



(51) International Patent Classification:

A61K 48/00 (2006.01) G06F 19/00 (2018.01)

(21) International Application Number:

PCT/US2018/041041

(22) International Filing Date:

06 July 2018 (06.07.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/529,573 07 July 2017 (07.07.2017) US

(71) Applicants: **THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, Massachusetts 02142 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US). **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, Massachusetts 02138 (US). **THE NEW YORK GENOME CENTER** [US/US]; 101 AVENUE OF THE AMERICAS, New York, New York 10013 (US). **NEW YORK UNIVERSITY** [US/US]; ONE PARK AVENUE, 6TH FLOOR, New York, New York 10016 (US).

(72) Inventors: **SANJANA, Neville**; c/o 101 AVENUE OF THE AMERICAS, New York, New York 10013 (US). **ZHANG, Feng**; c/o 415 MAIN STREET, CAMBRIDGE, Massachusetts 02142 (US). **MEIER, Joshua A.**; c/o 17 QUINCY STREET, CAMBRIDGE, Massachusetts 02138 (US).

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

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

(84) Designated States (unless otherwise indicated, for every kind of regional protection available): ARIPO (BW, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK,

(54) Title: METHODS FOR DESIGNING GUIDE SEQUENCES FOR GUIDED NUCLEASES

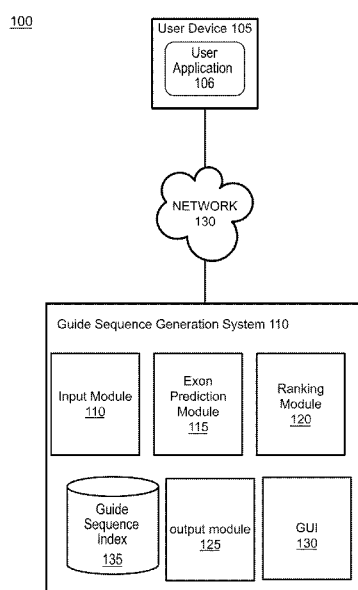


FIG. 1

(57) Abstract: Embodiments disclosed herein provide methods, including computer-implemented methods, for designing guide sequence which may be incorporated into custom, large scale guide sequence libraries. The methods require only a list of target genes as input and utilize on target and off target scores to generate an optimal set of guide sequences for a set of target genes. In certain embodiments, the methods may also utilize multi-tissue RNA-sequencing data and/or protein annotation to design targets to genes that are highly expressed and/or contain a functional protein domain. The invention further comprises guide libraries, cells comprising said guide libraries. Computer-implemented embodiments further improve computer system function by reducing excessive user wait time through the use of data structures that reduce search from linear to logarithmic time.

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

METHODS FOR DESIGNING GUIDE SEQUENCES FOR GUIDED NUCLEASES

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0001] This invention was made with government support under grant numbers MH100706 MH110049, and HG008171 granted by the National Institutes of Health. The government has certain rights in the invention.

TECHNICAL FIELD

[0002] The subject matter disclosed herein is generally directed to methods and systems for designing guide sequences and guide sequence libraries for guided nucleases.

BACKGROUND

[0003] Genome-scale CRISPR-Cas9 knockout libraries have emerged as powerful tools for unbiased, phenotypic screening. These libraries contain a fixed number of Cas9 single-guide RNAs (sgRNAs) targeting each gene in the genome and typically require large numbers of cells (>10⁸) to maintain genome-scale representation. However, there are many applications where it would be preferable to design a custom library targeting specific genes sets (*e.g.*, kinases, transcription factors, chromatin modifiers, the druggable genome) with higher coverage for these specific genes.

SUMMARY

[0004] In one aspect, the invention provides a computer-implemented method for making CRISPR-Cas system guide sequences comprising: identifying, using one or more computing devices, exons in one or more target genes; identifying, by the one or more computing devices, potential CRISPR-Cas system target regions in each exon; ranking, by the one or more computing devices, each identified potential CRISPR-Cas system target regions based at least in part by an off-target avoidance score and an on-target efficiency score determined using a classifier, and optionally, by the presence of a protein domain in the cutting region; generating, by the one or more computing devices, a set of guide sequences based at least in part on the ranked CRISPR-Cas targeting regions, wherein the set of guide sequences are designed to target

regions ranked above a defined cut-off value. In one embodiment, exons are defined as a union of consensus CDS (CCDS) regions. In another embodiment, the one or more target genes are identified based on an input nucleic acid sequence, a gene symbol, or combination thereof. In another embodiment, the potential CRISPR-Cas target regions are identified by proximity to a PAM. In still another embodiment, the exons are selected based on tissue-specific expression data to select exons with higher expression. In another embodiment, selecting coding regions based on tissue-specific expression data comprises determining a median expression for each exon in each tissue and determining a median expression value for each exon in each tissue for all samples of a same tissue type, and normalizing the median read counts by coding region base pair size. In further embodiments, the off-target avoidance score is determined by taking the sum of a cutting frequency determination score for each off-target site identified in an exome, or the on-target efficiency score is determined using a classifier. In still further embodiments, the classifier is applied to local sequence preferences learned from saturation mutagenesis studies. In another embodiment, the classifier is a boosted regression tree classifier. In another embodiment, the method further comprises one or more of eliminating guide sequences with homopolymer repeats, excluding the last exon in a coding region, eliminating guide sequences with transcriptional terminators, or a combination thereof. In another embodiment, the method further comprises generating non-targeting guide sequences to serve as a negative control. In another embodiment, the identified exons are species specific. Other embodiments provide guide sequences that are based on a full length sgRNA scaffold, an E+F scaffold, or a truncated guide sequence scaffold. In another embodiment, the CRISPR-Cas system is a CRISPR-Cas Type II system, a CRISPR-Cas V system, or a CRISPR-Cas VI system. In other embodiments, the method further comprises preparing a guide library comprising the set of guide sequences, such as RNA, DNA, DNA-RNA hybrid, or chemically modified guide sequences.

[0005] In another aspect, the invention provides a guide library made using the methods as described herein. In a further aspect, the invention provides a guide library comprising guide sequences to one or more target regions in one or more exons of one or more target genes, wherein individual guide sequences in the library are included based on optimization of an off-target avoidance score and an on-target efficiency score, and optionally, by the presence of a protein domain in the target region. In one embodiment, the exons are selected based on tissue-

specific expression data to select exons with higher expression. In another embodiment, the off-target avoidance score is determined by taking the sum of a cutting frequency determination score for each off-target site identified in an exome of the one or more target genes. In another embodiment, the on-target efficiency is determined by use of a classifier applied to local sequence preferences learned from saturation mutagenesis studies. In another embodiment, the classifier is a boosted regression tree classifier. Other embodiments provide guide sequences that exclude guide sequences targeting homopolymer regions, targeting the last exon in a coding region, include target regions with transcriptional terminators, or a combination thereof. In still further embodiments, the guide sequences are full length guide sequences, truncated guide sequences, full length sgRNA sequences, truncated sgRNA sequence, or E+F sgRNA sequences, or the guide sequences are RNA, DNA, DNA-RNA hybrid, chemically modified, or a combination thereof.

[0006] In another aspect, the invention provides a composition comprising (i) a population of cells and (ii) a guide sequence library as described herein, where each of the cells contains one or more of the guide sequences and thus the guides sequences of the library are integrated into the population of cells. In one embodiment, the population of cells is a eukaryotic population of cells.

[0007] In another aspect, the invention provides a kit comprising a guide sequence library as described herein, and/or a composition as described herein.

[0008] These and other aspects, objects, features, and advantages of the example embodiments will become apparent to those having ordinary skill in the art upon consideration of the following detailed description of illustrated example embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0009] **FIG. 1** – Shows a block diagram depicting a system for generating a guide sequence library in accordance with certain example embodiments.

[0010] **FIG. 2** – Shows a block diagram depicting a method or making guide nucleic acids for nucleic acid guided nucleases in accordance with certain example embodiments.

[0011] **FIG. 3** – Shows a block diagram depicting a computing machine and a module, in accordance with certain example embodiments.

[0012] FIG. 4 – Provides a graph showing integration of GTEx data shifts sgRNA targeting toward highly expressed exons. 1000 GUIDES-generated libraries using data for median expression across all GTEx tissues (red) or without using GTEx data (blue) to choose exons to target. For each library, 500 genes were selected at random from the human genome and GUIDES was instructed to design 5 sgRNAs per gene in the selected exons. On average, incorporation of GTEx data increases average expression of targeted exons by a factor of 1.5.

[0013] FIG. 5 – Provides a graph showing aggregate cut-frequency determination (CFD) off-target score reduces predicted off-target sites for selected sgRNAs. GUIDES calculate the sum of the cut-frequency determination (CFD) score 4 for all 0-3 bp mismatches in the human/mouse exome (“G-score”). When optimizing sgRNAs for off target avoidance/specificity, designed human and mouse libraries (n = 2,000 genes) have fewer sgRNAs with a high G-score (i.e. fewer sgRNAs with 0-3 bp potential exome off-targets).

[0014] FIG. 6 – Provides a graph showing on-target scores at each stage of the GUIDES pipeline. Average on-target scores for a sgRNA library targeting 2,654 transcription factors in the human genome with 3 sgRNAs per gene (a) or 6 sgRNAs per gene (b). On-target (efficiency) scores were calculated using the Microsoft Azimuth algorithm as in ref. 4. Despite on-target optimization being the last stage of the GUIDES pipeline (see Supplementary methods), on-target score optimization increases the average on-target score of chosen sgRNAs even after other optimizations such as off-target- and protein domain-based prioritization.

[0015] FIG. 7 – Provides a graph showing selected guide sequences that targeted essential genes display greater depletion in a meta-analysis of 77 pooled CRISPR screens. Cumulative density function of GenomeCRISPR sgRNA effect scores for GUIDES-selected sgRNAs versus a matched-size sample of sgRNAs targeting the same genes randomly chosen from the GenomeCRISPR database. The average increase in depletion by using GUIDES-generated sgRNAs over the size-matched randomly selected sets was 0.73 sgRNA effect (~10% increased depletion, n= 403 genes examined in 77 genome-scale screens using 61 different cell lines), which is significantly greater depletion ($p= 5e-07$, $t= -5.1$, $df= 409$, two-sample paired t-test).

[0016] FIG. 8 – Provides a graph showing library generation time scales linearly with number of genes targeted. The indicated number of genes was selected from the human genome and the time required for library generation was tracked as a function of the number of sgRNAs

per gene. Gene count and generation time are linearly correlated ($r^2 > 0.99$ over a range of 5-20,000 genes targeted for 3, 6, 10, and 20 sgRNAs/gene). Since all potential sgRNAs for each gene are precomputed, generation times are not affected by the number of sgRNAs. For benchmarking, GUIDES was run on a computer with a 2.5 GHz Intel Core i7 processor and 16 GB of memory running Linux (Ubuntu 14).

[0017] FIG 9. – Provides a screenshot of an example GUI output using the methods disclosed herein. The interactive designer allows adding and deleting guide sequences. For each guide sequence, the tool displays the exon targeted, the functional domain(s) in the protein targeted (if any), the efficiency (on-target) score, and specificity (off-target score). The three-panel example GUID allows rapid navigation between gene-level and guide sequence-level changes.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

General Definitions

[0018] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in *Molecular Cloning: A Laboratory Manual*, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); *Molecular Cloning: A Laboratory Manual*, 4th edition (2012) (Green and Sambrook); *Current Protocols in Molecular Biology* (1987) (F.M. Ausubel et al. eds.); the series *Methods in Enzymology* (Academic Press, Inc.); *PCR 2: A Practical Approach* (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.); *Antibodies, A Laboratory Manual* (1988) (Harlow and Lane, eds.); *Antibodies A Laboratory Manual*, 2nd edition 2013 (E.A. Greenfield ed.); *Animal Cell Culture* (1987) (R.I. Freshney, ed.); Benjamin Lewin, *Genes IX*, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), *The Encyclopedia of Molecular Biology*, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), *Molecular Biology and Biotechnology: a Comprehensive Desk Reference*, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, *Dictionary of Microbiology and Molecular Biology* 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, *Advanced Organic Chemistry Reactions, Mechanisms and Structure* 4th ed., John Wiley & Sons (New York, N.Y.

1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011) .

[0019] As used herein, the singular forms “a”, “an”, and “the” include both singular and plural referents unless the context clearly dictates otherwise.

[0020] The term “optional” or “optionally” means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[0021] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0022] The terms “about” or “approximately” as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier “about” or “approximately” refers is itself also specifically, and preferably, disclosed.

[0023] Reference throughout this specification to “one embodiment”, “an embodiment,” “an example embodiment,” means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases “in one embodiment,” “in an embodiment,” or “an example embodiment” in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

[0024] For ease of reference, the methods disclosed herein may be referred to in various instances as “GUIDES.”

[0025] All publications, published patent documents, and patent applications cited in this application are indicative of the level of skill in the art(s) to which the application pertains. All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

Overview

[0026] Embodiments disclosed herein provide methods, including computer-implemented methods, for designing guide sequence which may be incorporated into custom, large scale guide sequence libraries. The methods require only a list of target genes as input and utilize on target and off target scores to generate an optimal set of guide sequences for a set of target genes. In certain embodiments, the methods may also utilize multi-tissue RNA-sequencing data and/or protein annotation to design targets to genes that are highly expressed and/or contain a functional protein domain. The invention further comprises guide libraries, and cells comprising said guide libraries. Computer-implemented embodiments further improve computer system function by reducing excessive user wait time through the use of data structures that reduce search from linear to logarithmic time.

Guide Sequences

[0027] The guide molecule or guide RNA of a Class 2 type V CRISPR-Cas protein comprises a tracr-mate sequence (encompassing a “direct repeat” in the context of an endogenous CRISPR system) and a guide sequence (also referred to as a “spacer” in the context of an endogenous CRISPR system). Indeed, in contrast to the type II CRISPR-Cas proteins, the Cas13 protein does not rely on the presence of a tracr sequence. In some embodiments, the CRISPR-Cas system or complex as described herein does not comprise and/or does not rely on the presence of a tracr sequence (e.g. if the Cas protein is Cas13). In certain embodiments, the guide molecule may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence.

[0028] In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence. In the context of formation of a CRISPR complex, “target sequence” refers to a sequence to which a guide sequence is designed to have

complementarity, where hybridization between a target DNA sequence and a guide sequence promotes the formation of a CRISPR complex.

[0029] The terms “guide molecule” and “guide RNA” are used interchangeably herein to refer to RNA-based molecules that are capable of forming a complex with a CRISPR-Cas protein and comprises a guide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of the complex to the target nucleic acid sequence. The guide molecule or guide RNA specifically encompasses RNA-based molecules having one or more chemical modifications (e.g., by chemical linking two ribonucleotides or by replacement of one or more ribonucleotides with one or more deoxyribonucleotides), as described herein.

[0030] As used herein, the term "crRNA" or "guide RNA" or "single guide RNA" or "sgRNA" or "one or more nucleic acid components" of a Type V or Type VI CRISPR-Cas locus effector protein comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. In some embodiments, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid-targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the

target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting guide may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmic RNA (scRNA). In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[0031] In some embodiments, the guide molecule comprises a guide sequence that is designed to have at least one mismatch with the target sequence, such that an RNA duplex formed between the guide sequence and the target sequence comprises a non-pairing C in the guide sequence opposite to the target A for deamination on the target sequence. In some embodiments, aside from this A-C mismatch, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more.

[0032] As used herein, the term "crRNA" or "guide RNA" or "single guide RNA" or "sgRNA" or "one or more nucleic acid components" of a Type V or Type VI CRISPR-Cas locus effector protein comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid

sequence. In some embodiments, the degree of complementarity, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable algorithm for aligning sequences, non-limiting example of which include the Smith-Waterman algorithm, the Needleman-Wunsch algorithm, algorithms based on the Burrows-Wheeler Transform (e.g., the Burrows Wheeler Aligner), ClustalW, Clustal X, BLAT, Novoalign (Novocraft Technologies; available at www.novocraft.com), ELAND (Illumina, San Diego, CA), SOAP (available at soap.genomics.org.cn), and Maq (available at maq.sourceforge.net). The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid-targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting guide may be selected to target any target nucleic acid sequence. The target sequence may be DNA. The target sequence may be any RNA sequence. In some embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of messenger RNA (mRNA), pre-mRNA, ribosomal RNA (rRNA), transfer RNA (tRNA), micro-RNA (miRNA), small interfering RNA (siRNA), small nuclear RNA (snRNA), small nucleolar RNA (snoRNA), double stranded RNA (dsRNA), non-coding RNA (ncRNA), long non-coding RNA (lncRNA), and small cytoplasmatic RNA (scRNA). In some preferred embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of mRNA, pre-mRNA, and rRNA. In some preferred

embodiments, the target sequence may be a sequence within a RNA molecule selected from the group consisting of ncRNA, and lncRNA. In some more preferred embodiments, the target sequence may be a sequence within an mRNA molecule or a pre-mRNA molecule.

[0033] In some embodiments, a nucleic acid-targeting guide is selected to reduce the degree of secondary structure within the nucleic acid-targeting guide. In some embodiments, about or less than about 75%, 50%, 40%, 30%, 25%, 20%, 15%, 10%, 5%, 1%, or fewer of the nucleotides of the nucleic acid-targeting guide 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).

[0034] In certain embodiments, a guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat (DR) sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or crRNA may comprise, consist essentially of, or consist of a direct repeat sequence fused or linked to a guide sequence or spacer sequence. In certain embodiments, the direct repeat sequence may be located upstream (i.e., 5') from the guide sequence or spacer sequence. In other embodiments, the direct repeat sequence may be located downstream (i.e., 3') from the guide sequence or spacer sequence.

[0035] In certain embodiments, the crRNA comprises a stem loop, preferably a single stem loop. In certain embodiments, the direct repeat sequence forms a stem loop, preferably a single stem loop.

[0036] In certain embodiments, the spacer length of the guide RNA is from 15 to 35 nt. In certain embodiments, the spacer length of the guide RNA is at least 15 nucleotides. In certain embodiments, the spacer length is from 15 to 17 nt, e.g., 15, 16, or 17 nt, from 17 to 20 nt, e.g., 17, 18, 19, or 20 nt, from 20 to 24 nt, e.g., 20, 21, 22, 23, or 24 nt, from 23 to 25 nt, e.g., 23, 24, or 25 nt, from 24 to 27 nt, e.g., 24, 25, 26, or 27 nt, from 27-30 nt, e.g., 27, 28, 29, or 30 nt, from 30-35 nt, e.g., 30, 31, 32, 33, 34, or 35 nt, or 35 nt or longer.

[0037] The "tracrRNA" sequence or analogous terms includes any polynucleotide sequence that has sufficient complementarity with a crRNA sequence to hybridize. In some embodiments, the degree of complementarity between the tracrRNA sequence and crRNA 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 crRNA 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.

[0038] In general, degree of complementarity is with reference to the optimal alignment of the sca sequence and tracr sequence, along the length of the shorter of the two sequences. Optimal alignment may be determined by any suitable alignment algorithm, and may further account for secondary structures, such as self-complementarity within either the sca sequence or tracr sequence. In some embodiments, the degree of complementarity between the tracr sequence and sca sequence along the length of the shorter of the two when optimally aligned is about or more than about 25%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 95%, 97.5%, 99%, or higher.

[0039] In general, the CRISPR-Cas or CRISPR system may be as used in the foregoing documents, such as WO 2014/093622 (PCT/US2013/074667) and refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated ("Cas") genes, including sequences encoding a Cas gene, in particular a Cas13 gene in the case of CRISPR-Cas13, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a "direct repeat" and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a "spacer" in the context of an endogenous CRISPR system), or

"RNA(s)" as that term is herein used (e.g., RNA(s) to guide Cas13, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). In the context of formation of a CRISPR complex, "target sequence" refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. The section of the guide sequence through which complementarity to the target sequence is important for cleavage activity is referred to herein as the seed sequence. A target sequence may comprise any polynucleotide, such as DNA or RNA polynucleotides. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell, and may include nucleic acids in or from mitochondrial, organelles, vesicles, liposomes or particles present within the cell. In some embodiments, especially for non-nuclear uses, NLSs are not preferred. In some embodiments, a CRISPR system comprises one or more nuclear exports signals (NESs). In some embodiments, a CRISPR system comprises one or more NLSs and one or more NESs. In some embodiments, direct repeats may be identified in silico by searching for repetitive motifs that fulfill any or all of the following criteria: 1. found in a 2Kb window of genomic sequence flanking the type II CRISPR locus; 2. span from 20 to 50 bp; and 3. interspaced by 20 to 50 bp. In some embodiments, 2 of these criteria may be used, for instance 1 and 2, 2 and 3, or 1 and 3. In some embodiments, all 3 criteria may be used.

[0040] In embodiments of the invention the terms guide sequence and guide RNA, i.e. RNA capable of guiding Cas to a target genomic locus, are used interchangeably as in foregoing cited documents such as WO 2014/093622 (PCT/US2013/074667). In general, a guide sequence is any polynucleotide sequence having sufficient complementarity with a target polynucleotide sequence to hybridize with the target sequence and direct sequence-specific binding of a CRISPR complex to the target sequence. In some embodiments, the degree of complementarity between a guide sequence and its corresponding target sequence, when optimally aligned using a suitable alignment algorithm, is about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or more. Optimal alignment may be determined with the use of any suitable

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

[0041] In some embodiments of CRISPR-Cas systems, the degree of complementarity between a guide sequence and its corresponding target sequence can be about or more than about 50%, 60%, 75%, 80%, 85%, 90%, 95%, 97.5%, 99%, or 100%; a guide or RNA or sgRNA can be about or more than about 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, 75, or more nucleotides in length; or guide or RNA or sgRNA can be less than about 75, 50, 45, 40, 35, 30, 25, 20, 15, 12, or fewer nucleotides in length; and advantageously tracr RNA is 30 or 50 nucleotides in length. However, an aspect of the invention is to reduce off-target interactions, e.g., reduce the guide interacting with a target sequence having low complementarity. Indeed, in the examples, it is shown that the invention involves mutations that result in the CRISPR-Cas system being able to distinguish between target and off-

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

[0042] In particularly preferred embodiments according to the invention, the guide RNA (capable of guiding Cas to a target locus) may comprise (1) a guide sequence capable of hybridizing to a genomic target locus in the eukaryotic cell; (2) a tracr sequence; and (3) a tracr mate sequence. All (1) to (3) may reside in a single RNA, i.e. an sgRNA (arranged in a 5' to 3' orientation), or the tracr RNA may be a different RNA than the RNA containing the guide and tracr sequence. The tracr hybridizes to the tracr mate sequence and directs the CRISPR/Cas complex to the target sequence. Where the tracr RNA is on a different RNA than the RNA containing the guide and tracr sequence, the length of each RNA may be optimized to be shortened from their respective native lengths, and each may be independently chemically modified to protect from degradation by cellular RNase or otherwise increase stability.

[0043] The methods according to the invention as described herein comprehend inducing one or more mutations in a eukaryotic cell (in vitro, i.e. in an isolated eukaryotic cell) as herein discussed comprising delivering to cell a vector as herein discussed. The mutation(s) can include the introduction, deletion, or substitution of one or more nucleotides at each target sequence of cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1-75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 1, 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29,

30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 5, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations include the introduction, deletion, or substitution of 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s). The mutations can include the introduction, deletion, or substitution of 40, 45, 50, 75, 100, 200, 300, 400 or 500 nucleotides at each target sequence of said cell(s) via the guide(s) RNA(s) or sgRNA(s).

[0044] For minimization of toxicity and off-target effect, it may be important to control the concentration of Cas mRNA and guide RNA delivered. Optimal concentrations of Cas mRNA and guide RNA can be determined by testing different concentrations in a cellular or non-human eukaryote animal model and using deep sequencing to analyze the extent of modification at potential off-target genomic loci. Alternatively, to minimize the level of toxicity and off-target effect, Cas nickase mRNA (for example *S. pyogenes* Cas9 with the D10A mutation) can be delivered with a pair of guide RNAs targeting a site of interest. Guide sequences and strategies to minimize toxicity and off-target effects can be as in WO 2014/093622 (PCT/US2013/074667); or, via mutation as herein.

[0045] 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.

Guide Modifications

[0046] In certain embodiments, guides of the invention comprise non-naturally occurring nucleic acids and/or non-naturally occurring nucleotides and/or nucleotide analogs, and/or chemically modifications. Non-naturally occurring nucleic acids can include, for example, mixtures of naturally and non-naturally occurring nucleotides. Non-naturally occurring nucleotides and/or nucleotide analogs may be modified at the ribose, phosphate, and/or base moiety. In an embodiment of the invention, a guide nucleic acid comprises ribonucleotides and non-ribonucleotides. In one such embodiment, a guide comprises one or more ribonucleotides and one or more deoxyribonucleotides. In an embodiment of the invention, the guide comprises one or more non-naturally occurring nucleotide or nucleotide analog such as a nucleotide with phosphorothioate linkage, boranophosphate linkage, a locked nucleic acid (LNA) nucleotides comprising a methylene bridge between the 2' and 4' carbons of the ribose ring, peptide nucleic acids (PNA), or bridged nucleic acids (BNA). Other examples of modified nucleotides include 2'-O-methyl analogs, 2'-deoxy analogs, 2-thiouridine analogs, N6-methyladenosine analogs, or 2'-fluoro analogs. Further examples of modified nucleotides include linkage of chemical moieties at the 2' position, including but not limited to peptides, nuclear localization sequence (NLS), peptide nucleic acid (PNA), polyethylene glycol (PEG), triethylene glycol, or tetraethyleneglycol (TEG). Further examples of modified bases include, but are not limited to, 2-aminopurine, 5-bromo-uridine, pseudouridine (E®), N1-methylpseudouridine (me1E®), 5-methoxyuridine(5moU), inosine, 7-methylguanosine. Examples of guide RNA chemical modifications include, without limitation, incorporation of 2'-O-methyl (M), 2'-O-methyl-3'-phosphorothioate (MS), phosphorothioate (PS), S-constrained ethyl(cEt), 2'-O-methyl-3'-thioPACE (MSP), or 2'-O-methyl-3'-phosphonoacetate (MP) at one or more terminal nucleotides. Such chemically modified guides can comprise increased stability and increased activity as compared to unmodified guides, though on-target vs. off-target specificity is not predictable. (See, Hendel, 2015, Nat Biotechnol. 33(9):985-9, doi: 10.1038/nbt.3290, published online 29 June 2015; Ragdarm et al., 0215, PNAS, E7110-E7111; Allerson et al., J. Med. Chem. 2005, 48:901-904; Bramsen et al., Front. Genet., 2012, 3:154; Deng et al., PNAS, 2015, 112:11870-11875; Sharma et al., MedChemComm., 2014, 5:1454-1471; Hendel et al., Nat. Biotechnol. (2015) 33(9): 985-989; Li et al., Nature Biomedical Engineering, 2017, 1, 0066

DOI:10.1038/s41551-017-0066; Ryan et al., *Nucleic Acids Res.* (2018) 46(2): 792-803). In some embodiments, the 5' and/or 3' end of a guide RNA is modified by a variety of functional moieties including fluorescent dyes, polyethylene glycol, cholesterol, proteins, or detection tags. (See Kelly et al., 2016, *J. Biotech.* 233:74-83). In certain embodiments, a guide comprises ribonucleotides in a region that binds to a target DNA and one or more deoxyribonucleotides and/or nucleotide analogs in a region that binds to Cas9, Cpf1, C2c1, or Cas13. In an embodiment of the invention, deoxyribonucleotides and/or nucleotide analogs are incorporated in engineered guide structures, such as, without limitation, 5' and/or 3' end, stem-loop regions, and the seed region. In certain embodiments, the modification is not in the 5'-handle of the stem-loop regions. Chemical modification in the 5'-handle of the stem-loop region of a guide may abolish its function (see Li, et al., *Nature Biomedical Engineering*, 2017, 1:0066). In certain embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 35, 40, 45, 50, or 75 nucleotides of a guide is chemically modified. In some embodiments, 3-5 nucleotides at either the 3' or the 5' end of a guide is chemically modified. In some embodiments, only minor modifications are introduced in the seed region, such as 2'-F modifications. In some embodiments, 2'-F modification is introduced at the 3' end of a guide. In certain embodiments, three to five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-O-methyl (M), 2'-O-methyl-3'-phosphorothioate (MS), S-constrained ethyl(cEt), 2'-O-methyl-3'-thioPACE (MSP), or 2'-O-methyl-3'-phosphonoacetate (MP). Such modification can enhance genome editing efficiency (see Hendel et al., *Nat. Biotechnol.* (2015) 33(9): 985-989; Ryan et al., *Nucleic Acids Res.* (2018) 46(2): 792-803). In certain embodiments, all of the phosphodiester bonds of a guide are substituted with phosphorothioates (PS) for enhancing levels of gene disruption. In certain embodiments, more than five nucleotides at the 5' and/or the 3' end of the guide are chemically modified with 2'-O-Me, 2'-F or S-constrained ethyl(cEt). Such chemically modified guide can mediate enhanced levels of gene disruption (see Ragdarm et al., 0215, *PNAS*, E7110-E7111). In an embodiment of the invention, a guide is modified to comprise a chemical moiety at its 3' and/or 5' end. Such moieties include, but are not limited to amine, azide, alkyne, thio, dibenzocyclooctyne (DBCO), Rhodamine, peptides, nuclear localization sequence (NLS), peptide nucleic acid (PNA), polyethylene glycol (PEG), triethylene glycol, or tetraethyleneglycol (TEG). In certain

embodiment, the chemical moiety is conjugated to the guide by a linker, such as an alkyl chain. In certain embodiments, the chemical moiety of the modified guide can be used to attach the guide to another molecule, such as DNA, RNA, protein, or nanoparticles. Such chemically modified guide can be used to identify or enrich cells generically edited by a CRISPR system (see Lee et al., *eLife*, 2017, 6:e25312, DOI:10.7554). In some embodiments, 3 nucleotides at each of the 3' and 5' ends are chemically modified. In a specific embodiment, the modifications comprise 2'-O-methyl or phosphorothioate analogs. In a specific embodiment, 12 nucleotides in the tetraloop and 16 nucleotides in the stem-loop region are replaced with 2'-O-methyl analogs. Such chemical modifications improve in vivo editing and stability (see Finn et al., *Cell Reports* (2018), 22: 2227-2235). In some embodiments, more than 60 or 70 nucleotides of the guide are chemically modified. In some embodiments, this modification comprises replacement of nucleotides with 2'-O-methyl or 2'-fluoro nucleotide analogs or phosphorothioate (PS) modification of phosphodiester bonds. In some embodiments, the chemical modification comprises 2'-O-methyl or 2'-fluoro modification of guide nucleotides extending outside of the nuclease protein when the CRISPR complex is formed or PS modification of 20 to 30 or more nucleotides of the 3'-terminus of the guide. In a particular embodiment, the chemical modification further comprises 2'-O-methyl analogs at the 5' end of the guide or 2'-fluoro analogs in the seed and tail regions. Such chemical modifications improve stability to nuclease degradation and maintain or enhance genome-editing activity or efficiency, but modification of all nucleotides may abolish the function of the guide (see Yin et al., *Nat. Biotech.* (2018), 35(12): 1179-1187). Such chemical modifications may be guided by knowledge of the structure of the CRISPR complex, including knowledge of the limited number of nuclease and RNA 2'-OH interactions (see Yin et al., *Nat. Biotech.* (2018), 35(12): 1179-1187). In some embodiments, one or more guide RNA nucleotides may be replaced with DNA nucleotides. In some embodiments, up to 2, 4, 6, 8, 10, or 12 RNA nucleotides of the 5'-end tail/seed guide region are replaced with DNA nucleotides. In certain embodiments, the majority of guide RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, 16 guide RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, 8 guide RNA nucleotides of the 5'-end tail/seed region and 16 RNA nucleotides at the 3' end are replaced with DNA nucleotides. In particular embodiments, guide RNA nucleotides that extend outside of the

nuclease protein when the CRISPR complex is formed are replaced with DNA nucleotides. Such replacement of multiple RNA nucleotides with DNA nucleotides leads to decreased off-target activity but similar on-target activity compared to an unmodified guide; however, replacement of all RNA nucleotides at the 3' end may abolish the function of the guide (see Yin et al., Nat. Chem. Biol. (2018) 14, 311-316). Such modifications may be guided by knowledge of the structure of the CRISPR complex, including knowledge of the limited number of nuclease and RNA 2'-OH interactions (see Yin et al., Nat. Chem. Biol. (2018) 14, 311-316).

[0047] In one aspect of the invention, the guide comprises a modified crRNA for Cpf1, having a 5'-handle and a guide segment further comprising a seed region and a 3'-terminus. In some embodiments, the modified guide can be used with a Cpf1 of any one of *Acidaminococcus* sp. BV3L6 Cpf1 (AsCpf1); *Francisella tularensis* subsp. *Novicida* U112 Cpf1 (FnCpf1); *L. bacterium* MC2017 Cpf1 (Lb3Cpf1); *Butyrivibrio proteoclasticus* Cpf1 (BpCpf1); *Parcubacteria bacterium* GWC2011_GWC2_44_17 Cpf1 (PbCpf1); *Peregrinibacteria bacterium* GW2011_GWA_33_10 Cpf1 (PeCpf1); *Leptospira inadai* Cpf1 (LiCpf1); *Smithella* sp. SC_K08D17 Cpf1 (SsCpf1); *L. bacterium* MA2020 Cpf1 (Lb2Cpf1); *Porphyromonas crevioricanis* Cpf1 (PcCpf1); *Porphyromonas macacae* Cpf1 (PmCpf1); *Candidatus Methanoplasma termitum* Cpf1 (CMtCpf1); *Eubacterium eligens* Cpf1 (EeCpf1); *Moraxella bovoculi* 237 Cpf1 (MbCpf1); *Prevotella disiens* Cpf1 (PdCpf1); or *L. bacterium* ND2006 Cpf1 (LbCpf1).

[0048] In some embodiments, the modification to the guide is a chemical modification, an insertion, a deletion or a split. In some embodiments, the chemical modification includes, but is not limited to, incorporation of 2'-O-methyl (M) analogs, 2'-deoxy analogs, 2-thiouridine analogs, N6-methyladenosine analogs, 2'-fluoro analogs, 2-aminopurine, 5-bromo-uridine, pseudouridine (E®), N1-methylpseudouridine (me1E®), 5-methoxyuridine(5moU), inosine, 7-methylguanosine, 2'-O-methyl-3'-phosphorothioate (MS), S-constrained ethyl(cEt), phosphorothioate (PS), 2'-O-methyl-3'-thioPACE (MSP), or 2'-O-methyl-3'-phosphonoacetate (MP). In some embodiments, the guide comprises one or more of phosphorothioate modifications. In certain embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 25 nucleotides of the guide are chemically modified. In some embodiments, all nucleotides are chemically modified. In certain embodiments, one or more nucleotides in the

seed region are chemically modified. In certain embodiments, one or more nucleotides in the 3'-terminus are chemically modified. In certain embodiments, none of the nucleotides in the 5'-handle is chemically modified. In some embodiments, the chemical modification in the seed region is a minor modification, such as incorporation of a 2'-fluoro analog. In a specific embodiment, one nucleotide of the seed region is replaced with a 2'-fluoro analog. In some embodiments, 5 or 10 nucleotides in the 3'-terminus are chemically modified. Such chemical modifications at the 3'-terminus of the Cpf1 CrRNA improve gene cutting efficiency (see Li, et al., Nature Biomedical Engineering, 2017, 1:0066). In a specific embodiment, 5 nucleotides in the 3'-terminus are replaced with 2'-fluoro analogues. In a specific embodiment, 10 nucleotides in the 3'-terminus are replaced with 2'-fluoro analogues. In a specific embodiment, 5 nucleotides in the 3'-terminus are replaced with 2'-O-methyl (M) analogs. In some embodiments, 3 nucleotides at each of the 3' and 5' ends are chemically modified. In a specific embodiment, the modifications comprise 2'-O-methyl or phosphorothioate analogs. In a specific embodiment, 12 nucleotides in the tetraloop and 16 nucleotides in the stem-loop region are replaced with 2'-O-methyl analogs. Such chemical modifications improve in vivo editing and stability (see Finn et al., Cell Reports (2018), 22: 2227-2235).

[0049] In some embodiments, the loop of the 5'-handle of the guide is modified. In some embodiments, the loop of the 5'-handle of the guide is modified to have a deletion, an insertion, a split, or chemical modifications. In certain embodiments, the loop comprises 3, 4, or 5 nucleotides. In certain embodiments, the loop comprises the sequence of UCUU, UUUU, UAUU, or UGUU. In some embodiments, the guide molecule forms a stemloop with a separate non-covalently linked sequence, which can be DNA or RNA.

Synthetically linked guide

[0050] In one aspect, the guide comprises a tracr sequence and a tracr mate sequence that are chemically linked or conjugated via a non-phosphodiester bond. In one aspect, the guide comprises a tracr sequence and a tracr mate sequence that are chemically linked or conjugated via a non-nucleotide loop. In some embodiments, the tracr and tracr mate sequences are joined via a non-phosphodiester covalent linker. Examples of the covalent linker include but are not limited to a chemical moiety selected from the group consisting of carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters,

phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, fulfones, sulfoxides, ureas, thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[0051] In some embodiments, the tracr and tracr mate sequences are first synthesized using the standard phosphoramidite synthetic protocol (Herdewijn, P., ed., *Methods in Molecular Biology* Col 288, *Oligonucleotide Synthesis: Methods and Applications*, Humana Press, New Jersey (2012)). In some embodiments, the tracr or tracr mate sequences can be functionalized to contain an appropriate functional group for ligation using the standard protocol known in the art (Hermanson, G. T., *Bioconjugate Techniques*, Academic Press (2013)). Examples of functional groups include, but are not limited to, hydroxyl, amine, carboxylic acid, carboxylic acid halide, carboxylic acid active ester, aldehyde, carbonyl, chlorocarbonyl, imidazolylcarbonyl, hydrozide, semicarbazide, thio semicarbazide, thiol, maleimide, haloalkyl, sulfonyl, ally, propargyl, diene, alkyne, and azide. Once the tracr and the tracr mate sequences are functionalized, a covalent chemical bond or linkage can be formed between the two oligonucleotides. Examples of chemical bonds include, but are not limited to, those based on carbamates, ethers, esters, amides, imines, amidines, aminotrizines, hydrozone, disulfides, thioethers, thioesters, phosphorothioates, phosphorodithioates, sulfonamides, sulfonates, fulfones, sulfoxides, ureas, thioureas, hydrazide, oxime, triazole, photolabile linkages, C-C bond forming groups such as Diels-Alder cyclo-addition pairs or ring-closing metathesis pairs, and Michael reaction pairs.

[0052] In some embodiments, the tracr and tracr mate sequences can be chemically synthesized. In some embodiments, the chemical synthesis uses automated, solid-phase oligonucleotide synthesis machines with 2'-acetoxyethyl orthoester (2'-ACE) (Scaringe et al., *J. Am. Chem. Soc.* (1998) 120: 11820-11821; Scaringe, *Methods Enzymol.* (2000) 317: 3-18) or 2'-thionocarbamate (2'-TC) chemistry (Dellinger et al., *J. Am. Chem. Soc.* (2011) 133: 11540-11546; Hendel et al., *Nat. Biotechnol.* (2015) 33:985-989).

[0053] In some embodiments, the tracr and tracr mate sequences can be covalently linked using various bioconjugation reactions, loops, bridges, and non-nucleotide links via modifications of sugar, internucleotide phosphodiester bonds, purine and pyrimidine residues. Sletten et al., *Angew. Chem. Int. Ed.* (2009) 48:6974-6998; Manoharan, M. *Curr. Opin. Chem.*

Biol. (2004) 8: 570-9; Behlke et al., Oligonucleotides (2008) 18: 305-19; Watts, et al., Drug Discov. Today (2008) 13: 842-55; Shukla, et al., ChemMedChem (2010) 5: 328-49.

[0054] In some embodiments, the tracer and tracer mate sequences can be covalently linked using click chemistry. In some embodiments, the tracer and tracer mate sequences can be covalently linked using a triazole linker. In some embodiments, the tracer and tracer mate sequences can be covalently linked using Huisgen 1,3-dipolar cycloaddition reaction involving an alkyne and azide to yield a highly stable triazole linker (He et al., ChemBioChem (2015) 17: 1809-1812; WO 2016/186745). In some embodiments, the tracer and tracer mate sequences are covalently linked by ligating a 5'-hexyne tracerRNA and a 3'-azide crRNA. In some embodiments, either or both of the 5'-hexyne tracerRNA and a 3'-azide crRNA can be protected with 2'-acetoxyethyl orthoester (2'-ACE) group, which can be subsequently removed using Dharmacon protocol (Scaringe et al., J. Am. Chem. Soc. (1998) 120: 11820-11821; Scaringe, Methods Enzymol. (2000) 317: 3-18).

[0055] In some embodiments, the tracer and tracer mate sequences can be covalently linked via a linker (e.g., a non-nucleotide loop) that comprises a moiety such as spacers, attachments, bioconjugates, chromophores, reporter groups, dye labeled RNAs, and non-naturally occurring nucleotide analogues. More specifically, suitable spacers for purposes of this invention include, but are not limited to, polyethers (e.g., polyethylene glycols, polyalcohols, polypropylene glycol or mixtures of ethylene and propylene glycols), polyamines group (e.g., spennine, spermidine and polymeric derivatives thereof), polyesters (e.g., poly(ethyl acrylate)), polyphosphodiester, alkylenes, and combinations thereof. Suitable attachments include any moiety that can be added to the linker to add additional properties to the linker, such as but not limited to, fluorescent labels. Suitable bioconjugates include, but are not limited to, peptides, glycosides, lipids, cholesterol, phospholipids, diacyl glycerols and dialkyl glycerols, fatty acids, hydrocarbons, enzyme substrates, steroids, biotin, digoxigenin, carbohydrates, polysaccharides. Suitable chromophores, reporter groups, and dye-labeled RNAs include, but are not limited to, fluorescent dyes such as fluorescein and rhodamine, chemiluminescent, electrochemiluminescent, and bioluminescent marker compounds. The design of example linkers conjugating two RNA components are also described in WO 2004/015075.

[0056] The linker (e.g., a non-nucleotide loop) can be of any length. In some embodiments, the linker has a length equivalent to about 0-16 nucleotides. In some embodiments, the linker has a length equivalent to about 0-8 nucleotides. In some embodiments, the linker has a length equivalent to about 0-4 nucleotides. In some embodiments, the linker has a length equivalent to about 2 nucleotides. Example linker design is also described in WO2011/008730.

[0057] A typical Type II Cas sgRNA comprises (in 5' to 3' direction): a guide sequence, a poly U tract, a first complimentary stretch (the "repeat"), a loop (tetraloop), a second complimentary stretch (the "anti-repeat" being complimentary to the repeat), a stem, and further stem loops and stems and a poly A (often poly U in RNA) tail (terminator). In preferred embodiments, certain aspects of guide architecture are retained, certain aspect of guide architecture can be modified, for example by addition, subtraction, or substitution of features, whereas certain other aspects of guide architecture are maintained. Preferred locations for engineered sgRNA modifications, including but not limited to insertions, deletions, and substitutions include guide termini and regions of the sgRNA that are exposed when complexed with CRISPR protein and/or target, for example the tetraloop and/or loop2.

[0058] In certain embodiments, guides of the invention comprise specific binding sites (e.g. aptamers) for adapter proteins, which may comprise one or more functional domains (e.g. via fusion protein). When such a guide forms a CRISPR complex (i.e. CRISPR enzyme binding to guide and target) the adapter proteins bind and, the functional domain associated with the adapter protein is positioned in a spatial orientation which is advantageous for the attributed function to be effective. For example, if the functional domain is a transcription activator (e.g. VP64 or p65), the transcription activator is placed in a spatial orientation which allows it to affect the transcription of the target. Likewise, a transcription repressor will be advantageously positioned to affect the transcription of the target and a nuclease (e.g. FokI) will be advantageously positioned to cleave or partially cleave the target.

[0059] The skilled person will understand that modifications to the guide which allow for binding of the adapter + functional domain but not proper positioning of the adapter + functional domain (e.g. due to steric hindrance within the three-dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified guide may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein,

preferably at either the tetra loop or stem loop 2, and most preferably at both the tetra loop and stem loop 2.

[0060] The repeat:anti repeat duplex will be apparent from the secondary structure of the sgRNA. It may be typically a first complimentary stretch after (in 5' to 3' direction) the poly U tract and before the tetraloop; and a second complimentary stretch after (in 5' to 3' direction) the tetraloop and before the poly A tract. The first complimentary stretch (the "repeat") is complimentary to the second complimentary stretch (the "anti-repeat"). As such, they Watson-Crick base pair to form a duplex of dsRNA when folded back on one another. As such, the anti-repeat sequence is the complimentary sequence of the repeat and in terms to A-U or C-G base pairing, but also in terms of the fact that the anti-repeat is in the reverse orientation due to the tetraloop.

[0061] In an embodiment of the invention, modification of guide architecture comprises replacing bases in stemloop 2. For example, in some embodiments, "actt" ("acuu" in RNA) and "aagt" ("aagu" in RNA) bases in stemloop2 are replaced with "cgcc" and "gcgg". In some embodiments, "actt" and "aagt" bases in stemloop2 are replaced with complimentary GC-rich regions of 4 nucleotides. In some embodiments, the complimentary GC-rich regions of 4 nucleotides are "cgcc" and "gcgg" (both in 5' to 3' direction). In some embodiments, the complimentary GC-rich regions of 4 nucleotides are "gcgg" and "cgcc" (both in 5' to 3' direction). Other combination of C and G in the complimentary GC-rich regions of 4 nucleotides will be apparent including CCCC and GGGG.

[0062] In one aspect, the stemloop 2, e.g., "ACTTgtttAAGT" can be replaced by any "XXXXgtttYYYY", e.g., where XXXX and YYYY represent any complementary sets of nucleotides that together will base pair to each other to create a stem.

[0063] In one aspect, the stem comprises at least about 4bp comprising complementary X and Y sequences, although stems of more, e.g., 5, 6, 7, 8, 9, 10, 11 or 12 or fewer, e.g., 3, 2, base pairs are also contemplated. Thus, for example X2-12 and Y2-12 (wherein X and Y represent any complementary set of nucleotides) may be contemplated. In one aspect, the stem made of the X and Y nucleotides, together with the "gttt," will form a complete hairpin in the overall secondary structure; and, this may be advantageous and the amount of base pairs can be any amount that forms a complete hairpin. In one aspect, any complementary X:Y basepairing

sequence (e.g., as to length) is tolerated, so long as the secondary structure of the entire sgRNA is preserved. In one aspect, the stem can be a form of X:Y basepairing that does not disrupt the secondary structure of the whole sgRNA in that it has a DR:tracr duplex, and 3 stemloops. In one aspect, the "gttt" tetraloop that connects ACTT and AAGT (or any alternative stem made of X:Y basepairs) can be any sequence of the same length (e.g., 4 basepair) or longer that does not interrupt the overall secondary structure of the sgRNA. In one aspect, the stemloop can be something that further lengthens stemloop2, e.g. can be MS2 aptamer. In one aspect, the stemloop3 "GGCACCGagtCGGTGC" can likewise take on a "XXXXXXXXagtYYYYYYY" form, e.g., wherein X7 and Y7 represent any complementary sets of nucleotides that together will base pair to each other to create a stem. In one aspect, the stem comprises about 7bp comprising complementary X and Y sequences, although stems of more or fewer basepairs are also contemplated. In one aspect, the stem made of the X and Y nucleotides, together with the "agt", will form a complete hairpin in the overall secondary structure. In one aspect, any complementary X:Y basepairing sequence is tolerated, so long as the secondary structure of the entire sgRNA is preserved. In one aspect, the stem can be a form of X:Y basepairing that doesn't disrupt the secondary structure of the whole sgRNA in that it has a DR:tracr duplex, and 3 stemloops. In one aspect, the "agt" sequence of the stemloop 3 can be extended or be replaced by an aptamer, e.g., a MS2 aptamer or sequence that otherwise generally preserves the architecture of stemloop3. In one aspect for alternative Stemloops 2 and/or 3, each X and Y pair can refer to any basepair. In one aspect, non-Watson Crick basepairing is contemplated, where such pairing otherwise generally preserves the architecture of the stemloop at that position.

[0064] In one aspect, the DR:tracrRNA duplex can be replaced with the form: gYYYYag(N)NNNNxxxxNNNN(AAN)uuRRRRu (using standard IUPAC nomenclature for nucleotides), wherein (N) and (AAN) represent part of the bulge in the duplex, and "xxxx" represents a linker sequence. NNNN on the direct repeat can be anything so long as it basepairs with the corresponding NNNN portion of the tracrRNA. In one aspect, the DR:tracrRNA duplex can be connected by a linker of any length (xxxx...), any base composition, as long as it doesn't alter the overall structure.

[0065] In one aspect, the sgRNA structural requirement is to have a duplex and 3 stemloops. In most aspects, the actual sequence requirement for many of the particular base requirements

are lax, in that the architecture of the DR:tracrRNA duplex should be preserved, but the sequence that creates the architecture, i.e., the stems, loops, bulges, etc., may be altered.

Aptamers

[0066] One guide with a first aptamer/RNA-binding protein pair can be linked or fused to an activator, whilst a second guide with a second aptamer/RNA-binding protein pair can be linked or fused to a repressor. The guides are for different targets (loci), so this allows one gene to be activated and one repressed. For example, the following schematic shows such an approach:

[0067] Guide 1- MS2 aptamer-----MS2 RNA-binding protein-----VP64 activator; and

[0068] Guide 2 - PP7 aptamer-----PP7 RNA-binding protein-----SID4x repressor.

[0069] The present invention also relates to orthogonal PP7/MS2 gene targeting. In this example, sgRNA targeting different loci are modified with distinct RNA loops in order to recruit MS2-VP64 or PP7-SID4X, which activate and repress their target loci, respectively. PP7 is the RNA-binding coat protein of the bacteriophage Pseudomonas. Like MS2, it binds a specific RNA sequence and secondary structure. The PP7 RNA-recognition motif is distinct from that of MS2. Consequently, PP7 and MS2 can be multiplexed to mediate distinct effects at different genomic loci simultaneously. For example, an sgRNA targeting locus A can be modified with MS2 loops, recruiting MS2-VP64 activators, while another sgRNA targeting locus B can be modified with PP7 loops, recruiting PP7-SID4X repressor domains. In the same cell, dCas13 can thus mediate orthogonal, locus-specific modifications. This principle can be extended to incorporate other orthogonal RNA-binding proteins such as Q-beta.

[0070] An alternative option for orthogonal repression includes incorporating non-coding RNA loops with transactive repressive function into the guide (either at similar positions to the MS2/PP7 loops integrated into the guide or at the 3' terminus of the guide). For instance, guides were designed with non-coding (but known to be repressive) RNA loops (e.g. using the Alu repressor (in RNA) that interferes with RNA polymerase II in mammalian cells). The Alu RNA sequence was located: in place of the MS2 RNA sequences as used herein (e.g. at tetraloop and/or stem loop 2); and/or at 3' terminus of the guide. This gives possible combinations of MS2, PP7 or Alu at the tetraloop and/or stemloop 2 positions, as well as, optionally, addition of Alu at the 3' end of the guide (with or without a linker).

[0071] The use of two different aptamers (distinct RNA) allows an activator-adaptor protein fusion and a repressor-adaptor protein fusion to be used, with different guides, to activate expression of one gene, whilst repressing another. They, along with their different guides can be administered together, or substantially together, in a multiplexed approach. A large number of such modified guides can be used all at the same time, for example 10 or 20 or 30 and so forth, whilst only one (or at least a minimal number) of Cas13s to be delivered, as a comparatively small number of Cas13s can be used with a large number modified guides. The adaptor protein may be associated (preferably linked or fused to) one or more activators or one or more repressors. For example, the adaptor protein may be associated with a first activator and a second activator. The first and second activators may be the same, but they are preferably different activators. For example, one might be VP64, whilst the other might be p65, although these are just examples and other transcriptional activators are envisaged. Three or more or even four or more activators (or repressors) may be used, but package size may limit the number being higher than 5 different functional domains. Linkers are preferably used, over a direct fusion to the adaptor protein, where two or more functional domains are associated with the adaptor protein. Suitable linkers might include the GlySer linker.

[0072] It is also envisaged that the enzyme-guide complex as a whole may be associated with two or more functional domains. For example, there may be two or more functional domains associated with the enzyme, or there may be two or more functional domains associated with the guide (via one or more adaptor proteins), or there may be one or more functional domains associated with the enzyme and one or more functional domains associated with the guide (via one or more adaptor proteins).

[0073] The fusion between the adaptor protein and the activator or repressor may include a linker. For example, GlySer linkers GGGS can be used. They can be used in repeats of 3 ((GGGS)3) or 6, 9 or even 12 or more, to provide suitable lengths, as required. Linkers can be used between the RNA-binding protein and the functional domain (activator or repressor), or between the CRISPR Enzyme (Cas13) and the functional domain (activator or repressor). The linkers the user to engineer appropriate amounts of "mechanical flexibility".

[0074] *Dead guides: Guide RNAs comprising a dead guide sequence may be used in the present invention*

[0075] In one aspect, the invention provides guide sequences which are modified in a manner which allows for formation of the CRISPR complex and successful binding to the target, while at the same time, not allowing for successful nuclease activity (i.e. without nuclease activity / without indel activity). For matters of explanation such modified guide sequences are referred to as "dead guides" or "dead guide sequences". These dead guides or dead guide sequences can be thought of as catalytically inactive or conformationally inactive with regard to nuclease activity. Nuclease activity may be measured using surveyor analysis or deep sequencing as commonly used in the art, preferably surveyor analysis. Similarly, dead guide sequences may not sufficiently engage in productive base pairing with respect to the ability to promote catalytic activity or to distinguish on-target and off-target binding activity. Briefly, the surveyor assay involves purifying and amplifying a CRISPR target site for a gene and forming heteroduplexes with primers amplifying the CRISPR target site. After re-anneal, the products are treated with SURVEYOR nuclease and SURVEYOR enhancer S (Transgenomics) following the manufacturer's recommended protocols, analyzed on gels, and quantified based upon relative band intensities.

[0076] Hence, in a related aspect, the invention provides a non-naturally occurring or engineered composition Cas13 CRISPR-Cas system comprising a functional Cas13 as described herein, and guide RNA (gRNA) wherein the gRNA comprises a dead guide sequence whereby the gRNA is capable of hybridizing to a target sequence such that the Cas13 CRISPR-Cas system is directed to a genomic locus of interest in a cell without detectable indel activity resultant from nuclease activity of a non-mutant Cas13 enzyme of the system as detected by a SURVEYOR assay. For shorthand purposes, a gRNA comprising a dead guide sequence whereby the gRNA is capable of hybridizing to a target sequence such that the Cas13 CRISPR-Cas system is directed to a genomic locus of interest in a cell without detectable indel activity resultant from nuclease activity of a non-mutant Cas13 enzyme of the system as detected by a SURVEYOR assay is herein termed a "dead gRNA". It is to be understood that any of the gRNAs according to the invention as described herein elsewhere may be used as dead gRNAs / gRNAs comprising a dead guide sequence as described herein below. Any of the methods, products, compositions and uses as described herein elsewhere is equally applicable with the

dead gRNAs / gRNAs comprising a dead guide sequence as further detailed below. By means of further guidance, the following particular aspects and embodiments are provided.

[0077] The ability of a dead 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 dead guide sequence to be tested, may be provided to a host cell having the corresponding target sequence, such as by transfection with vectors encoding the components of the 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 dead guide sequence to be tested and a control guide sequence different from the test dead guide sequence, and comparing binding or rate of cleavage at the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A dead 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.

[0078] As explained further herein, several structural parameters allow for a proper framework to arrive at such dead guides. Dead guide sequences are shorter than respective guide sequences which result in active Cas13-specific indel formation. Dead guides are 5%, 10%, 20%, 30%, 40%, 50%, shorter than respective guides directed to the same Cas13 leading to active Cas13-specific indel formation.

[0079] As explained below and known in the art, one aspect of gRNA - Cas specificity is the direct repeat sequence, which is to be appropriately linked to such guides. In particular, this implies that the direct repeat sequences are designed dependent on the origin of the Cas. Thus, structural data available for validated dead guide sequences may be used for designing Cas specific equivalents. Structural similarity between, e.g., the orthologous nuclease domains RuvC of two or more Cas effector proteins may be used to transfer design equivalent dead guides. Thus, the dead guide herein may be appropriately modified in length and sequence to reflect such Cas specific equivalents, allowing for formation of the CRISPR complex and successful binding to the target, while at the same time, not allowing for successful nuclease activity.

[0080] The use of dead guides in the context herein as well as the state of the art provides a surprising and unexpected platform for network biology and/or systems biology in both in vitro, ex vivo, and in vivo applications, allowing for multiplex gene targeting, and in particular bidirectional multiplex gene targeting. Prior to the use of dead guides, addressing multiple targets, for example for activation, repression and/or silencing of gene activity, has been challenging and in some cases not possible. With the use of dead guides, multiple targets, and thus multiple activities, may be addressed, for example, in the same cell, in the same animal, or in the same patient. Such multiplexing may occur at the same time or staggered for a desired timeframe.

[0081] For example, the dead guides now allow for the first time to use gRNA as a means for gene targeting, without the consequence of nuclease activity, while at the same time providing directed means for activation or repression. Guide RNA comprising a dead guide may be modified to further include elements in a manner which allow for activation or repression of gene activity, in particular protein adaptors (e.g. aptamers) as described herein elsewhere allowing for functional placement of gene effectors (e.g. activators or repressors of gene activity). One example is the incorporation of aptamers, as explained herein and in the state of the art. By engineering the gRNA comprising a dead guide to incorporate protein-interacting aptamers (Konermann et al., "Genome-scale transcription activation by an engineered CRISPR-Cas9 complex," doi:10.1038/nature14136, incorporated herein by reference), one may assemble a synthetic transcription activation complex consisting of multiple distinct effector domains. Such may be modeled after natural transcription activation processes. For example, an aptamer, which selectively binds an effector (e.g. an activator or repressor; dimerized MS2 bacteriophage coat proteins as fusion proteins with an activator or repressor), or a protein which itself binds an effector (e.g. activator or repressor) may be appended to a dead gRNA tetraloop and/or a stem-loop 2. In the case of MS2, the fusion protein MS2-VP64 binds to the tetraloop and/or stem-loop 2 and in turn mediates transcriptional up-regulation, for example for Neurog2. Other transcriptional activators are, for example, VP64, P65, HSF1, and MyoD1. By mere example of this concept, replacement of the MS2 stem-loops with PP7-interacting stem-loops may be used to recruit repressive elements.

[0082] Thus, one aspect is a gRNA of the invention which comprises a dead guide, wherein the gRNA further comprises modifications which provide for gene activation or repression, as described herein. The dead gRNA may comprise one or more aptamers. The aptamers may be specific to gene effectors, gene activators or gene repressors. Alternatively, the aptamers may be specific to a protein which in turn is specific to and recruits / binds a specific gene effector, gene activator or gene repressor. If there are multiple sites for activator or repressor recruitment, it is preferred that the sites are specific to either activators or repressors. If there are multiple sites for activator or repressor binding, the sites may be specific to the same activators or same repressors. The sites may also be specific to different activators or different repressors. The gene effectors, gene activators, gene repressors may be present in the form of fusion proteins.

[0083] In an embodiment, the dead gRNA as described herein or the Cas13 CRISPR-Cas complex as described herein includes a non-naturally occurring or engineered composition comprising two or more adaptor proteins, wherein each protein is associated with one or more functional domains and wherein the adaptor protein binds to the distinct RNA sequence(s) inserted into the at least one loop of the dead gRNA.

[0084] Hence, an aspect provides a non-naturally occurring or engineered composition comprising a guide RNA (gRNA) comprising a dead guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell, wherein the dead guide sequence is as defined herein, a Cas13 comprising at least one or more nuclear localization sequences, wherein the Cas13 optionally comprises at least one mutation wherein at least one loop of the dead gRNA is modified by the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins, and wherein the adaptor protein is associated with one or more functional domains; or, wherein the dead gRNA is modified to have at least one non-coding functional loop, and wherein the composition comprises two or more adaptor proteins, wherein the each protein is associated with one or more functional domains.

[0085] In certain embodiments, the adaptor protein is a fusion protein comprising the functional domain, the fusion protein optionally comprising a linker between the adaptor protein and the functional domain, the linker optionally including a GlySer linker.

[0086] In certain embodiments, the at least one loop of the dead gRNA is not modified by the insertion of distinct RNA sequence(s) that bind to the two or more adaptor proteins.

[0087] In certain embodiments, the one or more functional domains associated with the adaptor protein is a transcriptional activation domain.

[0088] In certain embodiments, the one or more functional domains associated with the adaptor protein is a transcriptional activation domain comprising VP64, p65, MyoD1, HSF1, RTA or SET7/9.

[0089] In certain embodiments, the one or more functional domains associated with the adaptor protein is a transcriptional repressor domain.

[0090] In certain embodiments, the transcriptional repressor domain is a KRAB domain.

[0091] In certain embodiments, the transcriptional repressor domain is a NuE domain, NcoR domain, SID domain or a SID4X domain.

[0092] In certain embodiments, at least one of the one or more functional domains associated with the adaptor protein have one or more activities comprising methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, DNA integration activity RNA cleavage activity, DNA cleavage activity or nucleic acid binding activity.

[0093] In certain embodiments, the DNA cleavage activity is due to a Fok1 nuclease.

[0094] In certain embodiments, the dead gRNA is modified so that, after dead gRNA binds the adaptor protein and further binds to the Cas13 and target, the functional domain is in a spatial orientation allowing for the functional domain to function in its attributed function.

[0095] In certain embodiments, the at least one loop of the dead gRNA is tetra loop and/or loop2. In certain embodiments, the tetra loop and loop 2 of the dead gRNA are modified by the insertion of the distinct RNA sequence(s).

[0096] In certain embodiments, the insertion of distinct RNA sequence(s) that bind to one or more adaptor proteins is an aptamer sequence. In certain embodiments, the aptamer sequence is two or more aptamer sequences specific to the same adaptor protein. In certain embodiments, the aptamer sequence is two or more aptamer sequences specific to different adaptor protein.

[0097] In certain embodiments, the adaptor protein comprises MS2, PP7, Q β , F2, GA, fr, JP501, M12, R17, BZ13, JP34, JP500, KU1, M11, MX1, TW18, VK, SP, FI, ID2, NL95, TW19, AP205, ϕ Cb5, ϕ Cb8r, ϕ Cb12r, ϕ Cb23r, 7s, PRR1.

[0098] In certain embodiments, the cell is a eukaryotic cell. In certain embodiments, the eukaryotic cell is a mammalian cell, optionally a mouse cell. In certain embodiments, the mammalian cell is a human cell.

[0099] In certain embodiments, a first adaptor protein is associated with a p65 domain and a second adaptor protein is associated with a HSF1 domain.

[00100] In certain embodiments, the composition comprises a Cas13 CRISPR-Cas complex having at least three functional domains, at least one of which is associated with the Cas13 and at least two of which are associated with dead gRNA.

[00101] In certain embodiments, the composition further comprises a second gRNA, wherein the second gRNA is a live gRNA capable of hybridizing to a second target sequence such that a second Cas13 CRISPR-Cas system is directed to a second genomic locus of interest in a cell with detectable indel activity at the second genomic locus resultant from nuclease activity of the Cas13 enzyme of the system.

[00102] In certain embodiments, the composition further comprises a plurality of dead gRNAs and/or a plurality of live gRNAs.

[00103] One aspect of the invention is to take advantage of the modularity and customizability of the gRNA scaffold to establish a series of gRNA scaffolds with different binding sites (in particular aptamers) for recruiting distinct types of effectors in an orthogonal manner. Again, for matters of example and illustration of the broader concept, replacement of the MS2 stem-loops with PP7-interacting stem-loops may be used to bind / recruit repressive elements, enabling multiplexed bidirectional transcriptional control. Thus, in general, gRNA comprising a dead guide may be employed to provide for multiplex transcriptional control and preferred bidirectional transcriptional control. This transcriptional control is most preferred of genes. For example, one or more gRNA comprising dead guide(s) may be employed in targeting the activation of one or more target genes. At the same time, one or more gRNA comprising dead guide(s) may be employed in targeting the repression of one or more target genes. Such a sequence may be applied in a variety of different combinations, for example the target genes are first repressed and then at an appropriate period other targets are activated, or select genes are repressed at the same time as select genes are activated, followed by further activation and/or

repression. As a result, multiple components of one or more biological systems may advantageously be addressed together.

[00104] In an aspect, the invention provides nucleic acid molecule(s) encoding dead gRNA or the Cas13 CRISPR-Cas complex or the composition as described herein.

[00105] In an aspect, the invention provides a vector system comprising: a nucleic acid molecule encoding dead guide RNA as defined herein. In certain embodiments, the vector system further comprises a nucleic acid molecule(s) encoding Cas13. In certain embodiments, the vector system further comprises a nucleic acid molecule(s) encoding (live) gRNA. In certain embodiments, the nucleic acid molecule or the vector further comprises regulatory element(s) operable in a eukaryotic cell operably linked to the nucleic acid molecule encoding the guide sequence (gRNA) and/or the nucleic acid molecule encoding Cas13 and/or the optional nuclear localization sequence(s).

[00106] In another aspect, structural analysis may also be used to study interactions between the dead guide and the active Cas nuclease that enable DNA binding, but no DNA cutting. In this way amino acids important for nuclease activity of Cas are determined. Modification of such amino acids allows for improved Cas enzymes used for gene editing.

[00107] A further aspect is combining the use of dead guides as explained herein with other applications of CRISPR, as explained herein as well as known in the art. For example, gRNA comprising dead guide(s) for targeted multiplex gene activation or repression or targeted multiplex bidirectional gene activation / repression may be combined with gRNA comprising guides which maintain nuclease activity, as explained herein. Such gRNA comprising guides which maintain nuclease activity may or may not further include modifications which allow for repression of gene activity (e.g. aptamers). Such gRNA comprising guides which maintain nuclease activity may or may not further include modifications which allow for activation of gene activity (e.g. aptamers). In such a manner, a further means for multiplex gene control is introduced (e.g. multiplex gene targeted activation without nuclease activity / without indel activity may be provided at the same time or in combination with gene targeted repression with nuclease activity).

[00108] For example, 1) using one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) comprising dead guide(s) targeted to one or more genes and further

modified with appropriate aptamers for the recruitment of gene activators; 2) may be combined with one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) comprising dead guide(s) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene repressors. 1) and/or 2) may then be combined with 3) one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) targeted to one or more genes. This combination can then be carried out in turn with 1) + 2) + 3) with 4) one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene activators. This combination can then be carried out in turn with 1) + 2) + 3) + 4) with 5) one or more gRNA (e.g. 1-50, 1-40, 1-30, 1-20, preferably 1-10, more preferably 1-5) targeted to one or more genes and further modified with appropriate aptamers for the recruitment of gene repressors. As a result various uses and combinations are included in the invention. For example, combination 1) + 2); combination 1) + 3); combination 2) + 3); combination 1) + 2) + 3); combination 1) + 2) + 3) + 4); combination 1) + 3) + 4); combination 2) + 3) + 4); combination 1) + 2) + 4); combination 1) + 2) + 3) + 4) + 5); combination 1) + 3) + 4) + 5); combination 2) + 3) + 4) + 5); combination 1) + 2) + 4) + 5); combination 1) + 2) + 3) + 5); combination 1) + 3) + 5); combination 2) + 3) + 5); combination 1) + 2) + 5).

[00109] In an aspect, the invention provides an algorithm for designing, evaluating, or selecting a dead guide RNA targeting sequence (dead guide sequence) for guiding a Cas13 CRISPR-Cas system to a target gene locus. In particular, it has been determined that dead guide RNA specificity relates to and can be optimized by varying i) GC content and ii) targeting sequence length. In an aspect, the invention provides an algorithm for designing or evaluating a dead guide RNA targeting sequence that minimizes off-target binding or interaction of the dead guide RNA. In an embodiment of the invention, the algorithm for selecting a dead guide RNA targeting sequence for directing a CRISPR system to a gene locus in an organism comprises a) locating one or more CRISPR motifs in the gene locus, analyzing the 20 nt sequence downstream of each CRISPR motif by i) determining the GC content of the sequence; and ii) determining whether there are off-target matches of the 15 downstream nucleotides nearest to the CRISPR motif in the genome of the organism, and c) selecting the 15 nucleotide sequence for use in a dead guide RNA if the GC content of the sequence is 70% or less and no off-target matches are

identified. In an embodiment, the sequence is selected for a targeting sequence if the GC content is 60% or less. In certain embodiments, the sequence is selected for a targeting sequence if the GC content is 55% or less, 50% or less, 45% or less, 40% or less, 35% or less or 30% or less. In an embodiment, two or more sequences of the gene locus are analyzed and the sequence having the lowest GC content, or the next lowest GC content, or the next lowest GC content is selected. In an embodiment, the sequence is selected for a targeting sequence if no off-target matches are identified in the genome of the organism. In an embodiment, the targeting sequence is selected if no off-target matches are identified in regulatory sequences of the genome.

[00110] In an aspect, the invention provides a method of selecting a dead guide RNA targeting sequence for directing a functionalized CRISPR system to a gene locus in an organism, which comprises: a) locating one or more CRISPR motifs in the gene locus; b) analyzing the 20 nt sequence downstream of each CRISPR motif by: i) determining the GC content of the sequence; and ii) determining whether there are off-target matches of the first 15 nt of the sequence in the genome of the organism; c) selecting the sequence for use in a guide RNA if the GC content of the sequence is 70% or less and no off-target matches are identified. In an embodiment, the sequence is selected if the GC content is 50% or less. In an embodiment, the sequence is selected if the GC content is 40% or less. In an embodiment, the sequence is selected if the GC content is 30% or less. In an embodiment, two or more sequences are analyzed and the sequence having the lowest GC content is selected. In an embodiment, off-target matches are determined in regulatory sequences of the organism. In an embodiment, the gene locus is a regulatory region. An aspect provides a dead guide RNA comprising the targeting sequence selected according to the aforementioned methods.

[00111] In an aspect, the invention provides a dead guide RNA for targeting a functionalized CRISPR system to a gene locus in an organism. In an embodiment of the invention, the dead guide RNA comprises a targeting sequence wherein the GC content of the target sequence is 70% or less, and the first 15 nt of the targeting sequence does not match an off-target sequence downstream from a CRISPR motif in the regulatory sequence of another gene locus in the organism. In certain embodiments, the GC content of the targeting sequence 60% or less, 55% or less, 50% or less, 45% or less, 40% or less, 35% or less or 30% or less. In certain embodiments, the GC content of the targeting sequence is from 70% to 60% or from 60% to 50% or from 50%

to 40% or from 40% to 30%. In an embodiment, the targeting sequence has the lowest CG content among potential targeting sequences of the locus.

[00112] In an embodiment of the invention, the first 15 nt of the dead guide match the target sequence. In another embodiment, first 14 nt of the dead guide match the target sequence. In another embodiment, the first 13 nt of the dead guide match the target sequence. In another embodiment first 12 nt of the dead guide match the target sequence. In another embodiment, first 11 nt of the dead guide match the target sequence. In another embodiment, the first 10 nt of the dead guide match the target sequence. In an embodiment of the invention the first 15 nt of the dead guide does not match an off-target sequence downstream from a CRISPR motif in the regulatory region of another gene locus. In other embodiments, the first 14 nt, or the first 13 nt of the dead guide, or the first 12 nt of the guide, or the first 11 nt of the dead guide, or the first 10 nt of the dead guide, does not match an off-target sequence downstream from a CRISPR motif in the regulatory region of another gene locus. In other embodiments, the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt of the dead guide do not match an off-target sequence downstream from a CRISPR motif in the genome.

[00113] In certain embodiments, the dead guide RNA includes additional nucleotides at the 3'-end that do not match the target sequence. Thus, a dead guide RNA that includes the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt downstream of a CRISPR motif can be extended in length at the 3' end to 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, or longer.

[00114] The invention provides a method for directing a Cas13 CRISPR-Cas system, including but not limited to a dead Cas13 (dCas13) or functionalized Cas13 system (which may comprise a functionalized Cas13 or functionalized guide) to a gene locus. In an aspect, the invention provides a method for selecting a dead guide RNA targeting sequence and directing a functionalized CRISPR system to a gene locus in an organism. In an aspect, the invention provides a method for selecting a dead guide RNA targeting sequence and effecting gene regulation of a target gene locus by a functionalized Cas13 CRISPR-Cas system. In certain embodiments, the method is used to effect target gene regulation while minimizing off-target effects. In an aspect, the invention provides a method for selecting two or more dead guide RNA targeting sequences and effecting gene regulation of two or more target gene loci by a

functionalized Cas13 CRISPR-Cas system. In certain embodiments, the method is used to effect regulation of two or more target gene loci while minimizing off-target effects.

[00115] In an aspect, the invention provides a method of selecting a dead guide RNA targeting sequence for directing a functionalized Cas13 to a gene locus in an organism, which comprises: a) locating one or more CRISPR motifs in the gene locus; b) analyzing the sequence downstream of each CRISPR motif by: i) selecting 10 to 15 nt adjacent to the CRISPR motif, ii) determining the GC content of the sequence; and c) selecting the 10 to 15 nt sequence as a targeting sequence for use in a guide RNA if the GC content of the sequence is 40% or more. In an embodiment, the sequence is selected if the GC content is 50% or more. In an embodiment, the sequence is selected if the GC content is 60% or more. In an embodiment, the sequence is selected if the GC content is 70% or more. In an embodiment, two or more sequences are analyzed and the sequence having the highest GC content is selected. In an embodiment, the method further comprises adding nucleotides to the 3' end of the selected sequence which do not match the sequence downstream of the CRISPR motif. An aspect provides a dead guide RNA comprising the targeting sequence selected according to the aforementioned methods.

[00116] In an aspect, the invention provides a dead guide RNA for directing a functionalized CRISPR system to a gene locus in an organism wherein the targeting sequence of the dead guide RNA consists of 10 to 15 nucleotides adjacent to the CRISPR motif of the gene locus, wherein the CG content of the target sequence is 50% or more. In certain embodiments, the dead guide RNA further comprises nucleotides added to the 3' end of the targeting sequence which do not match the sequence downstream of the CRISPR motif of the gene locus.

[00117] In an aspect, the invention provides for a single effector to be directed to one or more, or two or more gene loci. In certain embodiments, the effector is associated with a Cas13, and one or more, or two or more selected dead guide RNAs are used to direct the Cas13-associated effector to one or more, or two or more selected target gene loci. In certain embodiments, the effector is associated with one or more, or two or more selected dead guide RNAs, each selected dead guide RNA, when complexed with a Cas13 enzyme, causing its associated effector to localize to the dead guide RNA target. One non-limiting example of such CRISPR systems modulates activity of one or more, or two or more gene loci subject to regulation by the same transcription factor.

[00118] In an aspect, the invention provides for two or more effectors to be directed to one or more gene loci. In certain embodiments, two or more dead guide RNAs are employed, each of the two or more effectors being associated with a selected dead guide RNA, with each of the two or more effectors being localized to the selected target of its dead guide RNA. One non-limiting example of such CRISPR systems modulates activity of one or more, or two or more gene loci subject to regulation by different transcription factors. Thus, in one non-limiting embodiment, two or more transcription factors are localized to different regulatory sequences of a single gene. In another non-limiting embodiment, two or more transcription factors are localized to different regulatory sequences of different genes. In certain embodiments, one transcription factor is an activator. In certain embodiments, one transcription factor is an inhibitor. In certain embodiments, one transcription factor is an activator and another transcription factor is an inhibitor. In certain embodiments, gene loci expressing different components of the same regulatory pathway are regulated. In certain embodiments, gene loci expressing components of different regulatory pathways are regulated.

[00119] In an aspect, the invention also provides a method and algorithm for designing and selecting dead guide RNAs that are specific for target DNA cleavage or target binding and gene regulation mediated by an active Cas13 CRISPR-Cas system. In certain embodiments, the Cas13 CRISPR-Cas system provides orthogonal gene control using an active Cas13 which cleaves target DNA at one gene locus while at the same time binds to and promotes regulation of another gene locus.

[00120] In an aspect, the invention provides an method of selecting a dead guide RNA targeting sequence for directing a functionalized Cas13 to a gene locus in an organism, without cleavage, which comprises a) locating one or more CRISPR motifs in the gene locus; b) analyzing the sequence downstream of each CRISPR motif by i) selecting 10 to 15 nt adjacent to the CRISPR motif, ii) determining the GC content of the sequence, and c) selecting the 10 to 15 nt sequence as a targeting sequence for use in a dead guide RNA if the GC content of the sequence is 30% more, 40% or more. In certain embodiments, the GC content of the targeting sequence is 35% or more, 40% or more, 45% or more, 50% or more, 55% or more, 60% or more, 65% or more, or 70% or more. In certain embodiments, the GC content of the targeting sequence is from 30% to 40% or from 40% to 50% or from 50% to 60% or from 60% to 70%. In an

embodiment of the invention, two or more sequences in a gene locus are analyzed and the sequence having the highest GC content is selected.

[00121] In an embodiment of the invention, the portion of the targeting sequence in which GC content is evaluated is 10 to 15 contiguous nucleotides of the 15 target nucleotides nearest to the PAM. In an embodiment of the invention, the portion of the guide in which GC content is considered is the 10 to 11 nucleotides or 11 to 12 nucleotides or 12 to 13 nucleotides or 13, or 14, or 15 contiguous nucleotides of the 15 nucleotides nearest to the PAM.

[00122] In an aspect, the invention further provides an algorithm for identifying dead guide RNAs which promote CRISPR system gene locus cleavage while avoiding functional activation or inhibition. It is observed that increased GC content in dead guide RNAs of 16 to 20 nucleotides coincides with increased DNA cleavage and reduced functional activation.

[00123] It is also demonstrated herein that efficiency of functionalized Cas13 can be increased by addition of nucleotides to the 3' end of a guide RNA which do not match a target sequence downstream of the CRISPR motif. For example, of dead guide RNA 11 to 15 nt in length, shorter guides may be less likely to promote target cleavage, but are also less efficient at promoting CRISPR system binding and functional control. In certain embodiments, addition of nucleotides that don't match the target sequence to the 3' end of the dead guide RNA increase activation efficiency while not increasing undesired target cleavage. In an aspect, the invention also provides a method and algorithm for identifying improved dead guide RNAs that effectively promote CRISPRP system function in DNA binding and gene regulation while not promoting DNA cleavage. Thus, in certain embodiments, the invention provides a dead guide RNA that includes the first 15 nt, or 14 nt, or 13 nt, or 12 nt, or 11 nt downstream of a CRISPR motif and is extended in length at the 3' end by nucleotides that mismatch the target to 12 nt, 13 nt, 14 nt, 15 nt, 16 nt, 17 nt, 18 nt, 19 nt, 20 nt, or longer.

[00124] In an aspect, the invention provides a method for effecting selective orthogonal gene control. As will be appreciated from the disclosure herein, dead guide selection according to the invention, taking into account guide length and GC content, provides effective and selective transcription control by a functional Cas13 CRISPR-Cas system, for example to regulate transcription of a gene locus by activation or inhibition and minimize off-target effects.

Accordingly, by providing effective regulation of individual target loci, the invention also provides effective orthogonal regulation of two or more target loci.

[00125] In certain embodiments, orthogonal gene control is by activation or inhibition of two or more target loci. In certain embodiments, orthogonal gene control is by activation or inhibition of one or more target locus and cleavage of one or more target locus.

[00126] In one aspect, the invention provides a cell comprising a non-naturally occurring Cas13 CRISPR-Cas system comprising one or more dead guide RNAs disclosed or made according to a method or algorithm described herein wherein the expression of one or more gene products has been altered. In an embodiment of the invention, the expression in the cell of two or more gene products has been altered. The invention also provides a cell line from such a cell.

[00127] In one aspect, the invention provides a multicellular organism comprising one or more cells comprising a non-naturally occurring Cas13 CRISPR-Cas system comprising one or more dead guide RNAs disclosed or made according to a method or algorithm described herein. In one aspect, the invention provides a product from a cell, cell line, or multicellular organism comprising a non-naturally occurring Cas13 CRISPR-Cas system comprising one or more dead guide RNAs disclosed or made according to a method or algorithm described herein.

[00128] A further aspect of this invention is the use of gRNA comprising dead guide(s) as described herein, optionally in combination with gRNA comprising guide(s) as described herein or in the state of the art, in combination with systems e.g. cells, transgenic animals, transgenic mice, inducible transgenic animals, inducible transgenic mice) which are engineered for either overexpression of Cas13 or preferably knock in Cas13. As a result a single system (e.g. transgenic animal, cell) can serve as a basis for multiplex gene modifications in systems / network biology. On account of the dead guides, this is now possible in both in vitro, ex vivo, and in vivo.

[00129] For example, once the Cas13 is provided for, one or more dead gRNAs may be provided to direct multiplex gene regulation, and preferably multiplex bidirectional gene regulation. The one or more dead gRNAs may be provided in a spatially and temporally appropriate manner if necessary or desired (for example tissue specific induction of Cas13 expression). On account that the transgenic / inducible Cas13 is provided for (e.g. expressed) in the cell, tissue, animal of interest, both gRNAs comprising dead guides or gRNAs comprising

guides are equally effective. In the same manner, a further aspect of this invention is the use of gRNA comprising dead guide(s) as described herein, optionally in combination with gRNA comprising guide(s) as described herein or in the state of the art, in combination with systems (e.g. cells, transgenic animals, transgenic mice, inducible transgenic animals, inducible transgenic mice) which are engineered for knockout Cas13 CRISPR-Cas.

[00130] As a result, the combination of dead guides as described herein with CRISPR applications described herein and CRISPR applications known in the art results in a highly efficient and accurate means for multiplex screening of systems (e.g. network biology). Such screening allows, for example, identification of specific combinations of gene activities for identifying genes responsible for diseases (e.g. on/off combinations), in particular gene related diseases. A preferred application of such screening is cancer. In the same manner, screening for treatment for such diseases is included in the invention. Cells or animals may be exposed to aberrant conditions resulting in disease or disease like effects. Candidate compositions may be provided and screened for an effect in the desired multiplex environment. For example a patient's cancer cells may be screened for which gene combinations will cause them to die, and then use this information to establish appropriate therapies.

[00131] In one aspect, the invention provides a kit comprising one or more of the components described herein. The kit may include dead guides as described herein with or without guides as described herein.

[00132] The structural information provided herein allows for interrogation of dead gRNA interaction with the target DNA and the Cas13 permitting engineering or alteration of dead gRNA structure to optimize functionality of the entire Cas13 CRISPR-Cas system. For example, loops of the dead gRNA may be extended, without colliding with the Cas13 protein by the insertion of adaptor proteins that can bind to RNA. These adaptor proteins can further recruit effector proteins or fusions which comprise one or more functional domains.

[00133] In some preferred embodiments, the functional domain is a transcriptional activation domain, preferably VP64. In some embodiments, the functional domain is a transcription repression domain, preferably KRAB. In some embodiments, the transcription repression domain is SID, or concatemers of SID (e.g. SID4X). In some embodiments, the functional domain is an epigenetic modifying domain, such that an epigenetic modifying enzyme is

provided. In some embodiments, the functional domain is an activation domain, which may be the P65 activation domain.

[00134] An aspect of the invention is that the above elements are comprised in a single composition or comprised in individual compositions. These compositions may advantageously be applied to a host to elicit a functional effect on the genomic level.

[00135] In general, the dead gRNA is modified in a manner that provides specific binding sites (e.g. aptamers) for adapter proteins comprising one or more functional domains (e.g. via fusion protein) to bind to. The modified dead gRNA is modified such that once the dead gRNA forms a CRISPR complex (i.e. Cas13 binding to dead gRNA and target) the adapter proteins bind and, the functional domain on the adapter protein is positioned in a spatial orientation which is advantageous for the attributed function to be effective. For example, if the functional domain is a transcription activator (e.g. VP64 or p65), the transcription activator is placed in a spatial orientation which allows it to affect the transcription of the target. Likewise, a transcription repressor will be advantageously positioned to affect the transcription of the target and a nuclease (e.g. FokI) will be advantageously positioned to cleave or partially cleave the target.

[00136] The skilled person will understand that modifications to the dead gRNA which allow for binding of the adapter + functional domain but not proper positioning of the adapter + functional domain (e.g. due to steric hindrance within the three dimensional structure of the CRISPR complex) are modifications which are not intended. The one or more modified dead gRNA may be modified at the tetra loop, the stem loop 1, stem loop 2, or stem loop 3, as described herein, preferably at either the tetra loop or stem loop 2, and most preferably at both the tetra loop and stem loop 2.

[00137] As explained herein the functional domains may be, for example, one or more domains from the group consisting of methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, DNA cleavage activity, nucleic acid binding activity, and molecular switches (e.g. light inducible). In some cases it is advantageous that additionally at least one NLS is provided. In some instances, it is advantageous to position the NLS at the N terminus. When more than one functional domain is included, the functional domains may be the same or different.

[00138] The dead gRNA may be designed to include multiple binding recognition sites (e.g. aptamers) specific to the same or different adapter protein. The dead gRNA may be designed to bind to the promoter region -1000 - +1 nucleic acids upstream of the transcription start site (i.e. TSS), preferably -200 nucleic acids. This positioning improves functional domains which affect gene activation (e.g. transcription activators) or gene inhibition (e.g. transcription repressors). The modified dead gRNA may be one or more modified dead gRNAs targeted to one or more target loci (e.g. at least 1 gRNA, at least 2 gRNA, at least 5 gRNA, at least 10 gRNA, at least 20 gRNA, at least 30 gRNA, at least 50 gRNA) comprised in a composition.

[00139] The adaptor protein may be any number of proteins that binds to an aptamer or recognition site introduced into the modified dead gRNA and which allows proper positioning of one or more functional domains, once the dead gRNA has been incorporated into the CRISPR complex, to affect the target with the attributed function. As explained in detail in this application such may be coat proteins, preferably bacteriophage coat proteins. The functional domains associated with such adaptor proteins (e.g. in the form of fusion protein) may include, for example, one or more domains from the group consisting of methylase activity, demethylase activity, transcription activation activity, transcription repression activity, transcription release factor activity, histone modification activity, RNA cleavage activity, DNA cleavage activity, nucleic acid binding activity, and molecular switches (e.g. light inducible). Preferred domains are Fok1, VP64, P65, HSF1, MyoD1. In the event that the functional domain is a transcription activator or transcription repressor it is advantageous that additionally at least an NLS is provided and preferably at the N terminus. When more than one functional domain is included, the functional domains may be the same or different. The adaptor protein may utilize known linkers to attach such functional domains.

[00140] Thus, the modified dead gRNA, the (inactivated) Cas13 (with or without functional domains), and the binding protein with one or more functional domains, may each individually be comprised in a composition and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or described herein for delivery to a host (e.g. lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g. for lentiviral gRNA selection) and

concentration of gRNA (e.g. dependent on whether multiple gRNAs are used) may be advantageous for eliciting an improved effect.

[00141] On the basis of this concept, several variations are appropriate to elicit a genomic locus event, including DNA cleavage, gene activation, or gene deactivation. Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic locus events. The compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling in vivo (e.g. gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling; the use the compositions of the invention to establish cell lines and transgenic animals for optimization and screening purposes).

[00142] The current invention comprehends the use of the compositions of the current invention to establish and utilize conditional or inducible CRISPR transgenic cell /animals, which are not believed prior to the present invention or application. For example, the target cell comprises Cas13 conditionally or inducibly (e.g. in the form of Cre dependent constructs) and/or the adapter protein conditionally or inducibly and, on expression of a vector introduced into the target cell, the vector expresses that which induces or gives rise to the condition of Cas13 expression and/or adaptor expression in the target cell. By applying the teaching and compositions of the current invention with the known method of creating a CRISPR complex, inducible genomic events affected by functional domains are also an aspect of the current invention. One example of this is the creation of a CRISPR knock-in / conditional transgenic animal (e.g. mouse comprising e.g. a Lox-Stop-polyA-Lox(LSL) cassette) and subsequent delivery of one or more compositions providing one or more modified dead gRNA (e.g. -200 nucleotides to TSS of a target gene of interest for gene activation purposes) as described herein (e.g. modified dead gRNA with one or more aptamers recognized by coat proteins, e.g. MS2), one or more adapter proteins as described herein (MS2 binding protein linked to one or more VP64) and means for inducing the conditional animal (e.g. Cre recombinase for rendering Cas13 expression inducible). Alternatively, the adaptor protein may be provided as a conditional or inducible element with a conditional or inducible Cas13 to provide an effective model for screening purposes, which advantageously only requires minimal design and administration of specific dead gRNAs for a broad number of applications.

[00143] In another aspect the dead guides are further modified to improve specificity. Protected dead guides may be synthesized, whereby secondary structure is introduced into the 3' end of the dead guide to improve its specificity. A protected guide RNA (pgRNA) comprises a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell and a protector strand, wherein the protector strand is optionally complementary to the guide sequence and wherein the guide sequence may in part be hybridizable to the protector strand. The pgRNA optionally includes an extension sequence. The thermodynamics of the pgRNA-target DNA hybridization is determined by the number of bases complementary between the guide RNA and target DNA. By employing 'thermodynamic protection', specificity of dead gRNA can be improved by adding a protector sequence. For example, one method adds a complementary protector strand of varying lengths to the 3' end of the guide sequence within the dead gRNA. As a result, the protector strand is bound to at least a portion of the dead gRNA and provides for a protected gRNA (pgRNA). In turn, the dead gRNA references herein may be easily protected using the described embodiments, resulting in pgRNA. The protector strand can be either a separate RNA transcript or strand or a chimeric version joined to the 3' end of the dead gRNA guide sequence.

Tandem guides and uses in a multiplex (tandem) targeting approach

[00144] The inventors have shown that CRISPR enzymes as defined herein can employ more than one RNA guide without losing activity. This enables the use of the CRISPR enzymes, systems or complexes as defined herein for targeting multiple DNA targets, genes or gene loci, with a single enzyme, system or complex as defined herein. The guide RNAs may be tandemly arranged, optionally separated by a nucleotide sequence such as a direct repeat as defined herein. The position of the different guide RNAs in the tandem does not influence the activity. It is noted that the terms "CRISPR-Cas system", "CRISPR-Cas complex", "CRISPR complex" and "CRISPR system" are used interchangeably. Also the terms "CRISPR enzyme", "Cas enzyme", or "CRISPR-Cas enzyme", can be used interchangeably. In preferred embodiments, said CRISPR enzyme, CRISPR-Cas enzyme or Cas enzyme is Cas13, or any one of the modified or mutated variants thereof described herein elsewhere.

[00145] In one aspect, the invention provides a non-naturally occurring or engineered CRISPR enzyme, preferably a class 2 CRISPR enzyme, preferably a Type V or VI CRISPR

enzyme as described herein, such as without limitation Cas13 as described herein elsewhere, used for tandem or multiplex targeting. It is to be understood that any of the CRISPR (or CRISPR-Cas or Cas) enzymes, complexes, or systems according to the invention as described herein elsewhere may be used in such an approach. Any of the methods, products, compositions and uses as described herein elsewhere are equally applicable with the multiplex or tandem targeting approach further detailed below. By means of further guidance, the following particular aspects and embodiments are provided.

[00146] In one aspect, the invention provides for the use of a Cas13 enzyme, complex or system as defined herein for targeting multiple gene loci. In one embodiment, this can be established by using multiple (tandem or multiplex) guide RNA (gRNA) sequences.

[00147] In one aspect, the invention provides methods for using one or more elements of a Cas13 enzyme, complex or system as defined herein for tandem or multiplex targeting, wherein said CRISPR system comprises multiple guide RNA sequences. Preferably, said gRNA sequences are separated by a nucleotide sequence, such as a direct repeat as defined herein elsewhere.

[00148] The Cas13 enzyme, system or complex as defined herein provides an effective means for modifying multiple target polynucleotides. The Cas13 enzyme, system or complex as defined herein has a wide variety of utility including modifying (e.g., deleting, inserting, translocating, inactivating, activating) one or more target polynucleotides in a multiplicity of cell types. As such the Cas13 enzyme, system or complex as defined herein of the invention has a broad spectrum of applications in, e.g., gene therapy, drug screening, disease diagnosis, and prognosis, including targeting multiple gene loci within a single CRISPR system.

[00149] In one aspect, the invention provides a Cas13 enzyme, system or complex as defined herein, i.e. a Cas13 CRISPR-Cas complex having a Cas13 protein having at least one destabilization domain associated therewith, and multiple guide RNAs that target multiple nucleic acid molecules such as DNA molecules, whereby each of said multiple guide RNAs specifically targets its corresponding nucleic acid molecule, e.g., DNA molecule. Each nucleic acid molecule target, e.g., DNA molecule can encode a gene product or encompass a gene locus. Using multiple guide RNAs hence enables the targeting of multiple gene loci or multiple genes. In some embodiments the Cas13 enzyme may cleave the RNA molecule encoding the gene product. In some embodiments expression of the gene product is altered. The Cas13 protein and

the guide RNAs do not naturally occur together. The invention comprehends the guide RNAs comprising tandemly arranged guide sequences. The invention further comprehends coding sequences for the Cas13 protein being codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell and in a more preferred embodiment the mammalian cell is a human cell. Expression of the gene product may be decreased. The Cas13 enzyme may form part of a CRISPR system or complex, which further comprises tandemly arranged guide RNAs (gRNAs) comprising a series of 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 25, 30, or more than 30 guide sequences, each capable of specifically hybridizing to a target sequence in a genomic locus of interest in a cell. In some embodiments, the functional Cas13 CRISPR system or complex binds to the multiple target sequences. In some embodiments, the functional CRISPR system or complex may edit the multiple target sequences, e.g., the target sequences may comprise a genomic locus, and in some embodiments there may be an alteration of gene expression. In some embodiments, the functional CRISPR system or complex may comprise further functional domains. In some embodiments, the invention provides a method for altering or modifying expression of multiple gene products. The method may comprise introducing into a cell containing said target nucleic acids, e.g., DNA molecules, or containing and expressing target nucleic acid, e.g., DNA molecules; for instance, the target nucleic acids may encode gene products or provide for expression of gene products (e.g., regulatory sequences).

[00150] In preferred embodiments the CRISPR enzyme used for multiplex targeting is Cas13, or the CRISPR system or complex comprises Cas13. In some embodiments, the CRISPR enzyme used for multiplex targeting is AsCas13, or the CRISPR system or complex used for multiplex targeting comprises an AsCas13. In some embodiments, the CRISPR enzyme is an LbCas13, or the CRISPR system or complex comprises LbCas13. In some embodiments, the Cas enzyme used for multiplex targeting cleaves both strands of DNA to produce a double strand break (DSB). In some embodiments, the CRISPR enzyme used for multiplex targeting is a nickase. In some embodiments, the Cas13 enzyme used for multiplex targeting is a dual nickase. In some embodiments, the Cas13 enzyme used for multiplex targeting is a Cas13 enzyme such as a DD Cas13 enzyme as defined herein elsewhere.

[00151] In some general embodiments, the Cas13 enzyme used for multiplex targeting is associated with one or more functional domains. In some more specific embodiments, the CRISPR enzyme used for multiplex targeting is a deadCas13 as defined herein elsewhere.

[00152] In an aspect, the present invention provides a means for delivering the Cas13 enzyme, system or complex for use in multiple targeting as defined herein or the polynucleotides defined herein. Non-limiting examples of such delivery means are e.g. particle(s) delivering component(s) of the complex, vector(s) comprising the polynucleotide(s) discussed herein (e.g., encoding the CRISPR enzyme, providing the nucleotides encoding the CRISPR complex). In some embodiments, the vector may be a plasmid or a viral vector such as AAV, or lentivirus. Transient transfection with plasmids, e.g., into HEK cells may be advantageous, especially given the size limitations of AAV and that while Cas13 fits into AAV, one may reach an upper limit with additional guide RNAs.

[00153] Also provided is a model that constitutively expresses the Cas13 enzyme, complex or system as used herein for use in multiplex targeting. The organism may be transgenic and may have been transfected with the present vectors or may be the offspring of an organism so transfected. In a further aspect, the present invention provides compositions comprising the CRISPR enzyme, system and complex as defined herein or the polynucleotides or vectors described herein. Also provided are Cas13 CRISPR systems or complexes comprising multiple guide RNAs, preferably in a tandemly arranged format. Said different guide RNAs may be separated by nucleotide sequences such as direct repeats.

[00154] Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing gene editing by transforming the subject with the polynucleotide encoding the Cas13 CRISPR system or complex or any of polynucleotides or vectors described herein and administering them to the subject. A suitable repair template may also be provided, for example delivered by a vector comprising said repair template. Also provided is a method of treating a subject, e.g., a subject in need thereof, comprising inducing transcriptional activation or repression of multiple target gene loci by transforming the subject with the polynucleotides or vectors described herein, wherein said polynucleotide or vector encodes or comprises the Cas13 enzyme, complex or system comprising multiple guide RNAs, preferably tandemly arranged.

Where any treatment is occurring ex vivo, for example in a cell culture, then it will be appreciated that the term 'subject' may be replaced by the phrase "cell or cell culture."

[00155] Compositions comprising Cas13 enzyme, complex or system comprising multiple guide RNAs, preferably tandemly arranged, or the polynucleotide or vector encoding or comprising said Cas13 enzyme, complex or system comprising multiple guide RNAs, preferably tandemly arranged, for use in the methods of treatment as defined herein elsewhere are also provided. A kit of parts may be provided including such compositions. Use of said composition in the manufacture of a medicament for such methods of treatment are also provided. Use of a Cas13 CRISPR system in screening is also provided by the present invention, e.g., gain of function screens. Cells which are artificially forced to overexpress a gene are able to down regulate the gene over time (re-establishing equilibrium) e.g. by negative feedback loops. By the time the screen starts the unregulated gene might be reduced again. Using an inducible Cas13 activator allows one to induce transcription right before the screen and therefore minimizes the chance of false negative hits. Accordingly, by use of the instant invention in screening, e.g., gain of function screens, the chance of false negative results may be minimized.

[00156] In one aspect, the invention provides an engineered, non-naturally occurring CRISPR system comprising a Cas13 protein and multiple guide RNAs that each specifically target a DNA molecule encoding a gene product in a cell, whereby the multiple guide RNAs each target their specific DNA molecule encoding the gene product and the Cas13 protein cleaves the target DNA molecule encoding the gene product, whereby expression of the gene product is altered; and, wherein the CRISPR protein and the guide RNAs do not naturally occur together. The invention comprehends the multiple guide RNAs comprising multiple guide sequences, preferably separated by a nucleotide sequence such as a direct repeat and optionally fused to a tracr sequence. In an embodiment of the invention the CRISPR protein is a type V or VI CRISPR-Cas protein and in a more preferred embodiment the CRISPR protein is a Cas13 protein. The invention further comprehends a Cas13 protein being codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of the gene product is decreased.

[00157] In another aspect, the invention provides an engineered, non-naturally occurring vector system comprising one or more vectors comprising a first regulatory element operably linked to the multiple Cas13 CRISPR system guide RNAs that each specifically target a DNA molecule encoding a gene product and a second regulatory element operably linked coding for a CRISPR protein. Both regulatory elements may be located on the same vector or on different vectors of the system. The multiple guide RNAs target the multiple DNA molecules encoding the multiple gene products in a cell and the CRISPR protein may cleave the multiple DNA molecules encoding the gene products (it may cleave one or both strands or have substantially no nuclease activity), whereby expression of the multiple gene products is altered; and, wherein the CRISPR protein and the multiple guide RNAs do not naturally occur together. In a preferred embodiment the CRISPR protein is Cas13 protein, optionally codon optimized for expression in a eukaryotic cell. In a preferred embodiment the eukaryotic cell is a mammalian cell, a plant cell or a yeast cell and in a more preferred embodiment the mammalian cell is a human cell. In a further embodiment of the invention, the expression of each of the multiple gene products is altered, preferably decreased.

[00158] In one aspect, the invention provides a vector system comprising one or more vectors. In some embodiments, the system comprises: (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the one or more guide sequence(s) direct(s) sequence-specific binding of the CRISPR complex to the one or more target sequence(s) in a eukaryotic cell, wherein the CRISPR complex comprises a Cas13 enzyme complexed with the one or more guide sequence(s) that is hybridized to the one or more target sequence(s); and (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas13 enzyme, preferably comprising at least one nuclear localization sequence and/or at least one NES; wherein components (a) and (b) are located on the same or different vectors of the system. Where applicable, a tracr sequence may also be provided. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a Cas13 CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the CRISPR complex

comprises one or more nuclear localization sequences and/or one or more NES of sufficient strength to drive accumulation of said Cas13 CRISPR complex in a detectable amount in or out of the nucleus of a eukaryotic cell. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, each of the guide sequences is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length.

[00159] Recombinant expression vectors can comprise the polynucleotides encoding the Cas13 enzyme, system or complex for use in multiple targeting as defined herein 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).

[00160] In some embodiments, a host cell is transiently or non-transiently transfected with one or more vectors comprising the polynucleotides encoding the Cas13 enzyme, system or complex for use in multiple targeting as defined 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 and exemplified herein elsewhere. 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 comprising the polynucleotides encoding the Cas13 enzyme, system or complex for use in multiple targeting as defined 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 Cas13 CRISPR system or complex for use in multiple targeting as described herein (such as by transient transfection of one or more vectors, or transfection with RNA), and modified through the activity of a Cas13 CRISPR system or 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 comprising the polynucleotides encoding the Cas13 enzyme, system or complex for use in multiple targeting as defined herein, or cell lines derived from such cells are used in assessing one or more test compounds.

[00161] The term "regulatory element" is as defined herein elsewhere.

[00162] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[00163] In one aspect, the invention provides a eukaryotic host cell comprising (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide RNA sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the guide sequence(s) direct(s) sequence-specific binding of the Cas13 CRISPR complex to the respective target sequence(s) in a eukaryotic cell, wherein the Cas13 CRISPR complex comprises a Cas13 enzyme complexed with the one or more guide sequence(s) that is hybridized to the respective target sequence(s); and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas13 enzyme comprising preferably at least one nuclear localization sequence and/or NES. In some embodiments, the host cell comprises components (a) and (b). Where applicable, a tracr sequence may also be provided. In some embodiments, component (a), component (b), or components (a) and (b) are stably integrated into a genome of the host eukaryotic cell. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, and optionally separated by a direct repeat, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a Cas13 CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas13 enzyme comprises one or more nuclear localization sequences and/or nuclear export sequences or NES of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in and/or out of the nucleus of a eukaryotic cell.

[00164] In some embodiments, the Cas13 enzyme is a type V or VI CRISPR system enzyme. In some embodiments, the Cas enzyme is a Cas13 enzyme. In some embodiments, the Cas13 enzyme is derived from *Francisella tularensis* 1, *Francisella tularensis* subsp. *novicida*, *Prevotella*

albensis, Lachnospiraceae bacterium MC2017 1, Butyrivibrio proteoclasticus, Peregrinibacteria bacterium GW2011_GWA2_33_10, Parcubacteria bacterium GW2011_GWC2_44_17, Smithella sp. SCADC, Acidaminococcus sp. BV3L6, Lachnospiraceae bacterium MA2020, Candidatus Methanoplasma termitum, Eubacterium eligens, Moraxella bovoculi 237, Leptospira inadai, Lachnospiraceae bacterium ND2006, Porphyromonas crevioricanis 3, Prevotella disiens, or Porphyromonas macacae Cas13, and may include further alterations or mutations of the Cas13 as defined herein elsewhere, and can be a chimeric Cas13. In some embodiments, the Cas13 enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the one or more guide sequence(s) is (are each) at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length. When multiple guide RNAs are used, they are preferably separated by a direct repeat sequence. In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example a mammal. Also, the organism may be an arthropod such as an insect. The organism also may be a plant. Further, the organism may be a fungus.

[00165] In one aspect, the invention provides a kit comprising one or more of the components described herein. In some embodiments, the kit comprises a vector system and instructions for using the kit. In some embodiments, the vector system comprises (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a Cas13 CRISPR complex to a target sequence in a eukaryotic cell, wherein the Cas13 CRISPR complex comprises a Cas13 enzyme complexed with the guide sequence that is hybridized to the target sequence; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence

encoding said Cas13 enzyme comprising a nuclear localization sequence. Where applicable, a tracr sequence may also be provided. In some embodiments, the kit comprises components (a) and (b) located on the same or different vectors of the system. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In some embodiments, the Cas13 enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In some embodiments, the CRISPR enzyme is a type V or VI CRISPR system enzyme. In some embodiments, the CRISPR enzyme is a Cas13 enzyme. In some embodiments, the Cas13 enzyme is derived from *Francisella tularensis* 1, *Francisella tularensis* subsp. *novicida*, *Prevotella albensis*, *Lachnospiraceae* bacterium MC2017 1, *Butyrivibrio proteoclasticus*, *Peregrinibacteria* bacterium GW2011_GWA2_33_10, *Parcubacteria* bacterium GW2011_GWC2_44_17, *Smithella* sp. SCADC, *Acidaminococcus* sp. BV3L6, *Lachnospiraceae* bacterium MA2020, *Candidatus Methanoplasma termitum*, *Eubacterium eligens*, *Moraxella bovoculi* 237, *Leptospira inadai*, *Lachnospiraceae* bacterium ND2006, *Porphyromonas crevioricanis* 3, *Prevotella disiens*, or *Porphyromonas macacae* Cas13 (e.g., modified to have or be associated with at least one DD), and may include further alteration or mutation of the Cas13, and can be a chimeric Cas13. In some embodiments, the DD-CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the DD-CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In some embodiments, the DD-CRISPR enzyme lacks or substantially DNA strand cleavage activity (e.g., no more than 5% nuclease activity as compared with a wild type enzyme or enzyme not having the mutation or alteration that decreases nuclease activity). In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the guide sequence is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length.

[00166] In one aspect, the invention provides a method of modifying multiple target polynucleotides in a host cell such as a eukaryotic cell. In some embodiments, the method

comprises allowing a Cas13 CRISPR complex to bind to multiple target polynucleotides, e.g., to effect cleavage of said multiple target polynucleotides, thereby modifying multiple target polynucleotides, wherein the Cas13 CRISPR complex comprises a Cas13 enzyme complexed with multiple guide sequences each of the being hybridized to a specific target sequence within said target polynucleotide, wherein said multiple guide sequences are linked to a direct repeat sequence. Where applicable, a tracr sequence may also be provided (e.g. to provide a single guide RNA, sgRNA). In some embodiments, said cleavage comprises cleaving one or two strands at the location of each of the target sequence by said Cas13 enzyme. In some embodiments, said cleavage results in decreased transcription of the multiple target genes. In some embodiments, the method further comprises repairing one or more of 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 one or more of said target polynucleotides. In some embodiments, said mutation results in one or more amino acid changes in a protein expressed from a gene comprising one or more of the target sequence(s). 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 Cas13 enzyme and the multiple guide RNA sequence linked to a direct repeat sequence. Where applicable, a tracr sequence may also be provided. 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.

[00167] In one aspect, the invention provides a method of modifying expression of multiple polynucleotides in a eukaryotic cell. In some embodiments, the method comprises allowing a Cas13 CRISPR complex to bind to multiple polynucleotides such that said binding results in increased or decreased expression of said polynucleotides; wherein the Cas13 CRISPR complex comprises a Cas13 enzyme complexed with multiple guide sequences each specifically hybridized to its own target sequence within said polynucleotide, wherein said guide sequences are linked to a direct repeat sequence. Where applicable, a tracr sequence may also be provided.

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 Cas13 enzyme and the multiple guide sequences linked to the direct repeat sequences. Where applicable, a tracr sequence may also be provided.

[00168] In one aspect, the invention provides a recombinant polynucleotide comprising multiple guide RNA sequences up- or downstream (whichever applicable) of a direct repeat sequence, wherein each of the guide sequences when expressed directs sequence-specific binding of a Cas13 CRISPR complex to its corresponding target sequence present in a eukaryotic cell. In some embodiments, the target sequence is a viral sequence present in a eukaryotic cell. Where applicable, a tracr sequence may also be provided. In some embodiments, the target sequence is a proto-oncogene or an oncogene.

[00169] Aspects of the invention encompass a non-naturally occurring or engineered composition that may comprise a guide RNA (gRNA) comprising a guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell and a Cas13 enzyme as defined herein that may comprise at least one or more nuclear localization sequences.

[00170] An aspect of the invention encompasses methods of modifying a genomic locus of interest to change gene expression in a cell by introducing into the cell any of the compositions described herein.

[00171] An aspect of the invention is that the above elements are comprised in a single composition or comprised in individual compositions. These compositions may advantageously be applied to a host to elicit a functional effect on the genomic level.

[00172] As used herein, the term "guide RNA" or "gRNA" has the meaning as used herein elsewhere and comprises any polynucleotide sequence having sufficient complementarity with a target nucleic acid sequence to hybridize with the target nucleic acid sequence and direct sequence-specific binding of a nucleic acid-targeting complex to the target nucleic acid sequence. Each gRNA may be designed to include multiple binding recognition sites (e.g., aptamers) specific to the same or different adapter protein. Each gRNA may be designed to bind to the promoter region -1000 - +1 nucleic acids upstream of the transcription start site (i.e. TSS), preferably -200 nucleic acids. This positioning improves functional domains which affect gene activation (e.g., transcription activators) or gene inhibition (e.g., transcription repressors). The

modified gRNA may be one or more modified gRNAs targeted to one or more target loci (e.g., at least 1 gRNA, at least 2 gRNA, at least 5 gRNA, at least 10 gRNA, at least 20 gRNA, at least 30 gRNA, at least 50 gRNA) comprised in a composition. Said multiple gRNA sequences can be tandemly arranged and are preferably separated by a direct repeat.

[00173] Thus, gRNA, the CRISPR enzyme as defined herein may each individually be comprised in a composition and administered to a host individually or collectively. Alternatively, these components may be provided in a single composition for administration to a host. Administration to a host may be performed via viral vectors known to the skilled person or described herein for delivery to a host (e.g., lentiviral vector, adenoviral vector, AAV vector). As explained herein, use of different selection markers (e.g., for lentiviral sgRNA selection) and concentration of gRNA (e.g., dependent on whether multiple gRNAs are used) may be advantageous for eliciting an improved effect. On the basis of this concept, several variations are appropriate to elicit a genomic locus event, including DNA cleavage, gene activation, or gene deactivation. Using the provided compositions, the person skilled in the art can advantageously and specifically target single or multiple loci with the same or different functional domains to elicit one or more genomic locus events. The compositions may be applied in a wide variety of methods for screening in libraries in cells and functional modeling in vivo (e.g., gene activation of lincRNA and identification of function; gain-of-function modeling; loss-of-function modeling; the use the compositions of the invention to establish cell lines and transgenic animals for optimization and screening purposes).

[00174] The current invention comprehends the use of the compositions of the current invention to establish and utilize conditional or inducible CRISPR transgenic cell /animals; see, e.g., Platt et al., Cell (2014), 159(2): 440-455, or PCT patent publications cited herein, such as WO 2014/093622 (PCT/US2013/074667). For example, cells or animals such as non-human animals, e.g., vertebrates or mammals, such as rodents, e.g., mice, rats, or other laboratory or field animals, e.g., cats, dogs, sheep, etc., may be 'knock-in' whereby the animal conditionally or inducibly expresses Cas13 akin to Platt et al. The target cell or animal thus comprises the CRISPR enzyme (e.g., Cas13) conditionally or inducibly (e.g., in the form of Cre dependent constructs), on expression of a vector introduced into the target cell, the vector expresses that which induces or gives rise to the condition of the CRISPR enzyme (e.g., Cas13) expression in

the target cell. By applying the teaching and compositions as defined herein with the known method of creating a CRISPR complex, inducible genomic events are also an aspect of the current invention. Examples of such inducible events have been described herein elsewhere.

[00175] In some embodiments, phenotypic alteration is preferably the result of genome modification when a genetic disease is targeted, especially in methods of therapy and preferably where a repair template is provided to correct or alter the phenotype.

[00176] In some embodiments diseases that may be targeted include those concerned with disease-causing splice defects.

[00177] In some embodiments, cellular targets include Hemopoietic Stem/Progenitor Cells (CD34+); Human T cells; and Eye (retinal cells) - for example photoreceptor precursor cells.

[00178] In some embodiments Gene targets include: Human Beta Globin - HBB (for treating Sickle Cell Anemia, including by stimulating gene-conversion (using closely related HBD gene as an endogenous template)); CD3 (T-Cells); and CEP920 - retina (eye).

[00179] In some embodiments disease targets also include: cancer; Sickle Cell Anemia (based on a point mutation); HBV, HIV; Beta-Thalassemia; and ophthalmic or ocular disease - for example Leber Congenital Amaurosis (LCA)-causing Splice Defect.

[0100] In some embodiments delivery methods include: Cationic Lipid Mediated "direct" delivery of Enzyme-Guide complex (RiboNucleoProtein) and electroporation of plasmid DNA.

[00180] Methods, products and uses described herein may be used for non-therapeutic purposes. Furthermore, any of the methods described herein may be applied in vitro and ex vivo.

[00181] In an aspect, provided is a non-naturally occurring or engineered composition comprising:

I. two or more CRISPR-Cas system polynucleotide sequences comprising

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

(b) a second guide sequence capable of hybridizing to a second target sequence in a polynucleotide locus,

(c) a direct repeat sequence,

and

II. a Cas13 enzyme or a second polynucleotide sequence encoding it,

wherein when transcribed, the first and the second guide sequences direct sequence-specific binding of a first and a second Cas13 CRISPR complex to the first and second target sequences respectively,

wherein the first CRISPR complex comprises the Cas13 enzyme complexed with the first guide sequence that is hybridizable to the first target sequence,

wherein the second CRISPR complex comprises the Cas13 enzyme complexed with the second guide sequence that is hybridizable to the second target sequence, and

wherein the first guide sequence directs cleavage of one strand of the DNA duplex near the first target sequence and the second guide sequence directs cleavage of the other strand near the second target sequence inducing a double strand break, thereby modifying the organism or the non-human or non-animal organism. Similarly, compositions comprising more than two guide RNAs can be envisaged e.g. each specific for one target, and arranged tandemly in the composition or CRISPR system or complex as described herein.

[00182] In another embodiment, the Cas13 is delivered into the cell as a protein. In another and particularly preferred embodiment, the Cas13 is delivered into the cell as a protein or as a nucleotide sequence encoding it. Delivery to the cell as a protein may include delivery of a Ribonucleoprotein (RNP) complex, where the protein is complexed with the multiple guides.

[00183] In an aspect, host cells and cell lines modified by or comprising the compositions, systems or modified enzymes of present invention are provided, including stem cells, and progeny thereof.

[00184] In an aspect, methods of cellular therapy are provided, where, for example, a single cell or a population of cells is sampled or cultured, wherein that cell or cells is or has been modified ex vivo as described herein, and is then re-introduced (sampled cells) or introduced (cultured cells) into the organism. Stem cells, whether embryonic or induce pluripotent or totipotent stem cells, are also particularly preferred in this regard. But, of course, in vivo embodiments are also envisaged.

[00185] Inventive methods can further comprise delivery of templates, such as repair templates, which may be dsODN or ssODN, see below. Delivery of templates may be via the contemporaneous or separate from delivery of any or all the CRISPR enzyme or guide RNAs and via the same delivery mechanism or different. In some embodiments, it is preferred that the

template is delivered together with the guide RNAs and, preferably, also the CRISPR enzyme. An example may be an AAV vector where the CRISPR enzyme is AsCas or LbCas.

[00186] Inventive methods can further comprise: (a) 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 -(b) 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. Inventive methods can be for the prevention or treatment of disease in an individual, optionally wherein said disease is caused by a defect in said locus of interest. Inventive methods can be 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.

[00187] The invention also comprehends products obtained from using CRISPR enzyme or Cas enzyme or Cas13 enzyme or CRISPR-CRISPR enzyme or CRISPR-Cas system or CRISPR-Cas13 system for use in tandem or multiple targeting as defined herein.

Escorted guides for the Cas13 CRISPR-Cas system according to the invention

[00188] In one aspect the invention provides escorted Cas13 CRISPR-Cas systems or complexes, especially such a system involving an escorted Cas13 CRISPR-Cas system guide. By "escorted" is meant that the Cas13 CRISPR-Cas system or complex or guide is delivered to a selected time or place within a cell, so that activity of the Cas13 CRISPR-Cas system or complex or guide is spatially or temporally controlled. For example, the activity and destination of the Cas13 CRISPR-Cas system or complex or guide may be controlled by an escort RNA aptamer sequence that has binding affinity for an aptamer ligand, such as a cell surface protein or other localized cellular component. Alternatively, the escort aptamer may for example be responsive to an aptamer effector on or in the cell, such as a transient effector, such as an external energy source that is applied to the cell at a particular time.

[00189] The escorted Cas13 CRISPR-Cas systems or complexes have a gRNA with a functional structure designed to improve gRNA structure, architecture, stability, genetic expression, or any combination thereof. Such a structure can include an aptamer.

[00190] Aptamers are biomolecules that can be designed or selected to bind tightly to other ligands, for example using a technique called systematic evolution of ligands by exponential

enrichment (SELEX; Tuerk C, Gold L: "Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase." *Science* 1990, 249:505-510). Nucleic acid aptamers can for example be selected from pools of random-sequence oligonucleotides, with high binding affinities and specificities for a wide range of biomedically relevant targets, suggesting a wide range of therapeutic utilities for aptamers (Keefe, Anthony D., Supriya Pai, and Andrew Ellington. "Aptamers as therapeutics." *Nature Reviews Drug Discovery* 9.7 (2010): 537-550). These characteristics also suggest a wide range of uses for aptamers as drug delivery vehicles (Levy-Nissenbaum, Etgar, et al. "Nanotechnology and aptamers: applications in drug delivery." *Trends in biotechnology* 26.8 (2008): 442-449; and, Hicke BJ, Stephens AW. "Escort aptamers: a delivery service for diagnosis and therapy." *J Clin Invest* 2000, 106:923-928.). Aptamers may also be constructed that function as molecular switches, responding to a cue by changing properties, such as RNA aptamers that bind fluorophores to mimic the activity of green fluorescent protein (Paige, Jeremy S., Karen Y. Wu, and Samie R. Jaffrey. "RNA mimics of green fluorescent protein." *Science* 333.6042 (2011): 642-646). It has also been suggested that aptamers may be used as components of targeted siRNA therapeutic delivery systems, for example targeting cell surface proteins (Zhou, Jiehua, and John J. Rossi. "Aptamer-targeted cell-specific RNA interference." *Silence* 1.1 (2010): 4).

[00191] Accordingly, provided herein is a gRNA modified, e.g., by one or more aptamer(s) designed to improve gRNA delivery, including delivery across the cellular membrane, to intracellular compartments, or into the nucleus. Such a structure can include, either in addition to the one or more aptamer(s) or without such one or more aptamer(s), moiety(ies) so as to render the guide deliverable, inducible or responsive to a selected effector. The invention accordingly comprehends an gRNA that responds to normal or pathological physiological conditions, including without limitation pH, hypoxia, O₂ concentration, temperature, protein concentration, enzymatic concentration, lipid structure, light exposure, mechanical disruption (e.g. ultrasound waves), magnetic fields, electric fields, or electromagnetic radiation.

[00192] An aspect of the invention provides non-naturally occurring or engineered composition comprising an escorted guide RNA (egRNA) comprising:

an RNA guide sequence capable of hybridizing to a target sequence in a genomic locus of interest in a cell; and,

an escort RNA aptamer sequence, wherein the escort aptamer has binding affinity for an aptamer ligand on or in the cell, or the escort aptamer is responsive to a localized aptamer effector on or in the cell, wherein the presence of the aptamer ligand or effector on or in the cell is spatially or temporally restricted.

[00193] The escort aptamer may for example change conformation in response to an interaction with the aptamer ligand or effector in the cell.

[00194] The escort aptamer may have specific binding affinity for the aptamer ligand.

[00195] The aptamer ligand may be localized in a location or compartment of the cell, for example on or in a membrane of the cell. Binding of the escort aptamer to the aptamer ligand may accordingly direct the egRNA to a location of interest in the cell, such as the interior of the cell by way of binding to an aptamer ligand that is a cell surface ligand. In this way, a variety of spatially restricted locations within the cell may be targeted, such as the cell nucleus or mitochondria.

[00196] Once intended alterations have been introduced, such as by editing intended copies of a gene in the genome of a cell, continued CRISPR/Cas13 expression in that cell is no longer necessary. Indeed, sustained expression would be undesirable in certain casein case of off-target effects at unintended genomic sites, etc. Thus time-limited expression would be useful. Inducible expression offers one approach, but in addition Applicants have engineered a Self-Inactivating Cas13 CRISPR-Cas system that relies on the use of a non-coding guide target sequence within the CRISPR vector itself. Thus, after expression begins, the CRISPR system will lead to its own destruction, but before destruction is complete it will have time to edit the genomic copies of the target gene (which, with a normal point mutation in a diploid cell, requires at most two edits). Simply, the self inactivating Cas13 CRISPR-Cas system includes additional RNA (i.e., guide RNA) that targets the coding sequence for the CRISPR enzyme itself or that targets one or more non-coding guide target sequences complementary to unique sequences present in one or more of the following: (a) within the promoter driving expression of the non-coding RNA elements, (b) within the promoter driving expression of the Cas13 gene, (c) within 100bp of the ATG translational start codon in the Cas13 coding sequence, (d) within the inverted terminal repeat (iTR) of a viral delivery vector, e.g., in an AAV genome.

[00197] The egRNA may include an RNA aptamer linking sequence, operably linking the escort RNA sequence to the RNA guide sequence.

[00198] In embodiments, the egRNA may include one or more photolabile bonds or non-naturally occurring residues.

[00199] In one aspect, the escort RNA aptamer sequence may be complementary to a target miRNA, which may or may not be present within a cell, so that only when the target miRNA is present is there binding of the escort RNA aptamer sequence to the target miRNA which results in cleavage of the egRNA by an RNA-induced silencing complex (RISC) within the cell.

[00200] In embodiments, the escort RNA aptamer sequence may for example be from 10 to 200 nucleotides in length, and the egRNA may include more than one escort RNA aptamer sequence.

[00201] It is to be understood that any of the RNA guide sequences as described herein elsewhere can be used in the egRNA described herein. In certain embodiments of the invention, the guide RNA or mature crRNA comprises, consists essentially of, or consists of a direct repeat sequence and a guide sequence or spacer sequence. In certain embodiments, the guide RNA or mature crRNA comprises, consists essentially of, or consists of a direct repeat sequence linked to a guide sequence or spacer sequence. In certain embodiments the guide RNA or mature crRNA comprises 19 nts of partial direct repeat followed by 23-25 nt of guide sequence or spacer sequence. In certain embodiments, the effector protein is a FnCas13 effector protein and requires at least 16 nt of guide sequence to achieve detectable DNA cleavage and a minimum of 17 nt of guide sequence to achieve efficient DNA cleavage in vitro. In certain embodiments, the direct repeat sequence is located upstream (i.e., 5') from the guide sequence or spacer sequence. In a preferred embodiment the seed sequence (i.e. the sequence essential critical for recognition and/or hybridization to the sequence at the target locus) of the FnCas13 guide RNA is approximately within the first 5 nt on the 5' end of the guide sequence or spacer sequence.

[00202] The egRNA may be included in a non-naturally occurring or engineered Cas13 CRISPR-Cas complex composition, together with a Cas13 which may include at least one mutation, for example a mutation so that the Cas13 has no more than 5% of the nuclease activity of a Cas13 not having the at least one mutation, for example having a diminished nuclease activity of at least 97%, or 100% as compared with the Cas13 not having the at least one

mutation. The Cas13 may also include one or more nuclear localization sequences. Mutated Cas13 enzymes having modulated activity such as diminished nuclease activity are described herein elsewhere.

[00203] The engineered Cas13 CRISPR-Cas composition may be provided in a cell, such as a eukaryotic cell, a mammalian cell, or a human cell.

[00204] In embodiments, the compositions described herein comprise a Cas13 CRISPR-Cas complex having at least three functional domains, at least one of which is associated with Cas13 and at least two of which are associated with egRNA.

[00205] The compositions described herein may be used to introduce a genomic locus event in a host cell, such as a eukaryotic cell, in particular a mammalian cell, or a non-human eukaryote, in particular a non-human mammal such as a mouse, in vivo. The genomic locus event may comprise affecting gene activation, gene inhibition, or cleavage in a locus. The compositions described herein may also be used to modify a genomic locus of interest to change gene expression in a cell. Methods of introducing a genomic locus event in a host cell using the Cas13 enzyme provided herein are described herein in detail elsewhere. Delivery of the composition may for example be by way of delivery of a nucleic acid molecule(s) coding for the composition, which nucleic acid molecule(s) is operatively linked to regulatory sequence(s), and expression of the nucleic acid molecule(s) in vivo, for example by way of a lentivirus, an adenovirus, or an AAV.

[00206] The present invention provides compositions and methods by which gRNA-mediated gene editing activity can be adapted. The invention provides gRNA secondary structures that improve cutting efficiency by increasing gRNA and/or increasing the amount of RNA delivered into the cell. The gRNA may include light labile or inducible nucleotides.

[00207] To increase the effectiveness of gRNA, for example gRNA delivered with viral or non-viral technologies, Applicants added secondary structures into the gRNA that enhance its stability and improve gene editing. Separately, to overcome the lack of effective delivery, Applicants modified gRNAs with cell penetrating RNA aptamers; the aptamers bind to cell surface receptors and promote the entry of gRNAs into cells. Notably, the cell-penetrating aptamers can be designed to target specific cell receptors, in order to mediate cell-specific delivery. Applicants also have created guides that are inducible.

[00208] Light responsiveness of an inducible system may be achieved via the activation and binding of cryptochrome-2 and CIB1. Blue light stimulation induces an activating conformational change in cryptochrome-2, resulting in recruitment of its binding partner CIB1. This binding is fast and reversible, achieving saturation in <15 sec following pulsed stimulation and returning to baseline <15 min after the end of stimulation. These rapid binding kinetics result in a system temporally bound only by the speed of transcription/translation and transcript/protein degradation, rather than uptake and clearance of inducing agents. Cryptochrome-2 activation is also highly sensitive, allowing for the use of low light intensity stimulation and mitigating the risks of phototoxicity. Further, in a context such as the intact mammalian brain, variable light intensity may be used to control the size of a stimulated region, allowing for greater precision than vector delivery alone may offer.

[00209] The invention contemplates energy sources such as electromagnetic radiation, sound energy or thermal energy to induce the guide. Advantageously, the electromagnetic radiation is a component of visible light. In a preferred embodiment, the light is a blue light with a wavelength of about 450 to about 495 nm. In an especially preferred embodiment, the wavelength is about 488 nm. In another preferred embodiment, the light stimulation is via pulses. The light power may range from about 0-9 mW/cm². In a preferred embodiment, a stimulation paradigm of as low as 0.25 sec every 15 sec should result in maximal activation.

[00210] Cells involved in the practice of the present invention may be a prokaryotic cell or a eukaryotic cell, advantageously an animal cell a plant cell or a yeast cell, more advantageously a mammalian cell.

[00211] The chemical or energy sensitive guide may undergo a conformational change upon induction by the binding of a chemical source or by the energy allowing it act as a guide and have the Cas13 CRISPR-Cas system or complex function. The invention can involve applying the chemical source or energy so as to have the guide function and the Cas13 CRISPR-Cas system or complex function; and optionally further determining that the expression of the genomic locus is altered.

[00212] There are several different designs of this chemical inducible system: 1. ABI-PYL based system inducible by Abscisc Acid (ABA) (see, e.g., <http://stke.sciencemag.org/cgi/content/abstract/sigtrans;4/164/rs2>), 2. FKBP-FRB based system

inducible by rapamycin (or related chemicals based on rapamycin) (see, e.g., <http://www.nature.com/nmeth/journal/v2/n6/full/nmeth763.html>), 3. GID1-GAI based system inducible by Gibberellin (GA) (see, e.g., <http://www.nature.com/nchembio/journal/v8/n5/full/nchembio.922.html>).

[00213] Another system contemplated by the present invention is a chemical inducible system based on change in sub-cellular localization. Applicants also developed a system in which the polypeptide include a DNA binding domain comprising at least five or more Transcription activator-like effector (TALE) monomers and at least one or more half-monomers specifically ordered to target the genomic locus of interest linked to at least one or more effector domains are further linker to a chemical or energy sensitive protein. This protein will lead to a change in the sub-cellular localization of the entire polypeptide (i.e. transportation of the entire polypeptide from cytoplasm into the nucleus of the cells) upon the binding of a chemical or energy transfer to the chemical or energy sensitive protein. This transportation of the entire polypeptide from one sub-cellular compartments or organelles, in which its activity is sequestered due to lack of substrate for the effector domain, into another one in which the substrate is present would allow the entire polypeptide to come in contact with its desired substrate (i.e. genomic DNA in the mammalian nucleus) and result in activation or repression of target gene expression.

[00214] This type of system could also be used to induce the cleavage of a genomic locus of interest in a cell when the effector domain is a nuclease.

[00215] A chemical inducible system can be an estrogen receptor (ER) based system inducible by 4-hydroxytamoxifen (4OHT) (see, e.g., <http://www.pnas.org/content/104/3/1027.abstract>). A mutated ligand-binding domain of the estrogen receptor called ERT2 translocates into the nucleus of cells upon binding of 4-hydroxytamoxifen. In further embodiments of the invention any naturally occurring or engineered derivative of any nuclear receptor, thyroid hormone receptor, retinoic acid receptor, estrogen receptor, estrogen-related receptor, glucocorticoid receptor, progesterone receptor, androgen receptor may be used in inducible systems analogous to the ER based inducible system.

[00216] Another inducible system is based on the design using Transient receptor potential (TRP) ion channel based system inducible by energy, heat or radio-wave (see, e.g., <http://www.sciencemag.org/content/336/6081/604>). These TRP family proteins respond to

different stimuli, including light and heat. When this protein is activated by light or heat, the ion channel will open and allow the entering of ions such as calcium into the plasma membrane. This influx of ions will bind to intracellular ion interacting partners linked to a polypeptide including the guide and the other components of the Cas13 CRISPR-Cas complex or system, and the binding will induce the change of sub-cellular localization of the polypeptide, leading to the entire polypeptide entering the nucleus of cells. Once inside the nucleus, the guide protein and the other components of the Cas13 CRISPR-Cas complex will be active and modulating target gene expression in cells.

[00217] This type of system could also be used to induce the cleavage of a genomic locus of interest in a cell; and, in this regard, it is noted that the Cas13 enzyme is a nuclease. The light could be generated with a laser or other forms of energy sources. The heat could be generated by raise of temperature results from an energy source, or from nano-particles that release heat after absorbing energy from an energy source delivered in the form of radio-wave.

[00218] While light activation may be an advantageous embodiment, sometimes it may be disadvantageous especially for in vivo applications in which the light may not penetrate the skin or other organs. In this instance, other methods of energy activation are contemplated, in particular, electric field energy and/or ultrasound which have a similar effect.

[00219] Electric field energy is preferably administered substantially as described in the art, using one or more electric pulses of from about 1 Volt/cm to about 10 kVolts/cm under in vivo conditions. Instead of or in addition to the pulses, the electric field may be delivered in a continuous manner. The electric pulse may be applied for between 1 μ s and 500 milliseconds, preferably between 1 μ s and 100 milliseconds. The electric field may be applied continuously or in a pulsed manner for 5 about minutes.

[00220] As used herein, 'electric field energy' is the electrical energy to which a cell is exposed. Preferably the electric field has a strength of from about 1 Volt/cm to about 10 kVolts/cm or more under in vivo conditions (see WO97/49450).

[00221] As used herein, the term "electric field" includes one or more pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave and/or modulated square wave forms. References to electric fields and electricity should be taken to include reference the presence of an electric potential difference in the environment of a cell.

Such an environment may be set up by way of static electricity, alternating current (AC), direct current (DC), etc, as known in the art. The electric field may be uniform, non-uniform or otherwise, and may vary in strength and/or direction in a time dependent manner.

[00222] Single or multiple applications of electric field, as well as single or multiple applications of ultrasound are also possible, in any order and in any combination. The ultrasound and/or the electric field may be delivered as single or multiple continuous applications, or as pulses (pulsatile delivery).

[00223] Electroporation has been used in both *in vitro* and *in vivo* procedures to introduce foreign material into living cells. With *in vitro* applications, a sample of live cells is first mixed with the agent of interest and placed between electrodes such as parallel plates. Then, the electrodes apply an electrical field to the cell/implant mixture. Examples of systems that perform *in vitro* electroporation include the Electro Cell Manipulator ECM600 product, and the Electro Square Porator T820, both made by the BTX Division of Genetronics, Inc (see U.S. Pat. No 5,869,326).

[00224] The known electroporation techniques (both *in vitro* and *in vivo*) function by applying a brief high voltage pulse to electrodes positioned around the treatment region. The electric field generated between the electrodes causes the cell membranes to temporarily become porous, whereupon molecules of the agent of interest enter the cells. In known electroporation applications, this electric field comprises a single square wave pulse on the order of 1000 V/cm, of about 100 μ s duration. Such a pulse may be generated, for example, in known applications of the Electro Square Porator T820.

[00225] Preferably, the electric field has a strength of from about 1 V/cm to about 10 kV/cm under *in vitro* conditions. Thus, the electric field may have a strength of 1 V/cm, 2 V/cm, 3 V/cm, 4 V/cm, 5 V/cm, 6 V/cm, 7 V/cm, 8 V/cm, 9 V/cm, 10 V/cm, 20 V/cm, 50 V/cm, 100 V/cm, 200 V/cm, 300 V/cm, 400 V/cm, 500 V/cm, 600 V/cm, 700 V/cm, 800 V/cm, 900 V/cm, 1 kV/cm, 2 kV/cm, 5 kV/cm, 10 kV/cm, 20 kV/cm, 50 kV/cm or more. More preferably from about 0.5 kV/cm to about 4.0 kV/cm under *in vitro* conditions. Preferably the electric field has a strength of from about 1 V/cm to about 10 kV/cm under *in vivo* conditions. However, the electric field strengths may be lowered where the number of pulses delivered to the target site are increased. Thus, pulsatile delivery of electric fields at lower field strengths is envisaged.

[00226] Preferably the application of the electric field is in the form of multiple pulses such as double pulses of the same strength and capacitance or sequential pulses of varying strength and/or capacitance. As used herein, the term "pulse" includes one or more electric pulses at variable capacitance and voltage and including exponential and/or square wave and/or modulated wave/square wave forms.

[00227] Preferably the electric pulse is delivered as a waveform selected from an exponential wave form, a square wave form, a modulated wave form and a modulated square wave form.

[00228] A preferred embodiment employs direct current at low voltage. Thus, Applicants disclose the use of an electric field which is applied to the cell, tissue or tissue mass at a field strength of between 1V/cm and 20V/cm, for a period of 100 milliseconds or more, preferably 15 minutes or more.

[00229] Ultrasound is advantageously administered at a power level of from about 0.05 W/cm² to about 100 W/cm². Diagnostic or therapeutic ultrasound may be used, or combinations thereof.

[00230] As used herein, the term "ultrasound" refers to a form of energy which consists of mechanical vibrations the frequencies of which are so high they are above the range of human hearing. Lower frequency limit of the ultrasonic spectrum may generally be taken as about 20 kHz. Most diagnostic applications of ultrasound employ frequencies in the range 1 and 15 MHz' (From Ultrasonics in Clinical Diagnosis, P. N. T. Wells, ed., 2nd. Edition, Publ. Churchill Livingstone [Edinburgh, London & NY, 1977]).

[00231] Ultrasound has been used in both diagnostic and therapeutic applications. When used as a diagnostic tool ("diagnostic ultrasound"), ultrasound is typically used in an energy density range of up to about 100 mW/cm² (FDA recommendation), although energy densities of up to 750 mW/cm² have been used. In physiotherapy, ultrasound is typically used as an energy source in a range up to about 3 to 4 W/cm² (WHO recommendation). In other therapeutic applications, higher intensities of ultrasound may be employed, for example, HIFU at 100 W/cm up to 1 kW/cm² (or even higher) for short periods of time. The term "ultrasound" as used in this specification is intended to encompass diagnostic, therapeutic and focused ultrasound.

[00232] Focused ultrasound (FUS) allows thermal energy to be delivered without an invasive probe (see Morocz et al 1998 Journal of Magnetic Resonance Imaging Vol.8, No. 1, pp.136-142.

Another form of focused ultrasound is high intensity focused ultrasound (HIFU) which is reviewed by Moussatov et al in *Ultrasonics* (1998) Vol.36, No.8, pp.893-900 and TranHuuHue et al in *Acustica* (1997) Vol.83, No.6, pp.1103-1106.

[00233] Preferably, a combination of diagnostic ultrasound and a therapeutic ultrasound is employed. This combination is not intended to be limiting, however, and the skilled reader will appreciate that any variety of combinations of ultrasound may be used. Additionally, the energy density, frequency of ultrasound, and period of exposure may be varied.

[00234] Preferably the exposure to an ultrasound energy source is at a power density of from about 0.05 to about 100 Wcm⁻². Even more preferably, the exposure to an ultrasound energy source is at a power density of from about 1 to about 15 Wcm⁻².

[00235] Preferably the exposure to an ultrasound energy source is at a frequency of from about 0.015 to about 10.0 MHz. More preferably the exposure to an ultrasound energy source is at a frequency of from about 0.02 to about 5.0 MHz or about 6.0 MHz. Most preferably, the ultrasound is applied at a frequency of 3 MHz.

[00236] Preferably the exposure is for periods of from about 10 milliseconds to about 60 minutes. Preferably the exposure is for periods of from about 1 second to about 5 minutes. More preferably, the ultrasound is applied for about 2 minutes. Depending on the particular target cell to be disrupted, however, the exposure may be for a longer duration, for example, for 15 minutes.

[00237] Advantageously, the target tissue is exposed to an ultrasound energy source at an acoustic power density of from about 0.05 Wcm⁻² to about 10 Wcm⁻² with a frequency ranging from about 0.015 to about 10 MHz (see WO 98/52609). However, alternatives are also possible, for example, exposure to an ultrasound energy source at an acoustic power density of above 100 Wcm⁻², but for reduced periods of time, for example, 1000 Wcm⁻² for periods in the millisecond range or less.

[00238] Preferably the application of the ultrasound is in the form of multiple pulses; thus, both continuous wave and pulsed wave (pulsatile delivery of ultrasound) may be employed in any combination. For example, continuous wave ultrasound may be applied, followed by pulsed wave ultrasound, or vice versa. This may be repeated any number of times, in any order and

combination. The pulsed wave ultrasound may be applied against a background of continuous wave ultrasound, and any number of pulses may be used in any number of groups.

[00239] Preferably, the ultrasound may comprise pulsed wave ultrasound. In a highly preferred embodiment, the ultrasound is applied at a power density of 0.7 Wcm⁻² or 1.25 Wcm⁻² as a continuous wave. Higher power densities may be employed if pulsed wave ultrasound is used.

[00240] Use of ultrasound is advantageous as, like light, it may be focused accurately on a target. Moreover, ultrasound is advantageous as it may be focused more deeply into tissues unlike light. It is therefore better suited to whole-tissue penetration (such as but not limited to a lobe of the liver) or whole organ (such as but not limited to the entire liver or an entire muscle, such as the heart) therapy. Another important advantage is that ultrasound is a non-invasive stimulus which is used in a wide variety of diagnostic and therapeutic applications. By way of example, ultrasound is well known in medical imaging techniques and, additionally, in orthopedic therapy. Furthermore, instruments suitable for the application of ultrasound to a subject vertebrate are widely available and their use is well known in the art.

[00241] The rapid transcriptional response and endogenous targeting of the instant invention make for an ideal system for the study of transcriptional dynamics. For example, the instant invention may be used to study the dynamics of variant production upon induced expression of a target gene. On the other end of the transcription cycle, mRNA degradation studies are often performed in response to a strong extracellular stimulus, causing expression level changes in a plethora of genes. The instant invention may be utilized to reversibly induce transcription of an endogenous target, after which point stimulation may be stopped and the degradation kinetics of the unique target may be tracked.

[00242] The temporal precision of the instant invention may provide the power to time genetic regulation in concert with experimental interventions. For example, targets with suspected involvement in long-term potentiation (LTP) may be modulated in organotypic or dissociated neuronal cultures, but only during stimulus to induce LTP, so as to avoid interfering with the normal development of the cells. Similarly, in cellular models exhibiting disease phenotypes, targets suspected to be involved in the effectiveness of a particular therapy may be modulated only during treatment. Conversely, genetic targets may be modulated only during a pathological

stimulus. Any number of experiments in which timing of genetic cues to external experimental stimuli is of relevance may potentially benefit from the utility of the instant invention.

[00243] The *in vivo* context offers equally rich opportunities for the instant invention to control gene expression. Photoinducibility provides the potential for spatial precision. Taking advantage of the development of optrode technology, a stimulating fiber optic lead may be placed in a precise brain region. Stimulation region size may then be tuned by light intensity. This may be done in conjunction with the delivery of the Cas13 CRISPR-Cas system or complex of the invention, or, in the case of transgenic Cas13 animals, guide RNA of the invention may be delivered and the optrode technology can allow for the modulation of gene expression in precise brain regions. A transparent Cas13 expressing organism, can have guide RNA of the invention administered to it and then there can be extremely precise laser induced local gene expression changes.

[00244] A culture medium for culturing host cells includes a medium commonly used for tissue culture, such as M199-earle base, Eagle MEM (E-MEM), Dulbecco MEM (DMEM), SC-UCM102, UP-SFM (GIBCO BRL), EX-CELL302 (Nichirei), EX-CELL293-S (Nichirei), TFBM-01 (Nichirei), ASF104, among others. Suitable culture media for specific cell types may be found at the American Type Culture Collection (ATCC) or the European Collection of Cell Cultures (ECACC). Culture media may be supplemented with amino acids such as L-glutamine, salts, anti-fungal or anti-bacterial agents such as Fungizone, penicillin-streptomycin, animal serum, and the like. The cell culture medium may optionally be serum-free.

[00245] The invention may also offer valuable temporal precision *in vivo*. The invention may be used to alter gene expression during a particular stage of development. The invention may be used to time a genetic cue to a particular experimental window. For example, genes implicated in learning may be overexpressed or repressed only during the learning stimulus in a precise region of the intact rodent or primate brain. Further, the invention may be used to induce gene expression changes only during particular stages of disease development. For example, an oncogene may be overexpressed only once a tumor reaches a particular size or metastatic stage. Conversely, proteins suspected in the development of Alzheimer's may be knocked down only at defined time points in the animal's life and within a particular brain region. Although these

examples do not exhaustively list the potential applications of the invention, they highlight some of the areas in which the invention may be a powerful technology.

Protected guides: Enzymes according to the invention can be used in combination with protected guide RNAs

[00246] In one aspect, an object of the current invention is to further enhance the specificity of Cas13 given individual guide RNAs through thermodynamic tuning of the binding specificity of the guide RNA to target DNA. This is a general approach of introducing mismatches, elongation or truncation of the guide sequence to increase / decrease the number of complimentary bases vs. mismatched bases shared between a genomic target and its potential off-target loci, in order to give thermodynamic advantage to targeted genomic loci over genomic off-targets.

[00247] In one aspect, the invention provides for the guide sequence being modified by secondary structure to increase the specificity of the Cas13 CRISPR-Cas system and whereby the secondary structure can protect against exonuclease activity and allow for 3' additions to the guide sequence.

[00248] In one aspect, the invention provides for hybridizing a “protector RNA” to a sequence of the guide molecule, wherein the “protector RNA” is an RNA strand complementary to the 3' end of the guide molecule to thereby generate a partially double-stranded guide RNA. In an embodiment of the invention, protecting mismatched bases (i.e. the bases of the guide molecule which do not form part of the guide sequence) with a perfectly complementary protector sequence decreases the likelihood of target RNA binding to the mismatched basepairs at the 3' end. In particular embodiments of the invention, additional sequences comprising an extended length may also be present within the guide molecule such that the guide comprises a protector sequence within the guide molecule. This “protector sequence” ensures that the guide molecule comprises a “protected sequence” in addition to an “exposed sequence” (comprising the part of the guide sequence hybridizing to the target sequence). In particular embodiments, the guide molecule is modified by the presence of the protector guide to comprise a secondary structure such as a hairpin. Advantageously there are three or four to thirty or more, e.g., about 10 or more, contiguous base pairs having complementarity to the protected sequence, the guide sequence or both. It is advantageous that the protected portion does not impede thermodynamics of the CRISPR-Cas system interacting with its target. By providing such an extension including

a partially double stranded guide molecule, the guide molecule is considered protected and results in improved specific binding of the CRISPR-Cas complex, while maintaining specific activity.

[00249] In particular embodiments, use is made of a truncated guide (tru-guide), i.e. a guide molecule which comprises a guide sequence which is truncated in length with respect to the canonical guide sequence length. As described by Nowak et al. (Nucleic Acids Res (2016) 44 (20): 9555-9564), such guides may allow catalytically active CRISPR-Cas enzyme to bind its target without cleaving the target RNA. In particular embodiments, a truncated guide is used which allows the binding of the target but retains only nickase activity of the CRISPR-Cas enzyme.

[00250] In some embodiments, the guide molecules disclosed herein may be used for preparing a library for screening drugable targets and/or identifying lead compounds. For example, the guide molecules may be used for manipulating the expression and/or the activities of components in one or more selected pathways in cells. The impacts of the manipulation on the effects of one or more small molecule compounds (e.g., a small molecule library) may be assessed. The assessments may be used for screening potential targets for the small molecule compounds and/or identifying pathways or component(s) therein for enhancing the effects or activities of the small molecule compounds.

CRISPR-Cas Enzymes

[00251] The embodiments disclosed here provide guide sequences suitable for use with CRISPR-Cas systems. In certain example embodiments, the methods disclosed herein may be used to design guide sequences for use with Type II, Type V, or Type VI CRISPR-Cas system.

[00252] The present invention may be further illustrated and extended based on aspects of CRISPR-Cas development and use as set forth in the following articles and particularly as relates to delivery of a CRISPR protein complex and uses of an RNA guided endonuclease in cells and organisms:

- Multiplex genome engineering using CRISPR-Cas systems. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., & Zhang, F. Science Feb 15;339(6121):819-23 (2013);

- RNA-guided editing of bacterial genomes using CRISPR-Cas systems. Jiang W., Bikard D., Cox D., Zhang F, Marraffini LA. *Nat Biotechnol* Mar;31(3):233-9 (2013);
- One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR-Cas-Mediated Genome Engineering. Wang H., Yang H., Shivalila CS., Dawlaty MM., Cheng AW., Zhang F., Jaenisch R. *Cell* May 9;153(4):910-8 (2013);
- Optical control of mammalian endogenous transcription and epigenetic states. Konermann S, Brigham MD, Trevino AE, Hsu PD, Heidenreich M, Cong L, Platt RJ, Scott DA, Church GM, Zhang F. *Nature*. Aug 22;500(7463):472-6. doi: 10.1038/Nature12466. Epub 2013 Aug 23 (2013);
- Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. Ran, FA., Hsu, PD., Lin, CY., Gootenberg, JS., Konermann, S., Trevino, AE., Scott, DA., Inoue, A., Matoba, S., Zhang, Y., & Zhang, F. *Cell* Aug 28. pii: S0092-8674(13)01015-5 (2013-A);
- DNA targeting specificity of RNA-guided Cas9 nucleases. Hsu, P., Scott, D., Weinstein, J., Ran, FA., Konermann, S., Agarwala, V., Li, Y., Fine, E., Wu, X., Shalem, O., Cradick, TJ., Marraffini, LA., Bao, G., & Zhang, F. *Nat Biotechnol* doi:10.1038/nbt.2647 (2013);
- Genome engineering using the CRISPR-Cas9 system. Ran, FA., Hsu, PD., Wright, J., Agarwala, V., Scott, DA., Zhang, F. *Nature Protocols* Nov;8(11):2281-308 (2013-B);
- Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. Shalem, O., Sanjana, NE., Hartenian, E., Shi, X., Scott, DA., Mikkelsen, T., Heckl, D., Ebert, BL., Root, DE., Doench, JG., Zhang, F. *Science* Dec 12. (2013);
- Crystal structure of cas9 in complex with guide RNA and target DNA. Nishimasu, H., Ran, FA., Hsu, PD., Konermann, S., Shehata, SI., Dohmae, N., Ishitani, R., Zhang, F., Nureki, O. *Cell* Feb 27, 156(5):935-49 (2014);
- Genome-wide binding of the CRISPR endonuclease Cas9 in mammalian cells. Wu X., Scott DA., Kriz AJ., Chiu AC., Hsu PD., Dadon DB., Cheng AW., Trevino AE., Konermann S., Chen S., Jaenisch R., Zhang F., Sharp PA. *Nat Biotechnol*. Apr 20. doi: 10.1038/nbt.2889 (2014);
- CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM,

- Jovanovic M, Graham DB, Jhunjhunwala S, Heidenreich M, Xavier RJ, Langer R, Anderson DG, Hacohen N, Regev A, Feng G, Sharp PA, Zhang F. Cell 159(2): 440-455 DOI: 10.1016/j.cell.2014.09.014(2014);
- Development and Applications of CRISPR-Cas9 for Genome Engineering, Hsu PD, Lander ES, Zhang F., Cell. Jun 5;157(6):1262-78 (2014).
 - Genetic screens in human cells using the CRISPR-Cas9 system, Wang T, Wei JJ, Sabatini DM, Lander ES., Science. January 3; 343(6166): 80-84. doi:10.1126/science.1246981 (2014);
 - Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation, Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE., (published online 3 September 2014) Nat Biotechnol. Dec;32(12):1262-7 (2014);
 - *In vivo* interrogation of gene function in the mammalian brain using CRISPR-Cas9, Swiech L, Heidenreich M, Banerjee A, Habib N, Li Y, Trombetta J, Sur M, Zhang F., (published online 19 October 2014) Nat Biotechnol. Jan;33(1):102-6 (2015);
 - Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex, Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F., Nature. Jan 29;517(7536):583-8 (2015).
 - A split-Cas9 architecture for inducible genome editing and transcription modulation, Zetsche B, Volz SE, Zhang F., (published online 02 February 2015) Nat Biotechnol. Feb;33(2):139-42 (2015);
 - Genome-wide CRISPR Screen in a Mouse Model of Tumor Growth and Metastasis, Chen S, Sanjana NE, Zheng K, Shalem O, Lee K, Shi X, Scott DA, Song J, Pan JQ, Weissleder R, Lee H, Zhang F, Sharp PA. Cell 160, 1246-1260, March 12, 2015 (multiplex screen in mouse), and
 - *In vivo* genome editing using *Staphylococcus aureus* Cas9, Ran FA, Cong L, Yan WX, Scott DA, Gootenberg JS, Kriz AJ, Zetsche B, Shalem O, Wu X, Makarova KS, Koonin EV, Sharp PA, Zhang F., (published online 01 April 2015), Nature. Apr 9;520(7546):186-91 (2015).

- Shalem et al., “High-throughput functional genomics using CRISPR-Cas9,” *Nature Reviews Genetics* 16, 299-311 (May 2015).
- Xu et al., “Sequence determinants of improved CRISPR sgRNA design,” *Genome Research* 25, 1147-1157 (August 2015).
- Parnas et al., “A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks,” *Cell* 162, 675-686 (July 30, 2015).
- Ramanan et al., CRISPR-Cas9 cleavage of viral DNA efficiently suppresses hepatitis B virus,” *Scientific Reports* 5:10833. doi: 10.1038/srep10833 (June 2, 2015)
- Nishimasu et al., Crystal Structure of *Staphylococcus aureus* Cas9,” *Cell* 162, 1113-1126 (Aug. 27, 2015)
- BCL11A enhancer dissection by Cas9-mediated in situ saturating mutagenesis, Canver et al., *Nature* 527(7577):192-7 (Nov. 12, 2015) doi: 10.1038/nature15521. Epub 2015 Sep 16.
- *Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System*, Zetsche et al., *Cell* 163, 759-771 (Sep 25, 2015).
- *Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems*, Shmakov et al., *Molecular Cell*, 60(3), 385–397 doi: 10.1016/j.molcel.2015.10.008 Epub October 22, 2015.
- *Rationally engineered Cas9 nucleases with improved specificity*, Slaymaker et al., *Science* 2016 Jan 1 351(6268): 84-88 doi: 10.1126/science.aad5227. Epub 2015 Dec 1.
- Gao *et al.*, “Engineered Cpf1 Enzymes with Altered PAM Specificities,” bioRxiv 091611; doi: <http://dx.doi.org/10.1101/091611> (Dec. 4, 2016).

each of which is incorporated herein by reference, may be considered in the practice of the instant invention, and discussed briefly below:

- Cong *et al.* engineered type II CRISPR-Cas systems for use in eukaryotic cells based on both *Streptococcus thermophilus* Cas9 and also *Streptococcus pyogenes* Cas9 and demonstrated that Cas9 nucleases can be directed by short RNAs to induce precise cleavage of DNA in human and mouse cells. Their study further showed that Cas9 as converted into a nicking enzyme can be used to facilitate homology-directed repair in eukaryotic cells with minimal mutagenic activity. Additionally, their study demonstrated

that multiple guide sequences can be encoded into a single CRISPR array to enable simultaneous editing of several at endogenous genomic loci sites within the mammalian genome, demonstrating easy programmability and wide applicability of the RNA-guided nuclease technology. This ability to use RNA to program sequence specific DNA cleavage in cells defined a new class of genome engineering tools. These studies further showed that other CRISPR loci are likely to be transplantable into mammalian cells and can also mediate mammalian genome cleavage. Importantly, it can be envisaged that several aspects of the CRISPR-Cas system can be further improved to increase its efficiency and versatility.

- Jiang *et al.* used the clustered, regularly interspaced, short palindromic repeats (CRISPR)–associated Cas9 endonuclease complexed with dual-RNAs to introduce precise mutations in the genomes of *Streptococcus pneumoniae* and *Escherichia coli*. The approach relied on dual-RNA:Cas9-directed cleavage at the targeted genomic site to kill unmutated cells and circumvents the need for selectable markers or counter-selection systems. The study reported reprogramming dual-RNA:Cas9 specificity by changing the sequence of short CRISPR RNA (crRNA) to make single- and multinucleotide changes carried on editing templates. The study showed that simultaneous use of two crRNAs enabled multiplex mutagenesis. Furthermore, when the approach was used in combination with recombineering, in *S. pneumoniae*, nearly 100% of cells that were recovered using the described approach contained the desired mutation, and in *E. coli*, 65% that were recovered contained the mutation.
- Wang *et al.* (2013) used the CRISPR-Cas system for the one-step generation of mice carrying mutations in multiple genes which were traditionally generated in multiple steps by sequential recombination in embryonic stem cells and/or time-consuming intercrossing of mice with a single mutation. The CRISPR-Cas system will greatly accelerate the *in vivo* study of functionally redundant genes and of epistatic gene interactions.
- Konermann *et al.* (2013) addressed the need in the art for versatile and robust technologies that enable optical and chemical modulation of DNA-binding domains based CRISPR Cas9 enzyme and also Transcriptional Activator Like Effectors

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

with target design, gene modifications can be achieved within as little as 1–2 weeks, and modified clonal cell lines can be derived within 2–3 weeks.

- Shalem *et al.* described a new way to interrogate gene function on a genome-wide scale. Their studies showed that delivery of a genome-scale CRISPR-Cas9 knockout (GeCKO) library targeted 18,080 genes with 64,751 unique guide sequences enabled both negative and positive selection screening in human cells. First, the authors showed use of the GeCKO library to identify genes essential for cell viability in cancer and pluripotent stem cells. Next, in a melanoma model, the authors screened for genes whose loss is involved in resistance to vemurafenib, a therapeutic that inhibits mutant protein kinase BRAF. Their studies showed that the highest-ranking candidates included previously validated genes NF1 and MED12 as well as novel hits NF2, CUL3, TADA2B, and 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 Å resolution. The structure revealed a bilobed architecture composed of target recognition and nuclease lobes, accommodating the sgRNA:DNA heteroduplex in a positively charged groove at their interface. Whereas the recognition lobe is essential for binding sgRNA and DNA, the nuclease lobe contains the HNH and RuvC nuclease domains, which are properly positioned for cleavage of the complementary and non-complementary strands of the target DNA, respectively. The nuclease lobe also contains a carboxyl-terminal domain responsible for the interaction with the protospacer adjacent motif (PAM). This high-resolution structure and accompanying functional analyses have revealed the molecular mechanism of RNA-guided DNA targeting by Cas9, thus paving the way for the rational design of new, versatile genome-editing technologies.
- 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.

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

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

- Canver *et al.* (2015) demonstrated a CRISPR-Cas9-based functional investigation of non-coding genomic elements. The authors we developed pooled CRISPR-Cas9 guide RNA libraries to perform *in situ* saturating mutagenesis of the human and mouse BCL11A enhancers which revealed critical features of the enhancers.
- Zetsche et al. (2015) reported characterization of Cpf1, a class 2 CRISPR nuclease from *Francisella novicida* U112 having features distinct from Cas9. Cpf1 is a single RNA-guided endonuclease lacking tracrRNA, utilizes a T-rich protospacer-adjacent motif, and cleaves DNA via a staggered DNA double-stranded break.
- Shmakov et al. (2015) reported three distinct Class 2 CRISPR-Cas systems. Two system CRISPR enzymes (C2c1 and C2c3) contain RuvC-like endonuclease domains distantly related to Cpf1. Unlike Cpf1, C2c1 depends on both crRNA and tracrRNA for DNA cleavage. The third enzyme (C2c2) contains two predicted HEPN RNase domains and is tracrRNA independent.
- Slaymaker et al (2016) reported the use of structure-guided protein engineering to improve the specificity of *Streptococcus pyogenes* Cas9 (SpCas9). The authors developed "enhanced specificity" SpCas9 (eSpCas9) variants which maintained robust on-target cleavage with reduced off-target effects.

[00253] The methods and tools provided herein may be designed for use with or Cas13, a type II nuclease that does not make use of tracrRNA. Orthologs of Cas13 have been identified in different bacterial species as described herein. Further type II nucleases with similar properties can be identified using methods described in the art (Shmakov et al. 2015, 60:385–397; Abudayeh et al. 2016, *Science*, 5:353(6299)). In particular embodiments, such methods for identifying novel CRISPR effector proteins may comprise the steps of selecting sequences from the database encoding a seed which identifies the presence of a CRISPR Cas locus, identifying loci located within 10 kb of the seed comprising Open Reading Frames (ORFs) in the selected sequences, selecting therefrom loci comprising ORFs of which only a single ORF encodes a novel CRISPR effector having greater than 700 amino acids and no more than 90% homology to a known CRISPR effector. In particular embodiments, the seed is a protein that is common to the CRISPR-Cas system, such as Cas1. In further embodiments, the CRISPR array is used as a seed to identify new effector proteins.

[00254] Also, “Dimeric CRISPR RNA-guided FokI nucleases for highly specific genome editing”, Shengdar Q. Tsai, Nicolas Wyvekens, Cyd Khayter, Jennifer A. Foden, Vishal Thapar, Deepak Reyon, Mathew J. Goodwin, Martin J. Aryee, J. Keith Joung Nature Biotechnology 32(6): 569-77 (2014), relates to dimeric RNA-guided FokI Nucleases that recognize extended sequences and can edit endogenous genes with high efficiencies in human cells.

[00255] With respect to general information on CRISPR/Cas Systems, components thereof, and delivery of such components, including methods, materials, delivery vehicles, vectors, particles, and making and using thereof, including as to amounts and formulations, as well as CRISPR-Cas-expressing eukaryotic cells, CRISPR-Cas expressing eukaryotes, such as a mouse, reference is made to: US Patents Nos. 8,999,641, 8,993,233, 8,697,359, 8,771,945, 8,795,965, 8,865,406, 8,871,445, 8,889,356, 8,889,418, 8,895,308, 8,906,616, 8,932,814, and 8,945,839; US Patent Publications US 2014-0310830 (US App. Ser. No. 14/105,031), US 2014-0287938 A1 (U.S. App. Ser. No. 14/213,991), US 2014-0273234 A1 (U.S. App. Ser. No. 14/293,674), US2014-0273232 A1 (U.S. App. Ser. No. 14/290,575), US 2014-0273231 (U.S. App. Ser. No. 14/259,420), US 2014-0256046 A1 (U.S. App. Ser. No. 14/226,274), US 2014-0248702 A1 (U.S. App. Ser. No. 14/258,458), US 2014-0242700 A1 (U.S. App. Ser. No. 14/222,930), US 2014-0242699 A1 (U.S. App. Ser. No. 14/183,512), US 2014-0242664 A1 (U.S. App. Ser. No. 14/104,990), US 2014-0234972 A1 (U.S. App. Ser. No. 14/183,471), US 2014-0227787 A1 (U.S. App. Ser. No. 14/256,912), US 2014-0189896 A1 (U.S. App. Ser. No. 14/105,035), US 2014-0186958 (U.S. App. Ser. No. 14/105,017), US 2014-0186919 A1 (U.S. App. Ser. No. 14/104,977), US 2014-0186843 A1 (U.S. App. Ser. No. 14/104,900), US 2014-0179770 A1 (U.S. App. Ser. No. 14/104,837) and US 2014-0179006 A1 (U.S. App. Ser. No. 14/183,486), US 2014-0170753 (US App Ser No 14/183,429); US 2015-0184139 (U.S. App. Ser. No. 14/324,960); 14/054,414 European Patent Applications EP 2 771 468 (EP13818570.7), EP 2 764 103 (EP13824232.6), and EP 2 784 162 (EP14170383.5); and PCT Patent Publications WO2014/093661 (PCT/US2013/074743), WO2014/093694 (PCT/US2013/074790), WO2014/093595 (PCT/US2013/074611), WO2014/093718 (PCT/US2013/074825), WO2014/093709 (PCT/US2013/074812), WO2014/093622 (PCT/US2013/074667), WO2014/093635 (PCT/US2013/074691), WO2014/093655 (PCT/US2013/074736), WO2014/093712 (PCT/US2013/074819), WO2014/093701 (PCT/US2013/074800),

WO2014/018423 (PCT/US2013/051418), WO2014/204723 (PCT/US2014/041790),
 WO2014/204724 (PCT/US2014/041800), WO2014/204725 (PCT/US2014/041803),
 WO2014/204726 (PCT/US2014/041804), WO2014/204727 (PCT/US2014/041806),
 WO2014/204728 (PCT/US2014/041808), WO2014/204729 (PCT/US2014/041809),
 WO2015/089351 (PCT/US2014/069897), WO2015/089354 (PCT/US2014/069902),
 WO2015/089364 (PCT/US2014/069925), WO2015/089427 (PCT/US2014/070068),
 WO2015/089462 (PCT/US2014/070127), WO2015/089419 (PCT/US2014/070057),
 WO2015/089465 (PCT/US2014/070135), WO2015/089486 (PCT/US2014/070175),
 WO2015/058052 (PCT/US2014/061077), WO2015/070083 (PCT/US2014/064663),
 WO2015/089354 (PCT/US2014/069902), WO2015/089351 (PCT/US2014/069897),
 WO2015/089364 (PCT/US2014/069925), WO2015/089427 (PCT/US2014/070068),
 WO2015/089473 (PCT/US2014/070152), WO2015/089486 (PCT/US2014/070175),
 WO2016/049258 (PCT/US2015/051830), WO2016/094867 (PCT/US2015/065385),
 WO2016/094872 (PCT/US2015/065393), WO2016/094874 (PCT/US2015/065396),
 WO2016/106244 (PCT/US2015/067177).

[00256] Mention is also made of US application 62/180,709, 17-Jun-15, PROTECTED GUIDE RNAS (PGRNAS); US application 62/091,455, filed, 12-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US application 62/096,708, 24-Dec-14, PROTECTED GUIDE RNAS (PGRNAS); US applications 62/091,462, 12-Dec-14, 62/096,324, 23-Dec-14, 62/180,681, 17-Jun-2015, and 62/237,496, 5-Oct-2015, DEAD GUIDES FOR CRISPR TRANSCRIPTION FACTORS; US application 62/091,456, 12-Dec-14 and 62/180,692, 17-Jun-2015, ESCORTED AND FUNCTIONALIZED GUIDES FOR CRISPR-CAS SYSTEMS; US application 62/091,461, 12-Dec-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR GENOME EDITING AS TO HEMATOPOETIC STEM CELLS (HSCs); US application 62/094,903, 19-Dec-14, UNBIASED IDENTIFICATION OF DOUBLE-STRAND BREAKS AND GENOMIC REARRANGEMENT BY GENOME-WISE INSERT CAPTURE SEQUENCING; US application 62/096,761, 24-Dec-14, ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED ENZYME AND GUIDE SCAFFOLDS FOR SEQUENCE MANIPULATION; US application 62/098,059, 30-Dec-14, 62/181,641, 18-Jun-2015, and 62/181,667, 18-Jun-2015, RNA-TARGETING SYSTEM; US

application 62/096,656, 24-Dec-14 and 62/181,151, 17-Jun-2015, CRISPR HAVING OR ASSOCIATED WITH DESTABILIZATION DOMAINS; US application 62/096,697, 24-Dec-14, CRISPR HAVING OR ASSOCIATED WITH AAV; US application 62/098,158, 30-Dec-14, ENGINEERED CRISPR COMPLEX INSERTIONAL TARGETING SYSTEMS; US application 62/151,052, 22-Apr-15, CELLULAR TARGETING FOR EXTRACELLULAR EXOSOMAL REPORTING; US application 62/054,490, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING PARTICLE DELIVERY COMPONENTS; US application 61/939,154, 12-F EB-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,484, 25-Sep-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,537, 4-Dec-14, SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/054,651, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US application 62/067,886, 23-Oct-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR MODELING COMPETITION OF MULTIPLE CANCER MUTATIONS IN VIVO; US applications 62/054,675, 24-Sep-14 and 62/181,002, 17-Jun-2015, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN NEURONAL CELLS/TISSUES; US application 62/054,528, 24-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS IN IMMUNE DISEASES OR DISORDERS; US application 62/055,454, 25-Sep-14, DELIVERY, USE AND THERAPEUTIC APPLICATIONS OF THE CRISPR-CAS SYSTEMS AND COMPOSITIONS FOR TARGETING DISORDERS AND DISEASES USING CELL PENETRATION PEPTIDES (CPP); US application 62/055,460, 25-Sep-14, MULTIFUNCTIONAL-CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; US application 62/087,475, 4-Dec-14 and 62/181,690,

18-Jun-2015, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/055,487, 25-Sep-14, FUNCTIONAL SCREENING WITH OPTIMIZED FUNCTIONAL CRISPR-CAS SYSTEMS; US application 62/087,546, 4-Dec-14 and 62/181,687, 18-Jun-2015, MULTIFUNCTIONAL CRISPR COMPLEXES AND/OR OPTIMIZED ENZYME LINKED FUNCTIONAL-CRISPR COMPLEXES; and US application 62/098,285, 30-Dec-14, CRISPR MEDIATED IN VIVO MODELING AND GENETIC SCREENING OF TUMOR GROWTH AND METASTASIS.

[00257] Mention is made of US applications 62/181,659, 18-Jun-2015 and 62/207,318, 19-Aug-2015, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS, ENZYME AND GUIDE SCAFFOLDS OF CAS9 ORTHOLOGS AND VARIANTS FOR SEQUENCE MANIPULATION. Mention is made of US applications 62/181,663, 18-Jun-2015 and 62/245,264, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS, US applications 62/181,675, 18-Jun-2015, 62/285,349, 22-Oct-2015, 62/296,522, 17-Feb-2016, and 62/320,231, 8-Apr-2016, NOVEL CRISPR ENZYMES AND SYSTEMS, US application 62/232,067, 24-Sep-2015, US Application 14/975,085, 18-Dec-2015, European application No. 16150428.7, US application 62/205,733, 16-Aug-2015, US application 62/201,542, 5-Aug-2015, US application 62/193,507, 16-Jul-2015, and US application 62/181,739, 18-Jun-2015, each entitled NOVEL CRISPR ENZYMES AND SYSTEMS and of US application 62/245,270, 22-Oct-2015, NOVEL CRISPR ENZYMES AND SYSTEMS. Mention is also made of US application 61/939,256, 12-Feb-2014, and WO 2015/089473 (PCT/US2014/070152), 12-Dec-2014, each entitled ENGINEERING OF SYSTEMS, METHODS AND OPTIMIZED GUIDE COMPOSITIONS WITH NEW ARCHITECTURES FOR SEQUENCE MANIPULATION. Mention is also made of PCT/US2015/045504, 15-Aug-2015, US application 62/180,699, 17-Jun-2015, and US application 62/038,358, 17-Aug-2014, each entitled GENOME EDITING USING CAS9 NICKASES.

[00258] Each of these patents, patent publications, and applications, and all documents cited therein or during their prosecution (“appln cited documents”) and all documents cited or referenced in the appln cited documents, together with any instructions, descriptions, product specifications, and product sheets for any products mentioned therein or in any document therein and incorporated by reference herein, are hereby incorporated herein by reference, and may be

employed in the practice of the invention. All documents (e.g., these patents, patent publications and applications and the appln cited documents) are incorporated herein by reference to the same extent as if each individual document was specifically and individually indicated to be incorporated by reference.

[00259] A description of methods for designing guide sequences for such systems follows.

Example Guide Sequence Library Generation Architectures

[00260] FIG. 1 is a block diagram depicting a guide sequence library generation system 110 for generating a library of guide sequences. As depicted in FIG. 1, the operating environment 100 includes network devices 105 and 110 that are configured to communicate with one another via one or more networks 130. In some embodiments, the user device or genomic sequence input device 105 may provide nucleic acid sequence data directly from user input, or from sequencing data, such as obtained from sequencing. In other example embodiments, the genomic sequence input device may be obtained from a remote server comprising said sequence information (not shown). In some embodiments, a user associated with a device must install an application 106 and/or make a feature selection to obtain the benefits of the techniques described herein. The guide sequence library generation system 110 receives sequence data and outputs a set of identified ranked guide sequence. The ranking of individual guide sequence reflects the likelihood of a given guide sequence being able to recognize, hybridize or bind to, and/or alter expression of a particular gene or exon in an organism such as a human or a mouse.

[00261] Each network device 105 and 110 includes a device having a communication module capable of transmitting and receiving data over the network 130. For example, each network device 105 and 110 can include a server, desktop computer, laptop computer, tablet computer, a television with one or more processors embedded therein and/or coupled thereto, smart phone, handheld computer, personal digital assistant (“PDA”), or any other wired or wireless, processor-driven device.

[00262] The genomic sequence input device 105 may generate nucleic acid sequence data files comprising information on the coding regions or exons, or genes, within a given biological sample. In one example embodiment, the genomic sequence input device 105 may directly communicate the data file to the guide sequence library generation system 110 across the network 130 and the guide sequence library generation and ranking is conducted in line with the

sequence input and/or analysis. In another example embodiment, the sequence data file may be stored on a data storage medium and later uploaded to the guide sequence library generation system 110 for further analysis.

[00263] The guide sequence library generation system 110 may comprise an input module 110, an exon prediction module 115, a ranking module 120, and a graphical user interface (GUI) module 125. The input module 110 receives input data from genomic sequence input device 105 and formats such data for further processing. The exon prediction module 115 takes the genomic input information and identifies exon sequences in order to identify an initial set of target regions. The ranking module 120 takes the identified target regions and generates a set of ranked guide sequences for each target region. The output module 125 then formats and displays this information to an end user. In certain example embodiments the out module 125 may be configured through GUI 130 to allow direct user interaction with guide library, for example by selecting a final set of guides or modifying certain input parameters to further refine the final guide sequence library produced. The guide sequence generation system 110 may further optionally comprise a guide sequence index 135 where guide sequence libraries are stored during and after guide sequence library production.

[00264] It will be appreciated that the network connections shown are examples and other means of establishing a communications link between the computers and devices can be used. Moreover, those having ordinary skill in the art having the benefit of the present disclosure will appreciate that the guide Library Generation System 110, can have any of several other suitable computer system configurations.

Example Processes

[00265] The method illustrated in FIG. 2 is described hereinafter with respect to the components of the example operating environment 100. The example method of FIG. 2 may also be performed with other systems and in other environments.

[00266] FIG. 2 is a block flow diagram depicting accordance with certain example embodiments. Method 200 begins at block 205, where the input module 110 receives gene input information for one or more target genes from the user device 105. The gene input information may be in the form of one or more target gene sequences, for example in fasta format. In other example embodiments, the input gene information may be in the form of a data file containing,

for example, gene symbols, Ensemble IDs, Entrez IDs or any other suitable identifier by which the guide sequence generation system 110 may lookup and obtain gene sequence information from a third-party database or other repository of sequence information. In certain other example embodiments, the input module 110 may provide via GUI 130 an interface, such as a web-based interface, where an end user may enter or upload the above information. The method then proceeds to block 210

[00267] At block 210, the exon prediction module 115 identifies exons in the one or more target genes. In certain example embodiments, the exonic region of a gene is identified by taking a union of Consensus CoDing Sequences (CCDS) regions associated with the gene. For example, the CCDS coordinates may be used by the exon prediction module 115 to lookup sequence data for each gene's coding exons (initiating ATG codon to stop codon). Upon initial provisioning, the exon prediction module 115 may discard all non-exonic genomic sequences and save the sequences encoding each exon in a separate file, for example by storage in index 135 or other suitable permanent or temporary data structure. Then, the exon module 115 may iterate through each CDS exon and determine all targetable sites. For example, in the context of a CRISPR effector system, the exon module 115 may determine all targetable sites by identifying the corresponding PAM sequence for a given CRISPR effector protein on either strand.

[00268] In certain example embodiments, the exon module 115 may then further select potential target regions by prioritizing those that have higher expression. For example, in the context of the human gene, the exon module 115 may use tissue-specific RNA-sequencing gene expression data. Example tissue-specific RNA-sequencing gene expression data may be obtained from the GTEx Consortium (GTEx Consortium. Human Genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648-660 (2015)) to select target regions with higher expression. The user may be presented with an option at input, or at this step of the process, to specify specific tissues or computer average expression over a set of tissues (or all tissues), which results in increased targeting to expressed exons. For GTEx data, the exon module 115 may use exon_reads files, which contain RNA-sequencing expression values (as read counts): GTEx_Analysis_v6_RNA-seq_RNA-SeQCv1.1.8_exon_reads_txt.gz. The exon module 115 may group individual GTEx data sample (e.g. patient-tissue pairs) by tissue and determine the median gene expression value for each

exon in each tissue, for example computed over all samples with the same GTEx SMTS tissue type. The median count reads are then normalized by exon size (base pairs) and then saved into a gene-specific file. For example, the gene-specific file may contain rows that denote different exons (by Ensemble GENE ID and exon number) and columns that denote different tissues. In certain example embodiments the following specific gene expression profiles from the following GTEx dataset tissues may be supported: adipose tissue, adrenal gland, bladder, blood, blood vessel, bone marrow, brain, breast, cervix uteri, colon, esophagus, fallopian tube, heart, kidney, liver, lung, muscle, nerve, ovary, pancreas, pituitary, prostate, salivary gland, skin, small intestine, spleen, stomach, testis, thyroid, uterus, and vagina. Exon expression values are determined based on the user-selected subset of tissues. For each exon, the expression value provided is the median of the exon's expression across selected tissue samples. The method then proceeds to block 215. Incorporation of tissue-specific RNA-sequencing gene expression data increased average expression of targeted exons by a factor of 1.5. See FIG. 4

[00269] At block 215, the ranking module 120 ranks each identified potential target region based at least in part by an off-target avoidance score, an on-target efficiency score, and optionally, by the presence of a protein domain in the target region. For example, potential target sites may be prioritized using a cutting frequency determination (CFD) score (Doench et al. Nat. Biotechnol. 34,184-191, 2016)). For each target site, the ranking module 120 identifies all sequences with 1, 2, or 3 base mismatches present in the exome and assigns an off-target score by taking the sum of the CFD scores for each potential off-target found. Target sites with perfect matches elsewhere in the exome are given a maximal off-target score and will be selected only in cases where not other guide sequences exist to target the gene. By using exome-wide CFD scoring as described herein, design of a library with approximately 2,000 genes, the percentage of designed guide sequences with predicted off-targets decreased from approximately 43% to approximately 4%. See FIG. 5. Targeting guides RNAs to exons constitutively expressed in the target cell type/tissue can be important, as mutations in alternatively-spliced exons may not result in protein knock-out (Murphy et al. Mol. Cell Biol. 35, 1860-1870, 2015).

[00270] In certain example embodiment an aggregate off-target score (G score) is determined for all guide sequences in the exome. The G-score may be computed as follows. The G-score utilizes the CFD scoring algorithm (Doench et al.), which computes the likelihood of a guide

sequence cutting at a particular off-target site based on experimental data from approximately 10,000 guide sequence mismatches, insertions and deletions. For a given guide sequence, the ranking module 120 finds all potential off-targets with up to three mismatches in the exosome (followed by a NGG motif) and calculates the CFD between the given guide sequence and potential off-target. The sum of these scores may then be linearly weighted by the number of times the potential off-target occurs in the exome and the results is returned as the G-score for the given guide sequence.

[00271] Mathematically, the G-score of guide sequence j may be defined as:

$$\text{G-Score}(j) = \sum_{i=1}^N c_i \cdot \text{CFD}(i, j)$$

where N is the number of potential off-targets with up to three mismatches, i , ranges over these mismatches and c_i is the number of times the mismatch occurs in the exome. For figures with the G-score, a small value (10^{-8}) was added to each G-score to avoid $\log(0)$.

[00272] Although indel mutations are introduced after CRISPR effector cutting and genome repair, not all indel mutations abolish protein expression, presumably due to in-frame mutations in regions tolerant to mutations. Saturation mutagenesis screens tiling over entire genes have shown increased knock-out efficiency when targeting protein functional domains (Shi et al. Nat. Biotechnol. 33, 661-667, 2015). To take advantage of this, an option may be presented to the user to preferentially choose guide sequences that target functional protein domains. The functional protein domains may be identified, for example, in the Protein Family (Pfam) database (Finn et al. Nucleic Acids Res. 44, D279-285, 2016). This may have a significant impact on library design since 90% of protein-coding genes in the human genome contain at least one Pfam-annotated domain (Mistry et al. Database J. Biol. Databases Curations 2013, bat023).

[00273] The ranking module 120 further ranks the potential guide sequences based on an on-target specificity analysis. In certain example embodiments, a boosted regression tree classifier is used to score target sites based on local sequence preferences learned from saturation mutagenesis screens and adds the highest-scoring guide sequences to the library. When targeting the same sets of genes, guide sequences designed with this criterion have an approximately 30% higher on-target efficiency score. See FIG. 6. In certain example embodiment, on-target scores

for all guide sequence in the exome were computed using the Azimuth 2.0 Python package from Microsoft Research (<https://github.com/MicrosoftResearch/Azimuth>) (Doench et al.). For each target site, on-target efficiencies may be computed using gradient-boosted regression trees on the surrounding 30 bp. On-target score may be generated with the `azimuth.model_comparison.predict` function. During initial provisioning, the ranking module 120 computes a table of on-target scores for all guide sequences in the exome.

[0101] Since homopolymeric regions can be difficult to synthesize and sequence accurately (Shi et al. Nat. Biotechnol. 33, 661-667, 2015), in certain example embodiments, the the ranking module 120 may also remove guide sequences containing stretches of 5 or more of the same base (A, T, C or G). Furthermore, in certain example embodiments, the ranking module 120 may exclude any guide sequence with 4 or more sequential T bases which can result in premature termination of Pol III transcription (Bogenhagen et al. Cell, 24, 261-270, 1981).

[0102] The method then proceeds to block 225. At block 225, the output module 125 prepares a final guide library for output to the user. In certain example embodiments, the output module 125 produces full-length, synthesis ready oligonucleotides that flank the guide sequence with overhangs for Gibson cloning into appropriate screening vectors (e.g. lentiCRISPRv2 or lentiGuide-Puro) (Sanjana et al. Nat. Methods 11, 783-784, 2014). These flanking sequences include the end of a U6 primer on the 5' side of the guide sequence and the beginning of the guide sequence scaffold on the 3' side of the guide sequence. In this format, the output can be directly synthesized using known synthesis techniques. In certain example embodiment, the output guide sequences may be based on a full length guide sequence scaffold (with 85 nt tracr RNA) or a modified version. In certain example embodiments, the modified version is an E+F modification with an A to U flip to prevent early Pol III termination and a 5bp (10nt) extension of the first stem loop (Chen et al. Cell 155, 1479-1491, 2013). The flanking sequences that may be used for synthesis ready oligonucleotides and appropriate primers for PCT and Gibson cloning for each scaffold are:

Full-length scaffold	GGAAAGGACGAAACACCGXXXXXXXXXXXXXXXXXXXX
73-nt including guide	XXGTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC
Full-length cloning F	TAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATCT
	TGTGGAAAGGACGAAACACCG
Full-length cloning R	ACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAAC

TTGCTATTTCTAGCTCTAAAAC

E+F modified scaffold	GGAAAGGACGAAACACCGXXXXXXXXXXXXXXXXXXXX
63-nt including <u>guide</u>	<u>XXGTTTAAGAGCTATGCTGGAAACAGC</u>
E+F cloning F	TAACTTGAAAGTATTTTCGATTTCTTGGCTTTATATATCT
	TGTGGAAAGGACGAAACACCG
E+F cloning R	GACTAGCCTTATTTAACTTGCTATGCTGTTTCCAGCA
	TAGCTCTTAAAC

[0103] For selecting guide sequences to return to each user, in certain example embodiments the output module 125 uses a data structure which enable processing in linear instead of linearithmic time. In this data structure, each gene is considered independently. For each gene, the method keeps track of two different list of guide sequences: a primary list and a secondary list. Each list contains all guide sequences from a particular group of exons and each exon can be on only one list at a time. The primary list is unordered, that is the guide sequences from all exons on the list are grouped together. The secondary list is ordered with guide from higher-ranked exons considered before guides from lower ranked exons.

[0104] During initial provisioning, a sorted list of guide sequences may be generated for each exon. The optimized, linear-time data structure keeps track of the top-ranked guide from each exon. During guide selection, the output module 125 selects the highest ranked guide sequence from the primary list. If the primary list is empty or all remaining guides have exact matches elsewhere in the exome, then the highest-ranked exon from the secondary list is moved to the primary list. This continues until all guide sequences are selected or both lists are empty.

[0105] For a gene containing N exons, when GTEx-based exon selection is disabled, the primary list contains exons 2 to N-2. The secondary list contains exon 1 followed by exon N. The last coding exon is given last priority since mRNA may escape nonsense-mediated decay when mutations are in the last exon (Popp et al. Annu. Rev. Genet. 47, 139-165 , 2013). When GTEx is enabled, the primary list contains the top M exons by RNA expression in the selected tissues (default: M=4) and the secondary list contains the remaining exons with exon 1 and exon N always placed second-to-last and last, respectively.

[0106] In addition to returning the top *i*, guide sequences (where *i* is the number of guide sequences per gene requested by the user), the output module 125 returns the next *j* (default:

$j=10$) guides as “unselected” to the front end to allow the user to further fine-tune the library. During guide selection, real-time updates are provided to the front-end to provide an accurate indication of remaining library generation time.

[0107] After library generation, the output module 125 may prompt the user to also include non-targeting (negative control) guide sequences in the library. By default, the output module 125 suggests adding a pool of non-targeting controls of a size equivalent to 5% of the number of targeting guide sequence with 1,000 gene-targeting guide sequences. For example, for a guide sequence library with 1000 gene-targeting guide sequences, the output module 125 will suggest adding 50 additional non-targeting guide sequences.

Other Example Embodiments

[0108] FIG. 3 depicts a computing machine 2000 and a module 2050 in accordance with certain example embodiments. The computing machine 2000 may correspond to any of the various computers, servers, mobile devices, embedded systems, or computing systems presented herein. The module 2050 may comprise one or more hardware or software elements configured to facilitate the computing machine 2000 in performing the various methods and processing functions presented herein. The computing machine 2000 may include various internal or attached components such as a processor 2010, system bus 2020, system memory 2030, storage media 2040, input/output interface 2060, and a network interface 2070 for communicating with a network 2080.

[0109] The computing machine 2000 may be implemented as a conventional computer system, an embedded controller, a laptop, a server, a mobile device, a smartphone, a set-top box, a kiosk, a router or other network node, a vehicular information system, one more processors associated with a television, a customized machine, any other hardware platform, or any combination or multiplicity thereof. The computing machine 2000 may be a distributed system configured to function using multiple computing machines interconnected via a data network or bus system.

[0110] The processor 2010 may be configured to execute code or instructions to perform the operations and functionality described herein, manage request flow and address mappings, and to perform calculations and generate commands. The processor 2010 may be configured to monitor and control the operation of the components in the computing machine 2000. The processor

2010 may be a general purpose processor, a processor core, a multiprocessor, a reconfigurable processor, a microcontroller, a digital signal processor (“DSP”), an application specific integrated circuit (“ASIC”), a graphics processing unit (“GPU”), a field programmable gate array (“FPGA”), a programmable logic device (“PLD”), a controller, a state machine, gated logic, discrete hardware components, any other processing unit, or any combination or multiplicity thereof. The processor 2010 may be a single processing unit, multiple processing units, a single processing core, multiple processing cores, special purpose processing cores, co-processors, or any combination thereof. According to certain embodiments, the processor 2010 along with other components of the computing machine 2000 may be a virtualized computing machine executing within one or more other computing machines.

[0111] The system memory 2030 may include non-volatile memories such as read-only memory (“ROM”), programmable read-only memory (“PROM”), erasable programmable read-only memory (“EPROM”), flash memory, or any other device capable of storing program instructions or data with or without applied power. The system memory 2030 may also include volatile memories such as random access memory (“RAM”), static random access memory (“SRAM”), dynamic random access memory (“DRAM”), and synchronous dynamic random access memory (“SDRAM”). Other types of RAM also may be used to implement the system memory 2030. The system memory 2030 may be implemented using a single memory module or multiple memory modules. While the system memory 2030 is depicted as being part of the computing machine 2000, one skilled in the art will recognize that the system memory 2030 may be separate from the computing machine 2000 without departing from the scope of the subject technology. It should also be appreciated that the system memory 2030 may include, or operate in conjunction with, a non-volatile storage device such as the storage media 2040.

[0112] The storage media 2040 may include a hard disk, a floppy disk, a compact disc read only memory (“CD-ROM”), a digital versatile disc (“DVD”), a Blu-ray disc, a magnetic tape, a flash memory, other non-volatile memory device, a solid state drive (“SSD”), any magnetic storage device, any optical storage device, any electrical storage device, any semiconductor storage device, any physical-based storage device, any other data storage device, or any combination or multiplicity thereof. The storage media 2040 may store one or more operating systems, application programs and program modules such as module 2050, data, or any other

information. The storage media 2040 may be part of, or connected to, the computing machine 2000. The storage media 2040 may also be part of one or more other computing machines that are in communication with the computing machine 2000 such as servers, database servers, cloud storage, network attached storage, and so forth.

[0113] The module 2050 may comprise one or more hardware or software elements configured to facilitate the computing machine 2000 with performing the various methods and processing functions presented herein. The module 2050 may include one or more sequences of instructions stored as software or firmware in association with the system memory 2030, the storage media 2040, or both. The storage media 2040 may therefore represent examples of machine or computer readable media on which instructions or code may be stored for execution by the processor 2010. Machine or computer readable media may generally refer to any medium or media used to provide instructions to the processor 2010. Such machine or computer readable media associated with the module 2050 may comprise a computer software product. It should be appreciated that a computer software product comprising the module 2050 may also be associated with one or more processes or methods for delivering the module 2050 to the computing machine 2000 via the network 2080, any signal-bearing medium, or any other communication or delivery technology. The module 2050 may also comprise hardware circuits or information for configuring hardware circuits such as microcode or configuration information for an FPGA or other PLD.

[0114] The input/output (“I/O”) interface 2060 may be configured to couple to one or more external devices, to receive data from the one or more external devices, and to send data to the one or more external devices. Such external devices along with the various internal devices may also be known as peripheral devices. The I/O interface 2060 may include both electrical and physical connections for operably coupling the various peripheral devices to the computing machine 2000 or the processor 2010. The I/O interface 2060 may be configured to communicate data, addresses, and control signals between the peripheral devices, the computing machine 2000, or the processor 2010. The I/O interface 2060 may be configured to implement any standard interface, such as small computer system interface (“SCSI”), serial-attached SCSI (“SAS”), fiber channel, peripheral component interconnect (“PCI”), PCI express (PCIe), serial bus, parallel bus, advanced technology attached (“ATA”), serial ATA (“SATA”), universal serial

bus (“USB”), Thunderbolt, FireWire, various video buses, and the like. The I/O interface 2060 may be configured to implement only one interface or bus technology. Alternatively, the I/O interface 2060 may be configured to implement multiple interfaces or bus technologies. The I/O interface 2060 may be configured as part of, all of, or to operate in conjunction with, the system bus 2020. The I/O interface 2060 may include one or more buffers for buffering transmissions between one or more external devices, internal devices, the computing machine 2000, or the processor 2010.

[0115] The I/O interface 2060 may couple the computing machine 2000 to various input devices including mice, touch-screens, scanners, biometric readers, electronic digitizers, sensors, receivers, touchpads, trackballs, cameras, microphones, keyboards, any other pointing devices, or any combinations thereof. The I/O interface 2060 may couple the computing machine 2000 to various output devices including video displays, speakers, printers, projectors, tactile feedback devices, automation control, robotic components, actuators, motors, fans, solenoids, valves, pumps, transmitters, signal emitters, lights, and so forth.

[0116] The computing machine 2000 may operate in a networked environment using logical connections through the network interface 2070 to one or more other systems or computing machines across the network 2080. The network 2080 may include wide area networks (WAN), local area networks (LAN), intranets, the Internet, wireless access networks, wired networks, mobile networks, telephone networks, optical networks, or combinations thereof. The network 2080 may be packet switched, circuit switched, of any topology, and may use any communication protocol. Communication links within the network 2080 may involve various digital or an analog communication media such as fiber optic cables, free-space optics, waveguides, electrical conductors, wireless links, antennas, radio-frequency communications, and so forth.

[0117] The processor 2010 may be connected to the other elements of the computing machine 2000 or the various peripherals discussed herein through the system bus 2020. It should be appreciated that the system bus 2020 may be within the processor 2010, outside the processor 2010, or both. According to some embodiments, any of the processor 2010, the other elements of the computing machine 2000, or the various peripherals discussed herein may be integrated

into a single device such as a system on chip (“SOC”), system on package (“SOP”), or ASIC device.

[0118] Embodiments may comprise a computer program that embodies the functions described and illustrated herein, wherein the computer program is implemented in a computer system that comprises instructions stored in a machine-readable medium and a processor that executes the instructions. However, it should be apparent that there could be many different ways of implementing embodiments in computer programming, and the embodiments should not be construed as limited to any one set of computer program instructions. Further, a skilled programmer would be able to write such a computer program to implement an embodiment of the disclosed embodiments based on the appended flow charts and associated description in the application text. Therefore, disclosure of a particular set of program code instructions is not considered necessary for an adequate understanding of how to make and use embodiments. Further, those skilled in the art will appreciate that one or more aspects of embodiments described herein may be performed by hardware, software, or a combination thereof, as may be embodied in one or more computing systems. Moreover, any reference to an act being performed by a computer should not be construed as being performed by a single computer as more than one computer may perform the act.

[0119] The example embodiments described herein can be used with computer hardware and software that perform the methods and processing functions described herein. The systems, methods, and procedures described herein can be embodied in a programmable computer, computer-executable software, or digital circuitry. The software can be stored on computer-readable media. For example, computer-readable media can include a floppy disk, RAM, ROM, hard disk, removable media, flash memory, memory stick, optical media, magneto-optical media, CD-ROM, etc. Digital circuitry can include integrated circuits, gate arrays, building block logic, field programmable gate arrays (FPGA), etc.

[0120] The example systems, methods, and acts described in the embodiments presented previously are illustrative, and, in alternative embodiments, certain acts can be performed in a different order, in parallel with one another, omitted entirely, and/or combined between different example embodiments, and/or certain additional acts can be performed, without departing from

the scope and spirit of various embodiments. Accordingly, such alternative embodiments are included in the invention claimed herein.

Guide Sequence Libraries

[0121] In another aspect, the invention is directed to guide libraries comprising the guide sequences generated using the methods disclosed herein. In some embodiments, the invention provides a guide library comprising guide sequences to one or more target regions in one or more exons of one or more target genes, wherein individual guide sequences in the library are included based on optimization of an off-target avoidance score and an on-target efficiency score, and optionally, by the presence of a protein domain in the target region. A guide library as described herein may be made using the methods as described herein and set forth in the Examples, and in at least FIG. 2. Exons may be selected for identification of guide sequences and/or a guide library as described herein based on expression in a subject of interest, for example, a human or mouse subject. Custom libraries of the invention may be produced to target specific gene sets (e.g., kinases, transcription factors, chromatin modifiers, the druggable genome) with higher coverage for these specific genes. For example, exons may be selected or identified based on tissue-specific expression, organ-specific expression, cell-type-specific expression, or any particular exon profiles for a particular application.

[0122] . Some guide libraries may include a plurality of guide sequences as defined herein, or may include a combination of such sequences. As would be understood by one of skill in the art, a guide library as described herein may encompass any number of such sequences, and may be of any composition or size appropriate for the particular application. In accordance with the invention, a guide sequence may be RNA, DNA, a DNA-RNA hybrid, a chemically modified RNA, a chemically modified DNA, a chemically modified DNA-RNA hybrid, or a combination of all of these.

[0123] As used herein, sgRNA refers to single-guide RNA, a nucleic acid molecule that combines the tracrRNA and crRNA, which are separate molecules in the native CRISPR/Cas9 system into a single RNA construct.

[0124] By “libraries” is meant a plurality. General methods for producing libraries as described herein are known in the art. In a preferred embodiment, the guide sequence libraries provided herein comprise any number of sequences, such as from about 50 to 100 individual

guide sequences or to between about 103 and about 109 sequences, or more. A library as described herein may be contained within eukaryotic cells or prokaryotic cells, and may encompass viruses, such as Lentivirus, BACs, YACs, artificial chromosomes, or any suitable vehicle appropriate for the particular application. A library may be contained within any appropriate means as described herein.

[0125] The initial nucleic acid or genomic material from which a guide sequence library is to be constructed may be mRNA, DNA, cDNA, or any type of nucleic acid. The starting material may be isolated or taken from a variety of different samples, where the sample will typically be derived from a physiological source. The physiological source may be derived from a variety of eukaryotic and/or prokaryotic sources. In addition, viral RNA may be used to serve as template for a guide library as described herein.

[0126] For example, a tissue type that may be analyzed with the present invention may include, but is not limited to, adipose tissue, adrenal gland, bladder, blood, blood vessel, bone marrow, brain, breast, cervix uteri, colon, esophagus, fallopian tube, heart, kidney, liver, lung, muscle, nerve, ovary, pancreas, pituitary, prostate, salivary gland, skin, small intestine, spleen, stomach, testis, thyroid, uterus, and/or vagina. Particular data for selection of exons may be on the basis of higher expression or lower expression, or may be based on a lack of expression when compared to a control population.

[0127] Screening of libraries is well-known in the art, and may be performed using any appropriate methods, including, but not limited to, electroporation, PCR, sequencing, such as Next-Generation Sequencing (NGS), electrophoresis, or any other methodology known and/or available in the art. Screening of a library as described herein may require large quantities of cells and therefore, may require cell culture techniques depending on the particular application. Viral vectors are often used in library construction as described herein, and therefore, techniques of which are known in the art and discussed in the Examples.

[0128] The nucleotide sequence of the gene or exon to be targeted will dictate the sequence of the guide sequence. Embodiments of functional guide sequences as described herein have sequence complementarity that need not be 100 percent, but is at least sufficient to permit hybridization to a target gene or exon, or a sequence of the target gene or exon, to form a duplex to permit a gene silencing mechanism. Thus, in embodiments, a guide sequence is designed to be

complementary to all or a portion of a desired target gene or exon sequence. For instance, the guide sequence may be essentially identical or essentially complementary to a sequence of, for example, 10 or more contiguous nucleotides in either the target gene or exon sequence or mRNA transcribed from the target gene or exon. By “essentially identical” is meant having 100 percent sequence identity or at least about 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent sequence identity when compared to the sequence of 10 or more contiguous nucleotides in either the target gene or exon, or the RNA transcribed from the target gene or exon. By “essentially complementary” is meant having 100 percent sequence complementarity or at least about 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99 percent sequence complementarity when compared to the sequence of 10 or more contiguous nucleotides in either the target gene or exon, or the RNA transcribed from the target gene or exon. In some embodiments, guide sequences may be designed to have 100 percent sequence identity with or complementarity to one allele or one family member of a given target gene (coding or non-coding sequence of a gene). In other embodiments, a guide sequence is designed to have 100 percent sequence identity with or complementarity to multiple alleles or family members of a given target gene.

[0129] In other embodiments, a guide sequence may have 0 mismatches with its target gene or exon, or the RNA transcribed from the target gene or exon, or may have 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more mismatches, without altering the activity of the guide sequence for its target. One of skill in the art will understand that the specificity of a guide sequence for its target may vary, depending on the particular application or depending on the target sequence

Vectors and Cell Lines Comprising Guide Libraries

[0130] A guide sequence library of the present invention may be amplified, packaged, and transduced into host cells as appropriate to generate cell lines containing individual guide sequences. The cell lines can then be screened under particular conditions depending on the particular application, for example, using positive or negative selection, or others known in the art. In such a way, important genes in a pathway may be identified, down-regulated, or knocked down entirely. In some embodiments, gene up-regulation or activation may be accomplished using the present technology, for example in instances where the target gene or exon pertains to a gene that is normally dormant or repressed, such as a gene involved in cancer biology. In some

embodiments, epigenetic changes can be targeted to a gene via chromatin or histone modifications, DNA methylations, or changes in genome structure or architecture.

[0131] In some embodiments, the invention provides a composition comprising a population of cells and a guide sequence library as described herein, wherein each of the cells contains one or more of the guide sequences and thus the guides sequences of the library are integrated into the population of cells. A population of cells as described herein may be a eukaryotic population of cells. In particular embodiments, eukaryotic cells may be human or mouse cells. Guide sequences as described herein may be contained within a vector, such as a plasmid, a virus, such as a phage, a cosmid, a BAC, a YAC, a P1 artificial chromosome, or any other type of vector or vehicle known and/or available in the art. For example, eukaryotic cells that may be useful in accordance with the invention may be human or mouse cell. In some embodiments, a population of cells may be prokaryotic, such as a population of bacterial cells. As required for the particular application, cellular machinery required for expression of a nucleic acid may be included in a prokaryotic population of cells to the extent expression is necessary for the particular application.

[0132] In some embodiments, the invention provides a vector comprising a guide sequence or sgRNA as described herein. Such a guide sequence may be present in single copy number, or may be present in multiple versions. In certain embodiments, the vector may be an expression vector. In certain embodiments, the vector may be a prokaryotic expression vector, or a eukaryotic expression vector. In certain embodiments, the vector may be an inducible, conditional, or constitutive (prokaryotic or eukaryotic) expression vector, as described herein elsewhere and known in the art.

[0133] In an aspect, the invention relates to a vector system comprising one or more vector, said one or more vector comprising a polynucleotide encoding a guide sequence or guide RNA, or sgRNA as described herein. In certain embodiments, a vector system of the invention may comprise one or more expression vectors, as defined above. A vector system in accordance with the invention may be a plurality of vectors, wherein each vector of the system comprises a guide sequence or sgRNA.

[0134] In certain embodiments, and by means of further guidance, a guide sequence or sgRNA as defined herein may comprise a guide sequence and a direct repeat sequence. In certain embodiments, such a guide sequence or sgRNA may comprise a polynucleotide sequence

comprising a 5' guide sequence and a 3' direct repeat sequence. In certain embodiments, such a guide sequence or sgRNA may be capable of hybridizing to a target nucleic acid sequence. In certain embodiments, such a sequence may be modified to alter functionality, specificity, and/or stability. In some embodiments, and without limitation, a guide sequence or sgRNA may be a dead gRNA, or may be a protected gRNA, or may be an escorted gRNA. By means of further example, and without limitation, said gRNA may be modified by addition of one or more aptamers, and may for instance be useable in a synergistic activator system (SAM).

[0135] In certain preferred embodiments, a guide sequence as described herein may be capable of binding a target sequence as described herein. In certain embodiments, a guide sequence as described herein may be capable of forming a functional complex with a target nucleic acid sequence. Such a guide sequence may further be capable of modifying or targeting a target DNA locus upon binding to said target DNA locus. In certain embodiments, binding of a guide sequence of the invention to its target nucleic acid sequence may affect modification of the target locus. Such complex may be capable of inducing DNA modifications, such as for instance single-stranded or double-stranded DNA cleavage, or may be capable of altering DNA architecture/structure, epigenetic modification, or gene expression.

[0136] In an aspect, the invention relates to a delivery system comprising a guide sequence as described herein. Such a delivery system may also include a polynucleotide, one or more vectors, vector system, or complex according to the invention as described herein. The delivery system may be configured for prokaryotic delivery, or for eukaryotic delivery. The delivery system may be configured for tissue-specific delivery, or non-tissue-specific delivery. The delivery system may be configured for inducible or non-inducible delivery. In certain embodiments, the delivery system may further include a liposomal, particle, exosomal, microvesicle, gene-gun, or viral delivery system.

[0137] The invention also provides a non-naturally occurring or engineered composition which is a composition having the characteristics as discussed herein or defined in any of the herein described methods.

[0138] The invention also provides a delivery system comprising one or more vectors or one or more polynucleotide molecules, the one or more vectors or polynucleotide molecules comprising one or more polynucleotide molecules encoding components of a non-naturally

occurring or engineered composition which is a composition having the characteristics as discussed herein or defined in any of the herein described methods.

[0139] The invention also provides a non-naturally occurring or engineered composition, or one or more polynucleotides encoding components of said composition, or delivery systems comprising one or more polynucleotides encoding components of said composition for use in a therapeutic method of treatment. The therapeutic method of treatment may comprise gene or genome editing, or gene therapy.

[0140] Delivery vehicles, vectors, particles, nanoparticles, formulations and components thereof for expression of one or more elements of a nucleic acid-targeting system are as used in the foregoing documents, such as WO 2014/093622 (PCT/US2013/074667). In some embodiments, a vector comprises one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a “cloning site”). In some embodiments, one or more insertion sites (e.g., about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors. When multiple different guide sequences are used, a single expression construct may be used to target nucleic acid-targeting activity to multiple different, corresponding target sequences within a cell. For example, a single vector may comprise about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 15, 20, or more guide sequences. In some embodiments, about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more such guide-sequence-containing vectors may be provided, and optionally delivered to a cell. In some embodiments, a vector comprises a regulatory element operably linked to an enzyme-coding sequence encoding a nucleic acid-targeting effector module. Nucleic acid-targeting effector module or nucleic acid-targeting guide RNA or RNA(s) can be delivered separately; and advantageously at least one of these is delivered via a particle complex. nucleic acid-targeting effector module mRNA can be delivered prior to the nucleic acid-targeting guide RNA to give time for nucleic acid-targeting effector module to be expressed. Nucleic acid-targeting effector module mRNA might be administered 1-12 hours (preferably around 2-6 hours) prior to the administration of nucleic acid-targeting guide RNA. Alternatively, nucleic acid-targeting effector module mRNA and nucleic acid-targeting guide RNA can be administered together. Advantageously, a second booster dose of guide RNA can be administered 1-12 hours (preferably around 2-6 hours) after the initial administration of nucleic

acid-targeting effector module mRNA + guide RNA. Additional administrations of nucleic acid-targeting effector module mRNA and/or guide RNA might be useful to achieve the most efficient levels of genome modification.

[0141] The invention also provides a non-naturally-occurring, engineered composition comprising:

a delivery system operably configured to deliver CRISPR-Cas effector module components or one or more polynucleotide sequences comprising or encoding said components into a cell, and wherein said CRISPR-Cas effector module is operable in the cell,

CRISPR-Cas effector module components or one or more polynucleotide sequences encoding for transcription and/or translation in the cell the CRISPR-Cas effector module components, comprising:

(I) the non-naturally-occurring CRISPR effector module (*e.g.* engineered Cas9, Cpf1, C2c2) as described herein;

(II) CRISPR-Cas effector module guide RNA comprising:

the guide sequence, and

a direct repeat sequence,

wherein the enzyme in the CRISPR complex has reduced capability of modifying one or more off-target loci as compared to an unmodified enzyme and/or whereby the enzyme in the CRISPR complex has increased capability of modifying the one or more target loci as compared to an unmodified enzyme.

[0142] In an aspect, the invention also provides in a system comprising the engineered CRISPR effector module as described herein, such as described in this section.

[0143] In any such compositions, the delivery system may comprise a yeast system, a lipofection system, a microinjection system, a biolistic system, virosomes, liposomes, immunoliposomes, polycations, lipid:nucleic acid conjugates or artificial virions, as defined herein elsewhere.

[0144] In any such compositions, the delivery system may comprise a vector system comprising one or more vectors, and wherein component (II) comprises a first regulatory element operably linked to a polynucleotide sequence which comprises the guide sequence, the

direct repeat sequence and optionally, and wherein component (I) comprises a second regulatory element operably linked to a polynucleotide sequence encoding the CRISPR enzyme.

[0145] In any such compositions, the delivery system may comprise a vector system comprising one or more vectors, and wherein component (II) comprises a first regulatory element operably linked to the guide sequence and the direct repeat sequence, and wherein component (I) comprises a second regulatory element operably linked to a polynucleotide sequence encoding the CRISPR enzyme.

[0146] In any such compositions, the composition may comprise more than one guide RNA, and each guide RNA has a different target whereby there is multiplexing.

[0147] In any such compositions, the polynucleotide sequence(s) may be on one vector.

[0148] The delivery can be of one or more polynucleotides encoding any one or more or all of the CRISPR-complex, advantageously linked to one or more regulatory elements for in vivo expression, e.g. via particle(s), containing a vector containing the polynucleotide(s) operably linked to the regulatory element(s). Any or all of the polynucleotide sequence encoding a CRISPR enzyme, guide sequence, direct repeat sequence, may be RNA. It will be appreciated that where reference is made to a polynucleotide, which is RNA and is said to 'comprise' a feature such a direct repeat sequence, the RNA sequence includes the feature. Where the polynucleotide is DNA and is said to comprise a feature such a direct repeat sequence, the DNA sequence is or can be transcribed into the RNA including 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).

[0149] In certain embodiments the invention provides a method of modifying an organism, e.g., mammal including human or a non-human mammal or organism by manipulation of a target sequence in a genomic locus of interest of an HSC e.g., wherein the genomic locus of interest is associated with a mutation associated with an aberrant protein expression or with a disease condition or state, comprising delivering, e.g., via contacting of a non-naturally occurring or engineered composition with the HSC, wherein the composition comprises one or more particles comprising viral, plasmid or nucleic acid molecule vector(s) (e.g. RNA) operably encoding a composition for expression thereof, wherein the composition comprises: (A) I. a first regulatory element operably linked to a CRISPR-Cas system RNA polynucleotide sequence, wherein the

polynucleotide sequence comprises (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, (b) a direct repeat 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 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 the guide sequence that is hybridized to the target sequence, or (B) a non-naturally occurring or engineered composition comprising a vector system comprising one or more vectors comprising I. a first regulatory element operably linked to (a) a guide sequence capable of hybridizing to a target sequence in a eukaryotic cell, and (b) at least one or more direct repeat sequences, II. a second regulatory element operably linked to an enzyme-coding sequence encoding a CRISPR enzyme, and optionally, where applicable, wherein components I, and II are located on the same or different vectors of the system, wherein when transcribed 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 the guide sequence that is hybridized to the target sequence; the method may optionally include also delivering a HDR template, e.g., via the particle contacting the HSC containing or contacting the HSC with another particle containing, the HDR template wherein the HDR template provides expression of a normal or less aberrant form of the protein; wherein "normal" is as to wild type, and "aberrant" can be a protein expression that gives rise to a condition or disease state; and optionally the method may include isolating or obtaining HSC from the organism or non-human organism, optionally expanding the HSC population, performing contacting of the particle(s) with the HSC to obtain a modified HSC population, optionally expanding the population of modified HSCs, and optionally administering modified HSCs to the organism or non-human organism. In some embodiments, components I, II and III are located on the same vector. In other embodiments, components I and II are located on the same vector, while component III is located on another vector. In other embodiments, components I and III are located on the same vector, while component II is located on another

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

[0150] The invention also provides an expression vector comprising any of the above-described polynucleotide molecules. The invention also provides such polynucleotide molecule(s), for instance such polynucleotide molecules operably configured to express the protein and/or the nucleic acid component(s), as well as such vector(s).

[0151] 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.

[0152] In particular embodiments, the CRISPR system components can be introduced in the plant cells using a plant viral vector (Scholthof et al. 1996, *Annu Rev Phytopathol.* 1996;34:299-323). In further particular embodiments, said viral vector is a vector from a DNA virus. For example, geminivirus (e.g., cabbage leaf curl virus, bean yellow dwarf virus, wheat dwarf virus, tomato leaf curl virus, maize streak virus, tobacco leaf curl virus, or tomato golden mosaic virus) or nanovirus (e.g., Faba bean necrotic yellow virus). In other particular embodiments, said viral vector is a vector from an RNA virus. For example, tobnavirus (e.g., tobacco rattle virus, tobacco mosaic virus), potexvirus (e.g., potato virus X), or hordevirus (e.g., barley stripe mosaic virus). The replicating genomes of plant viruses are non-integrative vectors.

[0153] In particular embodiments, the vector used for transient expression of CRISPR constructs is for instance a pEAQ vector, which is tailored for *Agrobacterium*-mediated transient expression (Sainsbury F. et al., *Plant Biotechnol J.* 2009 Sep;7(7):682-93) in the protoplast.

Precise targeting of genomic locations was demonstrated using a modified Cabbage Leaf Curl virus (CaLCuV) vector to express gRNAs in stable transgenic plants expressing a CRISPR enzyme (Scientific Reports 5, Article number: 14926 (2015), doi:10.1038/srep14926).

[0154] In the case of Agrobacterium-mediated transformation, the marker cassette may be adjacent to or between flanking T-DNA borders and contained within a binary vector. In another embodiment, the marker cassette may be outside of the T-DNA. A selectable marker cassette may also be within or adjacent to the same T-DNA borders as the expression cassette or may be somewhere else within a second T-DNA on the binary vector (e.g., a 2 T-DNA system).

[0155] In another embodiment, self-inactivating lentiviral vectors with an siRNA targeting a common exon shared by HIV tat/rev, a nucleolar-localizing TAR decoy, and an anti-CCR5-specific hammerhead ribozyme (see, e.g., DiGiusto et al. (2010) Sci Transl Med 2:36ra43) may be used/and or adapted to the CRISPR-Cas effector module system of the present invention. A minimum of 2.5×10^6 CD34+ cells per kilogram patient weight may be collected and prestimulated for 16 to 20 hours in X-VIVO 15 medium (Lonza) containing 2 μ mol/L-glutamine, stem cell factor (100 ng/ml), Flt-3 ligand (Flt-3L) (100 ng/ml), and thrombopoietin (10 ng/ml) (CellGenix) at a density of 2×10^6 cells/ml. Prestimulated cells may be transduced with lentiviral at a multiplicity of infection of 5 for 16 to 24 hours in 75-cm² tissue culture flasks coated with fibronectin (25 mg/cm²) (RetroNectin, Takara Bio Inc.).

[0156] In some embodiments, the invention relates to a host cell comprising a guide sequence or sgRNA as described herein. In other embodiments, a host cell may also include a polynucleotide, vector, vector system, complex, or delivery system according to the invention as described herein. In some embodiments, the invention relates to a host cell expressing or capable of expressing (such as a host cell comprising a polynucleic acid sequence encoding a guide sequence or sgRNA as described herein). A sequence as provided by the invention may advantageously be provided on a vector, such as a suitable expression vector. The host cell may be any type of host cell. By means of example, the host cell may be a prokaryotic host cell, or may be a eukaryotic host cell. In certain embodiments, the host cell may be an isolated host cell, i.e. a cell not residing in a (multicellular) organism, such as an isolated human or animal host cell. In other embodiments, a host cell may be a cell in culture conditions.

[0157] In some embodiments, the invention provides a composition comprising a guide sequence or sgRNA, polynucleotide, vector, vector system, complex, delivery system, or host cell according to the invention as described herein. In certain embodiments, said composition may be a pharmaceutical or non-pharmaceutical composition.

[0158] In an aspect, the invention provides a transgenic organism comprising a guide sequence or sgRNA, polynucleotide, vector, vector system, complex, delivery system, host cell, or composition according to the invention as described herein. In a further aspect, the invention provides a transgenic organism expressing or capable of expressing. For example, a transgenic organism of the invention may comprise a polynucleic acid sequence encoding a guide sequence or sgRNA as described herein, such as advantageously provided on a vector, such as a suitable expression vector, or a polynucleic acid sequence encoding a guide sequence or sgRNA of the invention, which is integrated in the genome, such as advantageously under control of a suitable promoter and optionally additional regulatory elements, as described herein elsewhere.

[0159] In some embodiments, the invention relates to a method of modifying or targeting a target DNA locus, the method comprising delivering to said locus a guide sequence, polynucleotide, vector, vector system, delivery system, complex, or composition according to the invention as described herein. In a further aspect, the invention relates to a method of modifying or targeting a target DNA locus, the method comprising delivering to said locus a guide sequence or sgRNA according to the invention, as described herein, wherein the guide sequence or sgRNA is capable of forming a complex with a target locus or nucleic acid sequence, and preferably wherein said target DNA locus is modified or targeted upon binding of said complex to said target DNA locus. In certain preferred embodiments, a guide sequence or sgRNA as described herein is capable of binding a target locus or nucleic acid as described herein. In certain embodiments, a guide sequence or sgRNA as described herein is capable of forming a (functional) complex with a target locus or nucleic acid sequence. In certain embodiments, the guide sequence or sgRNA as described herein in such a functional complex is capable of modifying or targeting a target DNA locus upon binding to said target DNA locus. In some embodiments, the guide sequence or sgRNA is capable of effecting modification of a target locus or nucleic acid sequence. Such a complex may be capable of inducing DNA modifications, such as for instance, single-stranded or double-stranded DNA cleavage, or may be capable of altering

DNA architecture/structure, epigenetic modification, or gene expression. In certain embodiments, said modifying or targeting a target locus comprises inducing a DNA strand break, such as a single strand break or a double strand break. In certain embodiments, said modifying or targeting a target locus comprises altering gene expression of one or more genes. In certain embodiments, said modifying or targeting a target locus comprises epigenetic modification of said target DNA locus, or comprises chromatin modification of said target DNA locus, or modification of the architecture of said target DNA locus.

[0160] In an aspect, the invention relates to a guide sequence or sgRNA identified by the methods according to the invention. In a further aspect, the invention relates to a polynucleotide encoding a guide sequence or sgRNA identified according to the methods of the invention, or vectors, vector systems, complexes, compositions, delivery systems, host cells, or transgenic organisms, as described herein elsewhere.

[0161] A guide sequence or sgRNA of the invention may be used for any of the methods or uses as described herein, such as without limitation in functional screening, genome-wide knockout screening, multiplexing, saturating mutagenesis, or modification of cells or organisms, as well as for therapeutic applications, such as those described herein.

[0162] In an aspect, the invention relates to the use of a guide sequence or sgRNA, polynucleotide, vector, vector system, complex, delivery system, host cell, or composition according to according to the invention as described herein for modifying or targeting a DNA target locus, preferably in vitro or ex vivo, or for modifying or targeting a DNA target locus in a non-human and/or non-animal organism, as described herein. In a further aspect, the invention relates to the use of a guide sequence or sgRNA, polynucleotide, vector, vector system, complex, delivery system, or host cell according to the invention as described herein for genome editing, preferably in vitro or ex vivo, or for genome editing of a non-human and/or non-animal organism, as described herein elsewhere. In another aspect, the invention relates to a guide sequence or sgRNA, polynucleotide, vector, vector system, complex, delivery system, host cell, or composition according to according to the invention as described herein for use in modifying or targeting a DNA target locus, such as in vitro, ex vivo, or in vivo, as described herein. In another aspect, the invention relates to a guide sequence or sgRNA, polynucleotide, vector, vector system, complex, delivery system, host cell, or composition according to according to the

invention as described herein for use in genome editing, such as in vitro, ex vivo, or in vivo, as described herein. In a further aspect, the invention relates to a guide sequence or sgRNA, polynucleotide, vector, vector system, complex, delivery system, host cell, or composition according to the invention as described herein for use in therapy or for use as a medicament, as described herein. In yet another aspect, the invention relates to the use of a guide sequence or sgRNA, polynucleotide, vector, vector system, complex, delivery system, host cell, or composition according to the invention as described herein for the manufacture of a medicament, as described herein elsewhere.

[0163] In an aspect, the invention relates to a nucleic acid targeting system comprising guide sequence or sgRNA (or polynucleic acid encoding such) according to the invention as described herein.

[0164] The term “nucleic acid-targeting system,” wherein nucleic acid is DNA, and in some aspects may also refer to DNA-RNA hybrids or derivatives thereof, refers collectively to transcripts and other elements involved in the expression of or directing the activity of DNA-targeting CRISPR-associated (“Cas”) genes, which may include sequences encoding a DNA-targeting Cas protein and a DNA-targeting guide RNA comprising a CRISPR RNA (crRNA) sequence and (in CRISPR-Cas9 system but not all systems) a trans-activating CRISPR-Cas system RNA (tracrRNA) sequence, or other sequences and transcripts from a DNA-targeting CRISPR locus.

[0165] In the DNA targeting RNA-guided endonuclease systems described herein, a tracrRNA sequence is not required. In general, a RNA-targeting system is characterized by elements that promote the formation of a RNA-targeting complex at the site of a target DNA sequence. In the context of formation of a DNA-targeting complex, “target sequence” refers to a DNA sequence to which a DNA or RNA-targeting guide RNA is designed to have complementarity, where hybridization between a target sequence and a RNA-targeting guide RNA promotes the formation of a RNA-targeting complex. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell.

[0166] In an aspect, the invention relates to a CRISPR-Cas complex, system, or targeting system, or CRISPR complex, system, or targeting system, or Cas complex, system, or targeting

system, comprising a guide sequence or sgRNA(or polynucleic acid encoding such) as described herein.

[0167] In an aspect of the invention, novel DNA targeting systems also referred to as DNA-targeting CRISPR-Cas or the CRISPR-Cas DNA-targeting system of the present application are based on identified Type V (e.g. subtype V-A and subtype V-B) Cas proteins which do not require the generation of customized proteins to target specific DNA sequences but rather a single effector protein or enzyme can be programmed by a RNA molecule to recognize a specific DNA target, in other words the enzyme can be recruited to a specific DNA target using said RNA molecule. Aspects of the invention particularly relate to DNA targeting RNA-guided CRISPR systems.

[0168] The nucleic acids-targeting systems, the vector systems, the vectors and the compositions described herein may be used in various nucleic acid-targeting applications, altering or modifying synthesis of a gene product, such as a protein, nucleic acids cleavage, nucleic acids editing, nucleic acids splicing; trafficking of target nucleic acids, tracing of target nucleic acids, isolation of target nucleic acids, visualization of target nucleic acids, etc.

[0169] As used herein, a Cas protein or a CRISPR enzyme refers to any of the proteins presented in the new classification of CRISPR-Cas systems. In an advantageous embodiment, the present invention encompasses effector proteins identified in a Type V CRISPR-Cas loci.

[0170] Aspects of the invention also encompass methods and uses of the compositions and systems described herein in genome engineering, e.g. for altering or manipulating the expression of one or more genes or the one or more gene products, in prokaryotic or eukaryotic cells, in vitro, in vivo, or ex vivo.

[0171] In certain aspects, the present invention involves vectors. As 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, and throughout this specification, 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). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as “expression vectors.” Vectors for and that result in expression in a eukaryotic cell can be referred to herein as “eukaryotic expression vectors.” Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0172] 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 regard to recombination and cloning methods, mention is made of U.S. Patent Application No. 10/815,730, published September 2, 2004 as U.S. 2004-0171156 A1, the contents of which are herein incorporated by reference in their entirety.

[0173] The term “regulatory element” is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g., transcription

termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter may direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g., liver, pancreas), or particular cell types (e.g., lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In some embodiments, a vector comprises one or more pol III promoter (e.g., 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g., 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g., 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) [see, e.g., Boshart et al, Cell, 41:521-530 (1985)], the SV40 promoter, the dihydrofolate reductase promoter, the β -actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1 α 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.

[0174] Advantageous vectors include lentiviruses and adeno-associated viruses, and types of such vectors can also be selected for targeting particular types of cells.

[0175] The ability of a guide sequence (within a nucleic acid-targeting guide RNA) to direct sequence-specific binding of a nucleic acid-targeting complex to a target nucleic acid sequence may be assessed by any suitable assay. For example, the components of a nucleic acid-targeting CRISPR system sufficient to form a nucleic acid-targeting complex, including the guide sequence to be tested, may be provided to a host cell having the corresponding target nucleic acid sequence, such as by transfection with vectors encoding the components of the nucleic acid-targeting complex, followed by an assessment of preferential targeting (e.g., cleavage) within the target nucleic acid sequence, such as by Surveyor assay as described herein. Similarly, cleavage of a target nucleic acid sequence (or a sequence in the vicinity thereof) may be evaluated in a test tube by providing the target nucleic acid sequence, components of a nucleic acid-targeting complex, including the guide sequence to be tested and a control guide sequence different from the test guide sequence, and comparing binding or rate of cleavage at or in the vicinity of the target sequence between the test and control guide sequence reactions. Other assays are possible, and will occur to those skilled in the art. A guide sequence, and hence a nucleic acid-targeting guide RNA may be selected to target any target nucleic acid sequence. The target sequence may be DNA. 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.

[0176] In an aspect, the invention provides methods as herein discussed wherein the host is a eukaryotic cell. In an aspect, the invention provides a method as herein discussed wherein the host is a mammalian cell. In an aspect, the invention provides a method as herein discussed, wherein the host is a non-human eukaryote cell. In an aspect, the invention provides a method as herein discussed, wherein the non-human eukaryote cell is a non-human mammal cell. In an aspect, the invention provides a method as herein discussed, wherein the non-human mammal cell may be including, but not limited to, primate bovine, ovine, porcine, canine, rodent, Leporidae such as monkey, cow, sheep, pig, dog, rabbit, rat or mouse cell. In an aspect, the invention provides a method as herein discussed, the cell may be a non-mammalian eukaryotic

cell such as poultry bird (e.g., chicken), vertebrate fish (e.g., salmon) or shellfish (e.g., oyster, clam, lobster, shrimp) cell. In an aspect, the invention provides a method as herein discussed, the non-human eukaryote cell is a plant cell. The plant cell may be of a monocot or dicot or of a crop or grain plant such as cassava, corn, sorghum, soybean, wheat, oat or rice. The plant cell may also be of an algae, tree or production plant, fruit or vegetable (e.g., trees such as citrus trees, e.g., orange, grapefruit or lemon trees; peach or nectarine trees; apple or pear trees; nut trees such as almond or walnut or pistachio trees; nightshade plants; plants of the genus Brassica; plants of the genus Lactuca; plants of the genus Spinacia; plants of the genus Capsicum; cotton, tobacco, asparagus, carrot, cabbage, broccoli, cauliflower, tomato, eggplant, pepper, lettuce, spinach, strawberry, blueberry, raspberry, blackberry, grape, coffee, cocoa, etc).

[0177] In one aspect, the invention provides a method for developing a biologically active agent that modulates a cell signaling event associated with a 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) contacting a test compound with a model cell of any one of the above-described embodiments; and (b) detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with said mutation in said disease gene, thereby developing said biologically active agent that modulates said cell signaling event associated with said disease gene.

[0178] In one aspect, the invention provides a recombinant polynucleotide comprising a guide sequence downstream of a direct repeat sequence, wherein the guide sequence when expressed directs sequence-specific binding of a CRISPR-Cas complex to a corresponding target sequence present in a host cell. In some embodiments, the target sequence is a viral sequence present in a eukaryotic cell. In some embodiments, the target sequence is a proto-oncogene or an oncogene.

[0179] In one aspect, the invention provides a vector system or eukaryotic host cell comprising (a) a first regulatory element operably linked to a direct repeat sequence and one or more insertion sites for inserting one or more guide sequences (including any guide sequences as described herein) downstream of the DR sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a CRISPR-Cas complex to a target sequence in a eukaryotic cell, wherein the CRISPR-Cas complex comprises any CRISPR-Cas9 enzymes, modified or not,

complexed with a guide sequence or sgRNA as described herein that is hybridized to the target sequence; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said CRISPR-Cas9 enzyme comprising a nuclear localization sequence and/or NES. In some embodiments, the host cell comprises components (a) and (b). In some embodiments, component (a), component (b), or components (a) and (b) are stably integrated into a genome of the host eukaryotic cell. In some embodiments, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR-Cas complex to a different target sequence in a eukaryotic cell. In some embodiments, the CRISPR enzyme comprises one or more nuclear localization sequences and/or nuclear export sequences or NES of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in and/or out of the nucleus of a eukaryotic cell. In some embodiments, the CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In some embodiments, the CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In a preferred embodiment, the strand break is a staggered cut with a 5' overhang. In some embodiments, the first regulatory element is a polymerase III promoter. In some embodiments, the second regulatory element is a polymerase II promoter. In some embodiments, the direct repeat has a minimum length of 16 nts and a single stem loop. In further embodiments, the direct repeat has a length longer than 16 nts, preferably more than 17 nts, and has more than one stem loop or optimized secondary structures. In some embodiments, the guide sequence is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length.

Kits

[0180] In still further embodiments, the present invention provides kits comprising a guide sequence library as described herein, and/or a composition as described herein. In some embodiments, a kit in accordance with the invention may be used with the methods described herein. For example, a kit of the invention may comprise a guide sequence library or a composition as described herein. A kit as described herein may include one or more containers, for example a vial, a tube, a flask, or a syringe.

[0181] In further embodiments, a kit may include one or more tubes or wells of a culture or microtiter plate into which cells or nucleic acids may be placed. The kit may allow for the assay of a single sample, or more than one sample. In some embodiments, the kit may include a plurality of plates or tubes, which allows for numerous samples concurrently or consecutively. In particular embodiments, the cells or nucleic acids as described herein may be human or mouse cells.

[0182] The treatment reagents of the kit may take any one of a variety of forms, including reagents with which to isolate nucleic acids, or reagents to grow cells in culture. Such reagents may include, but are not limited to, any cell culture reagents known in the art and useful with the methods of the invention, for example, fetal bovine serum, antibiotics, buffers, or the like. Detection labels that may be associated with and/or linked to a given nucleic acid or protein may be included in a kit according to the invention. Detectable labels that are associated with and/or attached to a secondary binding ligand are also contemplated. Exemplary secondary ligands are those secondary antibodies that have binding affinity for the first antibody. A number of exemplary labels are known in the art and/or all such labels may be employed in connection with the present invention.

[0183] A kit of the invention may optionally include a suitably aliquoted composition of, for example, a particular cell type as described herein to serve as a positive or negative control. The components of a kit may be packaged either in aqueous media and/or in lyophilized form, and may be suitable for storage for any length or at any temperature. In other embodiments, kits of the present invention may comprise instructions or written directions for use of the kit.

[0184] Although specific embodiments have been described above in detail, the description is merely for purposes of illustration. It should be appreciated, therefore, that many aspects described above are not intended as required or essential elements unless explicitly stated otherwise. Modifications of, and equivalent components or acts corresponding to, the disclosed aspects of the example embodiments, in addition to those described above, can be made by a person of ordinary skill in the art, having the benefit of the present disclosure, without departing from the spirit and scope of embodiments defined in the following claims, the scope of which is to be accorded the broadest interpretation so as to encompass such modifications and equivalent structures.

[0185] The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example 1 – Example Implementation and Software Framework

[0186] The methods disclosed herein may be implemented as follows. Back-end services for the method were implemented in Python using the Flask web framework with Eventlet-enabled concurrent network operations. The application was deployed on the Gunicorn HTTP Server. The interactive front-end visualization scheme was written in coffeescript using the AngularJS framework, with Asynchronous Javascript and XML (AJAX) for real-time interfacing with the server. A Redis-based storage system was implemented so that users can navigate way from a web-based interface during longer library runs.

[0187] Each library design may be processed in parallel using a Celery message broker. This generates the library in the background on top of the Redis store, while making concurrent progress accessible to the front-end. When the front-end observes that the concurrent routine has completed, it loads the finalized results via AJAX. The results are returned in JavaScript Object Notation (JSON) for front-end display. Additionally, the Celery broker uses smtpplib to notify the user of completion via email. Interactive charts were designed for guide sequence visualization using the open-source Chart.js library and HTML5 canvas element.

[0188] To increase the speed of gene expression computations, the GTEx data was divided into a separate pandas dataframe per gene and serialized using cPickle. This allowed for transitioning the overhead of gene lookups to a Linux filesystem, which is significantly faster than Pandas. Binary serialization was done using MessagePack for fast decoding. All other serialization were performed in pickle (for Python objects) or cPickle (for text only objects).

Example 2 – Comparison of Guide Sequences Selected Using Embodiments Herein in Genome-Scale Screens

[0189] Guide sequences designed using embodiments disclosed herein were tested to determine the guide sequences have consistently higher/lower activity using a meta-analysis of 77 pooled CRISPR screens. To do this, depletion screen results were analyzed from the

GenomeCRISPR database, which compiles data from multiple genome-scale screens (Rauscher et al. *Nucleic Acids Res.* 45, D679-D686, 2017). In this database, each guide sequence is normalized by depletion/enrichment within each screen using percentile rank to allow for relative comparison across the entire dataset (sgRNA effect). Using the previously computed sgRNA effect scores, comparison was made to guide sequences chosen using the embodiments disclosed herein have consistently high/lower sgRNA effect score than a size-matched control set chosen from all guide sequences targeting the same gene.

[0190] To first obtain a set of universally-essential genes, results from two recent studies that measure depletion using genome-scale CRISPR loss-of-function screens across multiple different cell types were combined (Wang et al. *Science* 350, 1096,1101, 2015; Hart et al. *Cell* 163, 1515-1526, 2015)). Hart et al. identify 829 genes as essential in all 5 cell lines (from diverse tissues) that they examined. The same number of top-ranked genes were retrieved from Wang et al. (ranked by average depletion p value across the 4 cell lines) and then computed the intersection of these 2 lists to find genes in common between these studies.

[0191] For each gene, the computer-implemented methods disclosed herein were used to generate a list of the top 50 guide sequences per genes (GTEx expression enabled using the average of all tissues, Pfam protein domain targeting enabled). For these guide sequences, a search for as many of these 50 guide sequence as possible was conducted in Genome CRISPR depletion experiments. (mean $\pm$ s.d.: 8 ± 6 sgRNAs per gene found in GenomeCRISPR). For each gene, the same number of sgRNAs were also selected from the GenomeCRISPR database (i.e. with no preference for GUIDES-ranking).

[0192] The average increase in depletion by using sgRNAs generated using the computer-based methods disclosed herein over the size-matched randomly selected sets was 0.73 sgRNA effect (~10% increased depletion, $n = 403$ genes examined in 77 genome-scale screens using 61 different cell lines), which is significantly greater depletion ($p = 5e-07$, $t = -5.1$, $df = 409$, two-sample paired t-test). The increased depletion of these sgRNAs over random selection of sgRNAs targeting the same gene in GenomeCRISPR can be visualized by examining the cumulative distribution of depletion scores (FIG. 7), where negative values of sgRNA effect indicate greater depletion.

Example 3 – Library Run-time Comparison

[0193] The library design run time based on targeting 500 genes in the human genome using an example embodiment disclosed herein was 15 seconds, compared to three minutes for GuideScan (Perez et al. Nature Biotechnology 35, 347-349, 2017), and 90 mins for CLD (Heigwer et al. Genome Biology 17, 55, 2016). It should be noted that GuideScan further requires manual conversion of gene list to genomic coordinates and this step was not include in the runtime cited. CLD require database setup and an a computationally intensive initial processing step.

[0194] Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention. This application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure come within known customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

CLAIMS

What is claimed is:

1. A computer-implemented method for making guide sequences comprising:
identifying, using one or more computing devices, exons in one or more target genes;
identifying, by the one or more computing devices, potential CRISPR-Cas system target regions in each exon;
ranking, by the one or more computing devices, each identified potential CRISPR-Cas system target regions based at least in part by an off-target avoidance score and an on-target efficiency score determined using a classifier, and optionally, by the presence of a protein domain in the cutting region;
generating, by the one or more computing devices, a set of guide sequences based at least in part on the ranked CRISPR-Cas targeting regions, wherein the set of guide sequences are designed to target regions ranked above a defined cut-off value.
2. The method of claim 1, wherein the exons are defined as a union of consensus CDS (CCDS) regions.
3. The method of claim 1 or 2, wherein the one or more target genes are identified based on an input nucleic acid sequence, a gene symbol, or a combination thereof.
4. The method of any of the proceeding claims, wherein the potential CRISPR-Cas target regions are identified by proximity to a PAM.
5. The method of any of the proceeding claims, wherein the exons are selected based on tissue-specific expression data to select exons with higher expression.
6. The method of claim 5, wherein selecting coding regions based on tissue-specific expression data comprises determining a median expression for each exon in each tissue and determining a median expression value for each exon in each tissue for all samples of a same tissue type, and normalizing the median read counts by coding region base pair size.

7. The method of any one of the proceeding claims, wherein the off-target avoidance score is determined by taking a sum of a cutting frequency determination score for each off-target site identified in an exome.
8. The method of any of the proceeding claims, wherein the classifier is applied to local sequence preferences learned from saturation mutagenesis studies.
9. The method of any of the proceeding claims, wherein the classifier is a boosted regression tree classifier.
10. The method of any of the proceeding claims, further comprising eliminating guide sequences with homopolymer repeats, excluding the last exon in a coding region, eliminating guide sequences with transcriptional terminators, or a combination thereof.
11. The method of any of the proceeding claims, further comprising generating non-targeting guide sequences to serve as a negative control.
12. The method of any of the proceeding claims, wherein the identified exons are species-specific.
13. The method of any of the proceeding claims, wherein the guide sequences are based on a full length sgRNA scaffold, an E+F scaffold, or a truncated guide sequence scaffold.
14. The method of any of the proceeding claims, wherein the CRISPR-Cas system is a CRISPR-Cas Type II system, a CRISPR-Cas V system, or a CRISPR-Cas VI system.
15. The method of any one of the proceeding claims, further comprising synthesizing the guide sequences.

16. The method of any of the proceeding claims, further comprising preparing a guide library comprising the set of guide sequences.

17. The method of claim 15, wherein the guide sequences are RNA, DNA, DNA-RNA hybrid, or chemically modified guide sequences.

18. A guide library made using the methods of any one of claims 1 to 17.

19. A guide library comprising guide sequences to one or more target regions in one or more exons of one or more target genes, wherein individual guide sequences in the library are included based on optimization of an off-target avoidance score and an on-target efficiency score, and optionally, by the presence of a protein domain in the target region.

20. The library of claim 19, wherein the exons are selected based on tissue-specific expression data to select exons with higher expression.

21. The library of any one of claims 19 to 20, wherein the off-target avoidance score is determined by taking a sum of a cutting frequency determination score for each off-target site identified in an exome of the one or more target genes.

22. The library of any one of claims 19 to 21, wherein the on-target efficiency score is determined by use of a classifier applied to local sequence preferences learned from saturation mutagenesis studies.

23. The library of claim 22, wherein the classifier is a boosted regression tree classifier.

24. The library of any one of claims 19 to 23, wherein the guide sequences exclude guide sequences targeting homopolymer regions, targeting the last exon in a coding region, include target regions with transcriptional terminators, or a combination thereof.

25. The library of any one of claims 19 to 24, wherein the guide sequences are full length guide sequences, truncated guide sequences, full length sgRNA sequences, truncated sgRNA sequence, or E+F sgRNA sequences.

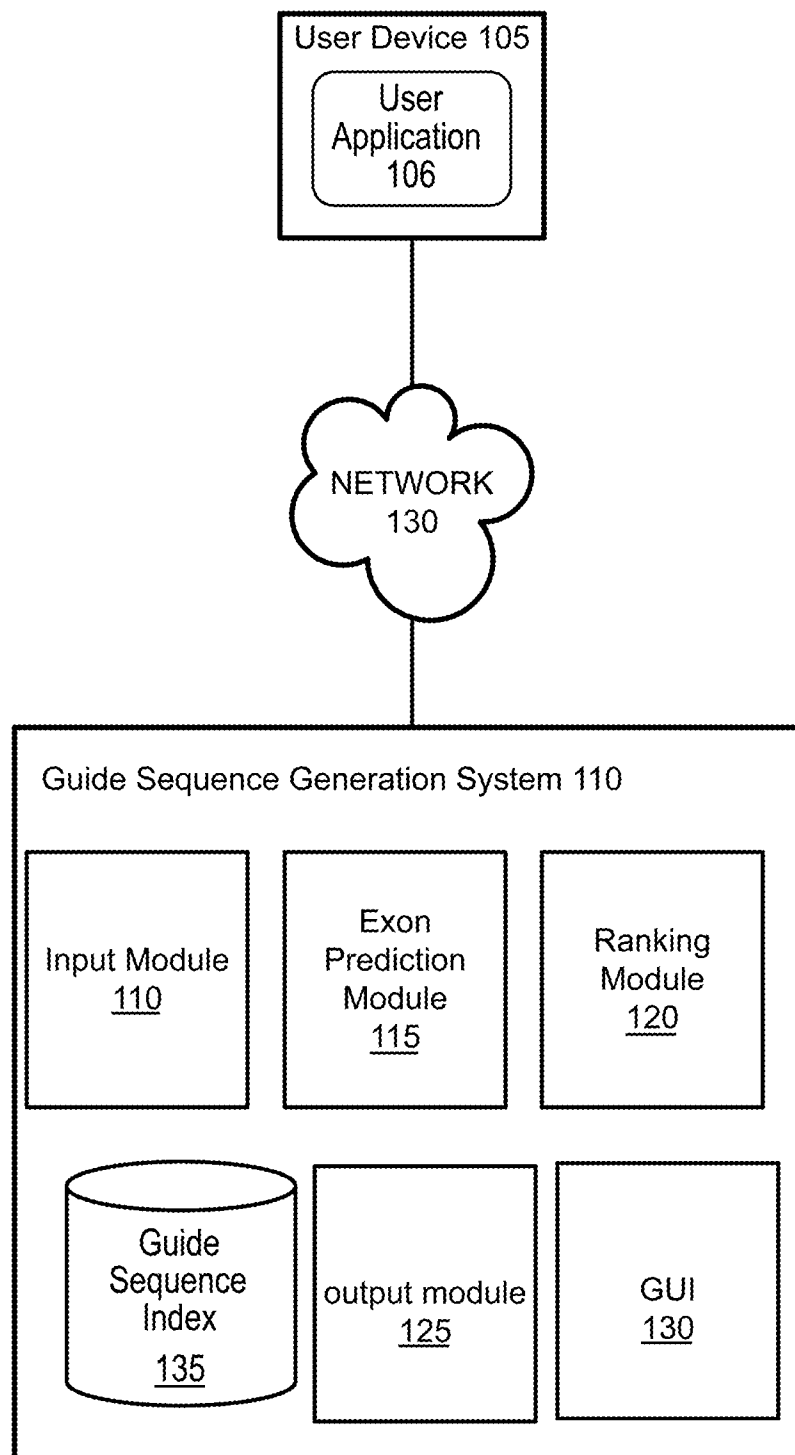
26. The library of any one of claims 19 to 25, wherein the guide sequences are RNA, DNA, DNA-RNA hybrid, chemically modified, or a combination thereof.

27. A composition comprising (i) a population of cells and (ii) the guide sequence library of any one of claims 18 to 26, wherein each of the cells contains one or more of the guide sequences and thus the guide sequences of the library are integrated into the population of cells.

28. The composition of claim 27, wherein the population of cells is a population of eukaryotic cells.

29. A kit comprising the guide sequence library of any one of claims 18 to 26 and/or the composition of claim 27 or 28.

1/9

100**FIG. 1**

2/9

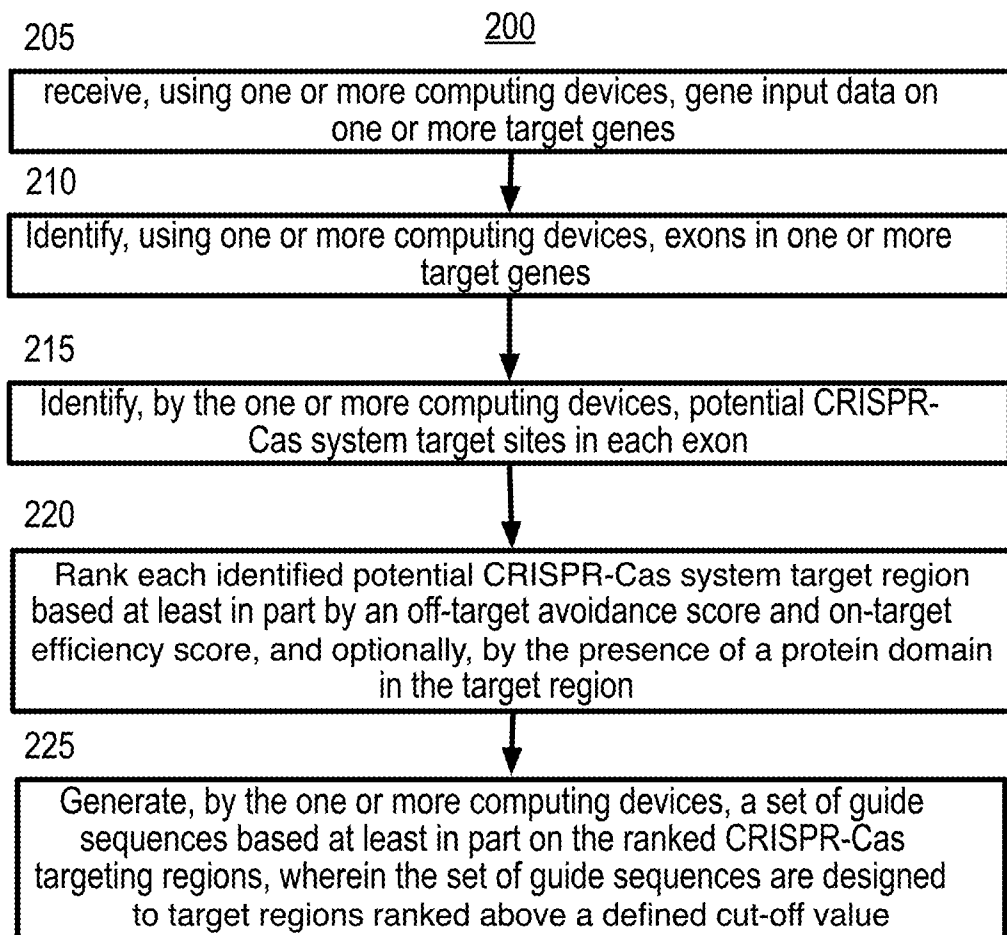
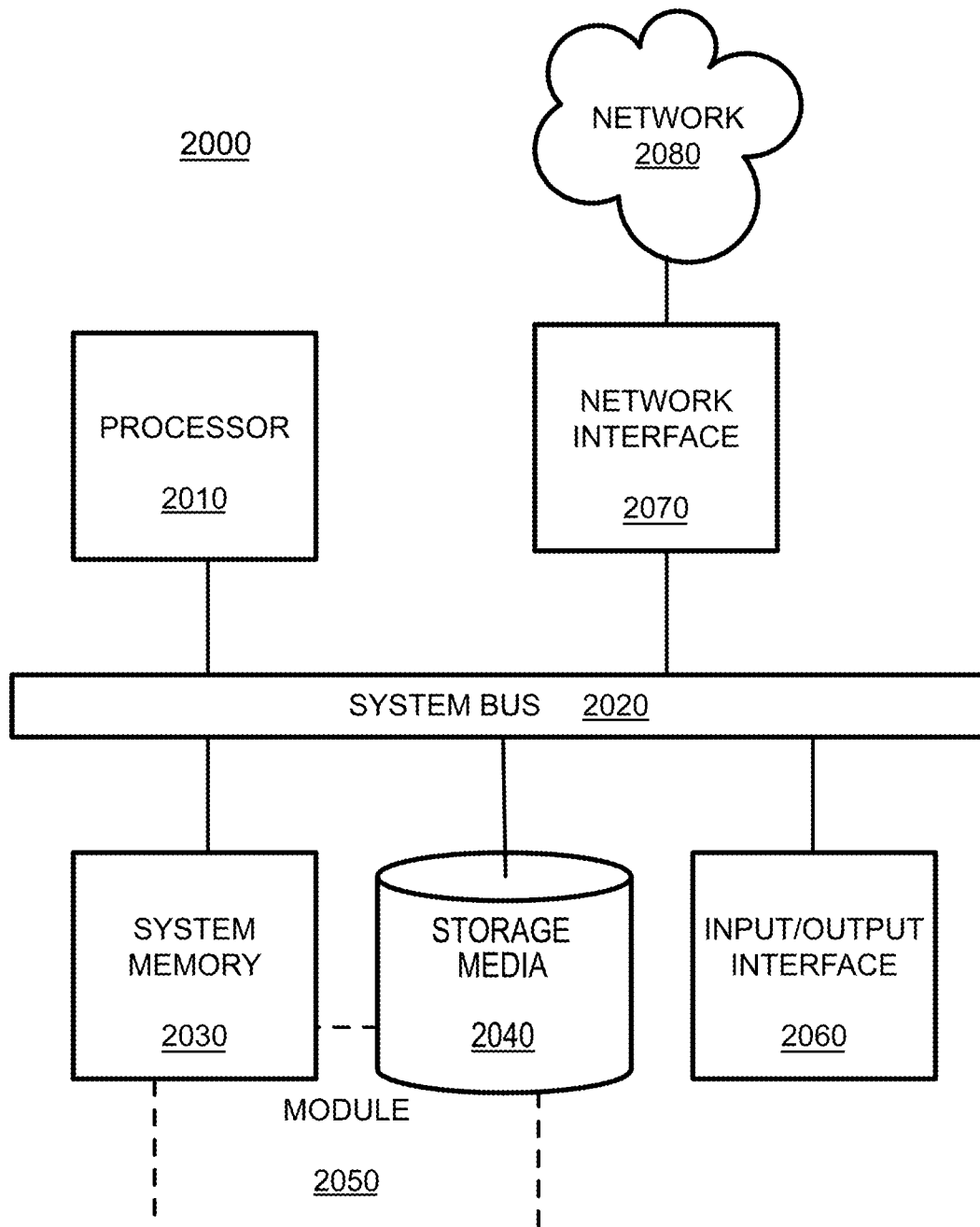


FIG. 2

3/9

**FIG. 3**

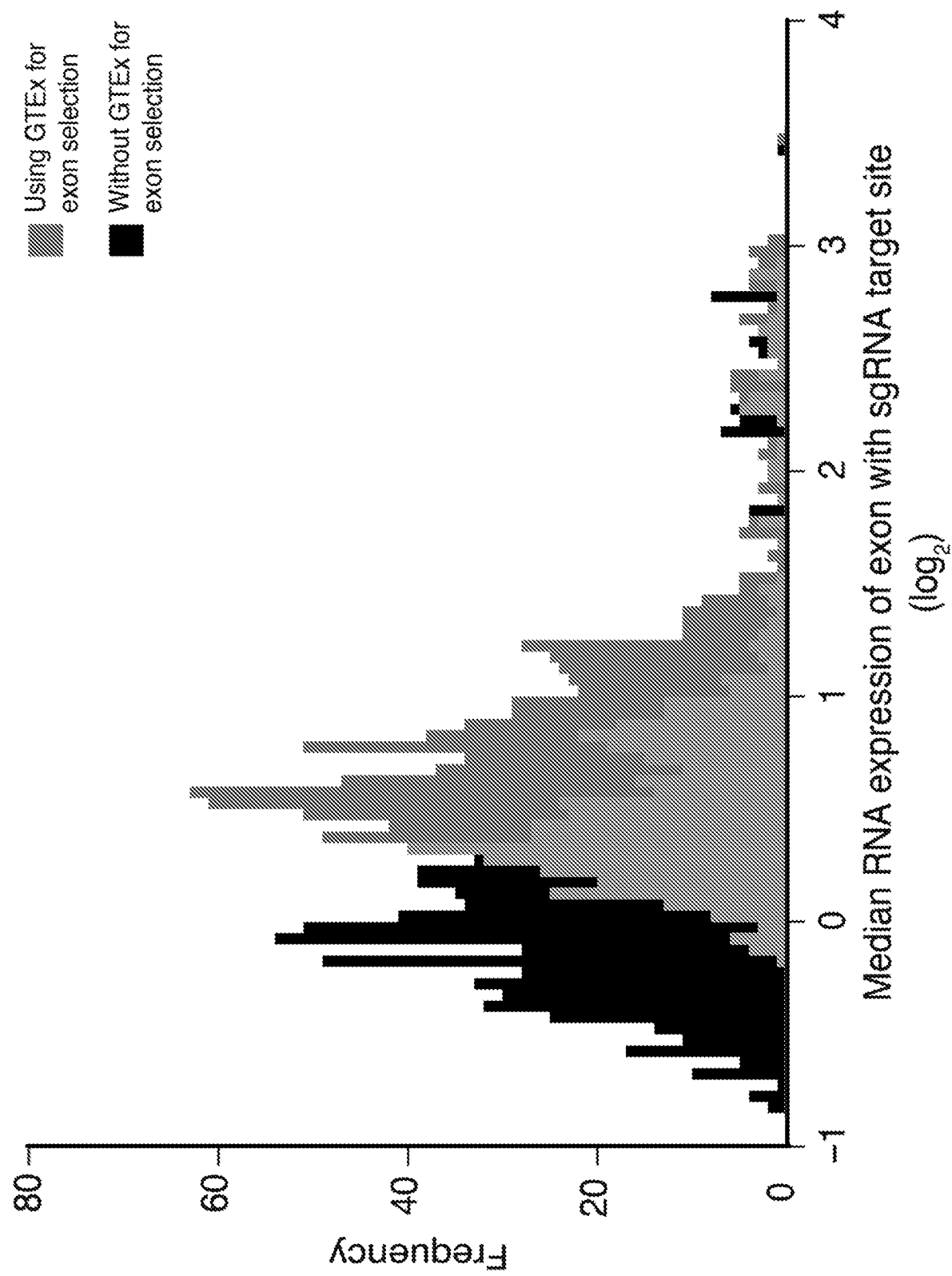


FIG. 4

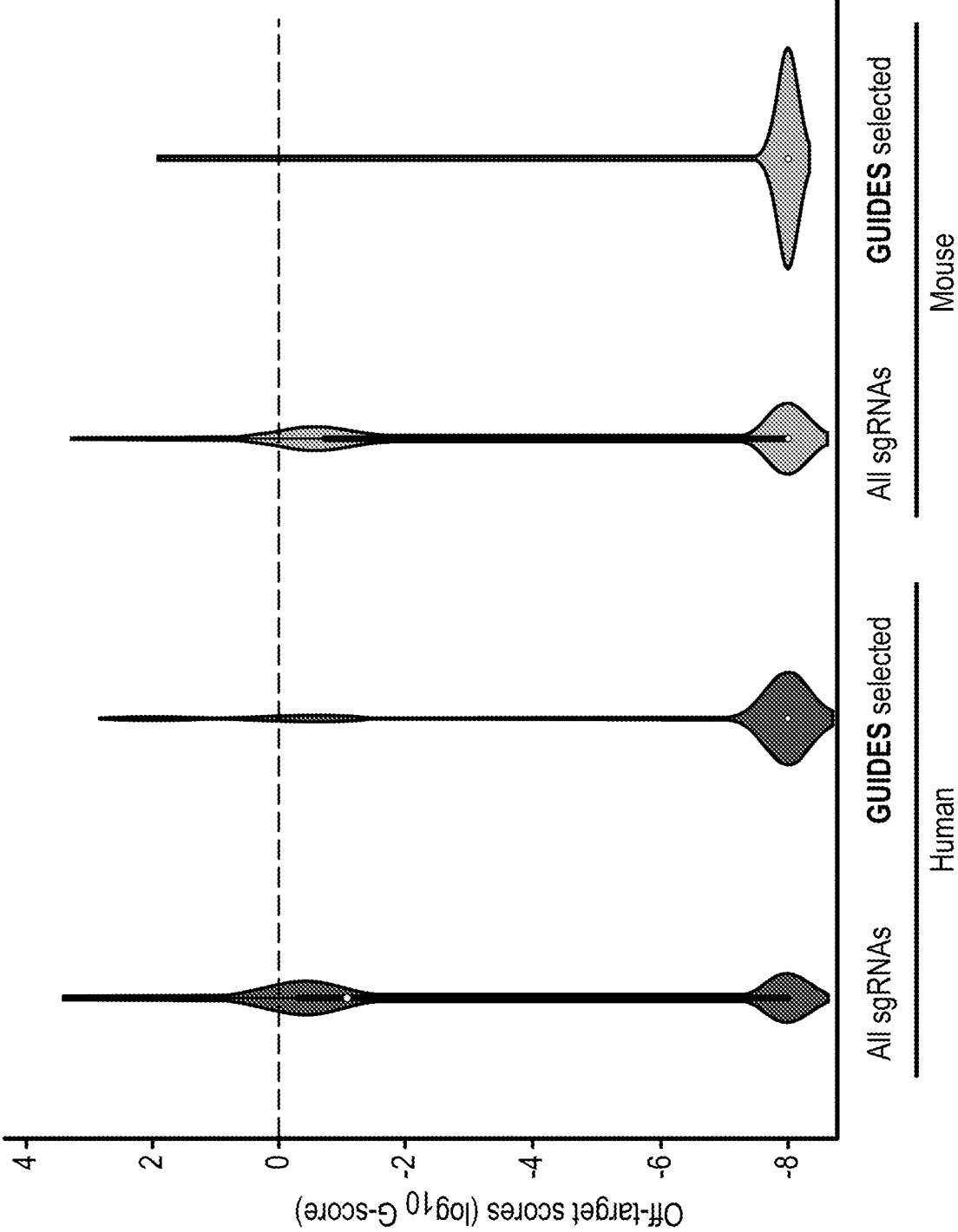


FIG. 5

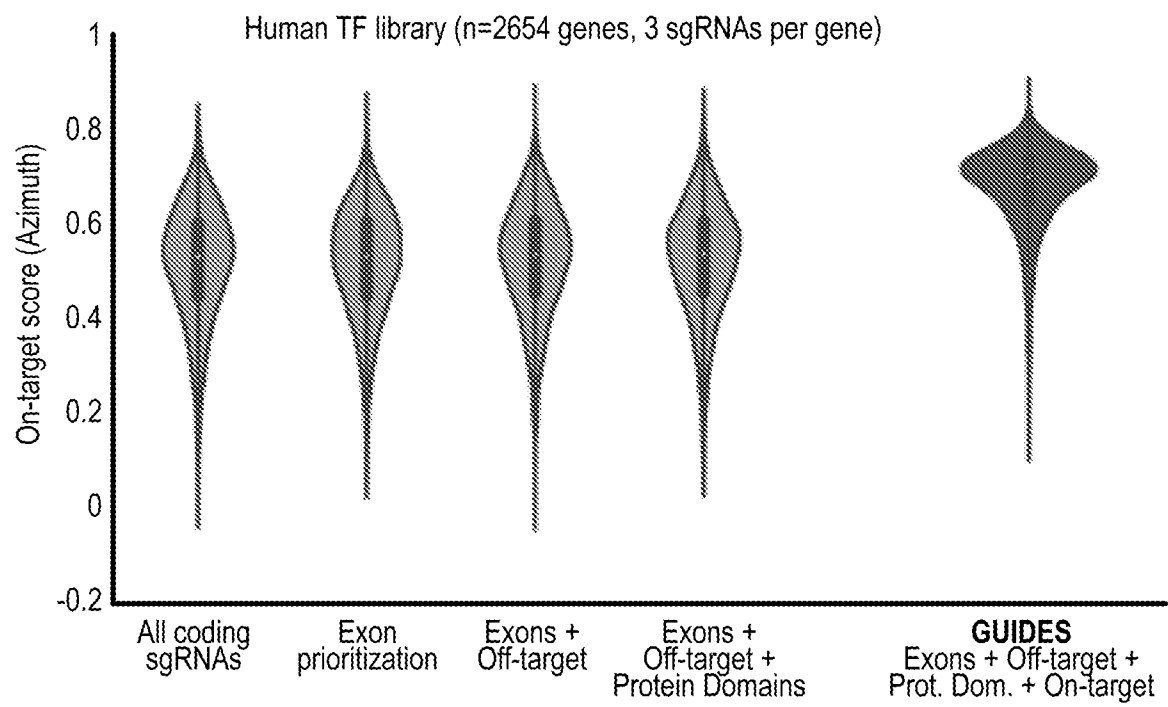


FIG. 6A

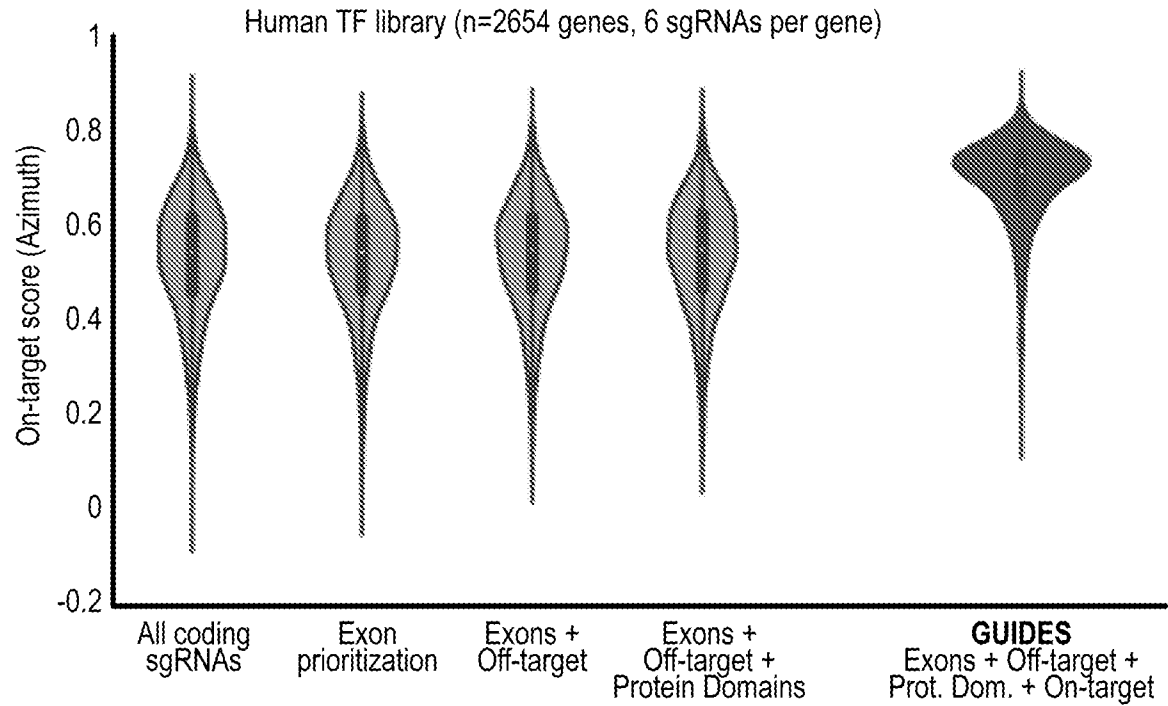


FIG. 6B

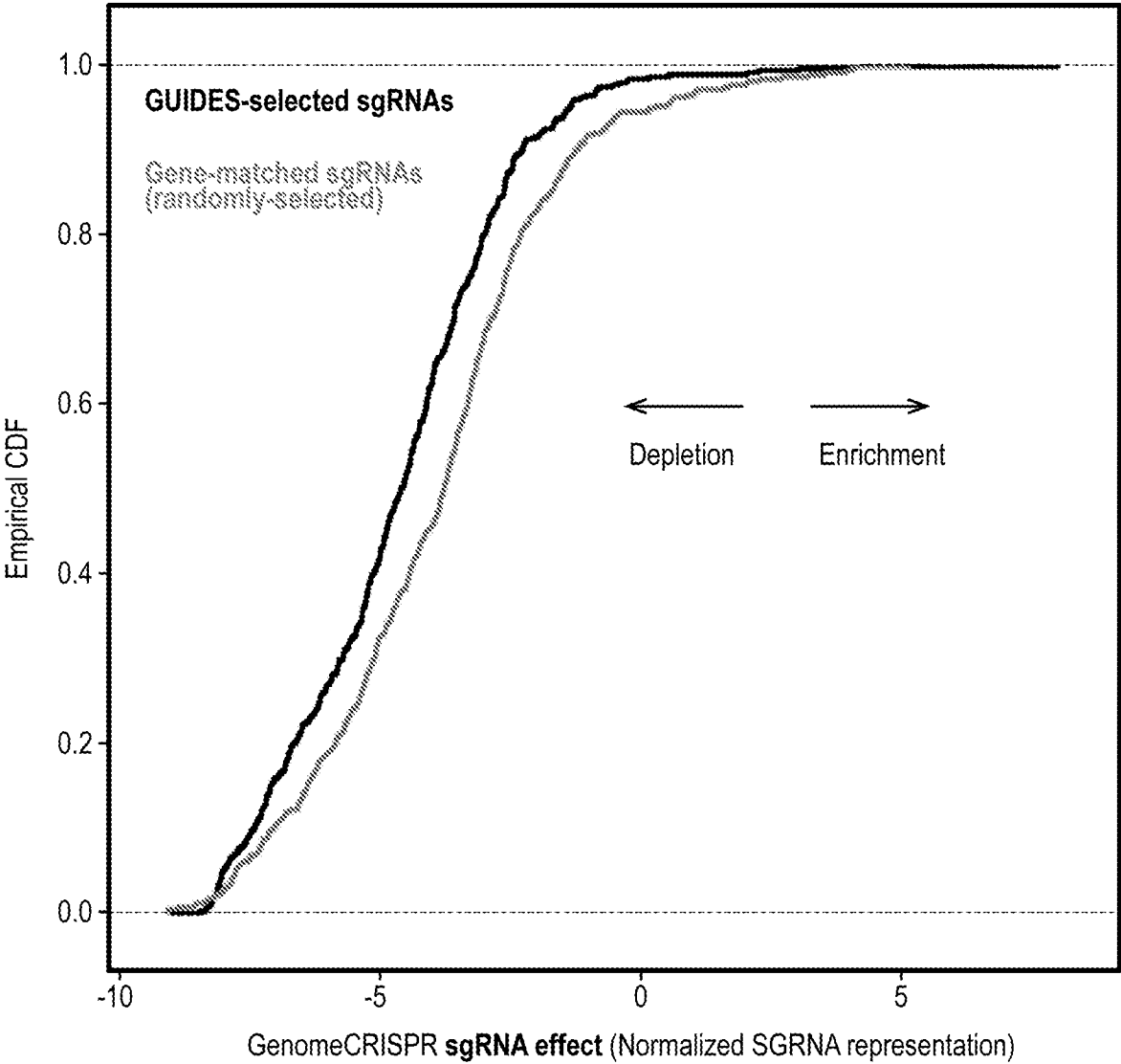


FIG. 7

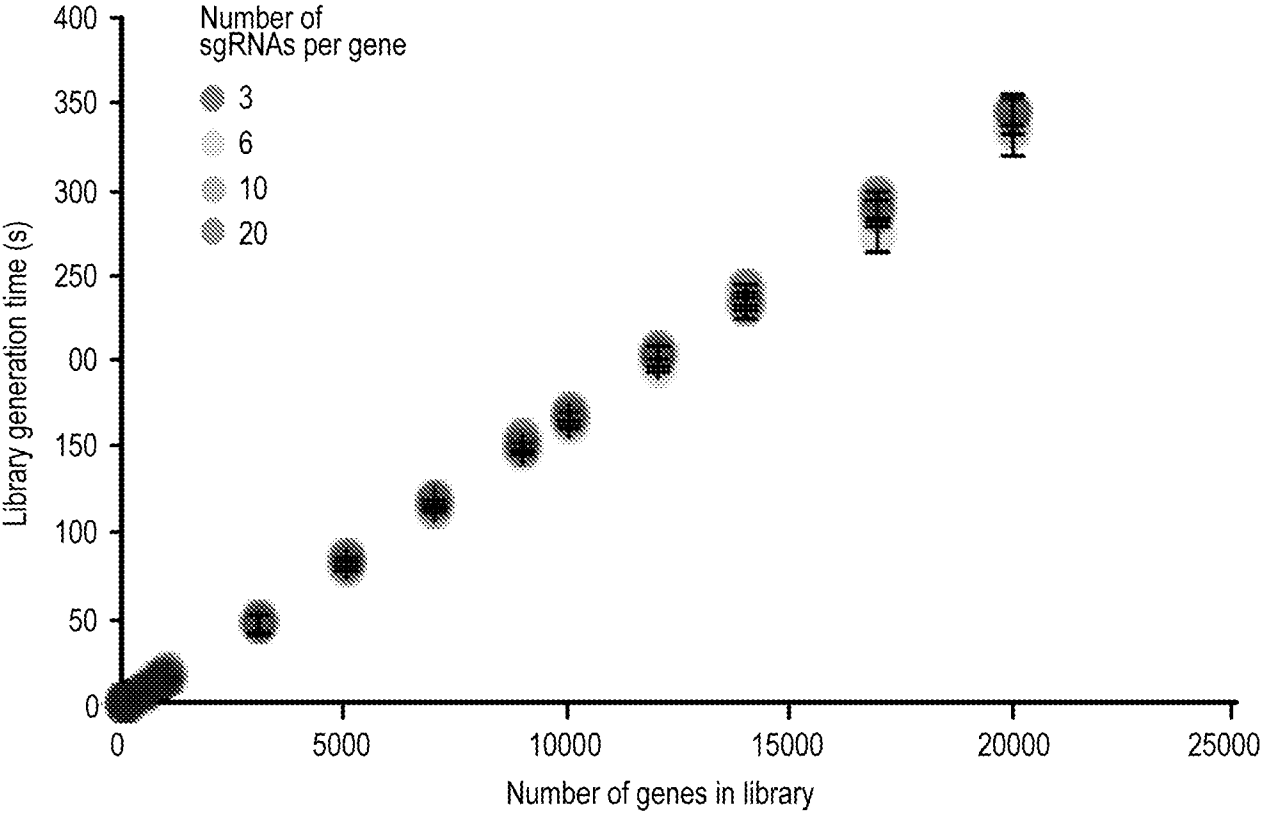


FIG. 8

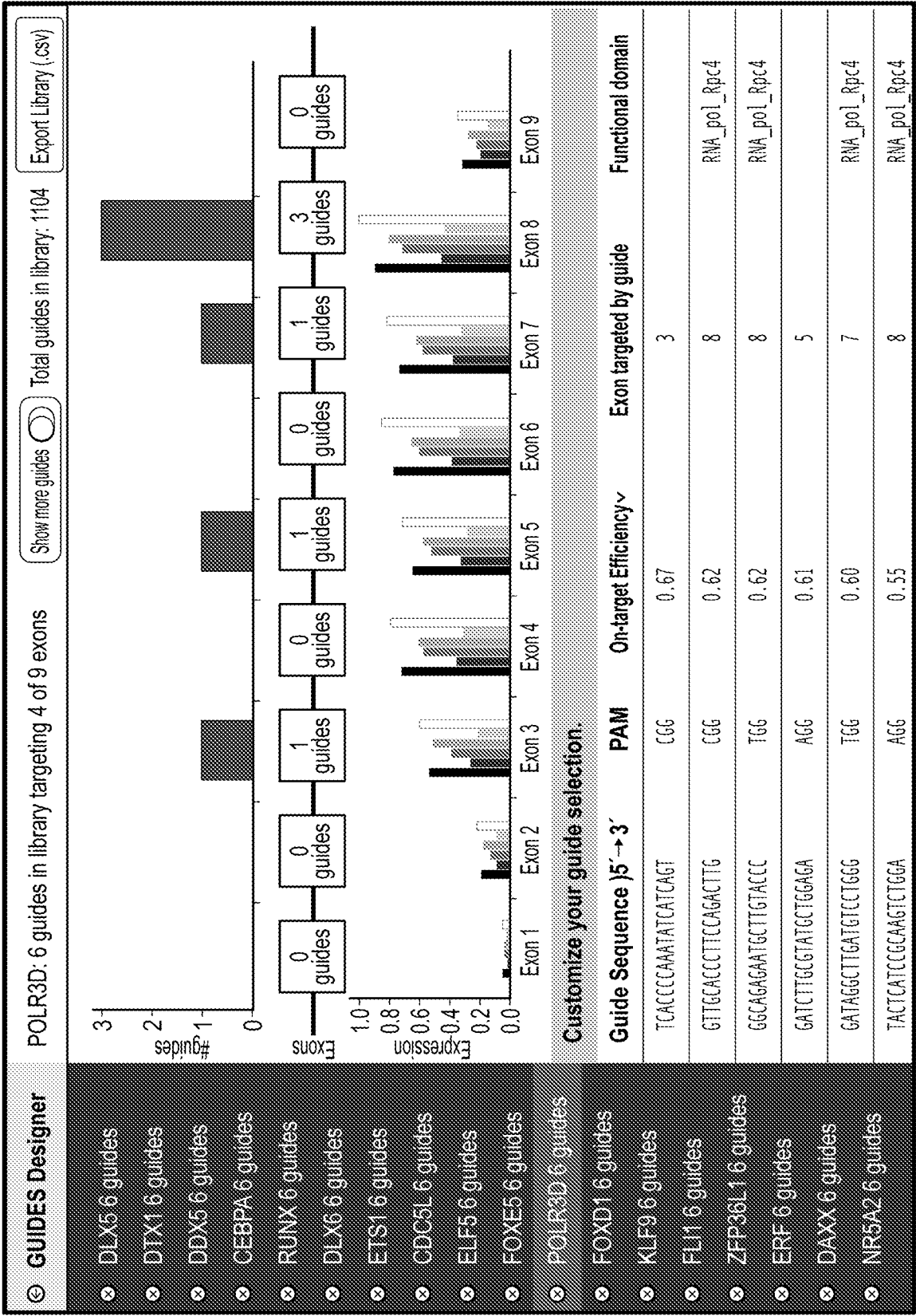


FIG. 9

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 18/41041

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - A61K 48/00, G06F 19/00 (2018.01)

CPC - A61K 38/465, G06F 19/18, A61K 31/7088, C12N 15/907, G06F 19/22

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History Document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

See Search History Document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

See Search History Document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X --- Y	US 2017/0053062 A1 (GEORGIA TECH RESEARCH CORPORATION) 23 February 2017 (23.02.2017) para [0004], [0011], [0015], [0017], [0024], [0086], [0093], [0103], [0216], [0226], [0236], [0280], [0281], [0288], [0290], [0305], [0359], [0385], [0387], [0416], Figs. 1A, 2A	1, 3, 19-21 ----- 2, 3/2
Y	PRUITT et al. The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. Genome res. July 2009, vol 19, no 7, pp 1316-1323, abstract	2, 3/2

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

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

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

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

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

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

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

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

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

"&" document member of the same patent family

Date of the actual completion of the international search

26 October 2018

Date of mailing of the international search report

14 NOV 2018

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer:

Lee W. Young

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 18/41041

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

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

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

☐ in the form of an Annex C/ST.25 text file.

☐ on paper or in the form of an image file.

b. ☐ furnished together with the international application under PCT Rule 13*ter*.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.

c. ☐ furnished subsequent to the international filing date for the purposes of international search only:

☐ in the form of an Annex C/ST.25 text file (Rule 13*ter*.1(a)).

☐ on paper or in the form of an image file (Rule 13*ter*.1(b) and Administrative Instructions, Section 713).

2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.

3. Additional comments:

The electronic sequence listing submitted on 20 August 2018 in response to the PCT/ISA/225 issued on 20 July 2018 is acknowledged; however, it does not comply with the standard provided for in Annex C of the Administrative Instructions. Therefore, the international search was not carried out on the basis of that non-compliant electronic sequence listing.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US 18/41041

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☒ Claims Nos.: 4-18, 22-29
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.