usually 400ul out of 500ul per well is replaced)

Thaw AAV supernatant in 37°C waterbath

Let equilibrate in incubator for 30min

Add 250 ul AAV supernatant to each well

Incubate 24h at 37°C

Remove media/supernatant and replace with fresh complete neurobasal

Expression starts to be visible after 48h, saturates around 6-7 Days Post Infection

Constructs for pAAV plasmid with GOI should not exceed 4.8kb including both ITRS.

[0895] Example of a human codon optimized sequence (i.e. being optimized for expression in humans) sequence : SaCas9 is provided below:

ACCGGTGCCACCATGTACCCATACGATGTTCCAGATTACGC
TTCGCCGAAGAAAAAGCGCAAGGTCGAAGCGTCCATGAAAAGGAACTACATTCT
GGGGCTGGACATCGGGATTACAAGCGTGGGGTATGGGATTATTGACTATGAAAC
AAGGGACGTGATCGACGCAGGCGTCAGACTGTTCAAGGAGGCCAACGTGGAAAA
CAATGAGGGACGGAGAAGCAAGAGGGGAGCCAGGCGCCTGAAACGACGGAGAA
GGCACAGAATCCAGAGGGTGAAGAAACTGCTGTTCGATTACAACCTGCTGACCG
ACCATTCTGAGCTGAGTGGAATTAATCCTTATGAAGCCAGGGTGAAAGGCCTGA
GTCAGAAGCTGTCAGAGGAAGAGTTTTCCGCAGCTCTGCTGCACCTGGCTAAGC
GCCGAGGAGTGCATAACGTCAATGAGGTGGAAGAGGACACCGGCAACGAGCTG
TCTACAAAGGAACAGATCTCACGCAATAGCAAAGCTCTGGAAGAGAAGTATGTC
GCAGAGCTGCAGCTGGAACGGCTGAAGAAAGATGGCGAGGTGAGAGGGTCAATT
AATAGGTTCAAGACAAGCGACTACGTCAAAGAAGCCAAGCAGCTGCTGAAAGTG
CAGAAGGCTTACCACCAGCTGGATCAGAGCTTCATCGATACTTATATCGACCTGC
TGGAGACTCGGAGAACCTACTATGAGGGACCAGGAGAAGGGAGCCCCTTCGGAT
GGAAAGACATCAAGGAATGGTACGAGATGCTGATGGGACATTGCACCTATTTTCC
AGAAGAGCTGAGAAGCGTCAAGTACGCTTATAACGCAGATCTGTACAACGCCCT
GAATGACCTGAACAACCTGGTCATCACCAGGGATGAAAACGAGAAACTGGAATA
CTATGAGAAGTTCCAGATCATCGAAAACGTGTTTAAGCAGAAGAAAAAGCCTAC
ACTGAAACAGATTGCTAAGGAGATCCTGGTCAACGAAGAGGACATCAAGGGCTA
CCGGGTGACAAGCACTGGAAAACCAGAGTTCACCAATCTGAAAGTGTATCACGA
TATTAAGGACATCACAGCACGGAAAGAAATCATTGAGAACGCCGAACTGCTGGA
TCAGATTGCTAAGATCCTGACTATCTACCAGAGCTCCGAGGACATCCAGGAAGAG
CTGACTAACCTGAACAGCGAGCTGACCCAGGAAGAGATCGAACAGATTAGTAAT
CTGAAGGGGTACACCGGAACACACAACCTGTCCCTGAAAGCTATCAATCTGATTC
TGGATGAGCTGTGGCATACAAACGACAATCAGATTGCAATCTTTAACCGGCTGAA
GCTGGTCCCAAAAAAGGTGGACCTGAGTCAGCAGAAAGAGATCCCAACCACACT
GGTGGACGATTTCATTCTGTCACCCGTGGTCAAGCGGAGCTTCATCCAGAGCATCAA
AGTGATCAACGCCATCATCAAGAAGTACGGCCTGCCCAATGATATCATTATCGAGCT
GGCTAGGGAGAAGAACAGCAAGGACGCACAGAAGATGATCAATGAGATGCAGAAA

CGAAACCGGCAGACCAATGAACGCATTGAAGAGATTATCCGAACTACCGGGAAAGA
GAACGCAAAGTACCTGATTGAAAAAATCAAGCTGCACGATATGCAGGAGGGAAAGT
GTCTGTATTCTCTGGAGGCCATCCCCCTGGAGGACCTGCTGAACAATCCATTCAACT
ACGAGGTCGATCATATTATCCCCAGAAGCGTGTCCTTCGACAATTCCTTTAACAACA
AGGTGCTGGTCAAGCAGGAAGAGAACTCTAAAAAGGGCAATAGGACTCCTTTCCAG
TACCTGTCTAGTTCAGATTCCAAGATCTCTTACGAAACCTTTAAAAAGCACATTCTG
AATCTGGCCAAAGGAAAGGGCCGCATCAGCAAGACCAAAAAGGAGTACCTGCTGG
AAGAGCGGGACATCAACAGATTCTCCGTCCAGAAGGATTTTATTAACCGGAATCTG
GTGGACACAAGATACGCTACTCGCGGCCTGATGAATCTGCTGCGATCCTATTTCCGG
GTGAACAATCTGGATGTGAAAGTCAAGTCCATCAACGGCGGGTTCACATCTTTTCTG
AGGCGCAAATGGAAGTTTAAAAAGGAGCGCAACAAAGGGTACAAGCACCATGCCG
AAGATGCTCTGATTATCGCAAATGCCGACTTCATCTTTAAGGAGTGGAAAAAGCTGG
ACAAAGCCAAGAAAGTGATGGAGAACCAGATGTTCGAAGAGAAGCAGGCCGAATC
TATGCCCGAAATCGAGACAGAACAGGAGTACAAGGAGATTTTCATCACTCCTCACC
AGATCAAGCATATCAAGGATTTCAAGGACTACAAGTACTCTCACCGGGTGGATAAA
AAGCCCAACAGAGAGCTGATCAATGACACCCTGTATAGTACAAGAAAAGACGATAA
GGGGAATACCCTGATTGTGAACAATCTGAACGGACTGTACGACAAAGATAATGACA
AGCTGAAAAAGCTGATCAACAAAAGTCCCGAGAAGCTGCTGATGTACCACCATGAT
CCTCAGACATATCAGAAACTGAAGCTGATTATGGAGCAGTACGGCGACGAGAAGAA
CCCACTGTATAAGTACTATGAAGAGACTGGGAACTACCTGACCAAGTATAGCAAAA
AGGATAATGGCCCCGTGATCAAGAAGATCAAGTACTATGGGAACAAGCTGAATGCC
CATCTGGACATCACAGACGATTACCCTAACAGTCGCAACAAGGTGGTCAAGCTGTC
ACTGAAGCCATACAGATTCGATGTCTATCTGGACAACGGCGTGTATAAATTTGTGAC
TGTCAAGAATCTGGATGTCATCAAAAAGGAGAACTACTATGAAGTGAATAGCAAGT
GCTACGAAGAGGCTAAAAAGCTGAAAAAGATTAGCAACCAGGCAGAGTTCATCGCC
TCCTTTTACAACAACGACCTGATTAAGATCAATGGCGAACTGTATAGGGTCATCGGG
GTGAACAATGATCTGCTGAACCGCATTGAAGTGAATATGATTGACATCACTTACCGA
GAGTATCTGGAAAACATGAATGATAAGCGCCCCCCTCGAATTATCAAAACAATTGC
CTCTAAGACTCAGAGTATCAAAAAGTACTCAACCGACATTCTGGGAAACCTGTATGA
GGTGAAGAGCAAAAAGCACCCTCAGATTATCAAAAAGGGCTAAGAATTC

Example 30: Minimizing off-target cleavage using Cas9 nickase and two guide RNAs

[0896] Cas9 is a RNA-guided DNA nuclease that may be targeted to specific locations in the genome with the help of a 20bp RNA guide. However the guide sequence may tolerate some mismatches between the guide sequence and the DNA-target sequence. The flexibility is undesirable due to the potential for off-target cleavage, when the guide RNA targets Cas9 to a an off-target sequence that has a few bases different from the guide sequence. For all experimental applications (gene targeting, crop engineering, therapeutic applications, etc) it is important to be able to improve the specificity of Cas9 mediated gene targeting and reduce the likelihood of off-target modification by Cas9.

[0897] Applicants developed a method of using a Cas9 nickase mutant in combination with two guide RNAs to facilitate targeted double strand breaks in the genome without off-target modifications. The Cas9 nickase mutant may be generated from a Cas9 nuclease by disabling its cleavage activity so that instead of both strands of the DNA duplex being cleaved only one strand is cleaved. The Cas9 nickase may be generated by inducing mutations in one ore more domains of the Cas9 nuclease, e.g. Ruvc1 or HNH. These mutations may include but are not limited to mutations in a Cas9 catalytic domain, e.g in SpCas9 these mutations may be at positions D10 or H840. These mutations may include but are not limited to D10A, E762A, H840A, N854A, N863A or D986A in SpCas9 but nickases may be generated by inducing mutations at corresponding positions in other CRISPR enzymes or Cas9 orthologs. In a most preferred embodiment of the invention the Cas9 nickase mutant is a SpCas9 nickase with a D10A mutation.

[0898] The way this works is that each guide RNA in combination with Cas9 nickase would induce the targeted single strand break of a duplex DNA target. Since each guide RNA nicks one strand, the net result is a double strand break. The reason this method eliminates off-target mutations is because it is very unlikely to have an off-target site that has high degrees of similarity for both guide sequences (20bp+2bp(PAM) = 22bp specificity for each guide, and two guides means any off-target site will have to have close to 44bp of homologous sequence). Although it is still likely that individual guides may have off-targets, but those off-targets will only be nicked, which is unlikely to be repaired by the mutagenic NHEJ process. Therefore the multiplexing of DNA double strand nicking provides a powerful way of introducing targeted DNA double strand breaks without off-target mutagenic effects.

[0899] Applicants carried out experiments involving the co-transfection of HEK293FT cells with a plasmid encoding Cas9(D10A) nickase as well as DNA expression cassettes for one or more guides. Applicants transfected cells using Lipofectamine 2000, and transfected cells were harvested 48 or 72 hours after transfections. Double nicking-induced NHEJ were detected using the SURVEYOR nuclease assay as described previously herein (Figures 33, 34 and 35).

[0900] Applicants have further identified parameters that relate to efficient cleavage by the Cas9 nickase mutant when combined with two guide RNAs and these parameters include but are not limited to the length of the 5' overhang. Efficient cleavage is reported for 5' overhang of at least 26 base pairs. In a preferred embodiment of the invention, the 5' overhang is at least 30 base pairs and more preferably at least 34 base pairs. Overhangs of up to 200 base pairs may be acceptable for cleavage, while 5' overhangs less than 100 base pairs are preferred and 5' overhangs less than 50 base pairs are most preferred (Figures 36 and 37).

Example 31: Tandem single guide RNAs (tsgRNAs)

[0901] The bacterial CRISPR-Cas system is an RNA-guided endonuclease that Applicants have shown can induce targeted modification of the eukaryotic genome predictably, precisely, and efficiently. The minimal components required to reconstitute this system in mammalian cells consist of delivery of the Cas9 endonuclease as well as a guide RNA carrying a 20bp specificity sequence complementary to the target site of interest. Until now, delivery of multiple individual guide RNAs was required to achieve multiplex genome editing. While co-transfection of multiple guides is relative simple in tissue culture, many *in vivo* and therapeutic applications for the CRISPR-Cas system would benefit from a single-vector system that could still allow multiplex editing.

[0902] Applicants mutagenized different parts of the original chimeric guide RNA architecture to determine what portions of the chimeric guide RNA would be amenable for alteration or improvement. From these studies, Applicants found that the two distal hairpins appear primarily to function in stabilization of the chimeric RNA. Alteration of the direct-repeat-and-tracrRNA duplex region, as well as the unstructured region between the duplex and the first hairpin of the tracrRNA abolished cleavage activity. However, increased stabilization of the two tracrRNA tail hairpins through introduction of G-C pairs (Architectures A and B in Figure 38A) did not decrease the ability of Cas9 to induce targeted indels.

[0903] In the native bacterial CRISPR system, Cas9 processes a continuous repetitive transcript of DR-flanked spacers into discrete units upon assembly with tracrRNA. Applicants reasoned that multiple preassembled sgRNA could likewise be processed into discrete functional units. To create tandem sgRNA (tsgRNA), an 8-nt linker was introduced between the tracrRNA end of the first sgRNA and the spacer of the following sgRNA (Figure 38B). Applicants tested the tsgRNA in two assays: microdeletion induction with wild-type Cas9 (Figure 38C) and indel formation with nickase Cas9n (Figure 38D). Applicants found that both tsgRNAs created with either Architecture A or B scaffolds were able to achieve targeted modification at levels comparable to individual delivery of both guide RNAs.

[0904] Tandem sgRNA are synthesized in a 2-round PCR as follows:

Round 1: Amplification using U6 promoter as template, U6-Fwd primer (as in previous PCR expression cassette experiments for sgRNA delivery), and a modified Reverse primer that contains from 5' to 3' (in reverse complement direction): spacer-2, sgRNA modified scaffold, spacer-1, U6 priming region.

Round 2: Amplifies using product from round 1 as template, using U6-Fwd primer as previously described, and a reverse primer 5' to 3' (in reverse complement direction): modified scaffold, spacer-2. (Figure 38).

[0905] After 2 rounds of PCR, the full-length tsgRNA product is purified and co-transfected with Cas9 for testing in cells.

Example 32: Tandem single guide RNAs (tsgRNAs)

[0906] The ability to simultaneously deliver more than a single guide RNA on a single vector is particularly tractable for lentiviral screening methodologies. The propensity of lentivirus to recombine has limited the ability to drive expression of multiple short RNAs from a single vector with the requirement of utilizing multiple unique promoters. A tandem approach may allow a single promoter to drive expression of at least two guide RNAs. In addition to allowing one to target combinations of genomic targets, placing two guide RNAs immediately adjacent to one another facilitates cloning in a single-step high-throughput fashion (Fig. 39A). To further decrease homology between adjacent scaffolds to lessen viral recombination and to increase guide stability within the tandem array, Applicants have created and are validating divergent scaffold sequences that conserve secondary structure (Fig. 39B). Finally, through Northern blot

analyses using a probe targeted against a particular spacer-scaffold combination, Applicants have observed that tandem guide RNAs are indeed processed to single sgRNA units and that the efficiency of processing is increased with the modified scaffolds relative to the wild-type+85 scaffold (Fig. 39C). Assembly of tandem guides can be accomplished by annealing of ultramers containing Spacer1-Scaffold-Spacer 2 (Fig. 39A), two-step PCR, or by Golden Gate assembly, which would be most amenable to construction of longer arrays.

Example 33: Characterization and optimization of the CRISPR/Cas system for applications in genome engineering.

[1000] Recent studies of Cas9 specificity have shown that although each base within the 20-nt guide sequence contributes to overall specificity, multiple mismatches between the guide RNA and its complementary DNA can be tolerated in a quantity-, position-, and base identity-sensitive manner. As a result, Cas9 can cleave genomic loci that share imperfect homology with the target 20-nt guide sequence, leading to off-target DSBs and NHEJ repair. Subsequent indel formation at off-target cleavage sites can lead to significant levels of unwanted mutations, which limit the utility of Cas9 for genome editing applications requiring high levels of precision, including generation of isogenic cell lines for testing causal genetic variations as well as *in vivo* and *ex vivo* genome editing-based therapies.

[0907] To improve the specificity of Cas9-mediated genome editing, Applicants developed a novel strategy that combines the D10A mutant nickase version of Cas9 (Cas9n) with a pair of offset sgRNAs targeting opposite strands of the target site. Nicking of both DNA strands at the target site by a pair of Cas9 nickases leads to site-specific DSBs, while individual nicks are predominantly repaired by the high-fidelity base excision repair pathway (BER) as opposed to error-prone NHEJ. This strategy would minimize off-target mutagenesis by each Cas9n-sgRNA complex while maximizing on-target NHEJ comparable to wild-type Cas9 and would be analogous to dimeric zinc finger nucleases (ZFNs) and transcription activator-like effector nucleases (TALENs), where DNA cleavage relies upon the synergistic interaction of two independent specificity-encoding modules. ZFNs and TALENs generate DSBs through the proximity-induced dimerization of two FokI monomers, each of which nicks one DNA strand. Similarly, Applicants paired Cas9n with two sgRNAs targeting opposite strands of a desired locus. This 'double nicking' strategy would effectively magnify the targeting specificity of Cas9 by requiring simultaneous targeting by two sgRNAs.

[0908] Finally, to facilitate the co-delivery of multiple sgRNA, Applicants have developed a system for expressing pairs of sgRNAs under a single promoter. Applicants first studied the structural components of sgRNA critical for function, and secondly used this knowledge to inform the design of sequence-divergent new sgRNA scaffolds that facilitate tandem sgRNA transcription.

[0909] Materials and Methods

[0910] PCR amplification of U6-promoter-driven sgRNAs and tandem sgRNAs.

[0911] Spacer selection for targeting by Cas9 and subsequent generation of PCR amplicon was performed as described in Ran et al. Briefly, oligo ultramers consisting of U6 priming site, spacer sequence, and guide RNA scaffold were synthesized by IDT for amplification of U6-driven PCR cassettes for cellular transfections. In both cases, either QiaQuick (Qiagen) or EconoSpin (Epoch Life Sciences) spin columns were used to clean up PCR reactions prior to transfections. Tandem sgRNA are synthesized in a 2-round PCR as follows: Round 1: Amplification using U6 promoter as template, U6-Fwd primer (as in previous PCR expression cassette experiments for sgRNA delivery), and a modified Reverse primer that contains from 5' to 3' (in reverse complement direction): spacer-2, sgRNA modified scaffold, spacer-1, U6 priming region. Round 2: Amplifies using product from round 1 as template, using U6-Fwd primer as previously described, and a reverse primer 5' to 3' (in reverse complement direction): modified scaffold, spacer-2. After 2 rounds of PCR, the full-length tsgRNA product is purified and co-transfected with Cas9 for testing in cells. In both cases, either QiaQuick (Qiagen) or EconoSpin (Epoch Life Sciences) spin columns were used to clean up PCR reactions prior to transfections. A list of the sgRNAs used and their genomic targets can be found in Table E.

[0912] Cell culture and transfection

[0913] Human embryonic kidney (HEK) cell line 293FT (Life Technologies) was maintained in Dulbecco's modified Eagle's Medium (DMEM) supplemented with 10% fetal bovine serum (HyClone), 2mM GlutaMAX (Life Technologies), 100U/mL penicillin, and 100μg/mL streptomycin at 37°C with 5% CO2 incubation. Cells are passaged at regular intervals and seeded onto 24-well plates (Corning) at a density of 120,000 cells/well, 24 hours prior to transfection. Cells were transfected using Lipofectamine 2000 (Life Technologies) at 80-90% confluency per manufacturer recommended protocol: A total of 400ng Cas9 plasmid and 100 ng of U6-sgRNA PCR product was transfected per well of a 24-well plate. For double-nicking

experiments or transfections involving more than a single guide, 100ng of each sgRNA was transfected. In the case of tandem sgRNAs, 200ng of purified U6-tsgRNA PCR product was transfected per well.

[0914] Human embryonic stem cell line HUES62 (Harvard Stem Cell Institute core) was maintained in feeder-free conditions on GelTrex (Life Technologies) in mTesR medium (Stemcell Technologies) supplemented with 100ug/ml Normocin (InvivoGen). HUES62 cells were transfected with Amaxa P3 Primary Cell 4-D Nucleofector Kit (Lonza) following the manufacturer's protocol.

[0915] SURVEYOR nuclease assay for genome modification

[0916] 293FT and HUES62 cells were transfected with DNA as described above. Cells were incubated at 37°C for 72 hours post-transfection prior to genomic DNA extraction. Genomic DNA was extracted using the QuickExtract DNA Extraction Solution (Epicentre) following the manufacturer's protocol. Briefly, pelleted cells were resuspended in QuickExtract solution and incubated at 65°C for 15 minutes, 68°C for 15 minutes, and 98°C for 10 minutes.

[0917] The genomic region flanking the CRISPR target site for each gene was PCR amplified (primers listed in Table G), and products were purified using QiaQuick Spin Column (Qiagen) following the manufacturer's protocol. 400ng total of the purified PCR products were mixed with 2μl 10X Taq DNA Polymerase PCR buffer (Enzymatics) and ultrapure water to a final volume of 20μl, and subjected to a re-annealing process to enable heteroduplex formation: 95°C for 10min, 95°C to 85°C ramping at – 2°C/s, 85°C to 25°C at – 0.25°C/s, and 25°C hold for 1 minute. After re-annealing, products were treated with SURVEYOR nuclease and SURVEYOR enhancer S (Transgenomics) following the manufacturer's recommended protocol, and analyzed on 4-20% Novex TBE poly-acrylamide gels (Life Technologies). Gels were stained with SYBR Gold DNA stain (Life Technologies) for 30 minutes and imaged with a Gel Doc gel imaging system (Bio-rad). Quantification was based on relative band intensities. Indel percentage was determined by the formula, 100 x (1 - (1 - (b + c) / (a + b + c))1/2), where a is the integrated intensity of the undigested PCR product, and b and c are the integrated intensities of each cleavage product.

[0918] Deep sequencing to assess targeting specificity

[0919] HEK 293FT cells were plated and transfected as described above, 72 hours prior to genomic DNA extraction. The genomic region flanking the CRISPR target site for each gene

was amplified (primers listed in Table H) by a fusion PCR method to attach the Illumina P5 adapters as well as unique sample-specific barcodes to the target. PCR products were purified using EconoSpin 96-well Filter Plates (Epoch Life Sciences) following the manufacturer's recommended protocol.

[0920] Barcoded and purified DNA samples were quantified by Qubit 2.0 Fluorometer (Life Technologies) and pooled in an equimolar ratio. Sequencing libraries were then sequenced with the Illumina MiSeq Personal Sequencer (Life Technologies).

[0921] Sequencing data analysis, indel detection, and homologous recombination detection

[0922] MiSeq reads were filtered by requiring an average Phred quality (Q score) of at least 30, as well as perfect sequence matches to barcodes and amplicon forward primers. Reads from on- and off-target loci were analyzed by performing Ratcliff-Obershelp string comparison, as implemented in the Python difflib module, against loci sequences that included 30 nucleotides upstream and downstream of the target site (a total of 80 bp). The resulting edit operations were parsed, and reads were counted as indels if insertion or deletion operations were found. Analyzed target regions were discarded if part of their alignment fell outside the MiSeq read itself or if more than 5 bases were uncalled.

[0923] Negative controls for each sample provided a gauge for the inclusion or exclusion of indels as putative cutting events. For quantification of homologous recombination, reads were first processed as in the indel detection workflow, and then checked for presence of homologous recombination template CCAGGCTTGG.

[0924] Flow cytometric analysis for Cas9 self-targeting indel induction

[0925] Cells were transfected as above using the Cas9 plasmid PX475 encoding SpCas9-t2a-GFP in the presence of guide RNAs targeting Cas9 itself. Three days following transfection, cells were washed once with PBS, trypsinized and triturated to single cell suspension, and re-suspended in PBS buffer supplemented with 5% FBS and 2mM EDTA. Fluorescent intensity was subsequently measured using the Accuri C6 flow cytometer.

[0926] Northern blot analysis of sgRNA processingCells were transfected as described above and incubated for 72 hrs at 37C. RNA was subsequently extracted from the cells per mirVana miRNA isolation kit protocol (Life Technologies) to enrich for small RNAs. Purified small RNAs were resolved on a denaturing gel, transferred to BrightStar Positively-Charged Nylon Membrane (Ambion), and probed overnight using radioactive or biotinylated oligonucleotides

targeted against specific sgRNA spacer sequences. Visualization was performed through the use of a Typhoon imager or Li-Cor CLx machine depending on probe modality.

[0927] Results

[0928] Cas9 nickase generates efficient NHEJ with closely approximated dual guide RNAs

[0929] The targeting specificity and activity of the Cas9 nuclease is dependent on base-pairing interaction between the 20nt guide sequence within the sgRNA and the target DNA. Applicants therefore reasoned that lengthening the guide sequence might increase guide:target basepairing and increase Cas9 targeting specificity. However, this failed to improve Cas9 targeting specificity as a majority of the lengthened guide sequence is processed back to a 20-nt length. Applicants therefore explored an alternate strategy for increasing the overall base-pairing length between guide sequence and DNA target based on simultaneous nicking of both strands of DNA by two separate Cas9-sgRNA complexes. Single-strand nicks by Cas9n are preferentially repaired by the BER pathway, which typically results in extremely low levels of mutagenesis. Applicants reasoned that two nicking enzymes directed by a pair of sgRNAs targeting opposite strands of a target locus, requiring double the number of sgRNA bases paired to target DNA, might still be able to mediate DSBs while loci nicked by a single sgRNA-Cas9 duplex would be perfectly repaired (schematized in Figure 40A). By co-transfecting sgRNAs and the Cas9 D10A nickase (Cas9n), which nicks the strand of DNA complementary to the sgRNA, into human embryonic kidney (HEK293FT) cells, Applicants observed that whereas Cas9n in combination with guide pairs could efficiently induce indel formation, Cas9n with single guides alone did not result in detectable modification of the target locus by SURVEYOR assay (Figure 40B).

[0930] Given that the double-nicking strategy requires two Cas9n-sgRNA complexes to simultaneously target the same locus, steric hindrance is likely to be of concern in determining whether any pair of sgRNAs targeting opposite strands of DNA may be used for generating DSBs. To thoroughly characterize the parameters of paired guide RNAs that would be amenable to indel formation, Applicants systematically designed sgRNA pairings targeting three different human genes separated by a range of offset distances from -200 to 200 bp, creating both 5'- and 3'-overhang products, and tested each for NHEJ (pairs listed in Table D). Significantly, across all three genes, Applicants observed substantial indel frequency (up to 40%) for sgRNA pair offsets from -4 to 20 bp (Figure 41A). Notably, indels formed by double-nicking with paired guide RNAs can result in larger and more varied types of mutations (representative indels

observed by deep sequencing shown in Figure 41B) than usually observed with single guides, which typically result in small deletions in the target sequence 4-6 bp upstream from the PAM. Occasionally, sgRNAs offset by up to 100 bp were observed to mediate on-target modification, which suggests a wide range of possible spacings for targeting. Importantly, all single sgRNAs transfections with wild-type Cas9, but not Cas9n, mediated efficient indel formation (summarized in Table D), consistent with relative spacing between guide pairs being the primary determinant of double-nickase induced genome modification. Impressively, double nickase indel frequencies were generally comparable to those mediated by wild-type Cas9 nuclease targeting the same locus. Taken together, these results indicate that double nicking can serve as a generalizable and predictable solution for efficiently mediating precisely targeted DSBs.

[0931] Double nicking allows high-efficiency homologous recombination

[0932] While induction of double-stranded DNA breaks can introduce mutagenic indels at targeted genomic loci and mediate gene knockout, it can also be a mechanism by which to facilitate homology directed repair (HDR) to enable highly precise editing or gene replacement of target sites. Given the wealth of SNP data that is being generated and the increasing association with and appreciation of small or single base-pair mutations in disease tissues through genome- or exome-wide sequencing efforts, the ability to reliably and efficiently alter small genomic regions for downstream functional testing or disease modeling would prove enormously useful.

[0933] Previously, Applicants have shown that Cas9n, when used with a single sgRNA to nick DNA, can initiate HDR. However, HDR occurs at a much lower frequency when mediated by nicking rather than DSB, which can further vary among cell types. To test the efficiency of HDR with using a double-nicking strategy, Applicants targeted the human EMX1 locus with two pairs of sgRNAs offset by -3 and 17 bp (generating 31- and 52-bp 5' overhangs, respectively) and introduced a single-stranded oligonucleotide (ssODN) bearing a HindIII restriction site as the HDR repair template in order to introduce a restriction fragment length polymorphism (RFLP) into the genomic locus (Figure 42A). Subsequent RFLP demonstrated that both sgRNA pairs were able to successfully introduce the HR template at frequencies significantly higher than those of single-guide Cas9n transfections and comparable to those of wild-type Cas9 (Figure 42B).

[0934] The growing interest and development in stem cells (ESC) or patient derived induced pluripotent stem cells (iPSC) biology represents simultaneously a key opportunity for generating new disease paradigms and developing new therapeutics, as well as an increasing need to develop ever more precise and efficient means of genome modification. While double-stranded breaks have been shown to efficiently facilitate HDR in ESC and iPSCs, there is still much interest in using nicking approaches for HDR in these sensitive applications due to their lower off-target activity. However, single nick approaches to inducing HDR in human embryonic stem cells using the CRISPR-Cas system have met with limited success. To improve HDR efficiency in ES cells, Applicants subsequently attempted double-nicking induced HDR in the HUES62 cell line observed significantly increased rates of incorporation of the HDR template (Figure 42C).

[0935] Analogous to defining optimal sgRNA spacing for indel generation by double-nicking, Applicants next sought to determine the ideal parameters for potentiating HDR. Applicants posited that to most efficiently facilitate strand invasion and subsequent conversion, at least one of the sgRNA paired RNAs should be targeted close to the site of integration. Applicants tested a variety of sgRNA pairs wherein at least one of the targeted cleavage sites was close to the site of recombination (Figure 43). Applicants observed that sgRNA pairs predicted to generate a 5' overhang with at least one target within 22bp of the site of integration were able to incorporate the provided HDR template at frequencies comparable to wild-type Cas9 nuclease mediated HDR. In contrast, sgRNA pairs that targeted the same strand of DNA, spaced by negative offsets, or that had neither sgRNAs close to the site of integration were unable to facilitate HDR at detectable levels.

[0936] Double nicking mediates highly specific genome editing

[0937] Having shown that double-nicking mediates high efficiency induction of both NHEJ and HDR at levels comparable to those induced by wild-type Cas9, Applicants next sought to determine whether this approach results in improved specificity over Cas9 through quantification of off-target activities. Applicants co-delivered Cas9n with two sgRNAs spaced by a 23-bp offset to target the human EMX1 locus (Figure 44A). As expected, this configuration of paired sgRNAs resulted in on-target indel levels comparable to those of wild-type Cas9 transfected with either sgRNA singly (Figure 44B, left panel). Strikingly, Applicants did not detect any modification by SURVEYOR assay at one of the sgRNA 1 off-target sites (OT-4) in the case of double-nicking where the wild-type Cas9 showed 10% modification (Figure 43, right panel).

Applicants subsequently used deep sequencing to assess modification at 5 different sgRNA 1 off-target loci and observed significant mutagenesis at all sites with wild-type Cas9 + sgRNA 1 alone (Figure 43). In contrast, off-target cleavage by Cas9n was barely detectable and difficult to distinguish from sequencing error. Normalized to a specificity ratio (on- to off-target indel percentage ratio), Cas9n with two sgRNAs could achieve over 100-fold greater specificity relative to wild-type Cas9 (Figure 43).

[0938] In summary, the strategy of using the nickase Cas9n with closely approximated pairs of guide RNAs is as efficient at inducing NHEJ and facilitating HDR as the wild-type nuclease, while achieving much higher targeting specificity. Furthermore, the relatively wide range of off-set distances between the double guides that is compatible with robust activity renders double-nicking an attractive and easily implemented method.

[0939] Systematic mutagenesis of sgRNA architecture identifies regions for further optimization

[0940] One of the critical elements of the type II CRISPR-Cas nuclease system is the trans-activating crRNA (tracrRNA), which shares partial sequence homology and base-pairs with the repeat region of the crRNA and is required for the assembly of the final Cas9-crRNA-tracrRNA complex. While elements from tracrRNA and crRNA have been adapted to form a single artificially linked sgRNA (hereafter referred to as the wild-type sp85 scaffold) (Jinek Science, HSU), the effects of sgRNA scaffold modification and tolerance towards mutagenesis has in general not been comprehensively studied.

[0941] The sgRNA scaffold can be functionally and structurally divided into several components. The crRNA portion includes the guide sequence and the direct repeat regions. The tracrRNA begins with a 14-nt anti-repeat that partially basepairs with the direct repeat to form a stem loop (stem loop 1), and further contains an 18-bp linker to two additional stem loops (stem loop 2 and 3). Importantly, there are several unpaired bases within the direct repeat and tracrRNA anti-repeat stem loop 1, which create a bulge separating the proximal and distal direct repeat regions (Figure 45A). Applicants hypothesized that optimization of the sgRNA architecture could improve the genome editing activity of Cas9 and subsequently performed a systematic interrogation of the sgRNA scaffold to gain a better functional understanding of each component.

[0942] Applicants first identified regions of the sgRNA likely important for binding and recognition by Cas9. Strikingly, replacement of the stem loop 1 bulge with perfectly base-pairing sequences completely abolished Cas9-mediated indel activity, while substituting other non-base pairing nucleotides and thus retaining the bulge structure still allowed maintenance of modest activity. Within stem loop 1, mutations in the proximal direct repeat was not uniformly tolerated: whereas shortening the proximal direct repeat duplex or mutating the poly-T tract to mixed pyrimidines and purines abolished Cas9 activity, mutating the poly-T tract to pyrimidines alone was well-tolerated (Figure 45B). Finally, truncation, shuffling, or randomization of the 18-bp linker sequence likewise resulted in complete loss of activity. However, it is possible that this longer linker forms additional secondary structures not predicted by RNA-folding, and further finer mapping mutagenesis experiments will be needed to elucidate its structural role.

[0943] Consistent with previous studies showing that stem loops 2 and 3 are not critical for Cas9 activity even though they significantly improve cleavage efficiency, alterations of the distal hairpins are largely well tolerated. For instance, both stem loops 2 and 3 could be largely replaced with G-C basepairs or extended in length without adversely affecting activity (Figure 45B). Together, these findings suggest that while the proximal direct repeat, bulge, and linker may be involved in Cas9 recognition and binding, the two distal hairpins are likely more important for sgRNA folding and stability. Indeed, simultaneous stabilization of both distal hairpins along with mutating the distal direct repeat region was well tolerated, yielding indel activity comparable to the original scaffold (Figure 46).

[1000] U6-driven tandem guide RNAs are able to deliver two functional sgRNAs

[1001] The programmable nature of the CRISPR-Cas system by small RNAs makes it inherently more tractable than purely protein module-based tools such as ZFNs and TALENs for applications requiring multiplex targeting. Indeed, Applicants and others have already shown that this can be readily achieved by co-delivering multiple sgRNAs in a variety of applications. While this approach works well for in vitro studies, in vivo or therapeutic applications would benefit from using a single vector system such as AAV. One of the major limitations of such systems is the amount genetic information that can be delivered (~4.8kb for AAV), above which the efficiency of AAV particle assembly rapidly declines. Furthermore, the alternative approach of using pooled delivery of independently transcribed sgRNAs is stochastic in nature and less reproducible than a single vector system, especially 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.

[1002] Applicants hypothesized that structurally stable sgRNA scaffolds would be more likely to fold into independent, functionally active units when multiple units are transcribed together in the same transcript. To test this, Applicants began by inserting an 8-nt linker between tandem adjacent sgRNAs (Figure 47A); for each the invariant sgRNA scaffold (non-guide region), we used either pairs of original sp85 sgRNA or scaffolds with stabilized distal hairpins (4558 and 4561). Strikingly, Applicants observed that when the tsgRNAs targeted closely approximated genomic loci previously shown to induce indels with Cas9 nickase, the stabilized scaffolds 4558 and 4561 were able to induce indels at frequencies similar to those induced by co-transfected individual sgRNAs (Figure 47B). 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 (Figure 47C).

[1003] Optimization of tandem sgRNA scaffold architecture

[1004] 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, we also tested linkers that encoded for either half of a direct repeat or a full-length direct repeat. Interestingly, Applicants observed 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 (Figure 48). However, it appeared that inclusion of direct repeat sequences adversely affected activity while there's a modest preference towards 12-nt linker length for cleavage efficiency, although more studies are needed to confirm these observations.

[1005] Processing of tandem sgRNAS into individual subunits occurs, is position-dependent

[1006] An obvious question to transcribing multiple sgRNAs under the same promoter is whether or not the co-transcribed tandem sgRNAs are processed to individual guide-scaffold units. To answer this, Applicants designed tandem sgRNAs that carried the same guide in either the first or second position (Figure 49A). 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 49B).

[1007] 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. However, when placed in the second position, there were only trace amounts of fully processed sgRNA present. Consistent with this, Applicants observed that reversing spacer order in microdeletion assays could significantly alter the efficiency of genomic modification (data not shown). Furthermore, when testing other pairs of sgRNAs targeting different genomic loci, Applicants observed that the same guide sequence typically has better activity when placed in the first rather than the second position (Figure 49C). 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.

[1008] Pairing of sequence-divergent scaffolds results in better second spacer activity

[1009] 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 can assess indel activity by measuring the fluorescence fraction and intensity of transfected cells (Figure 50A). Applicants observed that 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 50B).

[1010] Given that each sgRNA scaffold needs to fold into a stable secondary structure, Applicants hypothesized that 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. Applicants surmised that 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. To test this hypothesis, 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 50C). 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 51). In summary, 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 will likely aid processing and activity of tandem sgRNAs.once with PBS, trypsinized and triturated to single cell suspension, and re-suspended in PBS buffer supplemented with 5% FBS and 2mM EDTA. Fluorescent intensity was subsequently measured using the Accuri C6 flow cytometer.

[1011] Discussion

[1012] Double nicking approach to genome editing with CRISPR

[1013] Specificity is of paramount importance when introducing permanent genomic alterations, especially for highly sensitive applications such as gene therapy or studies aimed at linking causal genetic variants with biological processes or disease phenotypes. Designer nucleases such as ZFNs62 and TALENs63 have reported off-target activities over 15%. Given that both approaches are based on complex, evolved protein-DNA interactions, prediction or optimization of specificity through protein engineering can prove quite challenging. Nonetheless, efforts have been made to increase TALEN specificity, such as extending the number of bases recognized by protein monomers.

[1014] Strategies for improving the targeting accuracy of the CRISPR-Cas system can optimize either of its two essential components - the Cas9 nuclease or sgRNA. While Applicants' work has shown that extending the guide length does not improve specificity, it has

recently been reported that shorter guide lengths could potentially significantly decrease non-specific activity at known off-target sites. However, given that shorter guide sequence also increases the number of possible similar targets across a genome, it remains to be seen whether this strategy will decrease overall genome-wide off-target mutation frequencies. Here, Applicants have demonstrated that combining two sgRNAs with Cas9 nickases is able to effectively generate DSBs while avoiding mutagenic events arising from single-stranded DNA break mutations as they are typically repaired with high fidelity.

[1015] In the context of delivering gene repair or replacement templates, Cas9n nicking of DNA with a single sgRNA has been previously shown to facilitate HDR without generating indels. However, it is substantially less efficient at doing so relative to wild-type Cas9, and can be susceptible to differences in HDR efficiency among different cell types48,52. However, Applicants have demonstrated that using two closely approximated guides to target the Cas9n nickase to the same genomic locus can mediate HDR at high efficiencies while keeping off-target modifications to background levels. Moreover, the characterization of spacing parameters governing successful Cas9 double nickase-mediated gene targeting reveals an effective window of over 100-bp in which sgRNAs targeting opposite strands can be paired for double-nicking applications, allowing for a high degree of flexibility in their design. Applicants have additionally demonstrated that double nicking-mediated indel frequencies are comparable to those of wild-type Cas9 modification at multiple loci in both human and mouse cells, confirming the reproducibility of this strategy for high-precision genome engineering.

[1016] Though the ability to potentiate specific, targeted indel mutations greatly enables functional analyses by gene knock-out, the use of double-nicking to precisely target homologous recombination has practical implications in the generation of model systems and organisms. It has been reported that blastocyst injection of Cas9 nuclease with sgRNA and HDR template can generate conditional and reporter mice in a single step. While this finding immensely streamlines an otherwise laborious and prolonged process, the relatively high dose of Cas9 mRNA and guide RNA injected into each blastocyst can become a real concern for off-target modifications. Indeed, concurrent work by collaborators and other members of the lab has already demonstrated that analogous blastocyst delivery of Cas9n with two sgRNAs can induce efficient targeted gene modification at the mouse Mecp2 locus. Further studies investigating the efficiency and

specificity of the double-nicking approach in facilitating homologous recombination in the context of mouse model generation will be immensely informative.

[1017] While significant off-target mutagenesis has been previously reported for Cas9 nucleases in human cells, the double-nicking approach provides a generalizable solution for rapid and accurate genome editing. Even though double-nicking is conceptually similar to ZFN- and TALEN-based genome editing systems, which utilize hemi-nuclease domains to induce DSBs, the ease, flexibility, and improved predictably of using an RNA-guided DNA targeting system significantly increases its potential downstream applications. Given that it has been observed that cooperative nicking at off-target sites can still occur in the context of ZFNs and TALENs, significant and thorough characterization of the true genome-wide off-target activity of the CRISPR-Cas system is still prerequisite to its further development as means of efficient, ultrahigh-precision genome editing. Even so, Applicants believe that double nicking with Cas9n represents a solid step-forward in establishing CRISPR-Cas system as a versatile tool for genome manipulation in both basic science research and medicine.

[1018] sgRNA optimization and creation of tandem guide RNAs

[1019] Following the initial derivation of the chimeric sgRNA41 from elements of tracrRNA and crRNA, the subsequent sp85 sgRNA scaffold40,48 was developed from full-length tracrRNA and has become the most commonly architecture used for genome editing applications. However, aside from relatively few studies aimed at improving sgRNA stability, there has not yet been any reported fine mapping of sgRNA structure-function relationships or optimization of the sgRNA architecture through sequence replacement. The targeted, functional studies Applicants performed have identified a number of regions within the sgRNA that may be amenable to further modification or addition of functional groups that may broaden the range of applications for the CRISPR-Cas9 system. Notably, despite having tested a wide-range of scaffold modifications, Applicants observed few changes that significantly improve the indel activity of Cas9. Further elucidation of the structure-function relationship of the sgRNA interacting with its nuclease will be informative in making more targeted changes to both the sgRNA and Cas9 simultaneously that may allow further gains in on-target efficiency.

[1020] The ribonuclease RNAseIII has been shown as necessary for processing and maturation of the crRNA following binding to tracrRNA in the type II CRISPR systems. However, processing at the 5' end remains largely unknown in both microbial and eukaryotic

contexts. A recent report investigating processing of Neisseria spp. CRISPR RNA processing identified transcriptional promoters located in the direct repeat preceding each spacer sequence that drives transcription of individual crRNA units. Applicants have previously observed that lengthening of the spacer sequence to 30-nt does not result in a longer sgRNA: Northern blot analysis shows still the same length transcript as with sgRNAs with 20-nt guide sequences. Furthermore, Applicants' observation that sgRNA units can be fully processed – albeit at low levels – from the second position of the tsgRNA might point to the existence of potential endonucleases involved with end-maturation of sgRNAs.

[1021] RNA-sequencing analyses have shown that spacers located at the promoter-proximal end of the CRISPR arrays tend to be of higher abundance than those located more distally, suggesting that transcriptional processivity may be an important parameter in determining relative efficiency of mature crRNA units. Consistent with Applicants' findings, this study also reports that certain spacer sequences predicted to form secondary interactions with adjacent RNA are often under-represented. Thus, while developing synthetic CRISPR arrays, consideration of spacer sequence is likely to become of increasing importance.

[1022] There have been a number of recent studies using CRISPR-Cas9 lentiviral libraries for genome-wide knockout screens that have shown greater reproducibility and sensitivity than analogous RNAi-libraries. The ability to simultaneously deliver more than a single guide RNA on a single vector would be particularly interesting in the context of lentiviral screening methodologies, which would open the door for both high-throughput deletion as well as pairwise screens. In the context of commonly used shRNA libraries, the propensity of lentivirus to recombine has limited our ability to drive expression of multiple short RNAs from a single vector with the requirement of utilizing multiple unique promoters. Although it remains to be seen how many sgRNAs can be efficiently arrayed in tandem, this approach allows for a single promoter to drive expression of at least two sgRNAs. Furthermore, the knowledge that sequence-divergent yet structurally similar sgRNA scaffolds can remain active will be useful in a variety of applications where recombination between structural elements has been a limitation.

Table D. List of sgRNA pairs to identify optimal spacing.

Gene	Overhang Length (bp)	Overhang Type	Offset Length (bp)	Left sgRNA ID	Right sgRNA ID	Cas9n with left and right sgRNA indel (%)	left sgRNA target site		left sgRNA with wildtype Cas9 indel (%)	right sgRNA target site		right sgRNA with wildtype Cas9 indel (%)
							guide sequence (5' to 3')	PAM		guide sequence (5' to 3')	PAM	
EMX1	148	3'	-182	15	4	N.D.	TGCGCCACCGGTTGATGTGA	TGG	13.15	AGGCCCCAGTGGCTGCTCTG	GGG	27.29
EMX1	101	3'	-135	23	1	N.D.	ACTCTTGCCTCTTGTGGGTTTG	TGG	24.7	GAGTCCGAGCAGAAGAAGAA	GGG	21.9
EMX1	48	3'	-82	23	17	N.D.	ACTCTTGCCTCTTGTGGGTTTG	TGG	24.7	CACGAAGCAGGCCAATGGGG	AGG	13.57
EMX1	25	3'	-59	10	13	N.D.	CAAACGGCAGAAGCTGGAGG	AGG	26.15	GGAGCCCTTCTTCTTCTGCT	CGG	33.17
EMX1	15	3'	-49	4	5	N.D.	AGGCCCCAGTGGCTGCTCTG	GGG	27.29	GGGGCACAGAATGAGAAACTC	AGG	26.56
EMX1	8	3'	-42	7	9	N.D.	TGAAGGTGTGGTTCCAGAAC	CGG	36.02	GCCGTTTGTACTTTGTCCTC	CGG	30.49
EMX1	26	5'	-8	9	19	13.7 ± 1.27	GCCGTTTGTACTTTGTCCTC	CGG	9.82	GGCAGAGTGCTGCTTGCTGC	TGG	26.15
EMX1	30	5'	-4	9	10	19.72 ± 0.32	GCCGTTTGTACTTTGTCCTC	CGG	30.49	CAAACGGCAGAAGCTGGAGG	AGG	22.06
EMX1	31	5'	-3	6	7	21.35 ± 2.23	TCACCTGGGCCAGGGAGGGA	GGG	10.75	TGAAGGTGTGGTTCCAGAAC	CGG	36.02
EMX1	34	5'	0	15	16	26.89 ± 1.54	TGCGCCACCGGTTGATGTGA	TGG	13.15	TTGCCACGAAGCAGGCCAAT	GGG	13.77
EMX1	38	5'	4	15	17	36.31 ± 2.97	TGCGCCACCGGTTGATGTGA	TGG	14.49	CACGAAGCAGGCCAATGGGG	AGG	13.57
EMX1	51	5'	17	5	7	31.12 ± 0.25	GGGGCACAGATGAGAAACTC	AGG	26.56	TGAAGGTGTGGTTCCAGAAC	CGG	36.02
EMX1	54	5'	20	5	8	32.41 ± 3.68	GGGGCACAGATGAGAAACTC	AGG	26.56	AGGTGTGGTTCCAGAAGCTG	AGG	35.53
EMX1	65	5'	31	6	10	13.45 ± 1.99	TCACCTGGGCCAGGGAGGGA	GGG	10.75	CAAACGGCAGAAGCTGGAGG	AGG	26.15
EMX1	69	5'	35	6	11	12.39 ± 1.29	TCACCTGGGCCAGGGAGGGA	GGG	10.75	CGGCAGAAGCTGGAGGAGGA	AGG	22.06
EMX1	76	5'	42	9	14	21.71 ± 1.66	GCCGTTTGTACTTTGTCCTC	CGG	30.49	AGGGCTCCCATCACATCAAC	CGG	41.27
EMX1	85	5'	51	5	10	21.89 ± 1.88	GGGGCACAGATGAGAAACTC	AGG	26.56	CAAACGGCAGAAGCTGGAGG	AGG	26.15
EMX1	95	5'	61	6	12	5.88 ± 1.81	TCACCTGGGCCAGGGAGGGA	GGG	10.75	TGAGTCCGAGCAGAAGAAGA	AGG	29.06
EMX1	135	5'	101	5	14	15.78 ± 2.19	GGGGCACAGATGAGAAACTC	AGG	26.56	AGGGCTCCCATCACATCAAC	CGG	41.27
EMX1	145	5'	111	6	16	N.D.	TCACCTGGGCCAGGGAGGGA	GGG	10.75	TTGCCACGAAGCAGGCCAAT	GGG	13.77
EMX1	181	5'	147	6	18	N.D.	TCACCTGGGCCAGGGAGGGA	GGG	10.75	TCACCTCCAATGACTAGGGT	GGG	25.14
EMX1	201	5'	167	5	18	N.D.	GGGGCACACATGAGAAACTC	AGG	26.56	TCACCTCCAATGACTAGGGT	GGG	25.14
EMX1	222	5'	188	6	19	N.D.	TCACCTGGGCCAGGGAGGGA	GGG	10.75	GGCAGAGGTGCTGCTTGCTGC	TGG	10.75
EMX1	242	5'	208	5	19	N.D.	GGGGCACAGATGAGAAACTC	AGG	26.56	GGCAGAGGTGCTGCTTGCTGC	TGG	17.22
DYRK1A	164	3'	-198	34	47	N.D.	ATCTGGTCAGAATATGATAA	AGG	10.65 ± 2.65	AACCTCACTTATCTTCTTGT	AGG	19.02 ± 4.32
DYRK1A	105	3'	-139	35	47	N.D.	GTCACTGTACTGATGTGAAT	TGG	16.71 ± 2.47	AACCTCACTTATCTTCTTGT	AGG	17.04 ± 1.30
DYRK1A	66	3'	-100	36	48	N.D.	CATCTGAAGGCCAGCAGCAT	TGG	8.82 ± 1.01	CTCACTTATCTTCTTGTAGG	AGG	18.79 ± 2.71
DYRK1A	25	3'	-59	35	49	N.D.	GTCACTGTACTGATGTGAAT	TGG	17.83 ± 0.43	CCATGCTGCCTGGCTTCAGA	TGG	17.15 ± 3.29
DYRK1A	4	3'	-38	37	31	N.D.	TGATAAGGCAGAAACCTGTT	TGG	4.95 ± 0.66	GCCAAACATAAGTGACCAAC	AGG	16.38 ± 3.39
DYRK1A	17	5'	-17	38	47	N.D.	GAAGATAAGTGAGGTTTAAA	AGG	5.30 ± 1.98	AACCTCACTTATCTTCTTGT	AGG	24.18 ± 3.22
DYRK1A	21	5'	-13	39	33	10.54 ± 0.63	GTATCATTTGACATATCTAA	TGG	26.90 ± 1.17	TGTCAAATGATACAAACATT	AGG	29.69 ± 0.86
DYRK1A	25	5'	-9	40	49	2.33 ± 0.11	CAGCATGGAAGAAAAATGAC	CGG	3.33 ± 0.56	CCATGCTGCCTGGCTTCAGA	TGG	20.43 ± 2.40
DYRK1A	29	5'	-5	41	50	27.76 ± 0.84	GCAGCATGGAATGAAAAATGA	CGG	17.84 ± 5.46	GCTGCTGGCCTTCAGATGGC	TGG	21.92 ± 3.46
DYRK1A	33	5'	-1	34	51	10.42 ± 0.90	ATCTGGTCAGAATATGATAA	AGG	9.13 ± 2.32	TCAGCAACCTCTTAACTAACC	AGG	24.14 ± 2.95
DYRK1A	35	5'	1	42	52	7.63 ± 0.51	GTGCAAGCCGAACAGATGAA	AGG	6.65 ± 2.19	TCATTTTCATTCCATGCTGC	TGG	20.61 ± 3.64

N.D.: Not detected.

Gene	Overhang Length (bp)	Overhang Type	Offset Length (bp)	Left sgRNA ID	Right sgRNA ID	Cas9n with left and right sgRNA indel (%)	left sgRNA target site		left sgRNA	right sgRNA target site		right sgRNA
							guide sequence (5' to 3')	PAM	with wildtype Cas9 indel (%)	guide sequence (5' to 3')	PAM	with wildtype Cas9 indel (%)
DYRK1A	36	5'	2	28	29	38.46 ± 0.74	GGAGTATCAGAAATGACTAT	TGG	20.88 ± 9.09	GGAGTAECAGAAATGACTAT	TGG	30.3 ± 0.7
DYRK1A	41	5'	7	30	31	34.41 ± 0.87	GGTCACTGTACTGATGTGAA	TGG	25.68 ± 5.95	GCCAAACATAAGTGACCAAC	AGG	33.1 ± 0.4
DYRK1A	42	5'	8	43	31	38.36 ± 0.32	TCACTGTACTGATGTGAATG	GGG	24.68 ± 4.58	GCCAAACATAAGTGACCAAC	AGG	29.46 ± 3.30
DYRK1A	43	5'	9	32	33	28.97 ± 0.32	GTTCCTTAAATAAGAACTTT	AGG	23.60 ± 2.56	TGTCAAATGATACAAACATT	AGG	22.4 ± 1.6
DYRK1A	46	5'	12	44	53	11.90 ± 1.65	TCCTACAAGAAGATAAGTGA	AGG	6.57 ± 1.36	CATGCAAACCTTCATCTGTT	CGG	30.42 ± 1.14
DYRK1A	77	5'	43	36	31	6.63 ± 0.27	CATCTGAAGGCCAGCAGCAT	TGG	10.02 ± 1.17	GCCAAACATAAGTGACCAAC	AGG	22.92 ± 5.16
DYRK1A	86	5'	52	38	52	N.D.	GAAGATAAGTGAGGTTTAAA	AGG	2.90 ± 0.82	TCATTTCAATCCATGCTGC	TGG	17.30 ± 1.62
DYRK1A	97	5'	63	38	49	N.D.	GAAGATAAGTGAGGTTTAAA	AGG	2.16 ± 0.48	CCATGCTGCTGGCCTTCAGA	TGG	24.75 ± 2.56
DYRK1A	131	5'	97	45	52	N.D.	TATCATTTGACATATCTAAT	TGG	8.21 ± 2.83	TCATTTCAATCCATGCTGC	TGG	14.61 ± 4.10
DYRK1A	155	5'	121	44	31	N.D.	TCCTACAAGAAGATAAGTGA	AGG	9.99 ± 4.12	GCCAAACATAAGTGACCAAC	AGG	19.74 ± 2.91
DYRK1A	191	5'	157	46	52	N.D.	AACTTTTCTAACTACAAACA	AGG	5.74 ± 2.24	TCATTTCAATCCATGCTGC	TGG	21.37
GRIN2B	165	3'	-199	70	82	N.D.	CCAACACCAACCAGAACTTG	GGG	2.95 ± 0.21	CTGGTAGATGGAGTTGGGTT	TGG	17.25 ± 1.30
GRIN2B	67	3'	-101	71	83	N.D.	ACAGCAATGCCAATGCTGGG	GGG	18.00 ± 2.31	AGTGCTGTTCTCCCAAGTTC	TGG	28.64 ± 0.69
GRIN2B	42	3'	-76	72	84	N.D.	GTGGAAATCATCTTCTCGT	TGG	14.56 ± 7.84	GGCATTGCTGTCATCCTCGT	GGG	21.26 ± 2.68
GRIN2B	16	3'	-50	73	85	N.D.	TCTGCTGCCTGACACGGCCA	AGG	4.24 ± 0.79	TCCCAAGTTCTGGTTGGTGT	TGG	19.64 ± 0.23
GRIN2B	2	3'	-36	74	86	N.D.	CGAGCTCTGCTGCCTGACAC	CGG	2.99 ± 0.31	TTGGCCGTCCTGGCCGTGTC	AGG	4.74 ± 0.15
GRIN2B	9	5'	-25	75	87	1.04 ± 0.53	TCCTTGATGGCCACCTCGTC	CGG	2.25 ± 1.08	TTCCGACGAGGTGGCCATCA	AGG	17.13 ± 2.90
GRIN2B	18	5'	-16	76	88	5.93 ± 1.25	ATGACAGCAATGCCAATGCT	TGG	16.46 ± 2.28	TGGCATTGCTGTCATCCTCG	TGG	16.35 ± 1.25
GRIN2B	23	5'	-11	77	88	2.28 ± 0.34	AGCAATGCCAATGCTGGGGG	GGG	3.19 ± 0.51	TGGCATTGCTGTCATCCTCG	TGG	15.17 ± 2.02
GRIN2B	28	5'	-6	78	86	1.45 ± 0.12	GCCAACACCAACCAGAACTT	TGG	17.80 ± 2.30	TTGGCCGTCCTGGCCGTGTC	AGG	4.46 ± 1.35
GRIN2B	30	5'	-4	69	85	11.80 ± 0.29	GGAGAACAGCACTCCGCTCT	TGG	21.80 ± 1.40	TCCCAAGTTCTGGTTGGTGT	TGG	21.33 ± 0.63
GRIN2B	33	5'	-1	76	65	24.24 ± 0.23	ATGACAGCAATGCCAATGCT	TGG	19.48 ± 1.88	CCTCGTGGGGCACTTCCGACG	AGG	21.19 ± 3.42
GRIN2B	34	5'	0	79	65	20.83 ± 0.95	TGACAGCAATGCCAATGCTG	GGG	21.44 ± 3.02	CCTCGTGGGGCACTTCCGACG	AGG	24.11 ± 0.14
GRIN2B	36	5'	2	58	59	31.76 ± 1.00	CCGGCCAAGAACTTGAAGCC	AGG	32.50 ± 0.50	CTGGTTGTAGGATTTGAGTT	AGG	26.7 ± 2.9
GRIN2B	38	5'	4	54	55	34.45 ± 0.45	TATTACAGAATGGAGAGACTG	TGG	30.90 ± 1.40	TTATTTCTGAAAGAAATATTAA	AGG	27.6 ± 2.5
GRIN2B	38	5'	4	56	57	44.22 ± 0.55	AAAAGACCTAAACAAAAGAA	TGG	23.20 ± 2.10	TGTGTGAGGATAAAAGAGATT	GGG	29.4 ± 2.7
GRIN2B	38	5'	4	77	65	9.60 ± 0.25	AGCAATGCCAATGCTGGGGG	GGG	4.19 ± 0.58	CCTCGTGGGCACTTCCGACG	AGG	21.78 ± 1.70
GRIN2B	40	5'	6	60	61	42.54 ± 1.39	TCAGAGCTTCCTGACACCCA	TGG	14.20 ± 1.50	AATACCTAGTTACAGGCATT	TGG	24.8 ± 1.0
GRIN2B	45	5'	11	76	87	18.96 ± 0.93	ATGACAGCAATGCCAATGCT	TGG	20.45 ± 0.98	TTCCGACGAGGTGGCCATCA	AGG	13.21 ± 0.74
GRIN2B	50	5'	16	77	87	5.33 ± 0.57	AGCAATGCCAATGCTGGGGG	GGG	4.93 ± 2.06	TTCCGACGAGGTGGCCATCA	AGG	12.51 ± 1.21
GRIN2B	89	5'	55	80	89	7.31 ± 0.83	GAGAACAGCACTCCGCTCTG	GGG	3.09 ± 0.54	CAGAAGAGCCCCCCCAGCAT	TGG	25.02 ± 1.86
GRIN2B	105	5'	71	78	90	10.56 ± 1.21	GCCAACACCAACCAGAACTT	TGG	25.29 ± 1.65	CGTGGGCACTTCCGACGAGG	TGG	23.32 ± 0.78
GRIN2B	132	5'	98	81	67	2.66 ± 0.89	CTGCCTGACACGGCCAGGAC	CGG	4.34 ± 0.62	TGATTTCCACCATCTCTCCG	TGG	20.95 ± 0.79
GRIN2B	175	5'	141	80	67	N.D.	GAGAACAGCACTCCGCTCTG	GGG	2.96 ± 0.93	TGATTTCCACCATCTCTCCG	TGG	19.77 ± 2.20
GRIN2B	231	5'	197	81	91	N.D.	CTGCCTGACACGGCCAGGAC	CGG	6.17 ± 2.09	TGACCGGAAGATCCAGGGGG	TGG	23.36 ± 2.34

N.D.: Not detected.

Table E: List of sgRNAs used in this study

gene	sgRNA ID	guide sequence (5' to 3')	PAM	strand	species
EMX1	1	GAGTCCGAGCAGAAGAACAA	GGG	+	*H. sapiens*
EMX1	2	GGAAGGGCCTGAGTCCGAGCAGAAGAAGA	GGG	+	*H. sapiens*
EMX1	3	GGCCTCCAAGGAGTCCGAGCAGAAGAAGAA	GGG	+	*H. sapiens*
EMX1	4	AGGCCCCAGTGGCTGCTCTG	GGG	+	*H. sapiens*
EMX1	5	GGGGCACAGATGAGAAACTC	AGG	-	*H. sapiens*
EMX1	6	TCACCTGGGCCAGGGAGGGA	GGG	-	*H. sapiens*
EMX1	7	TGAAGGTGTGGTTCCAGAAC	CGG	+	*H. sapiens*
EMX1	8	AGGTGTGGTTCCAGAACCGG	AGG	+	*H. sapiens*
EMX1	9	GCCGTTTGTACTTTGTCCTC	CGG	-	*H. sapiens*
EMX1	10	CAAACGGCAGAAGCTGGAGG	AGG	+	*H. sapiens*
EMX1	11	CGGCAGAAGCTGGAGCAGGA	AGG	+	*H. sapiens*
EMX1	12	TGAGTCCGAGCAGAAGAAGA	AGG	+	*H. sapiens*
EMX1	13	GGAGCCCTTCTTCTTCTGCT	CGG	-	*H. sapiens*
EMX1	14	AGGGCTCCCATCACATCAAC	CGG	+	*H. sapiens*
EMX1	15	TGCGCCACCGGTTGATGTGA	TGG	-	*H. sapiens*
EMX1	16	TTGCCACGAAGCAGGCCAAT	GGG	+	*H. sapiens*
EMX1	17	CACGAAGCAGGCCAATGGGG	AGG	+	*H. sapiens*
EMX1	18	TCACCTCCAATGACTAGGGT	GGG	+	*H. sapiens*
EMX1	19	GGCAGAGTGCTGCTTGCTGC	TGG	+	*H. sapiens*
EMX1	20	GACATCGATGTCCTCCCCAT	TGG	-	*H. sapiens*
EMX1	21	GTCACCTCCAATGACTAGGG	TGG	+	*H. sapiens*
EMX1	22	GGCCAACCACAAACCCACGA	GGG	+	*H. sapiens*
EMX1	23	ACTCTGCCCTCGTGGGTTTG	TGG	-	*H. sapiens*
EMX1	24	CAAGCAGCACTCTGCCCTCG	TGG	-	*H. sapiens*
EMX1	25	TTCTTCTTCTGCTCGGACTC	AGG	-	*H. sapiens*
EMX1	26	CTCCCCATTGGCCTGCTTCG	AGG	-	*H. sapiens*
EMX1	27	GTCACCTCCAATGACTAGGG	TGG	+	*H. sapiens*
DYRK1A	28	GAACTTACCTCGTTACTTAC	AGG	-	*H. sapiens*
DYRK1A	29	GGAGTATCAGAAATGACTAT	TGG	+	*H. sapiens*
DYRK1A	30	GGTCACTGTACTGATGTGAA	TGG	-	*H. sapiens*
DYRK1A	31	GCCAAACATAAGTGACCAAC	AGG	+	*H. sapiens*
DYRK1A	32	GTTCCTTAAATAACAACTTT	AGG	-	*H. sapiens*
DYRK1A	33	TGTCAAATGATACAAACATT	AGG	+	*H. sapiens*
DYRK1A	34	ATCTGGTCAGAATATGATAA	AGG	-	*H. sapiens*
DYRK1A	35	GTCACTGTACTGATGTGAAT	TGG	-	*H. sapiens*
DYRK1A	36	CATCTGAAGGCCAGCAGCAT	TGG	-	*H. sapiens*
DYRK1A	37	TGATAAGGCAGAAACCTGTT	TGG	-	*H. sapiens*
DYRK1A	38	GAAGATAAGTGAGGTTTAAA	AGG	-	*H. sapiens*
DYRK1A	39	GTATCATTTGACATATCTAA	TGG	-	*H. sapiens*
DYRK1A	40	CAGCATGGAATGAAAATGAC	CGG	-	*H. sapiens*
DYRK1A	41	GCAGCATGGAATGAAAATGA	CGG	-	*H. sapiens*
DYRK1A	42	GTCCAAGCCCAACACATCAA	AGG	-	*H. sapiens*
DYRK1A	43	TCACTGTACTGATGTGAATG	GGG	-	*H. sapiens*
DYRK1A	44	TCCTACAAGAAGATAAGTGA	AGG	-	*H. sapiens*
DYRK1A	45	TATCATTTGACATATCTAAT	TGG	-	*H. sapiens*
DYRK1A	46	AACTTTTCTAACTACAAACA	AGG	-	*H. sapiens*

DYRK1A	47	AACCTCACTTATCTTCTTGT	AGG	+	*H. sapiens*
DYRK1A	48	CTCACTTATCTTCTTGTAGG	AGG	+	*H. sapiens*
DYRK1A	49	CCATGCTGCTGGCCTTCAGA	TGG	+	*H. sapiens*
DYRK1A	50	GCTGCTGGCCTTCAGATGGC	TGG	+	*H. sapiens*
DYRK1A	51	TCAGCAACCTCTAACTAACC	AGG	+	*H. sapiens*
DYRK1A	52	TCATTTTCATTCCATGCTGC	TGG	+	*H. sapiens*
DYRK1A	53	CATGCAAACCTTCATCTGTT	CGG	+	*H. sapiens*
DYRK1A	54	TATTACAGAATGAGAGACTG	TGG	-	*H. sapiens*
DYRK1A	55	TTATTTCTGAAGAATATTAA	AGG	+	*H. sapiens*
DYRK1A	56	AAAAGACCTAAACAAAAGAA	TGG	-	*H. sapiens*
DYRK1A	57	TGTGTGAGGATAAAAGAGTT	GGG	+	*H. sapiens*
DYRK1A	58	CCGGCCAAGACCTTGAAGCC	AGG	-	*H. sapiens*
DYRK1A	59	CTGGTTGTAGCATTTGAGTT	AGG	+	*H. sapiens*
DYRK1A	60	TCAGAGCTTCCTGACACCCA	TGG	-	*H. sapiens*
DYRK1A	61	AATACCTAGTTACAGGCATT	TGG	+	*H. sapiens*
GRIN2B	62	GGTGATGATGCTCTTTGGGT	CGG	-	*H. sapiens*
GRIN2B	63	TCTGTGATCTCATGTCTGAC	CGG	+	*H. sapiens*
GRIN2B	64	CAGCAATGCCAATGCTGGGG	GGG	-	*H. sapiens*
GRIN2B	65	CCTCGTGGGCACTTCCGACG	AGG	+	*H. sapiens*
GRIN2B	66	TTTCTCGTGGGCATCCTTGA	TGG	-	*H. sapiens*
GRIN2B	67	TGATTTCCACCATCTCTCCG	TGG	+	*H. sapiens*
GRIN2B	68	GGAGAACAGCACTCCGCTCT	GGG	-	*H. sapiens*
GRIN2B	69	CTGGTTGGTGTTGGCCGTCC	TGG	+	*H. sapiens*
GRIN2B	70	CCAACACCAACCAGAACTTC	CGG	-	*H. sapiens*
GRIN2B	71	ACAGCAATGCCAATGCTGGG	GGG	-	*H. sapiens*
GRIN2B	72	CTCGAAATCATCTTTCTCGT	TGG	-	*H. sapiens*
GRIN2B	73	TCTGCTGCCTGACACGGCCA	AGG	-	*H. sapiens*
GRIN2B	74	CGAGCTCTGCTGCCTGACAC	CGG	-	*H. sapiens*
GRIN2B	75	TCCTTGATGGCCACCTCGTC	CGG	-	*H. sapiens*
GRIN2B	76	ATGACAGCAATGCCAATGCT	TGG	-	*H. sapiens*
GRIN2B	77	AGCAATGCCAATGCTGGGGG	GGG	-	*H. sapiens*
GRIN2B	78	GCCAACACCAACCAGAACTT	TGG	-	*H. sapiens*
GRIN2B	79	TGACAGCAATGCCAATGCTG	GGG	-	*H. sapiens*
GRIN2B	80	GAGAACAGCACTCCGCTCTG	GGG	-	*H. sapiens*
GRIN2B	81	CTGCCTGACACGGCCAGGAC	CGG	-	*H. sapiens*
GRIN2B	82	CTGGTAGATGGAGTTGGGTT	TGG	+	*H. sapiens*
GRIN2B	83	AGTGCTGTTCTCCCAAGTTC	TGG	+	*H. sapiens*
GRIN2B	84	GGCATTGCTGTCATCCTCGT	CGG	+	*H. sapiens*
GRIN2B	85	TCCCAAGTTCTGGTTGGTGT	TGG	+	*H. sapiens*
GRIN2B	86	TTGGCCGTCCTGGCCGTGTC	AGG	+	*H. sapiens*
GRIN2B	87	TTCCGACGAGGTGGCCATCA	AGG	+	*H. sapiens*
GRIN2B	88	TGGCATTGCTGTCATCCTCG	TGG	+	*H. sapiens*
GRIN2B	89	CAGAAGAGCCCCCCCAGCAT	TGG	+	*H. sapiens*
GRIN2B	90	CGTGGGCACTTCCGACGAGG	TGG	+	*H. sapiens*
GRIN2B	91	TGACCGGAAGATCCAGGGGG	TGG	+	*H. sapiens*

'able F. List of sgRNA scaffolds used in this study

sgRNA Scaffold	Length	Sequence
Wild-type sp85	81	gttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttGAAAaagtggcaccgAGTcggtgcTTTTT
eplace bulge with G-C pair	79	gttttagCgctaGAAAtagcGttaaaataaggctagtccgttatcaacttGAAAaagtggcaccgAGTcggtgcTTTTT
Replace PDR with CGG trace, remove bulge	79	gCCGGCgCgctaGAAAtagcGtGCCGGtaaggctagtccgttatcaacttGAAAaagtggcaccgAGTcggtgcTTTTT
Replace PDR with CCGG tract, keep bulge	71	gCCGGCgCGAAAGtGCCGGtaaggctagtccgttatcaacttGAAAaagtggcaccgAGTcggtgcTTTTT
ıorten PDR, remove bulge	81	gCCGGCgagctaGAAAtagcaagtGCCGGtaaggctagtccgttatcaacttGAAAaagtggcaccgAGTcggtgcTTTTT
Replace PDR with CCCC tract	80	gccccagagctagaaatagcaagttggggtaaggctagtccgttatcaacttgaaaaagtggcaccgagtcggtgcttttt
ıge randomization	81	gttttagagctaGAAAtagcNNNNtaaaataaggctagtccgttatcaacttGAAAaagtggcaccgAGTcggtgcTTTTT
linker truncation, first half	72	gttttagagctaGAAAtagcaagttaaaataaggctagtacttGAAAaagtggcaccgAGTcggtgcTTTTT
.inker truncation, second half	63	gttttagagctaGAAAtagcaagttaaaatacttGAAAaagtggcaccgAGTcggtgcTTTTT
Linker scramble	81	gttttagagctaGAAAtagcaagttaaaatagatcgtatacatcattaacttGAAAaagtggcaccgAGTcggtgcTTTTT
Linker randomization	81	gttttagagctaGAAAtagcaagttaaaatNNNNNNNNNNNNNNNNNNacttGAAAaagtggcaccgAGTcggtgcTTTTT
eplace Stem 2 with G-C tract	81	gttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaCGCCGAAAGGCGggcaccgAGTcggtgcTTTTT
Extend Stem 2 by 2 bp	85	gttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttgaGAAAtcaagtggcaccgAGTcggtgcTTTTT
eplace Stem 3 with G-C tract	80	gttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgagcggcgggcttttt
Lengthen Stem 3 (add 3bp)	84	gttttagagctagaaatagcaagttaaaataaggctagtccgttatcaacttgaaaaagtggcaccgaaagtttcggtgcttttt
4558	82	gttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaCGCCGAAAGGCGggcaccgAGTcggtgccTTTTT
4561	92	gttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaccttgggccGAAAggcccaagggcgccgAGTcggcgccTTTT T
4638	82	gttttagagccgGAAAcggcaagttaaaataaggctagtccgttatcaCGCCGAAAGGCGggcaccgAGTcggtgccTTTTT
Opt1CCCC	81	gccccagCgctaGAAAtagcaagttggggtaaggctagtccgttatcaCGCCGAAAGGCGggcaccgAGTcggtgcTTTTT
Opt1CTTC	81	gcttcagCgctaGAAAtagcaagttgaagtaaggctagtccgttatcaCGCCGAAAGGCGggcaccgAGTcggtgcTTTTT
Scrmbl1	92	gttttagatgcaattatgcaaagttaaaataaggctagtccgttatcaacttatatcaataagtcatcgtcgagAAGctcgacgatgTTTT T
Scrmbl2	92	gttttagaactcattagagtaagttaaaataaggctagtccgttatcaacatgcatccagcatgttcaggattcaAACtgaatcctcaTTTT T
ShScrmbl1 PolyC	82	gccccagatgcaattatgcaaagttggggtaaggcaagtccgttatcaaccgaccacggtcacgtcgAAGcgacgtgTTTTT
ShScrmbl1 CTTC	82	gcttcagatgcaaattatgcaaagttgaagtaaggcaagtccgttatcacctgaccacagggtgtcgtAAGacgacacTTTTT
Shorter Scrmbl1 PolyC	77	gccccagaccattggaagttggggtaaggcaagtccgttatcaaccgaccacggtcacgtcgAAGcgacgtgTTTTT
Shorter Scrmbl1 CTTC	77	gcttcagagcattccaagttgaagtaaggcaagtccgttatcacctgaccacagggtgtcgtAAGacgacacTTTTT

Table G Primers used for SURVEYOR assays

Primer Name	Genomic Target	Sequence
SUV901	*EMX1*	CCATCCCCTTCTGTGAATGT
SUV902	*EMX1*	GGAGATTGGAGACACGGAGA
DYRK1A-F	*DYRK1A*	GGAGCTGGTCTGTTGGAGAA
DYRK1A-R	*DYRK1A*	TCCCAATCCATAATCCCACGTT
GRIN2B-F	*GRIN2B*	CAGGAGGGCCAGGAGATTTG
GRIN2B-R	*GRIN2B*	TGAAATCGAGGATCTGGGCG

Table H. Primers used to generate amplicons for NGS

Primer Name	Sequence
EMX1-F	GGAGGACAAAGTACAAACGGC
EMX1-R	ATCGATGTCCTCCCCATTGG
EMX1-HR-F	CCATCCCCTTCTGTGAATGT
EMX1-HR-R	GGAGATTGGAGACACGGAGA
EMX1-OT1.1-F	TGGGAGAGAGACCCCTTCTT
EMX1-OT1.1-R	TCCTGCTCTCACTTAGACTTTCTC
EMX1-OT1.2-F	GACATTCCTCCTGAGGGAAAA
EMX1-OT1.2-R	GATAAAATGTATTCCTTCTCACCATTC
EMX1-OT1.3-F	CCAGACTCAGTAAAGCCTGGA
EMX1-OT1.3-R	TGGCCCCAGTCTCTCTTCTA
EMX1-OT1.4-F	CACGGCCTTTGCAAATAGAG
EMX1-OT1.4-R	CATGACTTGGCCTTTGTAGGA

[0944] 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.

References

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

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

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

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

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

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

7. Porteus, M.H. & Baltimore, D. Chimeric nucleases stimulate gene targeting in human cells. Science 300, 763 (2003).

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

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

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

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

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

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

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

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

16. Moscou, M.J. & Bogdanove, A.J. A simple cipher governs DNA recognition by TAL effectors. Science 326, 1501 (2009).

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

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

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

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

21. 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).

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

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

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

25. 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).

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

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

28. 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).

29. 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).

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

31. 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).

32. Cong, L. et al. Multiplex Genome Engineering Using CRISPR-Cas Systems. Science 339, 819-823 (2013).

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

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

35. 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).

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

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

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

39. 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).

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

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

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

43. 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).

44. 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).

45. 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).

46. 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).

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

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

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

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

51. 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).

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

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

54. 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).

55. 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).

56. 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).

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

58. 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).

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

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

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

62. 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]

63. 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.

64. 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.

65. 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.

66. 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.

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

68. 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.

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

70. Capecchi, M. R. Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century. Nat Rev Genet 6, 507–512 (2005).

71. Bradley, A., Evans, M., Kaufman, M. H. & Robertson, E. Formation of germ-line chimaeras from embryo-derived teratocarcinoma cell lines. Nature 309, 255–256 (1984).

72. Smithies, R. O. & Koralewski, M. A. Insertion of DNA sequences into the human chromosomal B-globin locus by homologous recombination. Nature (1985).

73. Rudin, N., Sugarman, E. & Haber, J. E. Genetic and physical analysis of double-strand break repair and recombination in Saccharomyces cerevisiae. Genetics (1989).

74. Rouet, P., Smih, F. & Jasin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol Cell Biol (1994).

75. Choulika, A., Perrin, A. & Dujon, B. Induction of homologous recombination in mammalian chromosomes by using the I-SceI system of Saccharomyces cerevisiae. Molecular and cellular (1995).

76. Wyman, C. & Kanaar, R. DNA double-strand break repair: all's well that ends well. Annu. Rev. Genet. (2006).

77. Li, L., Wu, L. P. & Chandrasegaran, S. Functional domains in Fok I restriction endonuclease. in (1992).

78. Kim, Y. G. & Chandrasegaran, S. Chimeric restriction endonuclease. in (1994).

79. Miller, J., McLachlan, A. D. & Klug, A. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J (1985).

80. Pavletich, N. P. & Pabo, C. O. Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A. Science (1991).

81. Klug, A. The Discovery of Zinc Fingers and Their Applications in Gene Regulation and Genome Manipulation. Annu. Rev. Biochem. 79, 213–231 (2010).

82. Beerli, R. R. & Barbas, C. F. Engineering polydactyl zinc-finger transcription factors. Nat Biotech (2002).

83. Maeder, M. L., Thibodeau-Beganny, S., Osiak, A. & Wright, D. A. Rapid 'open-source' engineering of customized zinc-finger nucleases for highly efficient gene modification. Molecular Cell (2008).

84. Boch, J., Scholze, H., Schornack, S., Landgraf, A. & Hahn, S. Breaking the code of DNA binding specificity of TAL-type III effectors. Science (2009).

85. Gaj, T., Gersbach, C. A. & Barbas, C. F., III. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends in Biotechnology 1–9 (2013). doi:10.1016/j.tibtech.2013.04.004

86. Ding, Q. et al. A TALEN Genome-Editing System for Generating Human Stem Cell-Based Disease Models. Stem Cell 12, 238–251 (2013).

87. McManus, M. T. & Sharp, P. A. Gene silencing in mammals by small interfering RNAs. Nat Rev Genet 3, 737–747 (2002).

88. Montgomery, M. K., Kostas, S. A., Driver, S. E. & Mello, C. C. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. Nature (1998).

89. Vert, J.-P., Foveau, N., Lajaunie, C. & Vandenbrouck, Y. An accurate and interpretable model for siRNA efficacy prediction. BMC Bioinformatics 7, 520 (2006).

90. Bernstein, E., Caudy, A. A., Hammond, S. M. & Hannon, G. J. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature (2001).

91. Song, J., Giang, A., Lu, Y., Pang, S. & Chiu, R. Multiple shRNA expressing vector enhances efficiency of gene silencing. BMB Rep (2008).

92. van Parijs, L., Chen, J. & Sharp, P. A. Small interfering RNA-mediated gene silencing in T lymphocytes. The Journal of … (2002).

93. Grimm, D. & Kay, M. A. Combinatorial RNAi: A Winning Strategy for the Race Against Evolving Targets&quest. Mol Ther (2007).

94. Silva, J. M., Marran, K., Parker, J. S., Silva, J. & Golding, M. Profiling essential genes in human mammary cells by multiplex RNAi screening. Science (2008).

95. Stockwell, B. R., Hacohen, N., Hahn, W. C. & Lander, E. S. A lentiviral RNAi library for human and mouse genes applied to an arrayed viral high-content screen. Cell (2006).

96. Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. & Nakata, A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. Journal of Bacteriology 169, 5429–5433 (1987).

97. Spilman, M., Cocozaki, A., Hale, C., Shao, Y. & Ramia, N. Structure of an RNA silencing complex of the CRISPR-Cas immune system. Molecular Cell (2013).

98. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. Nature 482, 331–338 (2012).

99. Mojica, F. J. M., Díez-Villaseñor, C., García-Martínez, J. & Almendros, C. Short motif sequences determine the targets of the prokaryotic CRISPR defence system. Microbiology (Reading, Engl.) 155, 733–740 (2009).

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

101. Gilbert, L. A. et al. CRISPR-Mediated Modular RNA-Guided Regulationof Transcription in Eukaryotes. Cell 154, 442–451 (2013).

102. Konermann, S. et al. NOT FINAL PROOF. Nature 1–17 (2013). doi:10.1038/nature12466

103. Maeder, M. L. et al. nbt.2726. Nat Biotech 1–9 (2013). doi:10.1038/nbt.2726

104. Perez-Pinera, P. et al. RNA-guided gene activation by CRISPR-Cas9-based transcription factors. Nat Meth 10, 973–976 (2013).

105. Hsu, P. D., Scott, D. A., Weinstein, J. A. & Ran, F. A. DNA targeting specificity of RNA-guided Cas9 nucleases. Nature (2013).

106. Fu, Y. et al. High-frequency off-target mutagenesis induced by CrIsPr-Cas nucleases in human cells. Nat Biotech 1–6 (2013). doi:10.1038/nbt.2623.

107. Dianov, G. L. & Hübscher, U. Mammalian Base Excision Repair: the Forgotten Archangel. Nucleic Acids Res (2013).

108. Ran, F. A. et al. Genome engineering using the CRISPR-Cas9 system. Nature Protocols 8, 2281–2308 (2013).

109. Ran, F. A. et al. Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Cell 1–10 (2013). doi:10.1016/j.cell.2013.08.021

110. O'Roak, B. J., Deriziotis, P., Lee, C., Vives, L. & Schwartz, J. J. Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. Nature (2011).

111. Ng, S. B., Turner, E. H., Robertson, P. D. & Flygare, S. D. Targeted capture and massively parallel sequencing of 12 human exomes. Nature (2009).

112. Hsu, P. D. & Zhang, F. Dissecting Neural Function Using Targeted Genome Engineering Technologies. ACS Chemical Neuroscience (2012).

113. Hofacker, I. L., Fontana, W. & Stadler, P. F. Fast folding and comparison of RNA secondary structures. Monatshefte für Chemie/ … (1994).

114. Niu, Y. et al. Generation of Gene-Modified Cynomolgus Monkey via Cas9/RNA-MediatedGene Targeting in One-Cell Embryos. Cell 1–8 (2014). doi:10.1016/j.cell.2014.01.027

115. Grieger, J. C. & Samulski, R. J. Packaging capacity of adeno-associated virus serotypes: impact of larger genomes on infectivity and postentry steps. J Virol (2005).

116. Gabriel, R., Lombardo, A., Arens, A. & Miller, J. C. An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nature (2011).

117. Tesson, L., Usal, C., Ménoret, S., Leung, E. & Niles, B. J. Knockout rats generated by embryo microinjection of TALENs. Nature (2011).

118. Fu, Y., Sander, J. D., Reyon, D., Cascio, V. M. & Joung, J. K. Improving CRISPR-Cas nuclease specificity using truncated guide RNAs. Nat Biotech (2014). doi:10.1038/nbt.2808

119. Zhang, Y. et al. Processing-Independent CRISPR RNAs Limit Natural Transformation in Neisseria meningitidis. Molecular Cell 50, 488–503 (2013).

120. Zoephel, J. & Randau, L. RNA-Seq analyses reveal CRISPR RNA processing and regulation patterns. Biochem. Soc. Trans. 41, 1459–1463 (2013).

121. Shalem, O. et al. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Science 343, 84–87 (2014).

122. Wang, T., Wei, J. J., Sabatini, D. M. & Lander, E. S. Genetic screens in human cells using the CRISPR-Cas9 system. Science (2014).

123. Pattanayak, V., Lin, S., Guilinger, J. P. & Ma, E. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. Nature (2013).

2014281026 05 May 2020

The claims defining the invention are as follows:

1. A method of modifying a genomic locus of interest in a cell, the method comprising

delivering a non-naturally occurring or engineered composition comprising a tandem guideRNA construct comprising:

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

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

(b) a first tracr mate sequence,

(c) a first tracr sequence,

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

(e) a second tracr mate sequence, and

(f) a second tracr sequence, and

II. a polynucleotide sequence encoding a CRISPR enzyme comprising at least one or more nuclear localization sequences (NLS)s,

wherein the polynucleotide sequence comprises a linker sequence between the first tracr sequence and the second guide sequence, whereby the first guide sequence and the second guide sequence are in tandem,

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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence,

wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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 other strand near the second target sequence inducing a double strand break, thereby modifying the genomic locus.

2014281026 05 May 2020

2. A method of modifying an organism or a non-human organism 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, the method 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,

(b) a first tracr mate sequence,

(c) a first tracr sequence,

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

(e) a second tracr mate sequence, and

(f) a second tracr sequence, and

wherein a linker sequence is present between the first tracr sequence and the second guide sequence, whereby the first guide sequence and the second guide sequence are in tandem,

and

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

wherein components I and II are located on the same or different vectors of the system,

when transcribed, a first tracr mate sequence hybridizes to a first tracr sequence 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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence,

wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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 other strand near the

2014281026 05 May 2020

second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism.

3. A non-naturally occurring or engineered composition for modifying an organism or a non-human organism 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, the composition comprising:

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

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

(b) a first tracr mate sequence,

(c) a first tracr sequence,

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

(e) a second tracr mate sequence, and

(f) a second tracr sequence, and

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

wherein the polynucleotide sequence comprises a linker sequence between the first tracr sequence and the second guide sequence, whereby the first guide sequence and the second guide sequence are in tandem,

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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence,

wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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 other strand near the

2014281026 05 May 2020

second target sequence inducing a double strand break to modify the organism or the non-human organism, with a reduction in likelihood of off-target modifications.

4. The method or composition of any one of claims 1 to 3, wherein the modifying of the genomic locus, the modifying of the organism or the non-human organism, or the modifying of the organism or the non-human organism with a reduction in likelihood of off-target modifications, comprises inducing an indel and/or a microdeletion.

5. The method or composition of claim 4, wherein the microdeletion comprises deletion of a sequence between the first and second target sequences.

6. The method or composition of any one of claims 1 to 5, wherein the linker sequence:

(i) comprises at least 5 nucleotides;

(ii) comprises at least 10 nucleotides;

(iii) comprises at least 20 nucleotides; or

(iv) has 8 or 12 nucleotides.

7. The method or composition of any one of claims 1 to 6, wherein the first or second guide sequence, the first or second tracr sequence and/or the first or second tracr mate sequences are modified.

8. The method or composition of any one of claims 1 to 7, wherein the polynucleotides encoding 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 nanoparticles, exosomes, microvesicles, or a gene-gun.

9. The method or composition of any one of claims 1 to 8, wherein the CRISPR enzyme is a Cas9 enzyme functioning as a nuclease or a Cas9 enzyme functioning as a nickase.

10. A method of modifying an organism or a non-human organism 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, the method comprising

delivering a non-naturally occurring or engineered composition comprising:

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

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

(b) a first tracr mate sequence,

(c) a first tracr sequence,

2014281026 05 May 2020

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

(e) a second tracr mate sequence, and

(f) a second tracr sequence, and

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

wherein the polynucleotide sequence comprises a linker sequence between the first tracr sequence and the second guide sequence, whereby the first guide sequence and the second guide sequence are in tandem,

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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence,

wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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 other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism.

11. A non-naturally occurring or engineered composition for modifying an organism or a non-human organism 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, the composition comprising:

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

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

(b) a first tracr mate sequence,

(c) a first tracr sequence,

2014281026 05 May 2020

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

(e) a second tracr mate sequence, and

(f) a second tracr sequence, and

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

wherein the polynucleotide sequence comprises a linker sequence between the first tracr sequence and the second guide sequence, whereby the first guide sequence and the second guide sequence are in tandem,

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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence,

wherein the second CRISPR complex comprises the CRISPR enzyme complexed with (1) the second guide sequence that is hybridized to the second target sequence, and (2) the second tracr mate sequence that is hybridized 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 other strand near the second target sequence inducing a double strand break to modify the organism or the non-human organism.

12. The composition of claim 11, wherein the linker sequence:

(i) comprises at least 5 nucleotides;

(ii) comprises at least 10 nucleotides;

(iii) comprises at least 20 nucleotides; or

(iv) has 8 or 12 nucleotides.

13. The method or composition of any one of claims 7 to 12, wherein the modification comprises optimized first or second tracr sequence and/or optimized first or second guide sequence and/or, co-fold structure of first or second tracr sequence and/or first or second tracr

2014281026 05 May 2020

mate sequence(s) respectively and/or stabilizing secondary structures (hair pins) of first or second tracr sequence and/or first or second tracr sequence with a reduced region of base-pairing and/or first or second tracr sequence fused RNA elements.

14. A method of modifying an organism or a non-human organism 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,

(b) a first tracr mate sequence,

(c) a first tracr sequence,

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

(e) a second tracr mate sequence, and

(f) a second tracr sequence, and

wherein a linker sequence is present between the first tracr sequence and the second guide sequence, whereby the first guide sequence and the second guide sequence are in tandem,

and

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

wherein components I and II are located on the same or different vectors of the system,

when transcribed, a first tracr mate sequence hybridizes to a first tracr sequence 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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence,

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

wherein the polynucleotide sequence encoding a CRISPR enzyme is DNA or RNA, and

2014281026 05 May 2020

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 other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human organism.

15. The method of claim 10 or claim 14, wherein the linker sequence:

(i) comprises at least 5 nucleotides;

(ii) comprises at least 10 nucleotides;

(iii) comprises at least 20 nucleotides; or

(iv) has 8 nucleotides.

16. The method of claim 14 or claim 15, wherein one or more of the vectors are delivered via nanoparticles, exosomes, microvesicles, or a gene-gun.

17. The method or composition of any one of claims 10 to 16, wherein 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 other strand near the second target sequence results in a 5' overhang.

18. The method or composition of claim 17, wherein the 5' overhang is:

(i) at most 200 base pairs;

(ii) at most 100 base pairs;

(iii) at most 50 base pairs;

(iv) at least 26 base pairs;

(v) at least 30 base pairs; or

(vi) 34-50 base pairs.

19. The method or composition of any one of claims 10 to 18, wherein the CRISPR enzyme is a Cas9 enzyme.

20. The method or composition of claim 19, wherein the CRISPR enzyme comprises one or more mutations in a catalytic domain.

21. The method or composition of claim 20, wherein the CRISPR enzyme comprises one or more mutations selected from the group consisting of D10A, E762A, H840A, N854A, N863A and D986A.

22. The method or composition of claim 21, wherein the CRISPR enzyme has the D10A mutation.

2014281026 05 May 2020

23. The method or composition of any one of claims 1 to 22, wherein 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.

24. The method or composition of any one of claims 1 to 23, wherein:

(i) the first and second tracr mate sequence share 100% identity; and/or

(ii) the first and second tracr sequence share 100% identity.

25. The method or composition of any one of claims 1 to 24, wherein the CRISPR enzyme is SpCas9.

26. A method of modifying a DNA duplex at a locus of interest in a cell, the method comprising delivering to the cell:

I. a polynucleotide comprising:

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

(b) a first tracr mate sequence,

(c) a first tracr sequence,

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

(e) a second tracr mate sequence, and

(f) a second tracr sequence,

wherein (a), (b), (c), (d), (e) and (f) are arranged in a 5' to 3' orientation and a linker sequence is present between (c) and (d), whereby the first guide sequence and the second guide sequence are in tandem,

and

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

wherein the first target sequence is on a first strand of the DNA duplex and the second target sequence is on the opposite strand of the DNA duplex, and when the first and second guide sequences are hybridized to said target sequences in the duplex, the 5' ends of the first guide sequence and the second guide sequence are offset relative to each other by at least one base pair of the duplex;

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

2014281026 05 May 2020

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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to the first tracr sequence,

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

and wherein said first strand of the DNA duplex is cleaved near said first target sequence, and said opposite strand of the DNA duplex is cleaved near said second target sequence, resulting in a double strand break with a 5' overhang or a 3' overhang.

27. A method of modifying a DNA duplex at a locus of interest in a cell, the method comprising delivering to the cell a vector system comprising one or more vectors comprising:

I. a first polynucleotide comprising a regulatory element operably linked to

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

(b) a first tracr mate sequence,

(c) a first tracr sequence,

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

(e) a second tracr mate sequence, and

(f) a second tracr sequence, and

wherein (a), (b), (c), (d), (e) and (f) are arranged in a 5' to 3' orientation and a linker sequence is present between (c) and (d), whereby the first guide sequence and the second guide sequence are in tandem,

and

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

wherein components I and II are located on the same or different vectors of the system,

wherein the first target sequence is on a first strand of the DNA duplex and the second target sequence is on the opposite strand of the DNA duplex, and when the first and second guide sequences are hybridized to said target sequences in the duplex, the 5' ends of the first guide

2014281026 05 May 2020

sequence and the second guide sequence are offset relative to each other by at least one base pair of the duplex;

wherein when transcribed, the first and the second tracr mate sequences hybridize to a tracr sequence, and the first and the second guide sequences 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 to the first target sequence, and (2) the first tracr mate sequence that is hybridized to a tracr sequence,

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

and wherein said first strand of the DNA duplex is cleaved near said first target sequence, and said opposite strand of the DNA duplex is cleaved near said second target sequence, resulting in a double strand break with a 5' overhang or a 3' overhang.

28. The method of claim 26 or claim 27, wherein said CRISPR enzyme is a nickase enzyme, optionally a Cas9 enzyme comprising at least one mutation in a catalytic domain.

29. The method of claim 28, wherein said at least one mutation is in the RuvC domain and optionally is selected from the group consisting of D10A, E762A and D986A, or is in the HNH domain and optionally is selected from the group consisting of H840A, N854A and N863A.

30. The method of any one of claims 26 to 29, wherein said offset between the 5' ends of the first polynucleotide and the second polynucleotide is than -8 bp or -278 to +58 bp or -200 to +200 bp or up to or over 100 bp or -4 to 20 bp or +23 bp or +16 or +20 or +16 to +20 bp or -3 to +18 bp.

31. The method of any one of claims 26 to 30, wherein said cleavage of said first strand and of said opposite strand of the DNA duplex occurs 3' to a PAM (Protospacer adjacent motif) on each strand, and wherein said PAM on said first strand is separated from said PAM on said opposite strand by from 30 to 150 base pairs.

32. The method of any one of claims 26 to 31, wherein said overhang is at most 200 bases, at most 100 bases, or at most 50 bases.

2014281026 05 May 2020

33. The method of claim 32, wherein the overhang is: (i) at least 1 base, at least 10 bases, at least 15 bases, at least 26 bases or at least 30 bases; or (ii) between 34 and 50 bases or between 1 and 34 bases.

34. The method of any one of claims 26 to 33, wherein said double strand break induces a microdeletion in the locus of interest, preferably between the first and second target sequences.

35. The method of any one of claims 26 to 30, wherein the linker sequence comprises at least 5 nucleotides, at least 10 nucleotides, or at least 20 nucleotides, or wherein the linker consists of 8 nucleotides.

36. The method of any one of claims 26 to 31, wherein 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, and optionally wherein any or all of I, II and III are delivered via nanoparticles, exosomes, microvesicles, or a gene-gun.

37. The method of any one of claims 26 to 32, wherein the first and second tracr mate sequence share 100% identity and/or the first and second tracr sequence share 100% identity.

38. The method of claim 26, wherein each of I and II is provided in a vector, optionally wherein each is provided in the same or a different vector.

39. The method of any one of claims 26 to 33, wherein the first or second guide sequence, the first or second tracr sequence and/or the first or second tracr mate sequences are modified, optionally wherein the modification comprises optimized first or second tracr sequence and/or optimized first or second guide sequence and/or, co-fold structure of first or second tracr sequence and/or first or second tracr mate sequence(s) respectively and/or stabilizing secondary structures (hair pins) of first or second tracr sequence and/or first or second tracr sequence with a reduced region of base-pairing and/or first or second tracr sequence fused RNA elements.

40. The method of any one of claims 26 to 39, wherein said locus of interest comprises a gene and wherein said method results in a change in the expression of said gene, or in a change in the activity or function of the gene product.

41. The method of claim 40, wherein said gene product is a protein, and/or wherein said change in expression, activity or function is a reduction in said expression, activity or function.

42. The method of any one of claims 26 to 41, further comprising:

2014281026 05 May 2020

- delivering to the cell a double-stranded oligodeoxynucleotide (dsODN) comprising overhangs complimentary to the overhangs created by said double strand break, wherein said dsODN is integrated into the locus of interest; or

- delivering to the cell a single-stranded oligodeoxynucleotide (ssODN), wherein said ssODN acts as a template for homology directed repair of said double strand break.

43. The method of any one of claims 26 to 42, which is for the prevention or treatment of disease in an individual, optionally wherein said disease is caused by a defect in said locus of interest.

44. The method of claim 43, wherein the method is conducted *in vivo* in the individual or *ex vivo* on a cell taken from the individual, optionally wherein said cell is returned to the individual.

45. A kit or composition comprising:

I. a polynucleotide comprising:

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

(b) a first tracr mate sequence,

(c) a first tracr sequence,

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

(e) a second tracr mate sequence, and

(f) a second tracr sequence,

wherein (a), (b), (c), (d), (e) and (f) are arranged in a 5' to 3' orientation and a linker sequence is present between (c) and (d), whereby the first guide sequence and the second guide sequence are in tandem,

and

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

wherein the first target sequence is on a first strand of the DNA duplex and the second target sequence is on the opposite strand of the DNA duplex, and when the first and second guide sequences are hybridized to said target sequences in the duplex, the 5' ends of the first guide sequence and the second guide sequence are offset relative to each other by at least one base pair of the duplex;

and optionally wherein each of I and II is provided in the same or a different vector.

2014281026 05 May 2020

46. Use of the kit or composition of claim 45 in a method of any one of claims 26 to 44.

47. Use of the kit or composition of claim 45 in the manufacture of a medicament, optionally wherein said medicament is for the prevention or treatment of a disease caused by a defect in said locus of interest.

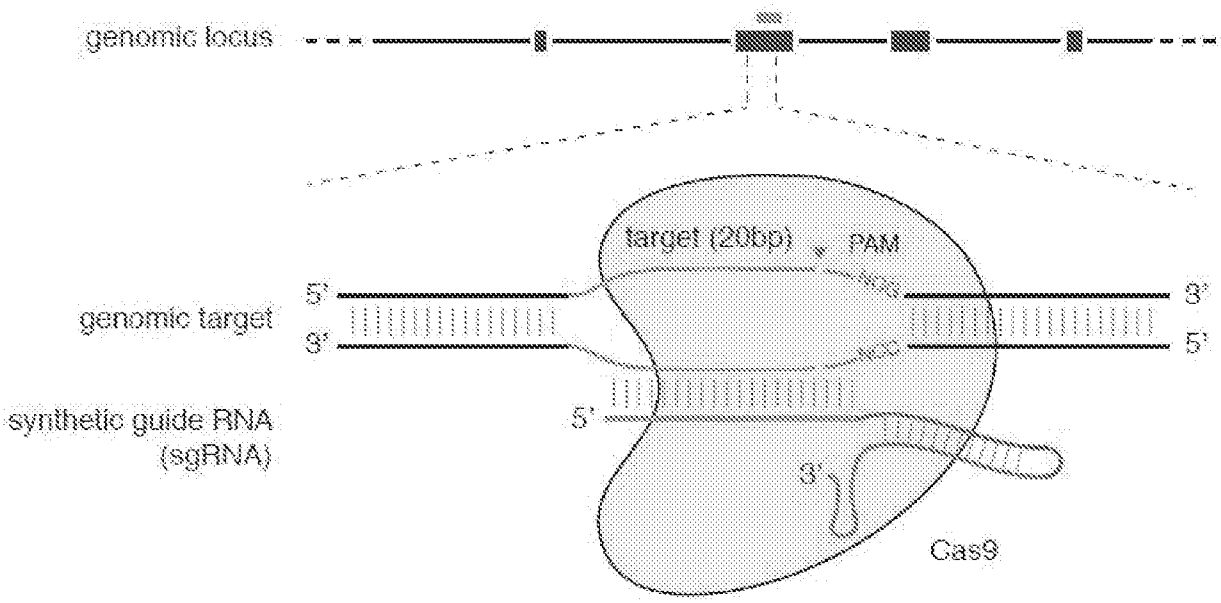

genomic locus
target (20bp)
PAM
5'
3'
genomic target
3'
5'
synthetic guide RNA
(sgRNA)
5'
3'
Cas9

FIG. 1

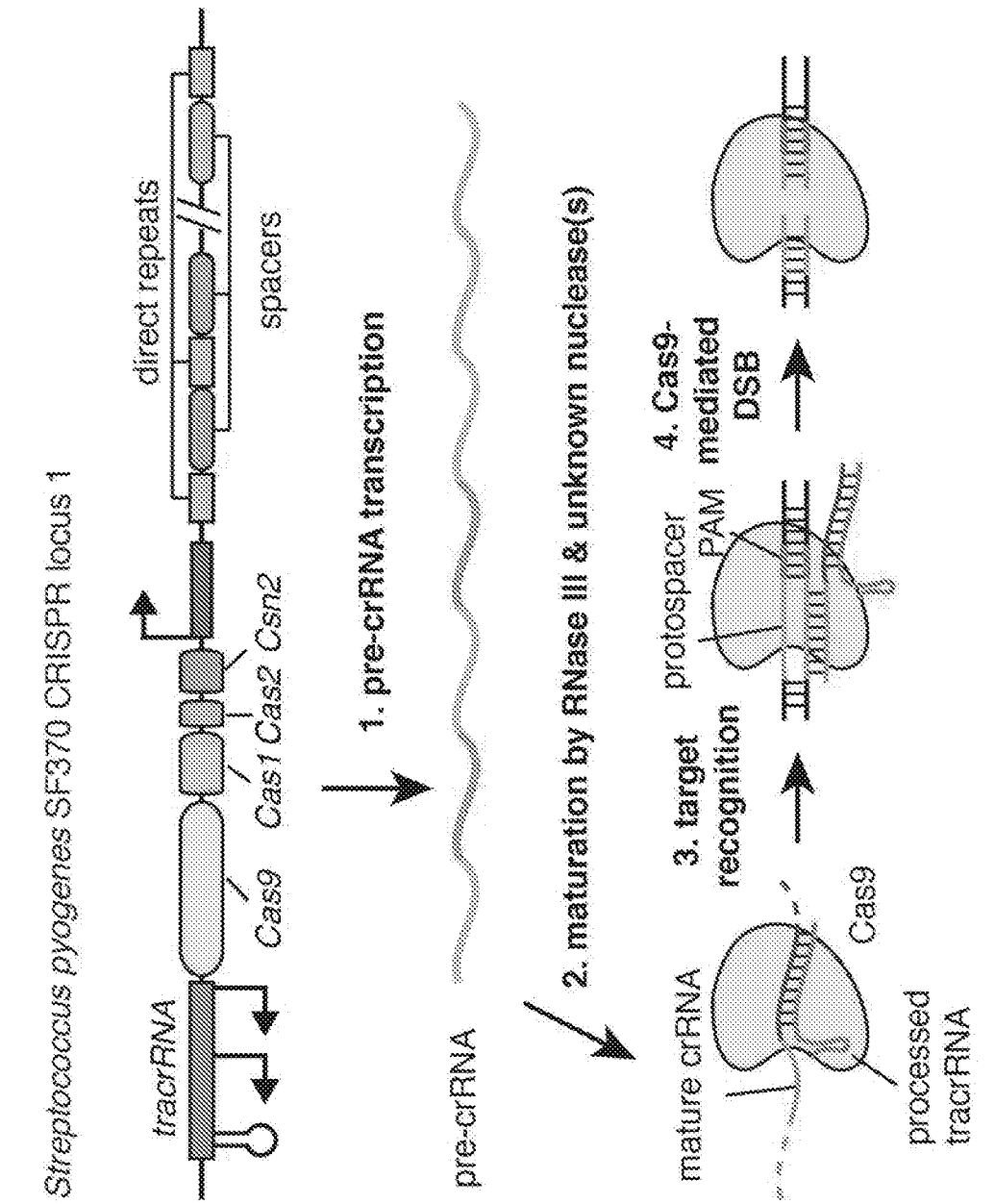

A
Streptococcus pyogenes SF370 CRISPR locus 1
tracrRNA
Cas9
Cas1
Cas2
Csn2
direct repeats
spacers
1. pre-crRNA transcription
pre-crRNA
2. maturation by RNase III & unknown nuclease(s)
mature crRNA
3. target recognition
protospacer
PAM
4. Cas9-mediated DSB
Cas9
processed tracrRNA

FIG. 2A

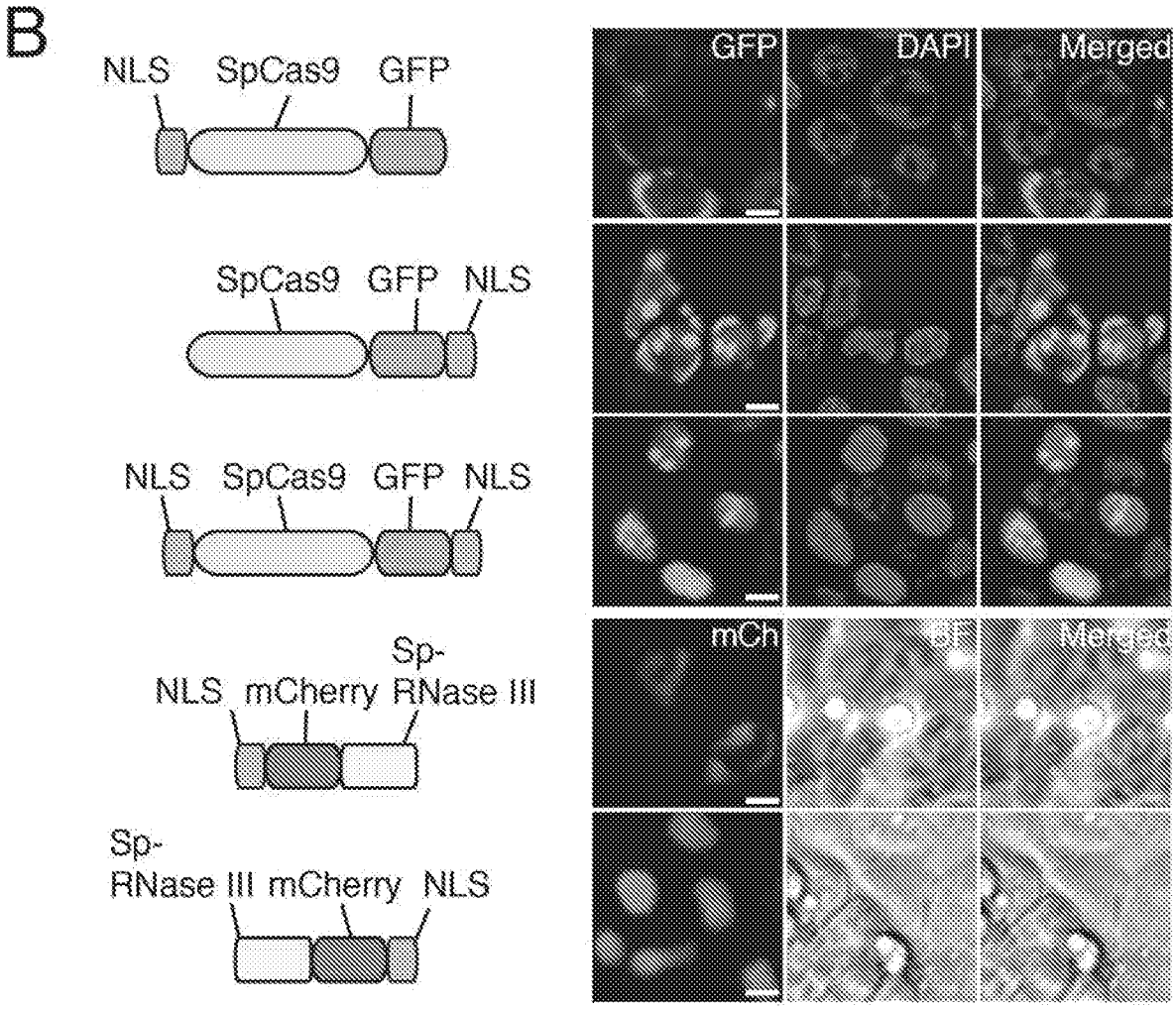

B
NLS SpCas9 GFP
SpCas9 GFP NLS
NLS SpCas9 GFP NLS
Sp-
NLS mCherry RNase III
Sp-
RNase III mCherry NLS
GFP
DAPI
Merged
mCh

FIG. 2B

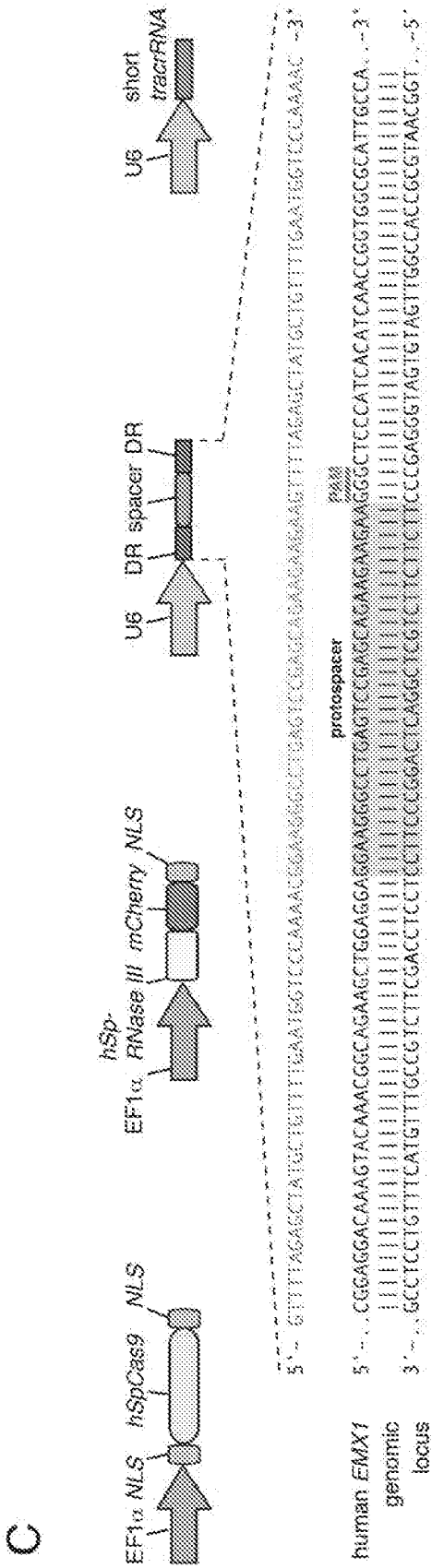

C
EF1α NLS hSpCas9 NLS
hSp
EF1α RNase III mCherry NLS
U6 DR spacer DR
U6 short tracrRNA
protospacer
PAM
human EMX1 genomic locus
5'- ...CGGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCA...-3'
3'- ...GCCTCCTGTTTCATGTTTGCCGTCTTCGACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCGAGGGTAGTGTAGTTGGCCACCGCGTAACGGT...-5'
5'- GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAACGGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC -3'

FIG. 2C

D

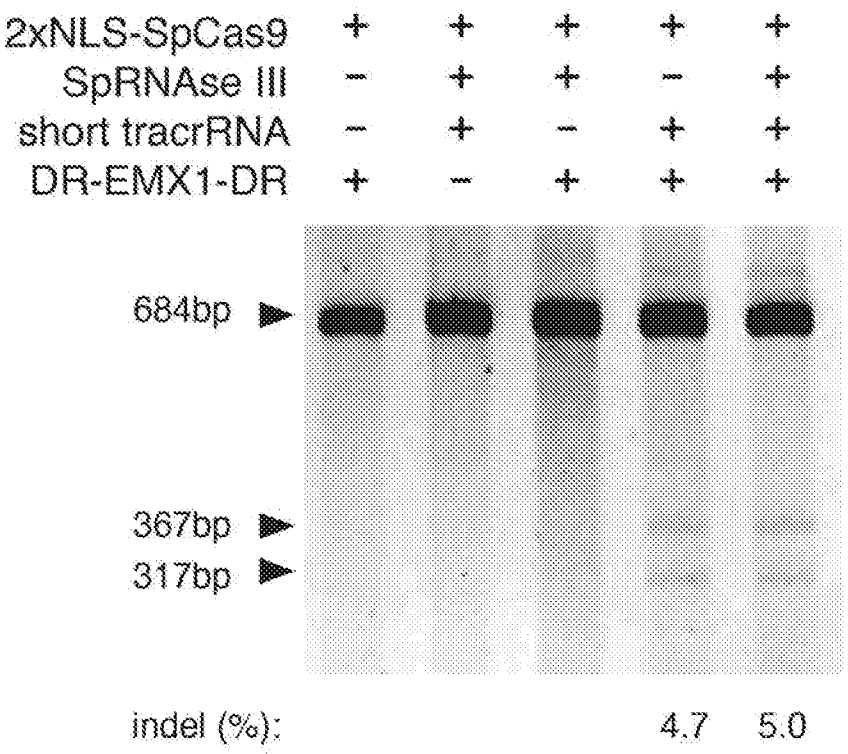

2xNLS-SpCas9 + + + + +
SpRNAse III − + + − +
short tracrRNA − + − + +
DR-EMX1-DR + − + + +
684bp
367bp
317bp
indel (%): 4.7 5.0

FIG. 2D

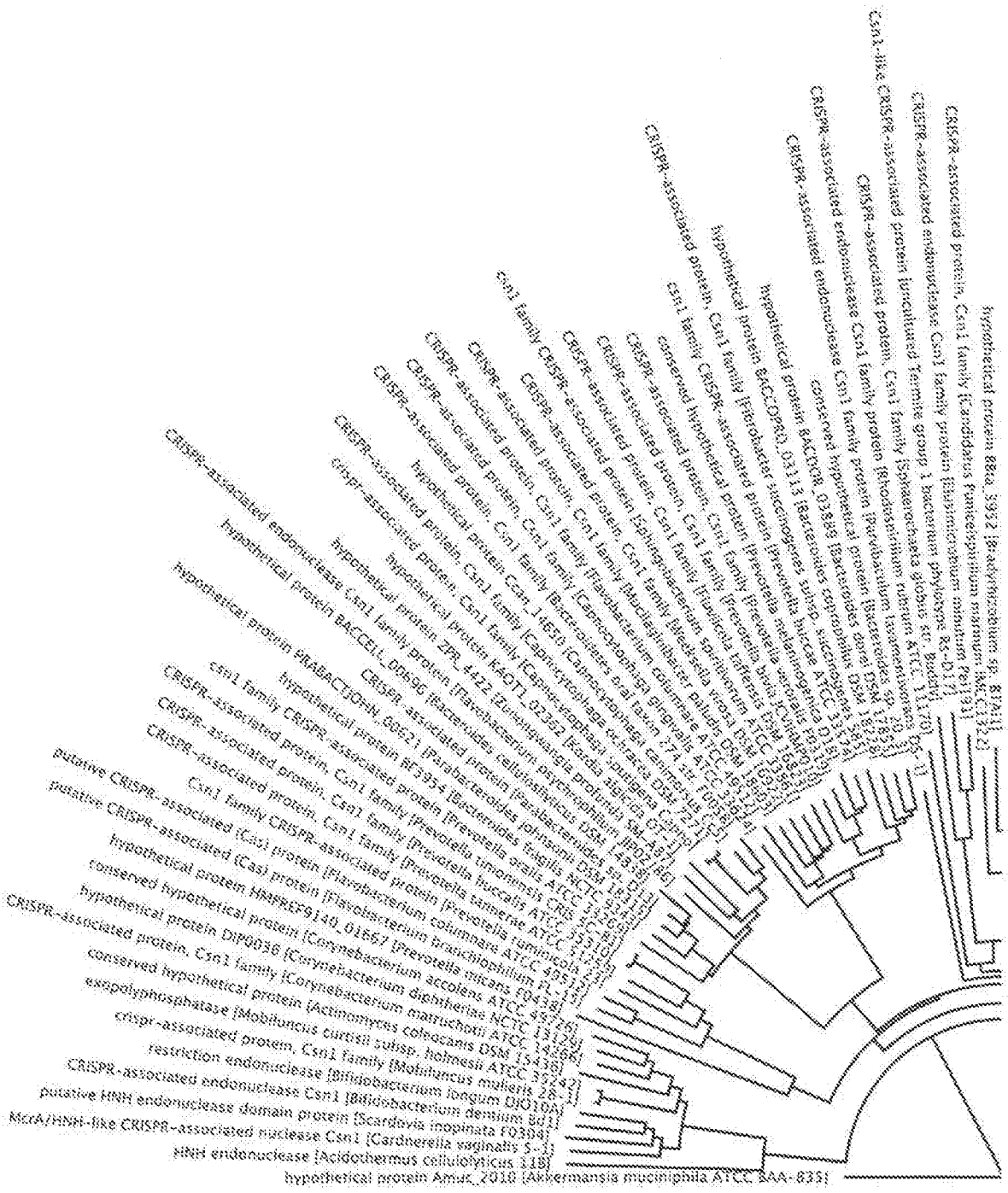
E

F

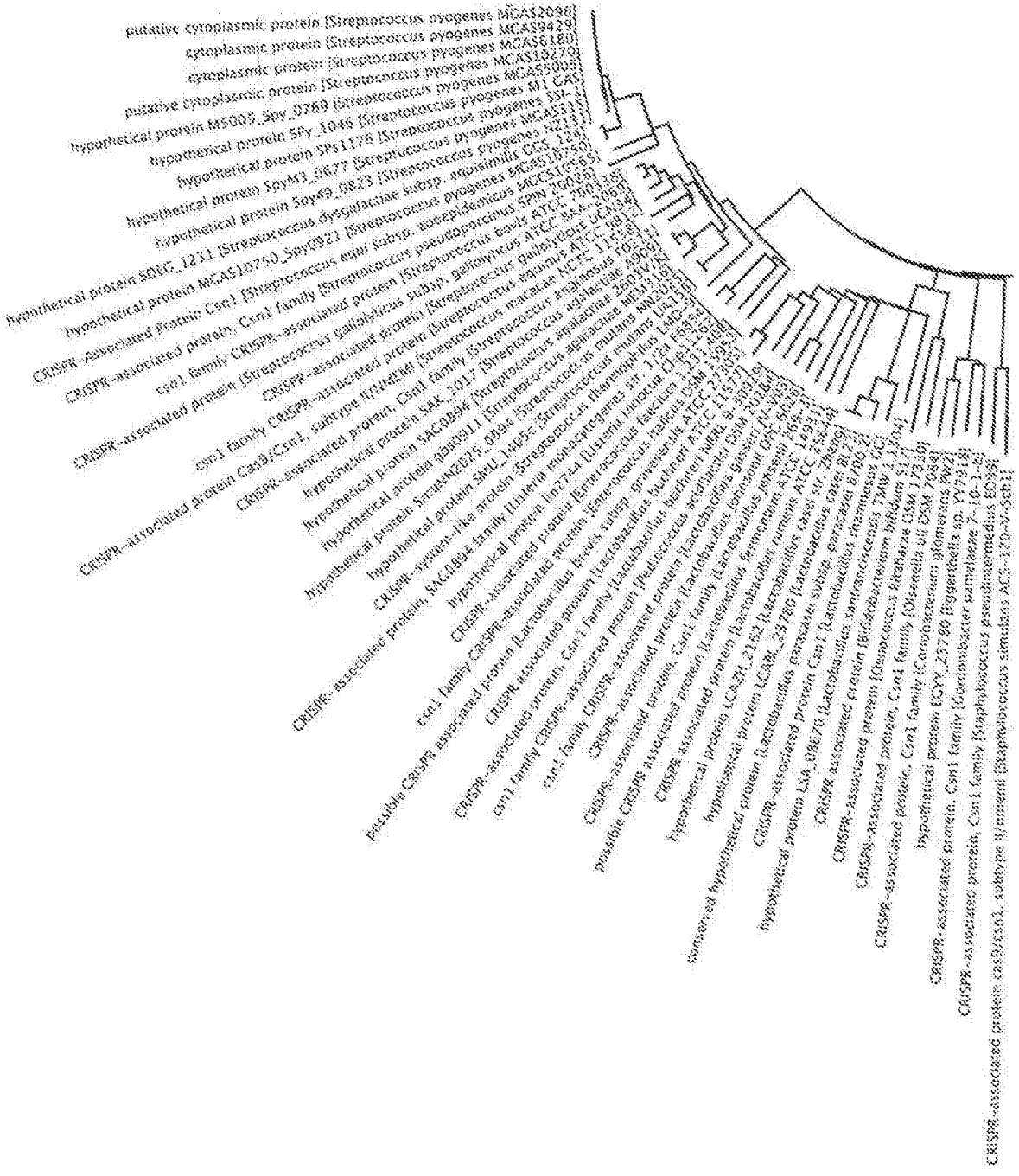
human *EMX1* protospacer target (mutation in 5 of 43 sequenced clones = 11.6%)

WT	5'-...CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG AAGAAGGGCTCCCATCACAT...-3'
Δ1	CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG---AGAAGGGCTCCCATCACAT
+1	CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGCTCCCATCACAT
Δ3	CTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAG-----AAGGGCTCCCATCACAT
m1, Δ6	CTGGAGGAGGAAGGGCCTGAGCCCGAGCAGAAG----------GGCTCCCATCACAT

FIG. 2E-F


hypothetical protein 88ca_3992 [Brachyspirum sp. 88AH]
CRISPR-associated protein, Csn1 family [Candidatus Puniceispirillum marinum IMCC1322]
CRISPR-associated endonuclease Csn1 family protein [Elusimicrobium minutum Pei191]
CRISPR-associated protein [uncultured Termite group 1 bacterium phylotype Rs-D17]
Csn1-like CRISPR-associated protein [Spirochaeta globosa str. Buddy 1]
CRISPR-associated protein, Csn1 family [Roseburia hominis A2-183]
CRISPR-associated endonuclease Csn1 family protein [Rhodanobacter lauranceae]
CRISPR-associated protein, Csn1 family [Parabacteroides sp. 20_3]
hypothetical protein BACCOPRO_03113 [Bacteroides coprophilus DSM 18228]
CRISPR-associated endonuclease Csn1 family protein [Parabacteroides goreii DSM 19448]
conserved hypothetical protein BACCOR_03848 [Bacteroides coprocola DSM 17136]
CRISPR-associated endonuclease Csn1 family protein [Prevotella oralis ATCC 33269]
hypothetical protein BACCOPRO_03113 [Bacteroides coprophilus DSM 18228]
hypothetical protein [Prevotella melaninogenica ATCC 25845]
CRISPR-associated protein, Csn1 family [Prevotella buccae ATCC 33574]
hypothetical protein BACCOPRO_03113 [Bacteroides coprophilus DSM 18228]
hypothetical protein [Prevotella melaninogenica ATCC 25845]
CRISPR-associated protein, Csn1 family [Prevotella buccae ATCC 33574]
hypothetical protein BACCOPRO_03113 [Bacteroides coprophilus DSM 18228]
CRISPR-associated protein, Csn1 family [Prevotella buccae ATCC 33574]
hypothetical protein [Prevotella melaninogenica ATCC 25845]
CRISPR-associated protein, Csn1 family [Prevotella buccae ATCC 33574]
hypothetical protein BACCOPRO_03113 [Bacteroides coprophilus DSM 18228


hypothetical protein Rham_4054 [Interhabitat hamburgensis K1]
CRISPR-associated protein, Csn1 family [Azospirillum sp. B510]
hypothetical protein [Enterobacter sakazakii ATCC 12868]
conserved hypothetical protein SJ517_3938 [Sphingomonas sp. S17]
conserved hypothetical protein [Ralstonia syzygii R24]
CRISPR-associated protein, Csn1 family [Haliangium ochraceum DSM 14365]
CRISPR-associated endonuclease Csn1 family [Verrucomicrobiae bacterium DG1235]
CRISPR-associated protein, Csn1 family [Verrucomicrobia bacterium IMCC26134]
CRISPR-associated protein, Csn1 family [Pedobacter sp. GW460-11-11-14-LB5]
CRISPR-associated protein, Csn1 family [Pedobacter sp. V48]
CRISPR-associated protein, Csn1 family [Flavobacterium sp. 52]
CRISPR-associated protein, Csn1 family [Flavobacterium johnsoniae UW101]
CRISPR-associated protein, Csn1 family [Niastella koreensis GR20-10]
CRISPR-associated protein, Csn1 family [Mucilaginibacter paludis DSM 18603]
CRISPR-associated protein, Csn1 family [Parafilimonas terrae]
CRISPR-associated protein, Csn1 family [Pedobacter saltans DSM 12145]
CRISPR-associated protein, Csn1 family [Pedobacter heparinus DSM 2366]
CRISPR-associated protein, Csn1 family [Pedobacter sp. BAL39]
CRISPR-associated protein, Csn1 family [Arcobacter nitrofigilis DSM 7299]
CRISPR-associated protein, Csn1 family [Sphaerochaeta globosa str. Buddy]
CRISPR-associated protein, Csn1 family [Sphaerochaeta pleomorpha str. Grapes]
CRISPR-associated protein, Csn1 family [Spirochaeta smaragdinae DSM 11293]
CRISPR-associated protein, Csn1 family [Leptospira interrogans serovar Lai str. 56601]
CRISPR-associated protein, Csn1 family [Leptospira interrogans serovar Copenhageni str. Fiocruz L1-130]
CRISPR-associated protein, Csn1 family [Leptospira interrogans serovar Lai str. 56601]
CRISPR-associated protein, Csn1 family [Leptospira interrogans serovar Pomona str. Pomona]
CRISPR-associated protein, Csn1 family [Leptospira interrogans serovar Pomona str. Kennewicki LC82-25


putative cytoplasmic protein [Streptococcus pyogenes MGAS2096]
cytoplasmic protein [Streptococcus pyogenes MGAS9429]
cytoplasmic protein [Streptococcus pyogenes MGAS10270]
putative cytoplasmic protein [Streptococcus pyogenes MGAS6180]
hypothetical protein MS005_Spy_0769 [Streptococcus pyogenes MGAS5005]
hypothetical protein SPy_1046 [Streptococcus pyogenes M1 GAS]
hypothetical protein SPs1178 [Streptococcus pyogenes MGAS315]
hypothetical protein SpyM3_0677 [Streptococcus pyogenes MGAS10750]
hypothetical protein Spy49_0823 [Streptococcus pyogenes MGAS49]
hypothetical protein SDEC_1231 [Streptococcus dysgalactiae subsp. equisimilis GGS_124]
hypothetical protein MGAS10750_Spy0921 [Streptococcus pyogenes MGAS10750]
CRISPR-associated protein Csn1 [Streptococcus equi subsp. zooepidemicus MGCS10565]
CRISPR-associated protein, Csn1 family [Streptococcus pseudoporcinus SPIN 20026]
csn1 family CRISPR-associated protein [Streptococcus gallolyticus UCN34]
CRISPR-associated protein [Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069]
CRISPR-associated protein [Streptococcus equinus ATCC 700338]
CRISPR-associated protein [Streptococcus macacae NCFC 11558]
csn1 family CRISPR-associated protein [Streptococcus agalactiae 2603V/R]
CRISPR-associated protein Cas9/Csn1, subtype II/NMEN [Streptococcus agalactiae A909]
hypothetical protein, Csn1 family [Streptococcus agalactiae NEM316]
hypothetical protein SAK_2017 [Streptococcus agalactiae 18RS21]
hypothetical protein spyM18_0594 [Streptococcus pyogenes MGAS8232]
hypothetical protein SmuNN2025_0594 [Streptococcus mutans NN2025]
CRISPR-associated protein, SAG1994 family [Streptococcus agalactiae ATCC 27591]
hypothetical protein SAG0894 [Streptococcus agalactiae 2603V/R]
CRISPR-system-like protein SkuCL_1485c [Streptococcus thermophilus LMD-9]
hypothetical protein [Streptococcus thermophilus LMD-9]
CRISPR-associated protein [Streptococcus mutans UA159]
hypothetical protein [Streptococcus mutans NN2025]
CRISPR-associated protein [Streptococcus mutans GS-5]
hypothetical protein [Streptococcus mutans 11A1]
CRISPR-associated protein [Streptococcus mutans 1SM1]
hypothetical protein [Streptococcus mutans U2A]
CRISPR-associated protein [Streptococcus mutans NV1996]
hypothetical protein [Streptococcus mutans UA130]
CRISPR-associated protein [Streptococcus mutans UA140]
hypothetical protein [Streptococcus mutans UA159]
CRISPR-associated protein [Streptococcus mutans GS-5]
hypothetical protein [Streptococcus mutans 11A1]
hypothetical protein [Streptococcus mutans 1SM1]
hypothetical protein [Streptococcus mutans U2A]
hypothetical protein [Streptococcus mutans NV1996]
hypothetical protein [Streptococcus mutans UA130]
hypothetical protein [Streptococcus mutans UA140]
CRISPR-associated protein [Streptococcus thermophilus LMD-9]
CRISPR-associated protein [Streptococcus thermophilus 1205]
CRISPR-associated protein [Streptococcus thermophilus MN-ZLW-002]
CRISPR-associated protein [Streptococcus thermophilus ND03]
CRISPR-associated protein [Streptococcus thermophilus JIM 8232]
CRISPR-associated protein [Streptococcus thermophilus MTH2-3L-3

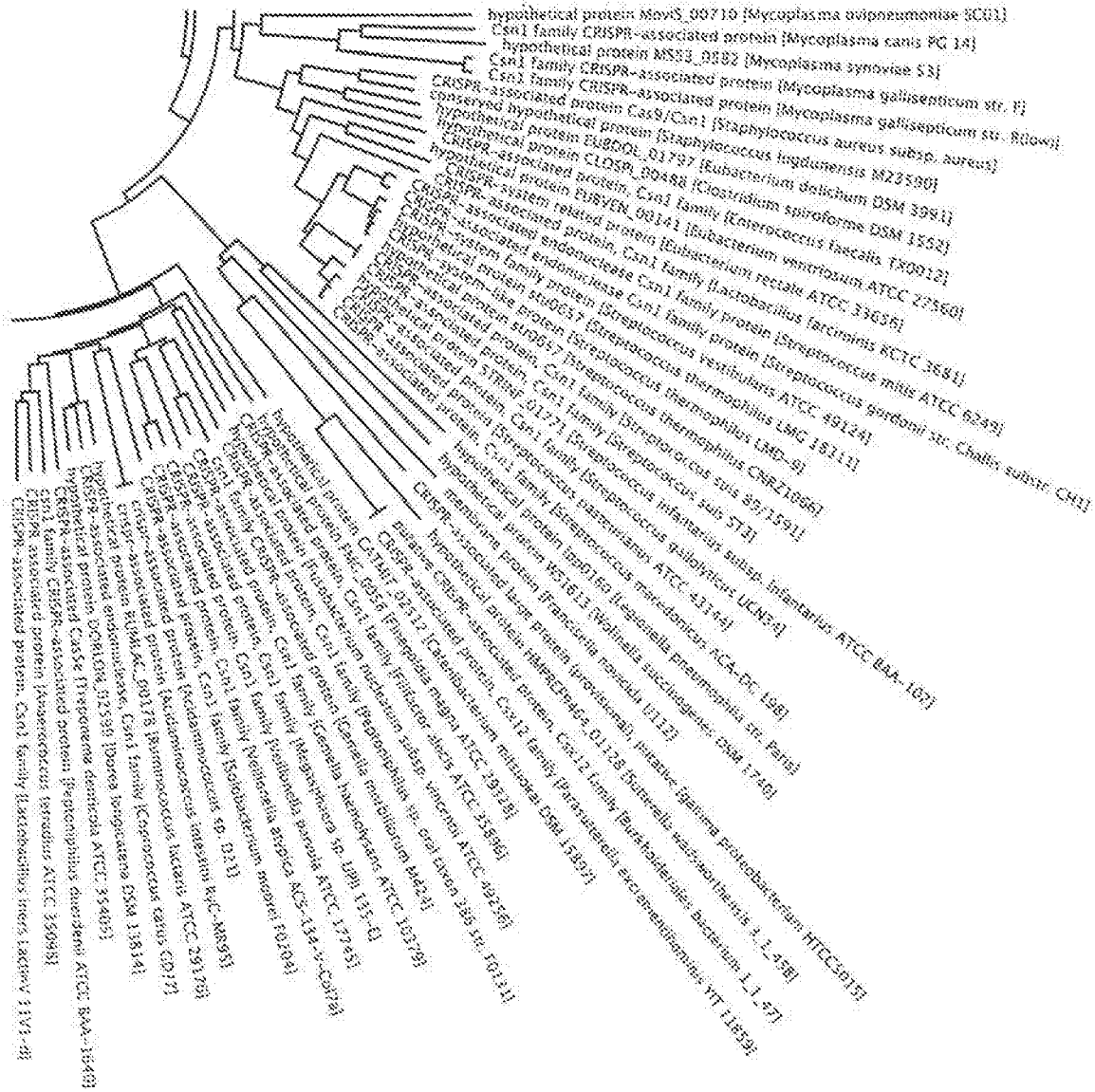

hypothetical protein MovS_0071D [Mycoplasma ovipneumoniae ECO1]
Csn1 family CRISPR-associated protein [Mycoplasma canis PG 14]
hypothetical protein MS53_0582 [Mycoplasma synoviae 53]
Csn1 family CRISPR-associated protein [Mycoplasma gallisepticum str. F]
Csn1 family CRISPR-associated protein [Mycoplasma gallisepticum str. Rlow]]
CRISPR-associated protein Csn9/Csn1 [Staphylococcus aureus subsp. aureus]
conserved hypothetical protein [Staphylococcus lugdunensis M23590]
hypothetical protein CLOSPI_00488 [Clostridium spiroforme DSM 1552]
hypothetical protein EUBVEN_00141 [Eubacterium ventriosum ATCC 27560]
CRISPR-associated protein, Csn1 family [Enterococcus faecalis TX0012]
CRISPR-system related protein [Eubacterium rectale ATCC 33656]
CRISPR-associated protein, Csn1 family [Lactobacillus jensenii JV-V03]
CRISPR-associated endonuclease Csn1 family protein [Streptococcus gordonii str. Challis substr. CH1]
CRISPR-system family protein [Streptococcus vestibularis ATCC 49124]
hypothetical protein sth0637 [Streptococcus thermophilus LMG 18311]
CRISPR-associated protein [Streptococcus thermophilus LMD-9]
hypothetical protein sth0637 [Streptococcus thermophilus CNRZ1066]
CRISPR-associated protein, Csn1 family [Streptococcus suis 89/1591]
hypothetical protein STRINF_01271 [Streptococcus infantarius subsp. infantarius ATCC BAA-102]
CRISPR-associated protein, Csn1 family [Streptococcus infantarius subsp. infantarius CJ18]
hypothetical protein [Streptococcus pasteurianus ATCC 43144]
CRISPR-associated protein, Csn1 family [Streptococcus thermophilus JCM 8649]
hypothetical protein [Streptococcus thermophilus ACA-DC 198]
membrane protein [Francisella novicida U112]
hypothetical protein WS1612 [Wolinella succinogenes DSM 1740]
putative CRISPR-associated protein [Campylobacter mucosalis CVM N29710]
hypothetical protein FRAPHE_01328 [Francisella philomiragia subsp. philomiragia ATCC 25017]
CRISPR-associated protein, Csn1 family [Haemophilus parainfluenzae T3T1]
hypothetical protein [Haemophilus sp. oral taxon 851 str. F0434]
CRISPR-associated protein, Csn1 family [Aggregatibacter aphrophilus NJ8700



FIG. 4A

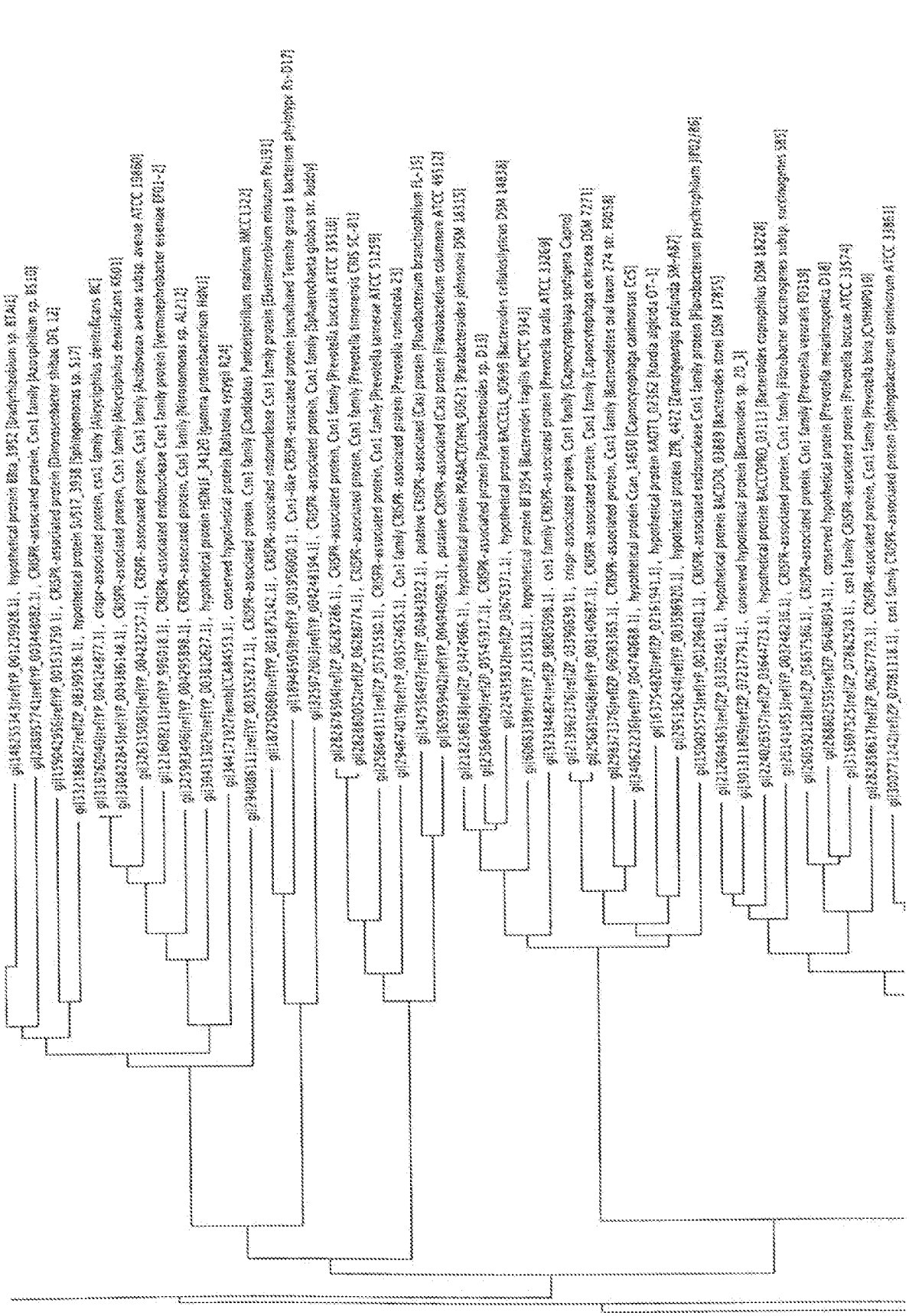


FIG. 4B

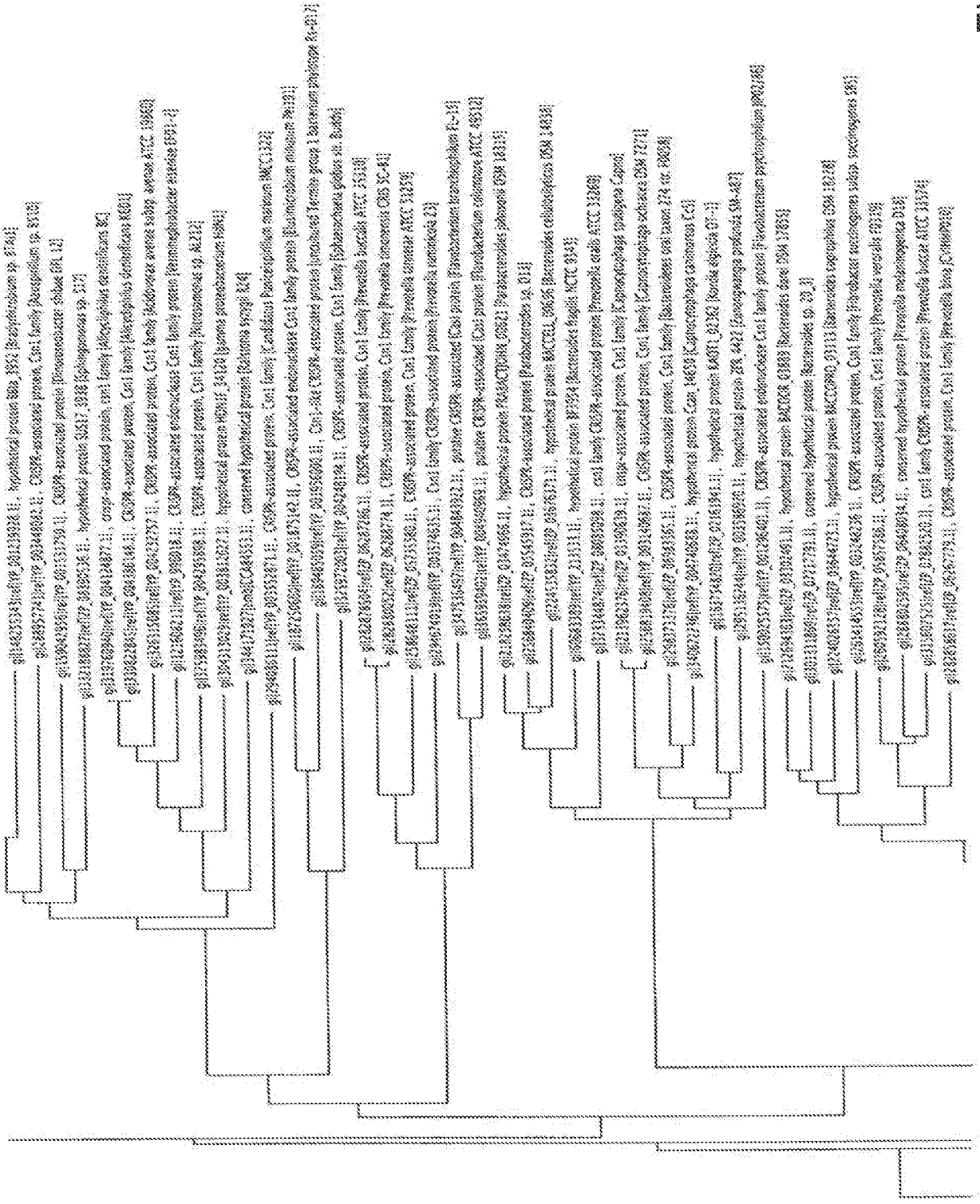


FIG. 4C

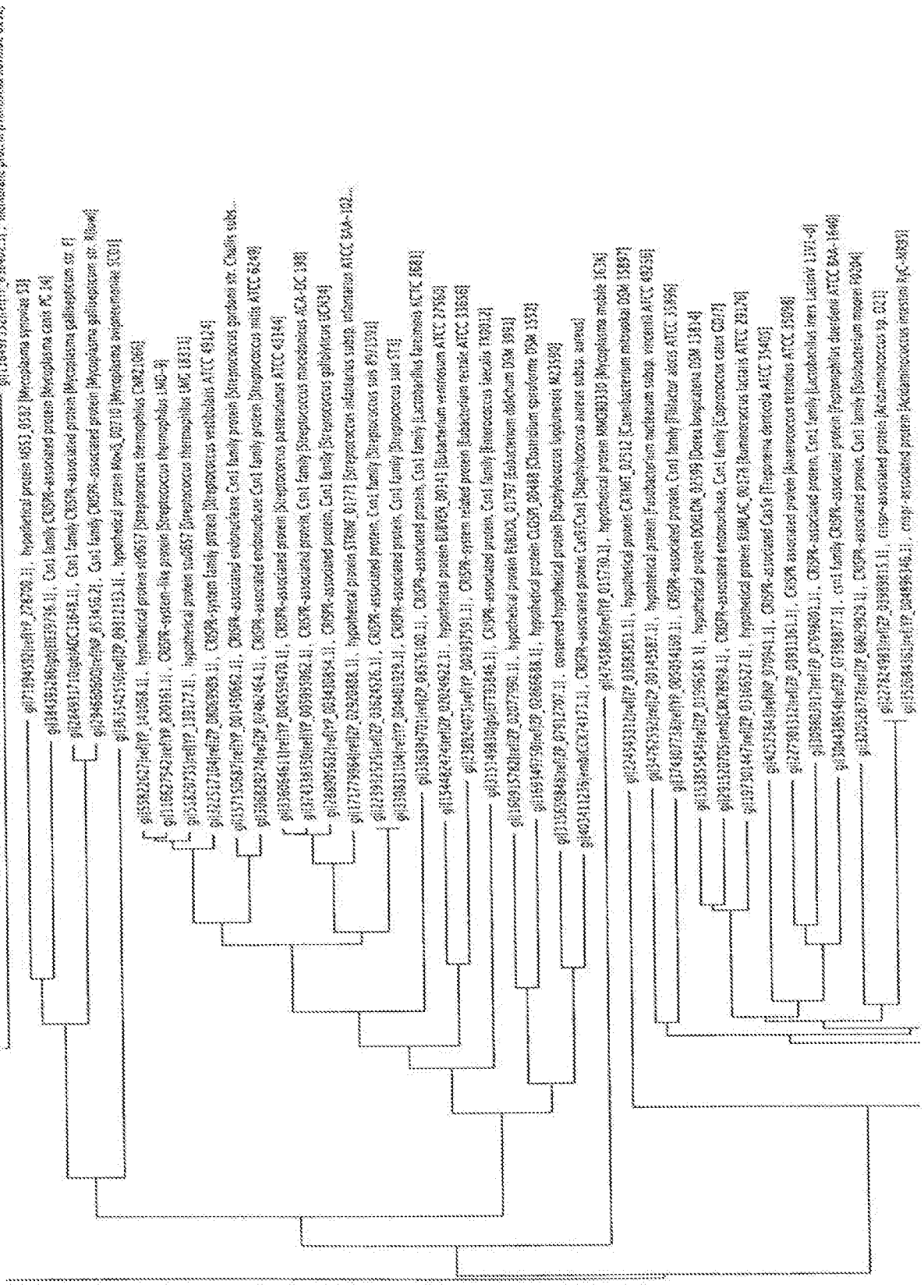


FIG. 4D

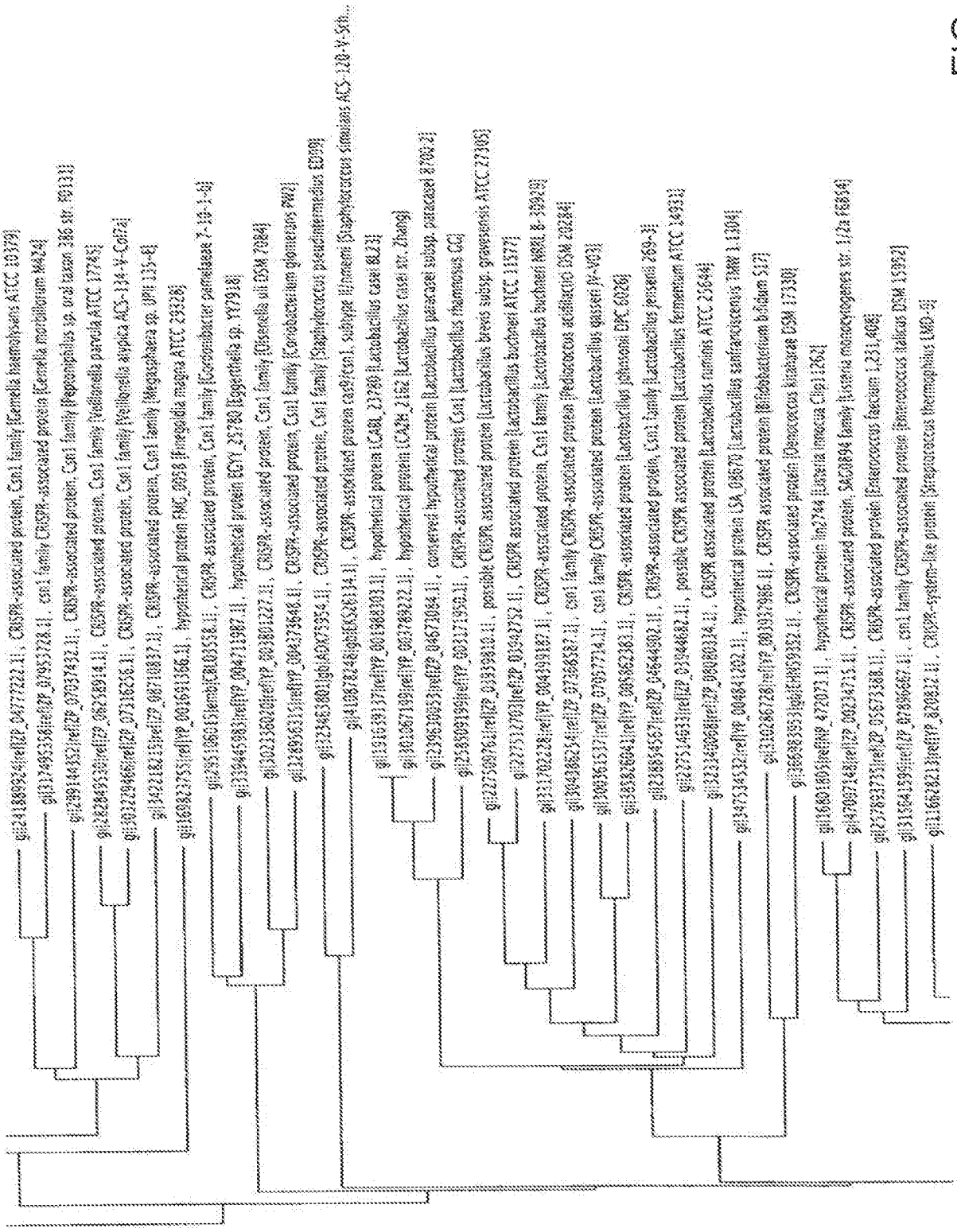


FIG. 4E

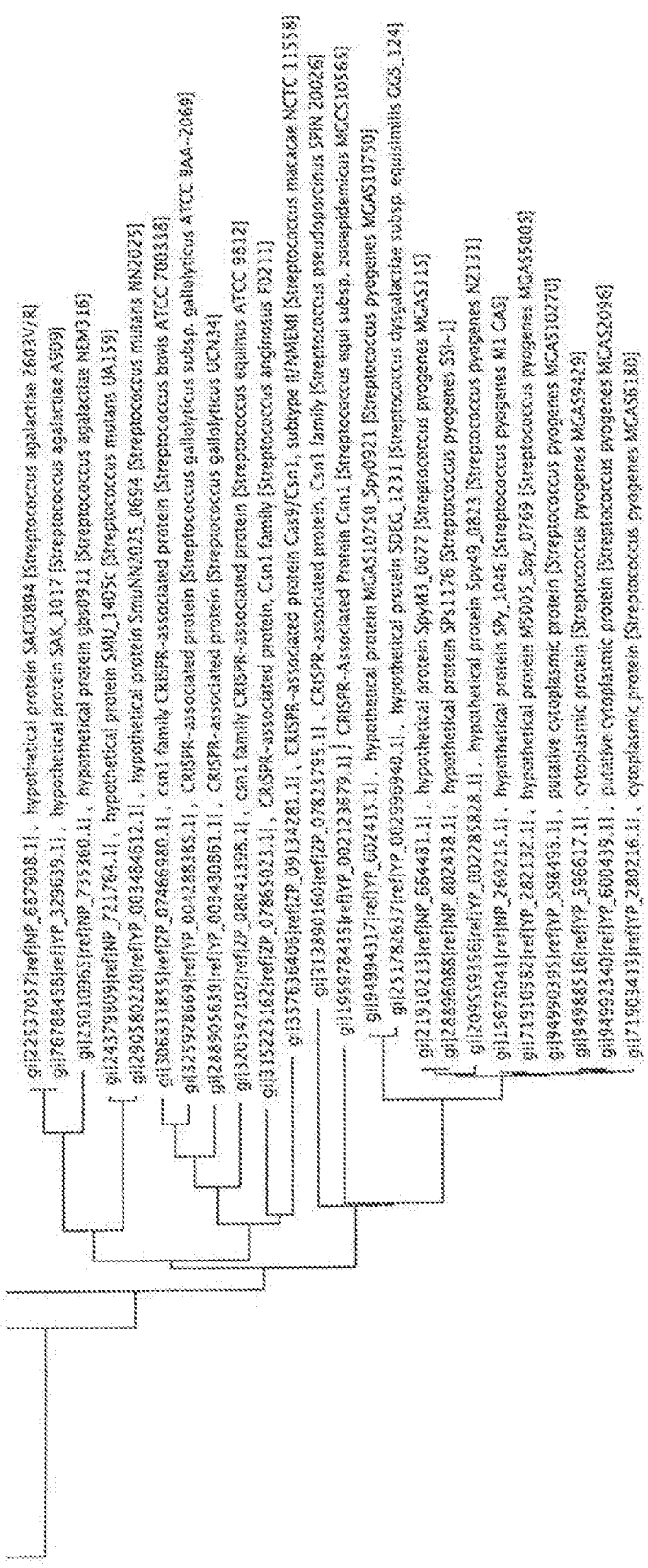

gi|22533705|ref|NP_667908.1| hypothetical protein SAG0894 [Streptococcus agalactiae 2603V/R]
gi|76788458|ref|YP_329639.1| hypothetical protein SAK_1017 [Streptococcus agalactiae A909]
gi|23101095|ref|NP_735360.1| hypothetical protein gbs0911 [Streptococcus agalactiae NEM316]
gi|24379809|ref|NP_721764.1| hypothetical protein SMU_1405c [Streptococcus mutans UA159]
gi|290580120|ref|YP_003484612.1| hypothetical protein SmuNN2025_0694 [Streptococcus mutans NN2025]
gi|339631855|ref|ZP_07466980.1| cas1 family CRISPR-associated protein [Streptococcus bovis ATCC 700338]
gi|321978869|ref|YP_004248365.1| CRISPR-associated protein [Streptococcus gallolyticus subsp. gallolyticus ATCC BAA-2069]
gi|288905616|ref|YP_003430861.1| CRISPR-associated protein [Streptococcus gallolyticus UCN34]
gi|320547102|ref|ZP_08041398.1| cas1 family CRISPR-associated protein [Streptococcus equinus ATCC 9812]
gi|313223162|ref|ZP_07665623.1| CRISPR-associated protein, Cas1 family [Streptococcus anginosus F0211]
gi|357636405|ref|ZP_09134281.1| CRISPR-associated protein Cas9/Csn1, subtype II/NMEN [Streptococcus massiae NCTC 11558]
gi|313890160|ref|ZP_07823795.1| CRISPR-associated protein, Csn1 family [Streptococcus pseudoporcinus SPIN 20026]
gi|156978435|ref|YP_002123879.1| CRISPR-Associated Protein Csn1 [Streptococcus equi subsp. zooepidemicus MGCS10565]
gi|94994317|ref|YP_602415.1| hypothetical protein MGAS10750_Spy0921 [Streptococcus pyogenes MGAS10750]
gi|251782637|ref|YP_002999940.1| hypothetical protein SDEG_1231 [Streptococcus dysgalactiae subsp. equisimilis GGS_124]
gi|21910213|ref|NP_664481.1| hypothetical protein SpyM1_0677 [Streptococcus pyogenes MGAS315]
gi|28896988|ref|NP_802838.1| hypothetical protein SPs1176 [Streptococcus pyogenes SSI-1]
gi|209559356|ref|YP_002285828.1| hypothetical protein Spy49_0823 [Streptococcus pyogenes NZ131]
gi|15676043|ref|NP_269215.1| hypothetical protein SPy_1045 [Streptococcus pyogenes M1 GAS]
gi|71910582|ref|YP_282132.1| hypothetical protein MGAS5005_Spy_0769 [Streptococcus pyogenes MGAS5005]
gi|94990335|ref|YP_598483.1| putative cytoplasmic protein [Streptococcus pyogenes MGAS10270]
gi|94988116|ref|YP_596617.1| cytoplasmic protein [Streptococcus pyogenes MGAS9429]
gi|94992340|ref|YP_600439.1| putative cytoplasmic protein [Streptococcus pyogenes MGAS2096]
gi|71903413|ref|YP_280216.1| cytoplasmic protein [Streptococcus pyogenes MGAS6180]

FIG. 4F

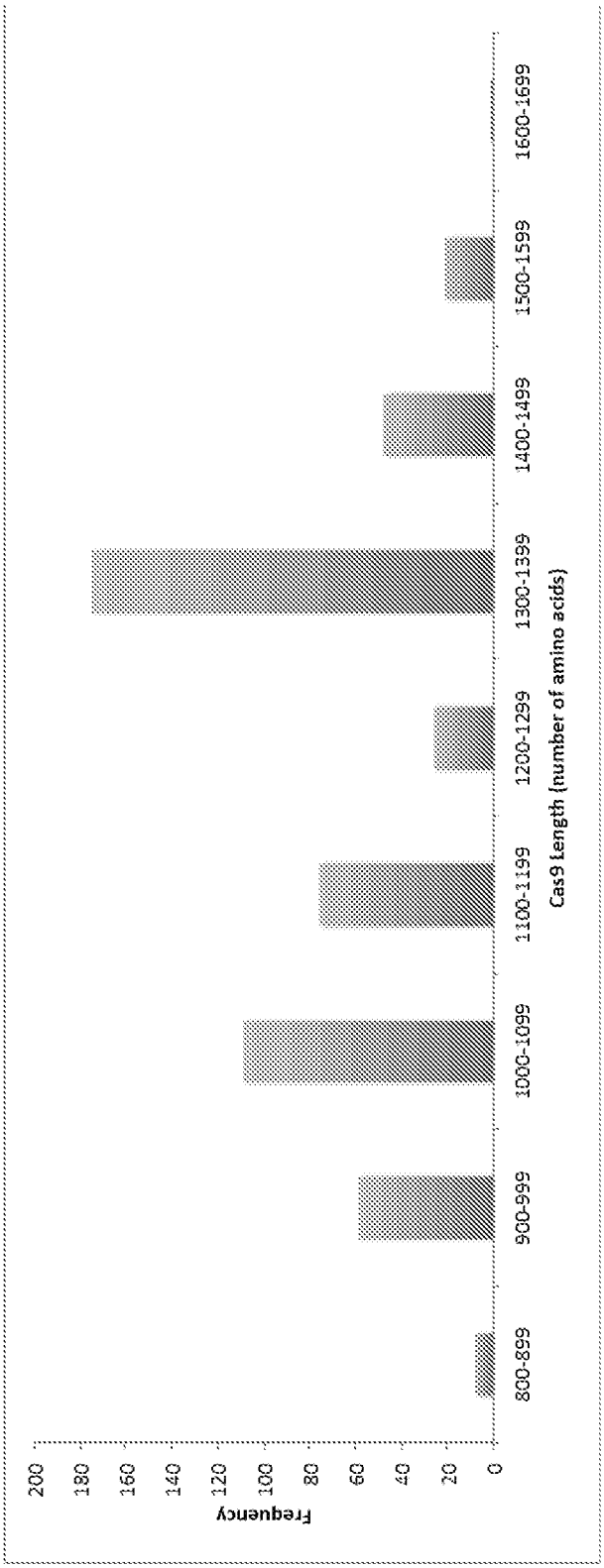

Frequency
200
180
160
140
120
100
80
60
40
20
0
800-899
900-999
1000-1099
1100-1199
1200-1299
1300-1399
1400-1499
1500-1599
1600-1699
Cas9 Length (number of amino acids)

FIG. 5

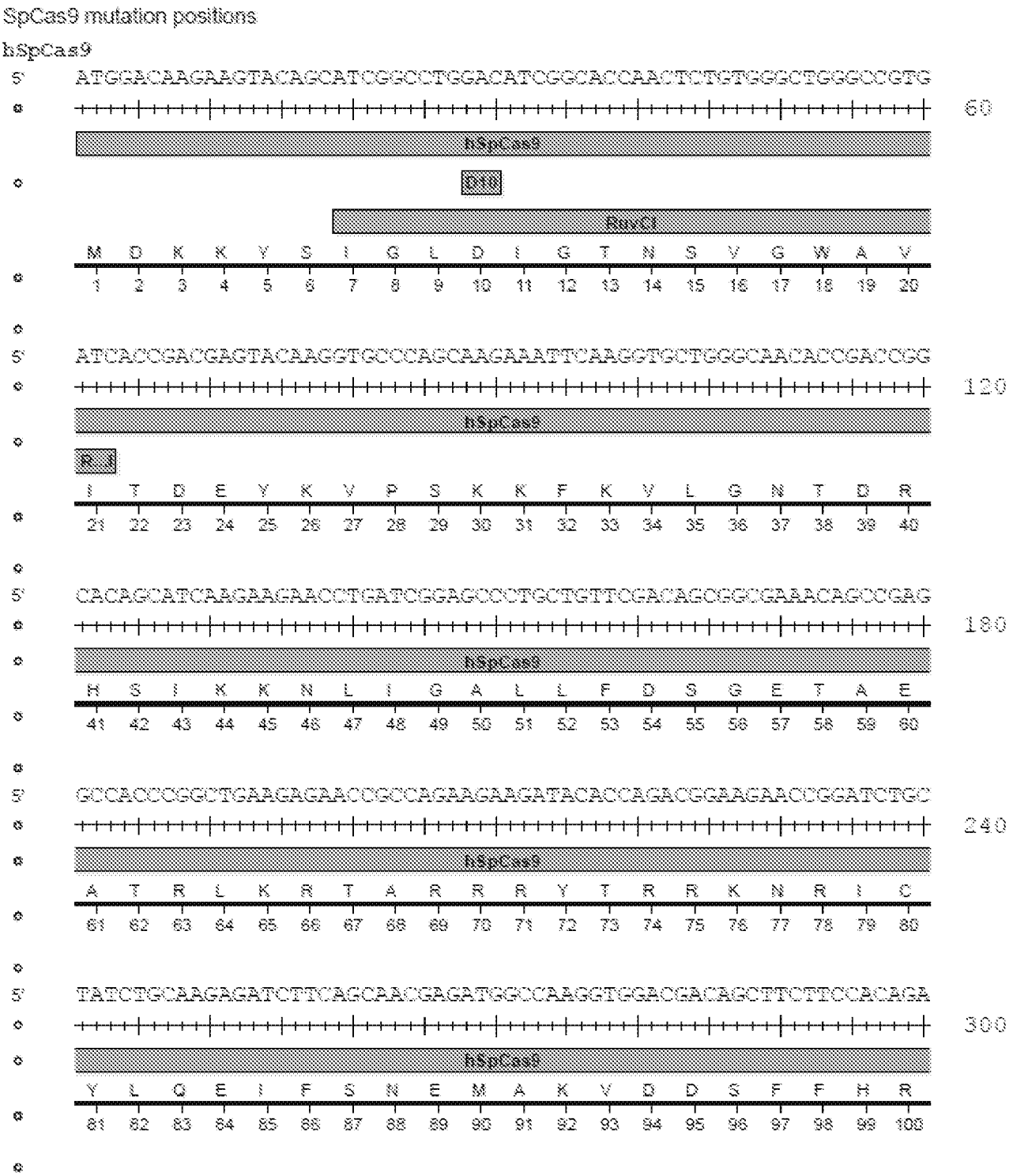

SpCas9 mutation positions
hSpCas9
5'
ATGGACAAGAAGTACAGCATCGGCCTGGACATCGGCACCAACTCTGTGGGCTGGGCCGTG
60
hSpCas9
D10
RuvC1
M D K K Y S I G L D I G T N S V G W A V
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20
5'
ATCACCGACGAGTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGG
120
hSpCas9
R_1
I T D E Y K V P S K K F K V L G N T D R
21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40
5'
CACAGCATCAAGAAGAACCTGATCGGAGCCCTGCTGTTCGACAGCGGCGAAACAGCCGAG
180
hSpCas9
H S I K K N L I G A L L F D S G E T A E
41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60
5'
GCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCGGATCTGC
240
hSpCas9
A T R L K R T A R R R Y T R R K N R I C
61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80
5'
TATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGA
300
hSpCas9
Y L Q E I F S N E M A K V D D S F F H R
81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100

FIG. 6A

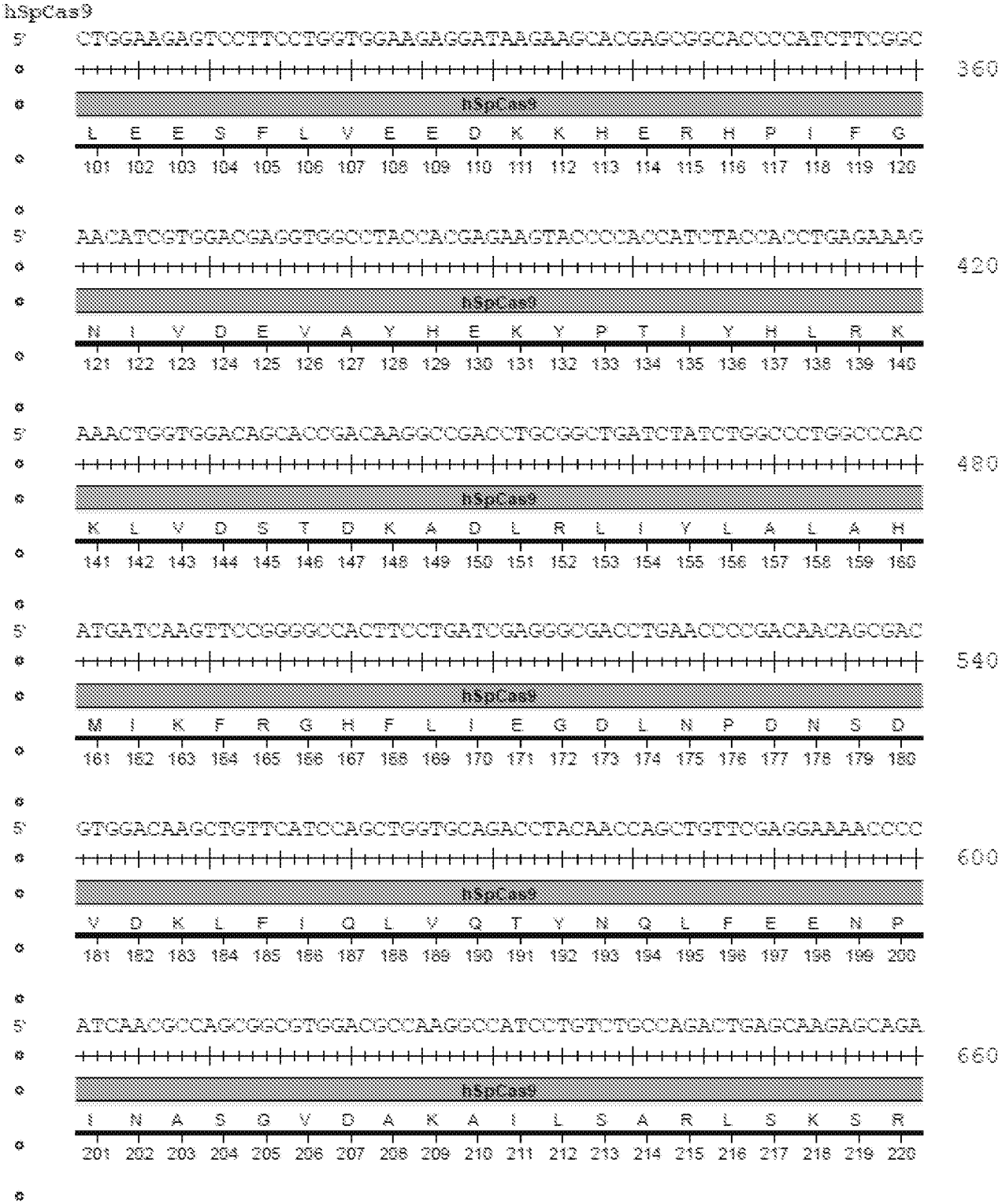

hSpCas9
5' CTGGAAGAGTCCTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGC
360
hSpCas9
L E E S F L V E E D K K H E R H P I F G
101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120
5' AACATCGTGGACGAGGTGGCCTACCACGAGAAGTACCCCACCATCTACCACCTGAGAAAAG
420
hSpCas9
N I V D E V A Y H E K Y P T I Y H L R K
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140
5' AAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCTGGCCCAC
480
hSpCas9
K L V D S T D K A D L R L I Y L A L A H
141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160
5' ATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGAC
540
hSpCas9
M I K F R G H F L I E G D L N P D N S D
161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180
5' GTGGACAAGCTGTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCC
600
hSpCas9
V D K L F I Q L V Q T Y N Q L F E E N P
181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200
5' ATCAACGCCAGCGGCGTGGACGCCAAGGCCATCCTGTCTGCCAGACTGAGCAAGAGCAGA
660
hSpCas9
I N A S G V D A K A I L S A R L S K S R
201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220

FIG. 6B

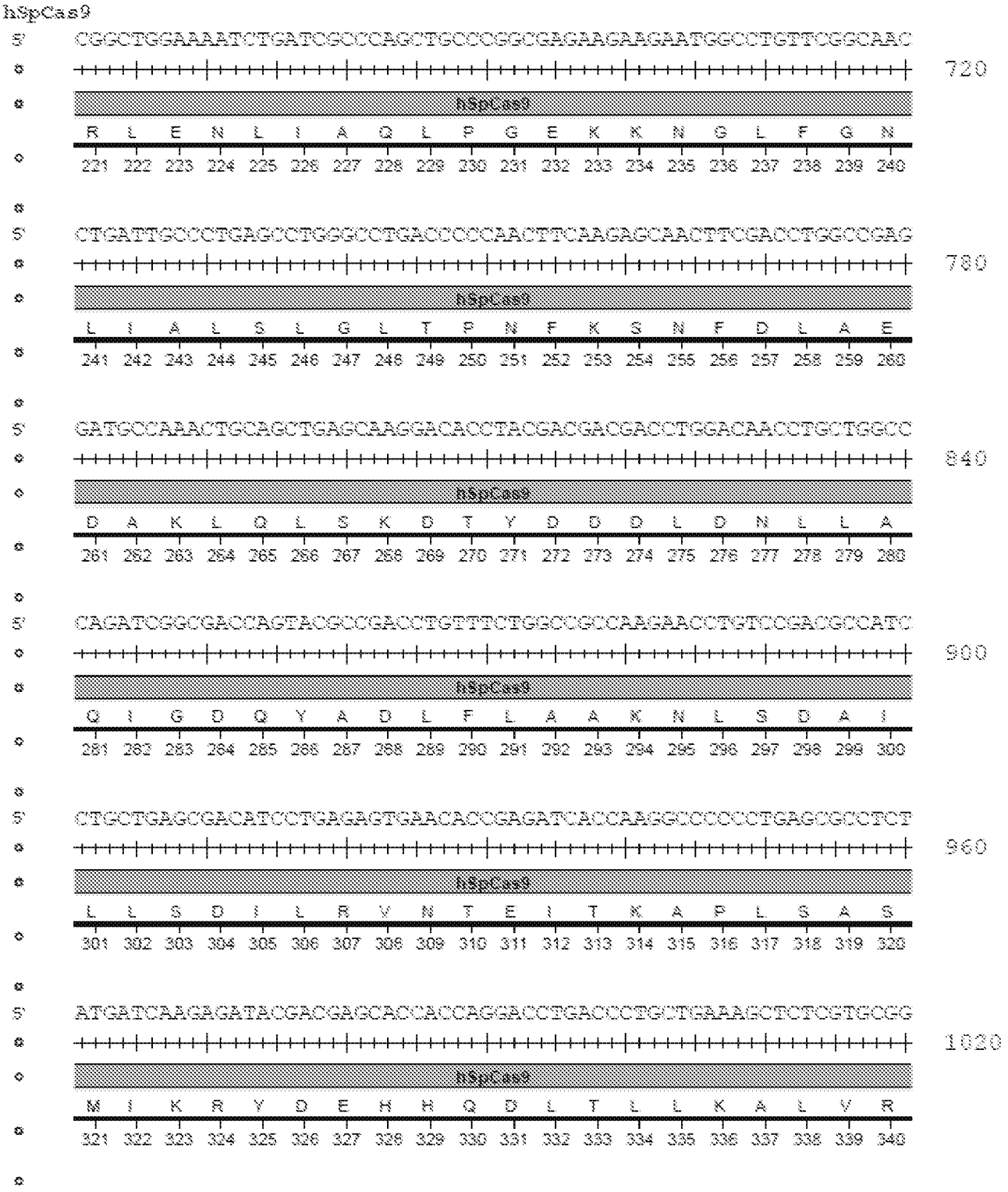

hSpCas9
5' CGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTTCGGCAAC 720
hSpCas9
R L E N L I A Q L P G E K K N G L F G N
221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240
5' CTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAG 780
hSpCas9
L I A L S L G L T P N F K S N F D L A E
241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260
5' GATGCCAAACTGCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCC 840
hSpCas9
D A K L Q L S K D T Y D D D L D N L L A
261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280
5' CAGATCGGCGACCAGTACGCCGACCTGTTTCTGGCCGCCAAGAACCTGTCCGACGCCATC 900
hSpCas9
Q I G D Q Y A D L F L A A K N L S D A I
281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300
5' CTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAGCGCCTCT 960
hSpCas9
L L S D I L R V N T E I T K A P L S A S
301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320
5' ATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGG 1020
hSpCas9
M I K R Y D E H H Q D L T L L K A L V R
321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340

FIG. 6C

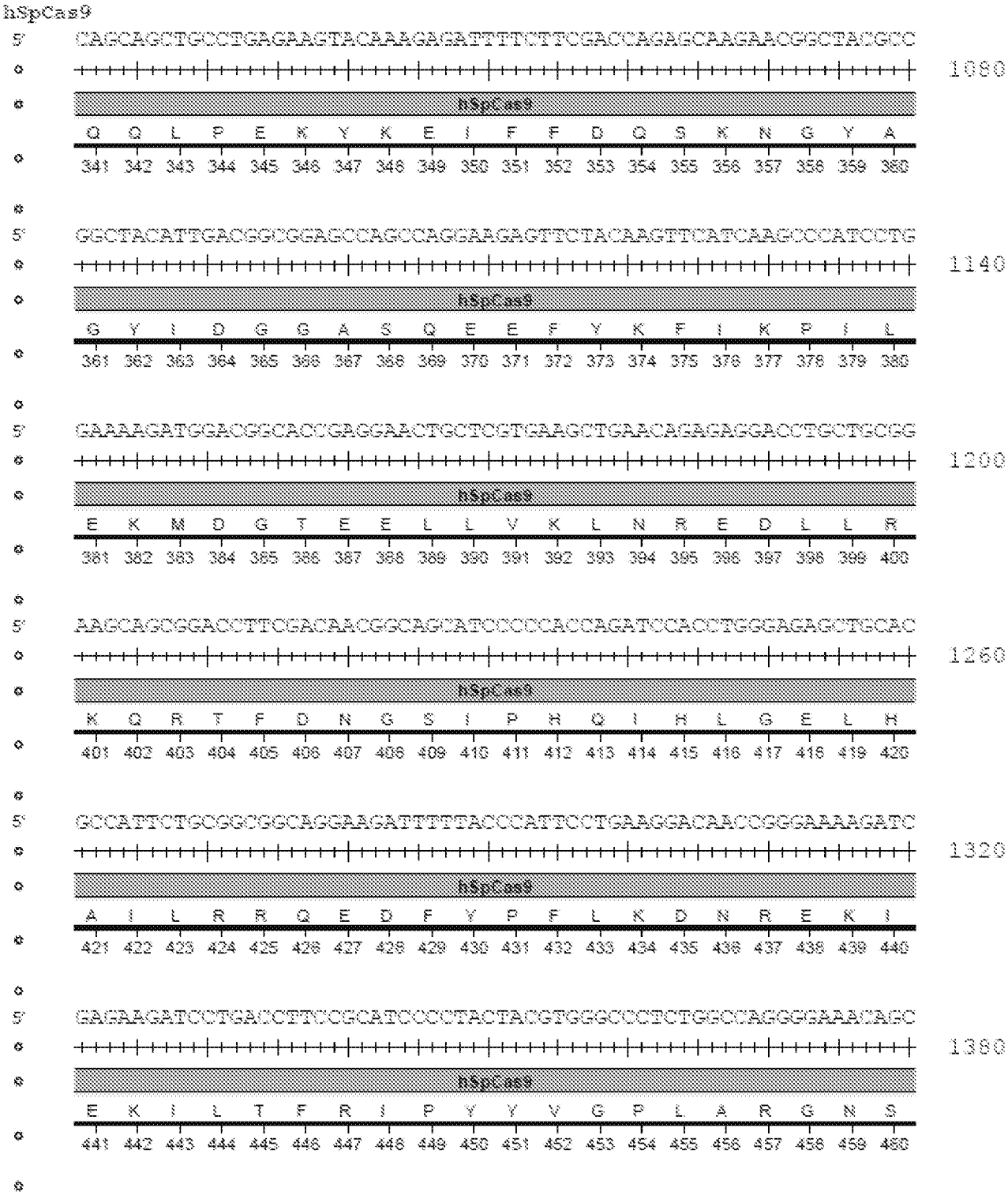

hSpCas9
5' CAGCAGCTGCCTGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCC
1080
hSpCas9
Q Q L P E K Y K E I F F D Q S K N G Y A
341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360
5' GGCTACATTGACGGCGGAGCCAGCCAGGAAGAGTTCTACAAGTTCATCAAGCCCATCCTG
1140
hSpCas9
G Y I D G G A S Q E E F Y K F I K P I L
361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380
5' GAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCTGCTGCGG
1200
hSpCas9
E K M D G T E E L L V K L N R E D L L R
381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400
5' AAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCAC
1260
hSpCas9
K Q R T F D N G S I P H Q I H L G E L H
401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420
5' GCCATTCTGCGGCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATC
1320
hSpCas9
A I L R R Q E D F Y P F L K D N R E K I
421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440
5' GAGAAGATCCTGACCTTCCGCATCCCCTACTACGTGGGCCCTCTGGCCAGGGGAAACAGC
1380
hSpCas9
E K I L T F R I P Y Y V G P L A R G N S
441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460

FIG. 6D

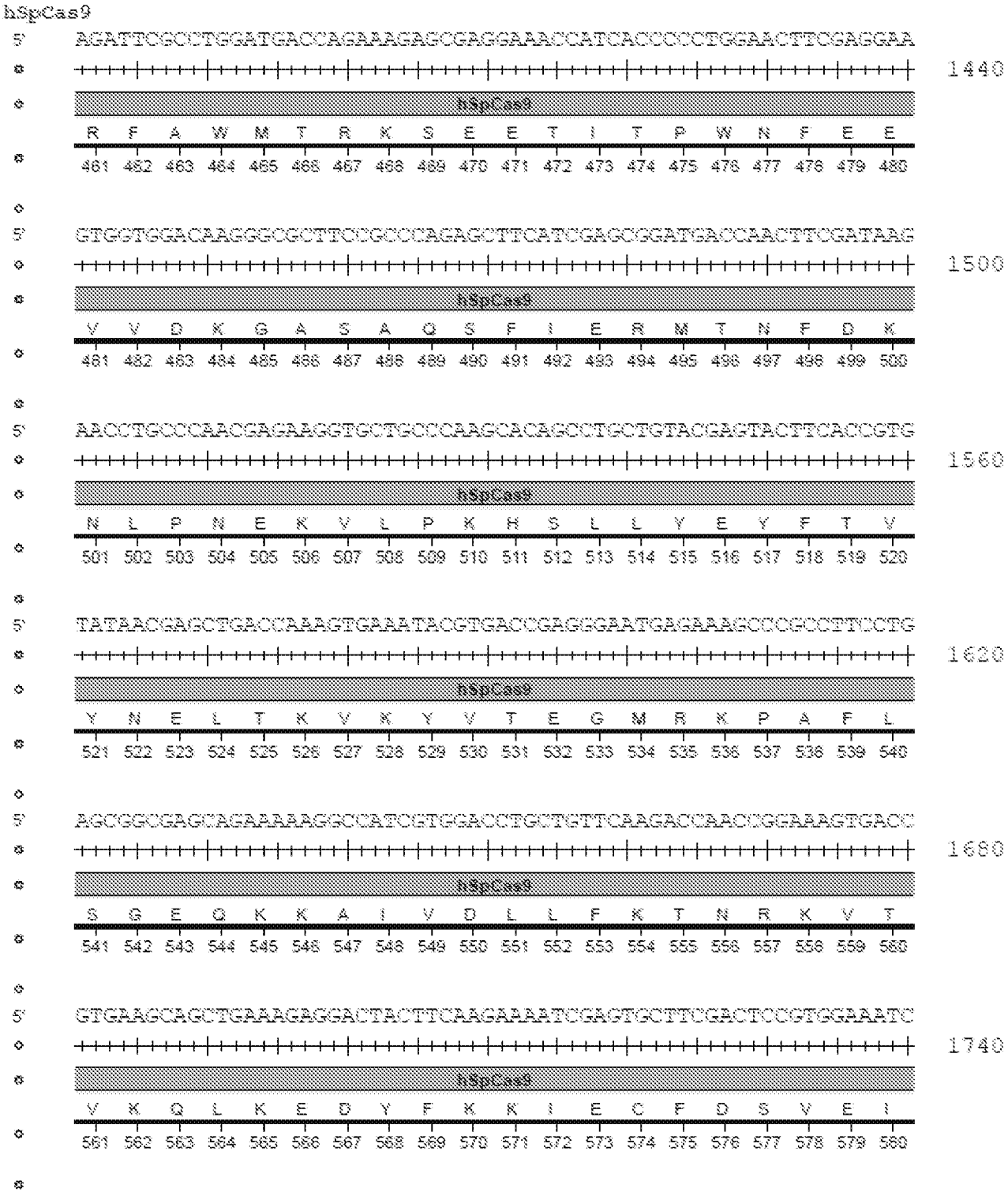

hSpCas9
5' AGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTTCGAGGAA
1440
hSpCas9
R F A W M T R K S E E T I T P W N F E E
461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480
5' GTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAG
1500
hSpCas9
V V D K G A S A Q S F I E R M T N F D K
481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500
5' AACCTGCCCAACGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTG
1560
hSpCas9
N L P N E K V L P K H S L L Y E Y F T V
501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520
5' TATAACGAGCTGACCAAAGTGAAATACGTGACCGAGGGAATGAGAAAGCCCGCCTTCCTG
1620
hSpCas9
Y N E L T K V K Y V T E G M R K P A F L
521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540
5' AGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAAAGTGACC
1680
hSpCas9
S G E Q K K A I V D L L F K T N R K V T
541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560
5' GTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATC
1740
hSpCas9
V K Q L K E D Y F K K I E C F D S V E I
561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580

FIG. 6E

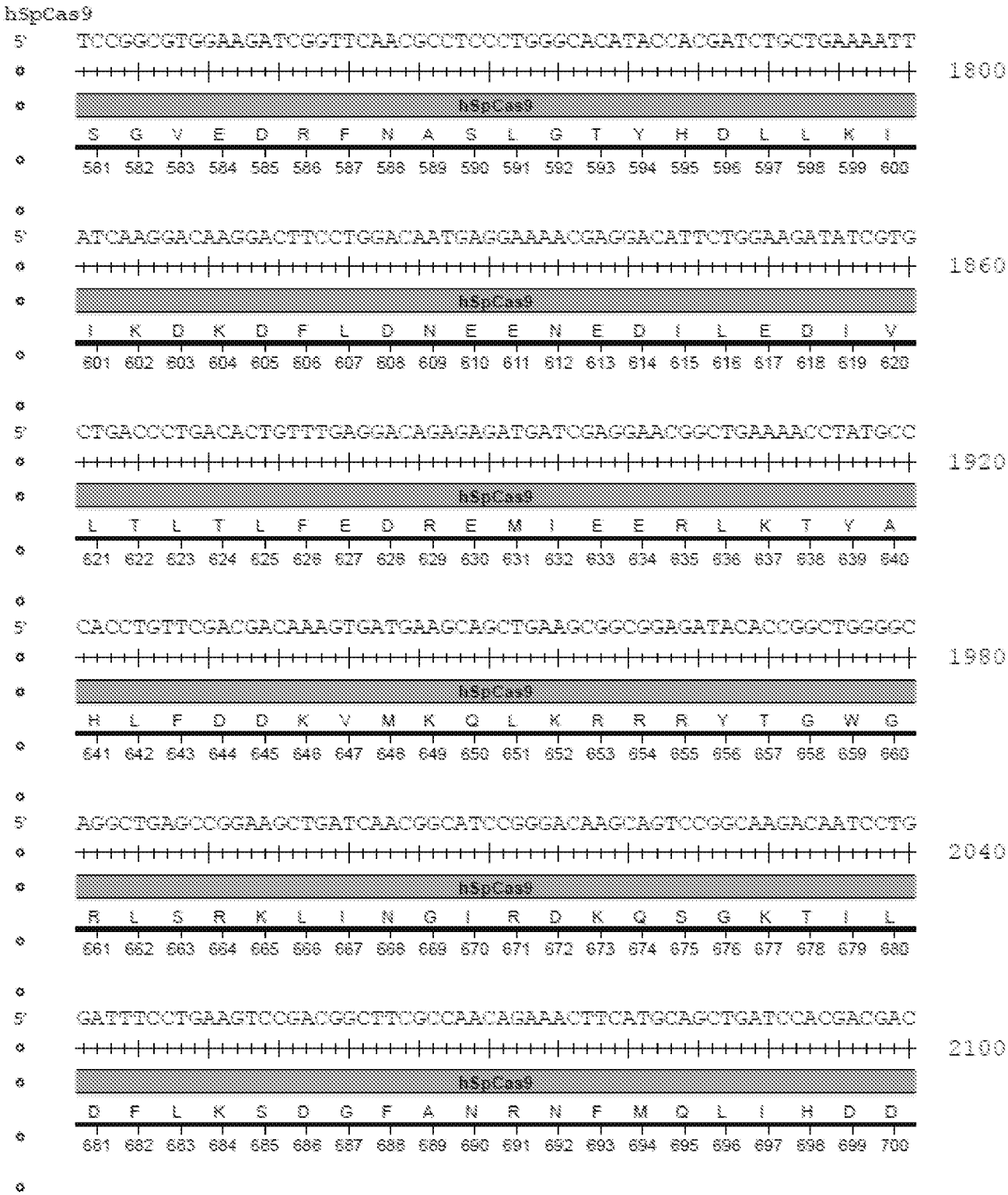

hSpCas9
5' TCCGGCGTGGAAGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATT
1800
hSpCas9
S G V E D R F N A S L G T Y H D L L K I
561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580
5' ATCAAGGACAAGGACTTCCTGGACAATGAGGAAAACGAGGACATTCTGGAAGATATCGTG
1860
hSpCas9
I K D K D F L D N E E N E D I L E D I V
581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600
5' CTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAACCTATGCC
1920
hSpCas9
L T L T L F E D R E M I E E R L K T Y A
621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640
5' CACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGC
1980
hSpCas9
H L F D D K V M K Q L K R R R Y T G W G
641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660
5' AGGCTGAGCCGGAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTG
2040
hSpCas9
R L S R K L I N G I R D K Q S G K T I L
661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680
5' GATTTCCTGAAGTCCGACGGCTTCGCCAACAGAAACTTCATGCAGCTGATCCACGACGAC
2100
hSpCas9
D F L K S D G F A N R N F M Q L I H D D
681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700

FIG. 6F

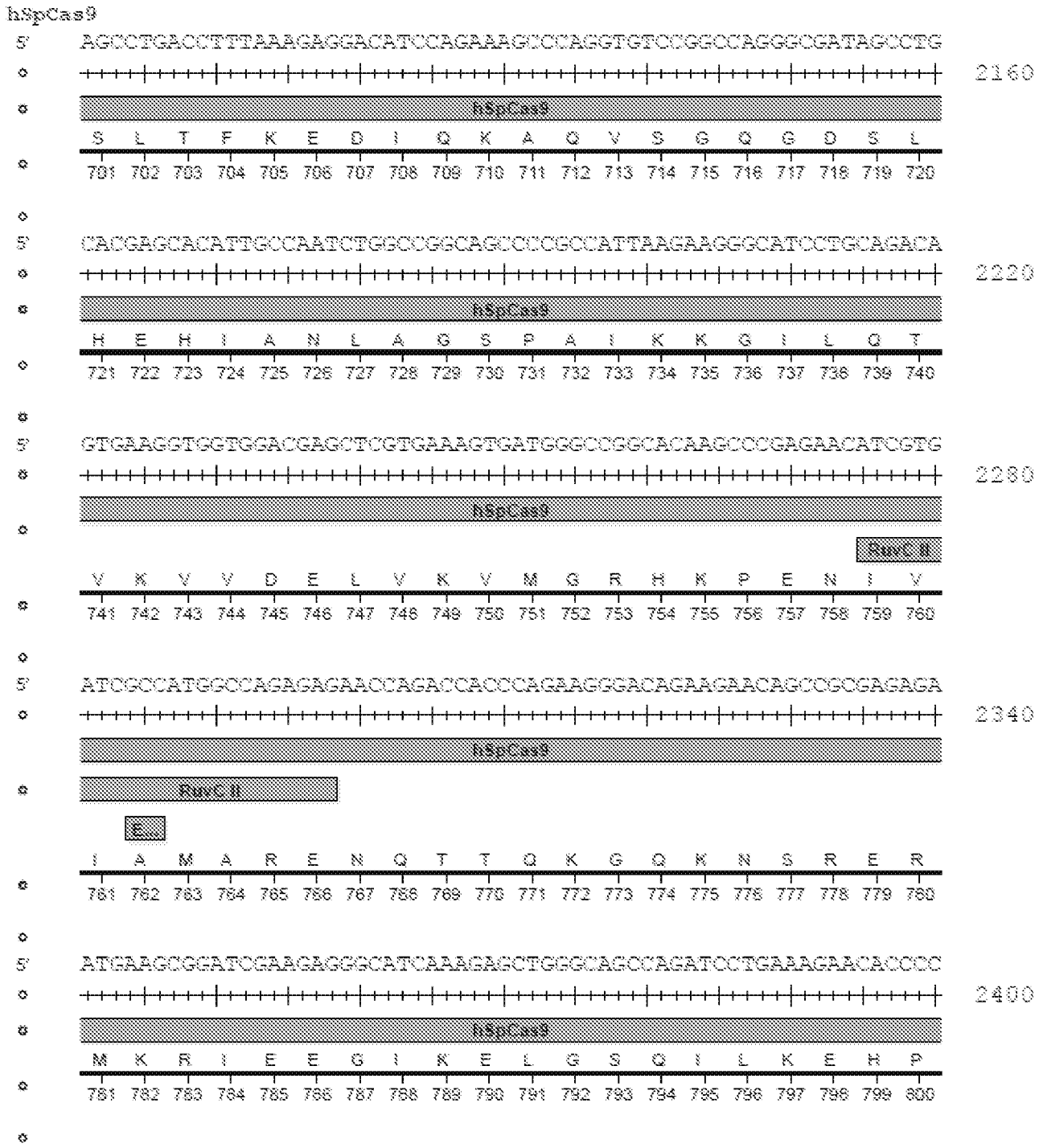

hSpCas9
5' AGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGATAGCCTG
2160
hSpCas9
S L T F K E D I Q K A Q V S G Q G D S L
721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740
5' CACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACA
2220
hSpCas9
H E H I A N L A G S P A I K K G I L Q T
721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740
5' GTGAAGGTGGTGGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTG
2280
hSpCas9
RuvC II
V K V V D E L V K V M G R H K P E N I V
741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760
5' ATCGCCATGGCCAGAGAGAACCAGACCACCCAGAAGGGACAGAAGAACAGCCGCGAGAGA
2340
hSpCas9
RuvC II
E...
I A M A R E N Q T T Q K G Q K N S R E R
761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780
5' ATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGAACACCCC
2400
hSpCas9
M K R I E E G I K E L G S Q I L K E H P
781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800

FIG. 6G

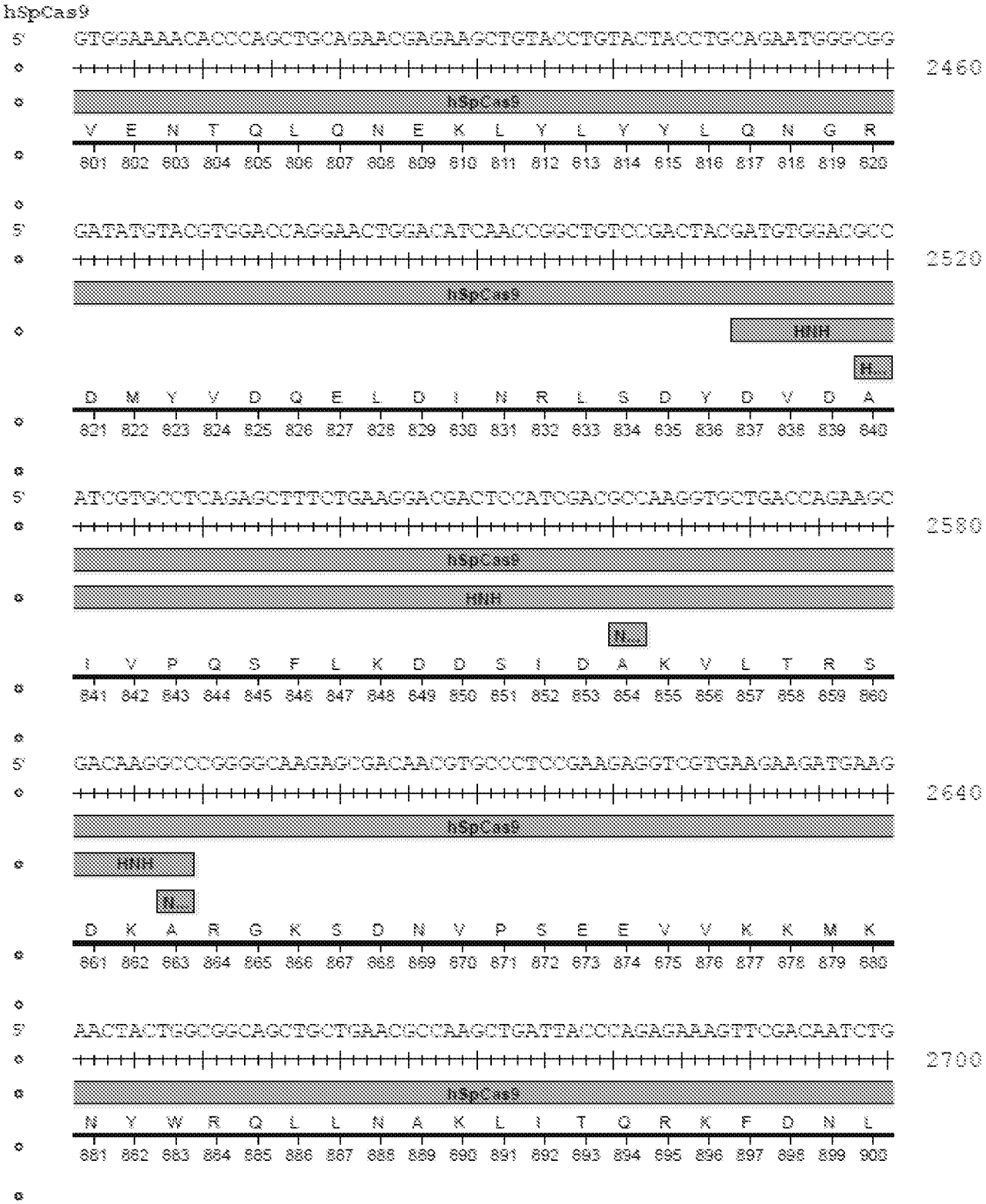

hSpCas9
5' GTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGG
2460
hSpCas9
V E N T Q L Q N E K L Y L Y Y L Q N G R
801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820
5' GATATGTACGTGGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACGCC
2520
hSpCas9
HNH
H...
D M Y V D Q E L D I N R L S D Y D V D A
821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840
5' ATCGTGCCTCAGAGCTTTCTGAAGGACGACTCCATCGACGCCAAGGTGCTGACCAGAAGC
2580
hSpCas9
HNH
N...
I V P Q S F L K D D S I D A K V L T R S
841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860
5' GACAAGGCCCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAAGATGAAG
2640
hSpCas9
HNH
N...
D K A R G K S D N V P S E E V V K K M K
861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880
5' AACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTG
2700
hSpCas9
N Y W R Q L L N A K L I T Q R K F D N L
881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900

FIG. 6H

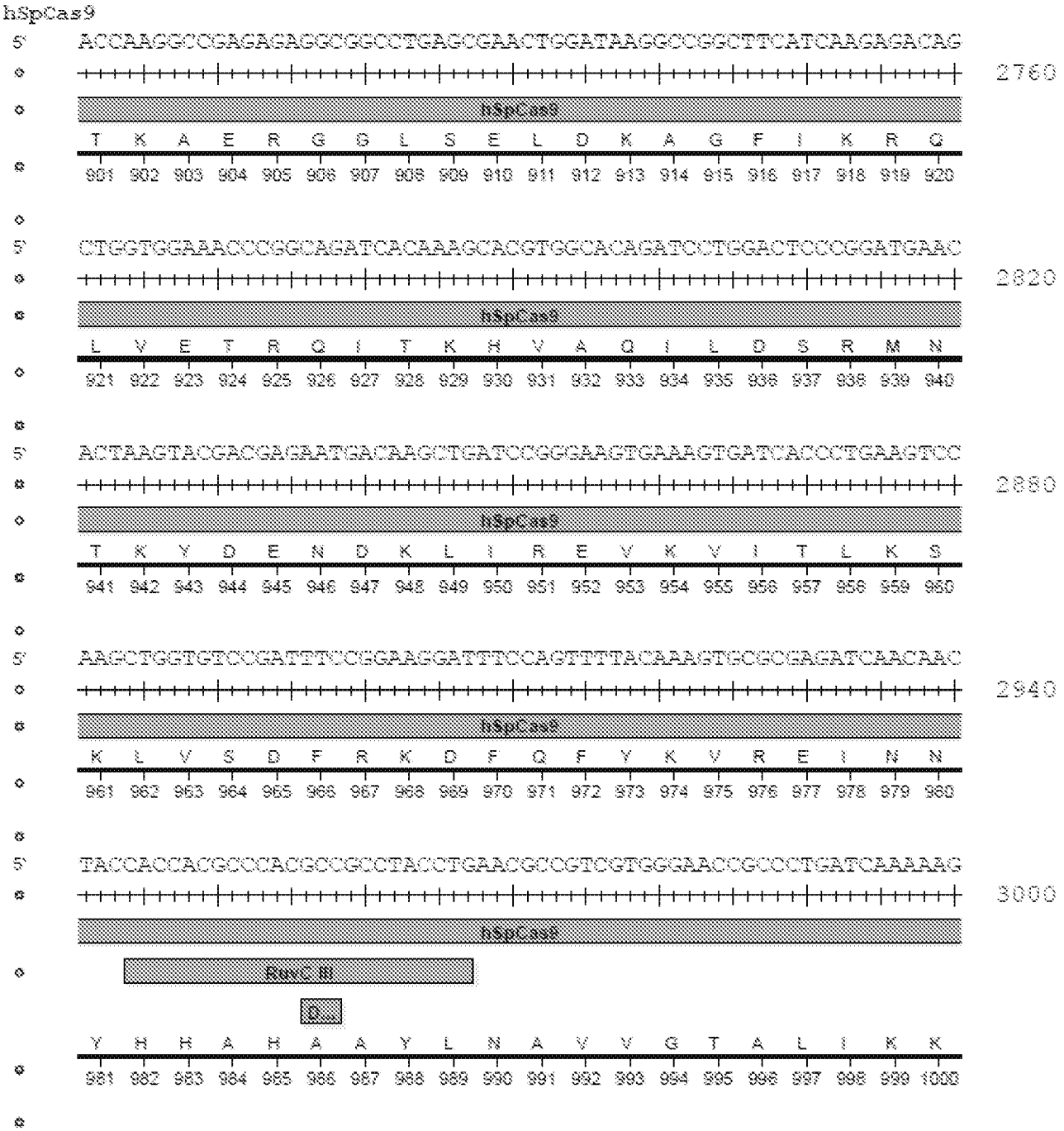

hSpCas9
5' ACCAAGGCCGAGAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAG
2760
hSpCas9
T K A E R G G L S E L D K A G F I K R Q
901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920
5' CTGGTGGAAACCCGGCAGATCACAAAGCACGTGGCACAGATCCTGGACTCCCGGATGAAC
2820
hSpCas9
L V E T R Q I T K H V A Q I L D S R M N
921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940
5' ACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCTGAAGTCC
2880
hSpCas9
T K Y D E N D K L I R E V K V I T L K S
941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960
5' AAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAAC
2940
hSpCas9
K L V S D F R K D F Q F Y K V R E I N N
961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980
5' TACCACCACGCCCACGCCGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAG
3000
hSpCas9
RuvC III
D...
Y H H A H A A Y L N A V V G T A L I K K
981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

FIG. 6I

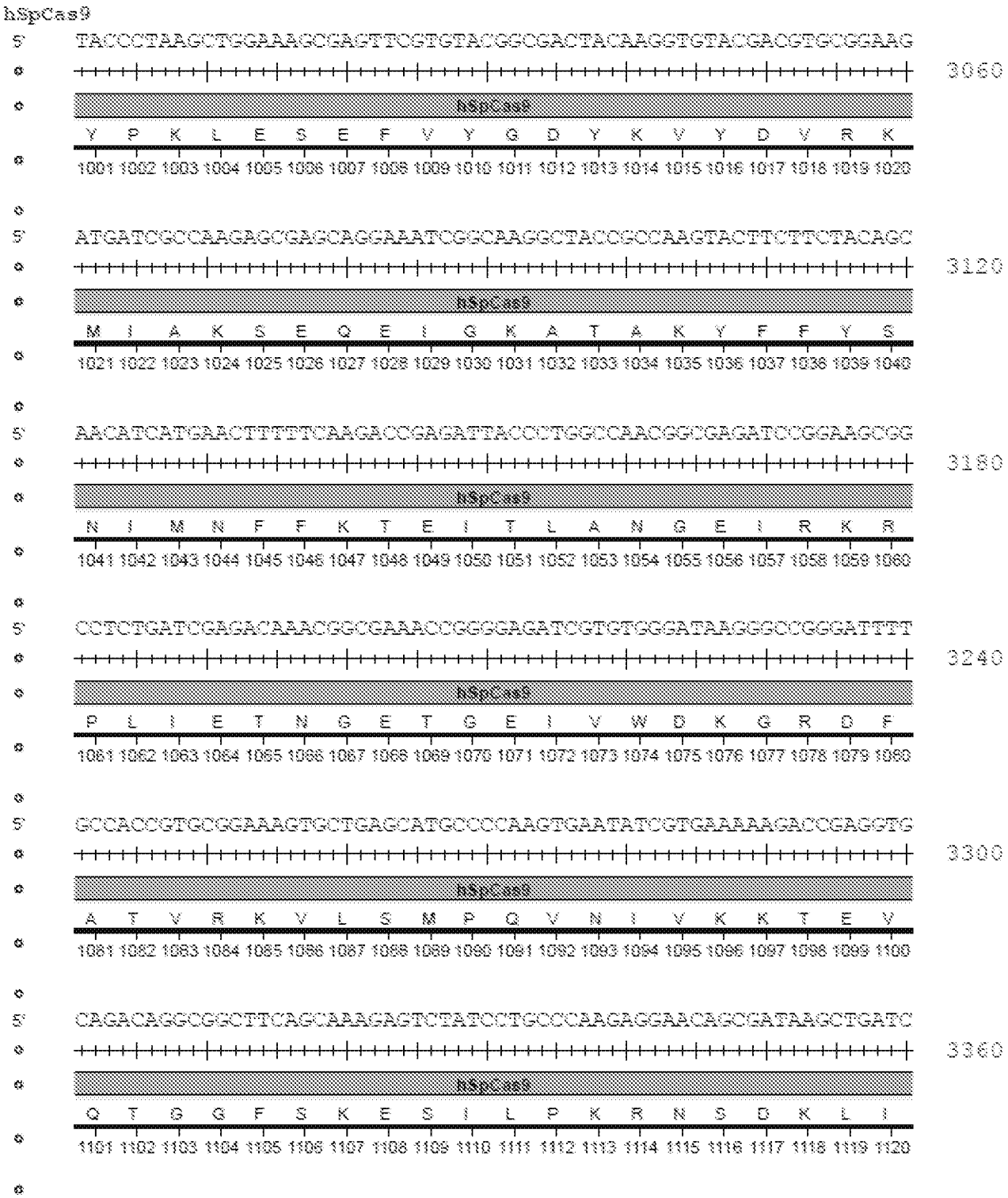

hSpCas9
5' TACCCTAAGCTGGAAAGCGAGTTCGTGTACGGCGACTACAAGGTGTACGACGTGCGGAAG
3060
nSpCas9
Y P K L E S E F V Y G D Y K V Y D V R K
1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020
5' ATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTTCTACAGC
3120
nSpCas9
M I A K S E Q E I G K A T A K Y F F Y S
1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040
5' AACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGG
3180
nSpCas9
N I M N F F K T E I T L A N G E I R K R
1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060
5' CCTCTGATCGAGACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTT
3240
nSpCas9
P L I E T N G E T G E I V W D K G R D F
1061 1062 1063 1064 1065 1066 1067 1068 1069 1070 1071 1072 1073 1074 1075 1076 1077 1078 1079 1080
5' GCCACCGTGCGGAAAGTGCTGAGCATGCCCCAAGTGAATATCGTGAAAAAGACCGAGGTG
3300
nSpCas9
A T V R K V L S M P Q V N I V K K T E V
1081 1082 1083 1084 1085 1086 1087 1088 1089 1090 1091 1092 1093 1094 1095 1096 1097 1098 1099 1100
5' CAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAAGCTGATC
3360
nSpCas9
Q T G G F S K E S I L P K R N S D K L I
1101 1102 1103 1104 1105 1106 1107 1108 1109 1110 1111 1112 1113 1114 1115 1116 1117 1118 1119 1120

FIG. 6J

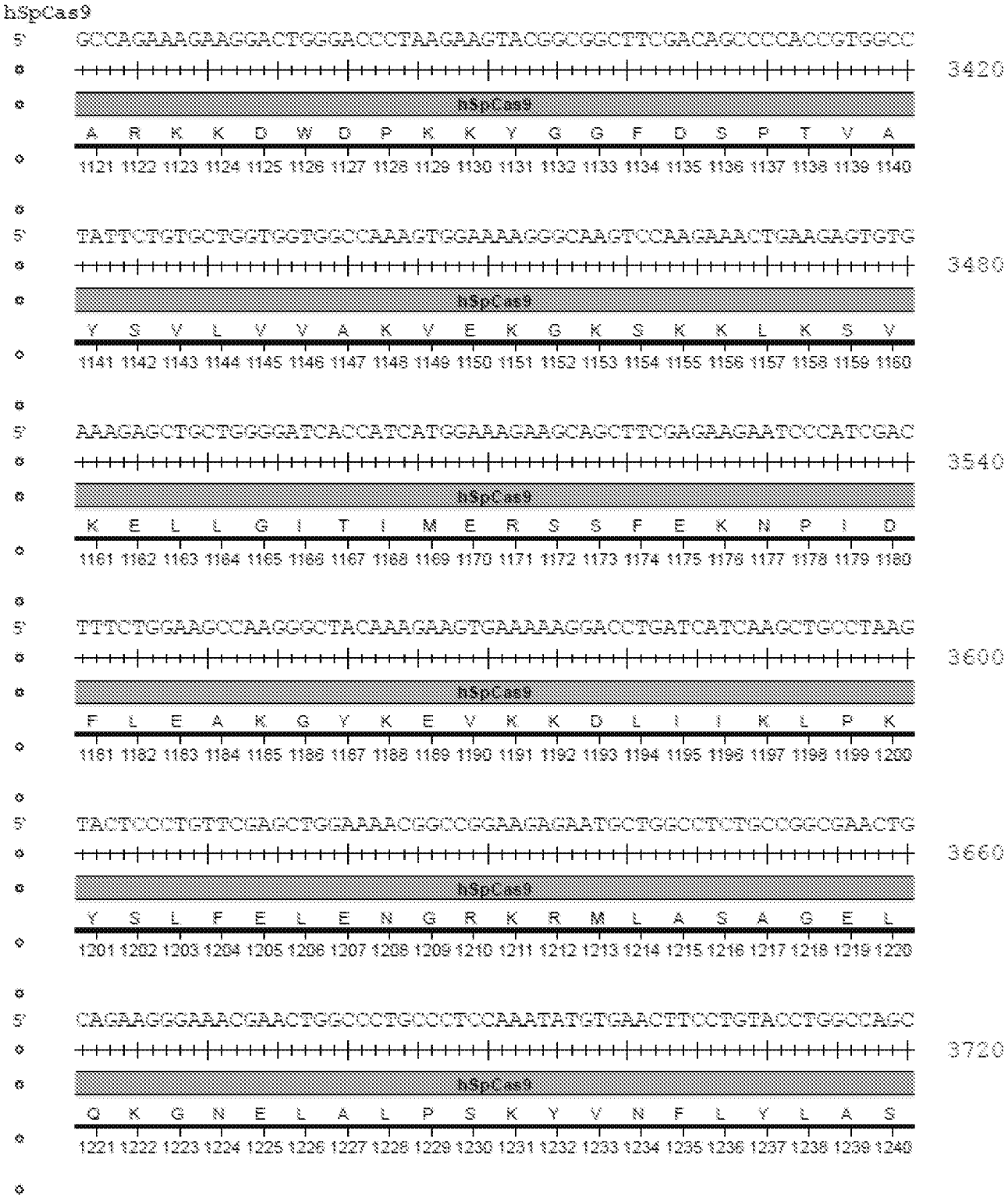

hSpCas9
5' GCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCC 3'
3420
hSpCas9
A R K K D W D P K K Y G G F D S P T V A
1121 1122 1123 1124 1125 1126 1127 1128 1129 1130 1131 1132 1133 1134 1135 1136 1137 1138 1139 1140
5' TATTCTGTGCTGGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTG 3'
3480
hSpCas9
Y S V L V V A K V E K G K S K K L K S V
1141 1142 1143 1144 1145 1146 1147 1148 1149 1150 1151 1152 1153 1154 1155 1156 1157 1158 1159 1160
5' AAAGAGCTGCTGGGGATCACCATCATGGAAAGAAGCAGCTTCGAGAAGAATCCCATCGAC 3'
3540
hSpCas9
K E L L G I T I M E R S S F E K N P I D
1161 1162 1163 1164 1165 1166 1167 1168 1169 1170 1171 1172 1173 1174 1175 1176 1177 1178 1179 1180
5' TTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCTGCCTAAG 3'
3600
hSpCas9
F L E A K G Y K E V K K D L I I K L P K
1181 1182 1183 1184 1185 1186 1187 1188 1189 1190 1191 1192 1193 1194 1195 1196 1197 1198 1199 1200
5' TACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTG 3'
3660
hSpCas9
Y S L F E L E N G R K R M L A S A G E L
1201 1202 1203 1204 1205 1206 1207 1208 1209 1210 1211 1212 1213 1214 1215 1216 1217 1218 1219 1220
5' CAGAAGGGAAACGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGC 3'
3720
hSpCas9
Q K G N E L A L P S K Y V N F L Y L A S
1221 1222 1223 1224 1225 1226 1227 1228 1229 1230 1231 1232 1233 1234 1235 1236 1237 1238 1239 1240

FIG. 6K

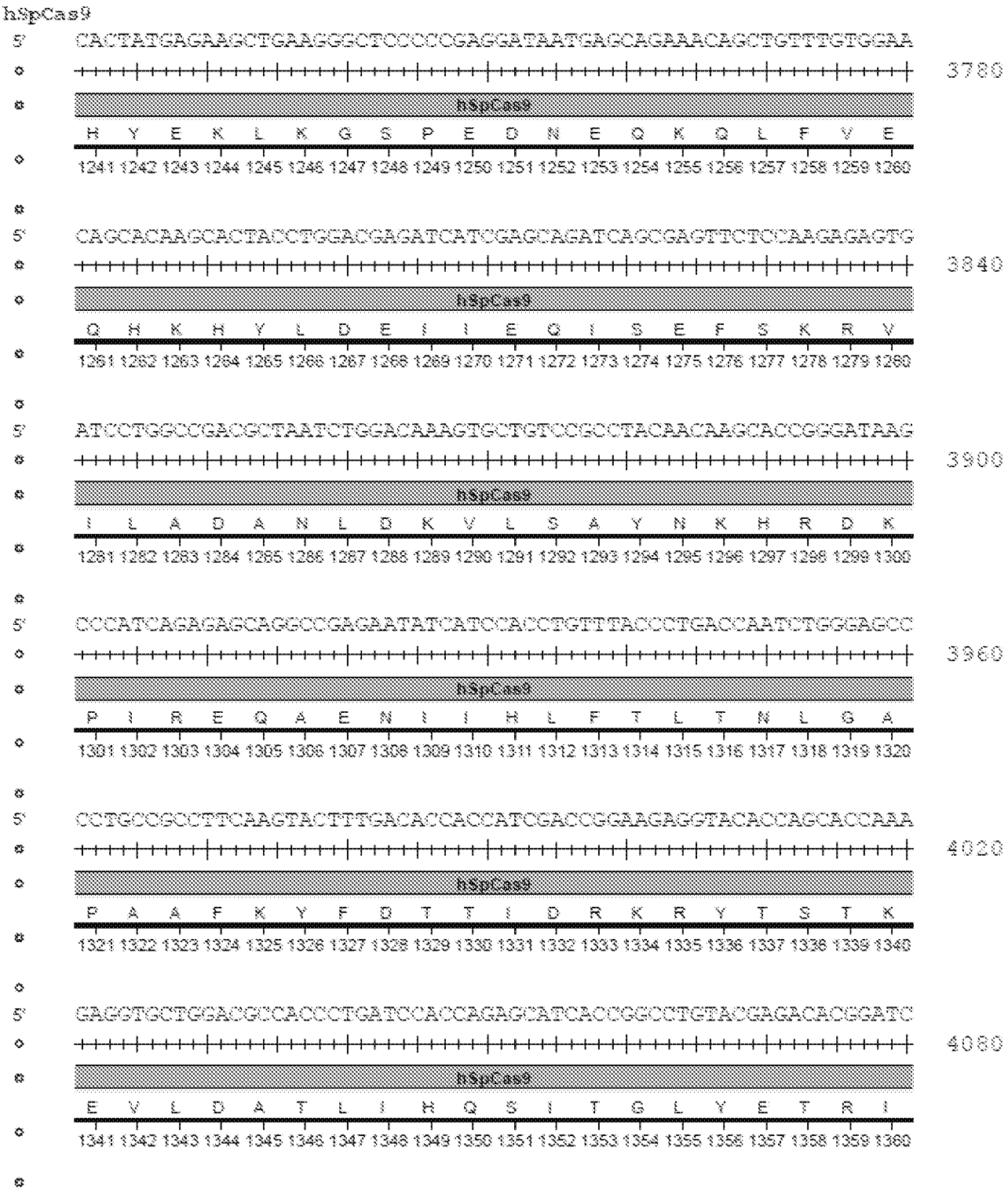

hSpCas9
5' CACTATGAGAAGCTGAAGGGCTCCCCCGAGGATAATGAGCAGAAACAGCTGTTTGTGGAA
3780
nSpCas9
H Y E K L K G S P E D N E Q K Q L F V E
1241 1242 1243 1244 1245 1246 1247 1248 1249 1250 1251 1252 1253 1254 1255 1256 1257 1258 1259 1260
5' CAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAAGAGAGTG
3840
nSpCas9
Q H K H Y L D E I I E Q I S E F S K R V
1261 1262 1263 1264 1265 1266 1267 1268 1269 1270 1271 1272 1273 1274 1275 1276 1277 1278 1279 1280
5' ATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAG
3900
nSpCas9
I L A D A N L D K V L S A Y N K H R D K
1281 1282 1283 1284 1285 1286 1287 1288 1289 1290 1291 1292 1293 1294 1295 1296 1297 1298 1299 1300
5' CCCATCAGAGAGCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCC
3960
nSpCas9
P I R E Q A E N I I H L F T L T N L G A
1301 1302 1303 1304 1305 1306 1307 1308 1309 1310 1311 1312 1313 1314 1315 1316 1317 1318 1319 1320
5' CCTGCCGCCTTCAAGTACTTTGACACCACCATCGACCGGAAGAGGTACACCAGCACCAAA
4020
nSpCas9
P A A F K Y F D T T I D R K R Y T S T K
1321 1322 1323 1324 1325 1326 1327 1328 1329 1330 1331 1332 1333 1334 1335 1336 1337 1338 1339 1340
5' GAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGACACGGATC
4080
nSpCas9
E V L D A T L I H Q S I T G L Y E T R I
1341 1342 1343 1344 1345 1346 1347 1348 1349 1350 1351 1352 1353 1354 1355 1356 1357 1358 1359 1360

FIG. 6L

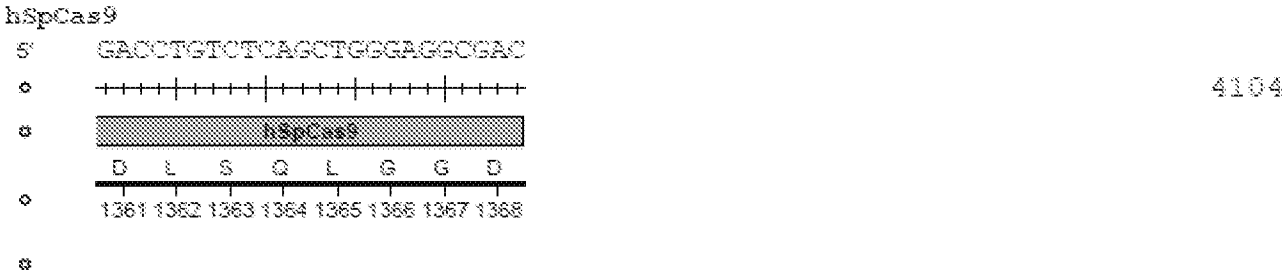

hSpCas9
5'
GACCTGTCTCAGCTGGGAGGCGAC
4104
hSpCas9
D L S Q L G G D
1361 1362 1363 1364 1365 1366 1367 1368

FIG. 6M

Cas9, Rosa26 targeting vector map

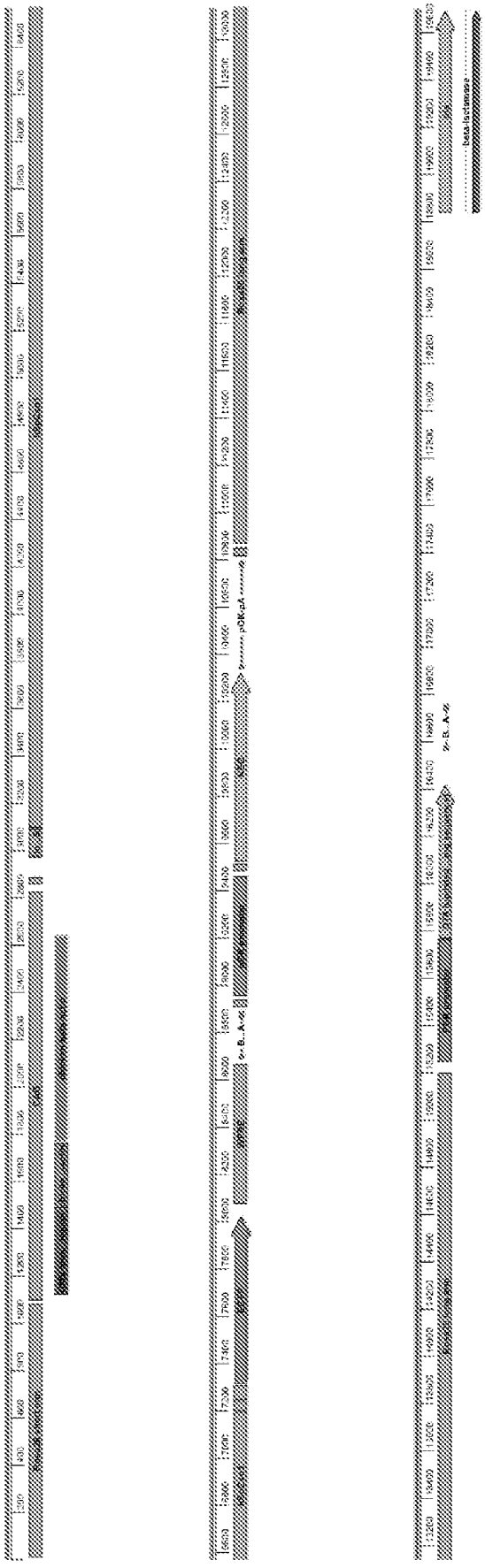


FIG. 7A

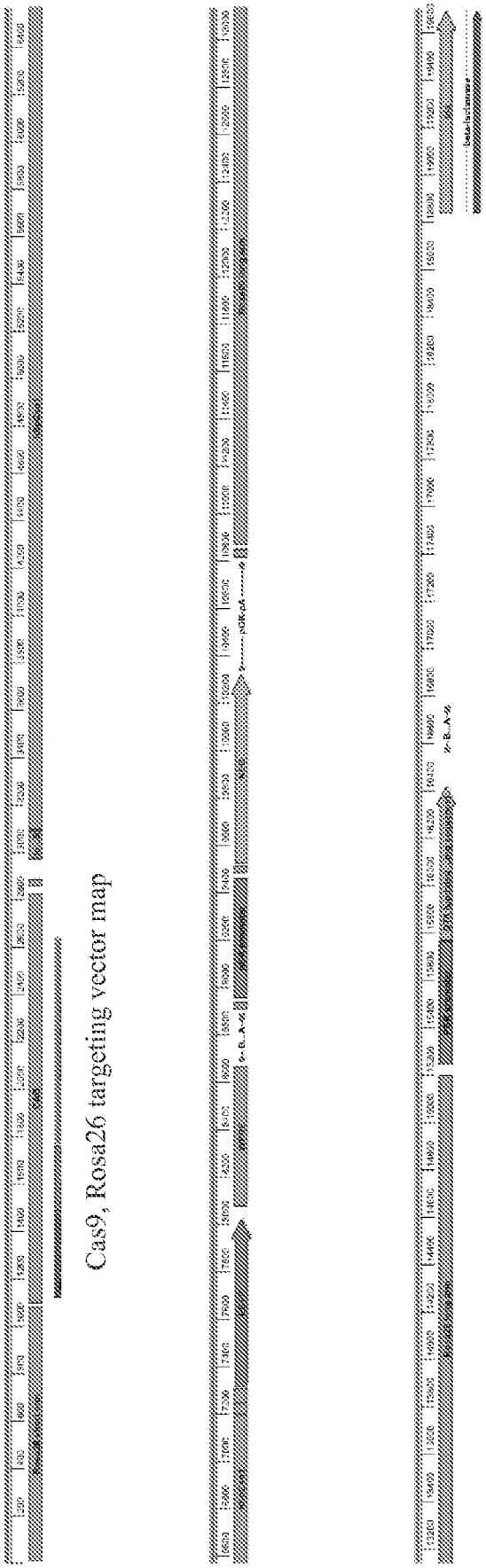


FIG. 7B

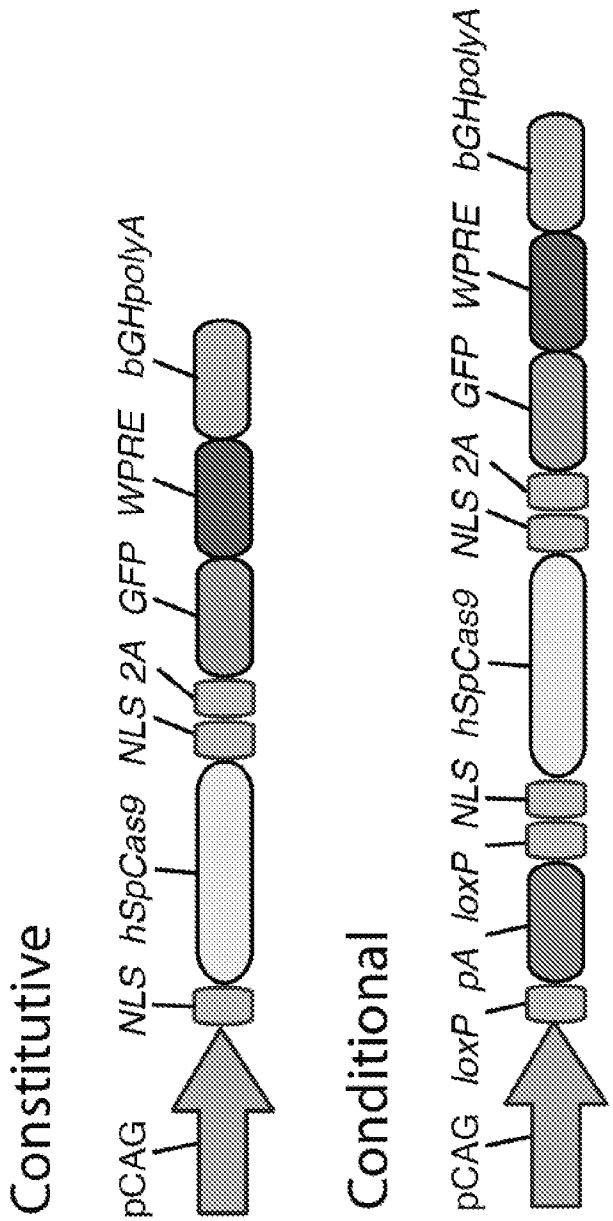
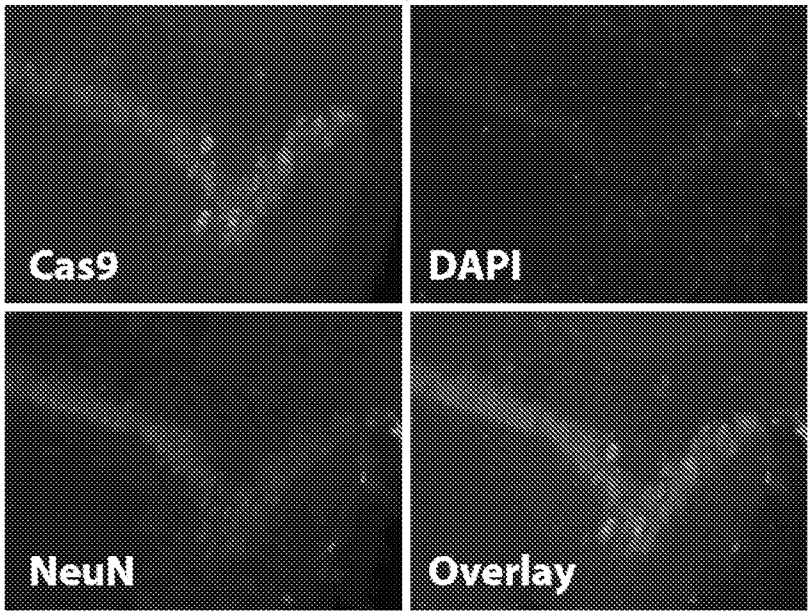

Constitutive
pCAG
NLS
hSpCas9
NLS
2A
GFP
WPRE
bGHpolyA
Conditional
pCAG
loxP
pA
loxP
NLS
hSpCas9
NLS
2A
GFP
WPRE
bGHpolyA

FIG. 8

Cas9 Expression in Mouse Hippocampus (AAV)



Cas9 Expression in Mouse Cortex (AAV)

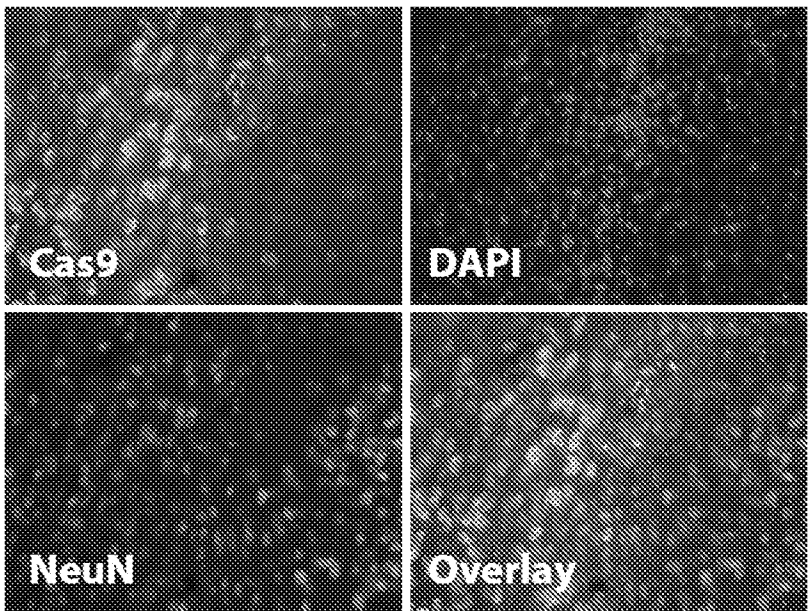


FIG. 9

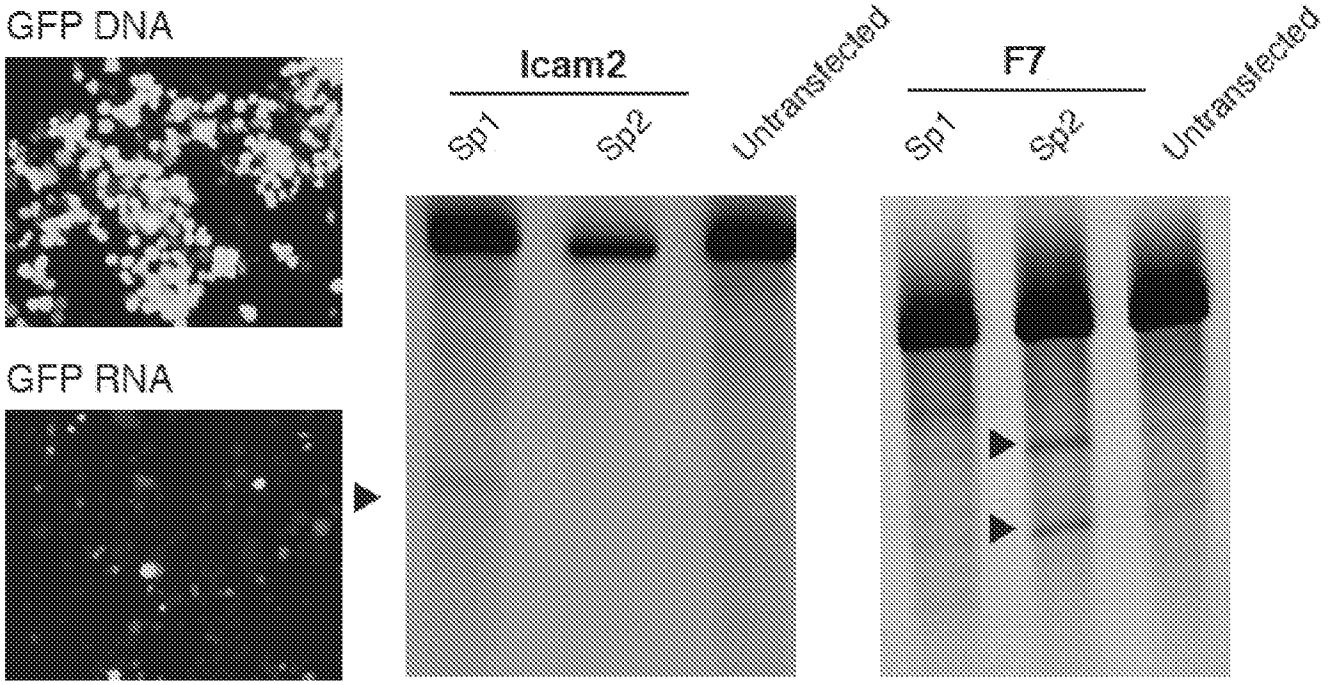

GFP DNA
GFP RNA
Icam2
Sp1
Sp2
Untransfected
F7
Sp1
Sp2
Untransfected

FIG. 10

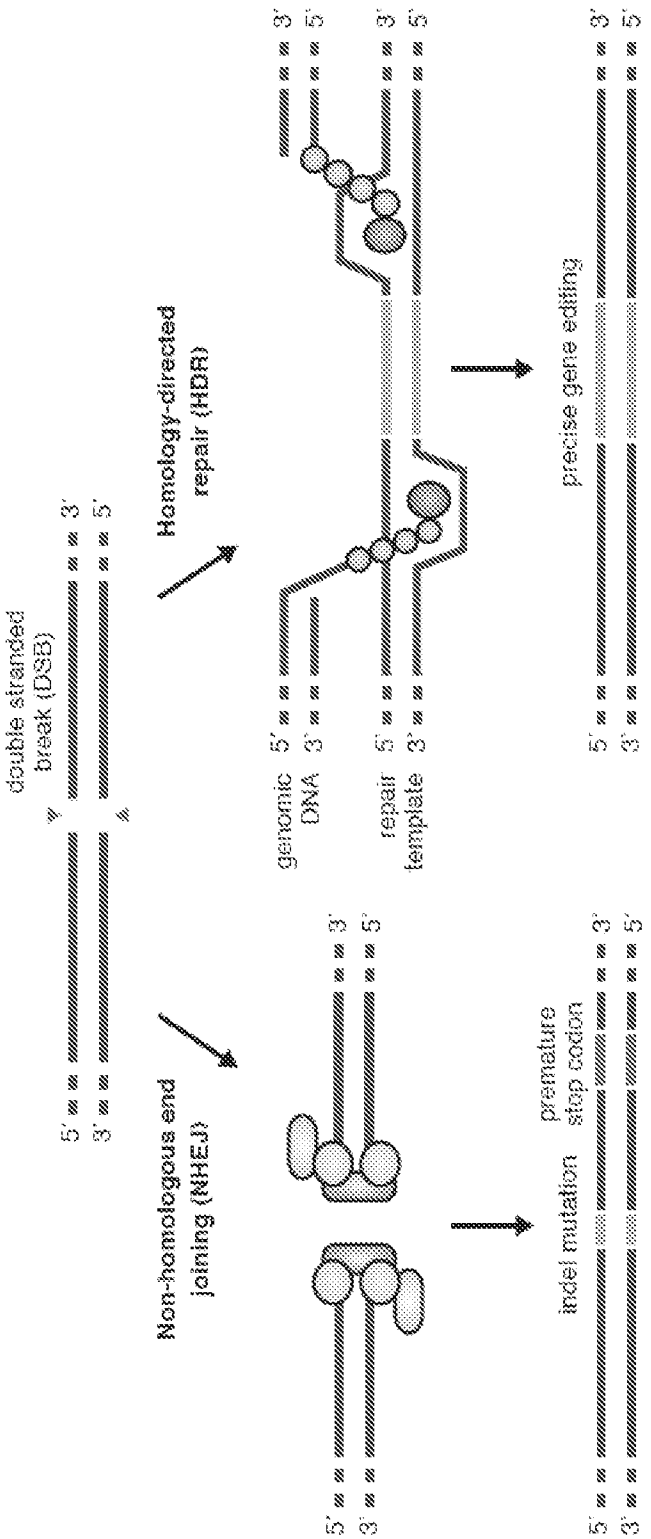

double stranded break (DSB)
5'
3'
Homology-directed repair (HDR)
genomic DNA
repair template
precise gene editing
Non-homologous end joining (NHEJ)
indel mutation
premature stop codon

FIG. 11

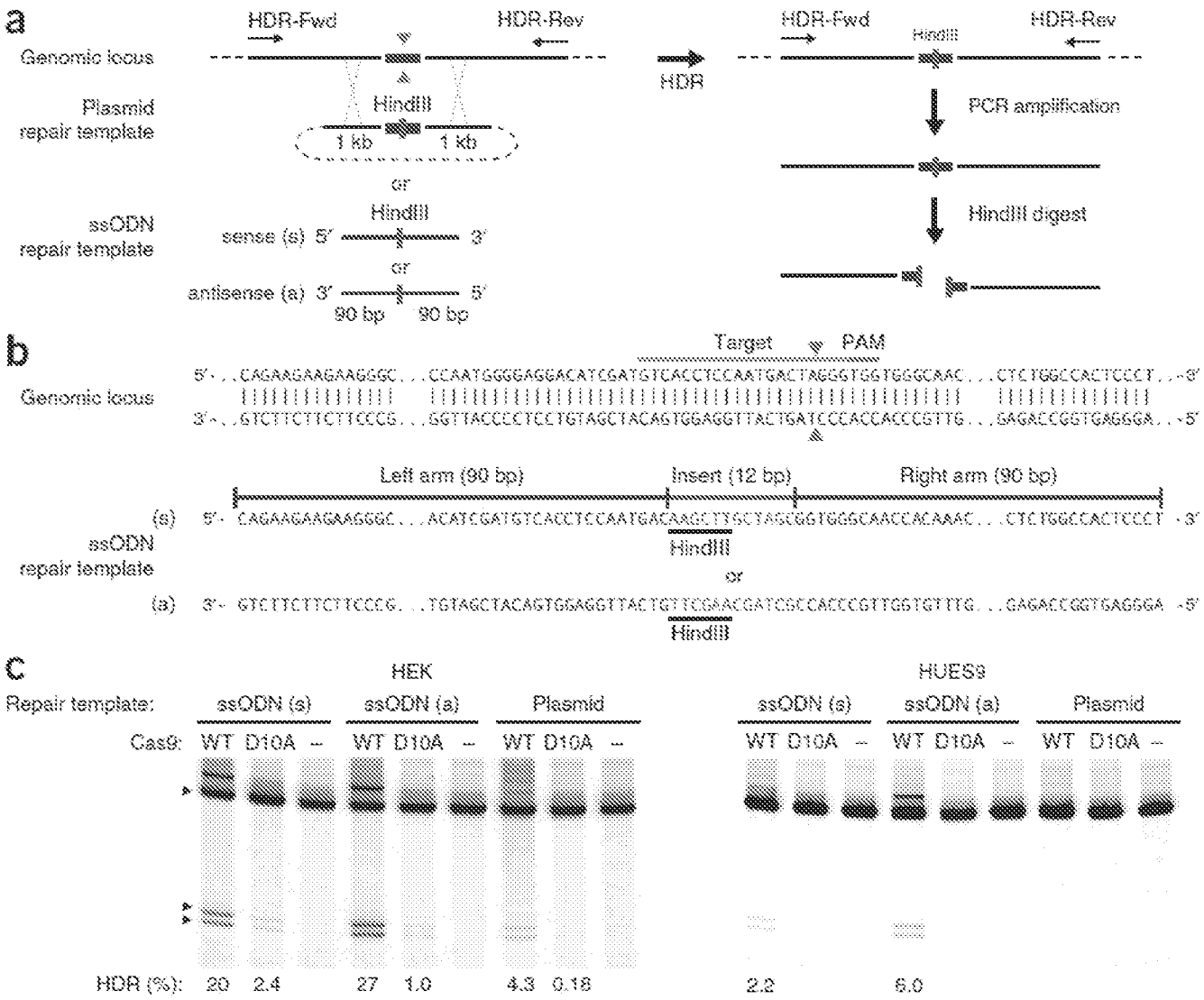

a
HDR-Fwd
HDR-Rev
Genomic locus
Plasmid repair template
HindIII
1 kb
1 kb
or
ssODN repair template
sense (s) 5′
3′
antisense (a) 3′
5′
90 bp
90 bp
HDR
PCR amplification
HindIII digest
b
Target
PAM
Genomic locus
Left arm (90 bp)
Insert (12 bp)
Right arm (90 bp)
(s)
(a)
ssODN repair template
HindIII
or
c
HEK
HUES9
Repair template:
ssODN (s)
Plasmid
Cas9:
WT
D10A
–
HDR (%):
20
2.4
27
1.0
4.3
0.18
2.2
6.0

FIG. 12A-C

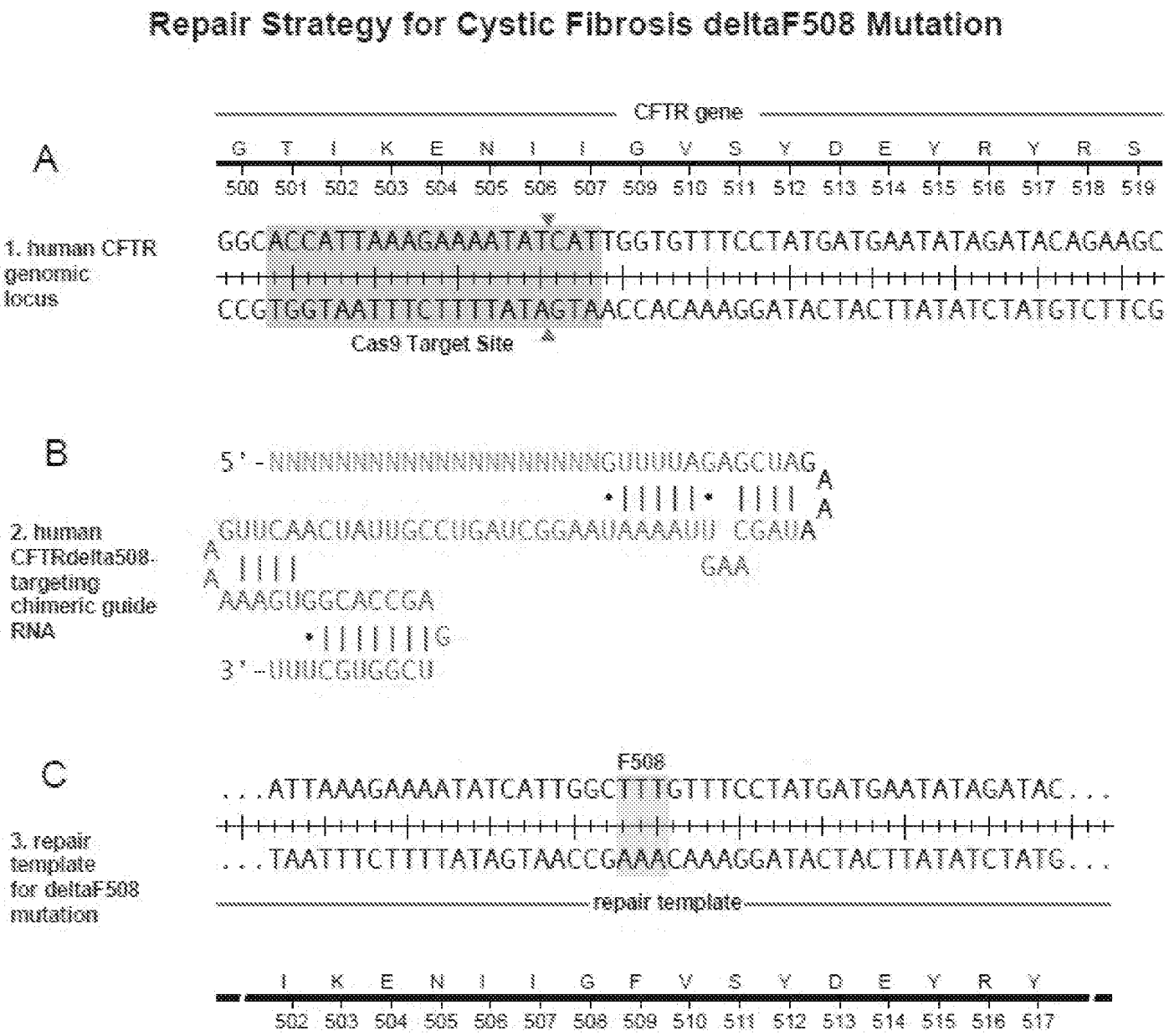
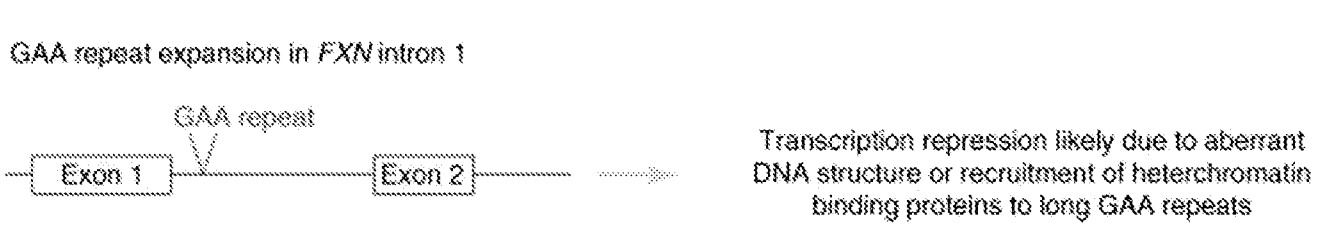

Repair Strategy for Cystic Fibrosis deltaF508 Mutation
A
CFTR gene
G T I K E N I I G V S Y D E Y R Y R S
500 501 502 503 504 505 506 507 509 510 511 512 513 514 515 516 517 518 519
1. human CFTR genomic locus
GGCACCATTAAAGAAAATATCATTGGTGTTTCCTATGATGAATATAGATACAGAAGC
CCGTGGTAATTTCTTTTATAGTAACCACAAAGGATACTACTTATATCTATGTCTTCG
Cas9 Target Site
B
5'-NNNNNNNNNNNNNNNNNNNNGUUUUAGAGCUAG
GUUCAACUAUUGCCUGAUCGGAAUAAAAUU CGAUA
AAAGUGGCACCGA
3'-UUUUCGUGGCU
2. human CFTRdelta508-targeting chimeric guide RNA
C
F508
...ATTAAAGAAAATATCATTGGCTTTGTTTCCTATGATGAATATAGATAC...
...TAATTTCTTTTATAGTAACCGAAACAAAGGATACTACTTATATCTATG...
repair template
I K E N I I G F V S Y D E Y R Y
502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517
3. repair template for deltaF508 mutation

FIG. 13A-C

A


GAA repeat expansion in *FXN* intron 1
GAA repeat
Exon 1
Exon 2
Transcription repression likely due to aberrant DNA structure or recruitment of heterchromatin binding proteins to long GAA repeats

B

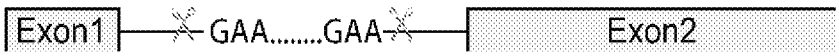

Exon1
GAA.....GAA
Exon2

FIG. 14A-B

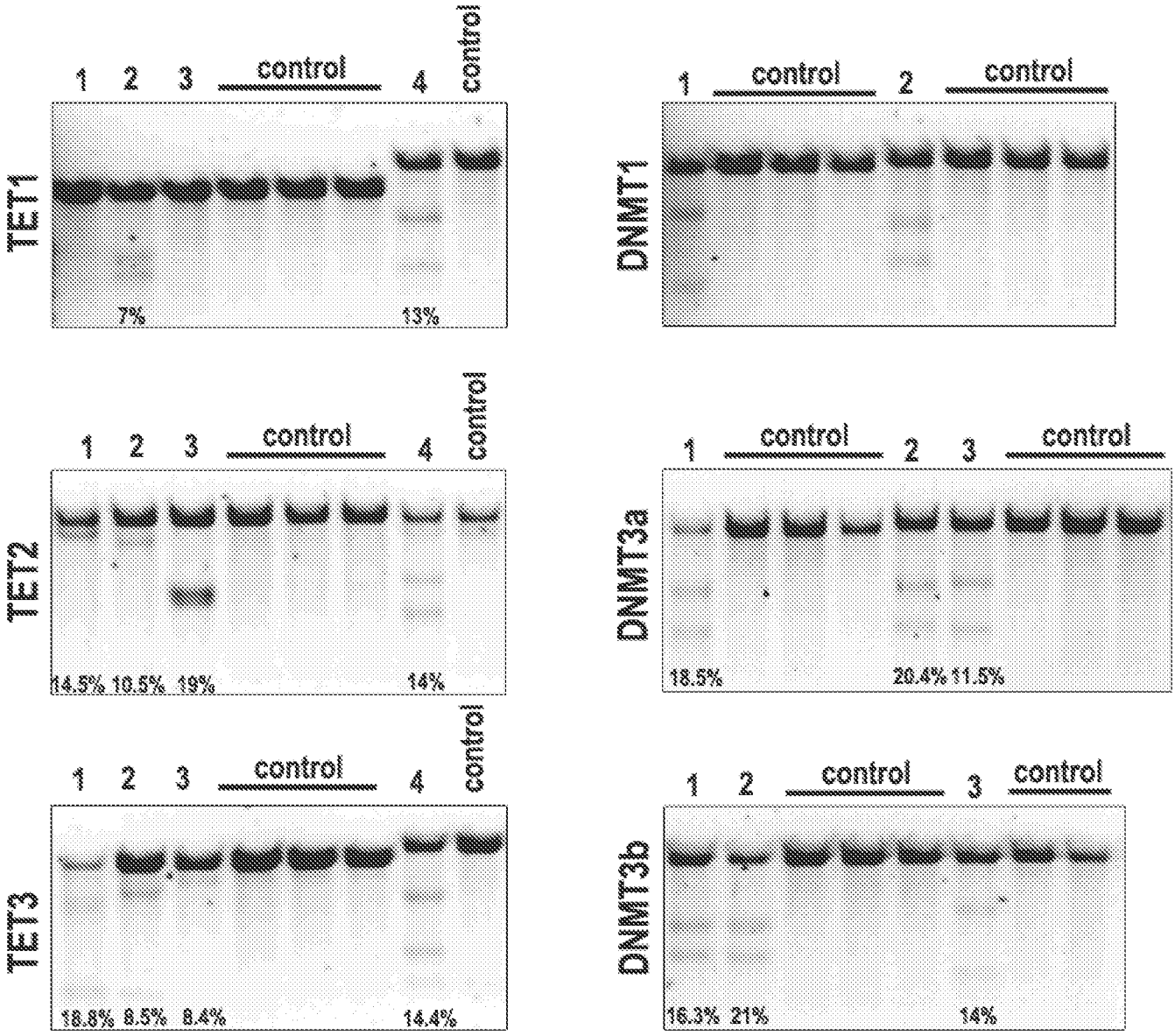

TET1
1 2 3 control 4 control
7%
13%
DNMT1
1 control 2 control
TET2
1 2 3 control 4 control
14.5% 10.5% 19%
14%
DNMT3a
1 control 2 3 control
18.5%
20.4% 11.5%
TET3
1 2 3 control 4 control
18.8% 8.5% 8.4%
14.4%
DNMT3b
1 2 control 3 control
16.3% 21%
14%

FIG. 15

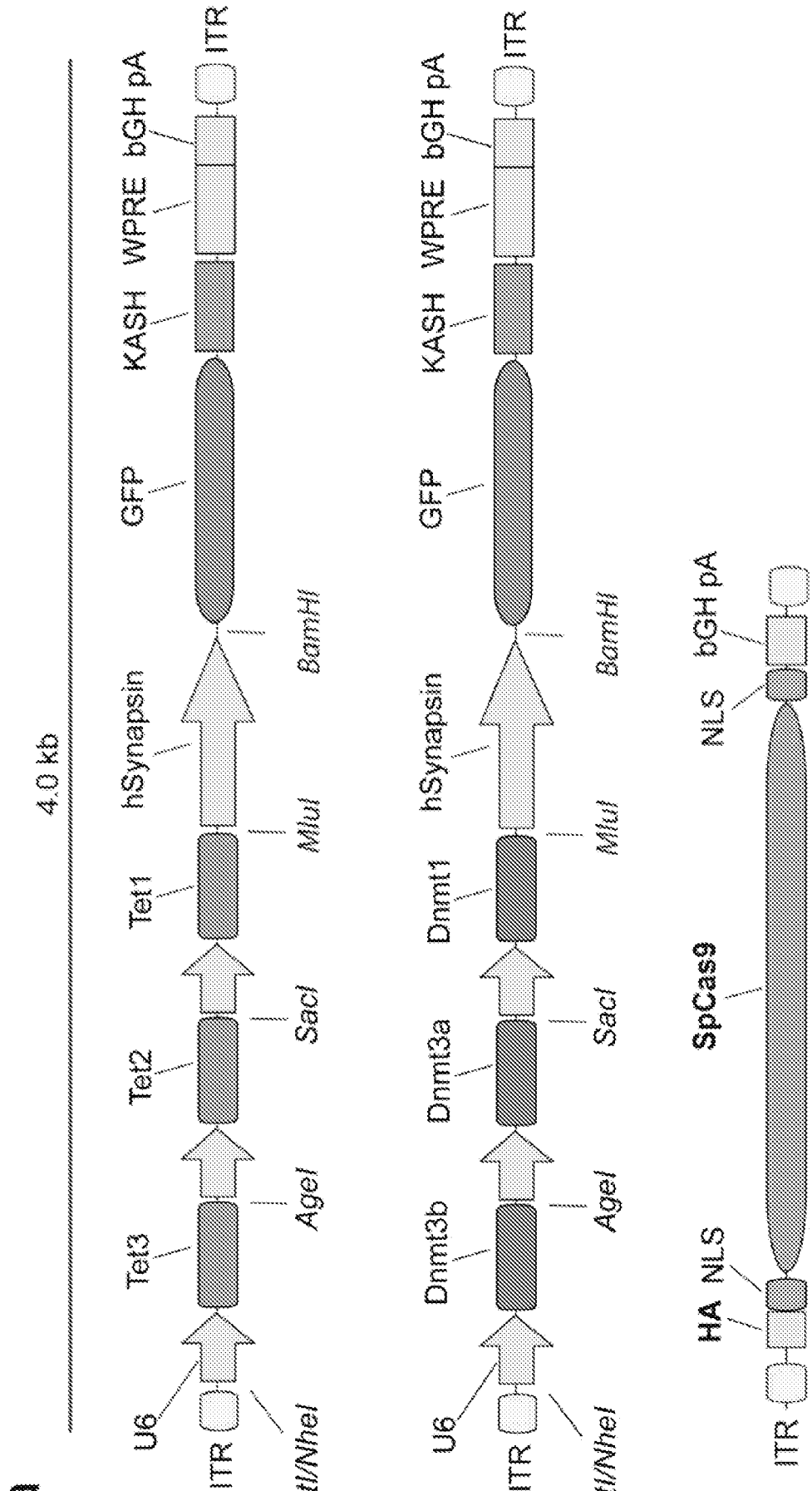

a
4.0 kb
U6
ITR
NotI/NheI
Tet3
AgeI
Tet2
SacI
Tet1
MluI
hSynapsin
GFP
KASH
WPRE
bGH pA
ITR
BamHI
Dnmt3b
Dnmt3a
Dnmt1
HA
NLS
SpCas9
NLS
bGH pA

FIG. 16

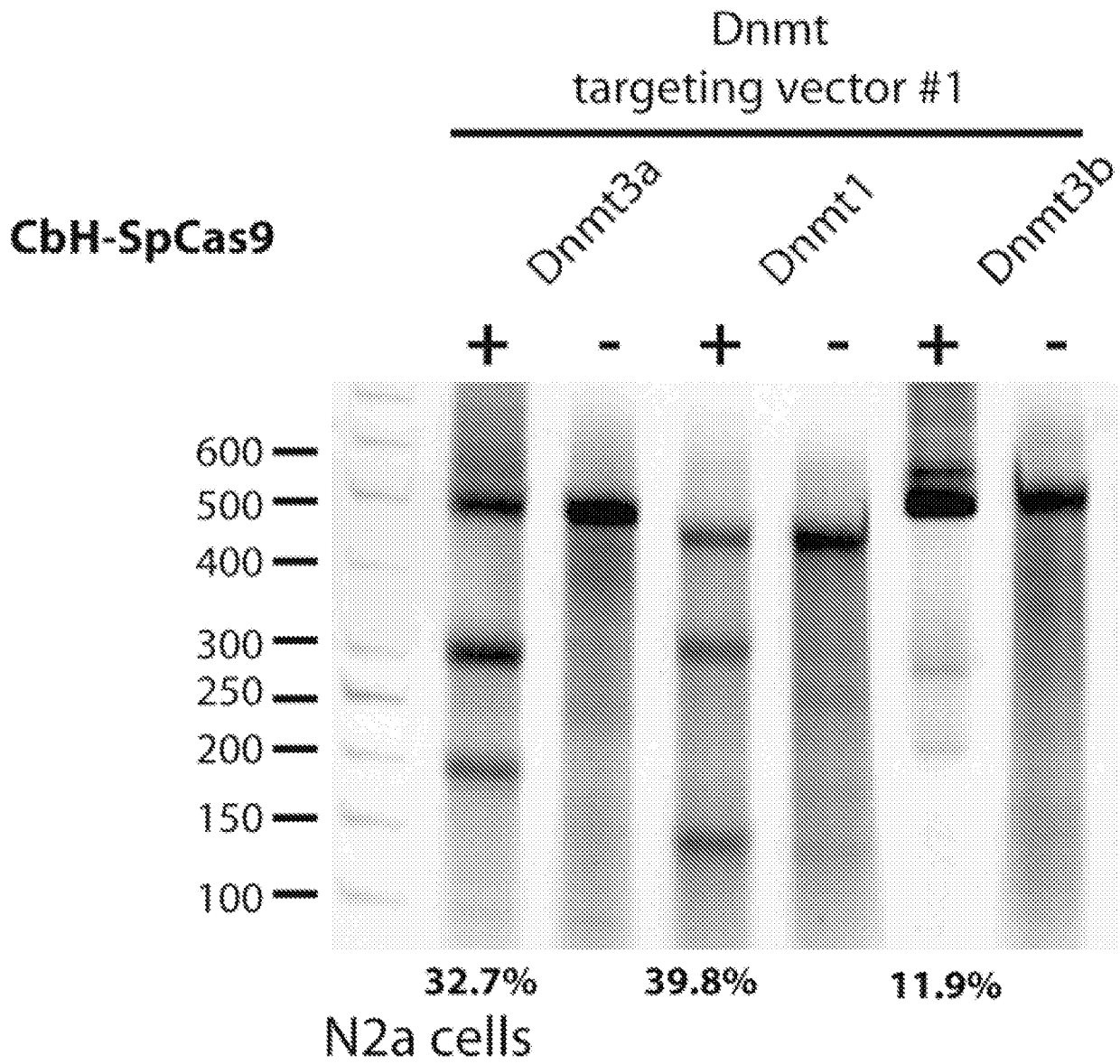

Dnmt
targeting vector #1
CbH-SpCas9
Dnmt3a
Dnmt1
Dnmt3b
+
-
+
-
+
-
600
500
400
300
250
200
150
100
32.7%
39.8%
11.9%
N2a cells

FIG. 17

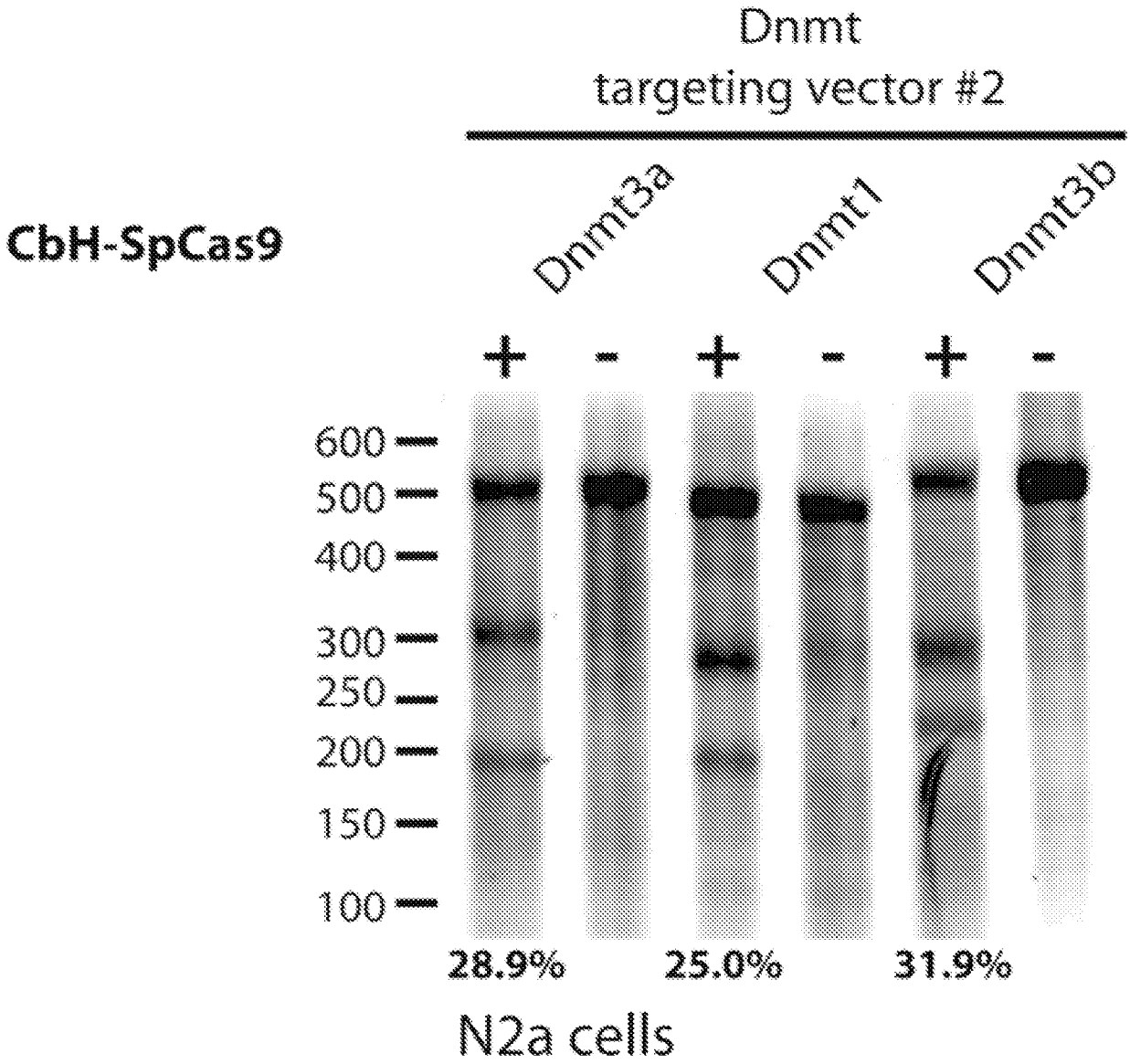

Dnmt
targeting vector #2
CbH-SpCas9
Dnmt3a
Dnmt1
Dnmt3b
+ - + - + -
600
500
400
300
250
200
150
100
28.9%
25.0%
31.9%
N2a cells

FIG. 18

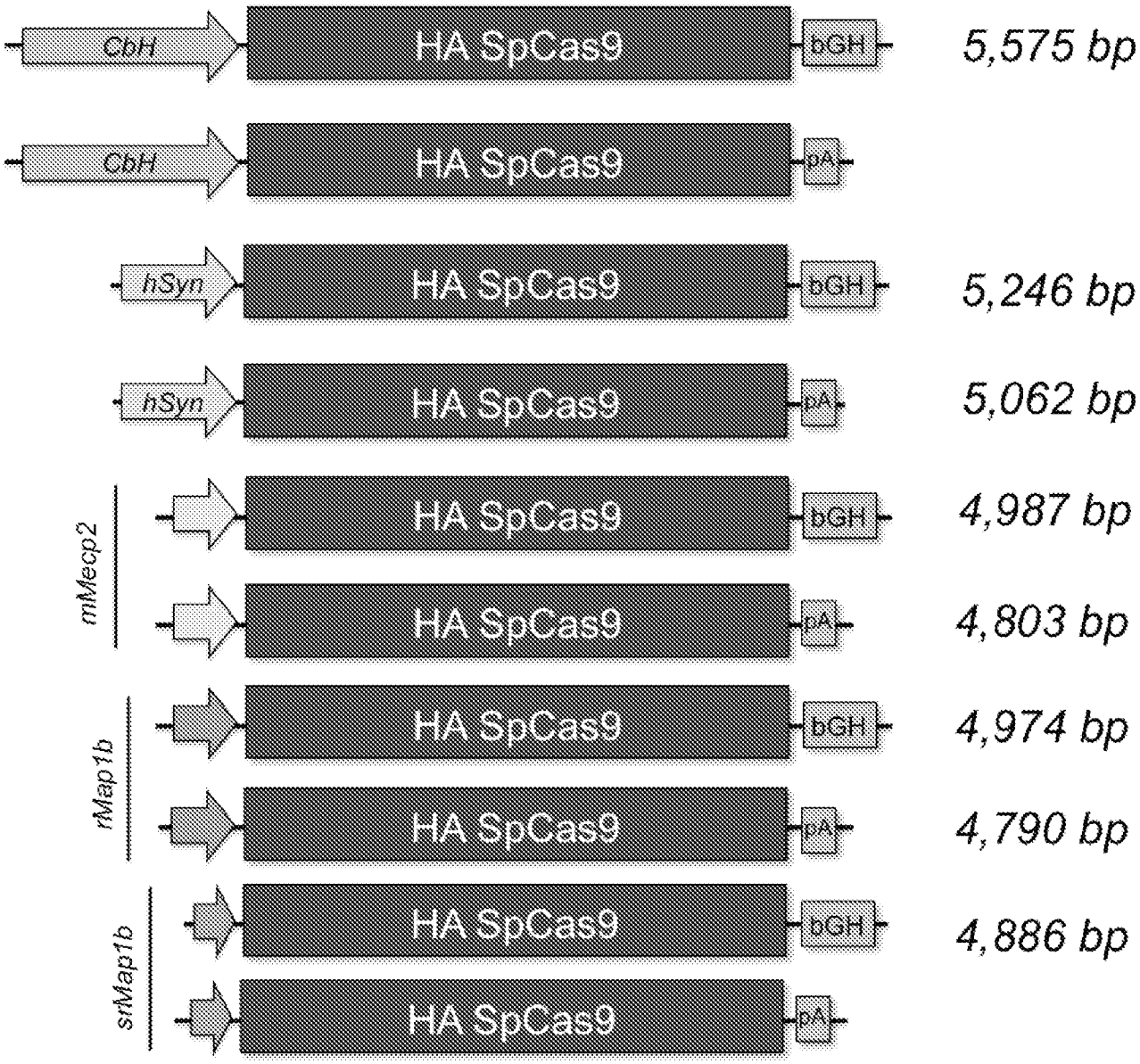

CbH
HA SpCas9
bGH
5,575 bp
CbH
HA SpCas9
pA
hSyn
HA SpCas9
bGH
5,246 bp
hSyn
HA SpCas9
pA
5,062 bp
mMecp2
HA SpCas9
bGH
4,987 bp
HA SpCas9
pA
4,803 bp
rMap1b
HA SpCas9
bGH
4,974 bp
HA SpCas9
pA
4,790 bp
srMap1b
HA SpCas9
bGH
4,886 bp
HA SpCas9
pA

FIG. 19

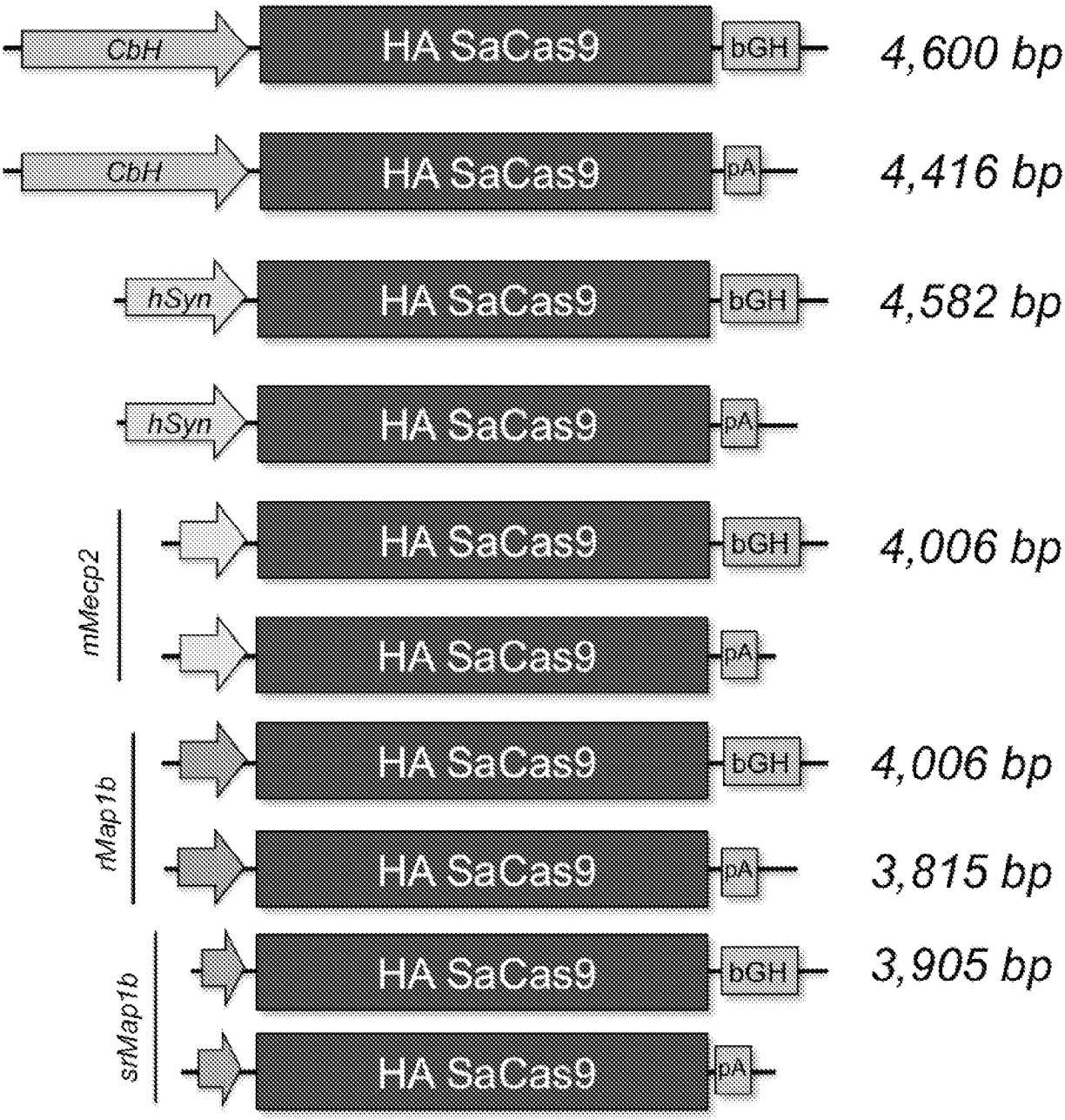

CbH
HA SaCas9
bGH
4,600 bp
CbH
HA SaCas9
pA
4,416 bp
hSyn
HA SaCas9
bGH
4,582 bp
hSyn
HA SaCas9
pA
mMecp2
HA SaCas9
bGH
4,006 bp
HA SaCas9
pA
rMap1b
HA SaCas9
bGH
4,006 bp
HA SaCas9
pA
3,815 bp
srMap1b
HA SaCas9
bGH
3,905 bp
HA SaCas9
pA

FIG. 20

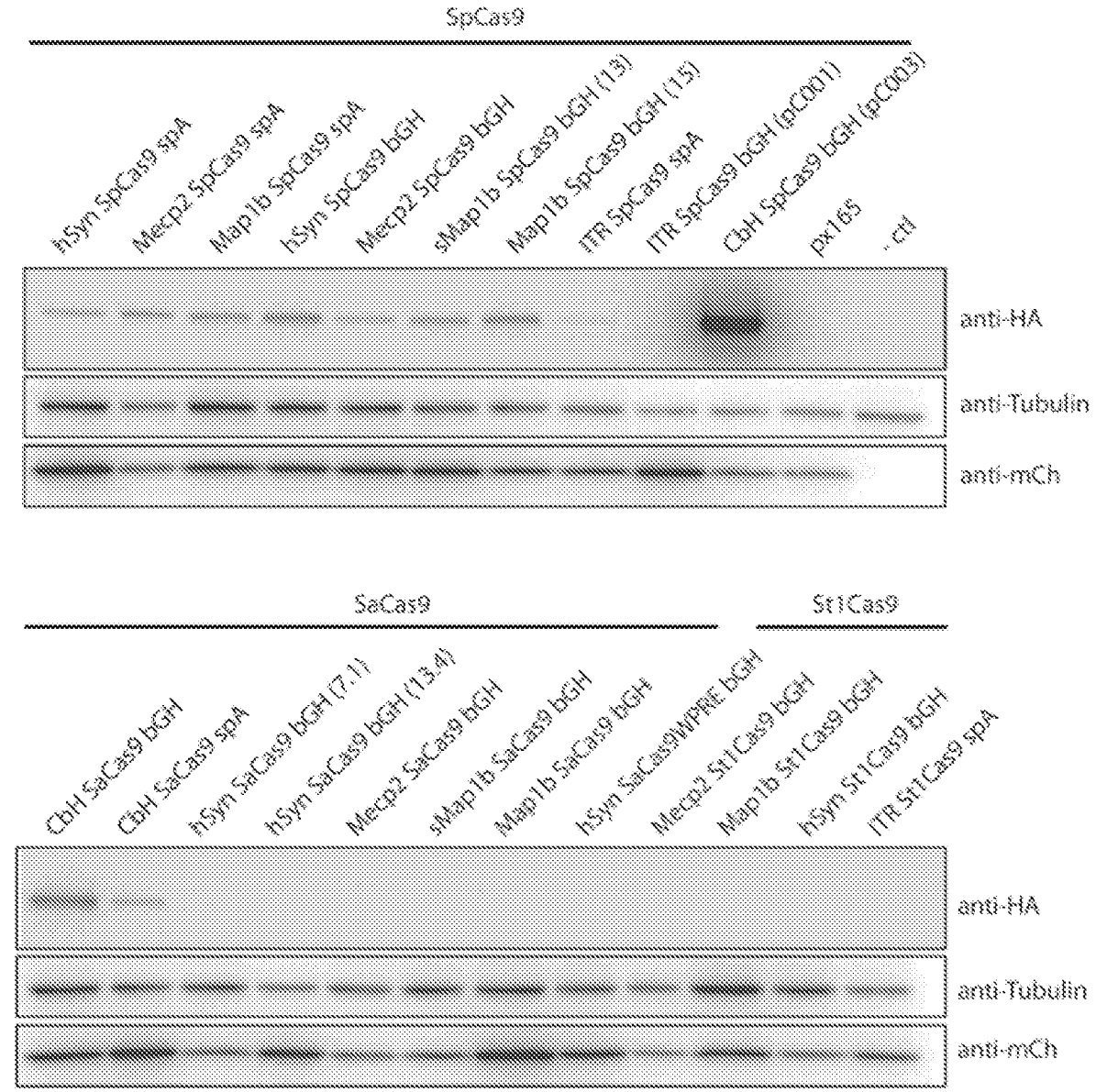

Expression of SpCas9 & SaCas9 in N2a cells
SpCas9
hSyn SpCas9 spA
Mecp2 SpCas9 spA
Map1b SpCas9 spA
hSyn SpCas9 bGH
Mecp2 SpCas9 bGH
sMap1b SpCas9 bGH (13)
Map1b SpCas9 bGH (15)
ITR SpCas9 spA
ITR SpCas9 bGH (pC001)
CbH SpCas9 bGH (pC003)
px165
- ctl
anti-HA
anti-Tubulin
anti-mCh
SaCas9
St1Cas9
CbH SaCas9 bGH
CbH SaCas9 spA
hSyn SaCas9 bGH (7.1)
hSyn SaCas9 bGH (13.4)
Mecp2 SaCas9 bGH
sMap1b SaCas9 bGH
Map1b SaCas9 bGH
hSyn SaCas9WPRE bGH
Mecp2 St1Cas9 bGH
Map1b St1Cas9 bGH
hSyn St1Cas9 bGH
ITR St1Cas9 spA
anti-HA
anti-Tubulin
anti-mCh

FIG. 21

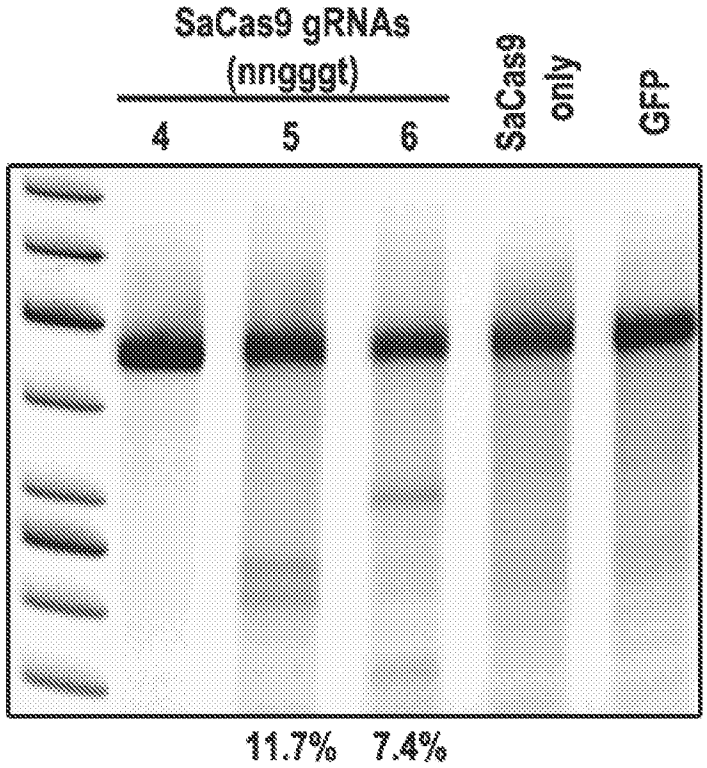


FIG. 22

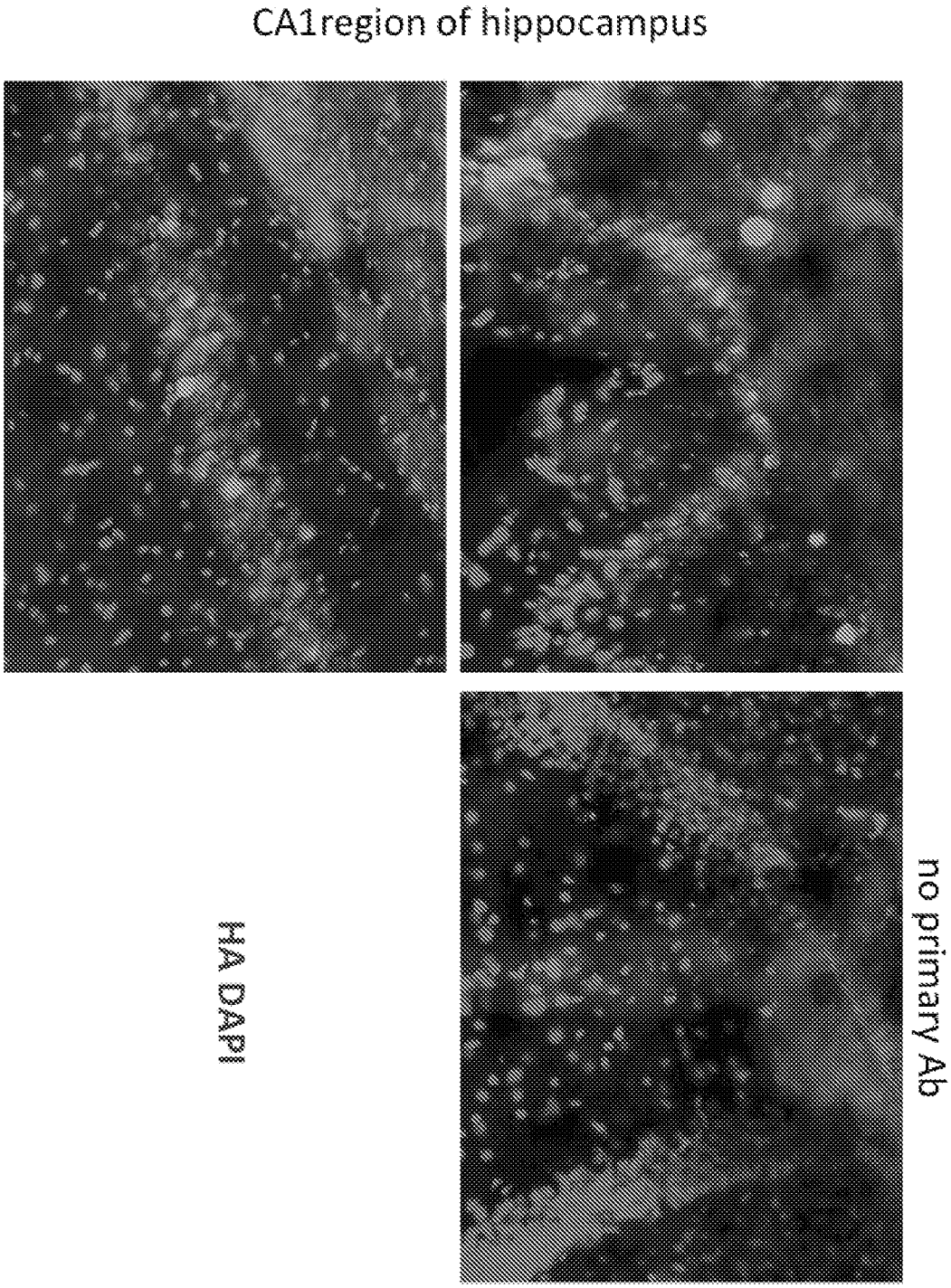


FIG. 23

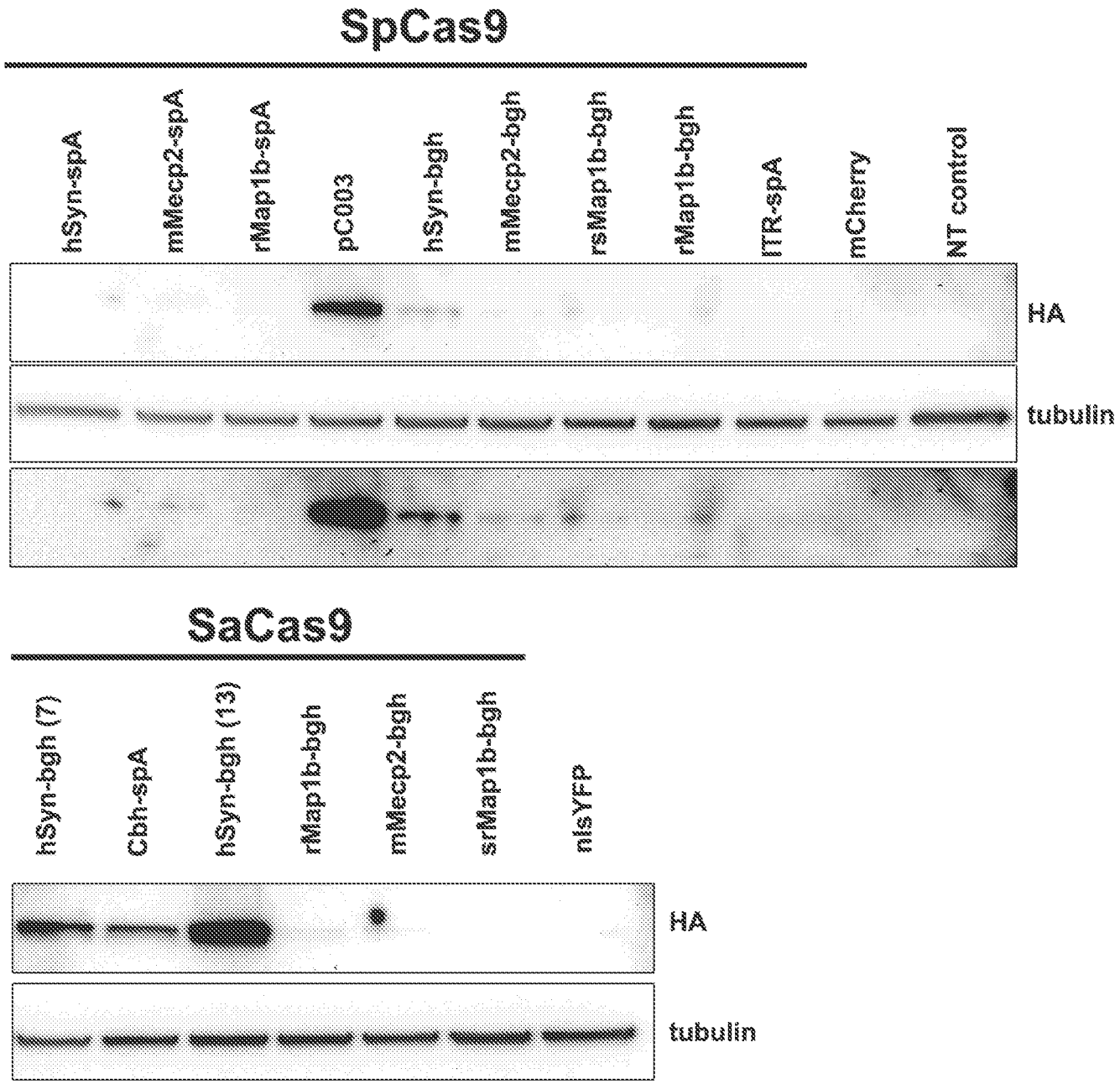

SpCas9
hSyn-spA
mMecp2-spA
rMap1b-spA
pC003
hSyn-bgh
mMecp2-bgh
rsMap1b-bgh
rMap1b-bgh
ITR-spA
mCherry
NT control
HA
tubulin
SaCas9
hSyn-bgh (7)
Cbh-spA
hSyn-bgh (13)
rMap1b-bgh
mMecp2-bgh
srMap1b-bgh
nlsYFP
HA
tubulin

FIG. 24

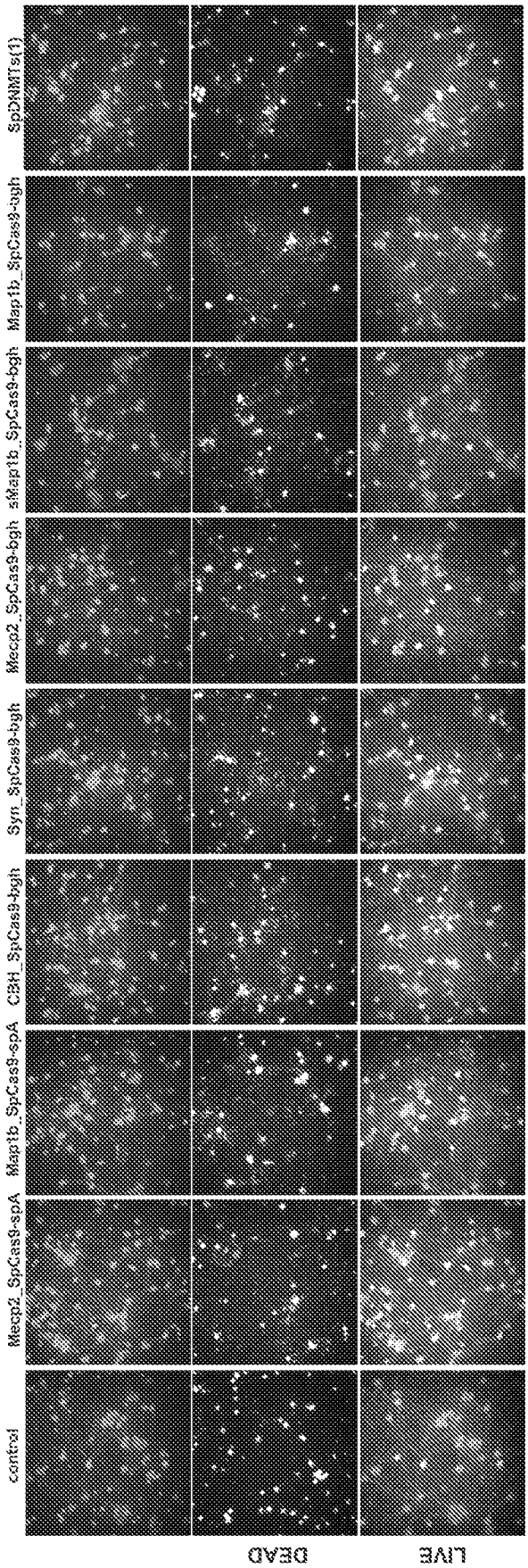

control
Mecp2_SpCas9-spA
Map1b_SpCas9-spA
CBH_SpCas9-bgh
Syn_SpCas9-bgh
Mecp2_SpCas9-bgh
sMap1b_SpCas9-bgh
Map1b_SpCas9-bgh
SpDNMTs(1)
DEAD
LIVE

FIG. 25

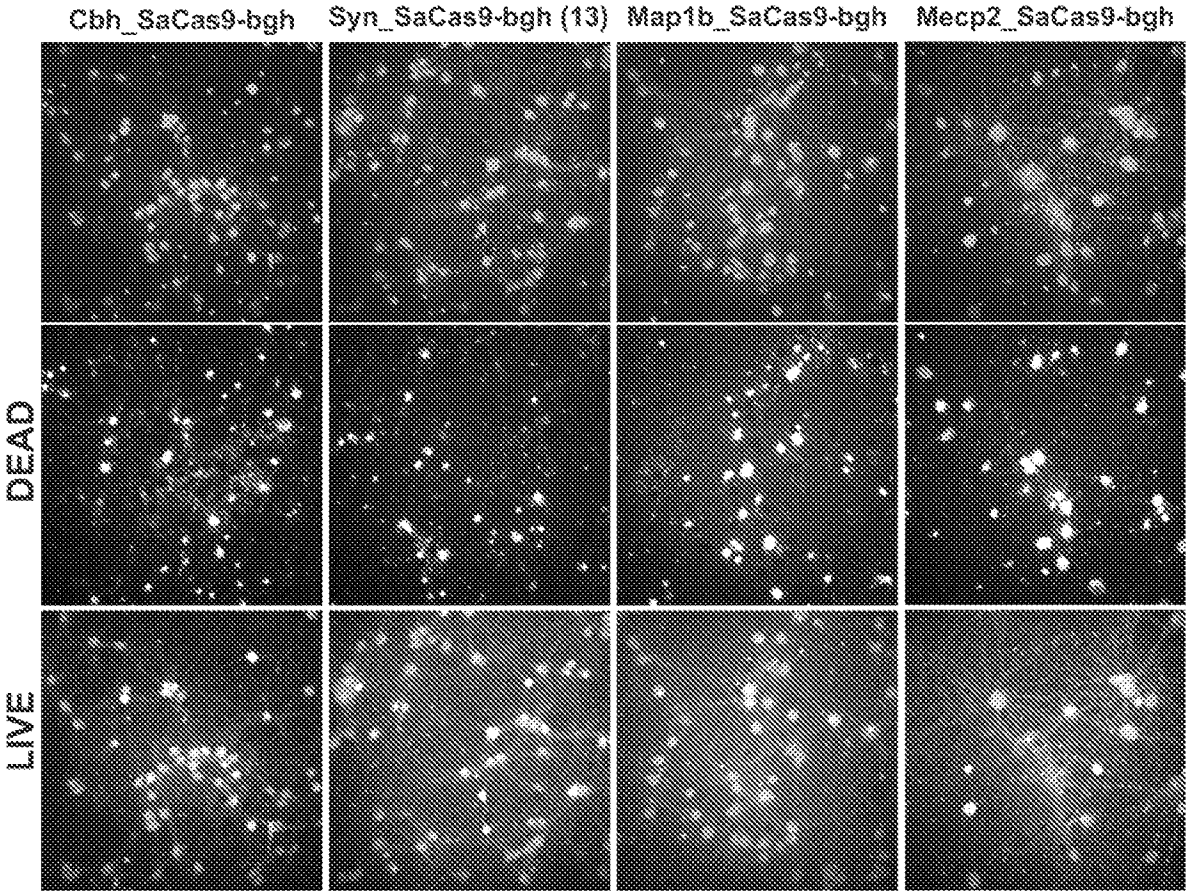

Cbh_SaCas9-bgh
Syn_SaCas9-bgh (13)
Map1b_SaCas9-bgh
Mecp2_SaCas9-bgh
DEAD
LIVE

FIG. 26

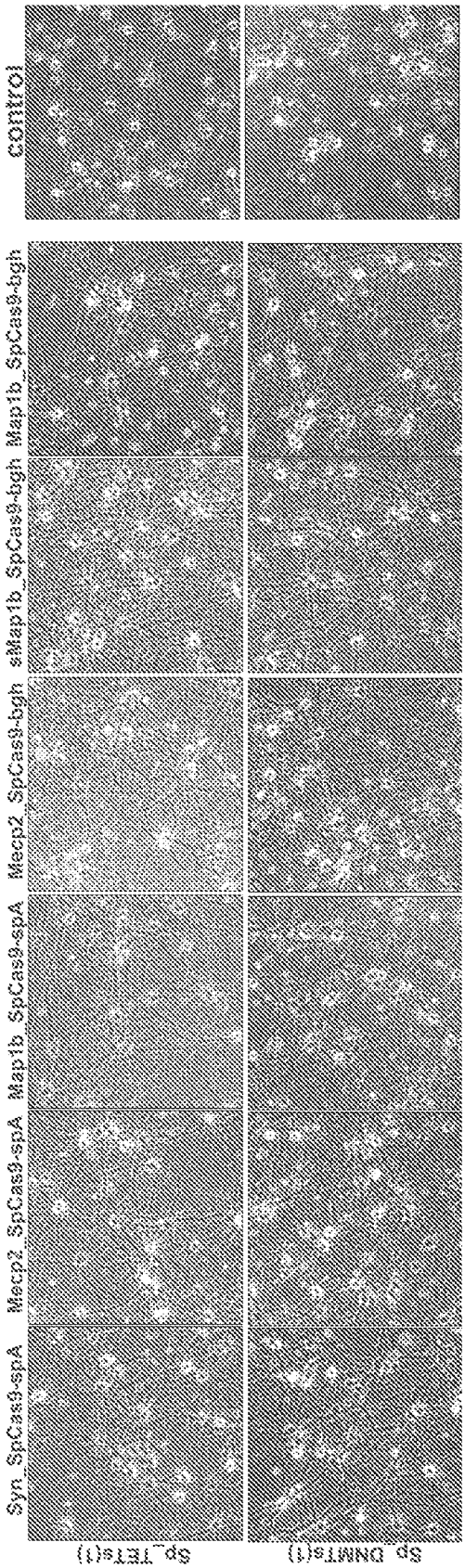

Sp_TETs(1)
Sp_DNMTs(1)
Syn_SpCas9-spA
Mecp2_SpCas9-spA
Map1b_SpCas9-spA
Mecp2_SpCas9-bgh
eMap1b_SpCas9-bgh
Map1b_SpCas9-bgh
control

FIG. 27

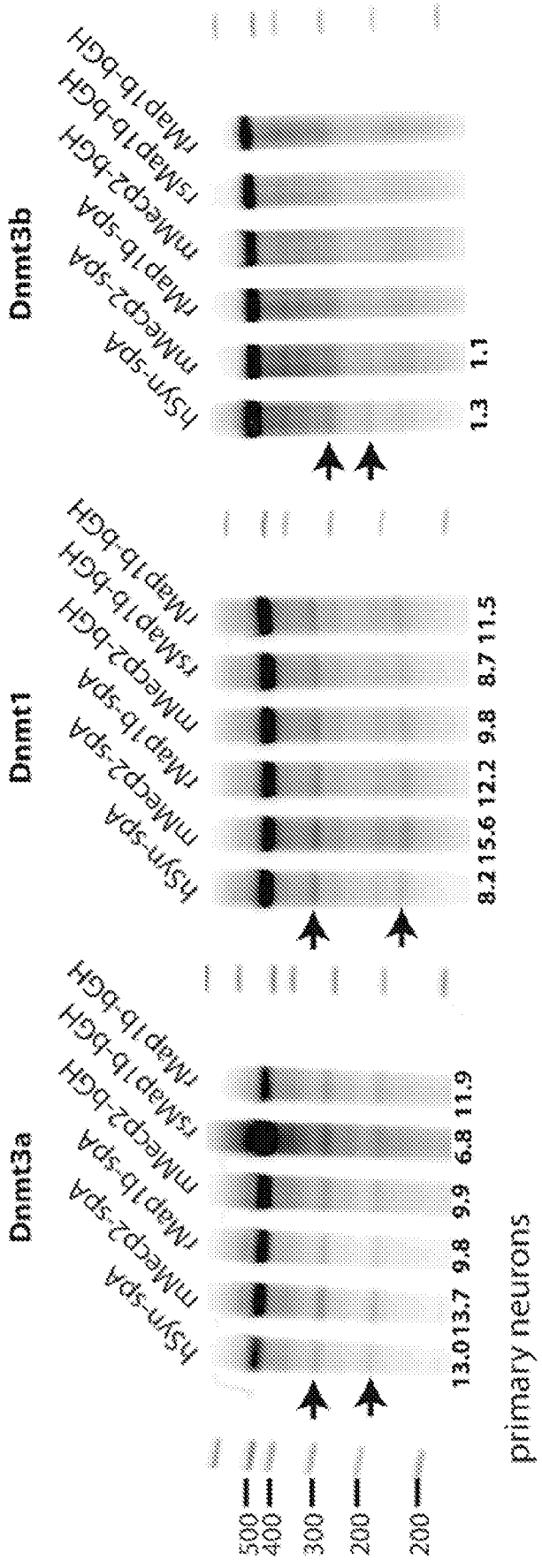

Dnmt3a
Dnmt1
Dnmt3b
hSyn-sPA
mMecp2-sPA
rMap1b-sPA
rMecp2-bGH
rsMap1b-bGH
rMap1b-bGH
13.0 13.7 9.8 9.9 6.8 11.9
8.2 15.6 12.2 9.8 8.7 11.5
1.3 1.1
500
400
300
200
200
primary neurons

FIG. 28

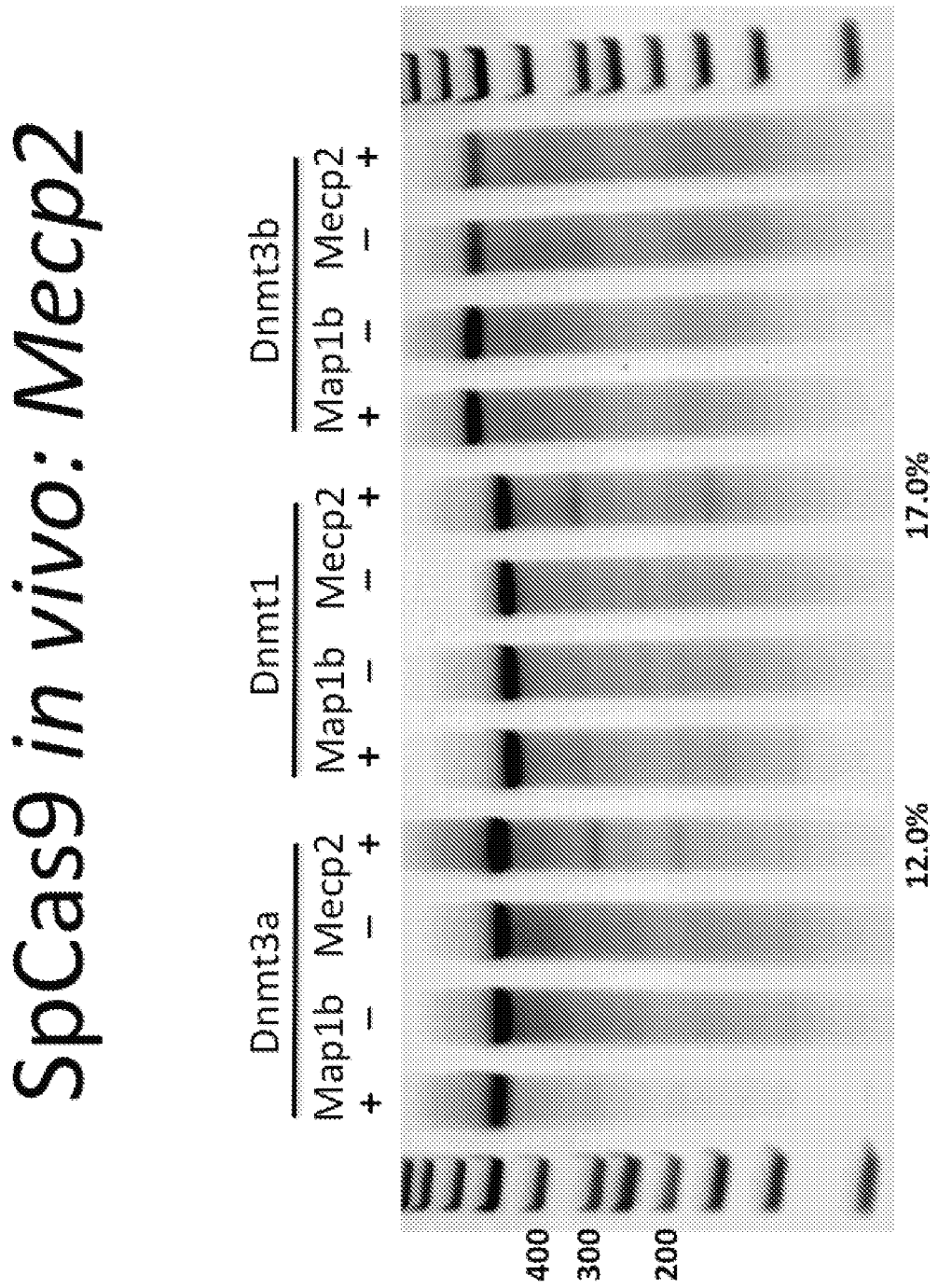

SpCas9 in vivo: Mecp2
Dnmt3a
Map1b + – –
Mecp2 – – +
Dnmt1
Map1b + – –
Mecp2 – – +
Dnmt3b
Map1b + – –
Mecp2 – – +
400
300
200
12.0%
17.0%

FIG. 29

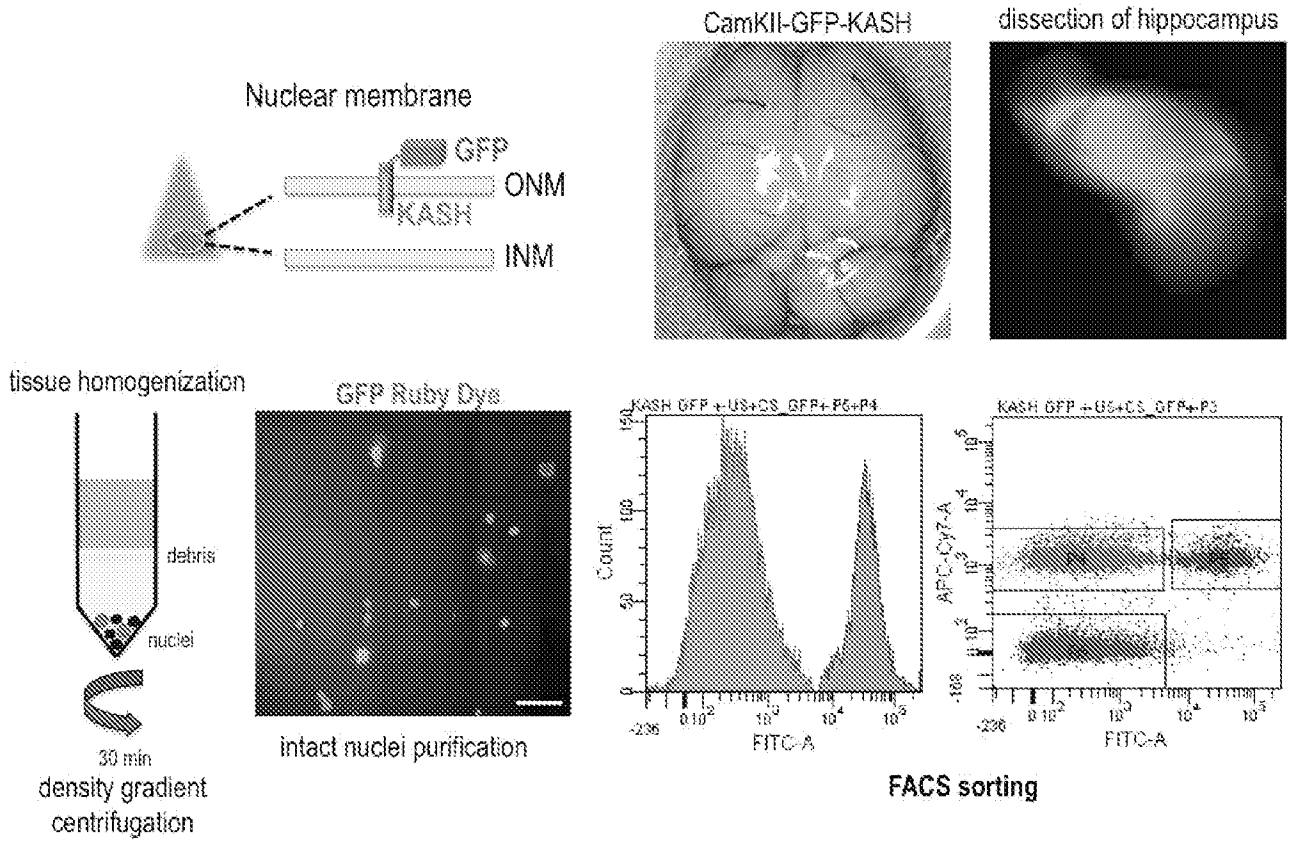

Purification of cell nuclei from brain
Nuclear membrane
GFP
ONM
KASH
INM
CamKII-GFP-KASH
dissection of hippocampus
tissue homogenization
debris
nuclei
30 min
density gradient centrifugation
GFP Ruby Dye
intact nuclei purification
Count
FITC-A
APC-Cy7-A
FACS sorting

FIG. 30

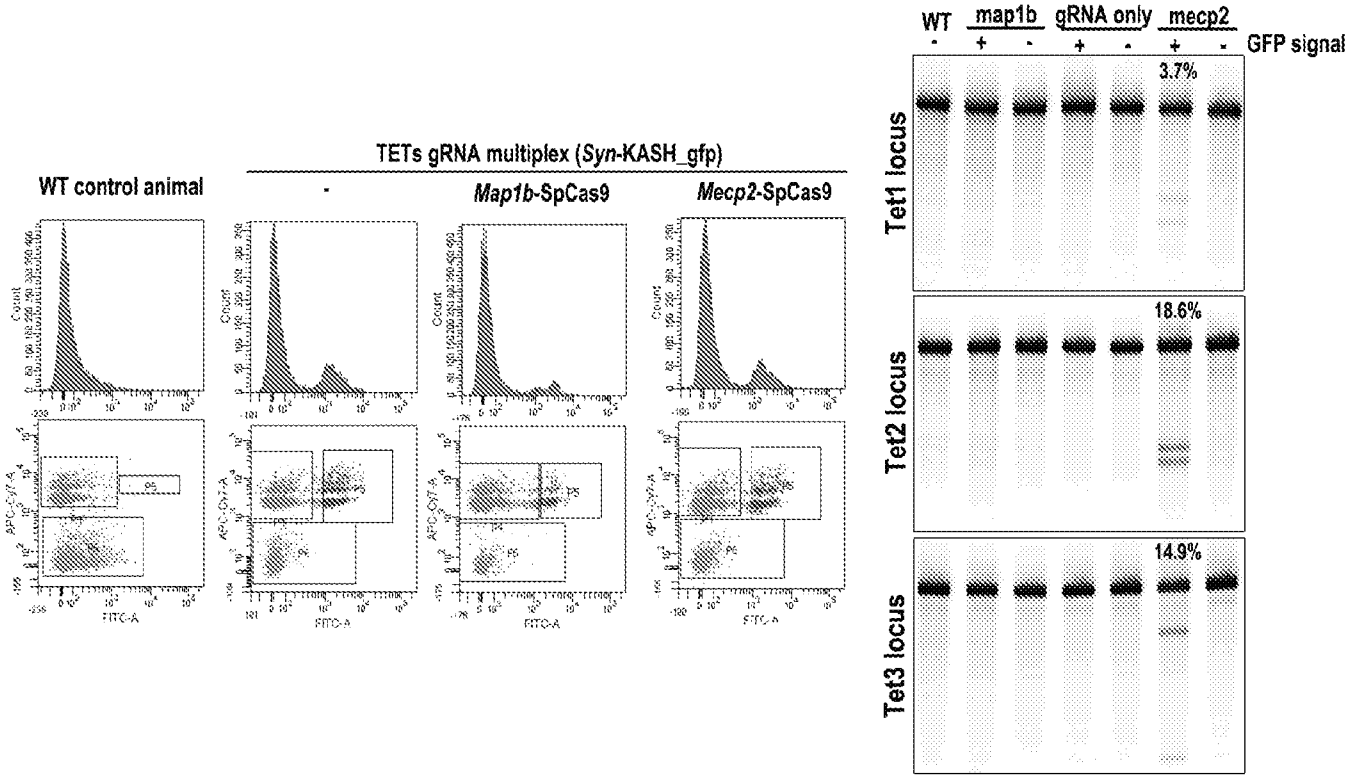

WT control animal
TETs gRNA multiplex (Syn-KASH_gfp)
-
Map1b-SpCas9
Mecp2-SpCas9
WT
map1b
gRNA only
mecp2
GFP signal
Tet1 locus
Tet2 locus
Tet3 locus
3.7%
18.6%
14.9%

FIG. 31

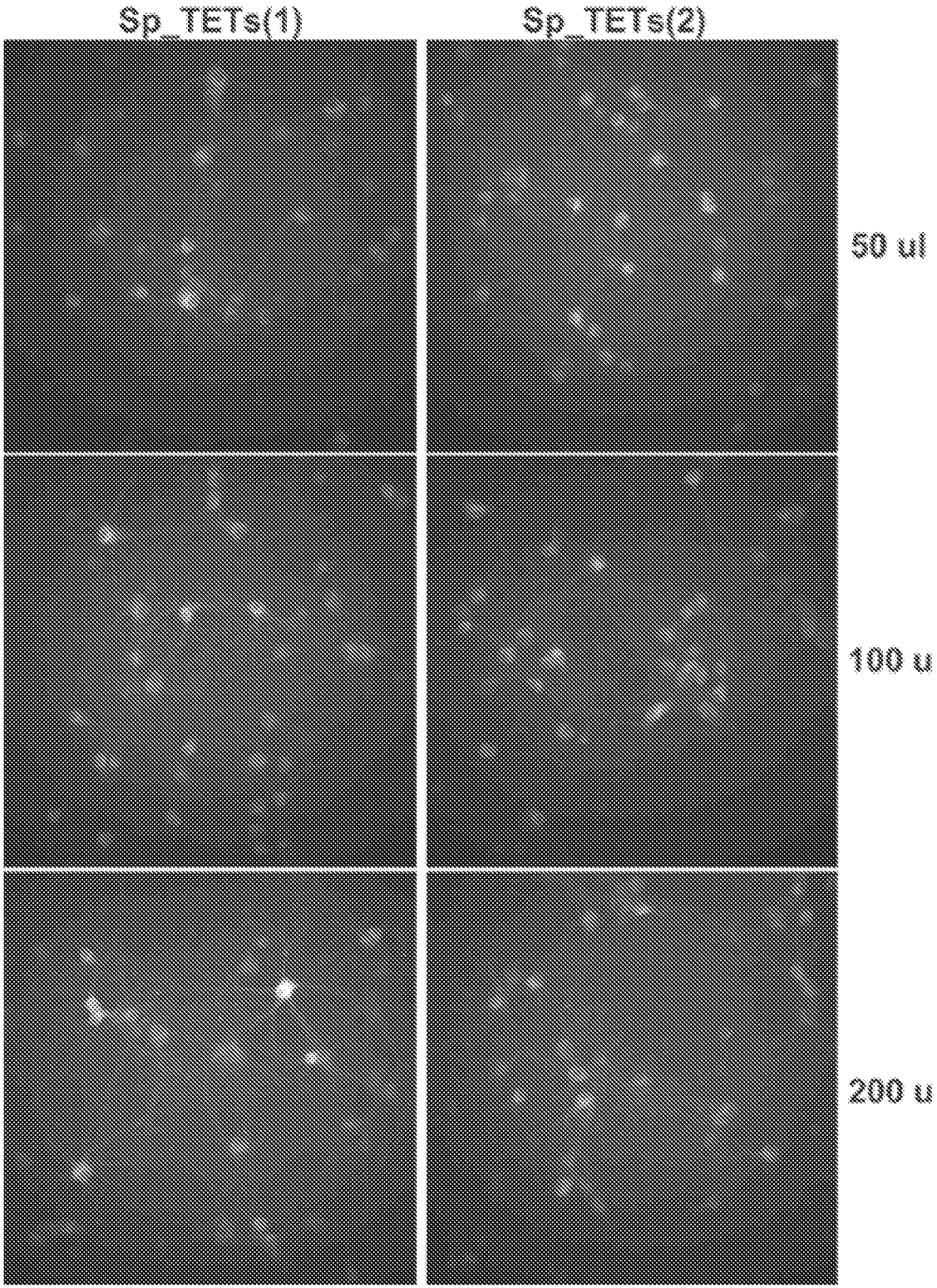
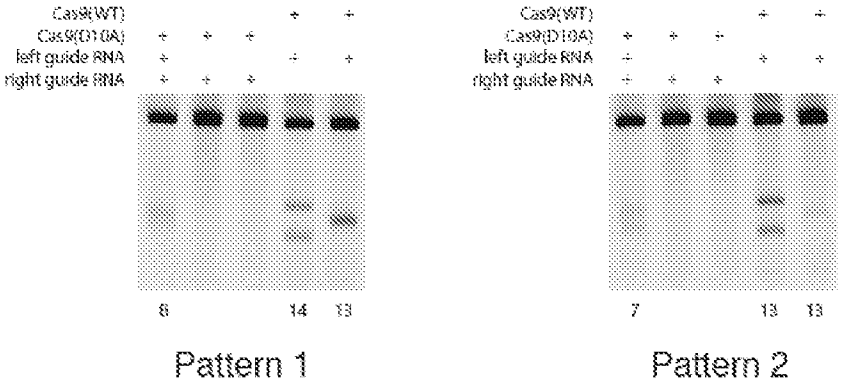


FIG. 32

	pattern	overhang
1	CCnNNNNNNNNNNNNNNNnnnnnnnnnnnnnNNNNNNNNNNNNNNNnGG	38
2	CCnNNNNNNNNNNNNNNNNnnnnnnnnnnnnNNNNNNNNNNNNNNNNnGG	34
3	CCnNNNNNNNNNNNNNNNNNnnnnnnnnnnNNNNNNNNNNNNNNNNNnGG	30
4	CCnNNNNNNNNNNNNNNNNNNnnnnnnnnnNNNNNNNNNNNNNNNNNNnGG	26
5	CCnNNNNNNNNNNNNNNNNNNNnnnnnnnnNNNNNNNNNNNNNNNNNNNnGG	22
6	CCnNNNNNNNNNNNNNNNNNNNNnnnnnnnNNNNNNNNNNNNNNNNNNNNnGG	18
7	CCnNNNNNNNNNNNNNNNNNNNNNnnnnnnNNNNNNNNNNNNNNNNNNNNNnGG	14
8	CCnNNNNNNNNNNNNNNNNNNNNNNnnnnnNNNNNNNNNNNNNNNNNNNNNNnGG	10
9	CCnNNNNNNNNNNNNNNNNNNNNNNNnnnnNNNNNNNNNNNNNNNNNNNNNNNnGG	6
10	CCnNNNNNNNNNNNNNNNNNNNNNNNNnnnNNNNNNNNNNNNNNNNNNNNNNNNnGG	4
11	CCnNNNNNNNNNNNNNNNNNNNNNNNNNnnNNNNNNNNNNNNNNNNNNNNNNNNNnGG	3
12	CCnNNNNNNNNNNNNNNNNNNNNNNNNNNnNNNNNNNNNNNNNNNNNNNNNNNNNNnGG	2
13	CCnNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNnGG	1
14	CCnNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNnGG	0
15	CCnNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNnGG	1
16	CCnNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNnGG	2
17	CCnNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNnGG	3
18	CCnnGG	4
19	CCGG	8
20	GGCC	10
21	GGNNCC	12
22	GGNNNCC	14
23	GGNNNNCC	16
24	GGNNNNNNNCC	20



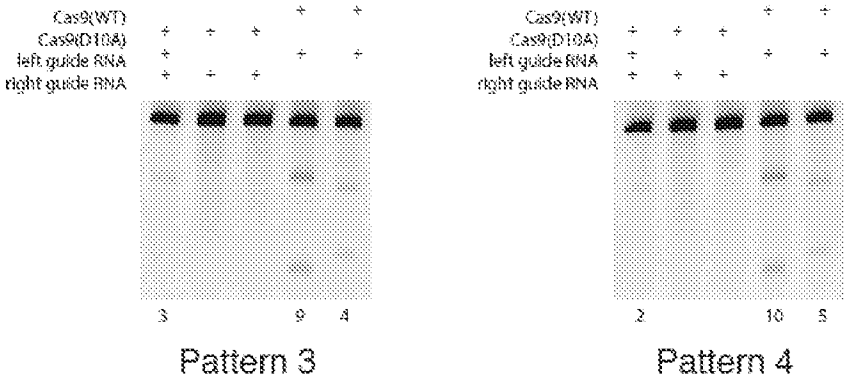


FIG. 33

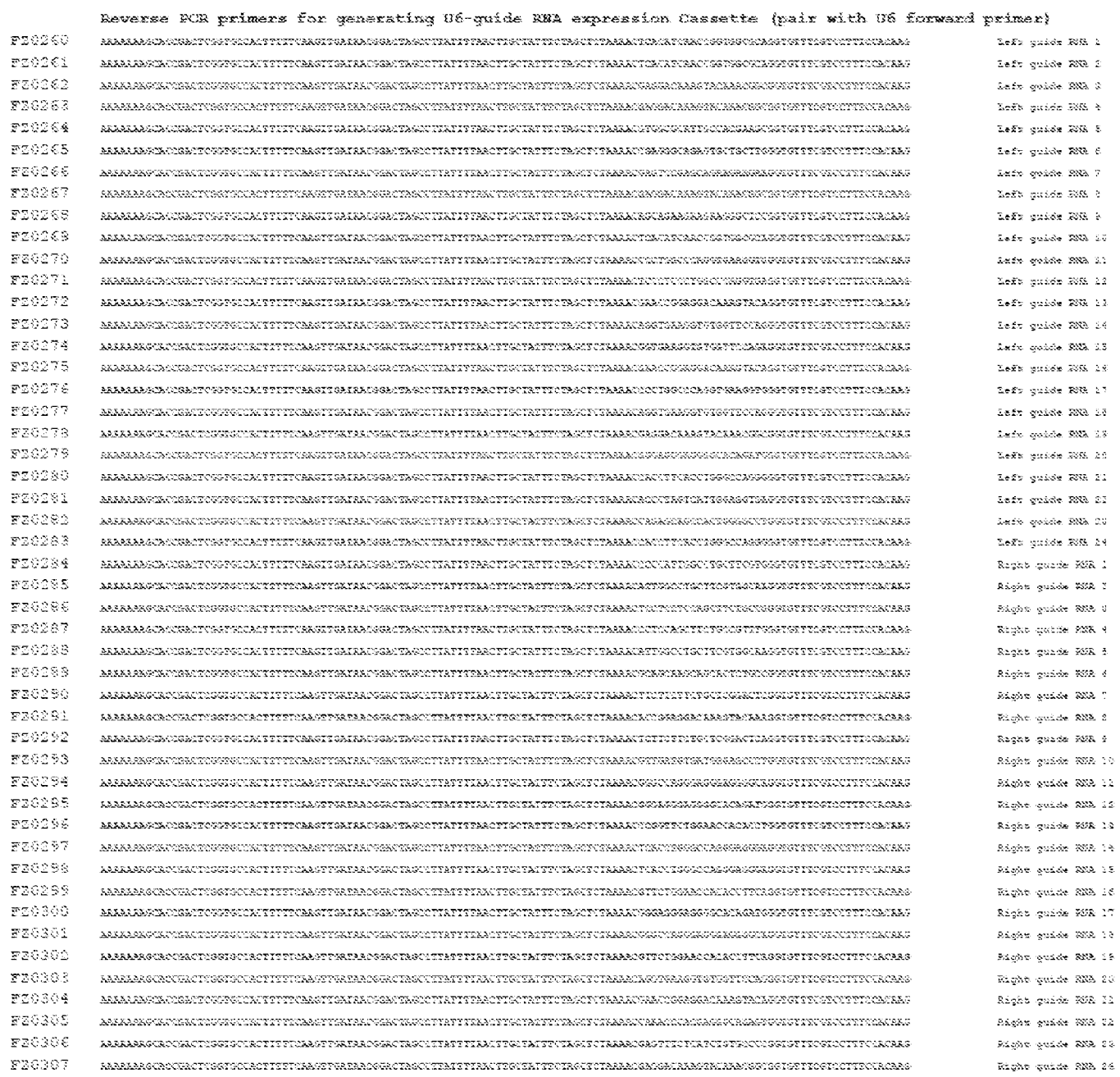

Reverse PCR primers for generating U6-guide RNA expression Cassette (pair with U6 forward primer)
Left guide RNA 1
Left guide RNA 2
Left guide RNA 3
Left guide RNA 4
Left guide RNA 5
Left guide RNA 6
Left guide RNA 7
Left guide RNA 8
Left guide RNA 9
Left guide RNA 10
Left guide RNA 11
Left guide RNA 12
Left guide RNA 13
Left guide RNA 14
Left guide RNA 15
Left guide RNA 16
Left guide RNA 17
Left guide RNA 18
Left guide RNA 19
Left guide RNA 20
Left guide RNA 21
Left guide RNA 22
Left guide RNA 23
Left guide RNA 24
Right guide RNA 1
Right guide RNA 2
Right guide RNA 3
Right guide RNA 4
Right guide RNA 5
Right guide RNA 6
Right guide RNA 7
Right guide RNA 8
Right guide RNA 9
Right guide RNA 10
Right guide RNA 11
Right guide RNA 12
Right guide RNA 13
Right guide RNA 14
Right guide RNA 15
Right guide RNA 16
Right guide RNA 17
Right guide RNA 18
Right guide RNA 19
Right guide RNA 20
Right guide RNA 21
Right guide RNA 22
Right guide RNA 23
Right guide RNA 24

FIG. 34



FIG. 35

lane #	Cas9 to use	Left gRNA	Right gRNA	Overhang (D10A) - number of bases protruding from 5'
1	D10A	left 23	left 12	-36
2	D10A	right 4	left 9	-25
3	D10A	left 23	right 23	-16
4	D10A	right 7	left 10	-15
5	D10A	right 16	left 3	-8
6	D10A	right 22	right 6	26
7	D10A	left 12	right 16	31
8	D10A	left 12	right 13	34
9	D10A	left 10	right 1	38
10	D10A	right 23	right 16	51
11	D10A	right 23	right 13	54
12	D10A	left 3	right 7	57
13	D10A	left 12	right 4	65
14	D10A	left 12	right 3	69
15	D10A	left 3	right 10	76
16	D10A	right 23	right 4	85
17	D10A	left 12	right 9	95
18	D10A	left 12	right 10	115
19	D10A	right 23	right 10	135
20	D10A	left 12	right 2	145
21	D10A	left 12	left 22	181
22	D10A	right 23	left 22	201
23	D10A	left 12	right 6	222
24	D10A	right 23	right 6	242

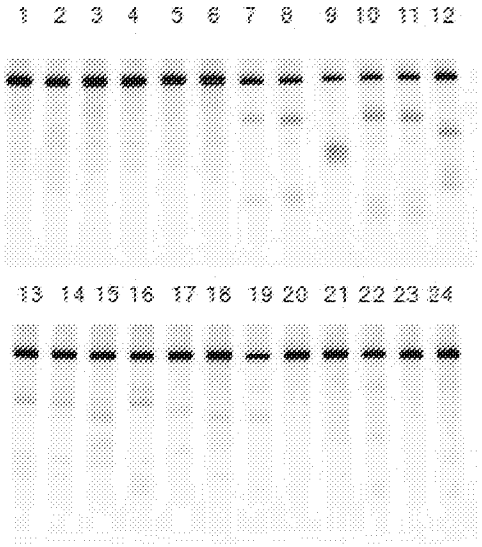


FIG. 36

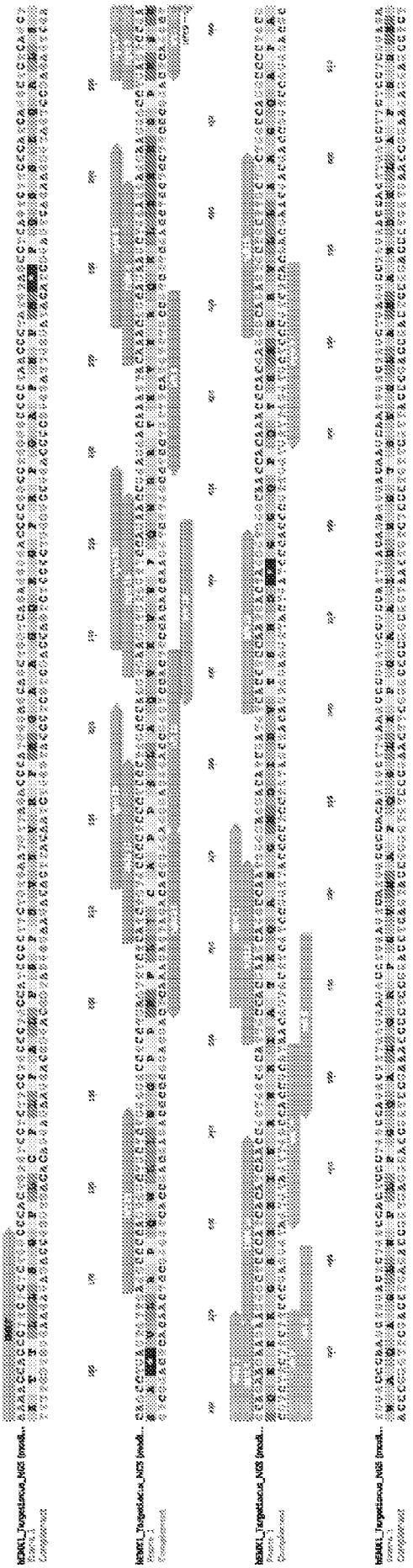
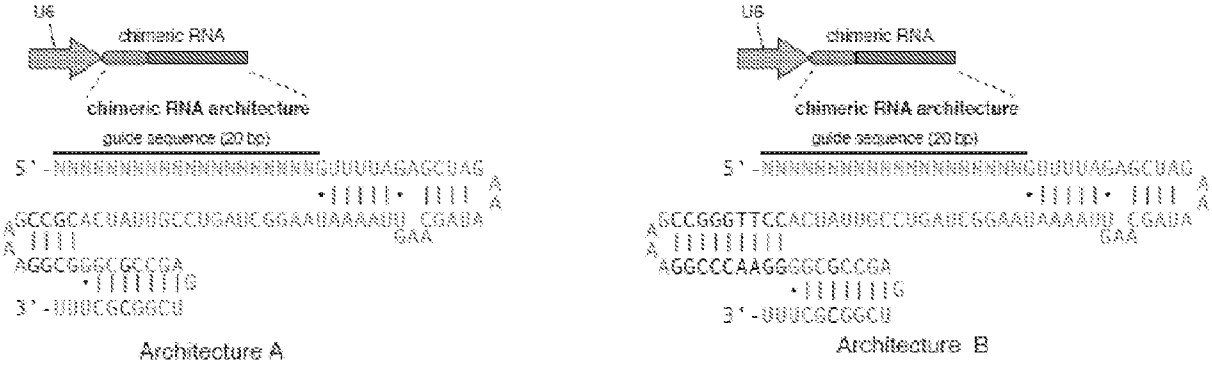


FIG. 37

A Stabilization of distal hairpins in existing chimeric RNA architecture



B Tandem single guide RNAs (tsgRNA) driven by a single U6 promoter

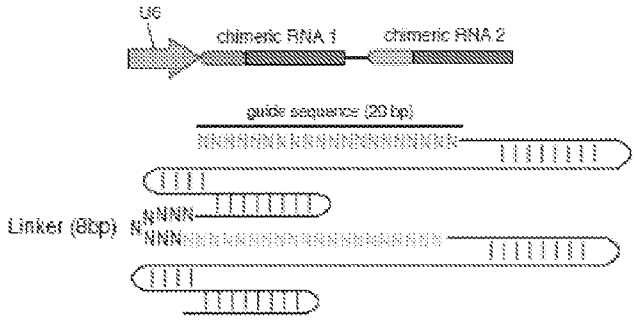
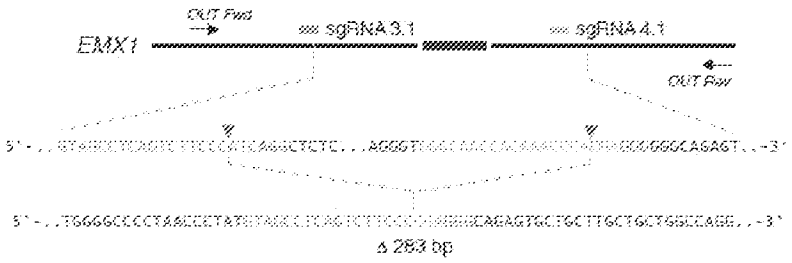
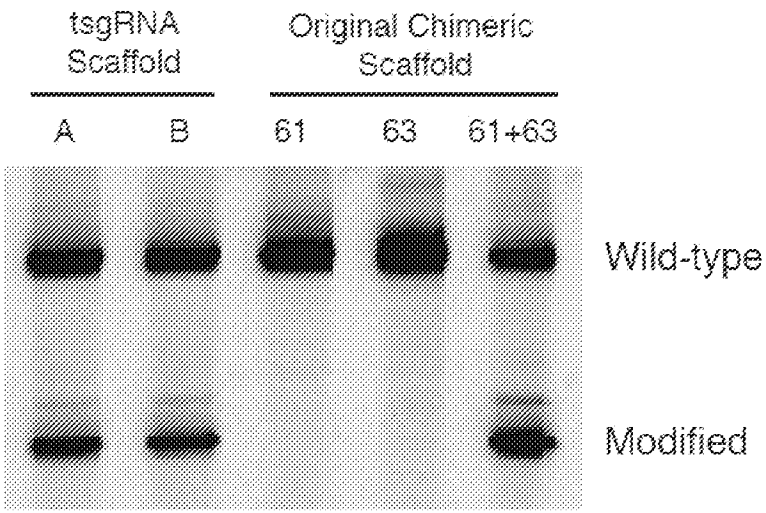


FIG. 38A-B

C Tandem sgRNA mediated microdeletion





D Tandem Double Nicking Activity with D10A Cas9

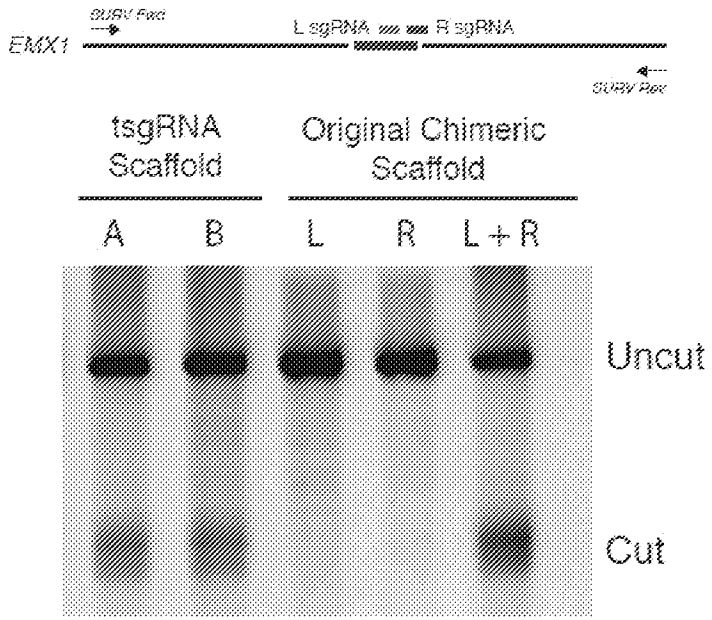
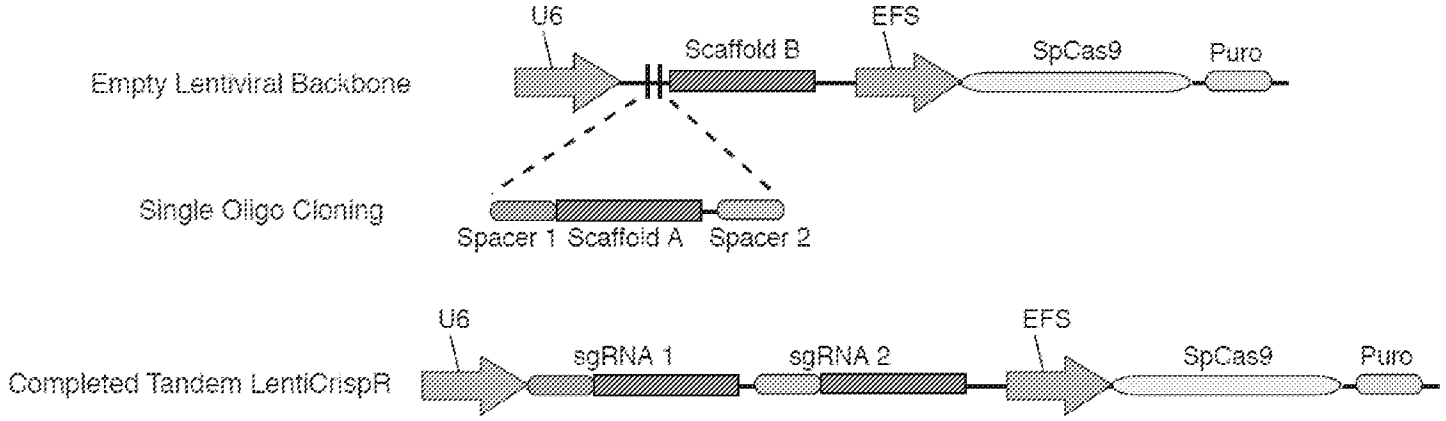


FIG. 38C-D

A. Tandem LentiCrispR Vector Design



B. Scaffold Sequences for LentiCrispr

Architecture	Length (nt)	Scaffold Sequence
A	77	gttttagagctaGAAAcggcaagttaaaataaggctagtccgttatcaCGCCGAAAGGCGggcgccgAGTcggcgcc
B	87	gttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaccttgggccGAAAggcccaaggggcgccgAGTcggcgcc
C	77	gttttagagccgGAAAcggcaagttaaaataaggctagtccgttatcaCGCCGAAAGGCGggcgccgAGTcggcgcc
D	87	gttttagatgcaattatgcaaagttaaaataaggctagtccgttatcaacttataccaataagtcatcgtcgagAAGctcgacgatg
E	87	gttttagaactcattagagtaagttaaaataaggctagtccgttatcaacatgcaccagcatgttcaggattcaAAGtgaatcctga

FIG. 39A-B

C. Tandem-guide RNAs constructed of modified scaffolds are more efficiently processed to individual units

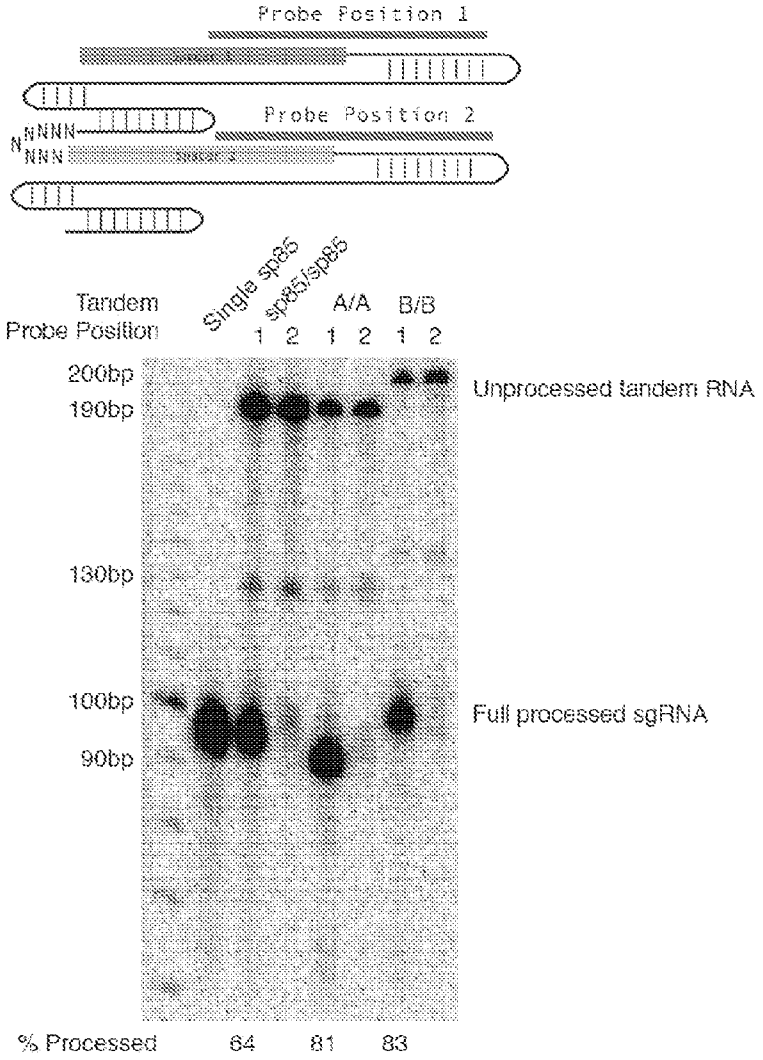


FIG. 39C

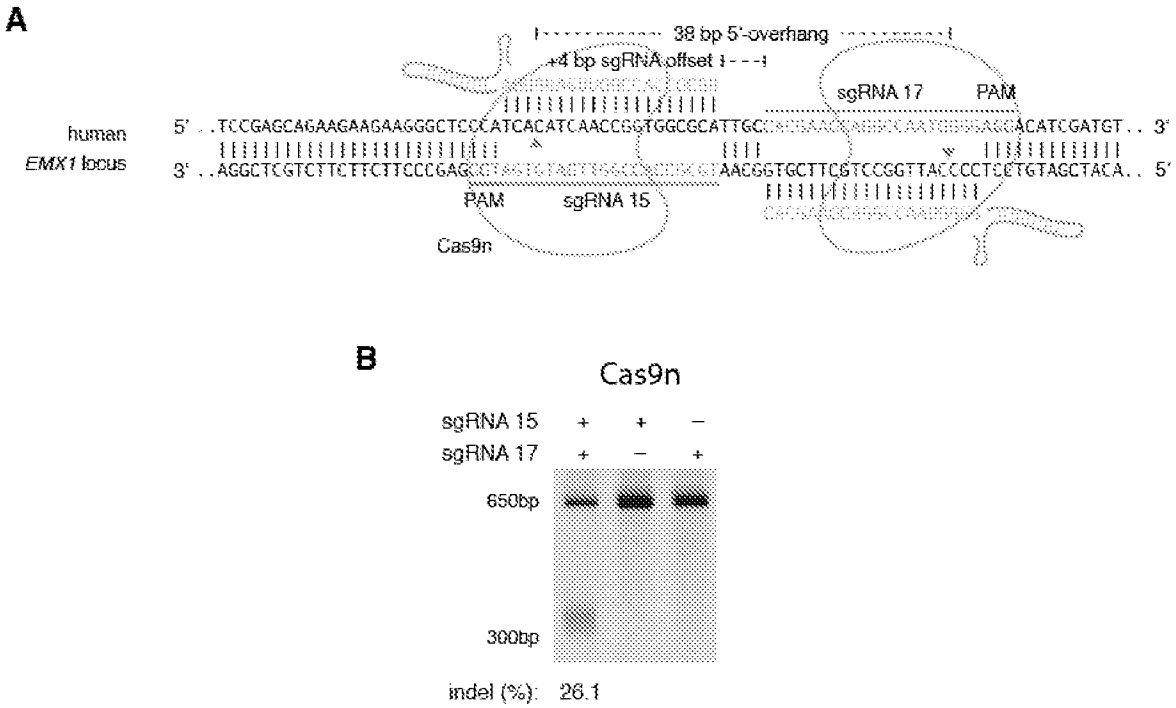
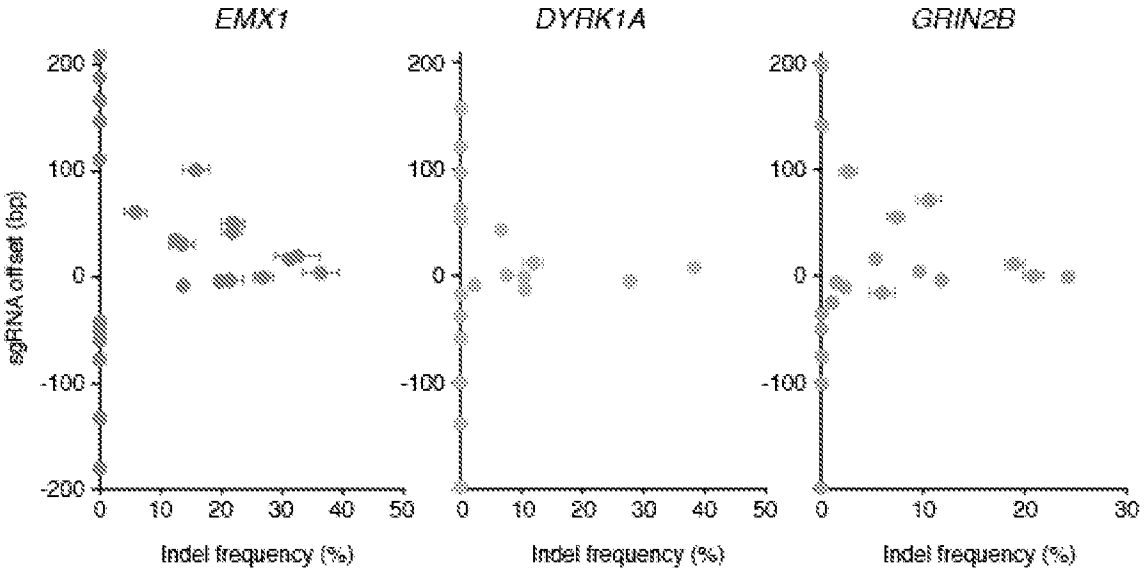

A
38 bp 5'-overhang
+4 bp sgRNA offset
sgRNA 17
PAM
human
EMX1 locus
5' . . TCCGAGCAGAAGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGC
3' . . AGGCTCGTCTTCTTCTTCCCGAG
AACGGTGCTTCGTCCGGTTACCCCTCCTGTAGCTACA . . 5'
PAM
sgRNA 15
Cas9n
B
Cas9n
sgRNA 15 + + −
sgRNA 17 + − +
650bp
300bp
indel (%): 26.1

FIG. 40A-B

A


EMX1
DYRK1A
GRIN2B
sgRNA offset (bp)
200
100
0
-100
-200
0 10 20 30 40 50
Indel frequency (%)

B

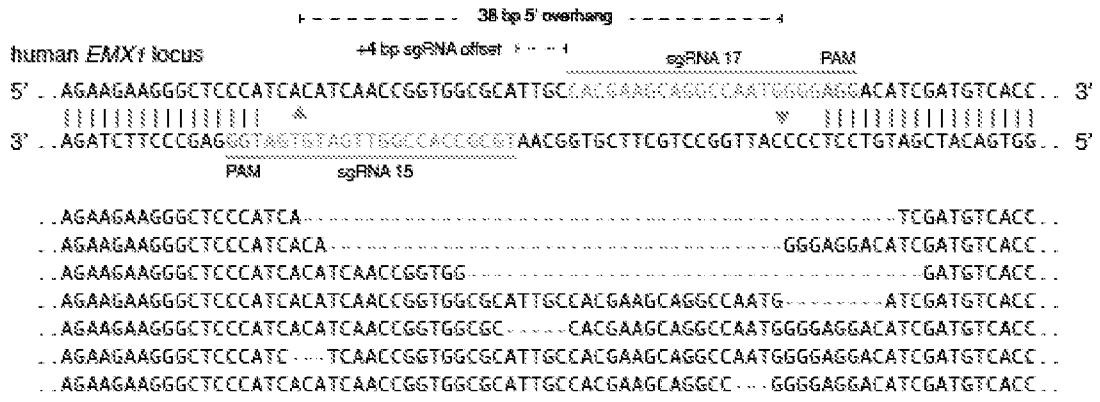

38 bp 5' overhang
+4 bp sgRNA offset
human EMX1 locus
sgRNA 17
PAM
5' ..AGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGACATCGATGTCACC.. 3'
3' ..AGATCTTCCCGAGGGTAGTGTAGTTGGCCACCGCGTAACGGTGCTTCGTCCGGTTACCCCTCCTGTAGCTACAGTGG.. 5'
PAM
sgRNA 15
..AGAAGAAGGGCTCCCATCA----------TCGATGTCACC..
..AGAAGAAGGGCTCCCATCACA----------GGGAGGACATCGATGTCACC..
..AGAAGAAGGGCTCCCATCACATCAACCGGTGG----------GATGTCACC..
..AGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAGGCCAATG--------ATCGATGTCACC..
..AGAAGAAGGGCTCCCATCACATCAACCGGTGGCGC-----CACGAAGCAGGCCAATGGGGAGGACATCGATGTCACC..
..AGAAGAAGGGCTCCCATC---TCAACCGGTGGCGCATTGCCACGAAGCAGGCCAATGGGGAGGACATCGATGTCACC..
..AGAAGAAGGGCTCCCATCACATCAACCGGTGGCGCATTGCCACGAAGCAGGCC---GGGGAGGACATCGATGTCACC..

FIG. 41A-B

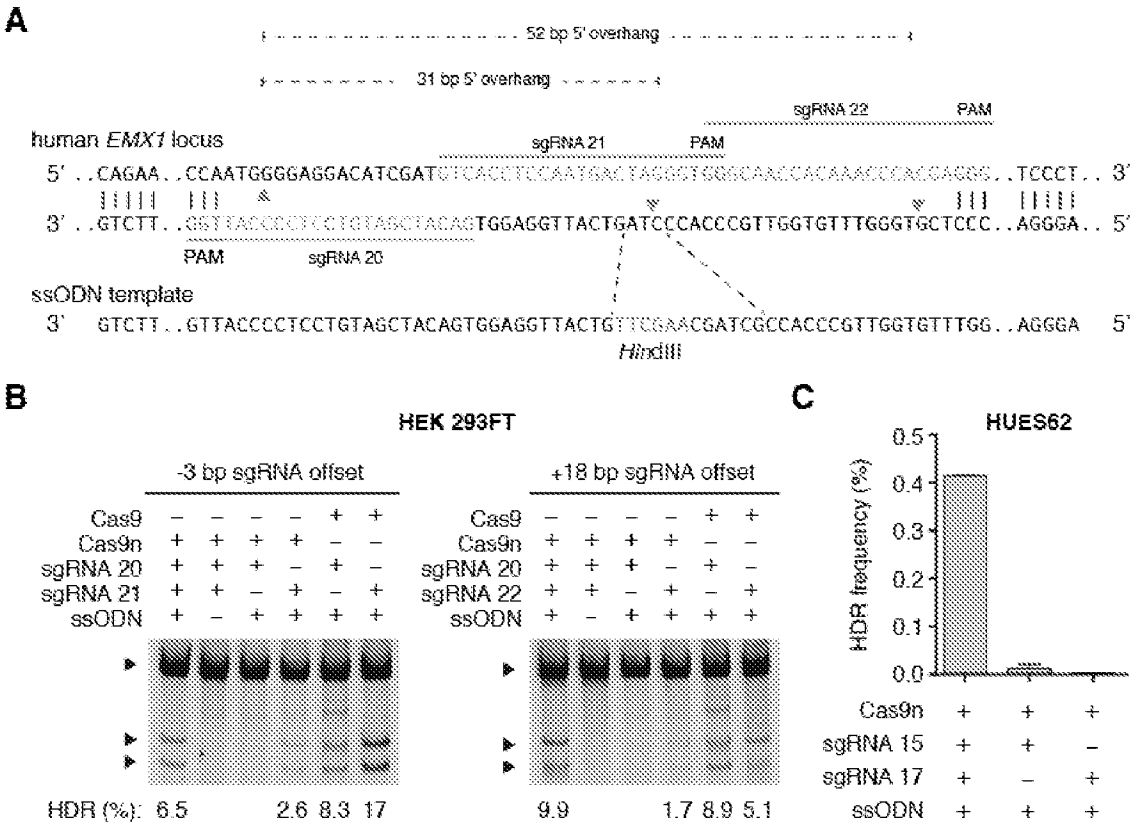

A
52 bp 5' overhang
31 bp 5' overhang
sgRNA 22
PAM
human EMX1 locus
sgRNA 21
PAM
5' ..CAGAA..CCAATGGGGAGGACATCGATGTCACCTCCAATGACTAGGGTGGGCAACCACAAACCCACGAGGG..TCCCT.. 3'
3' ..GTCTT..GGTTACCCCTCCTGTAGCTACAGTGGAGGTTACTGATCCCACCCGTTGGTGTTTGGGTGCTCCC..AGGGA.. 5'
PAM
sgRNA 20
ssODN template
3' GTCTT..GTTACCCCTCCTGTAGCTACAGTGGAGGTTACTGTTCGAACGATCGCCACCCGTTGGTGTTTGG..AGGGA 5'
HindIII
B
HEK 293FT
-3 bp sgRNA offset
Cas9
Cas9n
sgRNA 20
sgRNA 21
ssODN
HDR (%): 6.5 2.6 8.3 17
+18 bp sgRNA offset
Cas9
Cas9n
sgRNA 20
sgRNA 22
ssODN
9.9 1.7 8.9 5.1
C
HUES62
HDR frequency (%)
0.5
0.4
0.3
0.2
0.1
0.0
Cas9n
sgRNA 15
sgRNA 17
ssODN

FIG. 42A-C

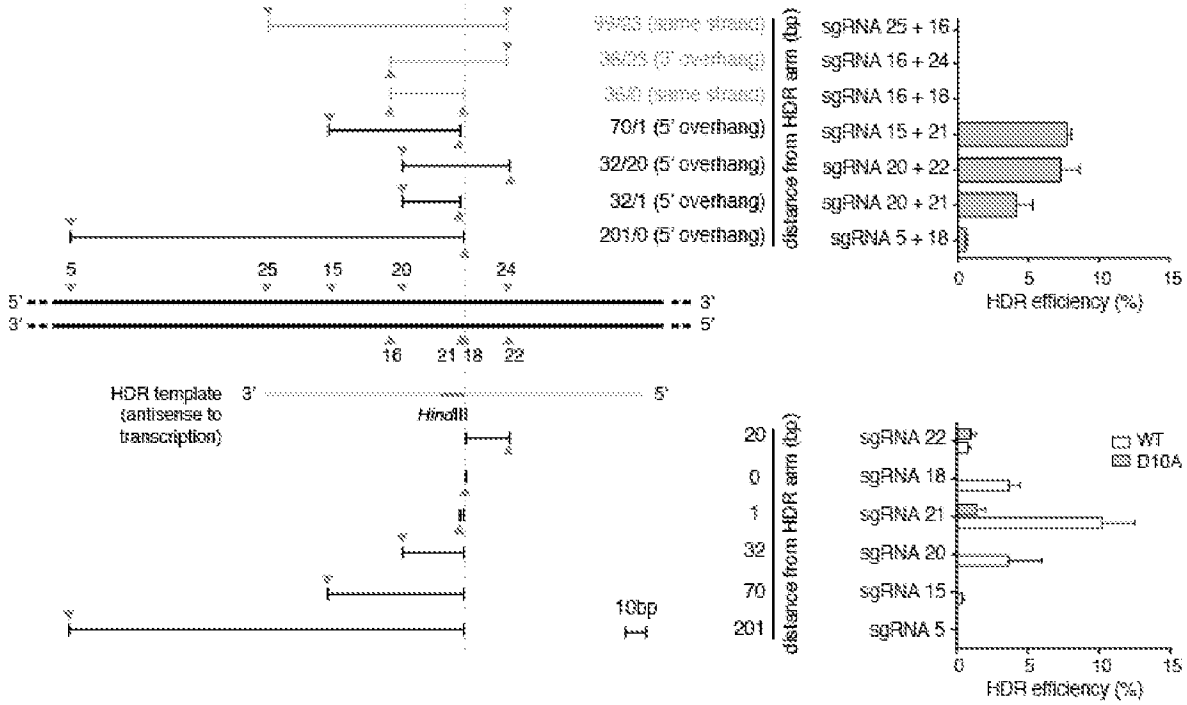

99/23 (same strand)
36/25 (5' overhang)
36/0 (same strand)
70/1 (5' overhang)
32/20 (5' overhang)
32/1 (5' overhang)
201/0 (5' overhang)
distance from HDR arm (bp)
sgRNA 25 + 16
sgRNA 16 + 24
sgRNA 16 + 18
sgRNA 15 + 21
sgRNA 20 + 22
sgRNA 20 + 21
sgRNA 5 + 18
0 5 10 15
HDR efficiency (%)
5 25 15 20 24
5' 3'
3' 5'
16 21 18 22
HDR template (antisense to transcription) 3' 5'
HindIII
10bp
20
0
1
32
70
201
distance from HDR arm (bp)
sgRNA 22
sgRNA 18
sgRNA 21
sgRNA 20
sgRNA 15
sgRNA 5
WT
D10A
0 5 10 15
HDR efficiency (%)

FIG. 43

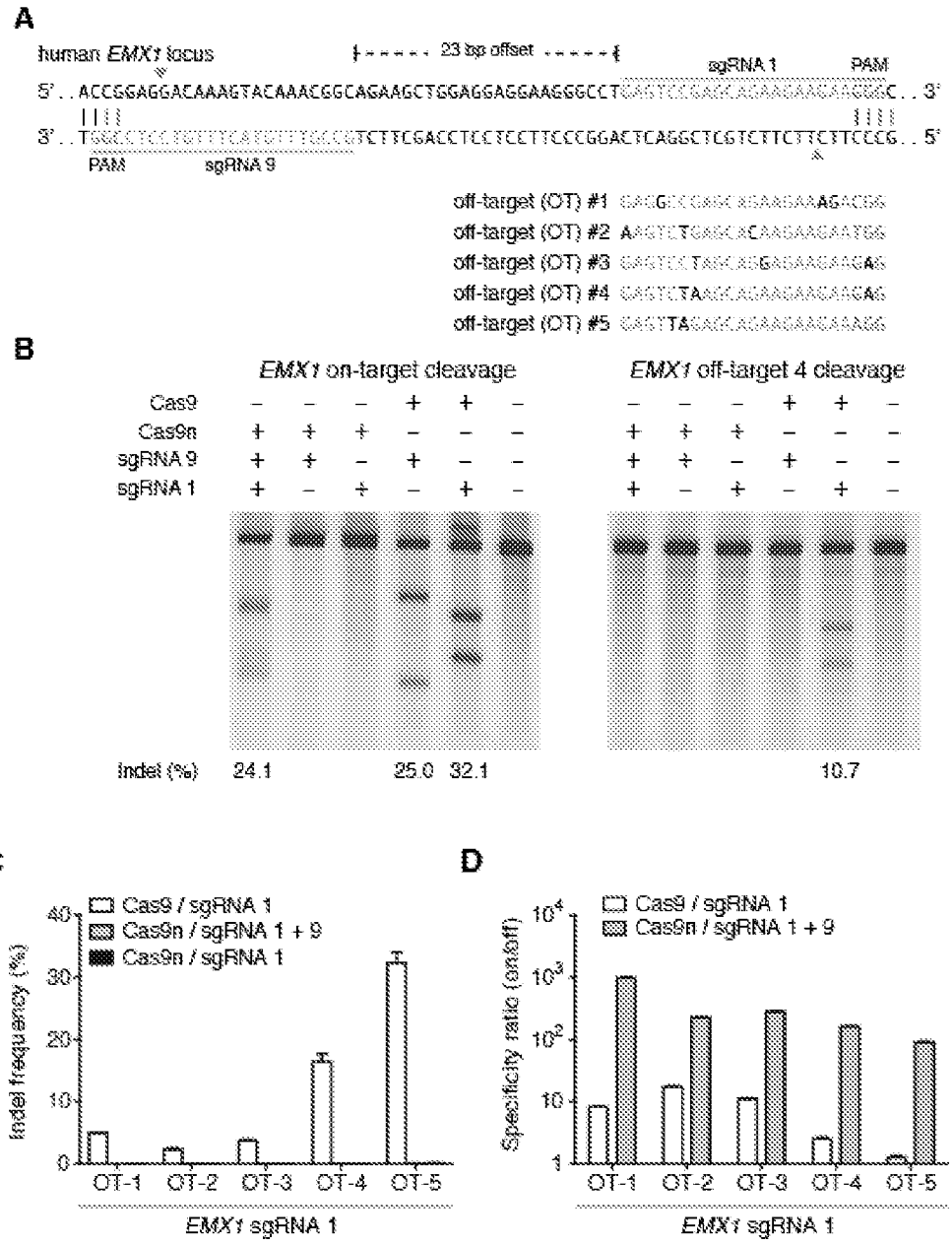
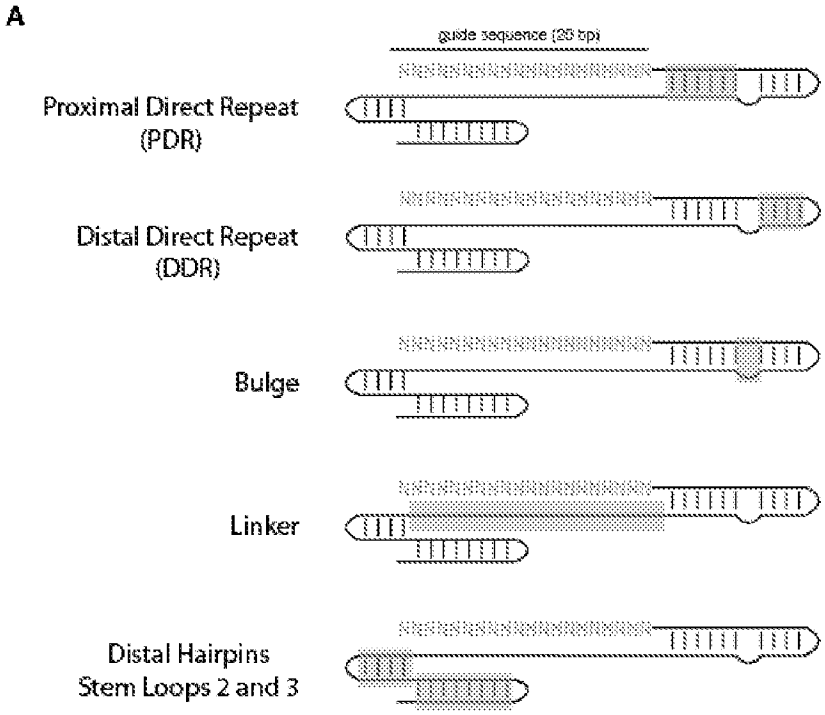

A
human EMX1 locus
23 bp offset
sgRNA 1
PAM
5'..ACCGGAGGACAAAGTACAAACGGCAGAAGCTGGAGGAGGAAGGGCCTGAGTCCGAGCAGAAGAAGAAGGGC..3'
3'..TGGCCTCCTGTTTCATGTTTGCCGTCTTCGACCTCCTCCTTCCCGGACTCAGGCTCGTCTTCTTCTTCCCG..5'
PAM
sgRNA 9
off-target (OT) #1
off-target (OT) #2
off-target (OT) #3
off-target (OT) #4
off-target (OT) #5
B
EMX1 on-target cleavage
EMX1 off-target 4 cleavage
Cas9
Cas9n
sgRNA 9
sgRNA 1
Indel (%)
24.1
25.0
32.1
10.7
C
Cas9 / sgRNA 1
Cas9n / sgRNA 1 + 9
Cas9n / sgRNA 1
Indel frequency (%)
OT-1
OT-2
OT-3
OT-4
OT-5
EMX1 sgRNA 1
D
Cas9 / sgRNA 1
Cas9n / sgRNA 1 + 9
Specificity ratio (on/off)
OT-1
OT-2
OT-3
OT-4
OT-5
EMX1 sgRNA 1

FIG. 44A-D


A
guide sequence (20 bp)
Proximal Direct Repeat (PDR)
Distal Direct Repeat (DDR)
Bulge
Linker
Distal Hairpins Stem Loops 2 and 3

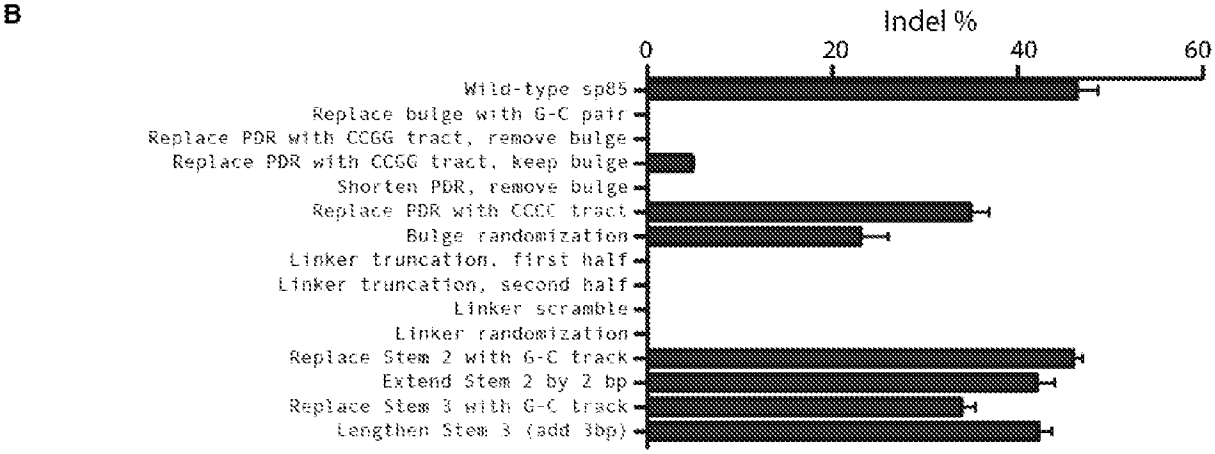

B
Indel %
0
20
40
60
Wild-type sp85
Replace bulge with G-C pair
Replace PDR with CCGG tract, remove bulge
Replace PDR with CCGG tract, keep bulge
Shorten PDR, remove bulge
Replace PDR with CCCC tract
Bulge randomization
Linker truncation, first half
Linker truncation, second half
Linker scramble
Linker randomization
Replace Stem 2 with G-C track
Extend Stem 2 by 2 bp
Replace Stem 3 with G-C track
Lengthen Stem 3 (add 3bp)

FIG. 45A-B

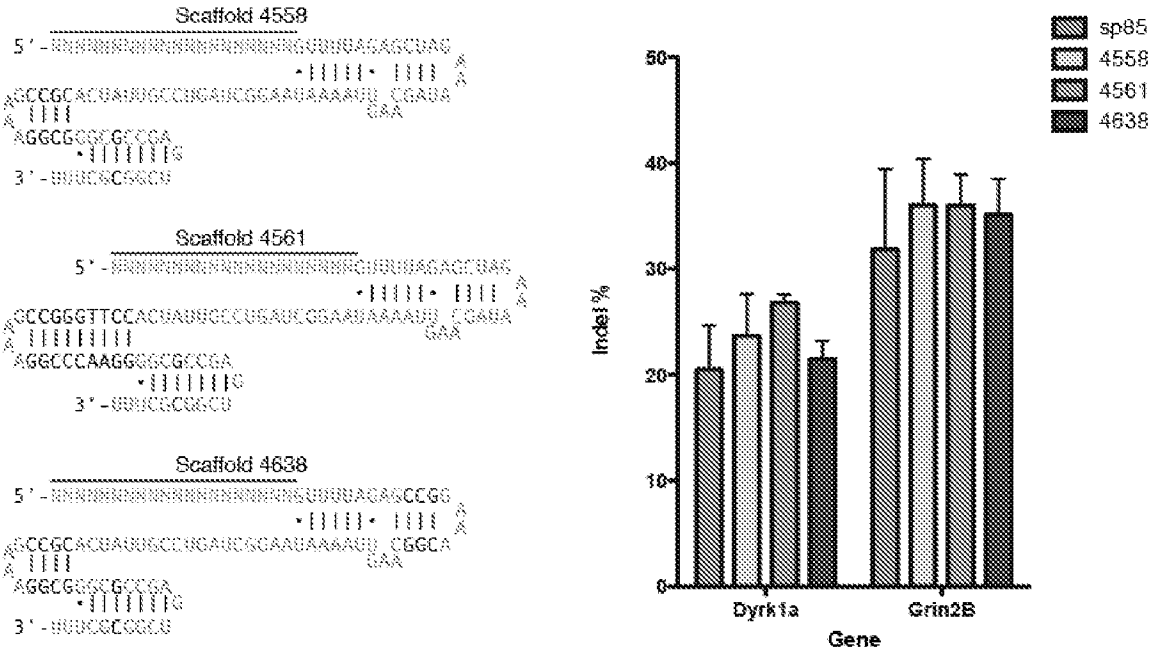

Scaffold 4558
Scaffold 4561
Scaffold 4638
sp85
4558
4561
4638
Indel %
50
40
30
20
10
0
Dyrk1a
Grin2B
Gene

FIG. 46

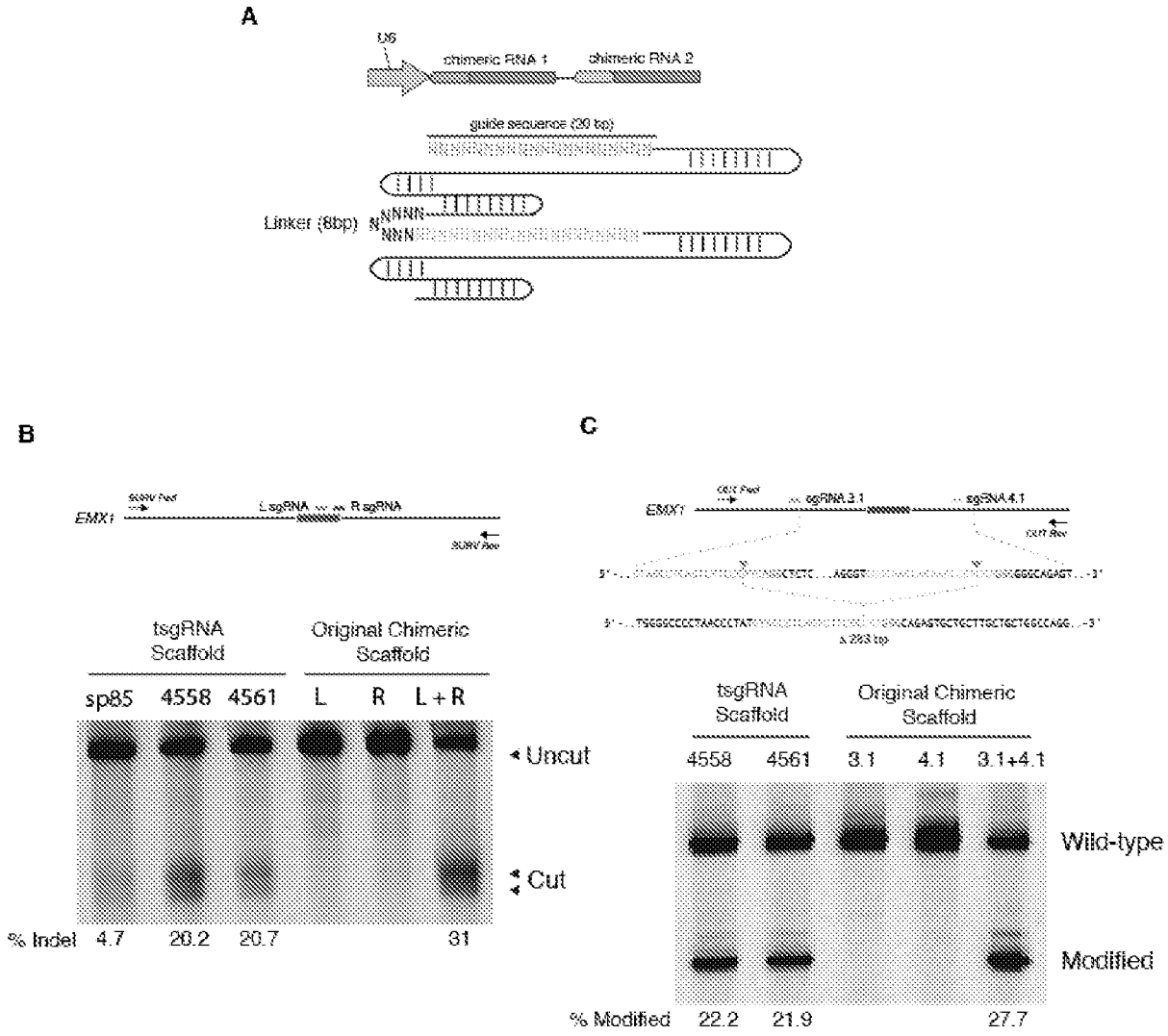
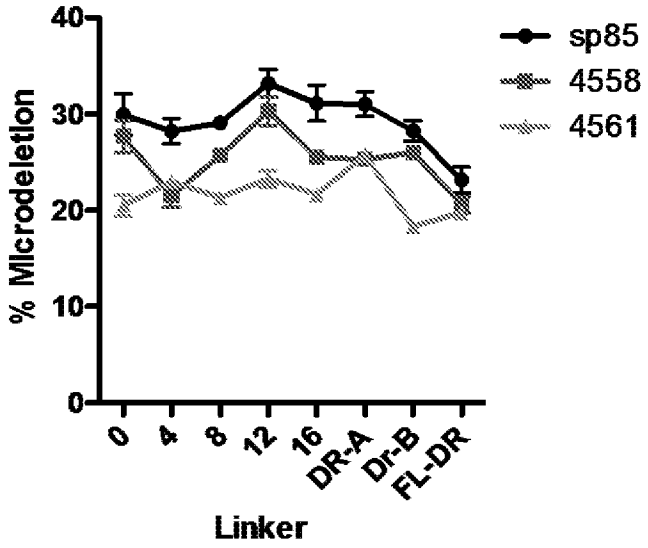

A
US
chimeric RNA 1
chimeric RNA 2
guide sequence (20 bp)
Linker (8bp)
B
EMX1
L sgRNA
R sgRNA
tsgRNA Scaffold
Original Chimeric Scaffold
sp85
4558
4561
L
R
L + R
Uncut
Cut
% Indel
4.7
26.2
20.7
31
C
EMX1
sgRNA 3.1
sgRNA 4.1
tsgRNA Scaffold
Original Chimeric Scaffold
4558
4561
3.1
4.1
3.1+4.1
Wild-type
Modified
% Modified
22.2
21.9
27.7

FIG. 47A-C



Tandem Linker	Length	Sequence
0	0	
4	4	ATTA
8	8	AATTATTA
12	12	AATTATTAATTA
16	16	AATTATTAATTATAAT
Direct Repeat A	16	GTTTTAGAGCTATGCT
Direct Repeat B	20	GTTTTGAATGGTCCCAAAAC
Full Direct Repeat	36	GTTTTAGAGCTATGCTGTTTTGAATGGTCCCAAAAC

FIG. 48

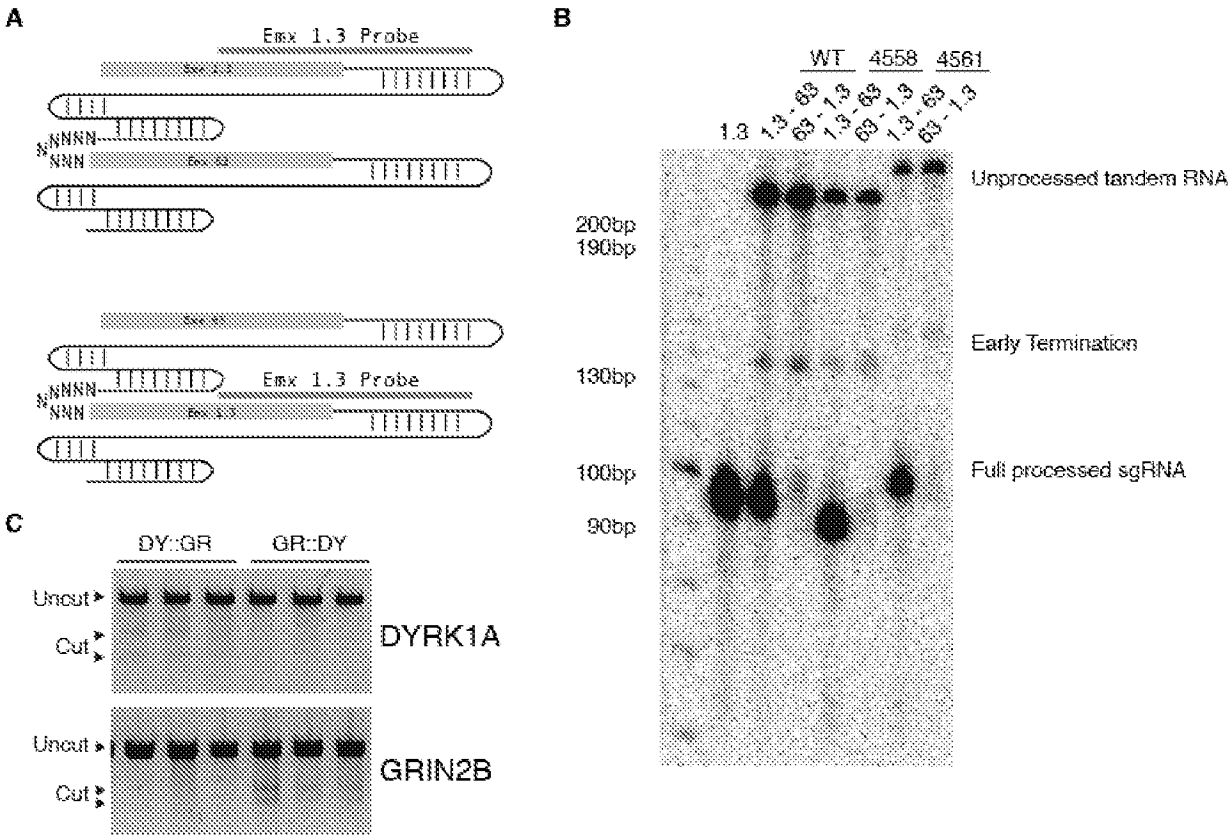

A
Emx 1.3 Probe
NNNN
NNN
Emx 1.3 Probe
B
WT
4558
4561
1.3
1.3 - 63
63 - 1.3
1.3 - 63
63 - 1.3
1.3 - 63
63 - 1.3
200bp
190bp
130bp
100bp
90bp
Unprocessed tandem RNA
Early Termination
Full processed sgRNA
C
DY::GR
GR::DY
Uncut
Cut
DYRK1A
Uncut
Cut
GRIN2B

FIG. 49A-C

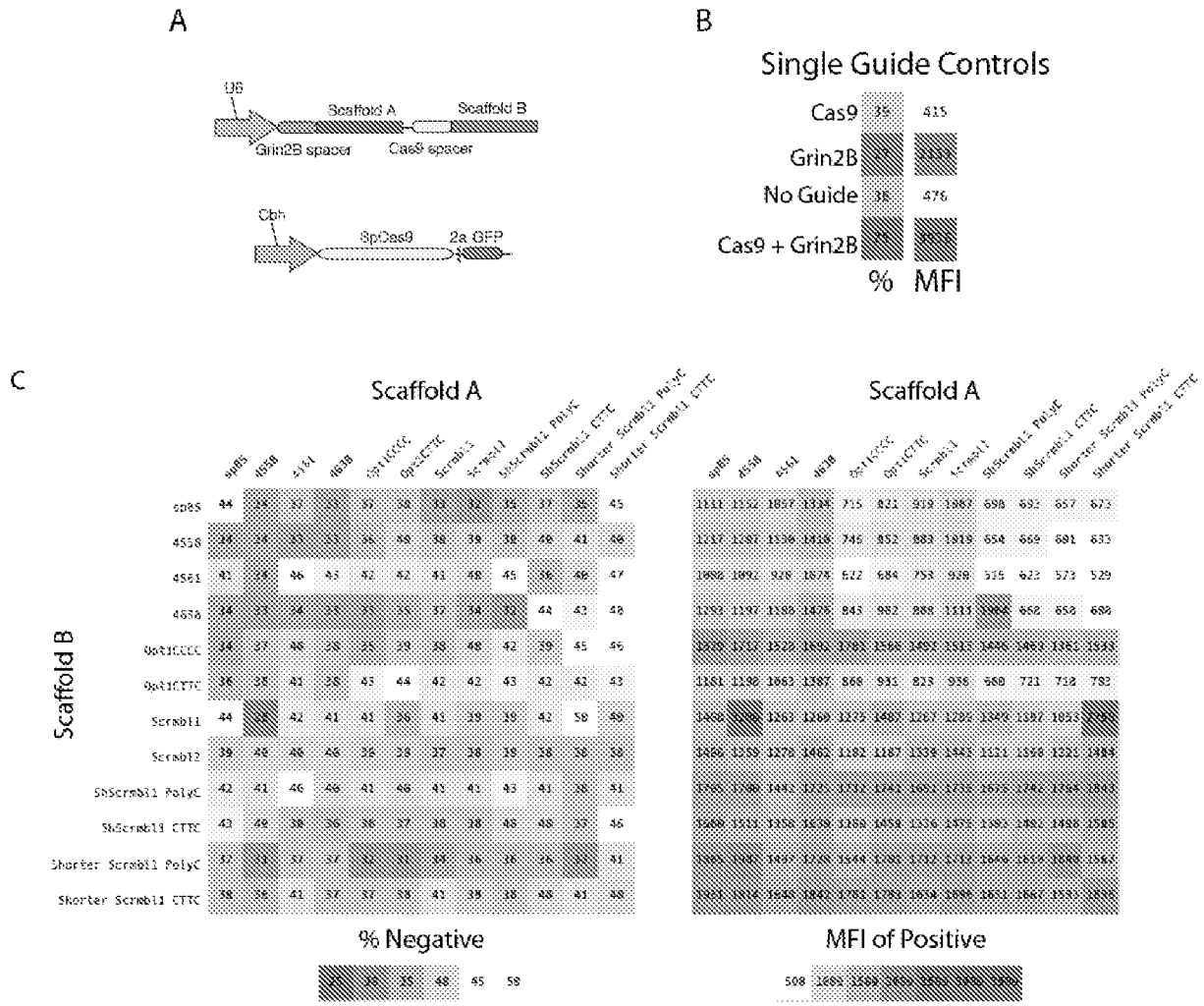

A
U6
Scaffold A
Scaffold B
Grin2B spacer
Cas9 spacer
Cbh
SpCas9
2a GFP
B
Single Guide Controls
Cas9
Grin2B
No Guide
Cas9 + Grin2B
% MFI
C
Scaffold A
Scaffold B
% Negative
MFI of Positive

FIG. 50A-C

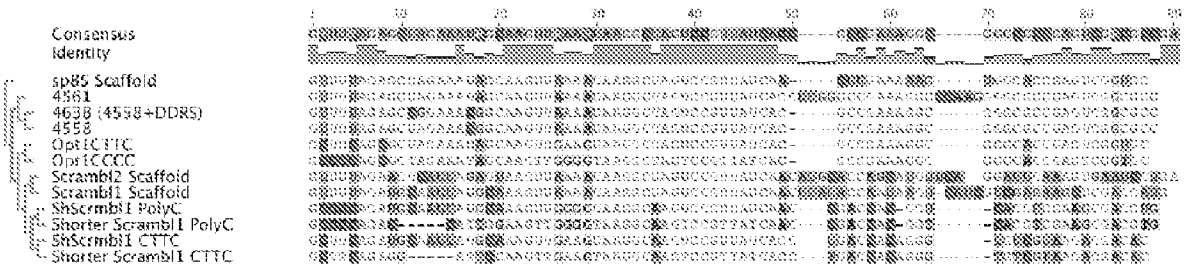
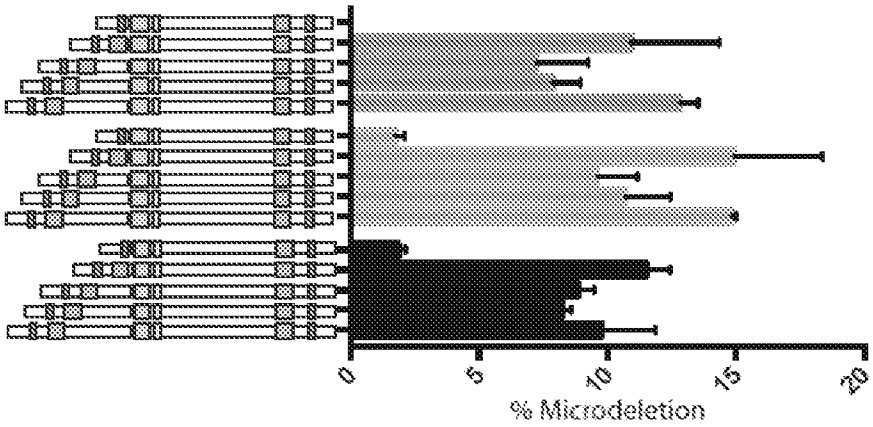

Consensus
Identity
sp85 Scaffold
4561
4638 (4558+DDRS)
4558
Opt1C CTTC
Opt1CCCC
Scramb12 Scaffold
Scramb11 Scaffold
ShScramb11 PolyC
Shorter Scramb11 PolyC
ShScramb11 CTTC
Shorter Scramb11 CTTC

FIG. 51

A. Truncated bi-directional U6 Promoter is able to induce micro-deletion.



B. Examination of indel activity from single guides being expressed off of bi-directional promoter

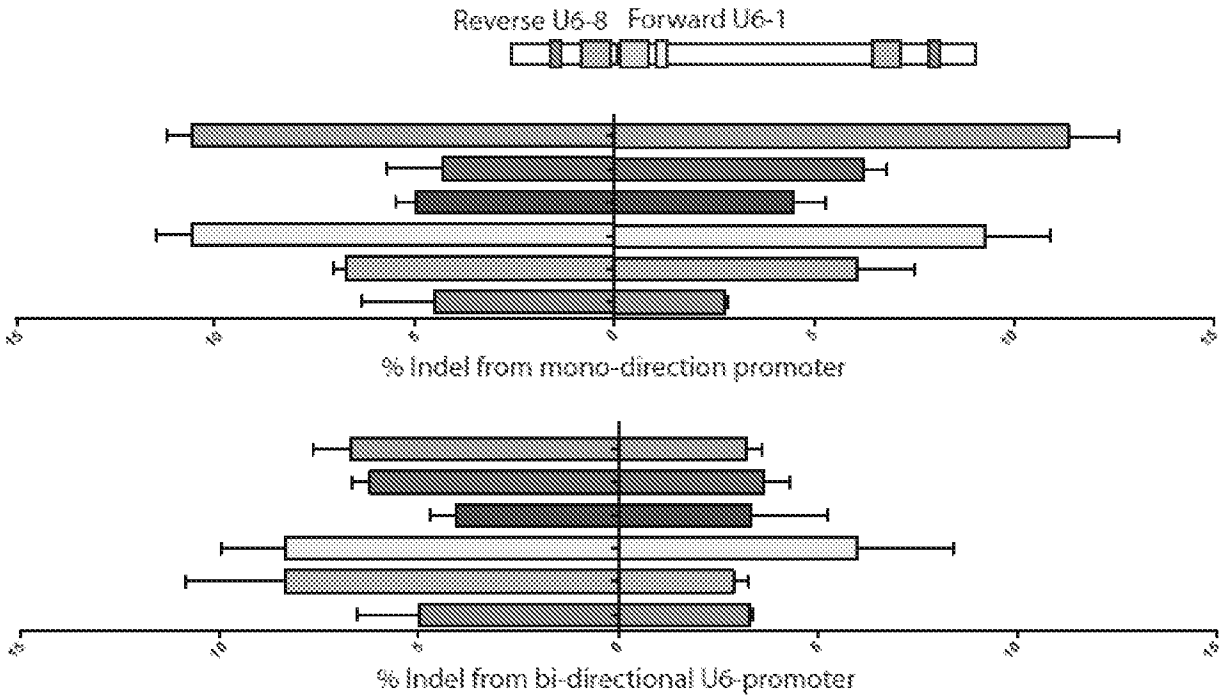


FIG. 52

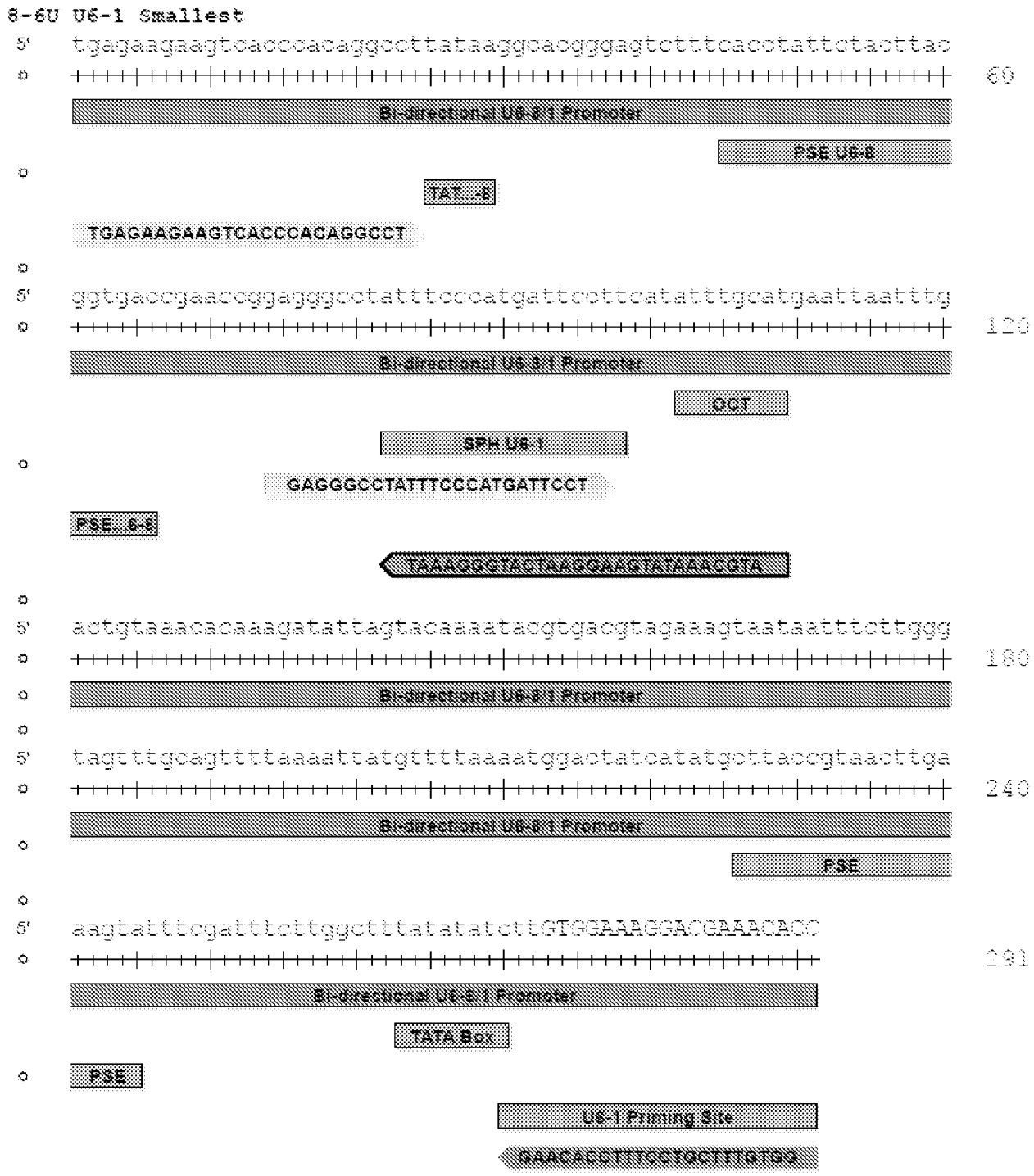

8-6U U6-1 Smallest
5' tgagaagaagtcacccacaggccttataaggcacgggagtctttcacctattctacttac
60
Bi-directional U6-8/1 Promoter
PSE U6-8
TAT -8
TGAGAAGAAGTCACCCACAGGCCT
5' ggtgaccgaaccggagggcctatttcccatgattccttcatatttgcatgaattaatttg
120
Bi-directional U6-8/1 Promoter
OCT
SPH U6-1
GAGGGCCTATTTCCCATGATTCCT
PSE 6-8
TAAAGGGTACTAAGGAAGTATAAACGTA
5' actgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttggg
180
Bi-directional U6-8/1 Promoter
5' tagtttgcagttttaaaattatgttttaaaatggactatcatatgcttaccgtaacttga
240
Bi-directional U6-8/1 Promoter
PSE
5' aagtatttcgatttcttggctttatatatcttGTGGAAAGGACGAAACACC
291
Bi-directional U6-8/1 Promoter
TATA Box
PSE
U6-1 Priming Site
GAACACCTTTCCTGCTTTGTGG

FIG. 53