



(51) International Patent Classification:

C07D 205/04 (2006.01) C12N 9/22 (2006.01)

(21) International Application Number:

PCT/US2019/047362

(22) International Filing Date:

20 August 2019 (20.08.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/765,352 20 August 2018 (20.08.2018) US

62/773,694 30 November 2018 (30.11.2018) US

(71) Applicants: THE BROAD INSTITUTE, INC. [US/US];

415 Main Street, Cambridge, Massachusetts 02142 (US).

THE BRIGHAM AND WOMEN'S HOSPITAL, INC.

[US/US]; 75 Francis Street, Boston, Massachusetts 02115

(US).

(72) Inventors; and

(71) Applicants: CHOUDHARY, Amit [US/US]; c/o 75 Fran-

cis Street, Boston, Massachusetts 02115 (US). COX, Kurt

[US/US]; c/o 415 Main Street, Cambridge, Massachusetts

02142 (US).

(74) Agent: RUTLEDGE, Rachel D. 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, 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).

(54) Title: 3-PHENYL-2-CYANO-AZETIDINE DERIVATIVES, INHIBITORS OF RNA-GUIDED NUCLEASE ACTIVITY

(57) Abstract: 3-Phenyl-2-cyano-azetidine derivatives and methods are provided for the inhibition of the function of RNA guided endonucleases, including the identification and use of such inhibitors. Methods of identifying inhibitor compounds of RNA guided nucleases, assays for detecting nuclease activity and compounds for use therein are also provided.



WO 2020/041384 A1

3-PHENYL-2-CYANO-AZETIDINE DERIVATIVES, INHIBITORS OF RNA-GUIDED NUCLEASE ACTIVITY

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 62/765,352, filed on August 20, 2018; and U.S. Provisional Application No. 62/773,694 filed on November 30, 2018. The entire contents of the above-identified applications are hereby fully incorporated herein by reference.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0002] The contents of the electronic sequence listing (BROD_4420_ST25.txt"; Size is 37,665 bytes and it was created on July 25, 2019) is herein incorporated by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0003] This invention was made with government support under Grant No. AI126239-01 awarded by the National Institutes of Health, N66001-17-2-4055 awarded by the Defense Advanced Research Projects Agency, and Grant No. Q911NF1610586 awarded by the United States Department of the Army. The government has certain rights in the invention.

TECHNICAL FIELD

[0004] The subject matter disclosed herein is generally directed to compositions and methods for the inhibition of the function of RNA guided endonucleases, including the identification and use of such inhibitors.

BACKGROUND

[0005] The CRISPR (clustered regularly interspaced short palindromic repeat) system is an adaptive immune system used by bacteria and archaea to defend against invading phages or mobile genetic elements. The most studied CRISPR system employs an RNA-guided endonuclease Cas9, which can cleave double-stranded target DNA in multiple cell types. Cas9 identifies the target sequence by two recognition mechanisms: (i) Watson-Crick base-pairing between the target DNA sequence and guide RNA and (ii) Protospacer Adjacent

Motif (PAM) sequence on the target DNA. Upon target recognition, Cas9 induces double-strand breaks in the target gene, which when repaired by non-homologous end joining (NHEJ) can result in frameshift mutations and gene knockdown. Alternatively, homology-directed repair (HDR) at the double-strand break site can allow insertion of a desired sequence.

[0006] Two common variants of Cas9 are SpCas9 and SaCas9, which naturally occur in *S. pyogenes* and *S. aureus*, respectively, and recently another endonuclease called Cas12a (Cpf1) has been reported. The relative ease of targeting Cas9/Cas12a to specific genomic loci has enabled the development of revolutionary biomedical technologies. For example, catalytically inactive Cas9 (called dCas9), when fused to transcriptional activators, has enabled genome-wide screening of gene targets. Further, by targeting dCas9 to the promoter or exonic sequences, transcriptional repression has been accomplished. In yet another example, a fusion of dCas9 to acetyltransferases has enabled epigenome editing. Imaging of specific genomic loci has been accomplished by fusing dCas9 to GFP.

[0007] There are multiple reasons to establish controls on CRISPR activity. First, as described by Paracelsus' "The dose makes the poison", dosable control of the therapeutic activity is important for effective therapeutic strategies. Indeed, Cas9 exhibits undesirable off-target editing and chromosomal translocations when present at high concentrations. Second, many gene delivery systems have constitutively active Cas9, which is important to be terminated rapidly following on-target gene-editing. Third, Cas9-based technologies (e.g., transcriptional regulation) would benefit from dosable and temporal control of Cas9 activity.

[0008] The rapid ascension of CRISPR-based genome editing technologies has raised serious biosafety and bioterrorism concerns, leading to calls for a moratorium and responsible conduct. In particular, much concern has surrounded CRISPR-based gene drives. In sexual reproduction, the progenies receive two versions of a gene, one from each parent. Gene drives enable replacement of one version of the gene with the other "selfish" version of the gene, thereby converting a heterozygous individual to homozygous individual. In laboratory settings, CRISPR-based gene drives have successfully enabled self-propagation of engineered genes in multiple organisms (e.g., mosquitoes) and complete annihilation of wild-type genes. For example, using gene drives engineered mosquitoes have been generated that can wipe out the entire species by ensuring that every female progeny is infertile. Gene drives can be used to propagate a particular trait in the entire ecosystem, which may find use in the

elimination of diseases (e.g., malaria, dengue fever) or invasive species, and reversing pesticide resistance in plants. On the other hand, there exists the malevolent use of gene drives in entomological and agricultural settings.

[0009] Reports of small-molecule controlled Cas9 activity are present in literature and involve fusing Cas9 to small-molecule controlled protein domains. Genetic-fusions of Cas9 to small-molecule controlled degrons (e.g., Wandless' destabilized domains) may allow aforementioned controls, but such fusions often have high background activity presumably owing to the large size of Cas9. These systems also do not ensure dosage control—the small molecules act merely as an inducer of Cas9 activity. Further, these “inducer” small molecules cannot control gene drives containing wild-type Cas9/Cas12a. A general approach would be desirable to control all variants of Cas9/Cas12a, including the wild type and engineered versions. The use of “inducible” systems to control gene drives is also questionable given that the “inducer” small molecules may be toxic at the organismal level (albeit not at the cellular level, where these systems were developed).

[0010] More importantly, large-sized genetic-fusion constructs are incompatible with the most common Cas9 gene delivery systems under investigation for therapeutic gene therapy. The application of these “inducible systems” in a therapeutic setting will be challenging as they involve fusion of large genes to Cas9 gene. Since Cas9 is a large protein, fitting even Cas9 gene into virus delivery systems (e.g., AAV) has been an enormous challenge. Even the smallest of the small-molecule controlled systems will magnify the delivery problems. Finally, many small-molecule “inducible” Cas9 constructs exist, but none allow dosable control. The reported “inducible” systems are not reversible upon removal of the small molecule, and therefore, do not allow complete temporal control.

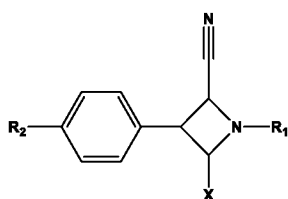
[0011] Accordingly, a need exists for compositions and methods for inhibiting one or more activities of RNA guided endonucleases (e.g., Cas9, Cas12a). Such compositions and methods are useful for regulating the activity of RNA guided endonucleases (e.g., in genome editing).

SUMMARY

[0012] In certain example embodiments, compositions and methods for inhibiting the activity of RNA guided endonucleases (e.g., Cas9, Cas12a), and methods of use therefor are

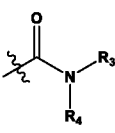
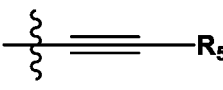
provided, including rapid, reversible, dosable, and/or temporal control of RNA guided endonuclease technologies. The compositions include broad-spectrum inhibitors of multiple CRISPR proteins. Also provided are high-throughput biochemical and cellular assays for detecting one or more activities of RNA guided endonucleases, and methods of using them to identify or screen agents that inhibit RNA guided nucleases.

[0013] In an aspect, the invention provides a method of inhibiting the activity of an RNA guided endonuclease comprising contacting the RNA guided endonuclease or a complex comprising the RNA guided endonuclease and a guide with a compound having the structure of Formula I:



wherein:

R₁ is selected from hydrogen, COR₆, CHR₇R₈, or SO₂R₉;

R₂ is selected from hydrogen, , , or a cyclic structure selected from cycloalkane, cycloalkene, aryl, or heteroaryl, each optionally substituted with one or more groups R₁₀;

R₃ is selected from hydrogen, C₁-C₁₂ alkyl, aryl, alkylaryl, optionally substituted with one or more groups selected from halogen, alkyl, alkoxy, or amine;

R₄ is selected from hydrogen, C₁-C₁₂ alkyl, aryl, alkylaryl, optionally substituted with one or more groups selected from halogen, alkyl, alkoxy, or amine;

R₅ is selected from C₁-C₁₂ alkyl and a cyclic structure selected from cycloalkane, cycloalkene, aryl, or heteroaryl, each optionally substituted with one or more groups R₁₁;

R₆ independently selected at each occurrence from C₁₋₁₂ alkyl, alkylcycloalkane, cycloalkane, aryl, heteroaryl, alkylaryl, alkylheteroaryl, or alkylamine;

R₇ is selected from hydrogen, C₁₋₁₂ alkyl, aryl, alkylaryl, hydroxyl, amine, alkoxy, or halogen;

R₈ is selected from hydrogen, C₁₋₁₂ alkyl, aryl, alkylaryl, hydroxyl, amine, alkoxy, or halogen;

R₉ is selected from C₁₋₁₂ alkyl or a cyclic structure selected from cycloalkane, cycloalkene, aryl, or heteroaryl, each optionally substituted with one or more groups R₁₂;

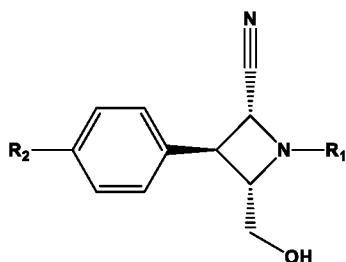
R₁₀ is independently selected at each occurrence from halogen, C₁-C₁₂ alkyl, haloalkyl, or alkoxy;

R₁₁ is independently selected at each occurrence from halogen, C₁-C₁₂ alkyl, haloalkyl, or alkoxy;

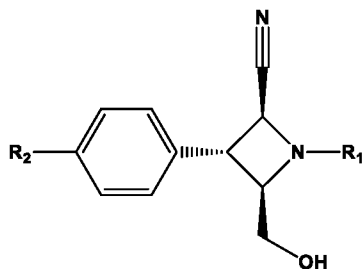
R₁₂ is independently selected at each occurrence from hydrogen, halogen, C₁-C₃ alkyl, haloalkyl, or alkoxy; and

X is selected from =O, hydroxyl, alkylalcohol, amine, or alkylamine.

[0014] In certain embodiments, the compound of Formula I has the structure:

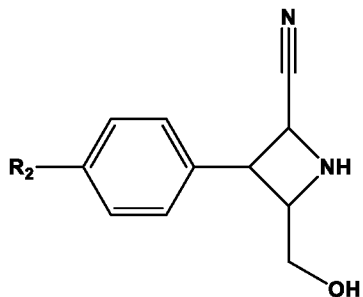


[0015] In one such embodiment, the compound of Formula I has the structure:



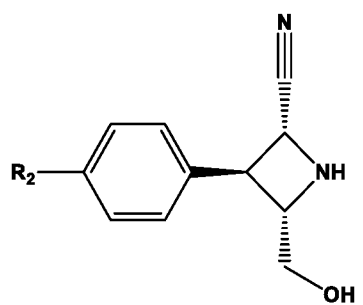
[0016]

[0017] In another embodiment, the compound of Formula I has the structure:



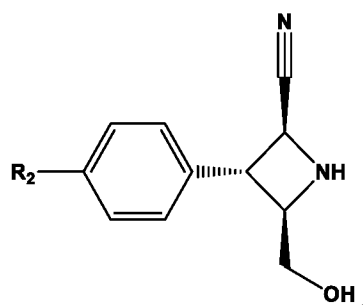
[0018]

[0019] In another embodiment, the compound of Formula I has the structure:



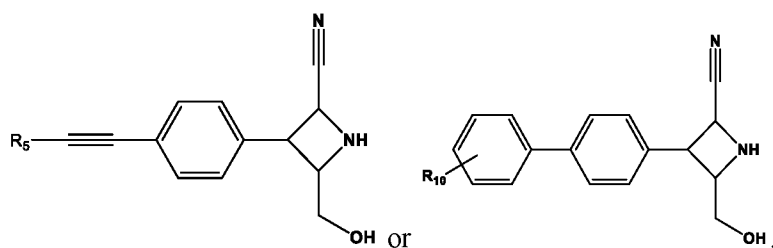
[0020]

[0021] In another embodiment, the compound of Formula I has the structure:



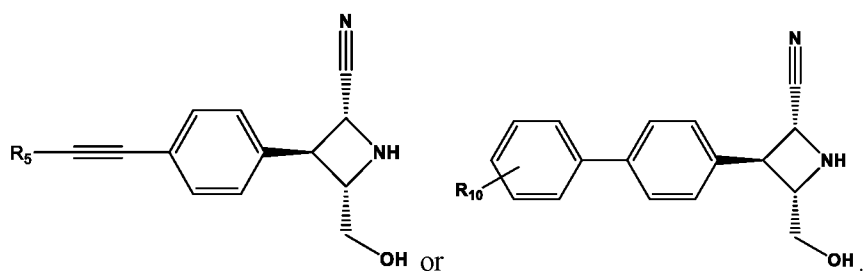
[0022]

[0023] In another embodiment, the compound of Formula I has the structure:



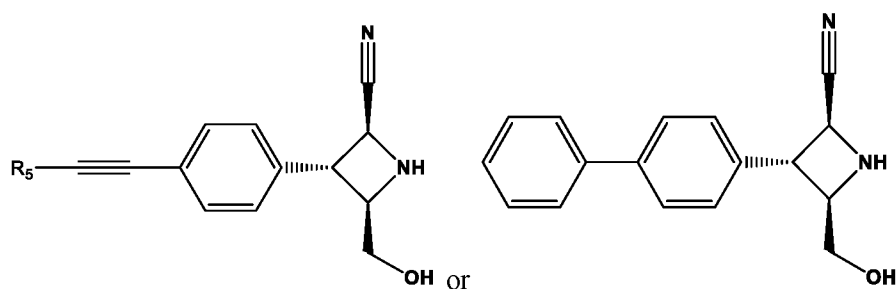
[0024]

[0025] In another embodiment, the compound of Formula I has the structure:



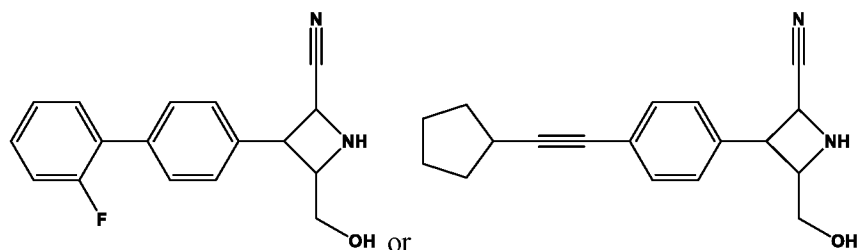
[0026]

[0027] In another embodiment, the compound of Formula I has the structure:



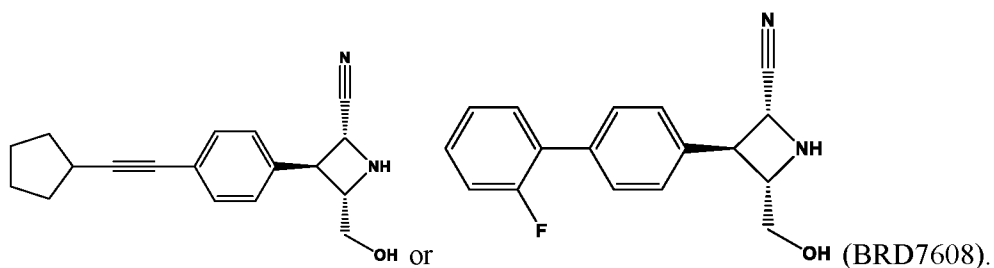
[0028]

[0029] In another embodiment, the compound of Formula I has the structure:



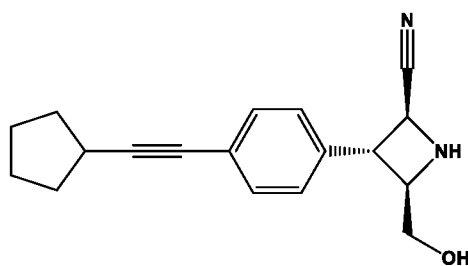
[0030]

[0031] In another embodiment, the compound of Formula I has the structure:



[0032]

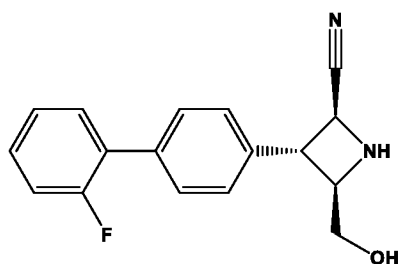
[0033] In another embodiment, the compound of Formula I has the structure:



[0034]

(BRD9822)

or



[0035] In certain embodiments, the method is performed *in vitro* or *in vivo*, for example in a cell. In certain embodiments, the cell is a prokaryotic cell, for example a bacterium. In

other embodiments, the cell is a eukaryotic cell, for example, a human, mammalian, insect, plant, or yeast cell.

[0036] In certain embodiments, the cell is in an organism, such as but not limited to a human, mammal, vertebrate, invertebrate, insect, or plant.

[0037] In certain embodiments, the RNA-guided endonuclease is Cas9, such as but not limited to SpCas9 or SaCas9. In certain embodiments, the RNA-guided endonuclease is Cas12a, such as but not limited to FnCas12a. The compounds include broad-spectrum inhibitors of multiple CRISPR proteins.

[0038] In another aspect, the invention provides a method of inhibiting the activity of an RNA guided endonuclease-guide complex, which comprises contacting the RNA guided endonuclease-guide complex with a small molecule. In certain embodiments of the invention, the small molecule is a compound having the structure of Formula I.

[0039] In certain embodiments, the invention provides a method of detecting nuclease activity of an RNA guided endonuclease-guide complex comprising: contacting a) an RNA guided endonuclease-guide complex, b) a double-stranded probe comprising a first oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 5' terminus and a second oligonucleotide complementary to the first oligonucleotide, wherein the second oligonucleotide is designed to hybridize to the guide whereby the first oligonucleotide and the second oligonucleotide are cleaved by the endonuclease, and c) a third oligonucleotide comprising a fluorescence quencher at the 3' terminus and designed to hybridize to and displace the cleaved first oligonucleotide from the cleaved second oligonucleotide, and detecting a decrease in fluorescence polarization relative to a reference, thereby detecting nuclease activity of the RNA guided endonuclease-guide complex.

[0040] The method is particularly useful for an RNA guided endonuclease system such as Cas9 wherein the PAM is located 3' to the target sequence along the non-target strand (NTS) of the DNA.

[0041] In certain embodiments, the invention provides a method of detecting nuclease activity of an RNA guided endonuclease-guide complex comprising: contacting a) an RNA guided endonuclease-guide complex, b) a double-stranded probe comprising a first oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 3' terminus and a second oligonucleotide complementary to the first oligonucleotide, wherein the second oligonucleotide is designed to hybridize to the guide whereby the first oligonucleotide and

the second oligonucleotide are cleaved by the endonuclease, and c) a third oligonucleotide comprising a fluorescence quencher at the 5' terminus and designed to hybridize to and displace the cleaved first oligonucleotide from the cleaved second oligonucleotide, and detecting a decrease in fluorescence polarization relative to a reference, thereby detecting nuclease activity of the RNA guided endonuclease-guide complex.

[0042] The method is particularly useful for an RNA guided endonuclease system such as Cas12a (Cpf1) wherein the PAM is located 5' to the target sequence along the non-target strand (NTS) of the DNA.

[0043] In another aspect, the invention provides a method of identifying an agent that inhibits the nuclease activity of an RNA guided endonuclease-guide complex.

[0044] In certain embodiments, the invention provides a method of identifying an agent that inhibits nuclease activity of an RNA guided endonuclease-guide complex comprising: contacting an RNA guided endonuclease-guide complex with an agent in the presence of a double-stranded probe comprising a first oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 5' terminus and a second oligonucleotide complementary to the first oligonucleotide, wherein the second oligonucleotide is designed to hybridize to the guide whereby the first oligonucleotide and the second oligonucleotide are cleaved by the nuclease, and a third oligonucleotide comprising a fluorescence quencher at the 3' terminus and designed to hybridize to and displace the cleaved first oligonucleotide from the cleaved second oligonucleotide, and detecting a decrease in the reduction of fluorescence polarization relative to a reference, thereby identifying the agent as an inhibitor of the nuclease activity of the RNA guided endonuclease-guide complex.

[0045] The method is particularly useful for identifying an inhibitor of an RNA guided endonuclease system such as Cas9 wherein the PAM is located 3' to the target sequence along the non-target strand (NTS) of the DNA.

[0046] In certain embodiments, the invention provides a method of identifying an agent that inhibits nuclease activity of an RNA guided endonuclease-guide complex comprising: contacting the RNA guided endonuclease-guide complex with an agent in the presence of a double-stranded probe comprising a first oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 3' terminus and a second oligonucleotide complementary to the first oligonucleotide, wherein the second oligonucleotide is designed to hybridize to the guide whereby the first oligonucleotide and the second oligonucleotide are cleaved by the

nuclease, and a third oligonucleotide comprising a fluorescence quencher at the 5' terminus and designed to hybridize to and displace the cleaved first oligonucleotide from the cleaved second oligonucleotide, and detecting a decrease in the reduction of fluorescence polarization relative to a reference, thereby identifying the agent as an inhibitor of the nuclease activity of the RNA guided endonuclease-guide complex.

[0047] The method is particularly useful for identifying an inhibitor of an RNA guided endonuclease system such as Cas12a (Cpf1) wherein the PAM is located 5' to the target sequence along the non-target strand (NTS) of the DNA.

[0048] In various embodiments, the double-stranded probe, i.e., the labeled strand of the double-stranded probe comprises an oligonucleotide comprising:

[0049] 5Alex647N/TAATACGACTCACTATAGGACGCGACCGAAATGGTGAAGGACGGGT-3' (SEQ ID NO:1).

[0050] 5Alex647N indicates linkage of a Alexa Fluor® 647 fluorophore. In other embodiments, other fluorophores can be employed.

[0051] In certain such embodiments, the complementary oligonucleotide comprises:

[0052] 5'-ACCCGTCCTTCACCA**TTTCGGTCGCGTCCTATAGT**GAGTCGTATTA-3' (SEQ ID NO:2)

[0053] In certain such embodiments, the displacer oligonucleotide comprises:

[0054] 5'-ATAGTGAGTCGTATTA/3IAbRQSp/ (SEQ ID NO:3)

[0055] 3IAbRQSp indicates linkage of a quencher. In other embodiments, other quenchers can be employed.

[0056] In the above sequences, the PAM is bold, protospacer italicized, and displacer binding site underlined.

[0057] 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

[0058] An understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention may be utilized, and the accompanying drawings of which:

[0059] **FIG. 1A-1B:** Fluorescence-based strand displacement assay (SDA). **FIG. 1A** Shown is a schematic of a fluorescence-based SDA for monitoring Cas9 nuclease activity. Following Cas9 cleavage, a fluorophore bearing double stranded oligo (DS-oligo) is displaced by a quencher (Q)-bearing displacer strand (Q-oligo), resulting in a decrease in fluorescent signal. The top cut is made by the HNH nuclease domain, and the bottom cut is made by the RuvC nuclease domain in both SpCas9 and SaCas9. **FIG. 1B** DS-fluor is not quenched in the presence of Disp-Q unless the duplex is disrupted by cleavage via an active SpCas9:gRNA complex (RNP).

[0060] **FIG. 2:** Overview of the exemplary screening protocol. Counterscreen steps and primary screen steps are as indicated. Plates are read on an Envision, preserving plate order throughout. Not all screens involve filtering by a kinetic stability step.

[0061] **FIG. 3A-3B:** Demonstration of SDA: **FIG. 3A** A PAM-sequence is required for quenching in the SDA. **FIG. 3B** Validation of SDA-substrate activity using a gel cleavage assay.

[0062] **FIG. 4:** Determining the Limit of Detection (LOD) of the SpCas9 DS-AF647 substrate/RedQ displacer strand pair. Z' is Z prime, and +/- is fold change between SS/Q and DS/Q. n=3

[0063] **FIG. 5:** Optimization of the ratio of SpCas9 to DS-AF647 and the time course of the reaction. DS-AF647 and RedQ were fixed at 1 nM and 5 nM, resp. n=3

[0064] **FIG. 6:** Kinetics of the SpCas9 SDA reaction at various ratios of SpCas9/DS-AF647 (related to Figure 6.2). n=3

[0065] **FIG. 7:** Optimizing the ratio of RedQ quencher strand to DS-AF647, and the time course of the reaction. DS-AF647 and SpCas9 were fixed at 1 nM and 5 nM, resp. n=3

[0066] **FIG. 8:** Fine tuning the optimized conditions at the Combi distribution level (n=32, all reagents were dispensed via Combi head to simulate the native screens).

[0067] **FIG. 9A-9C:** Substrate quenching and anti-CRISPR inhibition. **FIG. 9A** Time course of SpCas9/gRNA cleavage and substrate quenching at various DS-fluor:SpCas9/gRNA ratios. **FIG. 9B** Anti-CRISPR protein inhibition of SpCas9. **FIG. 9C** Z-prime factor of the SDA.

[0068] **FIG. 10A-10D.** Generalization to CRISPRs. **FIG. 10A** SDA assay readily translates to SaCas9. **FIG. 10B** The Cpf1/Cas12 protein/DNA complex has a different structure from Cas9 and requires optimization of the potential fluorophore attachment sites.

FIG. 10C Gel-cleavage screening of fluorophore attachment sites (NTS = non-targeting strand; TS = targeting strand) with different Cpf1 orthologs. Only FnCpf1 produced a resolvable product after cleavage. **FIG. 10D** Fluorophore attachment to the non-target strand (NTS) yields robust quenching compared to the target strand (TS). Activity is dependent on a TTTN PAM.

[0069] **FIG. 11:** Small molecule screening against Cas nucleases: Small molecule screening against SpCas9 (50,000 cpds), SaCas9 (10,000 cpds), and FnCpf1 (FnCas12) (25,000 cpds) is shown. Apo nuclease was used as an “inhibited” positive control (circles, appearing primarily in upper right quadrant of graph), and DMSO was used as the negative control (circles, appearing primarily in lower left quadrant of graph).

[0070] **FIG. 12A-12C:** An eGFP screen to identify cell-permeable inhibitors. **FIG. 12A** Breakdown of the hit rates from each subset of the DOS library. The azetidine nitrile library (~1000 compounds) was screened in U2OS cells using the eGFP assay (inset). SpCas9/gRNA complexes were either introduced through **FIG. 12B** plasmid or **FIG. 12C** ribonucleoprotein (RNP) nucleofection.

[0071] **FIG. 13:** 20 anti-CRISPR compounds active in both RNP and plasmid eGFP screens ordered by R₁ and R₂ group size are depicted.

[0072] **FIG. 14A-14C:** Activity and SAR of two SpCas9 inhibitors. **FIG. 14A** Structures of the top compound hits from the eGFP screens. **FIG. 14B** Retest of BRD7608 and BRD9822 in dose using the eGFP assay. Both plasmid and ribonucleoprotein (RNP) nucleofection were used to introduce SpCas9/gRNA into cells. **FIG. 14C** SAR of BRD7608 and BRD9822, based on the results of the eGFP screen.

[0073] **FIG. 15A-15B:** SAR of BRD7608 and BRD9822 orthologs. **FIG. 15A** N-capping (R₁) and aryl substituent (R₂) R groups. **FIG. 15B** Activity reductions compared to BRD7608 and BRD9822 based on eGFP screen results.

[0074] **FIG. 16A-16C:** Inhibition of SpCas9 by BRD7608 and BRD9822 in cells. **FIG. 16A** SpCas9 mediated knockdown of eGFP is reduced by inhibitors of SpCas9. **FIG. 16B** Dose dependent inhibition of SpCas9 mediated eGFP knockdown by BRD7608 (IC₅₀ = 3.9 μM) and by BRD9822 (IC₅₀ = 5.5 μM). **FIG. 16C** eGFP fluorescence is maintained in the presence of eGFP-targeted SpCas9 activity with inhibitors BRD7608 or BRD9822.

[0075] **FIG. 17A-17B:** Inhibition of CRISPR-mediated gene tagging. **FIG. 17A** depicts CRISPR mediated gene tagging with HiBiT, which is an 11 amino acid peptide tag that can

be attached to any protein and detected by bioluminescence. **FIG. 17B** shows bioluminescence resulting from recombination of the ssODN encoding HiBiT (DMSO) and inhibition by various concentrations of BRD7608 and BRD9822.

[0076] **FIG. 18A-18B**: Biochemical target engagement confirmed by **FIG. 18A** ¹⁹F- and **FIG. 18B** STD-NMR.

[0077] **FIG. 19A-19B**: Structure activity relationship of BRD7608 and compounds varying by substituent at R₁ or R₂ **FIG. 19A** or by stereochemistry **FIG. 19B**. * $P \leq 0.05$; *** $P \leq 0.001$

[0078] **FIG. 20A-20B**: Structure activity relationship of BRD9822 and compounds varying by substituent at R₁ or R₂ **FIG. 20A** or by stereochemistry **FIG. 20B** * $P \leq 0.05$; *** $P \leq 0.001$

[0079] **FIG. 21**: Library screening by eGFP fluorescence.

[0080] **FIG. 22A-22B**: Broad spectrum inhibition. BRD7608 and BRD9822 similarly inhibit SpCas9 **FIG. 22A** and FnCpf1 **FIG. 22B**.

[0081] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

General Definitions

[0082] 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) .

[0083] As used herein, the singular forms “a”, “an”, and “the” include both singular and plural referents unless the context clearly dictates otherwise.

[0084] 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.

[0085] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0086] 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.

[0087] As used herein, a “biological sample” may contain whole cells and/or live cells and/or cell debris. The biological sample may contain (or be derived from) a “bodily fluid”. The present invention encompasses embodiments wherein the bodily fluid is selected from amniotic fluid, aqueous humour, vitreous humour, bile, blood serum, breast milk, cerebrospinal fluid, cerumen (earwax), chyle, chyme, endolymph, perilymph, exudates, feces, female ejaculate, gastric acid, gastric juice, lymph, mucus (including nasal drainage and phlegm), pericardial fluid, peritoneal fluid, pleural fluid, pus, rheum, saliva, sebum (skin oil), semen, sputum, synovial fluid, sweat, tears, urine, vaginal secretion, vomit and mixtures of one or more thereof. Biological samples include cell cultures, bodily fluids, cell cultures from bodily fluids. Bodily fluids may be obtained from a mammal organism, for example by puncture, or other collecting or sampling procedures.

[0088] The terms “subject,” “individual,” and “patient” are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, and pets. Tissues, cells and their progeny of a biological entity obtained in vivo or cultured in vitro are also encompassed.

[0089] It will be understood that "substitution" or "substituted with" includes the implicit proviso that such substitution is in accordance with permitted valence of the substituted atom and the substituent, and that the substitution results in a stable compound, e.g., which does not spontaneously undergo transformation such as by rearrangement, cyclization, elimination, etc. In a broad aspect, the permissible substituents include acyclic and cyclic, branched and unbranched, carbocyclic and heterocyclic, aromatic and nonaromatic substituents of organic compounds. The permissible substituents can be one or more and the same or different for appropriate organic compounds. For purposes of this disclosure, the heteroatoms such as nitrogen may have hydrogen substituents and/or any permissible substituents of organic compounds described herein which satisfy the valences of the heteroatoms. Non-limiting examples of optional substituents as referred to herein include halogen, alkyl, aralkyl, heteroalkyl, alkenyl, alkynyl, cycloalkyl, hydroxyl, alkoxy, amino, amido, nitro, cyano, amido, carbonyl, carboxyl, silyl, ether, alkylthio, sulfonyl, sulfonamido, ketone, aldehyde, ester, heterocyclyl, aryl, and heteroaryl.

[0090] As used herein, the term “alkyl”, used either alone or in compound words such as “haloalkyl” includes straight-chain or branched alkyl, such as methyl, ethyl, n-propyl, i-propyl, or the different butyl, pentyl or hexyl isomers, etc.

[0091] As used herein, the term “alkoxy” includes, for example, methoxy, ethoxy, n-propyloxy, isopropyloxy and the different butoxy, pentoxy and hexyloxy isomers.

[0092] As used herein, the term “cycloalkane” denotes a cyclic saturated hydrocarbon and includes, for example, cyclopropane, cyclobutane, cyclopentane, cyclohexane, cycloheptane, cyclooctane, etc.

[0093] As used herein, the term “cycloalkene” is a cyclic unsaturated hydrocarbon, i.e., the cyclic structure contains at least one double bond, and includes, for example, cyclopropene, cyclobutene, cyclopentene, cyclohexene, cycloheptene, cyclooctene, etc.

[0094] As used herein, the term “aryl” used alone or as part of a larger moiety as in “alkylaryl” refers to aromatic ring groups having five to fourteen members, such as phenyl,

benzyl, phenethyl, 1-naphthyl, 2-naphthyl, 1-anthracyl and 2-anthracyl. The term “aryl” also refers to rings that are optionally substituted. The term “aryl” may be used interchangeably with the term “aryl ring”. “Aryl” also includes fused polycyclic aromatic ring systems in which an aromatic ring is fused to one or more rings. Examples include 1-naphthyl, 2-naphthyl, 1-anthracyl and 2-anthracyl. Also included within the scope of the term “aryl”, as it is used herein, is a group in which an aromatic ring is fused to one or more non-aromatic rings, such as in an indanyl, phenanthridinyl, or tetrahydronaphthyl, where the radical or point of attachment is on the aromatic ring.

[0095] The term “heteroaryl”, used alone or as part of a larger moiety as in “heteroaralkyl” or “heteroarylalkoxy”, refers to heteroaromatic ring groups having five to fourteen members. Examples of heteroaryl rings include 2-furanyl, 3-furanyl, 3-furazanyl, N-imidazolyl, 2-imidazolyl, 4-imidazolyl, 5-imidazolyl, 3-isoxazolyl, 4-isoxazolyl, 5-isoxazolyl, 2-oxadiazolyl, 5-oxadiazolyl, 2-oxazolyl, 4-oxazolyl, 5-oxazolyl, 1-pyrrolyl, 2-pyrrolyl, 3-pyrrolyl, 1-pyrazolyl, 2-pyrazolyl, 3-pyrazolyl, 2-pyridyl, 3-pyridyl, 4-pyridyl, 2-pyrimidyl, 4-pyrimidyl, 5-pyrimidyl, 3-pyridazinyl, 2-thiazolyl, 4-thiazolyl, 5-thiazolyl, 5-tetrazolyl, 2-triazolyl, 5-triazolyl, 2-thienyl, 3-thienyl, carbazolyl, benzimidazolyl, benzothienyl, benzofuranyl, indolyl, quinoliny, benzotriazolyl, benzothiazolyl, benzoisoxazolyl, benzimidazolyl, isoquinoliny, indazolyl, isoindolyl, acridinyl, or benzoisoxazolyl. Also included within the scope of the term “heteroaryl”, as it is used herein, is a group in which a heteroatomic ring is fused to one or more aromatic or nonaromatic rings where the radical or point of attachment is on the heteroaromatic ring. Examples include tetrahydroquinoliny, tetrahydroisoquinoliny, and pyrido[3,4-d]pyrimidinyl. The term “heteroaryl” also refers to rings that are optionally substituted. The term “heteroaryl” may be used interchangeably with the term “heteroaryl ring” or the term “heteroaromatic”.

[0096] As used herein, the term “halogen” or “halo” alone or as part of another group refers to chlorine, bromine, fluorine, and iodine, with chlorine or fluorine being preferred.

[0097] As used herein, the term “amine” alone or as part of another group such as “alkylamine” can be primary amine, secondary amine, or tertiary amine. Amine can be substituted or unsubstituted as specified in the specification.

[0098] When any variable (*e.g.*, R₁) occurs more than one time in any constituent or formula for a compound, its definition at each occurrence is independent of its definition at every other occurrence. Thus, for example, if a group is shown to be substituted with one or

more R₁ moieties, then R₁ at each occurrence is selected independently from the Markush group recited for R₁. Also, combinations of substituents and/or variables are permissible, but only if such combinations result in stable compounds within a designated atom's normal valency.

[0099] As used herein, “unsaturated” refers to compounds or structures having at least one degree of unsaturation (e.g., at least one double or triple bond).

[0100] The term “acid addition salt” refers to a salt of a compound prepared by reaction of a compound with a mineral or organic acid. For exemplification of pharmaceutically acceptable acid addition salts, see, e.g., Berge, S.M., Bighley, L.D., and Monkhouse, D.C., *J. Pharm. Sci.*, 66:1, 1977. For example, amine compounds are basic in nature and accordingly react with any of a number of inorganic and organic acids to form pharmaceutically acceptable acid addition salts.

[0101] Pharmaceutically acceptable acid addition salts can be formed by the reaction of a disclosed compound with an equimolar or excess amount of acid. Alternatively, hemi-salts can be formed by the reaction of a compound with the desired acid in a 2:1 ratio, compound to acid. The reactants are generally combined in a mutual solvent such as diethyl ether, tetrahydrofuran, methanol, ethanol, *iso*-propanol, benzene, or the like. The salts normally precipitate out of solution within, e.g., about one hour to about ten days and can be isolated by filtration or other conventional methods.

[0102] Inorganic acids commonly employed to form such salts include hydrochloric acid, hydrobromic acid, hydroiodic acid, sulfuric acid, phosphoric acid, and the like. Organic acids commonly employed to form such salts include p-toluenesulfonic acid, methanesulfonic acid, oxalic acid, p-bromophenylsulfonic acid, carbonic acid, succinic acid, citric acid, benzoic acid, acetic acid and the like. Examples of such pharmaceutically acceptable salts thus are the sulfate, pyrosulfate, bisulfate, sulfite, bisulfite, phosphate, monohydrogenphosphate, dihydrogenphosphate, metaphosphate, pyrophosphate, chloride, bromide, iodide, acetate, propionate, decanoate, caprylate, acrylate, formate, *iso*-butyrate, caproate, heptanoate, propiolate, oxalate, malonate, succinate, hemisuccinate, suberate, sebacate, fumarate, maleate, butyne-1,4-dioate, hexyne-1,6-dioate, benzoate, chlorobenzoate, methylbenzoate, dinitrobenzoate, hydroxybenzoate, methoxybenzoate, phthalate, sulfonate, xylenesulfonate, phenylacetate, phenylpropionate, phenylbutyrate, citrate, lactate, β -hydroxybutyrate,

glycolate, tartrate, methanesulfonate, propanesulfonate, naphthalene-1-sulfonate, naphthalene-2-sulfonate, mandalate and the like.

[0103] By “agent” is meant any small molecule chemical compound, antibody, nucleic acid molecule, or polypeptide, or fragments thereof.

[0104] By “decreases” is meant a negative alteration. Such alterations are by 5%, 10%, 25%, 50%, 75%, 85%, 90% or even by 100% of a reference value.

[0105] By “effective amount” is meant the amount of an agent required to ameliorate the symptoms of a condition, disease or disorder relative to an untreated subject or organism. In particular embodiments, the subject or organism expresses an active RNA guided endonuclease polypeptide. The effective amount of active agent(s) used to practice the present invention varies depending upon the manner of administration, the age, body weight, and general health of the subject.

[0106] By “fragment” is meant a portion of a polypeptide or nucleic acid molecule. This portion contains, preferably, at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the entire length of the reference nucleic acid molecule or polypeptide. A fragment may contain 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100, 200, 300, 400, 500, 600, 700, 800, 900, or 1000 nucleotides or amino acids.

[0107] By “identity” is meant the amino acid or nucleic acid sequence identity between a sequence of interest and a reference sequence. Sequence identity is typically measured using sequence analysis software (for example, Sequence Analysis Software Package of the Genetics Computer Group, University of Wisconsin Biotechnology Center, 1710 University Avenue, Madison, Wis. 53705, BLAST, BESTFIT, GAP, or PILEUP/PRETTYBOX programs). Such software matches identical or similar sequences by assigning degrees of homology to various substitutions, deletions, and/or other modifications. Conservative substitutions typically include substitutions within the following groups: glycine, alanine; valine, isoleucine, leucine; aspartic acid, glutamic acid, asparagine, glutamine; serine, threonine; lysine, arginine; and phenylalanine, tyrosine. In an exemplary approach to determining the degree of identity, a BLAST program may be used, with a probability score between e^{-3} and e^{-100} indicating a closely related sequence.

[0108] By “hybridize” is meant pair to form a double-stranded molecule between complementary polynucleotide sequences (e.g., genes listed in Tables 1 and 2), or portions

thereof, under various conditions of stringency. (See, e.g., Wahl, G. M. and S. L. Berger (1987) *Methods Enzymol.* 152:399; Kimmel, A. R. (1987) *Methods Enzymol.* 152:507).

[0109] For example, stringent salt concentration will ordinarily be less than about 750 mM NaCl and 75 mM trisodium citrate, preferably less than about 500 mM NaCl and 50 mM trisodium citrate, and most preferably less than about 250 mM NaCl and 25 mM trisodium citrate. Low stringency hybridization can be obtained in the absence of organic solvent, e.g., formamide, while high stringency hybridization can be obtained in the presence of at least about 35% formamide, and most preferably at least about 50% formamide. Stringent temperature conditions will ordinarily include temperatures of at least about 30° C., more preferably of at least about 37° C., and most preferably of at least about 42° C. Varying additional parameters, such as hybridization time, the concentration of detergent, e.g., sodium dodecyl sulfate (SDS), and the inclusion or exclusion of carrier DNA, are well known to those skilled in the art. Various levels of stringency are accomplished by combining these various conditions as needed. In a preferred embodiment, hybridization will occur at 30° C. in 750 mM NaCl, 75 mM trisodium citrate, and 1% SDS. In a more preferred embodiment, hybridization will occur at 37° C. in 500 mM NaCl, 50 mM trisodium citrate, 1% SDS, 35% formamide, and 100 µg/ml denatured salmon sperm DNA (ssDNA). In a most preferred embodiment, hybridization will occur at 42° C. in 250 mM NaCl, 25 mM trisodium citrate, 1% SDS, 50% formamide, and 200 µg/ml ssDNA. Useful variations on these conditions will be readily apparent to those skilled in the art.

[0110] For most applications, washing steps that follow hybridization will also vary in stringency. Wash stringency conditions can be defined by salt concentration and by temperature. As above, wash stringency can be increased by decreasing salt concentration or by increasing temperature. For example, stringent salt concentration for the wash steps will preferably be less than about 30 mM NaCl and 3 mM trisodium citrate, and most preferably less than about 15 mM NaCl and 1.5 mM trisodium citrate. Stringent temperature conditions for the wash steps will ordinarily include a temperature of at least about 25° C., more preferably of at least about 42° C., and most preferably of at least about 68° C. In a preferred embodiment, wash steps will occur at 25° C. in 30 mM NaCl, 3 mM trisodium citrate, and 0.1% SDS. In a more preferred embodiment, wash steps will occur at 42° C. in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. In a most preferred embodiment, wash steps will occur at 68° C. in 15 mM NaCl, 1.5 mM trisodium citrate, and 0.1% SDS. Additional

variations on these conditions will be readily apparent to those skilled in the art. Hybridization techniques are well known to those skilled in the art and are described, for example, in Benton and Davis (Science 196:180, 1977); Grunstein and Hogness (Proc. Natl. Acad. Sci., USA 72:3961, 1975); Ausubel et al. (Current Protocols in Molecular Biology, Wiley Interscience, New York, 2001); Berger and Kimmel (Guide to Molecular Cloning Techniques, 1987, Academic Press, New York); and Sambrook et al., Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory Press, New York.

[0111] By “increases” is meant a positive alteration. Such alterations are by 5%, 10%, 25%, 50%, 75%, 85%, 90% or even by 100% of a reference value.

[0112] In some aspects, the compound is an isomer. "Isomers" are different compounds that have the same molecular formula. "Stereoisomers" are isomers that differ only in the way the atoms are arranged in space. As used herein, the term "isomer" includes any and all geometric isomers and stereoisomers. For example, "isomers" include geometric double bond cis- and trans-isomers, also termed E- and Z-isomers; R- and S-enantiomers; diastereomers, (d)-isomers and (l)-isomers, racemic mixtures thereof, and other mixtures thereof, as falling within the scope of this disclosure.

[0113] Geometric isomers can be represented by the symbol ---- which denotes a bond that can be a single, double or triple bond as described herein. Provided herein are various geometric isomers and mixtures thereof resulting from the arrangement of substituents around a carbon-carbon double bond or arrangement of substituents around a carbocyclic ring. Substituents around a carbon-carbon double bond are designated as being in the "Z" or "E" configuration wherein the terms "Z" and "E" are used in accordance with IUPAC standards. Unless otherwise specified, structures depicting double bonds encompass both the "Z" and "E" isomers.

[0114] Substituents around a carbon-carbon double bond alternatively can be referred to as "cis" or "trans," where "cis" represents substituents on the same side of the double bond and "trans" represents substituents on opposite sides of the double bond. The arrangement of substituents around a carbocyclic ring can also be designated as "cis" or "trans." The term "cis" represents substituents on the same side of the plane of the ring, and the term "trans" represents substituents on opposite sides of the plane of the ring. Mixtures of compounds wherein the substituents are disposed on both the same and opposite sides of plane of the ring are designated "cis/trans."

[0115] The term "enantiomers" refers to a pair of stereoisomers that are non-superimposable mirror images of each other. An atom having an asymmetric set of substituents can give rise to an enantiomer. A mixture of a pair of enantiomers in any proportion can be known as a "racemic" mixture. The term "(±)" is used to designate a racemic mixture where appropriate. "Diastereoisomers" are stereoisomers that have at least two asymmetric atoms, but which are not mirror-images of each other. The absolute stereochemistry is specified according to the Cahn-Ingold-Prelog R-S system. When a compound is an enantiomer, the stereochemistry at each chiral carbon can be specified by either R or S. Resolved compounds whose absolute configuration is unknown can be designated (+) or (-) depending on the direction (dextro- or levorotatory) which they rotate plane polarized light at the wavelength of the sodium D line. Certain of the compounds described herein contain one or more asymmetric centers and can thus give rise to enantiomers, diastereomers, and other stereoisomeric forms that can be defined, in terms of absolute stereochemistry at each asymmetric atom, as (R)- or (S)-. The present chemical entities, pharmaceutical compositions and methods are meant to include all such possible isomers, including racemic mixtures, optically substantially pure forms and intermediate mixtures.

[0116] Optically active (R)- and (S)-isomers can be prepared, for example, using chiral synthons or chiral reagents, or resolved using conventional techniques. Enantiomers can be isolated from racemic mixtures by any method known to those skilled in the art, including chiral high pressure liquid chromatography (HPLC), the formation and crystallization of chiral salts, or prepared by asymmetric syntheses.

[0117] Optical isomers can be obtained by resolution of the racemic mixtures according to conventional processes, e.g., by formation of diastereoisomeric salts, by treatment with an optically active acid or base. Examples of appropriate acids are tartaric, diacetyltartaric, dibenzoyltartaric, ditoluoyltartaric, and camphorsulfonic acid. The separation of the mixture of diastereoisomers by crystallization followed by liberation of the optically active bases from these salts affords separation of the isomers. Another method involves synthesis of covalent diastereoisomeric molecules by reacting disclosed compounds with an optically pure acid in an activated form or an optically pure isocyanate. The synthesized diastereoisomers can be separated by conventional means such as chromatography, distillation, crystallization or sublimation, and then hydrolyzed to deliver the enantiomerically enriched compound.

Optically active compounds can also be obtained by using active starting materials. In some embodiments, these isomers can be in the form of a free acid, a free base, an ester or a salt.

[0118] In certain embodiments, a disclosed compound can be a tautomer. As used herein, the term "tautomer" is a type of isomer that includes two or more interconvertible compounds resulting from at least one formal migration of a hydrogen atom and at least one change in valency (e.g., a single bond to a double bond, a triple bond to a single bond, or vice versa). "Tautomerization" includes prototropic or proton-shift tautomerization, which is considered a subset of acid-base chemistry. "Prototropic tautomerization" or "proton-shift tautomerization" involves the migration of a proton accompanied by changes in bond order. The exact ratio of the tautomers depends on several factors, including temperature, solvent, and pH. Where tautomerization is possible (e.g., in solution), a chemical equilibrium of tautomers can be reached. Tautomerizations (i.e., the reaction providing a tautomeric pair) can be catalyzed by acid or base, or can occur without the action or presence of an external agent. Exemplary tautomerizations include, but are not limited to, keto-to-enol; amide-to-imide; lactam-to-lactim; enamine-to-imine; and enamine-to-(a different) enamine tautomerizations. A specific example of keto-enol tautomerization is the interconversion of pentane-2,4-dione and 4-hydroxypent-3-en-2-one tautomers. Another example of tautomerization is phenol-keto tautomerization. A specific example of phenol-keto tautomerization is the interconversion of pyridin-4-ol and pyridin-4(1H)-one tautomers.

[0119] All chiral, diastereomeric, racemic, and geometric isomeric forms of a structure are intended, unless specific stereochemistry or isomeric form is specifically indicated. All processes used to prepare compounds and intermediates made therein are encompassed by the present disclosure. All tautomers of shown or described compounds are also encompassed by the present disclosure.

[0120] The terms "pharmaceutical" or "pharmaceutically acceptable", when used herein as an adjective, mean substantially non-toxic and substantially non-deleterious to the subject.

[0121] As used herein, "pharmaceutically acceptable carrier" includes any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents and the like. The pharmaceutically acceptable carrier or excipient does not destroy the pharmacological activity of the disclosed compound and is nontoxic when administered in doses sufficient to deliver a therapeutic amount of the compound. The use of such media and agents for pharmaceutically active substances is well known in the art.

Except insofar as any conventional media or agent is incompatible with the active ingredient, its use in the therapeutic compositions as disclosed herein is contemplated. Non-limiting examples of pharmaceutically acceptable carriers and excipients include sugars such as lactose, glucose and sucrose; starches such as corn starch and potato starch; cellulose and its derivatives such as sodium carboxymethyl cellulose, ethyl cellulose and cellulose acetate; powdered tragacanth; malt; gelatin; talc; cocoa butter and suppository waxes; oils such as peanut oil, cottonseed oil, safflower oil, sesame oil, olive oil, corn oil and soybean oil; glycols, such as polyethylene glycol and propylene glycol; esters such as ethyl oleate and ethyl laurate; agar; buffering agents such as magnesium hydroxide and aluminum hydroxide; alginic acid; isotonic saline; Ringer's solution; ethyl alcohol; phosphate buffer solutions; non-toxic compatible lubricants such as sodium lauryl sulfate and magnesium stearate; coloring agents; releasing agents; coating agents; sweetening, flavoring and perfuming agents; preservatives; antioxidants; ion exchangers; alumina; aluminum stearate; lecithin; self-emulsifying drug delivery systems (SEDDS) such as d- α -tocopherol polyethyleneglycol 1000 succinate; surfactants used in pharmaceutical dosage forms such as Tweens or other similar polymeric delivery matrices; serum proteins such as human serum albumin; glycine; sorbic acid; potassium sorbate; partial glyceride mixtures of saturated vegetable fatty acids; water, salts or electrolytes such as protamine sulfate, disodium hydrogen phosphate, potassium hydrogen phosphate, sodium chloride, and zinc salts; colloidal silica; magnesium trisilicate; polyvinyl pyrrolidone; cellulose-based substances; polyacrylates; waxes; and polyethylene-polyoxypropylene-block polymers. Cyclodextrins such as α -, β -, and γ -cyclodextrin, or chemically modified derivatives such as hydroxyalkylcyclodextrins, including 2- and 3-hydroxypropyl-cyclodextrins, or other solubilized derivatives can also be used to enhance delivery of compounds described herein.

[0122] By "pharmaceutical formulation" it is further meant that the carrier, solvent, excipient(s) and/or salt must be compatible with the active ingredient of the formulation (e.g. a disclosed compound). It is understood by those of ordinary skill in this art that the terms "pharmaceutical formulation" and "pharmaceutical composition" are generally interchangeable, and they are so used for the purposes of this application.

[0123] By "prodrug" is meant any compound that must undergo bioactivation before exhibiting its intended pharmacological effects. Since prodrugs are known to enhance numerous desirable qualities of pharmaceuticals (e.g., solubility, bioavailability,

manufacturing, etc.), the disclosed compounds can be delivered in prodrug form. Thus, both prodrugs of the compounds, methods of delivering the same and compositions containing the same are disclosed herein. Prodrugs are intended to include any covalently bonded carriers that release an active parent drug (compound) *in vivo* when such prodrug is administered to a subject. Prodrugs are prepared, for example, by modifying functional groups present in the compound in such a way that the modifications are cleaved, either in routine manipulation or *in vivo*, to the parent compound. Prodrugs include compounds wherein a hydroxyl or amino group is bonded to any group that, when the prodrug is administered to a subject, it cleaves to form a free hydroxyl or free amino group, respectively. Examples of prodrugs include, but are not limited to, acetate, formate, and benzoate derivatives of alcohol and amine functional groups in the disclosed compounds. Examples of prodrugs include, but are not limited to, benzamide derivatives of an amine functional group in the active compound and the like. Other examples of prodrugs include compounds that comprise $-NO$, $-NO_2$, $-ONO$, or $-ONO_2$ moieties.

[0124] For example, if a disclosed compound or a pharmaceutically acceptable form of the compound contains a carboxylic acid functional group, a prodrug can comprise a pharmaceutically acceptable ester formed by the replacement of the hydrogen atom of the acid group with a group such as (C₁₋₈)alkyl, (C₁₋₁₂)alkanoyloxymethyl, 1-(alkanoyloxy)ethyl having from 4 to 9 carbon atoms, 1-methyl-1-(alkanoyloxy)-ethyl having from 5 to 10 carbon atoms, alkoxycarbonyloxymethyl having from 3 to 6 carbon atoms, 1-(alkoxycarbonyloxy)ethyl having from 4 to 7 carbon atoms, 1-methyl-1-(alkoxycarbonyloxy)ethyl having from 5 to 10 carbon atoms, N-(alkoxycarbonyl)aminomethyl having from 3 to 9 carbon atoms, 1-(N-(alkoxycarbonyl)amino)ethyl having from 4 to 10 carbon atoms, 3-phthalidyl, 4-crotonolactonyl, gamma-butyrolacton-4-yl, di-N,N-(C₁₋₂)alkylamino(C₂₋₃)alkyl (such as [3-dimethylaminoethyl], carbamoyl-(C₁₋₂)alkyl, N,N-di(C₁₋₂)alkylcarbamoyl-(C₁₋₂)alkyl and piperidino-, pyrrolidino- or morpholino(C₂₋₃)alkyl.

[0125] Similarly, if a disclosed compound or a pharmaceutically acceptable form of the compound contains an alcohol functional group, a prodrug can be formed by the replacement of the hydrogen atom of the alcohol group with a group such as (C₁₋₆)alkanoyloxymethyl, 1-((C₁₋₆)alkanoyloxy)ethyl, 1-methyl-1-((C₁₋₆)alkanoyloxy)ethyl, (C₁₋₆)alkoxycarbonyloxymethyl, N-(C₁₋₆)alkoxycarbonylaminomethyl, succinoyl, (C₁₋₆)alkanoyl,

α -amino(C₁₋₄)alkanoyl, arylacyl, and α -aminoacyl, or α -aminoacyl- α -aminoacyl, where each α -aminoacyl group is independently selected from the naturally occurring L-amino acids, -P(O)(OH)₂, -P(O)(O(C₁₋₆)alkyl)₂ or glycosyl (the radical resulting from the removal of a hydroxyl group of the hemiacetal form of a carbohydrate).

[0126] If a disclosed compound or a pharmaceutically acceptable form of the compound incorporates an amine functional group, a prodrug can be formed by the replacement of a hydrogen atom in the amine group with a group such as R-carbonyl, RO-carbonyl, NRR'-carbonyl where R and R' are each independently (C₁₋₁₀)alkyl, (C₃₋₇)cycloalkyl, benzyl, a natural α -aminoacyl or natural α -aminoacyl-natural- α -aminoacyl, -C(OH)C(O)OY¹ wherein Y¹ is H, (C₁₋₆)alkyl or benzyl, C(OY²)Y³ wherein Y² is (C₁₋₄)alkyl and Y³ is (C₁₋₆)alkyl, carboxy(C₁₋₆)alkyl, amino(C₁₋₄)alkyl or mono-N- or di-N,N- (C₁₋₆)alkylaminoalkyl, -C(Y⁴)Y⁵ wherein Y⁴ is H or methyl and Y⁵ is mono-N- or di-N,-(C₁₋₆)alkylamino, morpholino, piperidin-1-yl or pyrrolidin-1-yl.

[0127] In certain cases, a prodrug has improved physical and/or delivery properties over the parent compound. Prodrugs can increase the bioavailability of the compound when administered to a subject (e.g., by permitting enhanced absorption into the blood following oral administration) or which enhance delivery to a biological compartment of interest (e.g., the brain or lymphatic system) relative to the parent compound. Exemplary prodrugs include derivatives of a disclosed compound with enhanced aqueous solubility or active transport through the gut membrane, relative to the parent compound.

[0128] By "reference" is meant a standard or control condition.

[0129] A "reference sequence" is a defined sequence used as a basis for sequence comparison. In one embodiment, the reference sequence is Cas9.

[0130] By "small molecule" is meant any chemical compound.

[0131] The terms "selecting a subject" or "identifying a subject" are understood as choosing one or more members of a mixed population of individuals based on specific characteristics including, but not limited to, physical symptoms, and/or clinical characteristics as determined by diagnostic methods.

[0132] The term "solvate" means a solvent addition form that contains either a stoichiometric or non-stoichiometric amount of solvent. Some compounds have a tendency to trap a fixed molar ratio of solvent molecules in a solid state, thus forming a solvate. If the solvent is water, the solvate formed is a hydrate; when the solvent is alcohol, the solvate

formed is an alcoholate. Hydrates are formed by the combination of one or more molecules of water with one of the substances in which the water retains its molecular state as H₂O, such combination being able to form one or more hydrates.

[0133] The term “suitable solvent” refers to any solvent, or mixture of solvents, that may be inert to the ongoing reaction that sufficiently solubilizes the reactants to afford a medium within which to effect the desired reaction.

[0134] By “specifically binds” is meant recognizes and binds a polynucleotide or polypeptide of the invention, but which does not substantially recognize and bind other molecules in a sample.

[0135] By “subject” is meant an organism, including, but not limited to, a human or non-human mammal, such as a bovine, equine, canine, ovine, or feline; vertebrate; invertebrate, such as an insect; or plant; or any commercially relevant organism.

[0136] As used herein, the terms “treat,” “treating,” “treatment,” and the like refer to reducing or ameliorating a disorder and/or symptoms associated therewith. It will be appreciated that, although not precluded, treating a disorder or condition does not require that the disorder, condition or symptoms associated therewith be completely eliminated.

[0137] As used herein, the terms “prevent,” “preventing,” “prevention,” “prophylactic treatment” and the like refer to reducing the probability of developing a disorder or condition in a subject, who does not have, but is at risk of or susceptible to developing a disorder or condition.

[0138] Unless specifically stated or obvious from context, as used herein, the term “about” is understood as within a range of normal tolerance in the art, for example within 2 standard deviations of the mean. About can be understood as within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, 1%, 0.5%, 0.1%, 0.05%, or 0.01% of the stated value. Unless otherwise clear from context, all numerical values provided herein are modified by the term about.

[0139] Ranges provided herein are understood to be shorthand for all of the values within the range. For example, a range of 1 to 50 is understood to include any number, combination of numbers, or sub-range from the group consisting 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, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, or 50.

[0140] Any compounds, compositions, or methods provided herein can be combined with one or more of any of the other compositions and methods provided herein.

[0141] Unless specifically stated or obvious from context, as used herein, the term “or” is understood to be inclusive. Unless specifically stated or obvious from context, as used herein, the terms “a”, “an”, and “the” are understood to be singular or plural. Thus, for example, reference to “an amino acid substitution” includes reference to more than one amino acid substitution.

[0142] The term “including” is used herein to mean, and is used interchangeably with, the phrase “including but not limited to.”

[0143] As used herein, the terms “comprises,” “comprising,” “containing,” “having,” and the like can have the meaning ascribed to them in U.S. Patent law and can mean “includes,” “including,” and the like; “consisting essentially of” or “consists essentially” likewise has the meaning ascribed in U.S. Patent law and the term is open-ended, allowing for the presence of more than that which is recited so long as basic or novel characteristics of that which is recited is not changed by the presence of more than that which is recited, but excludes prior art embodiments.

[0144] Various embodiments are described hereinafter. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s). Reference throughout this specification to “one embodiment”, “an embodiment,” “an example embodiment,” means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases “in one embodiment,” “in an embodiment,” or “an example embodiment” in various places throughout this specification are not necessarily all referring to the same embodiment, 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.

[0145] 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

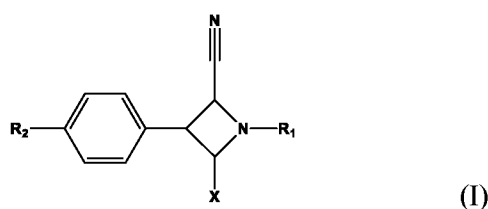
[0146] Embodiments disclosed herein provide compounds, compositions and methods for inhibiting the activity of RNA guided endonucleases (e.g., Cas9, Cpf1), and methods of use thereof, as well as to inhibit or prevent Cas9 genome editing. The invention is based, at least in part, on the discovery of small molecule inhibitor compounds of RNA guided endonucleases. As described herein, high-throughput biochemical and cellular assays, and workflows comprising combinations of such assays, were developed for screening and identifying small molecules with the ability to inhibit one or more activities of RNA guided endonucleases. Methods involving small molecule inhibitors of RNA guided endonucleases are useful for the modulation of RNA guided endonuclease activity, including rapid, reversible, dosage, and/or temporal control of RNA guided endonuclease technologies.

COMPOUNDS FOR INHIBITING NUCLEASE ACTIVITY

[0147] Compounds for inhibiting nuclease activity are provided that comprise contacting an endonuclease with a small molecule compounds that modulate (e.g., inhibit) RNA-guided endonuclease activity. The compounds may allow for rapid, dosable, and/or temporal control of RNA-guided endonuclease (e.g., Cas9) activity. In some embodiments, the compounds may increase the specificity of RNA-guided endonuclease and may enable external control and manipulation of gene targeting.

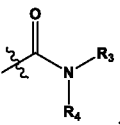
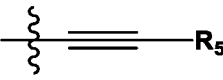
[0148] In embodiments, the nuclease is an RNA-guided endonucleases. By “RNA guided endonuclease” is meant a polypeptide having RNA binding activity, DNA binding activity, and/or DNA cleavage activity. RNA guided endonucleases form a complex with a guide RNA, which contains a sequence that is able to bind a target sequence on double stranded DNA. In some embodiments, the RNA guided endonuclease cleaves the double stranded target DNA.

[0149] Provided herein are compounds of Formula I and methods comprising use of compounds of Formula I. In an aspect, the invention provides a method of inhibiting the activity of an RNA guided endonuclease-guide complex comprising contacting the complex with a compound having the structure of Formula I:



wherein:

R₁ is selected from hydrogen, COR₆, CHR₇R₈, or SO₂R₉;

R₂ is selected from hydrogen, , , or a cyclic structure selected from cycloalkane, cycloalkene, aryl, or heteroaryl, each optionally substituted with one or more groups R₁₀;

R₃ is selected from hydrogen, C₁-C₁₂ alkyl, aryl, alkylaryl, optionally substituted with one or more groups selected from halogen, alkyl, alkoxy, or amine;

R₄ is selected from hydrogen, C₁-C₁₂ alkyl, aryl, alkylaryl, optionally substituted with one or more groups selected from halogen, alkyl, alkoxy, or amine;

R₅ is selected from C₁-C₁₂ alkyl and a cyclic structure selected from cycloalkane, cycloalkene, aryl, or heteroaryl, each optionally substituted with one or more groups R₁₁;

R₆ independently selected at each occurrence from C₁₋₁₂ alkyl, alkylcycloalkane, cycloalkane, aryl, heteroaryl, alkylaryl, alkylheteroaryl, or alkylamine;

R₇ is selected from hydrogen, C₁₋₁₂ alkyl, aryl, alkylaryl, hydroxyl, amine, alkoxy, or halogen;

R₈ is selected from hydrogen, C₁₋₁₂ alkyl, aryl, alkylaryl, hydroxyl, amine, alkoxy, or halogen;

R₉ is selected from C₁₋₁₂ alkyl or a cyclic structure selected from cycloalkane, cycloalkene, aryl, or heteroaryl, each optionally substituted with one or more groups R₁₂;

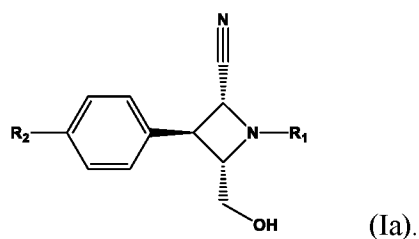
R₁₀ is independently selected at each occurrence from halogen, C₁-C₁₂ alkyl, haloalkyl, or alkoxy;

R₁₁ is independently selected at each occurrence from halogen, C₁-C₁₂ alkyl, haloalkyl, or alkoxy;

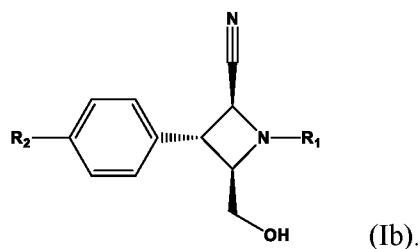
R₁₂ is independently selected at each occurrence from hydrogen, halogen, C₁-C₃ alkyl, haloalkyl, or alkoxy; and

X is selected from =O, hydroxyl, alkylalcohol, amine, or alkylamine.

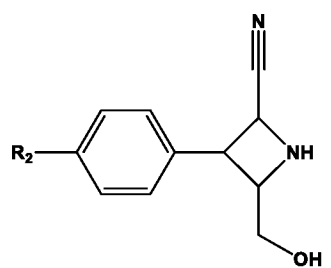
[0150] In certain embodiments, the compound of Formula I has the structure:



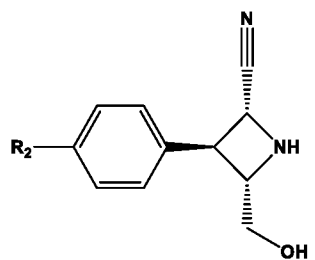
[0151] In one such embodiment, the compound of Formula I has the structure:



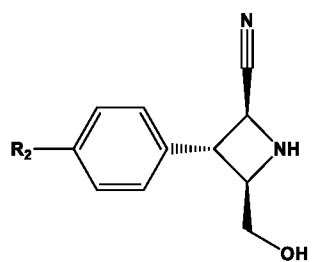
[0152] In another embodiment, the compound of Formula I has the structure:



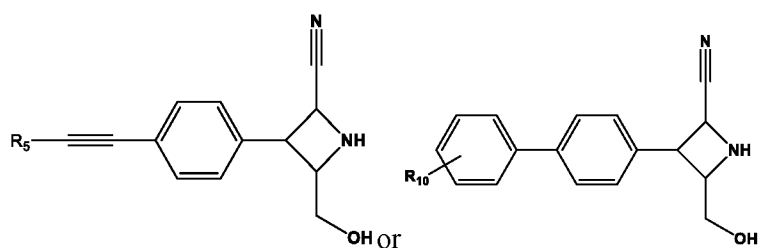
[0153] In another embodiment, the compound of Formula I has the structure:



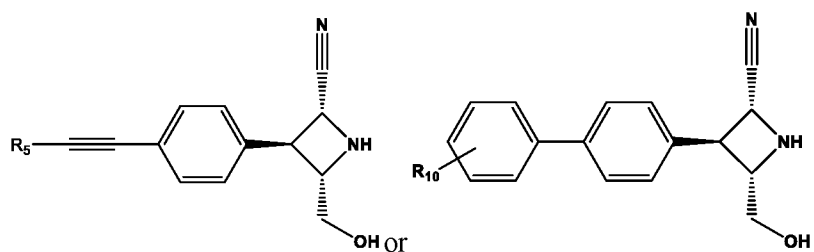
[0154] In another embodiment, the compound of Formula I has the structure:



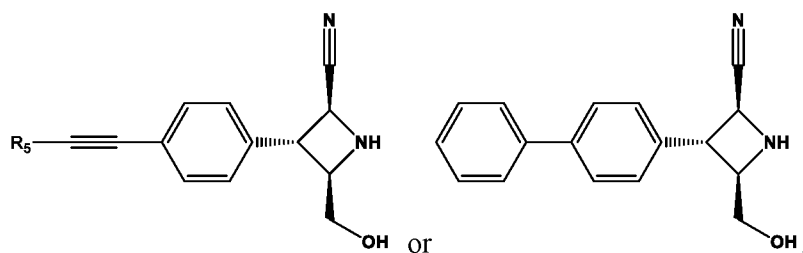
[0155] In another embodiment, the compound of Formula I has the structure:



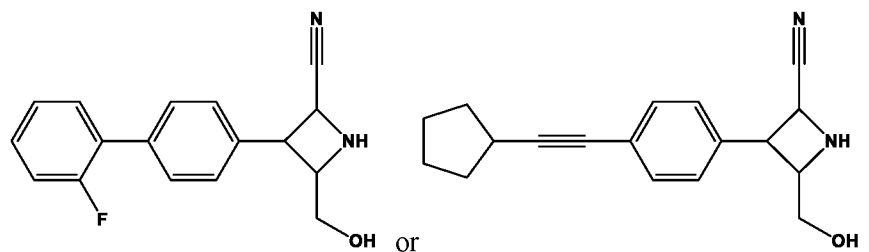
[0156] In another embodiment, the compound of Formula I has the structure:



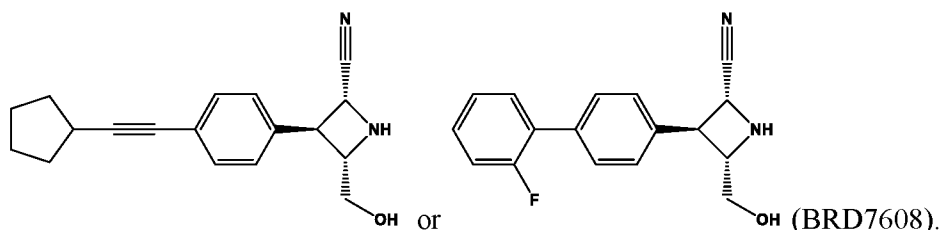
[0157] In another embodiment, the compound of Formula I has the structure:



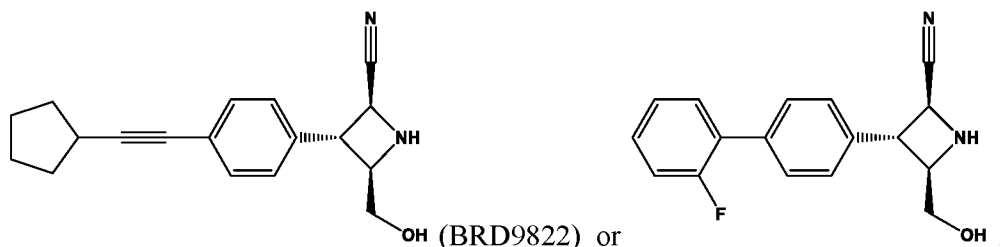
[0158] In another embodiment, the compound of Formula I has the structure:



[0159] In another embodiment, the compound of Formula I has the structure:



[0160] In another embodiment, the compound of Formula I has the structure:



[0161] BRD7608 is deposited in the PubChem Open Chemistry Database as PubChem CID: 54650473. BRD9822 is deposited in the PubChem Open Chemistry Database as PubChem CID: 54651106.

[0162] The disclosed compounds of Formula I can be in free base form unassociated with other ions or molecules, or they can be a pharmaceutically acceptable salt, solvate, or prodrug thereof. One aspect provides a disclosed compound or a pharmaceutically acceptable salt. One aspect provides a disclosed compound or a pharmaceutically acceptable salt or solvate thereof. One aspect provides a pharmaceutically acceptable salt of a disclosed compound. One aspect provides a solvate of a disclosed compound. One aspect provides a hydrate of a disclosed compound. One aspect provides a prodrug of a disclosed compound.

METHODS OF USE

[0163] The compounds described herein may be used as inhibitors of RNA guided endonucleases and can include use of the small molecule compounds with RNA guided endonucleases (e.g., Cas9) in potential therapeutic uses for regulating genome editing technologies involving RNA guided endonucleases. Dosable control of the therapeutic activity of RNA guided endonucleases introduced into a subject or cell of a subject is important for effective genome editing therapeutic strategies. In some examples, methods herein may include a method for treating a subject, comprising administering an RNA-guided endonuclease-RNA complex or a reagent causing expression of the RNA-guided

endonuclease-RNA complex to the subject, and administering one or more compounds described herein.

[0164] Small molecule inhibitors of RNA guided endonucleases can be administered to a subject undergoing RNA guided endonuclease-based gene therapy or any other RNA guided endonuclease-based therapy. In certain embodiments, the subject is a human or mammal. Small molecule inhibitors of RNA guided endonucleases eliminate or reduce undesirable off-target editing and chromosomal translocations when present at high concentrations. Furthermore, small molecule inhibitors of RNA guided endonucleases can be used to rapidly terminate constitutively active Cas9, following on-target gene-editing.

[0165] The compounds described herein were identified and developed to allow rapid, dosable, and/or temporal control of Cas9 activities. The reports of small-molecule controlled Cas9 activity are present in literature (Senis et al., *Biotechnol J* 2014, 9, 1402-12; Wright et al., *Proc Natl Acad Sci U S A.* 2015 Mar 10;112(10):2984-9; Gonzalez et al., *Cell Stem Cell* 2014, 15, 215-26; Davis et al., *Nat Chem Biol* 2015, 11, 316-8), but none of them ensure dosability—the small molecules act merely as inducers of Cas9 activity. Further, most of these small molecule systems are not reversible upon removal of the small molecule (Zetsche et al., *Nat Biotech* 2015, 33, 139-142; Davis et al., *Nat Chem Biol* 2015, 11, 316-8), and therefore, do not allow precise temporal control in transcriptional regulatory technologies as currently provided.

[0166] In certain embodiments, the method is performed *in vitro*, *ex vivo*, or *in vivo*. Alternatively or additionally, the methods may be performed *in vivo*. In some examples, the methods may be performed in a cell. The cell may be a germline cell. The cell may also be any types of cell, e.g., a stem cell such as an embryonic stem cell or a induced pluripotent stem cell. In certain examples, the methods may be performed in a cell in an organism (e.g., human, mammal, vertebrate, invertebrate, insect, plant). In some cases, the cell may be a prokaryotic cell, e.g., a bacterium. In certain cases, the cell may be a eukaryotic cell, e.g., a mammalian (e.g., human) cell, an insect cell, a plant cell, a fungal cell (e.g., a yeast cell). In certain embodiments, the cell is in an organism, such as but not limited to a human, mammal, vertebrate, invertebrate, insect, or plant.

[0167] Methods using the compounds disclosed herein may be performed in cell, plants and animal models that have been engineered to constitutively or inducibly express an RNA-guided nuclease. The compounds disclosed herein may be delivered concurrent or

subsequent to delivery of an RNA-guided nuclease, including in applications where RNA-guided nuclease is used for therapeutics. Accordingly, temporal control over editing may include a single dose, multiple doses over particular time frames, and/or a continuous dose of the compounds. Dosing is provided at amounts effective to impact editing outcomes. In embodiments, administration is provided to favor editing by HDR over NHEJ. Administration may be to cells or organisms with gene drives. In particular embodiments, the administration is provided in an amount effective to halt or reverse gene drive function.

[0168] Small molecule inhibitors of RNA guided endonucleases (e.g., Cas9) have potential therapeutic uses for regulating genome editing technologies involving RNA guided endonucleases, and in certain embodiments allow for reversible control of the systems. Dosable control of the therapeutic activity of RNA guided endonucleases introduced into a subject or cell of a subject is important for effective genome editing therapeutic strategies. Small molecule inhibitors of RNA guided endonucleases can be administered to a subject undergoing RNA guided endonuclease-based gene therapy or any other RNA guided endonuclease based therapy. In certain embodiments, the subject is a human or mammal. These small molecule compounds eliminate or reduce undesirable off-target editing and chromosomal translocations when present at high concentrations. Furthermore, small molecule inhibitors of RNA guided endonucleases can be used to rapidly terminate constitutively active Cas9, following on-target gene-editing.

[0169] The compounds disclosed herein can also be used to regulate genome editing technologies in other organisms, including invertebrates, plants, and unicellular organisms (e.g., bacteria). Potential uses include regulating gene drives for entomological and agricultural uses. In addition, it is anticipated that Cas9 inhibitors will be valuable probes to understand the role of Cas9 in CRISPR-mediated bacterial immunity (e.g., spacer acquisition) (Nunez et al., *Nature*. 2015 Mar 12;519(7542):193-8; Heler et al., *Nature* 2015, 519, 199-202). Along similar lines, Cas9 inhibitors can be deployed for directed evolution of Cas9. It is hypothesized that Cas9 inhibitors will disrupt bacterial immunity against bacteriophages (or toxic DNA) by interfering with the CRISPR-Cas9-based immune surveillance system in bacteria. Akin to the development of antibiotic resistance, bacteria will be forced to evolve Cas9 protein. Accordingly, the inhibitors may also be used as an anti-infective agent.

[0170] Agents described herein, including analogs thereof, and/or agents discovered to have medicinal value using the methods described herein are useful as a drug for inhibiting RNA guided endonucleases (e.g., Cas9, Cpf1). For therapeutic uses, the compositions or agents identified using the methods disclosed herein may be administered systemically, for example, formulated in a pharmaceutically-acceptable buffer such as physiological saline. Preferable routes of administration include, for example, subcutaneous, intravenous, intraperitoneally, intramuscular, or intradermal injections that provide continuous, sustained levels of the drug in the patient. Treatment of human patients or other animals will be carried out using a therapeutically effective amount of a therapeutic identified herein in a physiologically- acceptable carrier. Suitable carriers and their formulation are described, for example, in Remington's Pharmaceutical Sciences by E. W. Martin. The amount of the therapeutic agent to be administered varies depending upon the manner of administration, the age and body weight of the patient, and with the clinical symptoms. Generally, amounts will be in the range of those used for other agents used in the treatment of disease.

[0171] The disclosed compounds may be administered alone (e.g., in saline or buffer) or using any delivery vehicles known in the art. For instance the following delivery vehicles have been described: Cochleates; Emulsomes, ISCOMs; Liposomes; Live bacterial vectors (e.g., Salmonella, Escherichia coli, Bacillus calmette-guerin, Shigella, Lactobacillus); Live viral vectors (e.g., Vaccinia, adenovirus, Herpes Simplex); Microspheres; Nucleic acid vaccines; Polymers; Polymer rings; Proteosomes; Sodium Fluoride; Transgenic plants; Virosomes; Virus-like particles. Other delivery vehicles are known in the art and some additional examples are provided below.

[0172] The disclosed compounds may be administered by any route known, such as, for example, orally, transdermally, intravenously, cutaneously, subcutaneously, nasally, intramuscularly, intraperitoneally, intracranially, and intracerebroventricularly.

[0173] In certain embodiments, disclosed compounds are administered at dosage levels greater than about 0.001mg/kg, such as greater than about 0.01 mg/kg or greater than about 0.1 mg/kg. For example, the dosage level may be from about 0.001 mg/kg to about 50 mg/kg such as from about 0.01 mg/kg to about 25 mg/kg, from about 0.1 mg/kg to about 10 mg/kg, or from about 1 mg/kg to about 5 mg/kg of subject body weight per day, one or more times a day, to obtain the desired therapeutic effect. It will also be appreciated that dosages smaller

than about 0.001 mg/kg or greater than about 50 mg/kg (for example about 50-100 mg/kg) can also be administered to a subject.

[0174] In one embodiment, the compound is administered once-daily, twice-daily, or three- times daily. In one embodiment, the compound is administered continuously (e.g., every day) or intermittently (e.g., 3-5 days a week). In another embodiment, administration could be on an intermittent schedule.

[0175] Further, administration less frequently than daily, such as, for example, every other day may be chosen. In additional embodiments, administration with at least 2 days between doses may be chosen. By way of example only, dosing may be every third day, bi-weekly or weekly. As another example, a single, acute dose may be administered. Alternatively, compounds can be administered on a non-regular basis e.g., whenever symptoms begin. For any compound described herein the effective amount can be initially determined from animal models.

[0176] Toxicity and efficacy of the compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds that exhibit large therapeutic indices may have a greater effect when practicing the methods as disclosed herein. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0177] Data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage of the compounds disclosed herein for use in humans. The dosage of such agents lies within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage form employed and the route of administration utilized. For any compound used in the disclosed methods, the effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC₅₀ (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for

example, by high performance liquid chromatography. In certain embodiments, pharmaceutical compositions may comprise, for example, at least about 0.1% of an active compound. In other embodiments, the active compound may comprise between about 2% to about 75% of the weight of the unit, or between about 25% to about 60%, for example, and any range derivable therein. Multiple doses of the compounds are also contemplated.

[0178] The formulations disclosed herein are administered in pharmaceutically acceptable solutions, which may routinely contain pharmaceutically acceptable concentrations of salt, buffering agents, preservatives, compatible carriers, and optionally other therapeutic ingredients.

[0179] For use in therapy, an effective amount of one or more disclosed compounds can be administered to a subject by any mode that delivers the compound(s) to the desired surface, e.g., mucosal, systemic. Administering the pharmaceutical composition of the present disclosure may be accomplished by any means known to the skilled artisan. Disclosed compounds may be administered orally, transdermally, intravenously, cutaneously, subcutaneously, nasally, intramuscularly, intraperitoneally, intracranially, or intracerebroventricularly.

Nucleases

[0180] The inhibitor compounds herein may be used for modulating the activities and functions of Cas proteins in CRISPR-Cas systems. In certain example embodiments, the RNA guided nuclease is a Cas protein. In general, a CRISPR-Cas or CRISPR system as used in herein and in documents, such as WO 2014/093622 (PCT/US2013/074667), refers collectively to transcripts and other elements involved in the expression of or directing the activity of CRISPR-associated (“Cas”) genes, including sequences encoding a Cas gene, a tracr (trans-activating CRISPR) sequence (e.g. tracrRNA or an active partial tracrRNA), a tracr-mate sequence (encompassing a “direct repeat” and a tracrRNA-processed partial direct repeat in the context of an endogenous CRISPR system), a guide sequence (also referred to as a “spacer” in the context of an endogenous CRISPR system), or “RNA(s)” as that term is herein used (e.g., RNA(s) to guide Cas, such as Cas9, e.g. CRISPR RNA and transactivating (tracr) RNA or a single guide RNA (sgRNA) (chimeric RNA)) or other sequences and transcripts from a CRISPR locus. In general, a CRISPR system is characterized by elements that promote the formation of a CRISPR complex at the site of a target sequence (also referred to as a protospacer in the context of an endogenous CRISPR system). *See, e.g.*

Shmakov et al. (2015) “Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems”, *Molecular Cell*, DOI: [dx.doi.org/10.1016/j.molcel.2015.10.008](https://doi.org/10.1016/j.molcel.2015.10.008).

[0181] In certain embodiments, a protospacer adjacent motif (PAM) or PAM-like motif directs binding of the effector protein complex as disclosed herein to the target locus of interest. In some embodiments, the PAM may be a 5' PAM (i.e., located upstream of the 5' end of the protospacer). In other embodiments, the PAM may be a 3' PAM (i.e., located downstream of the 5' end of the protospacer). The term “PAM” may be used interchangeably with the term “PFS” or “protospacer flanking site” or “protospacer flanking sequence”.

[0182] In a preferred embodiment, the CRISPR effector protein may recognize a 3' PAM. In certain embodiments, the CRISPR effector protein may recognize a 3' PAM which is 5'H, wherein H is A, C or U.

[0183] By “protospacer adjacent motif (PAM)” is meant a nucleic acid sequence immediately adjacent the nucleic acid sequence targeted by an RNA guided endonuclease (e.g., Cas9, Cpf1). In certain embodiments, the PAM sequence is the Cas9 PAM sequence: 5'-NGC-3'.

[0184] The Cas protein is preferably a class 2 CRISPR Cas protein including Type II, Type V, and Type VI proteins. A Class 2 CRISPR-Cas system may be of a subtype, e.g., Type II-A, Type II-B, Type II-C, Type V-A, Type V-B, Type V-C, Type V-U, Type VI-A, Type VI-B, or Type VI-C CRISPR-Cas system. The definition and exemplary members of the CRISPR-Cas system include those described in Kira S. Makarova and Eugene V. Koonin, *Annotation and Classification of CRISPR-Cas systems*, *Methods Mol Biol.* 2015; 1311: 47–75; and Sergey Shmakov et al., *Diversity and evolution of class 2 CRISPR–Cas systems*, *Nat Rev Microbiol.* 2017 Mar; 15(3): 169–182.

[0185] Non-limiting examples of RNA-guided nucleases include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Cas12 (e.g., Cas12a, Cas12b, Cas12c, Cas12d), Cas13 (e.g., (Cas13a, Cas13b, Cas13c, Cas13d), Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologues thereof, or modified versions thereof. In certain embodiments, the RNA-guided nucleases may be the nuclease in any CRISPR-Cas system. For example, the CRISPR system may be a class 2 CRISPR system, including Type II, Type V and Type VI systems. In certain example embodiments, the

CRISPR system is a Cas9, a Cas12a, Cas12b, Cas12c, Cas12d, Cas13a, Cas13b, Cas13c, or Cas13d system. For example, the RNA-guided nuclease may be Cas9, a Cas12a, Cas12b, Cas12c, Cas12d, Cas13a, Cas13b, Cas13c, or Cas13d. The Cas protein is preferably a class 2 CRISPR Cas protein including Type II, Type V, and Type VI proteins. Accordingly, the methods can inhibit Cas9, including but not limited to SpCas9, SaCas9, StCas9 and other Cas9 orthologs. Similarly, the methods may inhibit Cas12, including but not limited to orthologs of Cas12a, Cas12b, and Cas12c, including FnCas12a. Moreover, the method may inhibit Cas13, including but not limited to orthologs of Cas13a, Cas13b, Cas13c, and Cas13d. Accordingly, the methods can inhibit Cas9, including but not limited to SpCas9, SaCas9, StCas9 and other Cas9 orthologs. Similarly, the methods may inhibit Cas12, including but not limited to orthologs of Cas12a, Cas12b, and Cas12c, including FnCas12a. Moreover, the method may inhibit Cas13, including but not limited to orthologs of Cas13a, Cas13b, Cas13c, and Cas13d.

[0186] In certain embodiments, the RNA-guided endonuclease is Cas9, such as but not limited to SpCas9 or SaCas9.

[0187] In certain example embodiments, the RNA-guided nuclease is naturally present in a prokaryotic genome within 20kb upstream or downstream of a Cas 1 gene. The terms "orthologue" (also referred to as "ortholog" herein) and "homologue" (also referred to as "homolog" herein) are well known in the art. By means of further guidance, a "homologue" of a protein as used herein is a protein of the same species which performs the same or a similar function as the protein it is a homologue of. Homologous proteins may but need not be structurally related, or are only partially structurally related. An "orthologue" of a protein as used herein is a protein of a different species which performs the same or a similar function as the protein it is an orthologue of. Orthologous proteins may but need not be structurally related, or are only partially structurally related.

[0188] Cas9 is a spectacular molecular machine with several fascinating attributes. First, the Cas9 employs molecular interactions and recognition between all the three elements of the central dogma—DNA, RNA, and protein. Second, Cas9 unwinds target DNA and facilitates strand invasion without utilizing ATP. Third, Cas9 efficiently induces DNA strand breaks in both prokaryotic and eukaryotic genomes despite enormous differences in their genome size, structure, and organization. Finally, unlike transcription factors that often employ 1D diffusion and hopping for target search, Cas9:gRNA complex accomplishes target

search by 3D diffusion only. Upon target recognition, Cas9 induces double strand breaks in the target gene, which when repaired by non-homologous end joining (NHEJ) can result in frameshift mutations and gene knockdown. Alternatively, homology-directed repair (HDR) at the double-strand break site can allow insertion of the desired sequence.

[0189] While Cas9 is a highly efficient molecular machine, its specificity is problematic. Using genomewide, unbiased identification of double strand breaks enabled by sequencing (GUIDE-seq), Joung and co-workers observed that Cas9's off-targets were present on nearly every chromosome for the on-target gene EMX1. In another specificity study, Zhang and co-workers systematically studied the effect of single-base mismatch between gRNA and target sequence on Cas9 cleavage efficiency for the EMX1 gene. They found that Cas9 tolerated mismatches at PAM-distal sites on gRNA for multiple locations on EMX1 gene. Similar trends in mismatch tolerance were also reported by Doudna and Joung laboratories. Finally, Alt and co-workers have shown that in addition to off-target editing, Cas9 induces chromosomal translocations leading to dicentric chromosomes, which will generate genomic instability. They and others also demonstrated an inverse correlation between Cas9 activity and specificity.

[0190] In embodiments, Cas9 recognizes foreign DNA using Protospacer Adjacent Motif (PAM) sequence and the base-pairing of the target DNA by the guide RNA (gRNA). The relative ease of inducing targeted strand breaks at any genomic loci by Cas9 has enabled efficient genome editing in multiple cell types and organisms. Cas9 derivatives can also be used as transcriptional activators/repressors. A challenge posed by Cas9 is that its cleavage selectivity is low. Off-target editing activity can result in undesired undesirable chromosomal translocation. This activity limits the use of Cas9 in a therapeutic setting due to unreliable gene manipulation and lack of ability to control the action of Cas9.

[0191] By "Cas9 (CRISPR associated protein 9)" is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to NCBI Accession No. NP_269215 and having RNA binding activity, DNA binding activity, and/or DNA cleavage activity (e.g., endonuclease or nickase activity). An exemplary Cas9 polypeptide sequence is provided below.

```
1 mdkkysigld igtnsvgwav itdeykvpsk kfkvlgntr hsikknliga llfdsgetae
61 atrlkrarr rytrkrnric ylqeifsnem akvddsfhr leesflveed kkherhpifg
121 nivdevayhe kyptiyhlrk klvdstdkad lrliylalah mikfrghfli egdlndnsd
```

181 vdklfqlvq tynqlfeenp inasgvdaka ilsarlsksr rlenliaqlp gekknlgfn
 241 lialslgltp nfksnfdlae daklqlskdt yddldnlla qigdqyadlf laaknlsdai
 301 llsdilrvnt eitkaplsas mikrydehhq dltilkalvr qqlpekykei ffdqskngya
 361 gyidggasqe efykfikpil ekmdgteell vknredllr kqrtfdngsi phqihlgelh
 421 ailrrqedfy pflkdnreki ekiltfipy yvgplargns rfawmtrkse eitpwnfee
 481 vvdkgasqs fiermtndfk nlpnekvlpk hsllyeyftv yneltkvkyv tegmrkpafl
 541 sgeqkkaivd llfktnrkvt vkqlkedyfk kiecfdsvei sgvedrfnas lgtyhdlki
 601 ikdkdfldne enedilediv ltltfedre mieerlktya hlfdkvmkq lkrtrytgwg
 661 rlsrklingi rdkqsgktil dflksdgan rnmqlihdd sltfkediql aqvsgqgds
 721 hehianlags paikkgilqt vkvvdelvkv mgrhkpeniv iemarenqtt qkgqknsrer
 781 mkrieegike lgsqilkehp ventqlqnek lylyylqngr dmyvdqeldi nrlsdydvdh
 841 ivpqsfkdd sidknvltrs dknrgksdnv pseevvkkmk nywrqllnak litqrkfdnl
 901 tkaergglse ldkagfikrq lvetrqitkh vaqildsrnm tkydendkli revkvitlks
 961 klvsdfrkdf qfykvreinn yhhahdayln avvgtalikk ypklesefvy gdykvydvk
 1021 miakseqeig katakyffys nimnffktei tlangeirkr plietngetg eivwdkgrdf
 1081 atvrkvlsmv qvnivkktev qtggfskesi lpkrnsdkli arkkdwdpkk yggdsptva
 1141 ysvlvvakve kgkskklksv kelligitime rssfeknpid fleakgykev kkdliiklpk
 1201 yslfelengr krmlasagel qkgnelalps kyvnflylas hyeklkgspe dneqkqlfve
 1261 qhkhyldiie eqisefskrv iladanldkv lsaynkhrdk pireqaenii hlftlnlga
 1321 paafkyfddt idrkrytstk evldatlihq sitglyetri dlsqlggd (SEQ ID NO:7)

[0192] "Cas9 function" can be defined by any of a number of assays including, but not limited to, fluorescence polarization-based nucleic acid bind assays, fluorescence polarization-based strand invasion assays, transcription assays, EGFP disruption assays, DNA cleavage assays, and/or Surveyor assays, for example, as described herein.

[0193] By "Cas 9 nucleic acid molecule" is meant a polynucleotide encoding a Cas9 polypeptide or fragment thereof. An exemplary Cas9 nucleic acid molecule sequence is provided at NCBI Accession No. NC_002737 and is shown below.

1 atggataaga aataactcaat aggcttagat atcggcacia atagcgtcgg atgggcggtg
 61 atcactgatg aatataaggt tccgtctaaa aagttaagg ttctgggaaa tacagaccgc
 121 cacagtatca aaaaaaatct tataggggct cttttatttg acagtggaga gacagcgga
 181 gcgactcgtc tcaaacggac agctcgtaga aggtatacac gtcggaagaa tcgtatttgt
 241 tatctacagg agatttttc aatgagatg gcgaaagtag atgatatgtt ctttcacga

301 ctgaagagt ctttttggg ggaagaagac aagaagcatg aacgtcatcc ttttttggg
361 aatatagtag atgaagtgc ttatcatgag aaatatcaa ctatctatca tctgcgaaaa
421 aaattggtag attctactga taaagcggat ttgcgcttaa tctatttggc cttagcgc
481 atgattaagt ttcgtggtca ttttttgatt gagggagatt taaatcctga taatagtat
541 gtggacaaac tatttatcca gttggtacaa acctacaatc aattatttga agaaaaccct
601 attaacgcaa gtggagtaga tgctaaagcg attcttctg cacgattgag taaatcaaga
661 cgattagaaa atctcattgc tcagctcccc ggtgagaaga aaaatggctt atttgggaat
721 ctattgctt tgcattggg ttgaccct aattttaa caaatttga tttggcagaa
781 gatgctaat tacagcttc aaaagatact tacgatgatg atttagataa tttattggcg
841 caaattggag atcaatatgc tgattgtt tttggcagta agaattatc agatgctatt
901 ttacttcag atacttaag agtaataact gaataacta aggtccct atcagctca
961 atgattaaac gctacgatga acatcatcaa gactgactc tttaaaagc tttagtcca
1021 caacaacttc cagaaaagta taaagaaatc tttttgatc aatcaaaaaa cggatatgca
1081 ggttatattg atgggggagc tagccaagaa gaatttata aatttatcaa accaatttta
1141 gaaaaaatgg atggtactga ggaattattg gtgaaactaa atcgtgaaga tttgctgcgc
1201 aagcaacgga ctttgacaa cggctctatt ccccatcaaa ttcactggg tgagctgcat
1261 gctatttga gaagacaaga agactttat ccattttta aagacaatcg tgagaagatt
1321 gaaaaaatct tgactttcg aattcctat tatgttggc cattggcgcg tggcaatagt
1381 cgtttgc atggtactcg gaagtctgaa gaaacaatta ccccatggaa tttgaagaa
1441 gttgctgata aagggtctc agctcaatca tttattgaac gcatgacaaa cttgataaa
1501 aatcttcaa atgaaaaagt actacaaaa catagtttgc tttatgagta tttacggt
1561 tataacgaat tgacaaagg ccaatatgtt actgaaggaa tgcgaaaacc agcattctt
1621 tcagggaac agaagaagc cattgtgat ttactctca aaacaatcg aaaagtaacc
1681 gtaagcaat taaaagaaga ttatttcaa aaaatagaat gtttgatag tttgaaatt
1741 tcaggagttg aagatagatt taatgctca ttaggtacct accatgatt gctaaaatt
1801 attaaagata aagattttt ggataatgaa gaaaatgaag atatctaga ggatattgt
1861 ttaacattga cttatttga agataggag atgattgagg aaagactaa aacatatgct
1921 cacctcttg atgataagg gatgaacag cttaaacgc gccgttatac tggttggga
1981 cgtttgc cgttattgaa taatggtatt agggataagc aatctggcaa aacaatatta
2041 gatttttga aatcagatgg ttttccaat cgcaattta tgcagctgat ccatgatg
2101 agtttgacat taaagaaga cattcaaaaa gcacaagtgt ctggacaagg cgatagtta
2161 catgaacata ttgcaaatt agctggtagc cctgtatta aaaaaggat tttacagact

2221 gtaaaagtgtg ttgatgaatt ggtcaaagta atggggcggc ataagccaga aaatatcggt
2281 attgaaatgg cacgtgaaaa tcagacaact caaaagggcc agaaaaattc gcgagagcgt
2341 atgaaacgaa tcgaagaagg tatcaaagaa ttaggaagtc agattcttaa agagcatcct
2401 gttgaaaata ctcaattgca aaatgaaaag ctctatctct attatctcca aaatggaaga
2461 gacatgtatg tggaccaaga attagatatt aatcggttaa gtgattatga tgcgatcac
2521 attgtccac aaagtttct taaagacgat tcaatagaca ataaggtctt aacgcgttct
2581 gataaaaatc gtggtaaatc ggataacgtt ccaagtgaag aagtagtcaa aaagatgaaa
2641 aactattgga gacaacttct aaacgccaag ttaatcactc aacgtaagtt tgataattta
2701 acgaaagctg aacgtggagg ttgagtgaa ctgataaag ctggtttat caaacgcaa
2761 ttggtgaaa ctgcctaat cactaagcat gtggcacaaa ttttgatag tcgcatgaat
2821 actaaatagc atgaaaatga taaacttatt cgagaggta aagtattac cttaaatct
2881 aaattagttt ctgactccg aaaagattc caattctata aagtacgtga gattaacaat
2941 taccatcatg cccatgatgc gtatctaat gccgtcgtg gaactgctt gattaagaaa
3001 tatccaaaac ttgaatcgga gttgtctat ggtgattata aagtttatga tgttcgtaa
3061 atgattgcta agtctgagca agaaataggc aaagcaaccg caaaatattt ctttactct
3121 aatatcatga acttctcaa aacagaaatt acattgcaa atggagagat tcgcaaaccg
3181 cctctaactg aaactaatgg ggaaactgga gaaattgtct gggataaagg gcgagattt
3241 gccacagtgc gcaaagtatt gtccatgcc caagtcaata ttgtcaagaa aacagaagta
3301 cagacaggcg gattctcaa ggagtcaatt ttacaaaaa gaaatcgga caagcttatt
3361 gctcgtaaaa aagactggga tccaaaaaa tatggtggtt ttgatagtc aacggtagct
3421 tattcagtc tagtggttgc taagggtgaa aaagggaat cgaagaagt aaaatccgt
3481 aaagagttac tagggatcac aattatggaa agaagttcct tgaaaaaaa tccgattgac
3541 ttttagaag ctaaaggata taaggaagt aaaaagact taatcattaa actacctaaa
3601 tatagtcttt ttgagttag aaacggtcgt aaacggatgc tggctagtgc cggagaatta
3661 caaaaaggaa atgagctggc tctgccaagc aaatatgtga atttttata ttagctagt
3721 cattatgaaa agttgaagg tagtccagaa gataacgaac aaaaacaatt gttgtggag
3781 cagcataagc attatttaga tgagattatt gagcaaatca gtgaatttc taagcgtgt
3841 atttagcag atgccaatt agataaagt cttagtcat ataacaaca tagagacaaa
3901 ccaatacgtg aacaagcaga aaatattatt cattattta cgttgacgaa tcttgagct
3961 cccgtcgtt taaatatt tgatacaaca atgatcgta aacgatatac gtctacaaa
4021 gaagtttag atgccactct tatccatcaa tccatcactg gtcttatga aacacgcatt
4081 gatttgagtc agctaggagg tgactga (SEQ ID NO:8)

[0194] In another aspect, the invention provides a method of inhibiting the activity of an RNA guided endonuclease-guide complex, for example by contacting the RNA guided endonuclease-guide complex with a small molecule. A nucleic acid guided nuclease-guide complex is comprised of the elements that form a complex at the site of a target sequence. See, e.g., Shmakov et al. (2015) “Discovery and Functional Characterization of Diverse Class 2 CRISPR-Cas Systems”, *Molecular Cell*, DOI: dx.doi.org/10.1016/j.molcel.2015.10.008. This complex comprises a nuclease, e.g. Cas polypeptide as described herein, and a guide molecule.

[0195] The Cas9 and other nuclease inhibitors disclosed herein provide rapid, dosable, and/or temporal control of Cas9 that increases Cas9 specificity and enables external control and manipulation of gene targeting. The nuclease inhibitors of the invention can be used in combination with other controls of CRISPR expression and/or activity.

[0196] By “Cpf1 (CRISPR associated protein Cpf1)” is meant a polypeptide or fragment thereof having at least about 85% amino acid identity to GenBank Accession No. AJ161006.1 and having RNA binding activity, DNA binding activity, and/or DNA cleavage activity (e.g., endonuclease or nickase activity). An exemplary Cas9 polypeptide sequence is provided below.

```

1  mslyqefvnk yslsktlrfe lipqgkltlen ikarglildd ekrakdykka kqiidkyhqf
61  f1eeilssvc isedllqnys dvyfklkksd ddnlqkdfks akdtikkqis eyikdsekfk
121 nlnfnqlida kkgqesdlil wlqskdngi elfkansdit didealeiik sfgkwtytyfk
181 gfhenrknvy ssndiptsii yrivddnlpk flenkakyes lkdkaepain yeqikkdlae
241 eltfddidyt sevnqrsvfl devfeianfn nylngsgitk fntiiggkfv ngentkrkgi
301 neyinlysqq indktlkkyk msvlfkqils dtesksfvid kleddsdvvt tmqsfyeqia
361 afktveeksi ketlsllfdd lkaqkldlsk iyfkndkslt dlsqqvfddy svigtavley
421 itqqiapknl dnpskkeqel iakktekaky lsletiklal eefnkhrrid kqcrfeeila
481 nfaaipmifd eiaqnkdnl qisikyqnqg kkdllqasae ddvkaikdl dqtlnllhkl
541 kifhisqsed kanildkdeh fylvfeecyf elanivplyn kirnyitqkp ysdekfklnf
601 enstlangwd knkepntai lfikddkyyl gymnkknki fddkaikenk gegykkivyk
661 llpgankmlp kvffsaksik fynpsedilr irnhsthtkn gspqkyekf efniedcrkf
721 idfykqsisk hpewkdfgfr fsdtqrynsi defyrevenq gykltfenis esyidsvvnq
781 gklylfqiyn kdfsayskgr pnlhtlywka lfdernlqdv vyklngeael fyrkqsipkk
841 ithpakeaia nknkdnpkke svfeydlikd krftedkfff hcpitinfsk sgankfndei

```

901 nlllkekand vhilsidrg rhlayytlvd gkgniikqdt fniigndrmk tnyhdklaai
 961 ekdrdsarkd wkkinnikem kegylsqvvh eiaklvieyn aivvfedlnf gfkrgfrkve
 1021 kqvyqklekm liekllylvf kdnefdktgg vlayqltap fetfkmgkq tgiyyvpag
 1081 ftskicpvtg fvnqlypkve svksqeffs kfdkicynld kgyfefsfdy knfgdkaakg
 1141 kwtfasfgr linfrnsdkn hnwdtrevyp tkelekllkd ysieyghgec ikaaicgesd
 1201 kkffakltsv lntilqmrns ktgteldyli spvadvngnf fdsrqapknm pqdadangay
 1261 higlkgmlml griknqegk klnlviknee yfefvqnrnn (SEQ ID NO:9)

[0197] "CpfI function" can be defined by any of a number of assays including, but not limited to, fluorescence polarization-based nucleic acid bind assays, fluorescence polarization-based strand invasion assays, transcription assays, EGFP disruption assays, DNA cleavage assays, and/or Surveyor assays, for example, as described herein.

[0198] By "CpfI nucleic acid molecule" is meant a polynucleotide encoding aCpfI polypeptide or fragment thereof. An exemplary CpfI nucleic acid molecule sequence is provided at GenBank Accession No. CP009633, nucleotides 652838 – 656740 and is shown below.

1 atgtcaatttcaagaatttgtaataaatatagtttaagtaaaactctaagatttgag
 61 ttaatccacagggttaaacacttgaaaacataaaagcaagaggtttgatttagatgat
 121 gagaaaagagctaaagactacaaaaggctaaacaataattgataaatatcatcagttt
 181 tttatagaggagatattaagttcggttgtattagcgaagattattacaaaactattct
 241 gatgtttattttaacttaaaaagagtgtatgataatctacaaaagattttaaaagt
 301 gcaaaagatacgataaagaacaaatatctgaatatataaaggactcagagaaatttaag
 361 aattgtttaatcaaaaccttatcgatgctaaaaaagggaagagtcagatttaattcta
 421 tggctaaagcaatctaaggataatggtatagaactatttaagccaatagtgatcacaca
 481 gatatagatgaggcgttagaataatcaaatcttttaagggtggacaactattttaag
 541 gggtttcatgaaaatagaaaaatgtttatagtagcaatgatattcctacatctattatt
 601 tataggatagtagatgataatttgcctaaatttctagaaaataaagctaagtatgagagt
 661 ttaaaagacaaagctccagaagctataaactatgaacaaattaaaaaagatttggcagaa
 721 gagctaacctttgatattgactacaaaacatctgaagttaatcaaagagtttttcactt
 781 gatgaagttttgagatagcaaactttaataattatctaaatcaaagtggtattactaaa
 841 ttaatactattattggtggttaattgtaaatggtgaaatacaaagagaaaaggata
 901 aatgaatatataaatctatactcacagcaaataatgataaaacactcaaaaaatataaa
 961 atgagtgtttttaagcaaatttaagtatacagaatctaaatcttttgtaattgat

1021 aagttagaagatgatagtgatgttagtacaacgatgcaaagttttatgagcaaatagca
1081 gcttttaaacagtagaagaaaaatctattaaagaacactatctttattttgatgat
1141 ttaaaagctcaaaaacttgatttgagtaaaattttttaaaatgataaatctcttact
1201 gatctatcacaacaagttttgatgattatagtgatttggtagcagcggtagtagaatat
1261 ataactcaaaaatagcacctaaaaatcttgataaccctagtaagaagagcaagaatta
1321 atagccaaaaaactgaaaagcaaaatacttctctagaaactataaagcttgcccta
1381 gaagaatttaataagcatagagatataagataaacagtgtaggttgagaataacttgca
1441 aactttgcggctattccgatgatatttgatgaatagctcaaaacaaagacaatttggca
1501 cagatatctatcaaatatcaaatcaaggtaaaaaagacactactcaagctagtgcggaa
1561 gatgatgttaaagctatcaaggatcttttagatcaaactaataatctcttacataaacta
1621 aaaatatttcatattagtcagtcagaagataaggcaaatatttagacaaggatgagcat
1681 tttatctagatatttgaggagtgtctactttgagctagcgaatatagtgcctctttataac
1741 aaaattagaactatataactcaaaagccatatagtgatgagaaatttaagctcaatttt
1801 gagaactcgactttggctaattggtgggataaaaaataaagagcctgacaatacggcaatt
1861 ttatttatcaaagatgataaatattatctgggtgtgatgaataagaaaaataacaaaata
1921 ttgatgataaagctatcaagaaaaataaaggcgagggtataaaaaaattgtttataaa
1981 cttttacctggcgcaataaaatgttacctaaggttttctttctgctaaatctataaaa
2041 tttataatcctagtgaagataacttagaataagaaatcattccacacatacaaaaaat
2101 ggtagtcctcaaaaaggatatgaaaaatttgagttaattgaagattgccgaaaattt
2161 atagattttataaacagctctataagtaagcatccggagtggaaagattttggatttaga
2221 tttctgatactcaaagatataattctatagatgaattttatagagaagttgaaatcaa
2281 ggctacaactaacttttgaatatatcagagagctatattgtagcgtagttaatcag
2341 ggtaaattgtacctattccaatctataataaagattttcagcttatagcaaagggcga
2401 ccaaattacatactttatattggaaagcgtgtttgatgagagaaatctcaagatgtg
2461 gttataagctaaatggtgaggcagagctttttatcgtaaacatcaatacctaaaaaa
2521 atcactcaccagctaaagaggcaatagctaataaaaacaaagataatcctaaaaaagag
2581 agtgttttgaaatgattttaatacaagataaacgcttactgaagataagttttcttt
2641 cactgtcctattacaatcaattttaaatctagtggagctaataagtttaatgatgaaatc
2701 aatttattgctaaaagaaaaagcaaatgatgttcataatattaagtatagatagagtgaa
2761 agacatttagcttactatactttggtagatggtaaggcaatatcatcaacaagatact
2821 ttcaacatcattggtaatgatagaatgaaaacaaactacatgataagcttgctgcaata
2881 gagaagatagggattcagctaggaaagactggaaaaagataataacatcaaagagatg

2941 aaagaggctatctatctcaggtagttcatgaaatagctaagctagttatagagtataat
 3001 gctattgtggttttgaggatttaattttgattaaaagagggcgttcaaggtagag
 3061 aagcaggctatcaaaagttagaaaaaatgctaattgagaaactaaactatctagttttc
 3121 aaagataatgagtttgataaaactgggggagtgcttagagcttatcagctaacagcacct
 3181 ttgagacttttaaaagatgggtaacaaacaggtattatctactatgtaccagctggt
 3241 ttacttcaaaaattgtcctgtaactggtttgtaaatcagttatatcctaagtatgaa
 3301 agtgtcagcaaatctcaagagtctttagtaagttgacaagattgttataaccttgat
 3361 aagggtattttgagtttagtttgattataaaaactttggtgacaaggctgccaaggc
 3421 aagtgactatagctagcttgggagtagattgattaactttagaaattcagataaaat
 3481 cataattgggatactcgagaagtttatccaactaaagagttggagaaattgctaaaagat
 3541 tattctatcgaaatgggcatggcgaaatgtatcaaaacagctatttgcggtgagagcgac
 3601 aaaaagtttttgctaagctaactagtgctctaaatactatcttacaatgcgtaactca
 3661 aaacaggtactgagtttagattatctaatttcaccagtagcagatgtaaatggcaatttc
 3721 ttgattcgcgacaggcgccaaaaaatatgcctcaagatgctgatccaatggtgcttat
 3781 catattgggctaaaaggctgatgctactaggtaggatcaaaaataatcaagagggcaaa
 3841 aaactcaatttggtatcaaaaatgaagagtattttgagttcgtgcagaataggaataac
 3901 taa (SEQ ID NO:10)

[0199] In the context of formation of a CRISPR complex, “target sequence” refers to a sequence to which a guide sequence is designed to have complementarity, where hybridization between a target sequence and a guide sequence promotes the formation of a CRISPR complex. A target sequence may comprise RNA polynucleotides. The term “target RNA” refers to a RNA polynucleotide being or comprising the target sequence. In other words, the target RNA may be a RNA polynucleotide or a part of a RNA polynucleotide to which a part of the gRNA, i.e. the guide sequence, is designed to have complementarity and to which the effector function mediated by the complex comprising CRISPR effector protein and a gRNA is to be directed. In some embodiments, a target sequence is located in the nucleus or cytoplasm of a cell.

[0200] In certain example embodiments, the CRISPR effector protein may be delivered using a nucleic acid molecule encoding the CRISPR effector protein. The nucleic acid molecule encoding a CRISPR effector protein, may advantageously be a codon optimized CRISPR effector protein. An example of a codon optimized sequence, is in this instance a sequence optimized for expression in eukaryote, e.g., humans (i.e. being optimized for

expression in humans), or for another eukaryote, animal or mammal as herein discussed; see, e.g., SaCas9 human codon optimized sequence in WO 2014/093622 (PCT/US2013/074667). Whilst this is preferred, it will be appreciated that other examples are possible and codon optimization for a host species other than human, or for codon optimization for specific organs is known. In some embodiments, an enzyme coding sequence encoding a CRISPR effector protein is a codon optimized for expression in particular cells, such as eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a plant or a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In some embodiments, processes for modifying the germ line genetic identity of human beings and/or processes for modifying the genetic identity of animals which are likely to cause them suffering without any substantial medical benefit to man or animal, and also animals resulting from such processes, may be excluded. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g. about or more than about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the “Codon Usage Database” available at kazusa.or.jp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. “Codon usage tabulated from the international DNA sequence databases: status for the year 2000” Nucl. Acids Res. 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. In some embodiments, one or more codons (e.g. 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or

more, or all codons) in a sequence encoding a Cas correspond to the most frequently used codon for a particular amino acid.

[0201] In certain embodiments, the effector protein (CRISPR enzyme; Cas9, Cas12a) according to the invention as described herein is associated with or fused to a destabilization domain (DD). In some embodiments, the DD is ER50. A corresponding stabilizing ligand for this DD is, in some embodiments, 4HT. As such, in some embodiments, one of the at least one DDs is ER50 and a stabilizing ligand therefor is 4HT or CMP8. In some embodiments, the DD is DHFR50. A corresponding stabilizing ligand for this DD is, in some embodiments, TMP. As such, in some embodiments, one of the at least one DDs is DHFR50 and a stabilizing ligand therefor is TMP. In some embodiments, the DD is ER50. A corresponding stabilizing ligand for this DD is, in some embodiments, CMP8. CMP8 may therefore be an alternative stabilizing ligand to 4HT in the ER50 system. While it may be possible that CMP8 and 4HT can/should be used in a competitive matter, some cell types may be more susceptible to one or the other of these two ligands, and from this disclosure and the knowledge in the art the skilled person can use CMP8 and/or 4HT.

[0202] In some embodiments, one or two DDs may be fused to the N- terminal end of a CRISPR enzyme with one or two DDs fused to the C- terminal of the CRISPR enzyme. In some embodiments, the at least two DDs are associated with the CRISPR enzyme and the DDs are the same DD, i.e. the DDs are homologous. Thus, both (or two or more) of the DDs could be ER50 DDs. This is preferred in some embodiments. Alternatively, both (or two or more) of the DDs could be DHFR50 DDs. This is also preferred in some embodiments. In some embodiments, the at least two DDs are associated with the CRISPR enzyme and the DDs are different DDs, i.e. the DDs are heterologous. Thus, one of the DDs could be ER50 while one or more of the DDs or any other DDs could be DHFR50. Having two or more DDs which are heterologous may be advantageous as it would provide a greater level of degradation control. A tandem fusion of more than one DD at the N or C-term may enhance degradation; and such a tandem fusion can be, for example ER50-ER50-C2c1. It is envisaged that high levels of degradation would occur in the absence of either stabilizing ligand, intermediate levels of degradation would occur in the absence of one stabilizing ligand and the presence of the other (or another) stabilizing ligand, while low levels of degradation would occur in the presence of both (or two of more) of the stabilizing ligands. Control may also be imparted by having an N-terminal ER50 DD and a C-terminal DHFR50 DD.

[0203] In certain embodiments, the methods as described herein may comprise providing a Cas transgenic cell in which one or more nucleic acids encoding one or more guide RNAs are provided or introduced operably connected in the cell with a regulatory element comprising a promoter of one or more gene of interest. As used herein, the term “Cas transgenic cell” refers to a cell, such as a eukaryotic cell, in which a Cas gene has been genomically integrated. The nature, type, or origin of the cell are not particularly limiting according to the present invention. Also the way the Cas transgene is introduced in the cell may vary and can be any method as is known in the art. In certain embodiments, the Cas transgenic cell is obtained by introducing the Cas transgene in an isolated cell. In certain other embodiments, the Cas transgenic cell is obtained by isolating cells from a Cas transgenic organism. By means of example, and without limitation, the Cas transgenic cell as referred to herein may be derived from a Cas transgenic eukaryote, such as a Cas knock-in eukaryote. Reference is made to WO 2014/093622 (PCT/US13/74667), incorporated herein by reference. Methods of US Patent Publication Nos. 20120017290 and 20110265198 assigned to Sangamo BioSciences, Inc. directed to targeting the Rosa locus may be modified to utilize the CRISPR Cas system of the present invention. Methods of US Patent Publication No. 20130236946 assigned to Collectis directed to targeting the Rosa locus may also be modified to utilize the CRISPR Cas system of the present invention. By means of further example reference is made to Platt et. al. (Cell; 159(2):440-455 (2014)), describing a Cas9 knock-in mouse, which is incorporated herein by reference. The Cas transgene can further comprise a Lox-Stop-polyA-Lox(LSL) cassette thereby rendering Cas expression inducible by Cre recombinase. Alternatively, the Cas transgenic cell may be obtained by introducing the Cas transgene in an isolated cell. Delivery systems for transgenes are well known in the art. By means of example, the Cas transgene may be delivered in for instance eukaryotic cell by means of vector (e.g., AAV, adenovirus, lentivirus) and/or particle and/or nanoparticle delivery, as also described herein elsewhere.

[0204] It will be understood by the skilled person that the cell, such as the Cas transgenic cell, as referred to herein may comprise further genomic alterations besides having an integrated Cas gene or the mutations arising from the sequence specific action of Cas when complexed with RNA capable of guiding Cas to a target locus.

[0205] In certain aspects the invention involves vectors, e.g. for delivering or introducing in a cell Cas and/or RNA capable of guiding Cas to a target locus (i.e. guide RNA), but also

for propagating these components (e.g. in prokaryotic cells). 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, the term “vector” refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g. circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a “plasmid,” which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally-derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g. retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses (AAVs)). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g. bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as “expression vectors.” Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0206] Recombinant expression vectors can comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which may be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, “operably linked” is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g. in an in vitro transcription/translation system or in a host cell when the vector is introduced into the

host cell). With regards to recombination and cloning methods, mention is made of U.S. patent application 10/815,730, published September 2, 2004 as US 2004-0171156 A1, the contents of which are herein incorporated by reference in their entirety. Thus, the embodiments disclosed herein may also comprise transgenic cells comprising the CRISPR effector system. In certain example embodiments, the transgenic cell may function as an individual discrete volume. In other words samples comprising a masking construct may be delivered to a cell, for example in a suitable delivery vesicle and if the target is present in the delivery vesicle the CRISPR effector is activated and a detectable signal generated.

[0207] The vector(s) can include the regulatory element(s), e.g., promoter(s). The vector(s) can comprise Cas encoding sequences, and/or a single, but possibly also can comprise at least 3 or 8 or 16 or 32 or 48 or 50 guide RNA(s) (e.g., sgRNAs) encoding sequences, such as 1-2, 1-3, 1-4 1-5, 3-6, 3-7, 3-8, 3-9, 3-10, 3-8, 3-16, 3-30, 3-32, 3-48, 3-50 RNA(s) (e.g., sgRNAs). In a single vector there can be a promoter for each RNA (e.g., sgRNA), advantageously when there are up to about 16 RNA(s); and, when a single vector provides for more than 16 RNA(s), one or more promoter(s) can drive expression of more than one of the RNA(s), e.g., when there are 32 RNA(s), each promoter can drive expression of two RNA(s), and when there are 48 RNA(s), each promoter can drive expression of three RNA(s). By simple arithmetic and well established cloning protocols and the teachings in this disclosure one skilled in the art can readily practice the invention as to the RNA(s) for a suitable exemplary vector such as AAV, and a suitable promoter such as the U6 promoter. For example, the packaging limit of AAV is ~4.7 kb. The length of a single U6-gRNA (plus restriction sites for cloning) is 361 bp. Therefore, the skilled person can readily fit about 12-16, e.g., 13 U6-gRNA cassettes in a single vector. This can be assembled by any suitable means, such as a golden gate strategy used for TALE assembly (genome-engineering.org/taleffectors/). The skilled person can also use a tandem guide strategy to increase the number of U6-gRNAs by approximately 1.5 times, e.g., to increase from 12-16, e.g., 13 to approximately 18-24, e.g., about 19 U6-gRNAs. Therefore, one skilled in the art can readily reach approximately 18-24, e.g., about 19 promoter-RNAs, e.g., U6-gRNAs in a single vector, e.g., an AAV vector. A further means for increasing the number of promoters and RNAs in a vector is to use a single promoter (e.g., U6) to express an array of RNAs separated by cleavable sequences. And an even further means for increasing the number of promoter-RNAs in a vector, is to express an array of promoter-RNAs separated by cleavable

sequences in the intron of a coding sequence or gene; and, in this instance it is advantageous to use a polymerase II promoter, which can have increased expression and enable the transcription of long RNA in a tissue specific manner. (see, e.g., nar.oxfordjournals.org/content/34/7/e53.short and nature.com/mt/journal/v16/n9/abs/mt2008144a.html). In an advantageous embodiment, AAV may package U6 tandem gRNA targeting up to about 50 genes. Accordingly, from the knowledge in the art and the teachings in this disclosure the skilled person can readily make and use vector(s), e.g., a single vector, expressing multiple RNAs or guides under the control or operatively or functionally linked to one or more promoters—especially as to the numbers of RNAs or guides discussed herein, without any undue experimentation.

[0208] The guide RNA(s) encoding sequences and/or Cas encoding sequences, can be functionally or operatively linked to regulatory element(s) and hence the regulatory element(s) drive expression. The promoter(s) can be constitutive promoter(s) and/or conditional promoter(s) and/or inducible promoter(s) and/or tissue specific promoter(s). The promoter can be selected from the group consisting of RNA polymerases, pol I, pol II, pol III, T7, U6, H1, retroviral Rous sarcoma virus (RSV) LTR promoter, the cytomegalovirus (CMV) promoter, the SV40 promoter, the dihydrofolate reductase promoter, the β -actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1 α promoter. An advantageous promoter is the promoter is U6.

[0209] Additional effectors for use according to the invention can be identified by their proximity to cas1 genes, for example, though not limited to, within the region 20 kb from the start of the cas1 gene and 20 kb from the end of the cas1 gene. In certain embodiments, the effector protein comprises at least one HEPN domain and at least 500 amino acids, and wherein the C2c2 effector protein is naturally present in a prokaryotic genome within 20 kb upstream or downstream of a Cas gene or a CRISPR array. Non-limiting examples of Cas proteins include Cas1, Cas1B, Cas2, Cas3, Cas4, Cas5, Cas6, Cas7, Cas8, Cas9 (also known as Csn1 and Csx12), Cas10, Csy1, Csy2, Csy3, Cse1, Cse2, Csc1, Csc2, Csa5, Csn2, Csm2, Csm3, Csm4, Csm5, Csm6, Cmr1, Cmr3, Cmr4, Cmr5, Cmr6, Csb1, Csb2, Csb3, Csx17, Csx14, Csx10, Csx16, CsaX, Csx3, Csx1, Csx15, Csf1, Csf2, Csf3, Csf4, homologues thereof, or modified versions thereof. In certain example embodiments, the C2c2 effector protein is naturally present in a prokaryotic genome within 20kb upstream or downstream of a Cas 1 gene. The terms “orthologue” (also referred to as “ortholog” herein) and

“homologue” (also referred to as “homolog” herein) are well known in the art. By means of further guidance, a “homologue” of a protein as used herein is a protein of the same species which performs the same or a similar function as the protein it is a homologue of. Homologous proteins may but need not be structurally related, or are only partially structurally related. An “orthologue” of a protein as used herein is a protein of a different species which performs the same or a similar function as the protein it is an orthologue of. Orthologous proteins may but need not be structurally related, or are only partially structurally related.

Guides

[0210] 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.

[0211] The terms “guide molecule,” “guide RNA,” and “guide” are used interchangeably herein to refer to nucleic acid-based molecules, including but not limited to RNA-based molecules that are capable of forming a complex with a CRISPR-Cas protein and comprise 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.

[0212] 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., 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. In the context of deaminase conjugates the target nucleic acid sequence or target sequence is the sequence comprising the target adenosine to be deaminated also

referred to herein as the “target adenosine”. In some embodiments, the complementarity described herein above excludes an intended mismatch, such as the dA-C mismatch described herein.

[0213] In some embodiments, a nucleic acid-targeting guide (direct repeat and/or spacer) is selected to reduce the degree 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).

[0214] 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.

[0215] 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 a heteroduplex 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.

[0216] In a particular embodiment the guide molecule comprises a guide sequence linked to a direct repeat sequence, wherein the direct repeat sequence comprises one or more stem loops or optimized secondary structures. In particular 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 loops or optimized secondary structures. In particular embodiments the guide molecule comprises or consists of the guide sequence linked to all or part of the natural direct repeat sequence. A typical Type V or Type VI CRISPR-cas guide molecule comprises (in 3' to 5' direction or in 5' to 3' direction): a guide sequence a first complimentary stretch (the "repeat"), a loop (which is typically 4 or 5 nucleotides long), a second complimentary stretch (the "anti-repeat" being complimentary to the repeat), and a poly A (often poly U in RNA) tail (terminator). In certain embodiments, the direct repeat sequence retains its natural architecture and forms a single stem loop. In particular embodiments, certain aspects of the 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 guide molecule modifications, including but not limited to insertions, deletions, and substitutions include guide termini and regions of the guide molecule that are exposed when complexed with the CRISPR-Cas protein and/or target, for example the stemloop of the direct repeat sequence.

[0217] In particular embodiments, 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 X2-10 and Y2-10 (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 loop 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 guide molecule is preserved. In one aspect, the loop that connects the stem made of X:Y basepairs can be any sequence of the same length (e.g., 4 or 5 nucleotides) or longer that does not interrupt the overall secondary structure of the guide molecule. In one aspect, the stemloop can further comprise, e.g. an MS2 aptamer. In one aspect, the stem comprises about 5-7bp comprising complementary X and Y sequences, although stems of more or fewer basepairs are also contemplated. In one aspect, non-Watson Crick basepairing is contemplated, where such pairing otherwise generally preserves the architecture of the stemloop at that position.

[0218] In particular embodiments the natural hairpin or stemloop structure of the guide molecule is extended or replaced by an extended stemloop. It has been demonstrated that extension of the stem can enhance the assembly of the guide molecule with the CRISPR-Cas protein (Chen et al. *Cell*. (2013); 155(7): 1479–1491). In particular embodiments the stem of the stemloop is extended by at least 1, 2, 3, 4, 5 or more complementary basepairs (i.e. corresponding to the addition of 2, 4, 6, 8, 10 or more nucleotides in the guide molecule). In particular embodiments these are located at the end of the stem, adjacent to the loop of the stemloop.

[0219] In particular embodiments, the susceptibility of the guide molecule to RNases or to decreased expression can be reduced by slight modifications of the sequence of the guide molecule which do not affect its function. For instance, in particular embodiments, premature termination of transcription, such as premature transcription of U6 Pol-III, can be removed by modifying a putative Pol-III terminator (4 consecutive U's) in the guide molecules sequence. Where such sequence modification is required in the stemloop of the guide molecule, it is preferably ensured by a basepair flip.

[0220] In a particular embodiment the direct repeat may be modified to comprise one or more protein-binding RNA aptamers. In a particular embodiment, one or more aptamers may be included such as part of optimized secondary structure. Such aptamers may be capable of binding a bacteriophage coat protein as detailed further herein.

[0221] In some embodiments, the guide molecule forms a duplex with a target RNA comprising at least one target cytosine residue to be edited. Upon hybridization of the guide RNA molecule to the target RNA, the cytidine deaminase binds to the single strand RNA in the duplex made accessible by the mismatch in the guide sequence and catalyzes deamination of one or more target cytosine residues comprised within the stretch of mismatching nucleotides.

[0222] 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 mRNA.

[0223] In certain embodiments, the guide sequence or spacer length of the guide molecules is from 15 to 50 nt, more particularly from 15 to 35 nt in length. 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. In certain example embodiment, the guide sequence is 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 nt.

[0224] In some embodiments, the guide sequence is an RNA sequence of between 10 to 50 nt in length, but more particularly of about 20-30 nt advantageously about 20 nt, 23-25 nt or 24 nt. The guide sequence is selected so as to ensure that it hybridizes to the target sequence. This is described more in detail below. Selection can encompass further steps which increase efficacy and specificity.

[0225] In some embodiments, the guide sequence has a canonical length (e.g., about 15-30 nt) is used to hybridize with the target RNA or DNA. In some embodiments, a guide molecule is longer than the canonical length (e.g., >30 nt) is used to hybridize with the target RNA or DNA, such that a region of the guide sequence hybridizes with a region of the RNA or DNA strand outside of the Cas-guide target complex. This can be of interest where additional modifications, such as deamination of nucleotides is of interest. In alternative embodiments, it is of interest to maintain the limitation of the canonical guide sequence length.

[0226] In some embodiments, it is of interest to reduce the susceptibility of the guide molecule to RNA cleavage, such as to cleavage by Cas13. Accordingly, in particular embodiments, the guide molecule is adjusted to avoid cleavage by Cas13 or other RNA-cleaving enzymes.

[0227] 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 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 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.

[0228] 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.

[0229] 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.

[0230] In general, degree of complementarity is with reference to the optimal alignment of the guide 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 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.

[0231] 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 (Ψ), N¹-methylpseudouridine (me¹ Ψ), 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, or C2c1. 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).

[0232] In some embodiments, the modification to the guide is a chemical modification, an insertion, a deletion or a split. 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 some embodiments, the chemical modification includes, but is not limited to, incorporation of 2'-O-methyl (M) analogs, 2'-deoxy analogs, 2-thiouridine analogs, N6-methyladenosine analogs, 2'-fluoro analogs, 2-aminopurine, 5-bromo-uridine, pseudouridine (Ψ), N1-methylpseudouridine (me1 Ψ), 5-methoxyuridine(5moU), inosine, 7-methylguanosine, 2'-O-methyl 3'phosphorothioate (MS), S-constrained ethyl(cEt), phosphorothioate (PS), or 2'-O-methyl 3'thioPACE (MSP). In some embodiments, the guide comprises one or more of phosphorothioate modifications. In certain embodiments, at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, or 25 nucleotides of the guide are chemically modified. In certain embodiments, one or more nucleotides in the seed region are chemically modified. In certain embodiments, one or more nucleotides in the 3'-terminus are chemically modified. In certain embodiments, none of the nucleotides in the 5'-handle is chemically modified. In some embodiments, the chemical

modification in the seed region is a minor modification, such as incorporation of a 2'-fluoro analog. In a specific embodiment, one nucleotide of the seed region is replaced with a 2'-fluoro analog. In some embodiments, 5 to 10 nucleotides in the 3'-terminus are chemically modified. Such chemical modifications at the 3'-terminus of the Cas13 CrRNA may improve Cas13 activity. In a specific embodiment, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides in the 3'-terminus are replaced with 2'-fluoro analogues. In a specific embodiment, 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 nucleotides in the 3'-terminus are replaced with 2'-O-methyl (M) analogs.

[0233] 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 modified loop comprises 3, 4, or 5 nucleotides. In certain embodiments, the loop comprises the sequence of UCUU, UUUU, UAUU, or UGUU.

[0234] In some embodiments, the guide molecule forms a stemloop with a separate non-covalently linked sequence, which can be DNA or RNA. In particular embodiments, the sequences forming the guide are first synthesized using the standard phosphoramidite synthetic protocol (Herdewijn, P., ed., *Methods in Molecular Biology* Col 288, *Oligonucleotide Synthesis: Methods and Applications*, Humana Press, New Jersey (2012)). In some embodiments, these 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 this sequence is functionalized, a covalent chemical bond or linkage can be formed between this sequence and the direct repeat sequence. 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.

[0235] In some embodiments, these stem-loop forming 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).

[0236] In certain embodiments, the guide molecule comprises (1) a guide sequence capable of hybridizing to a target locus and (2) a tracr mate or direct repeat sequence whereby the direct repeat sequence is located upstream (i.e., 5') from the guide sequence. In a particular embodiment the seed sequence (i.e. the sequence essential critical for recognition and/or hybridization to the sequence at the target locus) of the guide sequence is approximately within the first 10 nucleotides of the guide sequence.

[0237] 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).

[0238] In certain embodiments, the target sequence should be associated with a PAM (protospacer adjacent motif) or PFS (protospacer flanking sequence or site); that is, a short sequence recognized by the CRISPR complex. Depending on the nature of the CRISPR-Cas protein, the target sequence should be selected such that its complementary sequence in the DNA duplex (also referred to herein as the non-target sequence) is upstream or downstream of the PAM. In the embodiments of the present invention where the CRISPR-Cas protein is a Cas13 protein, the complementary sequence of the target sequence is downstream or 3' of the PAM or upstream or 5' of the PAM. The precise sequence and length requirements for the PAM differ depending on the Cas13 protein used, but PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence). Examples of the natural PAM sequences for different Cas13 orthologues are provided herein below and the skilled person will be able to identify further PAM sequences for use with a given Cas13 protein.

[0239] Further, engineering of the PAM Interacting (PI) domain may allow programming of PAM specificity, improve target site recognition fidelity, and increase the versatility of the CRISPR-Cas protein, for example as described for Cas9 in Kleinstiver BP et al. Engineered CRISPR-Cas9 nucleases with altered PAM specificities. *Nature*. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592. As further detailed herein, the skilled person will understand that Cas13 proteins may be modified analogously.

[0240] In particular embodiment, the guide is an escorted guide. By "escorted" is meant that the CRISPR-Cas system or complex or guide is delivered to a selected time or place

within a cell, so that activity of the CRISPR-Cas system or complex or guide is spatially or temporally controlled. For example, the activity and destination of the 3 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.

[0241] The escorted CRISPR-Cas systems or complexes have a guide molecule with a functional structure designed to improve guide molecule structure, architecture, stability, genetic expression, or any combination thereof. Such a structure can include an aptamer.

[0242] 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).

[0243] Accordingly, in particular embodiments, the guide molecule is modified, e.g., by one or more aptamer(s) designed to improve guide molecule 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 molecule deliverable, inducible or responsive to a selected effector. The invention accordingly comprehends an guide molecule 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.

[0244] 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.

[0245] 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.

[0246] 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.

[0247] There are several different designs of this chemical inducible system: 1. ABI-PYL based system inducible by Absciscic 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>).

[0248] 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.

[0249] 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.

[0250] 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.

[0251] 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.

[0252] 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).

[0253] 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.

[0254] 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).

[0255] 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).

[0256] 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.

[0257] 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.

[0258] 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.

[0259] 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.

[0260] 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.

[0261] 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.

[0262] 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]).

[0263] 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.

[0264] 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.

[0265] 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.

[0266] 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.

[0267] 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.

[0268] 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.

[0269] 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.

[0270] 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.

[0271] 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.

[0272] 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.

[0273] In particular embodiments, the guide molecule is modified by a secondary structure to increase the specificity of the CRISPR-Cas system and the secondary structure can protect against exonuclease activity and allow for 5' additions to the guide sequence also referred to herein as a protected guide molecule.

[0274] 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.

[0275] 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.

Base Editing

[0276] The compositions herein may comprise a Cas protein or a variant thereof (e.g., inactive or dead Cas) fused with one or more functional domains. The Cas protein may be a dead Cas protein or a Cas nickase protein. For example, the compositions herein may comprise one or more components of a base editing system. In some embodiments, compositions herein comprise nucleotide sequence comprising encoding sequences for one or more components of a base editing system. A base-editing system may comprise a deaminase (e.g., an adenosine deaminase or cytidine deaminase) fused with a Cas protein or a variant

thereof. In certain examples, the system comprises a mutated form of an adenosine deaminase fused with a dead CRISPR-Cas or CRISPR-Cas nickase. The mutated form of the adenosine deaminase may have both adenosine deaminase and cytidine deaminase activities. Examples of base editing systems include those described in WO2019071048, WO2019084063, WO2019126716, WO2019126709, WO2019126762, WO2019126774, Cox DBT, et al., RNA editing with CRISPR-Cas13, *Science*. 2017 Nov 24;358(6366):1019-1027; Abudayyeh OO, et al., A cytosine deaminase for programmable single-base RNA editing, *Science* 26 Jul 2019: Vol. 365, Issue 6451, pp. 382-386; Gaudelli NM et al., Programmable base editing of A•T to G•C in genomic DNA without DNA cleavage, *Nature* volume 551, pages 464–471 (23 November 2017); Komor AC, et al., Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. 2016 May 19;533(7603):420-4.

Applications in non-animal organisms and animals

[0277] The small molecule compounds herein may be used to modulate the functions and activities of RNA-guided nuclease (e.g., Cas proteins), variants thereof, and fragments thereof in animals and non-animal organisms. In some examples, the animals and non-animal organisms may have been engineered to constitutively or inducibly express an RNA-guided nuclease (e.g., Cas protein). In some examples, the compounds herein may modulate the activities of the RNA-guided nucleases or their interaction with other molecules, e.g., their binding with target polynucleotides.

Applications in non-animal organisms

[0278] The small molecule compounds herein can be used on non-animal organisms, such as plants and fungi. The small molecule compounds herein can be used to perform efficient and cost effective plant gene or genome interrogation or editing or manipulation—for instance, for rapid investigation and/or selection and/or interrogations and/or comparison and/or manipulations and/or transformation of plant genes or genomes; e.g., to create, identify, develop, optimize, or confer trait(s) or characteristic(s) to plant(s) or to transform a plant genome. There can accordingly be improved production of plants, new plants with new combinations of traits or characteristics or new plants with enhanced traits.

[0279] The compounds disclosed herein may be used on plants. In general, the term “plant” relates to any various photosynthetic, eukaryotic, unicellular or multicellular organism of the kingdom Plantae characteristically growing by cell division, containing

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

[0280] The small molecule compounds herein can be used with nucleases to confer desired traits on essentially any plant. A wide variety of plants and plant cell systems may be engineered for the desired physiological and agronomic characteristics described herein using the nucleic acid constructs of the present disclosure and the various transformation methods mentioned above. In preferred embodiments, target plants and plant cells include, but are not limited to, those monocotyledonous and dicotyledonous plants, such as crops including grain crops (e.g., wheat, maize, rice, millet, barley), fruit crops (e.g., tomato, apple, pear, strawberry, orange), forage crops (e.g., alfalfa), root vegetable crops (e.g., carrot, potato, sugar beets, yam), leafy vegetable crops (e.g., lettuce, spinach); flowering plants (e.g., petunia, rose, chrysanthemum), conifers and pine trees (e.g., pine fir, spruce); plants used in phytoremediation (e.g., heavy metal accumulating plants); oil crops (e.g., sunflower, rape seed) and plants used for experimental purposes (e.g., Arabidopsis). Plant cells and tissues for engineering include, without limitation, roots, stems, leaves, flowers, and reproductive structures, undifferentiated meristematic cells, parenchyma, collenchyma, sclerenchyma,

xylem, phloem, epidermis, and germplasm. Thus, the methods and systems can be used over a broad range of plants, such as for example with dicotyledonous plants belonging to the orders Magnoliales, Illiciales, Laurales, Piperales, Aristochiales, Nymphaeales, Ranunculales, Papaverales, Sarraceniaceae, Trochodendrales, Hamamelidales, Eucomiales, Leitneriales, Myricales, Fagales, Casuarinales, Caryophyllales, Batales, Polygonales, Plumbaginales, Dilleniales, Theales, Malvales, Urticales, Lecythidales, Violales, Salicales, Capparales, Ericales, Diapensales, Ebenales, Primulales, Rosales, Fabales, Podostemales, Haloragales, Myrtales, Cornales, Proteales, Santales, Rafflesiales, Celastrales, Euphorbiales, Rhamnales, Sapindales, Juglandales, Geraniales, Polygalales, Umbellales, Gentianales, Polemoniales, Lamiales, Plantaginales, Scrophulariales, Campanulales, Rubiales, Dipsacales, and Asterales; the methods and systems can be used with monocotyledonous plants such as those belonging to the orders Alismatales, Hydrocharitales, Najadales, Triuridales, Commelinales, Eriocaulales, Restionales, Poales, Juncals, Cyperales, Typhales, Bromeliales, Zingiberales, Arecales, Cyclanthales, Pandanales, Arales, Lilliales, and Orchidales, or with plants belonging to Gymnospermae, e.g. those belonging to the orders Pinales, Ginkgoales, Cycadales, Araucariales, Cupressales and Gnetales.

[0281] The small molecule inhibitors herein can be used over a broad range of plant species, included in the non-limitative list of dicot, monocot or gymnosperm genera hereunder: *Atropa*, *Alseodaphne*, *Anacardium*, *Arachis*, *Beilschmiedia*, *Brassica*, *Carthamus*, *Cocculus*, *Croton*, *Cucumis*, *Citrus*, *Citrullus*, *Capsicum*, *Catharanthus*, *Cocos*, *Coffea*, *Cucurbita*, *Daucus*, *Duguetia*, *Eschscholzia*, *Ficus*, *Fragaria*, *Glaucium*, *Glycine*, *Gossypium*, *Helianthus*, *Hevea*, *Hyoscyamus*, *Lactuca*, *Landolphia*, *Linum*, *Litsea*, *Lycopersicon*, *Lupinus*, *Manihot*, *Majorana*, *Malus*, *Medicago*, *Nicotiana*, *Olea*, *Parthenium*, *Papaver*, *Persea*, *Phaseolus*, *Pistacia*, *Pisum*, *Pyrus*, *Prunus*, *Raphanus*, *Ricinus*, *Senecio*, *Sinomenium*, *Stephania*, *Sinapis*, *Solanum*, *Theobroma*, *Trifolium*, *Trigonella*, *Vicia*, *Vinca*, *Vitis*, and *Vigna*; and the genera *Allium*, *Andropogon*, *Aragrostis*, *Asparagus*, *Avena*, *Cynodon*, *Elaeis*, *Festuca*, *Festulolium*, *Heterocallis*, *Hordeum*, *Lemna*, *Lolium*, *Musa*, *Oryza*, *Panicum*, *Pennisetum*, *Phleum*, *Poa*, *Secale*, *Sorghum*, *Triticum*, *Zea*, *Abies*, *Cunninghamia*, *Ephedra*, *Picea*, *Pinus*, and *Pseudotsuga*.

[0282] The small molecule compounds herein can also be used over a broad range of "algae" or "algae cells"; including for example algae selected from several eukaryotic phyla, including the Rhodophyta (red algae), Chlorophyta (green algae), Phaeophyta (brown algae),

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

[0283] The small molecule compounds herein can also be used over a broad range of fungi. As used herein, a "fungal cell" refers to any type of eukaryotic cell within the kingdom of fungi. Phyla within the kingdom of fungi include Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota, Glomeromycota, Microsporidia, and Neocallimastigomycota. Fungal cells may include yeasts, molds, and filamentous fungi. In some embodiments, the fungal cell is a yeast cell. As used herein, the term "yeast" refers to any fungal cell within the phyla Ascomycota and Basidiomycota. Yeast cells may include budding yeast cells, fission yeast cells, and mold cells. Without being limited to these organisms, many types of yeast used in laboratory and industrial settings are part of the phylum Ascomycota. In some embodiments, the yeast cell is an *S. cerevisiae*, *Kluyveromyces marxianus*, or *Issatchenkia orientalis* cell. Other yeast cells may include without limitation *Candida* spp. (e.g., *Candida albicans*), *Yarrowia* spp. (e.g., *Yarrowia lipolytica*), *Pichia* spp. (e.g., *Pichia pastoris*), *Kluyveromyces* spp. (e.g., *Kluyveromyces lactis* and *Kluyveromyces marxianus*), *Neurospora* spp. (e.g., *Neurospora crassa*), *Fusarium* spp. (e.g., *Fusarium oxysporum*), and *Issatchenkia* spp. (e.g., *Issatchenkia orientalis*, a.k.a. *Pichia kudriavzevii* and *Candida acidothermophilum*). In some embodiments, the fungal cell is a filamentous fungal cell. As used herein, the term "filamentous fungal cell" refers to any type of fungal cell that grows in filaments, i.e., hyphae or mycelia. Examples of filamentous fungal cells may include without limitation *Aspergillus* spp. (e.g., *Aspergillus niger*), *Trichoderma* spp. (e.g., *Trichoderma reesei*), *Rhizopus* spp. (e.g., *Rhizopus oryzae*), and *Mortierella* spp. (e.g., *Mortierella isabellina*).

Plant cultures and regeneration

[0284] In particular embodiments, plant cells which have a modified genome and that are produced or obtained may be treated with the small molecule compounds herein, and can be cultured to regenerate a whole plant which possesses the transformed or modified genotype and thus the desired phenotype. Conventional regeneration techniques are well known to those skilled in the art. Particular examples of such regeneration techniques rely on manipulation of certain phytohormones in a tissue culture growth medium, and typically relying on a biocide and/or herbicide marker which has been introduced together with the desired nucleotide sequences. In further particular embodiments, plant regeneration is obtained from cultured protoplasts, plant callus, explants, organs, pollens, embryos or parts thereof (see e.g. Evans et al. (1983), Handbook of Plant Cell Culture, Klee et al (1987) Ann. Rev. of Plant Phys.).

[0285] In particular embodiments, transformed or improved plants to be treated with the small molecule compounds as described herein can be self-pollinated to provide seed for homozygous improved plants of the invention (homozygous for the DNA modification) or crossed with non-transgenic plants or different improved plants to provide seed for heterozygous plants. Where a recombinant DNA was introduced into the plant cell, the resulting plant of such a crossing is a plant which is heterozygous for the recombinant DNA molecule. Both such homozygous and heterozygous plants obtained by crossing from the improved plants and comprising the genetic modification (which can be a recombinant DNA) are referred to herein as "progeny". Progeny plants are plants descended from the original transgenic plant and containing the genome modification or recombinant DNA molecule introduced by the methods provided herein. Alternatively, genetically modified plants can be obtained by one of the methods described supra using the Cfp1 enzyme whereby no foreign DNA is incorporated into the genome. Progeny of such plants, obtained by further breeding may also contain the genetic modification. Breedings are performed by any breeding methods that are commonly used for different crops (e.g., Allard, Principles of Plant Breeding, John Wiley & Sons, NY, U. of CA, Davis, CA, 50-98 (1960).

Generation of plants with enhanced agronomic traits

[0286] The small molecule compounds herein can be used to treat plants that are introduced with targeted double-strand or single-strand breaks and/or introduced gene activator and or repressor systems and without being limitative, can be used for gene

targeting, gene replacement, targeted mutagenesis, targeted deletions or insertions, targeted inversions and/or targeted translocations. By co-expression of multiple targeting RNAs directed to achieve multiple modifications in a single cell, multiplexed genome modification can be ensured. This technology can be used to high-precision engineering of plants with improved characteristics, including enhanced nutritional quality, increased resistance to diseases and resistance to biotic and abiotic stress, and increased production of commercially valuable plant products or heterologous compounds.

[0287] In embodiments, the small molecule compounds may be used to treat organisms that are introduced targeted double-strand breaks (DSB) in an endogenous DNA sequence. The DSB activates cellular DNA repair pathways, which can be harnessed to achieve desired DNA sequence modifications near the break site. This is of interest where the inactivation of endogenous genes can confer or contribute to a desired trait. In particular embodiments, homologous recombination with a template sequence is promoted at the site of the DSB, in order to introduce a gene of interest.

[0288] The small molecule compounds may be used to treat improved plants. Improved plants may have one or more desirable traits compared to the wildtype plant. The small molecule compounds may further modulate these traits. In particular embodiments, the plants, plant cells or plant parts obtained are transgenic plants, comprising an exogenous DNA sequence incorporated into the genome of all or part of the cells of the plant. In particular embodiments, non-transgenic genetically modified plants, plant parts or cells are obtained, in that no exogenous DNA sequence is incorporated into the genome of any of the plant cells of the plant. In such embodiments, the improved plants are non-transgenic. Where only the modification of an endogenous gene is ensured and no foreign genes are introduced or maintained in the plant genome, the resulting genetically modified crops contain no foreign genes and can thus basically be considered non-transgenic.

Modification of polyploid plants

[0289] Many plants are polyploid, which means they carry duplicate copies of their genomes—sometimes as many as six, as in wheat. The methods according to the present invention, which make use of the small molecule inhibitors can be “multiplexed” to affect all copies of a gene, or to target dozens of genes at once. For instance, in particular embodiments, the methods of the present invention are used to simultaneously ensure a loss of function mutation in different genes responsible for suppressing defenses against a disease.

In particular embodiments, the methods of the present invention are used to simultaneously suppress the expression of the TaMLO-A1, TaMLO-B1 and TaMLO-D1 nucleic acid sequence in a wheat plant cell and regenerating a wheat plant therefrom, in order to ensure that the wheat plant is resistant to powdery mildew (see also WO2015109752).

Creating male sterile plants

[0290] Hybrid plants typically have advantageous agronomic traits compared to inbred plants. However, for self-pollinating plants, the generation of hybrids can be challenging. In different plant types, genes have been identified which are important for plant fertility, more particularly male fertility. For instance, in maize, at least two genes have been identified which are important in fertility (Amitabh Mohanty International Conference on New Plant Breeding Molecular Technologies Technology Development And Regulation, Oct 9-10, 2014, Jaipur, India; Svitashv et al. Plant Physiol. 2015 Oct;169(2):931-45; Djukanovic et al. Plant J. 2013 Dec;76(5):888-99). The small molecule compounds herein can be used to target genes required for male fertility so as to generate male sterile plants which can easily be crossed to generate hybrids. In particular embodiments, the system provided herein is used for targeted mutagenesis of the cytochrome P450-like gene (MS26) or the meganuclease gene (MS45) thereby conferring male sterility to the maize plant. Maize plants which are as such genetically altered can be used in hybrid breeding programs.

Increasing the fertility stage in plants

[0291] In particular embodiments, the small molecule inhibitors herein are used to prolong the fertility stage of a plant such as of a rice plant. For instance, a rice fertility stage gene such as Ehd3 can be targeted in order to generate a mutation in the gene and plantlets can be selected for a prolonged regeneration plant fertility stage (as described in CN 104004782)

Regulating fruit-ripening

[0292] Ripening is a normal phase in the maturation process of fruits and vegetables. Only a few days after it starts it renders a fruit or vegetable inedible. This process brings significant losses to both farmers and consumers. In particular embodiments, the small molecule compounds herein can be used with systems to reduce ethylene production. This is ensured by ensuring one or more of the following: a. Suppression of ACC synthase gene expression. ACC (1-aminocyclopropane-1-carboxylic acid) synthase is the enzyme responsible for the conversion of S-adenosylmethionine (SAM) to ACC; the second to the

last step in ethylene biosynthesis. Enzyme expression is hindered when an antisense (“mirror-image”) or truncated copy of the synthase gene is inserted into the plant’s genome; b. Insertion of the ACC deaminase gene. The gene coding for the enzyme is obtained from *Pseudomonas chlororaphis*, a common nonpathogenic soil bacterium. It converts ACC to a different compound thereby reducing the amount of ACC available for ethylene production; c. Insertion of the SAM hydrolase gene. This approach is similar to ACC deaminase wherein ethylene production is hindered when the amount of its precursor metabolite is reduced; in this case SAM is converted to homoserine. The gene coding for the enzyme is obtained from *E. coli* T3 bacteriophage and d. Suppression of ACC oxidase gene expression. ACC oxidase is the enzyme which catalyzes the oxidation of ACC to ethylene, the last step in the ethylene biosynthetic pathway. Using the methods described herein, down regulation of the ACC oxidase gene results in the suppression of ethylene production, thereby delaying fruit ripening. In particular embodiments, additionally or alternatively to the modifications described above, the methods described herein are used to modify ethylene receptors, so as to interfere with ethylene signals obtained by the fruit. In particular embodiments, expression of the ETR1 gene, encoding an ethylene binding protein is modified, more particularly suppressed. In particular embodiments, additionally or alternatively to the modifications described above, the methods described herein are used to modify expression of the gene encoding Polygalacturonase (PG), which is the enzyme responsible for the breakdown of pectin, the substance that maintains the integrity of plant cell walls. Pectin breakdown occurs at the start of the ripening process resulting in the softening of the fruit. Accordingly, in particular embodiments, the methods described herein are used to introduce a mutation in the PG gene or to suppress activation of the PG gene in order to reduce the amount of PG enzyme produced thereby delaying pectin degradation.

[0293] Thus in particular embodiments, the small molecule compounds herein may be used to ensure one or more modifications of the genome of a plant cell such as described above, and regenerating a plant therefrom. In particular embodiments, the plant is a tomato plant.

Increasing storage life of plants

[0294] In particular embodiments, the small molecule inhibitors herein are used to modify genes involved in the production of compounds which affect storage life of the plant or plant part. More particularly, the modification is in a gene that prevents the accumulation

of reducing sugars in potato tubers. Upon high-temperature processing, these reducing sugars react with free amino acids, resulting in brown, bitter-tasting products and elevated levels of acrylamide, which is a potential carcinogen. In particular embodiments, the methods provided herein are used to reduce or inhibit expression of the vacuolar invertase gene (VInv), which encodes a protein that breaks down sucrose to glucose and fructose (Clasen et al. DOI: 10.1111/pbi.12370).

The use of the system to ensure a value added trait

[0295] In particular embodiments the small molecule inhibitors herein is used to produce nutritionally improved agricultural crops. In particular embodiments, the methods provided herein are adapted to generate “functional foods”, i.e. a modified food or food ingredient that may provide a health benefit beyond the traditional nutrients it contains and or “nutraceutical”, i.e. substances that may be considered a food or part of a food and provides health benefits, including the prevention and treatment of disease. In particular embodiments, the nutraceutical is useful in the prevention and/or treatment of one or more of cancer, diabetes, cardiovascular disease, and hypertension.

Biofuel production

[0296] The small molecule compounds herein may be used to modulate biofuel production in plants, fungi, and other organism, e.g., those have been engineered by expressing an RNA- guide nucleases. The term “biofuel” as used herein is an alternative fuel made from plant and plant-derived resources. Renewable biofuels can be extracted from organic matter whose energy has been obtained through a process of carbon fixation or are made through the use or conversion of biomass. This biomass can be used directly for biofuels or can be converted to convenient energy containing substances by thermal conversion, chemical conversion, and biochemical conversion. This biomass conversion can result in fuel in solid, liquid, or gas form. There are two types of biofuels: bioethanol and biodiesel. Bioethanol is mainly produced by the sugar fermentation process of cellulose (starch), which is mostly derived from maize and sugar cane. Biodiesel on the other hand is mainly produced from oil crops such as rapeseed, palm, and soybean. Biofuels are used mainly for transportation.

Enhancing plant properties for biofuel production

[0297] In particular embodiments, the small molecule inhibitors herein are used to alter the properties of the cell wall in order to facilitate access by key hydrolyzing agents for a

more efficient release of sugars for fermentation. In particular embodiments, the biosynthesis of cellulose and/or lignin are modified. Cellulose is the major component of the cell wall. The biosynthesis of cellulose and lignin are co-regulated. By reducing the proportion of lignin in a plant the proportion of cellulose can be increased. In particular embodiments, the methods described herein are used to downregulate lignin biosynthesis in the plant so as to increase fermentable carbohydrates. More particularly, the methods described herein are used to downregulate at least a first lignin biosynthesis gene selected from the group consisting of 4-coumarate 3-hydroxylase (C3H), phenylalanine ammonia-lyase (PAL), cinnamate 4-hydroxylase (C4H), hydroxycinnamoyl transferase (HCT), caffeic acid O-methyltransferase (COMT), caffeoyl CoA 3-O-methyltransferase (CCoAOMT), ferulate 5-hydroxylase (F5H), cinnamyl alcohol dehydrogenase (CAD), cinnamoyl CoA-reductase (CCR), 4-coumarate-CoA ligase (4CL), monolignol-lignin-specific glycosyltransferase, and aldehyde dehydrogenase (ALDH) as disclosed in WO 2008064289 A2.

[0298] In particular embodiments, the small molecule inhibitors herein are used to produce plant mass that produces lower levels of acetic acid during fermentation (see also WO 2010096488). More particularly, the methods disclosed herein are used to generate mutations in homologs to CasIL to reduce polysaccharide acetylation.

Modifying yeast for Biofuel production

[0299] In particular embodiments, the small molecule compounds herein may be used to modulate Cas enzymes used for bioethanol production by recombinant micro-organisms. For instance, Cas can be used to engineer micro-organisms, such as yeast, to generate biofuel or biopolymers from fermentable sugars and optionally to be able to degrade plant-derived lignocellulose derived from agricultural waste as a source of fermentable sugars. More particularly, the invention provides methods whereby the Cas CRISPR complex is used to introduce foreign genes required for biofuel production into micro-organisms and/or to modify endogenous genes which may interfere with the biofuel synthesis. More particularly the methods involve introducing into a micro-organism such as a yeast one or more nucleotide sequence encoding enzymes involved in the conversion of pyruvate to ethanol or another product of interest. In particular embodiments the methods ensure the introduction of one or more enzymes which allows the micro-organism to degrade cellulose, such as a cellulase. In yet further embodiments, the Cas proteins may be those used to modify endogenous metabolic pathways which compete with the biofuel production pathway.

Modifying Algae and plants for production of vegetable oils or biofuels

[0300] Transgenic algae or other plants such as rape may be particularly useful in the production of vegetable oils or biofuels such as alcohols (especially methanol and ethanol), for instance. These may be engineered to express or overexpress high levels of oil or alcohols for use in the oil or biofuel industries.

[0301] According to particular embodiments of the invention, the system is used to generate lipid-rich diatoms which are useful in biofuel production.

[0302] In particular embodiments it is envisaged to specifically modify genes that are involved in the modification of the quantity of lipids and/or the quality of the lipids produced by the algal cell. Examples of genes encoding enzymes involved in the pathways of fatty acid synthesis can encode proteins having for instance acetyl-CoA carboxylase, fatty acid synthase, 3-ketoacyl-acyl- carrier protein synthase III, glycerol-3-phosphate deshydrogenase (G3PDH), Enoyl-acyl carrier protein reductase (Enoyl-ACP-reductase), glycerol-3-phosphate acyltransferase, lysophosphatidic acyl transferase or diacylglycerol acyltransferase, phospholipid:diacylglycerol acyltransferase, phosphatidate phosphatase, fatty acid thioesterase such as palmitoyl protein thioesterase, or malic enzyme activities. In further embodiments it is envisaged to generate diatoms that have increased lipid accumulation. This can be achieved by targeting genes that decrease lipid catabolisation. Of particular interest for use in the methods of the present invention are genes involved in the activation of both triacylglycerol and free fatty acids, as well as genes directly involved in β -oxidation of fatty acids, such as acyl-CoA synthetase, 3-ketoacyl-CoA thiolase, acyl-CoA oxidase activity and phosphoglucosmutase. The system and methods described herein can be used to specifically activate such genes in diatoms as to increase their lipid content.

[0303] Organisms such as microalgae are widely used for synthetic biology. Stovicek et al. (Metab. Eng. Comm., 2015; 2:13 describes genome editing of industrial yeast, for example, *Saccharomyces cerevisiae*, to efficiently produce robust strains for industrial production. Stovicek used a CRISPR-Cas9 system codon-optimized for yeast to simultaneously disrupt both alleles of an endogenous gene and knock in a heterologous gene. Cas9 and gRNA were expressed from genomic or episomal 2 μ -based vector locations. The authors also showed that gene disruption efficiency could be improved by optimization of the levels of Cas9 and gRNA expression. Hlavová et al. (Biotechnol. Adv. 2015) discusses development of species or strains of microalgae using techniques such as CRISPR to target

nuclear and chloroplast genes for insertional mutagenesis and screening. The methods of Stovicek and Hlavová may be applied to the Cas effector protein system of the present invention.

[0304] US 8,945,839 describes a method for engineering Micro-Algae (*Chlamydomonas reinhardtii* cells) species) using Cas9. Using similar tools, the methods of the system described herein can be applied on *Chlamydomonas* species and other algae. In particular embodiments, Cas and guide RNA are introduced in algae expressed using a vector that expresses Cas under the control of a constitutive promoter such as Hsp70A-Rbc S2 or Beta2 - tubulin. Guide RNA will be delivered using a vector containing T7 promoter. Alternatively, Cas mRNA and in vitro transcribed guide RNA can be delivered to algal cells. Electroporation protocol follows standard recommended protocol from the GeneArt *Chlamydomonas* Engineering kit.

The use of systems in the generation of micro-organisms capable of fatty acid production

[0305] In particular embodiments, the methods of the invention are used for the generation of genetically engineered micro-organisms capable of the production of fatty esters, such as fatty acid methyl esters ("FAME") and fatty acid ethyl esters ("FAEE"),

[0306] Typically, host cells can be engineered to produce fatty esters from a carbon source, such as an alcohol, present in the medium, by expression or overexpression of a gene encoding a thioesterase, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase. Accordingly, the methods provided herein are used to modify a micro-organisms so as to overexpress or introduce a thioesterase gene, a gene encoding an acyl-CoA synthase, and a gene encoding an ester synthase. In particular embodiments, the thioesterase gene is selected from *tesA*, *'tesA*, *tesB*, *fatB*, *fatB2*, *fatB3*, *fatA1*, or *fatA*. In particular embodiments, the gene encoding an acyl-CoA synthase is selected from *fadDJadK*, *BH3103*, *pfl-4354*, *EAV15023*, *fadD1*, *fadD2*, *RPC_4074*, *fadDD35*, *fadDD22*, *faa39*, or an identified gene encoding an enzyme having the same properties. In particular embodiments, the gene encoding an ester synthase is a gene encoding a synthase/acyl-CoA:diacylglycerol acyltransferase from *Simmondsia chinensis*, *Acinetobacter* sp. ADP, *Alcanivorax borkumensis*, *Pseudomonas aeruginosa*, *Fundibacter jadensis*, *Arabidopsis thaliana*, or *Alkaligenes eutrophus*, or a variant thereof.

[0307] Additionally, or alternatively, the small molecule compounds herein are used to decrease expression in said micro-organism of at least one of a gene encoding an acyl-

CoA dehydrogenase, a gene encoding an outer membrane protein receptor, and a gene encoding a transcriptional regulator of fatty acid biosynthesis. In particular embodiments one or more of these genes is inactivated, such as by introduction of a mutation.

[0308] In particular embodiments, the gene encoding an acyl-CoA dehydrogenase is *fadE*. In particular embodiments, the gene encoding a transcriptional regulator of fatty acid biosynthesis encodes a DNA transcription repressor, for example, *fabR*.

[0309] Additionally, or alternatively, said micro-organism is modified to reduce expression of at least one of a gene encoding a pyruvate formate lyase, a gene encoding a lactate dehydrogenase, or both. In particular embodiments, the gene encoding a pyruvate formate lyase is *pflB*. In particular embodiments, the gene encoding a lactate dehydrogenase is *IdhA*. In particular embodiments one or more of these genes is inactivated, such as by introduction of a mutation therein.

[0310] In particular embodiments, the micro-organism is selected from the genus *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Synechococcus*, *Synechocystis*, *Pseudomonas*, *Aspergillus*, *Trichoderma*, *Neurospora*, *Fusarium*, *Humicola*, *Rhizomucor*, *Kluyveromyces*, *Pichia*, *Mucor*, *Myceliophthora*, *Penicillium*, *Phanerochaete*, *Pleurotus*, *Trametes*, *Chrysosporium*, *Saccharomyces*, *Stenotrophomonas*, *Schizosaccharomyces*, *Yarrowia*, or *Streptomyces*.

[0311] The small molecule compounds herein are further used to engineer micro-organisms capable of organic acid production, more particularly from pentose or hexose sugars. In particular embodiments, the methods comprise introducing into a micro-organism an exogenous LDH gene. In particular embodiments, the organic acid production in said micro-organisms is additionally or alternatively increased by inactivating endogenous genes encoding proteins involved in an endogenous metabolic pathway which produces a metabolite other than the organic acid of interest and/or wherein the endogenous metabolic pathway consumes the organic acid. In particular embodiments, the modification ensures that the production of the metabolite other than the organic acid of interest is reduced. According to particular embodiments, the methods are used to introduce at least one engineered gene deletion and/or inactivation of an endogenous pathway in which the organic acid is consumed or a gene encoding a product involved in an endogenous pathway which produces a metabolite other than the organic acid of interest. In particular embodiments, the at least one engineered gene deletion or inactivation is in one or more gene encoding an enzyme selected

from the group consisting of pyruvate decarboxylase (pdc), fumarate reductase, alcohol dehydrogenase (adh), acetaldehyde dehydrogenase, phosphoenolpyruvate carboxylase (ppc), D-lactate dehydrogenase (d-ldh), L-lactate dehydrogenase (l-ldh), lactate 2-monooxygenase.

[0312] In further embodiments the at least one engineered gene deletion and/or inactivation is in an endogenous gene encoding pyruvate decarboxylase (pdc).

[0313] In further embodiments, the micro-organism is engineered to produce lactic acid and the at least one engineered gene deletion and/or inactivation is in an endogenous gene encoding lactate dehydrogenase. Additionally or alternatively, the micro-organism comprises at least one engineered gene deletion or inactivation of an endogenous gene encoding a cytochrome-dependent lactate dehydrogenase, such as a cytochrome B2-dependent L-lactate dehydrogenase.

The use of the small molecule compounds in the generation of improved xylose or cellobiose utilizing yeasts strains

[0314] In particular embodiments, the small molecule compounds herein may be applied to select for improved xylose or cellobiose utilizing yeast strains. Error-prone PCR can be used to amplify one (or more) genes involved in the xylose utilization or cellobiose utilization pathways. Examples of genes involved in xylose utilization pathways and cellobiose utilization pathways may include, without limitation, those described in Ha, S.J., et al. (2011) Proc. Natl. Acad. Sci. USA 108(2):504-9 and Galazka, J.M., et al. (2010) Science 330(6000):84-6. Resulting libraries of double-stranded DNA molecules, each comprising a random mutation in such a selected gene could be co-transformed with the components of the system into a yeast strain (for instance S288C) and strains can be selected with enhanced xylose or cellobiose utilization capacity, as described in WO2015138855.

The use of the small molecule compounds in the generation of improved yeasts strains for use in isoprenoid biosynthesis

[0315] Tadas Jakočiūnas et al. described the successful application of a multiplex CRISPR/Cas9 system for genome engineering of up to 5 different genomic loci in one transformation step in baker's yeast *Saccharomyces cerevisiae* (Metabolic Engineering Volume 28, March 2015, Pages 213–222) resulting in strains with high mevalonate production, a key intermediate for the industrially important isoprenoid biosynthesis pathway. In particular embodiments, the system may be applied in a multiplex genome

engineering method as described herein for identifying additional high producing yeast strains for use in isoprenoid synthesis.

The use of the small molecule compounds in the generation of lactic acid producing yeasts strains

[0316] In another embodiment, successful application of a multiplex system is encompassed. In analogy with Vratislav Stovicek et al. (Metabolic Engineering Communications, Volume 2, December 2015, Pages 13–22), improved lactic acid-producing strains can be designed and obtained in a single transformation event. In a particular embodiment, the system is used for simultaneously inserting the heterologous lactate dehydrogenase gene and disruption of two endogenous genes PDC1 and PDC5 genes.

Further applications of the small molecule compounds in plants

[0317] In particular embodiments, the small molecule inhibitors herein, can be used for visualization of genetic element dynamics. For example, the small molecule inhibitors may regulate the CRISPR imaging that can visualize either repetitive or non-repetitive genomic sequences, report telomere length change and telomere movements and monitor the dynamics of gene loci throughout the cell cycle (Chen et al., Cell, 2013). These methods may also be applied to plants.

[0318] Other applications of the the small molecule inhibitors herein, is the targeted gene disruption positive-selection screening in vitro and in vivo (Malina et al., Genes and Development, 2013). These methods may also be applied to plants.

[0319] In particular embodiments, fusion of inactive Cas endonucleases with histone-modifying enzymes can introduce custom changes in the complex epigenome (Rusk et al., Nature Methods, 2014). The small molecule inhibitors may also be applied to plants.

[0320] In particular embodiments, the small molecule inhibitors herein, can be used to purify a specific portion of the chromatin and identify the associated proteins, thus elucidating their regulatory roles in transcription (Waldrip et al., Epigenetics, 2014). These methods may also be applied to plants.

[0321] In particular embodiments, the small molecule inhibitors can be used as a therapy for virus removal in plant systems as it is able to cleave both viral DNA and RNA. Previous studies in human systems have demonstrated the success of utilizing CRISPR in targeting the single strand RNA virus, hepatitis C (A. Price, et al., Proc. Natl. Acad. Sci, 2015) as well as the double stranded DNA virus, hepatitis B (V. Ramanan, et al., Sci. Rep, 2015).

[0322] In particular embodiments, the small molecule inhibitors may be used to alter genome complexity. In further particular embodiment, the small molecule inhibitors herein, can be used to disrupt or alter chromosome number and generate haploid plants, which only contain chromosomes from one parent. Such plants can be induced to undergo chromosome duplication and converted into diploid plants containing only homozygous alleles (Karimi-Ashtiyani et al., PNAS, 2015; Anton et al., Nucleus, 2014). These methods may also be applied to plants.

[0323] In particular embodiments, the small molecule compounds can be used for regulating self-cleavage. In these embodiments, the promotor of the Cas enzyme and gRNA can be a constitutive promotor and a second gRNA is introduced in the same transformation cassette, but controlled by an inducible promoter. This second gRNA can be designated to induce site-specific cleavage in the Cas gene in order to create a non-functional Cas. In a further particular embodiment, the second gRNA induces cleavage on both ends of the transformation cassette, resulting in the removal of the cassette from the host genome. This system offers a controlled duration of cellular exposure to the Cas enzyme and further minimizes off-target editing. Furthermore, cleavage of both ends of a CRISPR/Cas cassette can be used to generate transgene-free T0 plants with bi-allelic mutations (as described for Cas9 e.g. Moore et al., Nucleic Acids Research, 2014; Schaeffer et al., Plant Science, 2015). The methods of Moore et al. may be applied to the systems described herein.

Applications in non-human animals

[0324] The compounds described herein, alternatively referenced as small molecule compounds, may be used on non-human animals, those expressing an RNA-guided nuclease (e.g., constitutively or inducibly). The inhibitors may regulate the functions and activities of the RNA-guided nucleases in the animals.

[0325] In an aspect, the invention provides a non-human eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. In other aspects, the invention provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell according to any of the described embodiments. The organism in some embodiments of these aspects may be an animal; for example, a mammal. Also, the organism may be an arthropod such as an insect. The present invention may also be extended to other agricultural applications such as, for example, farm and production animals. For example, pigs have many features that

make them attractive as biomedical models, especially in regenerative medicine. In particular, pigs with severe combined immunodeficiency (SCID) may provide useful models for regenerative medicine, xenotransplantation (discussed also elsewhere herein), and tumor development and will aid in developing therapies for human SCID patients. Lee et al., (Proc Natl Acad Sci U S A. 2014 May 20;111(20):7260-5) utilized a reporter-guided transcription activator-like effector nuclease (TALEN) system to generated targeted modifications of recombination activating gene (RAG) 2 in somatic cells at high efficiency, including some that affected both alleles. The Type V effector protein may be applied to a similar system.

[0326] The small molecule compounds herein may be used (e.g., with an RNA-guided nuclease) to create a platform to model a disease or disorder of an animal, in some embodiments a mammal, in some embodiments a human. In certain embodiments, such models and platforms are rodent based, in non-limiting examples rat or mouse. Such models and platforms can take advantage of distinctions among and comparisons between inbred rodent strains. In certain embodiments, such models and platforms primate, horse, cattle, sheep, goat, swine, dog, cat or bird-based, for example to directly model diseases and disorders of such animals or to create modified and/or improved lines of such animals. Advantageously, in certain embodiments, an animal-based platform or model is created to mimic a human disease or disorder. For example, the similarities of swine to humans make swine an ideal platform for modeling human diseases. Compared to rodent models, development of swine models has been costly and time intensive. On the other hand, swine and other animals are much more similar to humans genetically, anatomically, physiologically and pathophysiologically. The small molecule compounds herein may be used to provide a high efficiency platform for targeted gene and genome editing, gene and genome modification and gene and genome regulation to be used in such animal platforms and models. Though ethical standards block development of human models and in many cases models based on non-human primates, the present invention is used with in vitro systems, including but not limited to cell culture systems, three dimensional models and systems, and organoids to mimic, model, and investigate genetics, anatomy, physiology and pathophysiology of structures, organs, and systems of humans. The platforms and models provide manipulation of single or multiple targets.

[0327] In certain embodiments, the present invention is applicable to disease models like that of Schomberg et al. (FASEB Journal, April 2016; 30(1):Suppl 571.1). For example,

small molecule compounds herein may be used to regulate the CRISPR systems used herein. To model the inherited disease neurofibromatosis type 1 (NF-1) Schomberg used CRISPR-Cas9 to introduce mutations in the swine neurofibromin 1 gene by cytosolic microinjection of CRISPR/Cas9 components into swine embryos. CRISPR guide RNAs (gRNA) were created for regions targeting sites both upstream and downstream of an exon within the gene for targeted cleavage by Cas9 and repair was mediated by a specific single-stranded oligodeoxynucleotide (ssODN) template to introduce a 2500 bp deletion. The small molecules herein may also be used to engineer swine with specific NF-1 mutations or clusters of mutations, and further can be used to engineer mutations that are specific to or representative of a given human individual. The invention is similarly used to develop animal models, including but not limited to swine models, of human multigenic diseases. According to the invention, multiple genetic loci in one gene or in multiple genes are simultaneously targeted using multiplexed guides and optionally one or multiple templates.

[0328] The present invention is also applicable to modifying SNPs of other animals, such as cows. Tan et al. (Proc Natl Acad Sci U S A. 2013 Oct 8; 110(41): 16526–16531) expanded the livestock gene editing toolbox to include transcription activator-like (TAL) effector nuclease (TALEN)- and clustered regularly interspaced short palindromic repeats (CRISPR)/Cas9- stimulated homology-directed repair (HDR) using plasmid, rAAV, and oligonucleotide templates. Gene specific gRNA sequences were cloned into the Church lab gRNA vector (Addgene ID: 41824) according to their methods (Mali P, et al. (2013) RNA-Guided Human Genome Engineering via Cas9. Science 339(6121):823-826). The Cas9 nuclease was provided either by co-transfection of the hCas9 plasmid (Addgene ID: 41815) or mRNA synthesized from RCIScript-hCas9. This RCIScript-hCas9 was constructed by sub-cloning the XbaI-AgeI fragment from the hCas9 plasmid (encompassing the hCas9 cDNA) into the RCIScript plasmid.

[0329] Heo et al. (Stem Cells Dev. 2015 Feb 1;24(3):393-402. doi: 10.1089/scd.2014.0278. Epub 2014 Nov 3) reported highly efficient gene targeting in the bovine genome using bovine pluripotent cells and clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 nuclease. First, Heo et al. generate induced pluripotent stem cells (iPSCs) from bovine somatic fibroblasts by the ectopic expression of yamanaka factors and GSK3 β and MEK inhibitor (2i) treatment. Heo et al. observed that these bovine iPSCs are highly similar to naïve pluripotent stem cells with regard to gene expression and

developmental potential in teratomas. Moreover, CRISPR-Cas9 nuclease, which was specific for the bovine NANOG locus, showed highly efficient editing of the bovine genome in bovine iPSCs and embryos.

Livestock - Pigs

[0330] Viral targets in livestock may include, in some embodiments, porcine CD163, for example on porcine macrophages. CD163 is associated with infection (thought to be through viral cell entry) by PRRSV (Porcine Reproductive and Respiratory Syndrome virus, an arterivirus). Infection by PRRSV, especially of porcine alveolar macrophages (found in the lung), results in a previously incurable porcine syndrome (“Mystery swine disease” or “blue ear disease”) that causes suffering, including reproductive failure, weight loss and high mortality rates in domestic pigs. Opportunistic infections, such as enzootic pneumonia, meningitis and ear oedema, are often seen due to immune deficiency through loss of macrophage activity. It also has significant economic and environmental repercussions due to increased antibiotic use and financial loss (an estimated \$660m per year).

[0331] As reported by Kristin M Whitworth and Dr Randall Prather et al. (Nature Biotech 3434 published online 07 December 2015) at the University of Missouri and in collaboration with Genus Plc, CD163 was targeted using CRISPR-Cas9 and the offspring of edited pigs were resistant when exposed to PRRSV. One founder male and one founder female, both of whom had mutations in exon 7 of CD163, were bred to produce offspring. The founder male possessed an 11-bp deletion in exon 7 on one allele, which results in a frameshift mutation and missense translation at amino acid 45 in domain 5 and a subsequent premature stop codon at amino acid 64. The other allele had a 2-bp addition in exon 7 and a 377-bp deletion in the preceding intron, which were predicted to result in the expression of the first 49 amino acids of domain 5, followed by a premature stop code at amino acid 85. The sow had a 7 bp addition in one allele that when translated was predicted to express the first 48 amino acids of domain 5, followed by a premature stop codon at amino acid 70. The sow’s other allele was unamplifiable. Selected offspring were predicted to be a null animal (CD163–/–), i.e. a CD163 knock out.

[0332] Accordingly, in some embodiments, porcine alveolar macrophages may be targeted by the CRISPR protein and regulated by small molecule compounds herein. In some embodiments, porcine CD163 may be targeted by the CRISPR protein. In some embodiments, porcine CD163 may be knocked out through induction of a DSB or through

insertions or deletions, for example targeting deletion or modification of exon 7, including one or more of those described above, or in other regions of the gene, for example deletion or modification of exon 5.

[0333] An edited pig and its progeny are also envisaged, for example a CD163 knock out pig. This may be for livestock, breeding or modelling purposes (i.e. a porcine model). Semen comprising the gene knock out is also provided.

[0334] CD163 is a member of the scavenger receptor cysteine-rich (SRCR) superfamily. Based on in vitro studies SRCR domain 5 of the protein is the domain responsible for unpackaging and release of the viral genome. As such, other members of the SRCR superfamily may also be targeted in order to assess resistance to other viruses. PRRSV is also a member of the mammalian arterivirus group, which also includes murine lactate dehydrogenase-elevating virus, simian hemorrhagic fever virus and equine arteritis virus. The arteriviruses share important pathogenesis properties, including macrophage tropism and the capacity to cause both severe disease and persistent infection. Accordingly, arteriviruses, and in particular murine lactate dehydrogenase-elevating virus, simian hemorrhagic fever virus and equine arteritis virus, may be targeted, for example through porcine CD163 or homologues thereof in other species, and murine, simian and equine models and knockout also provided.

[0335] Indeed, this approach may be extended to viruses or bacteria that cause other livestock diseases that may be transmitted to humans, such as Swine Influenza Virus (SIV) strains which include influenza C and the subtypes of influenza A known as H1N1, H1N2, H2N1, H3N1, H3N2, and H2N3, as well as pneumonia, meningitis and edema mentioned above.

Generation of Models of Genetic and Epigenetic Conditions

[0336] The small molecule compounds herein may be used, e.g., with an RNA-guided nuclease, to create a plant, an animal or cell that may be used to model and/or study genetic or epigenetic conditions of interest, such as a through a model of mutations of interest or a disease model. In some embodiments, the models may be generated using the RNA-guided nuclease, and the characters of the models may be further modulated and controlled using the small molecule compounds herein.

[0337] As used herein, “disease” refers to a disease, disorder, or indication in a subject. For example, a method of the invention may be used to create an animal or cell that

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

[0338] In some methods, the disease model can be used to study the effects of mutations on the animal or cell and development and/or progression of the disease using measures commonly used in the study of the disease. Alternatively, such a disease model is useful for studying the effect of a pharmaceutically active compound on the disease.

[0339] In some methods, the disease model can be used to assess the efficacy of a potential gene therapy strategy. That is, a disease-associated gene or polynucleotide can be modified such that the disease development and/or progression is inhibited or reduced. In particular, the method comprises modifying a disease-associated gene or polynucleotide such that an altered protein is produced and, as a result, the animal or cell has an altered response. Accordingly, in some methods, a genetically modified animal may be compared with an animal predisposed to development of the disease such that the effect of the gene therapy event may be assessed.

[0340] In another embodiment, this invention provides a method of developing a biologically active agent that modulates a cell signaling event associated with a disease gene. The method comprises contacting a test compound with a cell comprising one or more vectors that drive expression of one or more of components of the system; and detecting a change in a readout that is indicative of a reduction or an augmentation of a cell signaling event associated with, e.g., a mutation in a disease gene contained in the cell.

[0341] A cell model or animal model can be constructed in combination with the method of the invention for screening a cellular function change. Such a model may be used to study the effects of a genome sequence modified by the systems and methods herein on a cellular function of interest. For example, a cellular function model may be used to study the effect of a modified genome sequence on intracellular signaling or extracellular signaling. Alternatively, a cellular function model may be used to study the effects of a modified genome sequence on sensory perception. In some such models, one or more genome sequences associated with a signaling biochemical pathway in the model are modified.

[0342] Several disease models have been specifically investigated. These include *de novo* autism risk genes CHD8, KATNAL2, and SCN2A; and the syndromic autism (Angelman Syndrome) gene UBE3A. These genes and resulting autism models are of course preferred, but serve to show the broad applicability of the invention across genes and corresponding models. An altered expression of one or more genome sequences associated with a signaling biochemical pathway can be determined by assaying for a difference in the mRNA levels of the corresponding genes between the test model cell and a control cell, when they are contacted with a candidate agent. Alternatively, the differential expression of the sequences associated with a signaling biochemical pathway is determined by detecting a difference in the level of the encoded polypeptide or gene product.

[0343] To assay for an agent-induced alteration in the level of mRNA transcripts or corresponding polynucleotides, nucleic acid contained in a sample is first extracted according to standard methods in the art. For instance, mRNA can be isolated using various lytic enzymes or chemical solutions according to the procedures set forth in Sambrook et al. (1989), or extracted by nucleic-acid-binding resins following the accompanying instructions provided by the manufacturers. The mRNA contained in the extracted nucleic acid sample is then detected by amplification procedures or conventional hybridization assays (e.g. Northern blot analysis) according to methods widely known in the art or based on the methods exemplified herein.

[0344] 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.

[0345] The small molecule compounds herein may regulate the effects (e.g., binding, cleavage) of a CRISPR-complex on a target polynucleotide. The target polynucleotide of a CRISPR complex can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA).

[0346] Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A “disease-associated” gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non-disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the etiology of a disease. The transcribed or translated products may be known or unknown, and may be at a normal or abnormal level.

[0347] The target polynucleotide of the system herein can be any polynucleotide endogenous or exogenous to the eukaryotic cell. For example, the target polynucleotide can be a polynucleotide residing in the nucleus of the eukaryotic cell. The target polynucleotide can be a sequence coding a gene product (e.g., a protein) or a non-coding sequence (e.g., a regulatory polynucleotide or a junk DNA). Without wishing to be bound by theory, it is believed that the target sequence should be associated with a PAM (protospacer adjacent motif); that is, a short sequence recognized by the CRISPR complex. The precise sequence and length requirements for the PAM differ depending on the CRISPR enzyme used, but

PAMs are typically 2-5 base pair sequences adjacent the protospacer (that is, the target sequence) Examples of PAM sequences are given in the examples section below, and the skilled person will be able to identify further PAM sequences for use with a given CRISPR enzyme. Further, engineering of the PAM Interacting (PI) domain may allow programming of PAM specificity, improve target site recognition fidelity, and increase the versatility of the Cas, e.g. Cas9, genome engineering platform. Cas proteins, such as Cas9 proteins may be engineered to alter their PAM specificity, for example as described in Kleinstiver BP et al. *Engineered CRISPR-Cas9 nucleases with altered PAM specificities*. Nature. 2015 Jul 23;523(7561):481-5. doi: 10.1038/nature14592.

[0348] The target polynucleotide of the system may include a number of disease-associated genes and polynucleotides as well as signaling biochemical pathway-associated genes and polynucleotides as listed in US provisional patent applications 61/736,527 and 61/748,427 having Broad reference BI-2011/008/WSGR Docket No. 44063-701.101 and BI-2011/008/WSGR Docket No. 44063-701.102 respectively, both entitled SYSTEMS METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION filed on December 12, 2012 and January 2, 2013, respectively, and PCT Application PCT/US2013/074667, entitled DELIVERY, ENGINEERING AND OPTIMIZATION OF SYSTEMS, METHODS AND COMPOSITIONS FOR SEQUENCE MANIPULATION AND THERAPEUTIC APPLICATIONS, filed December 12, 2013, the contents of all of which are herein incorporated by reference in their entirety.

[0349] Examples of target polynucleotides include a sequence associated with a signaling biochemical pathway, e.g., a signaling biochemical pathway-associated gene or polynucleotide. Examples of target polynucleotides include a disease associated gene or polynucleotide. A “disease-associated” gene or polynucleotide refers to any gene or polynucleotide which is yielding transcription or translation products at an abnormal level or in an abnormal form in cells derived from a disease-affected tissues compared with tissues or cells of a non-disease control. It may be a gene that becomes expressed at an abnormally high level; it may be a gene that becomes expressed at an abnormally low level, where the altered expression correlates with the occurrence and/or progression of the disease. A disease-associated gene also refers to a gene possessing mutation(s) or genetic variation that is directly responsible or is in linkage disequilibrium with a gene(s) that is responsible for the

etiology of a disease. The transcribed or translated products may be known or unknown, and may be at a normal or abnormal level.

THERAPEUTIC APPLICATIONS

[0350] The small molecule compounds herein may be used for treatment in a variety of diseases and disorders. The inhibitors may be used to modulate the function and activity of an RNA-guided nuclease (e.g., a Cas protein) used for treating a disease. For example, the inhibitors may be used for regulating the strength, efficacy, timing, dosage of the therapeutic RNA-guided nuclease.

[0351] In some cases, a small molecule compound herein may be administered to a subject concurrently with an RNA-guided nuclease. Alternatively or additionally, a small molecule compound herein may be administered to a subject prior to the administration of an RNA-guided nuclease. Alternatively or additionally, a small molecule compound herein may be administered to a subject after the administration of an RNA-guided nuclease. In some examples, the small molecule compounds herein are used for modulating CRISPR gene editing (e.g., by modulating Cas protein of the CRISPR system).

[0352] The compounds herein may be administered as one or more doses as needed. In some examples, the small molecule compounds may be administered as a single dose. In certain examples, the compounds may be administered as multiple doses, e.g., 2, 3, 4, 5, 6, 7, 8, 9, 10 or more doses. The multi-dose regime may be used to achieve optimal efficacy and/or temporal control of the activity and function of the RNA-guided nuclease.

Exemplary Therapies

[0353] The small molecule compounds herein may be used for treatment in a variety of diseases and disorders. The inhibitors may be used to modulate the function and activity of an RNA-guided nuclease (e.g., a Cas protein) used for treating a disease.

[0354] In embodiments, the compounds can be used in method for therapy in which cells are edited *ex vivo*, *in vivo* or *in vitro* using CRISPR systems to modulate at least one gene. In embodiments, *in vitro* methods may include with subsequent administration of the edited cells to a patient in need thereof. In some embodiments, the CRISPR editing involves knocking in, knocking out or knocking down expression of at least one target gene in a cell. In particular embodiments, the small molecule compounds herein can modulate CRISPR editing inserts an exogenous, gene, minigene or sequence, which may comprise one or more exons and introns or natural or synthetic introns into the locus of a target gene, a hot-spot

locus, a safe harbor locus of the gene genomic locations where new genes or genetic elements can be introduced without disrupting the expression or regulation of adjacent genes, or correction by insertions or deletions one or more mutations in DNA sequences that encode regulatory elements of a target gene.

[0355] In embodiments, the treatment is for disease/disorder of an organ, including liver disease, eye disease, muscle disease, heart disease, blood disease, brain disease, kidney disease, or may comprise treatment for an autoimmune disease, central nervous system disease, cancer and other proliferative diseases, neurodegenerative disorders, inflammatory disease, metabolic disorder, musculoskeletal disorder and the like.

[0356] Particular diseases/disorders include chondroplasia, achromatopsia, acid maltase deficiency, adrenoleukodystrophy, aicardi syndrome, alpha-1 antitrypsin deficiency, alpha-thalassemia, androgen insensitivity syndrome, apert syndrome, arrhythmogenic right ventricular, dysplasia, ataxia telangiectasia, Barth syndrome, beta-thalassemia, blue rubber bleb nevus syndrome, Canavan disease, chronic granulomatous diseases (CGD), cri du chat syndrome, cystic fibrosis, Dercum's disease, ectodermal dysplasia, Fanconi anemia, fibrodysplasia ossificans progressive, fragile X syndrome, galactosemia, Gaucher's disease, generalized gangliosidoses (e.g., GM1), hemochromatosis, the hemoglobin C mutation in the 6th codon of beta-globin (HbC), hemophilia, Huntington's disease, Hurler Syndrome, hypophosphatasia, Klinefelter syndrome, Krabbes Disease, Langer-Giedion Syndrome, leukodystrophy, long QT syndrome, Marfan syndrome, Moebius syndrome, mucopolysaccharidosis (MPS), nail patella syndrome, nephrogenic diabetes insipidus, neurofibromatosis, Niemann-Pick disease, osteogenesis imperfecta, porphyria, Prader-Willi syndrome, progeria, Proteus syndrome, retinoblastoma, Rett syndrome, Rubinstein-Taybi syndrome, Sanfilippo syndrome, severe combined immunodeficiency (SCID), Shwachman syndrome, sickle cell disease (sickle cell anemia), Smith-Magenis syndrome, Stickler syndrome, Tay-Sachs disease, Thrombocytopenia Absent Radius (TAR) syndrome, Treacher Collins syndrome, trisomy, tuberous sclerosis, Turner's syndrome, urea cycle disorder, von Hippel-Landau disease, Waardenburg syndrome, Williams syndrome, Wilson's disease, and Wiskott-Aldrich syndrome.

[0357] In embodiments, the disease is associated with expression of a tumor antigen, e.g., a proliferative disease, a precancerous condition, a cancer, or a non-cancer related indication associated with expression of the tumor antigen, which may in some embodiments comprise

a target selected from B2M, CD247, CD3D, CD3E, CD3G, TRAC, TRBC1, TRBC2, HLA-A, HLA-B, HLA-C, DCK, CD52, FKBP1A, CIITA, NLRC5, RFXANK, RFX5, RFXAP, or NR3C1, HAVCR2, LAG3, PDCD1, PD-L2, CTLA4, CEACAM (CEACAM-1, CEACAM-3 and/or CEACAM-5), VISTA, BTLA, TIGIT, LAIR1, CD160, 2B4, CD80, CD86, B7-H3 (CD113), B7-H4 (VTCN1), HVEM (TNFRSF14 or CD107), KIR, A2aR, MHC class I, MHC class II, GAL9, adenosine, and TGF beta, or PTPN11 DCK, CD52, NR3C1, LILRB1, CD19; CD123; CD22; CD30; CD171; CS-1 (also referred to as CD2 subset 1, CRACC, SLAMF7, CD319, and 19A24); C-type lectin-like molecule-1 (CLL-1 or CLECL1); CD33; epidermal growth factor receptor variant III (EGFRvIII); ganglioside G2 (GD2); ganglioside GD3 (aNeu5Ac(2-8)aNeu5Ac(2-3)bDGalp(1-4)bDGlc(1-1)Cer); TNF receptor family member B cell maturation (BCMA); Tn antigen ((Tn Ag) or (GalNAc-Ser/Thr)); prostate-specific membrane antigen (PSMA); Receptor tyrosine kinase-like orphan receptor 1 (ROR1); Fms-Like Tyrosine Kinase 3 (FLT3); Tumor-associated glycoprotein 72 (TAG72); CD38; CD44v6; Carcinoembryonic antigen (CEA); Epithelial cell adhesion molecule (EPCAM); B7H3 (CD276); KIT (CD117); Interleukin-13 receptor subunit alpha-2 (IL-13Ra2 or CD213A2); Mesothelin; Interleukin 11 receptor alpha (IL-11Ra); prostate stem cell antigen (PSCA); Protease Serine 21 (Testisin or PRSS21); vascular endothelial growth factor receptor 2 (VEGFR2); Lewis(Y) antigen; CD24; Platelet-derived growth factor receptor beta (PDGFR-beta); Stage-specific embryonic antigen-4 (SSEA-4); CD20; Folate receptor alpha; Receptor tyrosine-protein kinase ERBB2 (Her2/neu); n kinase ERBB2 (Her2/neu); Mucin 1, cell surface associated (MUC1); epidermal growth factor receptor (EGFR); neural cell adhesion molecule (NCAM); Prostase; prostatic acid phosphatase (PAP); elongation factor 2 mutated (ELF2M); Ephrin B2; fibroblast activation protein alpha (FAP); insulin-like growth factor 1 receptor (IGF-I receptor), carbonic anhydrase IX (CAIX); Proteasome (Prosome, Macropain) Subunit, Beta Type, 9 (LMP2); glycoprotein 100 (gp100); oncogene fusion protein consisting of breakpoint cluster region (BCR) and Abelson murine leukemia viral oncogene homolog 1 (Abl) (bcr-abl); tyrosinase; ephrin type-A receptor 2 (EphA2); Fucosyl GM1; sialyl Lewis adhesion molecule (sLe); ganglioside GM3 (aNeu5Ac(2-3)bDGalp(1-4)bDGlc(1-1)Cer); transglutaminase 5 (TGS5); high molecular weight-melanoma-associated antigen (HMWMAA); o-acetyl-GD2 ganglioside (OAcGD2); Folate receptor beta; tumor endothelial marker 1 (TEM1/CD248); tumor endothelial marker 7-related (TEM7R); claudin 6 (CLDN6); thyroid stimulating hormone receptor (TSHR); G protein-coupled

receptor class C group 5, member D (GPC5D); chromosome X open reading frame 61 (CXORF61); CD97; CD179a; anaplastic lymphoma kinase (ALK); Polysialic acid; placenta-specific 1 (PLAC1); hexasaccharide portion of globoH glycosphingolipid (GloboH); mammary gland differentiation antigen (NY-BR-1); uroplakin 2 (UPK2); Hepatitis A virus cellular receptor 1 (HAVCR1); adrenoceptor beta 3 (ADRB3); pannexin 3 (PANX3); G protein-coupled receptor 20 (GPR20); lymphocyte antigen 6 complex, locus K 9 (LY6K); Olfactory receptor 51E2 (OR51E2); TCR Gamma Alternate Reading Frame Protein (TARP); Wilms tumor protein (WT1); Cancer/testis antigen 1 (NY-ESO-1); Cancer/testis antigen 2 (LAGE-1a); Melanoma-associated antigen 1 (MAGE-A1); ETS translocation-variant gene 6, located on chromosome 12p (ETV6-AML); sperm protein 17 (SPA17); X Antigen Family, Member 1A (XAGE1); angiopoietin-binding cell surface receptor 2 (Tie 2); melanoma cancer testis antigen-1 (MAD-CT-1); melanoma cancer testis antigen-2 (MAD-CT-2); Fos-related antigen 1; tumor protein p53 (p53); p53 mutant; prostatein; surviving; telomerase; prostate carcinoma tumor antigen-1 (PCTA-1 or Galectin 8), melanoma antigen recognized by T cells 1 (MelanA or MART1); Rat sarcoma (Ras) mutant; human Telomerase reverse transcriptase (hTERT); sarcoma translocation breakpoints; melanoma inhibitor of apoptosis (ML-IAP); ERG (transmembrane protease, serine 2 (TMPRSS2) ETS fusion gene); N-Acetyl glucosaminyl-transferase V (NA17); paired box protein Pax-3 (PAX3); Androgen receptor; Cyclin B1; v-myc avian myelocytomatosis viral oncogene neuroblastoma derived homolog (MYCN); Ras Homolog Family Member C (RhoC); Tyrosinase-related protein 2 (TRP-2); Cytochrome P450 1B1 (CYP1B1); CCCTC-Binding Factor (Zinc Finger Protein)-Like (BORIS or Brother of the Regulator of Imprinted Sites), Squamous Cell Carcinoma Antigen Recognized By T Cells 3 (SART3); Paired box protein Pax-5 (PAX5); proacrosin binding protein sp32 (OY-TES1); lymphocyte-specific protein tyrosine kinase (LCK); A kinase anchor protein 4 (AKAP-4); synovial sarcoma, X breakpoint 2 (SSX2); Receptor for Advanced Glycation Endproducts (RAGE-1); renal ubiquitous 1 (RU1); renal ubiquitous 2 (RU2); legumain; human papilloma virus E6 (HPV E6); human papilloma virus E7 (HPV E7); intestinal carboxyl esterase; heat shock protein 70-2 mutated (mut hsp70-2); CD79a; CD79b; CD72; Leukocyte-associated immunoglobulin-like receptor 1 (LAIR1); Fc fragment of IgA receptor (FCAR or CD89); Leukocyte immunoglobulin-like receptor subfamily A member 2 (LILRA2); CD300 molecule-like family member f (CD300LF); C-type lectin domain family 12 member A (CLEC12A); bone marrow stromal cell antigen 2 (BST2); EGF-like module-

containing mucin-like hormone receptor-like 2 (EMR2); lymphocyte antigen 75 (LY75); Glypican-3 (GPC3); Fc receptor-like 5 (FCRL5); and immunoglobulin lambda-like polypeptide 1 (IGLL1), CD19, BCMA, CD70, G6PC, Dystrophin, including modification of exon 51 by deletion or excision, DMPK, CFTR (cystic fibrosis transmembrane conductance regulator). In embodiments, the targets comprise CD70, or a Knock-in of CD33 and Knock-out of B2M. In embodiments, the targets comprise a knockout of TRAC and B2M, or TRAC B2M and PD1, with or without additional target genes. In certain embodiments, the disease is cystic fibrosis with targeting of the SCNN1A gene, e.g., the non-coding or coding regions, e.g., a promoter region, or a transcribed sequence, e.g., intronic or exonic sequence, targeted knock-in at CFTR sequence within intron 2, into which, e.g., can be introduced CFTR sequence that codes for CFTR exons 3-27; and sequence within CFTR intron 10, into which sequence that codes for CFTR exons 11-27 can be introduced.

[0358] In embodiments, the disease is Metachromatic Leukodystrophy, and the target is Arylsulfatase A, the disease is Wiskott-Aldrich Syndrome and the target is Wiskott-Aldrich Syndrome protein, the disease is Adreno leukodystrophy and the target is ATP-binding cassette DI, the disease is Human Immunodeficiency Virus and the target is receptor type 5-C-C chemokine or CXCR4 gene, the disease is Beta-thalassemia and the target is Hemoglobin beta subunit, the disease is X-linked Severe Combined ID receptor subunit gamma and the target is interleukin-2 receptor subunit gamma, the disease is Multisystemic Lysosomal Storage Disorder cystinosis and the target is cystinosis, the disease is Diamond-Blackfan anemia and the target is Ribosomal protein S19, the disease is Fanconi Anemia and the target is Fanconi anemia complementation groups (e.g. FNACA, FNACB, FANCC, FANCD1, FANCD2, FANCE, FANCF, RAD51C), the disease is Shwachman-Bodian-Diamond syndrome and the target is Shwachman syndrome gene, the disease is Gaucher's disease and the target is Glucocerebrosidase, the disease is Hemophilia A and the target is Anti-hemophilic factor OR Factor VIII, Christmas factor, Serine protease, Factor Hemophilia B IX, the disease is Adenosine deaminase deficiency (ADA-SCID) and the target is Adenosine deaminase, the disease is GM1 gangliosidosis and the target is beta-galactosidase, the disease is Glycogen storage disease type II, Pompe disease, the disease is acid maltase deficiency and the target is alpha-glucosidase, the disease is Niemann-Pick disease, SMPD1-associated (Types Sphingomyelin phosphodiesterase 1 OR A and B) and the target is sphingomyelinase, the disease is Krabbe disease, globoid cell

leukodystrophy and the target is Galactosylceramidase or galactosylceramide lipidosis and the target is galactercerebrosidase, Human leukocyte antigens DR-15, DQ-6, the disease is Multiple Sclerosis (MS) DRB1, the disease is Herpes Simplex Virus 1 or 2 and the target is knocking down of one, two or three of RS1, RL2 and/or LAT genes. In embodiments, the disease is an HPV associated cancer with treatment including edited cells comprising binding molecules, such as TCRs or antigen binding fragments thereof and antibodies and antigen-binding fragments thereof, such as those that recognize or bind human papilloma virus. The disease can be Hepatitis B with a target of one or more of PreC, C, X, PreS1, PreS2, S, P and/or SP gene(s).

[0359] In embodiments, the immune disease is severe combined immunodeficiency (SCID), Omenn syndrome, and in one aspect the target is Recombination Activating Gene 1 (RAG1) or an interleukin-7 receptor (IL7R). In particular embodiments, the disease is Transthyretin Amyloidosis (ATTR), Familial amyloid cardiomyopathy, and in one aspect, the target is the TTR gene, including one or more mutations in the TTR gene. In embodiments, the disease is Alpha-1 Antitrypsin Deficiency (AATD) or another disease in which Alpha-1 Antitrypsin is implicated, for example GvHD, Organ transplant rejection, diabetes, liver disease, COPD, Emphysema and Cystic Fibrosis, in particular embodiments, the target is SERPINA1.

[0360] In embodiments, the disease is primary hyperoxaluria, which, in certain embodiments, the target comprises one or more of Lactate dehydrogenase A (LDHA) and hydroxy Acid Oxidase 1 (HAO 1). In embodiments, the disease is primary hyperoxaluria type 1 (ph1) and other alanine-glyoxylate aminotransferase (agxt) gene related conditions or disorders, such as Adenocarcinoma, Chronic Alcoholic Intoxication, Alzheimer's Disease, Cooley's anemia, Aneurysm, Anxiety Disorders, Asthma, Malignant neoplasm of breast, Malignant neoplasm of skin, Renal Cell Carcinoma, Cardiovascular Diseases, Malignant tumor of cervix, Coronary Arteriosclerosis, Coronary heart disease, Diabetes, Diabetes Mellitus, Diabetes Mellitus Non- Insulin-Dependent, Diabetic Nephropathy, Eclampsia, Eczema, Subacute Bacterial Endocarditis, Glioblastoma, Glycogen storage disease type II, Sensorineural Hearing Loss (disorder), Hepatitis, Hepatitis A, Hepatitis B, Homocystinuria, Hereditary Sensory Autonomic Neuropathy Type 1, Hyperaldosteronism, Hypercholesterolemia, Hyperoxaluria, Primary Hyperoxaluria, Hypertensive disease, Inflammatory Bowel Diseases, Kidney Calculi, Kidney Diseases, Chronic Kidney Failure,

leiomyosarcoma, Metabolic Diseases, Inborn Errors of Metabolism, Mitral Valve Prolapse Syndrome, Myocardial Infarction, Neoplasm Metastasis, Nephrotic Syndrome, Obesity, Ovarian Diseases, Periodontitis, Polycystic Ovary Syndrome, Kidney Failure, Adult Respiratory Distress Syndrome, Retinal Diseases, Cerebrovascular accident, Turner Syndrome, Viral hepatitis, Tooth Loss, Premature Ovarian Failure, Essential Hypertension, Left Ventricular Hypertrophy, Migraine Disorders, Cutaneous Melanoma, Hypertensive heart disease, Chronic glomerulonephritis, Migraine with Aura, Secondary hypertension, Acute myocardial infarction, Atherosclerosis of aorta, Allergic asthma, pineoblastoma, Malignant neoplasm of lung, Primary hyperoxaluria type I, Primary hyperoxaluria type 2, Inflammatory Breast Carcinoma, Cervix carcinoma, Restenosis, Bleeding ulcer, Generalized glycogen storage disease of infants, Nephrolithiasis, Chronic rejection of renal transplant, Urolithiasis, pricking of skin, Metabolic Syndrome X, Maternal hypertension, Carotid Atherosclerosis, Carcinogenesis, Breast Carcinoma, Carcinoma of lung, Nephronophthisis, Microalbuminuria, Familial Retinoblastoma, Systolic Heart Failure Ischemic stroke, Left ventricular systolic dysfunction, Cauda Equina Paraganglioma, Hepatocarcinogenesis, Chronic Kidney Diseases, Glioblastoma Multiforme, Non-Neoplastic Disorder, Calcium Oxalate Nephrolithiasis, Ablepharon-Macrostomia Syndrome, Coronary Artery Disease, Liver carcinoma, Chronic kidney disease stage 5, Allergic rhinitis (disorder), Crigler Najjar syndrome type 2, and Ischemic Cerebrovascular Accident. In certain embodiments, treatment is targeted to the liver. In embodiments, the gene is AGXT, with a a cytogenetic location of 2q37.3 and the genomic coordinate are on Chromosome 2 on the forward strand at position 240,868,479-240,880,502.

[0361] Treatment can also target collagen type vii alpha 1 chain (col7a1) gene related conditions or disorders, such as Malignant neoplasm of skin, Squamous cell carcinoma, Colorectal Neoplasms, Crohn Disease, Epidermolysis Bullosa, Indirect Inguinal Hernia, Pruritus, Schizophrenia, Dermatologic disorders, Genetic Skin Diseases, Teratoma, Cockayne-Touraine Disease, Epidermolysis Bullosa Acquisita, Epidermolysis Bullosa Dystrophica, Junctional Epidermolysis Bullosa, Hallopeau- Siemens Disease, Bullous Skin Diseases, Agenesis of corpus callosum, Dystrophia unguium, Vesicular Stomatitis, Epidermolysis Bullosa With Congenital Localized Absence Of Skin And Deformity Of Nails, Juvenile Myoclonic Epilepsy, Squamous cell carcinoma of esophagus, Poikiloderma of Kindler, pretibial Epidermolysis bullosa, Dominant dystrophic epidermolysis bullosa

albopapular type (disorder), Localized recessive dystrophic epidermolysis bullosa, Generalized dystrophic epidermolysis bullosa, Squamous cell carcinoma of skin, Epidermolysis Bullosa Pruriginosa, Mammary Neoplasms, Epidermolysis Bullosa Simplex Superficialis, Isolated Toenail Dystrophy, Transient bullous dermolysis of the newborn, Autosomal Recessive Epidermolysis Bullosa Dystrophica Localisata Variant, and Autosomal Recessive Epidermolysis Bullosa Dystrophica Inversa.

[0362] In embodiments, the disease is acute myeloid leukemia (AML), targeting Wilms Tumor I (WTI) and HLA expressing cells. In embodiments, the therapy is T cell therapy, as described elsewhere herein, comprising engineered T cells with WTI specific TCRs. In certain embodiments, the target is CD157 in AML.

[0363] In embodiments, the disease is a blood disease. In certain embodiments, the disease is hemophilia, in one aspect the target is Factor XI. In other embodiments, the disease is a hemoglobinopathy, such as sickle cell disease, sickle cell trait, hemoglobin C disease, hemoglobin C trait, hemoglobin S/C disease, hemoglobin D disease, hemoglobin E disease, a thalassemia, a condition associated with hemoglobin with increased oxygen affinity, a condition associated with hemoglobin with decreased oxygen affinity, unstable hemoglobin disease, methemoglobinemia. Hemostasis and Factor X and XII deficiencies can also be treated. In embodiments, the target is BCL11A gene (e.g., a human BCL11a gene), a BCL11a enhancer (e.g., a human BCL11a enhancer), or a HFPH region (e.g., a human HPFH region), beta globulin, fetal hemoglobin, γ -globin genes (e.g., HBG1, HBG2, or HBG1 and HBG2), the erythroid specific enhancer of the BCL11A gene (BCL11Ae), or a combination thereof.

[0364] In embodiments, the target locus can be one or more of RAC, TRBC1, TRBC2, CD3E, CD3G, CD3D, B2M, CIITA, CD247, HLA-A, HLA-B, HLA-C, DCK, CD52, FKBP1A, NLRC5, RFXANK, RFX5, RFXAP, NR3C1, CD274, HAVCR2, LAG3, PDCD1, PD-L2, HCF2, PAI, TFPI, PLAT, PLAUI, PLG, RPOZ, F7, F8, F9, F2, F5, F7, F10, F11, F12, F13A1, F13B, STAT1, FOXP3, IL2RG, DCLRE1C, ICOS, MHC2TA, GALNS, HGSNAT, ARSB, RFXAP, CD20, CD81, TNFRSF13B, SEC23B, PKLR, IFNG, SPTB, SPTA, SLC4A1, EPO, EPB42, CSF2 CSF3, VFW, SERPINCA1, CTLA4, CEACAM (e.g., CEACAM-1, CEACAM-3 and/or CEACAM-5), VISTA, BTLA, TIGIT, LAIR1, CD160, 2B4, CD80, CD86, B7-H3 (CD113), B7-H4 (VTCN1), HVEM (TNFRSF14 or CD107), KIR, A2aR, MHC class I, MHC class II, GAL9, adenosine, and TGF beta, PTPN11, and

combinations thereof. In embodiments, the target sequence within the genomic nucleic acid sequence at Chr1 1:5,250,094-5,250,237, - strand, hg38; Chr1 1:5,255,022-5,255,164, - strand, hg38; nondeletional HFPH region; Chr1 1:5,249,833 to Chr1 1:5,250,237, - strand, hg38; Chr1 1:5,254,738 to Chr1 1:5,255,164, - strand, hg38; Chr1 1:5,249,833-5,249,927, - strand, hg38; Chr1 1:5,254,738-5,254,851, - strand, hg38; Chr1 1:5,250,139-5,250,237, - strand, hg38.

[0365] In embodiments, the disease is associated with high cholesterol, and regulation of cholesterol is provided, in some embodiments, regulation is effected by modification in the target PCSK9. Other diseases in which PCSK9 can be implicated, and thus would be a target for the systems and methods described herein include Abetaipoproteinemia, Adenoma, Arteriosclerosis, Atherosclerosis, Cardiovascular Diseases, Cholelithiasis, Coronary Arteriosclerosis, Coronary heart disease, Non-Insulin-Dependent Diabetes Mellitus, Hypercholesterolemia, Familial Hypercholesterolemia, Hyperinsulinism, Hyperlipidemia, Familial Combined Hyperlipidemia, Hypobetalipoproteinemias, Chronic Kidney Failure, Liver diseases, Liver neoplasms, melanoma, Myocardial Infarction, Narcolepsy, Neoplasm Metastasis, Nephroblastoma, Obesity, Peritonitis, Pseudoxanthoma Elasticum, Cerebrovascular accident, Vascular Diseases, Xanthomatosis, Peripheral Vascular Diseases, Myocardial Ischemia, Dyslipidemias, Impaired glucose tolerance, Xanthoma, Polygenic hypercholesterolemia, Secondary malignant neoplasm of liver, Dementia, Overweight, Hepatitis C, Chronic, Carotid Atherosclerosis, Hyperlipoproteinemia Type Ha, Intracranial Atherosclerosis, Ischemic stroke, Acute Coronary Syndrome, Aortic calcification, Cardiovascular morbidity, Hyperlipoproteinemia Type lib, Peripheral Arterial Diseases, Familial Hyperaldosteronism Type II, Familial hypobetalipoproteinemia, Autosomal Recessive Hypercholesterolemia, Autosomal Dominant Hypercholesterolemia 3, Coronary Artery Disease, Liver carcinoma, Ischemic Cerebrovascular Accident, and Arteriosclerotic cardiovascular disease NOS. In embodiments, the treatment can be targeted to the liver, the primary location of activity of PCSK9.

[0366] In embodiments, the disease or disorder is Hyper IGM syndrome or a disorder characterized by defective CD40 signaling. In certain embodiments, the insertion of CD40L exons are used to restore proper CD40 signaling and B cell class switch recombination. In particular embodiments, the target is CD40 ligand (CD40L)-edited at one or more of exons 2-5 of the CD40L gene, in cells, e.g., T cells or hematopoietic stem cells (HSCs).

[0367] In embodiments, the disease is merosin-deficient congenital muscular dystrophy (mdcmd) and other laminin, alpha 2 (lama2) gene related conditions or disorders. The therapy can be targeted to the muscle, for example, skeletal muscle, smooth muscle, and/or cardiac muscle. In certain embodiments, the target is Laminin, Alpha 2 (LAMA2) which may also be referred to as Laminin- 12 Subunit Alpha, Laminin-2 Subunit Alpha, Laminin-4 Subunit Alpha 3, Merosin Heavy Chain, Laminin M Chain, LAMM, Congenital Muscular Dystrophy and Merosin. LAMA2 has a cytogenetic location of 6q22.33 and the genomic coordinate are on Chromosome 6 on the forward strand at position 128,883, 141-129,516,563. In embodiments, the disease treated can be Merosin-Deficient Congenital Muscular Dystrophy (MDCMD), Amyotrophic Lateral Sclerosis, Bladder Neoplasm, Charcot-Marie-Tooth Disease, Colorectal Carcinoma, Contracture, Cyst, Duchenne Muscular Dystrophy, Fatigue, Hyperopia, Renovascular Hypertension, melanoma, Mental Retardation, Myopathy, Muscular Dystrophy, Myopia, Myositis, Neuromuscular Diseases, Peripheral Neuropathy, Refractive Errors, Schizophrenia, Severe mental retardation (I.Q. 20-34), Thyroid Neoplasm, Tobacco Use Disorder, Severe Combined Immunodeficiency, Synovial Cyst, Adenocarcinoma of lung (disorder), Tumor Progression, Strawberry nevus of skin, Muscle degeneration, Microdontia (disorder), Walker-Warburg congenital muscular dystrophy, Chronic Periodontitis, Leukoencephalopathies, Impaired cognition, Fukuyama Type Congenital Muscular Dystrophy, Scleroatonic muscular dystrophy, Eichsfeld type congenital muscular dystrophy, Neuropathy, Muscle eye brain disease, Limb-Muscular Dystrophies, Girdle, Congenital muscular dystrophy (disorder), Muscle fibrosis, cancer recurrence, Drug Resistant Epilepsy, Respiratory Failure, Myxoid cyst, Abnormal breathing, Muscular dystrophy congenital merosin negative, Colorectal Cancer, Congenital Muscular Dystrophy due to Partial LAMA2 Deficiency, and Autosomal Dominant Craniometaphyseal Dysplasia.

[0368] In certain embodiments, the target is an AAVS1 (PPPIR12C), an ALB gene, an Angptl3 gene, an ApoC3 gene, an ASGR2 gene, a CCR5 gene, a FIX (F9) gene, a G6PC gene, a Gys2 gene, an HGD gene, a Lp(a) gene, a Pcsk9 gene, a Serpinal gene, a TF gene, and a TTR gene). Assessment of efficiency of HDR/NHEJ mediated knock-in of cDNA into the first exon can utilize cDNA knock-in into “safe harbor” sites such as: single-stranded or double-stranded DNA having homologous arms to one of the following regions, for example: ApoC3 (chr11:116829908-116833071), Angptl3 (chr1:62,597,487-62,606,305), Serpinal

(chr14:94376747-94390692), Lp(a) (chr6:160531483-160664259), Pcsk9 (chr1:55,039,475-55,064,852), FIX (chrX:139,530,736-139,563,458), ALB (chr4:73,404,254-73,421,411), TTR (chr1 8:31,591,766-31,599,023), TF (chr3:133,661,997-133,779,005), G6PC (chr17:42,900,796-42,914,432), Gys2 (chr12:21,536,188-21,604,857), AAVS1 (PPP1R12C) (chr19:55,090,912-55,117,599), HGD (chr3:120,628,167-120,682,570), CCR5 (chr3:46,370,854-46,376,206), or ASGR2 (chr17:7,101,322-7,114,310).

[0369] In one aspect, the target is superoxide dismutase 1, soluble (SOD1), which can aid in treatment of a disease or disorder associated with the gene. In particular embodiments, the disease or disorder is associated with SOD1, and can be, for example, Adenocarcinoma, Albuminuria, Chronic Alcoholic Intoxication, Alzheimer's Disease, Amnesia, Amyloidosis, Amyotrophic Lateral Sclerosis, Anemia, Autoimmune hemolytic anemia, Sickle Cell Anemia, Anoxia, Anxiety Disorders, Aortic Diseases, Arteriosclerosis, Rheumatoid Arthritis, Asphyxia Neonatorum, Asthma, Atherosclerosis, Autistic Disorder, Autoimmune Diseases, Barrett Esophagus, Behcet Syndrome, Malignant neoplasm of urinary bladder, Brain Neoplasms, Malignant neoplasm of breast, Oral candidiasis, Malignant tumor of colon, Bronchogenic Carcinoma, Non-Small Cell Lung Carcinoma, Squamous cell carcinoma, Transitional Cell Carcinoma, Cardiovascular Diseases, Carotid Artery Thrombosis, Neoplastic Cell Transformation, Cerebral Infarction, Brain Ischemia, Transient Ischemic Attack, Charcot-Marie-Tooth Disease, Cholera, Colitis, Colorectal Carcinoma, Coronary Arteriosclerosis, Coronary heart disease, Infection by *Cryptococcus neoformans*, Deafness, Cessation of life, Deglutition Disorders, Presenile dementia, Depressive disorder, Contact Dermatitis, Diabetes, Diabetes Mellitus, Experimental Diabetes Mellitus, Insulin-Dependent Diabetes Mellitus, Non-Insulin-Dependent Diabetes Mellitus, Diabetic Angiopathies, Diabetic Nephropathy, Diabetic Retinopathy, Down Syndrome, Dwarfism, Edema, Japanese Encephalitis, Toxic Epidermal Necrolysis, Temporal Lobe Epilepsy, Exanthema, Muscular fasciculation, Alcoholic Fatty Liver, Fetal Growth Retardation, Fibromyalgia, Fibrosarcoma, Fragile X Syndrome, Giardiasis, Glioblastoma, Glioma, Headache, Partial Hearing Loss, Cardiac Arrest, Heart failure, Atrial Septal Defects, Helminthiasis, Hemochromatosis, Hemolysis (disorder), Chronic Hepatitis, HIV Infections, Huntington Disease, Hypercholesterolemia, Hyperglycemia, Hyperplasia, Hypertensive disease, Hyperthyroidism, Hypopituitarism, Hypoproteinemia, Hypotension, natural Hypothermia, Hypothyroidism, Immunologic Deficiency Syndromes, Immune System Diseases, Inflammation, Inflammatory

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

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

[0370] In particular embodiments, the disease is associated with the gene ATXN1, ATXN2, or ATXN3, which may be targeted for treatment. In some embodiments, the CAG repeat region located in exon 8 of ATXN1, exon 1 of ATXN2, or exon 10 of the ATXN3 is targeted. In embodiments, the disease is spinocerebellar ataxia 3 (sca3), sca1, or sca2 and other related disorders, such as Congenital Abnormality, Alzheimer's Disease, Amyotrophic Lateral Sclerosis, Ataxia, Ataxia Telangiectasia, Cerebellar Ataxia, Cerebellar Diseases, Chorea, Cleft Palate, Cystic Fibrosis, Mental Depression, Depressive disorder, Dystonia, Esophageal Neoplasms, Exotropia, Cardiac Arrest, Huntington Disease, Machado- Joseph Disease, Movement Disorders, Muscular Dystrophy, Myotonic Dystrophy, Narcolepsy, Nerve Degeneration, Neuroblastoma, Parkinson Disease, Peripheral Neuropathy, Restless Legs Syndrome, Retinal Degeneration, Retinitis Pigmentosa, Schizophrenia, Shy-Drager Syndrome, Sleep disturbances, Hereditary Spastic Paraplegia, Thromboembolism, Stiff-Person Syndrome, Spinocerebellar Ataxia, Esophageal carcinoma, Polyneuropathy, Effects of heat, Muscle twitch, Extrapyramidal sign, Ataxic, Neurologic Symptoms, Cerebral atrophy, Parkinsonian Disorders, Protein S Deficiency, Cerebellar degeneration, Familial Amyloid Neuropathy Portuguese Type, Spastic syndrome, Vertical Nystagmus, Nystagmus End-Position, Antithrombin III Deficiency, Atrophic, Complicated hereditary spastic paraplegia, Multiple System Atrophy, Pallidolusian degeneration, Dystonia Disorders, Pure Autonomic Failure, Thrombophilia, Protein C, Deficiency, Congenital Myotonic Dystrophy, Motor symptoms, Neuropathy, Neurodegenerative Disorders, Malignant neoplasm of esophagus, Visual disturbance, Activated Protein C Resistance, Terminal illness, Myokymia, Central neuroblastoma, Dyssomnias, Appendicular Ataxia, Narcolepsy-Cataplexy Syndrome, Machado- Joseph Disease Type I, Machado- Joseph Disease Type II, Machado- Joseph Disease Type III, Dentatorubral-Pallidolusian Atrophy, Gait Ataxia, Spinocerebellar Ataxia Type 1, Spinocerebellar Ataxia Type 2, Spinocerebellar Ataxia Type 6 (disorder), Spinocerebellar Ataxia Type 7, Muscular Spinobulbar Atrophy, Genomic Instability, Episodic ataxia type 2 (disorder), Bulbo-Spinal Atrophy X-Linked, Fragile X Tremor/ Ataxia Syndrome, Thrombophilia Due to Activated Protein C Resistance (Disorder), Amyotrophic Lateral Sclerosis 1, Neuronal Intranuclear Inclusion Disease, Hereditary Antithrombin Iii Deficiency, and Late-Onset Parkinson Disease.

[0371] In embodiments, the disease is associated with expression of a tumor antigen-cancer or non-cancer related indication, for example acute lymphoid leukemia, diffuse large

B cell lymphoma, follicular lymphoma, chronic lymphocytic leukemia, Hodgkin lymphoma, non-Hodgkin lymphoma. In embodiments, the target can be TET2 intron, a TET2 intron-exon junction, a sequence within a genomic region of chr4.

[0372] In embodiments, neurodegenerative diseases can be treated. In particular embodiments, the target is Synuclein, Alpha (SNCA). In certain embodiments, the disorder treated is a pain related disorder, including congenital pain insensitivity, Compressive Neuropathies, Paroxysmal Extreme Pain Disorder, High grade atrioventricular block, Small Fiber Neuropathy, and Familial Episodic Pain Syndrome 2. In certain embodiments, the target is Sodium Channel, Voltage Gated, Type X Alpha Subunit (SCN10A).

[0373] In certain embodiments, hematopoietic stem cells and progenitor stem cells are edited, including knock-ins. In particular embodiments, the knock-in is for treatment of lysosomal storage diseases, glycogen storage diseases, mucopolysaccharidoses, or any disease in which the secretion of a protein will ameliorate the disease. In one embodiment, the disease is sickle cell disease (SCD). In another embodiment, the disease is β -thalassemia.

[0374] In certain embodiments, the T cell or NK cell is used for cancer treatment and may include T cells comprising the recombinant receptor (e.g. CAR) and one or more phenotypic markers selected from CCR7+, 4-1BB+ (CD137+), TIM3+, CD27+, CD62L+, CD127+, CD45RA+, CD45RO-, t-bet^{hi}, IL-7R α +, CD95+, IL-2R β +, CXCR3+ or LFA-1+. In certain embodiments the editing of a T cell for cancer immunotherapy comprises altering one or more T-cell expressed gene, e.g., one or more of FAS, BID, CTLA4, PDCD1, CBLB, PTPN6, B2M, TRAC and TRBC gene. In some embodiments, editing includes alterations introduced into, or proximate to, the CBLB target sites to reduce CBLB gene expression in T cells for treatment of proliferative diseases and may include larger insertions or deletions at one or more CBLB target sites. T cell editing of TGFBR2 target sequence can be, for example, located in exon 3, 4, or 5 of the TGFBR2 gene and utilized for cancers and lymphoma treatment.

[0375] Cells for transplantation can be edited and may include allele-specific modification of one or more immunogenicity genes (e.g., an HLA gene) of a cell, e.g., HLA-A, HLA-B, HLA-C, HLA-DRB1, HLA-DRB3/4/5, HLA-DQ, and HLA-DP MHCs, and any other MHC Class I or Class II genes or loci, which may include delivery of one or more matched recipient HLA alleles into the original position(s) where the one or more mismatched donor HLA alleles are located, and may include inserting one or more matched

recipient HLA alleles into a “safe harbor” locus. In an embodiment, the method further includes introducing a chemotherapy resistance gene for in vivo selection in a gene.

[0376] Methods and systems can target Dystrophia Myotonica-Protein Kinase (DMPK) for editing, in particular embodiments, the target is the CTG trinucleotide repeat in the 3' untranslated region (UTR) of the DMPK gene. Disorders or diseases associated with DMPK include Atherosclerosis, Azoospermia, Hypertrophic Cardiomyopathy, Celiac Disease, Congenital chromosomal disease, Diabetes Mellitus, Focal glomerulosclerosis, Huntington Disease, Hypogonadism, Muscular Atrophy, Myopathy, Muscular Dystrophy, Myotonia, Myotonic Dystrophy, Neuromuscular Diseases, Optic Atrophy, Paresis, Schizophrenia, Cataract, Spinocerebellar Ataxia, Muscle Weakness, Adrenoleukodystrophy, Centronuclear myopathy, Interstitial fibrosis, myotonic muscular dystrophy, Abnormal mental state, X-linked Charcot- Marie-Tooth disease 1, Congenital Myotonic Dystrophy, Bilateral cataracts (disorder), Congenital Fiber Type Disproportion, Myotonic Disorders, Multisystem disorder, 3- Methylglutaconic aciduria type 3, cardiac event, Cardiogenic Syncope, Congenital Structural Myopathy, Mental handicap, Adrenomyeloneuropathy, Dystrophia myotonica 2, and Intellectual Disability.

[0377] In embodiments, the disease is an inborn error of metabolism. The disease may be selected from Disorders of Carbohydrate Metabolism (glycogen storage disease, G6PD deficiency), Disorders of Amino Acid Metabolism (phenylketonuria, maple syrup urine disease, glutaric acidemia type 1), Urea Cycle Disorder or Urea Cycle Defects (carbamoyl phosphate synthase I deficiency), Disorders of Organic Acid Metabolism (alkaptonuria, 2-hydroxyglutaric acidurias), Disorders of Fatty Acid Oxidation/Mitochondrial Metabolism (Medium-chain acyl-coenzyme A dehydrogenase deficiency), Disorders of Porphyrin metabolism (acute intermittent porphyria), Disorders of Purine/Pyrimidine Metabolism (Lesch-Nyhan syndrome), Disorders of Steroid Metabolism (lipoid congenital adrenal hyperplasia, congenital adrenal hyperplasia), Disorders of Mitochondrial Function (Kearns-Sayre syndrome), Disorders of Peroxisomal function (Zellweger syndrome), or Lysosomal Storage Disorders (Gaucher's disease, Niemann-Pick disease).

[0378] In embodiments, the target can comprise Recombination Activating Gene 1 (RAG1), BCL11 A, PCSK9, laminin, alpha 2 (lama2), ATXN3, alanine-glyoxylate aminotransferase (AGXT), collagen type vii alpha 1 chain (COL7a1), spinocerebellar ataxia type 1 protein (ATXN1), Angiopoietin-like 3 (ANGPTL3), Frataxin (FXN), Superoxidase

Dismutase 1, soluble (SOD1), Synuclein, Alpha (SNCA), Sodium Channel, Voltage Gated, Type X Alpha Subunit (SCN10A), Spinocerebellar Ataxia Type 2 Protein (ATXN2), Dystrophin Myotonic-Protein Kinase (DMPK), beta globin locus on chromosome 11, acyl-coenzyme A dehydrogenase for medium chain fatty acids (ACADM), long-chain 3-hydroxyl-coenzyme A dehydrogenase for long chain fatty acids (HADHA), acyl-coenzyme A dehydrogenase for very long-chain fatty acids (ACADVL), Apolipoprotein C3 (APOCIII), Transthyretin (TTR), Angiopoietin-like 4 (ANGPTL4), Sodium Voltage-Gated Channel Alpha Subunit 9 (SCN9A), Interleukin-7 receptor (IL7R), glucose-6-phosphatase, catalytic (G6PC), haemochromatosis (HFE), SERPINA1, C9ORF72, β -globin, dystrophin, γ -globin.

[0379] In certain embodiments, the disease or disorder is associated with Apolipoprotein C3 (APOCIII), which can be targeted for editing. In embodiments, the disease or disorder may be Dyslipidemias, Hyperalphalipoproteinemia Type 2, Lupus Nephritis, Wilms Tumor 5, Morbid obesity and spermatogenic, Glaucoma, Diabetic Retinopathy, Arthrogryposis renal dysfunction cholestasis syndrome, Cognition Disorders, Altered response to myocardial infarction, Glucose Intolerance, Positive regulation of triglyceride biosynthetic process, Renal Insufficiency, Chronic, Hyperlipidemias, Chronic Kidney Failure, Apolipoprotein C-III Deficiency, Coronary Disease, Neonatal Diabetes Mellitus, Neonatal, with Congenital Hypothyroidism, Hypercholesterolemia Autosomal Dominant 3, Hyperlipoproteinemia Type III, Hyperthyroidism, Coronary Artery Disease, Renal Artery Obstruction, Metabolic Syndrome X, Hyperlipidemia, Familial Combined, Insulin Resistance, Transient infantile hypertriglyceridemia, Diabetic Nephropathies, Diabetes Mellitus (Type 1), Nephrotic Syndrome Type 5 with or without ocular abnormalities, and Hemorrhagic Fever with renal syndrome.

[0380] In certain embodiments, the target is Angiopoietin-like 4 (ANGPTL4). Diseases or disorders associated with ANGPTL4 that can be treated include ANGPTL4 is associated with dyslipidemias, low plasma triglyceride levels, regulator of angiogenesis and modulate tumorigenesis, and severe diabetic retinopathy. both proliferative diabetic retinopathy and non-proliferative diabetic retinopathy.

[0381] In embodiments, editing can be used for the treatment of fatty acid disorders. In certain embodiments, the target is one or more of ACADM, HADHA, ACADVL. In embodiments, the targeted edit is the activity of a gene in a cell selected from the acyl-coenzyme A dehydrogenase for medium chain fatty acids (ACADM) gene, the long-chain 3-

hydroxyl-coenzyme A dehydrogenase for long chain fatty acids (HADHA) gene, and the acyl-coenzyme A dehydrogenase for very long-chain fatty acids (ACADVL) gene. In one aspect, the disease is medium chain acyl-coenzyme A dehydrogenase deficiency (MCADD), long-chain 3-hydroxyl-coenzyme A dehydrogenase deficiency (LCHADD), and/or very long-chain acyl-coenzyme A dehydrogenase deficiency (VLCADD).

Adoptive Cell Therapies

[0382] The small molecule compounds herein may modulate RNA-guided nuclease that modifies cells for adoptive therapies. Aspects of the invention accordingly involve the adoptive transfer of immune system cells, such as T cells, specific for selected antigens, such as tumor associated antigens (see Maus et al., 2014, Adoptive Immunotherapy for Cancer or Viruses, Annual Review of Immunology, Vol. 32: 189-225; Rosenberg and Restifo, 2015, Adoptive cell transfer as personalized immunotherapy for human cancer, Science Vol. 348 no. 6230 pp. 62-68; and, Restifo et al., 2015, Adoptive immunotherapy for cancer: harnessing the T cell response. Nat. Rev. Immunol. 12(4): 269-281; and Jenson and Riddell, 2014, Design and implementation of adoptive therapy with chimeric antigen receptor-modified T cells. Immunol Rev. 257(1): 127-144). Various strategies may for example be employed to genetically modify T cells by altering the specificity of the T cell receptor (TCR) for example by introducing new TCR α and β chains with selected peptide specificity (see U.S. Patent No. 8,697,854; PCT Patent Publications: WO2003020763, WO2004033685, WO2004044004, WO2005114215, WO2006000830, WO2008038002, WO2008039818, WO2004074322, WO2005113595, WO2006125962, WO2013166321, WO2013039889, WO2014018863, WO2014083173; U.S. Patent No. 8,088,379).

[0383] As an alternative to, or addition to, TCR modifications, chimeric antigen receptors (CARs) may be used in order to generate immunoresponsive cells, such as T cells, specific for selected targets, such as malignant cells, with a wide variety of receptor chimera constructs having been described (see U.S. Patent Nos. 5,843,728; 5,851,828; 5,912,170; 6,004,811; 6,284,240; 6,392,013; 6,410,014; 6,753,162; 8,211,422; and, PCT Publication WO9215322). Alternative CAR constructs may be characterized as belonging to successive generations. First-generation CARs typically consist of a single-chain variable fragment of an antibody specific for an antigen, for example comprising a VL linked to a VH of a specific antibody, linked by a flexible linker, for example by a CD8 α hinge domain and a CD8 α transmembrane domain, to the transmembrane and intracellular signaling domains of either

CD3 ζ or FcR γ (scFv-CD3 ζ or scFv-FcR γ ; see U.S. Patent No. 7,741,465; U.S. Patent No. 5,912,172; U.S. Patent No. 5,906,936). Second-generation CARs incorporate the intracellular domains of one or more costimulatory molecules, such as CD28, OX40 (CD134), or 4-1BB (CD137) within the endodomain (for example scFv-CD28/OX40/4-1BB-CD3 ζ ; see U.S. Patent Nos. 8,911,993; 8,916,381; 8,975,071; 9,101,584; 9,102,760; 9,102,761). Third-generation CARs include a combination of costimulatory endodomains, such a CD3 ζ -chain, CD97, GDI la-CD18, CD2, ICOS, CD27, CD154, CDS, OX40, 4-1BB, or CD28 signaling domains (for example scFv-CD28-4-1BB-CD3 ζ or scFv-CD28-OX40-CD3 ζ ; see U.S. Patent No. 8,906,682; U.S. Patent No. 8,399,645; U.S. Pat. No. 5,686,281; PCT Publication No. WO2014134165; PCT Publication No. WO2012079000). Alternatively, costimulation may be orchestrated by expressing CARs in antigen-specific T cells, chosen so as to be activated and expanded following engagement of their native $\alpha\beta$ TCR, for example by antigen on professional antigen-presenting cells, with attendant costimulation. In addition, additional engineered receptors may be provided on the immunoresponsive cells, for example to improve targeting of a T-cell attack and/or minimize side effects.

[0384] Alternative techniques may be used to transform target immunoresponsive cells, such as protoplast fusion, lipofection, transfection or electroporation. A wide variety of vectors may be used, such as retroviral vectors, lentiviral vectors, adenoviral vectors, adeno-associated viral vectors, plasmids or transposons, such as a Sleeping Beauty transposon (see U.S. Patent Nos. 6,489,458; 7,148,203; 7,160,682; 7,985,739; 8,227,432), may be used to introduce CARs, for example using 2nd generation antigen-specific CARs signaling through CD3 ζ and either CD28 or CD137. Viral vectors may for example include vectors based on HIV, SV40, EBV, HSV or BPV.

[0385] Cells that are targeted for transformation, and can be treated by the small molecule compounds herein, may for example include T cells, Natural Killer (NK) cells, cytotoxic T lymphocytes (CTL), regulatory T cells, human embryonic stem cells, tumor-infiltrating lymphocytes (TIL) or a pluripotent stem cell from which lymphoid cells may be differentiated. T cells expressing a desired CAR may for example be selected through co-culture with γ -irradiated activating and propagating cells (AaPC), which co-express the cancer antigen and co-stimulatory molecules. The engineered CAR T-cells may be expanded, for example by co-culture on AaPC in presence of soluble factors, such as IL-2 and IL-21. This expansion may for example be carried out so as to provide memory CAR⁺ T cells

(which may for example be assayed by non-enzymatic digital array and/or multi-panel flow cytometry). In this way, CAR T cells may be provided that have specific cytotoxic activity against antigen-bearing tumors (optionally in conjunction with production of desired chemokines such as interferon- γ). CAR T cells of this kind may for example be used in animal models, for example to treat tumor xenografts.

[0386] Approaches such as the foregoing may be adapted to provide methods of treating and/or increasing survival of a subject having a disease, such as a neoplasia, for example by administering an effective amount of an immunoresponsive cell comprising an antigen recognizing receptor that binds a selected antigen, wherein the binding activates the immunoresponsive cell, thereby treating or preventing the disease (such as a neoplasia, a pathogen infection, an autoimmune disorder, or an allogeneic transplant reaction). Dosing in CAR T cell therapies may for example involve administration of from 10^6 to 10^9 cells/kg, with or without a course of lymphodepletion, for example with cyclophosphamide.

[0387] In one embodiment, the treatment can be administered into patients undergoing an immunosuppressive treatment. The cells or population of cells, may be made resistant to at least one immunosuppressive agent due to the inactivation of a gene encoding a receptor for such immunosuppressive agent. Not being bound by a theory, the immunosuppressive treatment should help the selection and expansion of the immunoresponsive or T cells according to the invention within the patient.

[0388] The administration of the cells or population of cells according to the present invention may be carried out in any convenient manner, including by aerosol inhalation, injection, ingestion, transfusion, implantation or transplantation. The cells or population of cells may be administered to a patient subcutaneously, intradermally, intratumorally, intranodally, intramedullary, intramuscularly, by intravenous or intralymphatic injection, or intraperitoneally. In one embodiment, the cell compositions of the present invention are preferably administered by intravenous injection.

[0389] The administration of the cells or population of cells can consist of the administration of 10^4 - 10^9 cells per kg body weight, preferably 10^5 to 10^6 cells/kg body weight including all integer values of cell numbers within those ranges. Dosing in CAR T cell therapies may for example involve administration of from 10^6 to 10^9 cells/kg, with or without a course of lymphodepletion, for example with cyclophosphamide. The cells or population of cells can be administered in one or more doses. In another embodiment, the

effective amount of cells are administrated as a single dose. In another embodiment, the effective amount of cells are administrated as more than one dose over a period time. Timing of administration is within the judgment of managing physician and depends on the clinical condition of the patient. The cells or population of cells may be obtained from any source, such as a blood bank or a donor. While individual needs vary, determination of optimal ranges of effective amounts of a given cell type for a particular disease or conditions are within the skill of one in the art. An effective amount means an amount which provides a therapeutic or prophylactic benefit. The dosage administrated will be dependent upon the age, health and weight of the recipient, kind of concurrent treatment, if any, frequency of treatment and the nature of the effect desired.

[0390] In another embodiment, the effective number of cells or composition comprising those cells are administrated parenterally. The administration can be an intravenous administration. The administration can be directly done by injection within a tumor.

[0391] To guard against possible adverse reactions, engineered immunoresponsive cells may be equipped with a transgenic safety switch, in the form of a transgene that renders the cells vulnerable to exposure to a specific signal. For example, the herpes simplex viral thymidine kinase (TK) gene may be used in this way, for example by introduction into allogeneic T lymphocytes used as donor lymphocyte infusions following stem cell transplantation (Greco, et al., Improving the safety of cell therapy with the TK-suicide gene. *Front. Pharmacol.* 2015; 6: 95). In such cells, administration of a nucleoside prodrug such as ganciclovir or acyclovir causes cell death. Alternative safety switch constructs include inducible caspase 9, for example triggered by administration of a small-molecule dimerizer that brings together two nonfunctional icasp9 molecules to form the active enzyme. A wide variety of alternative approaches to implementing cellular proliferation controls have been described (see U.S. Patent Publication No. 20130071414; PCT Patent Publication WO2011146862; PCT Patent Publication WO2014011987; PCT Patent Publication WO2013040371; Zhou et al. *BLOOD*, 2014, 123/25:3895 – 3905; Di Stasi et al., *The New England Journal of Medicine* 2011; 365:1673-1683; Sadelain M, *The New England Journal of Medicine* 2011; 365:1735-173; Ramos et al., *Stem Cells* 28(6):1107-15 (2010)).

[0392] In a further refinement of adoptive therapies, genome editing with a system as described herein may be used to tailor immunoresponsive cells to alternative implementations, for example providing edited CAR T cells (see Poirot et al., 2015,

Multiplex genome edited T-cell manufacturing platform for "off-the-shelf" adoptive T-cell immunotherapies, Cancer Res 75 (18): 3853). For example, immunoresponsive cells may be edited to delete expression of some or all of the class of HLA type II and/or type I molecules, or to knockout selected genes that may inhibit the desired immune response, such as the PD1 gene.

[0393] Cells may be edited using any system and method of use thereof as described herein. systems may be delivered to an immune cell by any method described herein. In preferred embodiments, cells are edited ex vivo and transferred to a subject in need thereof. Immunoresponsive cells, CAR T cells or any cells used for adoptive cell transfer may be edited. Editing may be performed to eliminate potential alloreactive T-cell receptors (TCR), disrupt the target of a chemotherapeutic agent, block an immune checkpoint, activate a T cell, and/or increase the differentiation and/or proliferation of functionally exhausted or dysfunctional CD8⁺ T-cells (see PCT Patent Publications: WO2013176915, WO2014059173, WO2014172606, WO2014184744, and WO2014191128). Editing may result in inactivation of a gene.

[0394] By inactivating a gene it is intended that the gene of interest is not expressed in a functional protein form. In a particular embodiment, the system specifically catalyzes cleavage in one targeted gene thereby inactivating said targeted gene. The nucleic acid strand breaks caused are commonly repaired through the distinct mechanisms of homologous recombination or non-homologous end joining (NHEJ). However, NHEJ is an imperfect repair process that often results in changes to the DNA sequence at the site of the cleavage. Repair via non-homologous end joining (NHEJ) often results in small insertions or deletions (Indel) and can be used for the creation of specific gene knockouts. Cells in which a cleavage induced mutagenesis event has occurred can be identified and/or selected by well-known methods in the art.

[0395] T cell receptors (TCR) are cell surface receptors that participate in the activation of T cells in response to the presentation of antigen. The TCR is generally made from two chains, α and β , which assemble to form a heterodimer and associates with the CD3-transducing subunits to form the T cell receptor complex present on the cell surface. Each α and β chain of the TCR consists of an immunoglobulin-like N-terminal variable (V) and constant (C) region, a hydrophobic transmembrane domain, and a short cytoplasmic region. As for immunoglobulin molecules, the variable region of the α and β chains are generated by

V(D)J recombination, creating a large diversity of antigen specificities within the population of T cells. However, in contrast to immunoglobulins that recognize intact antigen, T cells are activated by processed peptide fragments in association with an MHC molecule, introducing an extra dimension to antigen recognition by T cells, known as MHC restriction. Recognition of MHC disparities between the donor and recipient through the T cell receptor leads to T cell proliferation and the potential development of graft versus host disease (GVHD). The inactivation of TCR α or TCR β can result in the elimination of the TCR from the surface of T cells preventing recognition of alloantigen and thus GVHD. However, TCR disruption generally results in the elimination of the CD3 signaling component and alters the means of further T cell expansion.

[0396] Allogeneic cells are rapidly rejected by the host immune system. It has been demonstrated that, allogeneic leukocytes present in non-irradiated blood products will persist for no more than 5 to 6 days (Boni, Muranski et al. 2008 Blood 1;112(12):4746-54). Thus, to prevent rejection of allogeneic cells, the host's immune system usually has to be suppressed to some extent. However, in the case of adoptive cell transfer the use of immunosuppressive drugs also have a detrimental effect on the introduced therapeutic T cells. Therefore, to effectively use an adoptive immunotherapy approach in these conditions, the introduced cells would need to be resistant to the immunosuppressive treatment. Thus, in a particular embodiment, the present invention further comprises a step of modifying T cells to make them resistant to an immunosuppressive agent, preferably by inactivating at least one gene encoding a target for an immunosuppressive agent. An immunosuppressive agent is an agent that suppresses immune function by one of several mechanisms of action. An immunosuppressive agent can be, but is not limited to a calcineurin inhibitor, a target of rapamycin, an interleukin-2 receptor α -chain blocker, an inhibitor of inosine monophosphate dehydrogenase, an inhibitor of dihydrofolic acid reductase, a corticosteroid or an immunosuppressive antimetabolite. The present invention allows conferring immunosuppressive resistance to T cells for immunotherapy by inactivating the target of the immunosuppressive agent in T cells. As non-limiting examples, targets for an immunosuppressive agent can be a receptor for an immunosuppressive agent such as: CD52, glucocorticoid receptor (GR), a FKBP family gene member and a cyclophilin family gene member.

[0397] Immune checkpoints are inhibitory pathways that slow down or stop immune reactions and prevent excessive tissue damage from uncontrolled activity of immune cells. In certain embodiments, the immune checkpoint targeted is the programmed death-1 (PD-1 or CD279) gene (PDCD1). In other embodiments, the immune checkpoint targeted is cytotoxic T-lymphocyte-associated antigen (CTLA-4). In additional embodiments, the immune checkpoint targeted is another member of the CD28 and CTLA4 Ig superfamily such as BTLA, LAG3, ICOS, PDL1 or KIR. In further additional embodiments, the immune checkpoint targeted is a member of the TNFR superfamily such as CD40, OX40, CD137, GITR, CD27 or TIM-3.

[0398] Additional immune checkpoints include Src homology 2 domain-containing protein tyrosine phosphatase 1 (SHP-1) (Watson HA, et al., SHP-1: the next checkpoint target for cancer immunotherapy? *Biochem Soc Trans.* 2016 Apr 15;44(2):356-62). SHP-1 is a widely expressed inhibitory protein tyrosine phosphatase (PTP). In T-cells, it is a negative regulator of antigen-dependent activation and proliferation. It is a cytosolic protein, and therefore not amenable to antibody-mediated therapies, but its role in activation and proliferation makes it an attractive target for genetic manipulation in adoptive transfer strategies, such as chimeric antigen receptor (CAR) T cells. Immune checkpoints may also include T cell immunoreceptor with Ig and ITIM domains (TIGIT/Vstm3/WUCAM/VSIG9) and VISTA (Le Mercier I, et al., (2015) Beyond CTLA-4 and PD-1, the generation Z of negative checkpoint regulators. *Front. Immunol.* 6:418).

[0399] WO2014172606 relates to the use of MT1 and/or MT1 inhibitors to increase proliferation and/or activity of exhausted CD8⁺ T-cells and to decrease CD8⁺ T-cell exhaustion (e.g., decrease functionally exhausted or unresponsive CD8⁺ immune cells). In certain embodiments, metallothioneins are targeted by gene editing in adoptively transferred T cells.

[0400] In certain embodiments, targets of gene editing may be at least one targeted locus involved in the expression of an immune checkpoint protein. Such targets may include, but are not limited to CTLA4, PPP2CA, PPP2CB, PTPN6, PTPN22, PDCD1, ICOS (CD278), PDL1, KIR, LAG3, HAVCR2, BTLA, CD160, TIGIT, CD96, CRTAM, LAIR1, SIGLEC7, SIGLEC9, CD244 (2B4), TNFRSF10B, TNFRSF10A, CASP8, CASP10, CASP3, CASP6, CASP7, FADD, FAS, TGFBR1, TGFBR2, SMAD2, SMAD3, SMAD4, SMAD10, SKI, SKIL, TGIF1, IL10RA, IL10RB, HMOX2, IL6R, IL6ST, EIF2AK4, CSK, PAG1, SIT1,

FOXP3, PRDM1, BATF, VISTA, GUCY1A2, GUCY1A3, GUCY1B2, GUCY1B3, MT1, MT2, CD40, OX40, CD137, GITR, CD27, SHP-1 or TIM-3. In preferred embodiments, the gene locus involved in the expression of PD-1 or CTLA-4 genes is targeted. In other preferred embodiments, combinations of genes are targeted, such as but not limited to PD-1 and TIGIT.

[0401] In other embodiments, at least two genes are edited. Pairs of genes may include, but are not limited to PD1 and TCR α , PD1 and TCR β , CTLA-4 and TCR α , CTLA-4 and TCR β , LAG3 and TCR α , LAG3 and TCR β , Tim3 and TCR α , Tim3 and TCR β , BTLA and TCR α , BTLA and TCR β , BY55 and TCR α , BY55 and TCR β , TIGIT and TCR α , TIGIT and TCR β , B7H5 and TCR α , B7H5 and TCR β , LAIR1 and TCR α , LAIR1 and TCR β , SIGLEC10 and TCR α , SIGLEC10 and TCR β , 2B4 and TCR α , 2B4 and TCR β .

[0402] Whether prior to or after genetic modification of the T cells, the T cells can be activated and expanded generally using methods as described, for example, in U.S. Patents 6,352,694; 6,534,055; 6,905,680; 5,858,358; 6,887,466; 6,905,681; 7,144,575; 7,232,566; 7,175,843; 5,883,223; 6,905,874; 6,797,514; 6,867,041; and 7,572,631. T cells can be expanded in vitro or in vivo.

[0403] The practice of the present invention employs, unless otherwise indicated, conventional techniques of immunology, biochemistry, chemistry, molecular biology, microbiology, cell biology, genomics and recombinant DNA, which are within the skill of the art. See MOLECULAR CLONING: A LABORATORY MANUAL, 2nd edition (1989) (Sambrook, Fritsch and Maniatis); MOLECULAR CLONING: A LABORATORY MANUAL, 4th edition (2012) (Green and Sambrook); CURRENT PROTOCOLS IN MOLECULAR BIOLOGY (1987) (F. M. Ausubel, et al. eds.); the series METHODS IN ENZYMOLOGY (Academic Press, Inc.); PCR 2: A PRACTICAL APPROACH (1995) (M.J. MacPherson, B.D. Hames and G.R. Taylor eds.); ANTIBODIES, A LABORATORY MANUAL (1988) (Harlow and Lane, eds.); ANTIBODIES A LABORATORY MANUAL, 2nd edition (2013) (E.A. Greenfield ed.); and ANIMAL CELL CULTURE (1987) (R.I. Freshney, ed.).

[0404] The practice of the present invention employs, unless otherwise indicated, conventional techniques for generation of genetically modified mice. See Marten H. Hofker and Jan van Deursen, TRANSGENIC MOUSE METHODS AND PROTOCOLS, 2nd edition (2011).

[0405] In some embodiments, the invention described herein relates to a method for adoptive immunotherapy, in which T cells are edited ex vivo by CRISPR to modulate at least one gene and subsequently administered to a patient in need thereof. In some embodiments, the CRISPR editing comprising knocking-out or knocking-down the expression of at least one target gene in the edited T cells. In some embodiments, in addition to modulating the target gene, the T cells are also edited ex vivo by CRISPR to (1) knock-in an exogenous gene encoding a chimeric antigen receptor (CAR) or a T-cell receptor (TCR), (2) knock-out or knock-down expression of an immune checkpoint receptor, (3) knock-out or knock-down expression of an endogenous TCR, (4) knock-out or knock-down expression of a human leukocyte antigen class I (HLA-I) proteins, and/or (5) knock-out or knock-down expression of an endogenous gene encoding an antigen targeted by an exogenous CAR or TCR.

[0406] In some embodiments, the T cells are contacted ex vivo with an adeno-associated virus (AAV) vector encoding a CRISPR effector protein, and a guide molecule comprising a guide sequence hybridizable to a target sequence, a tracr mate sequence, and a tracr sequence hybridizable to the tracr mate sequence. In some embodiments, the T cells are contacted ex vivo (e.g., by electroporation) with a ribonucleoprotein (RNP) comprising a CRISPR effector protein complexed with a guide molecule, wherein the guide molecule comprising a guide sequence hybridizable to a target sequence, a tracr mate sequence, and a tracr sequence hybridizable to the tracr mate sequence. See Rupp et al., *Scientific Reports* 7:737 (2017); Liu et al., *Cell Research* 27:154-157 (2017). In some embodiments, the T cells are contacted ex vivo (e.g., by electroporation) with an mRNA encoding a CRISPR effector protein, and a guide molecule comprising a guide sequence hybridizable to a target sequence, a tracr mate sequence, and a tracr sequence hybridizable to the tracr mate sequence. See Eyquem et al., *Nature* 543:113-117 (2017). In some embodiments, the T cells are not contacted ex vivo with a lentivirus or retrovirus vector.

[0407] In some embodiments, the small molecule compounds herein may modulate the editing of T cells ex vivo by CRISPR to knock-in an exogenous gene encoding a CAR, thereby allowing the edited T cells to recognize cancer cells based on the expression of specific proteins located on the cell surface. In some embodiments, T cells are edited ex vivo by CRISPR to knock-in an exogenous gene encoding a TCR, thereby allowing the edited T cells to recognize proteins derived from either the surface or inside of the cancer cells. In some embodiments, the method comprising providing an exogenous CAR-encoding or TCR-

encoding sequence as a donor sequence, which can be integrated by homology-directed repair (HDR) into a genomic locus targeted by a CRISPR guide sequence. In some embodiments, targeting the exogenous CAR or TCR to an endogenous TCR α constant (TRAC) locus can reduce tonic CAR signaling and facilitate effective internalization and re-expression of the CAR following single or repeated exposure to antigen, thereby delaying effector T-cell differentiation and exhaustion. See Eyquem et al., *Nature* 543:113-117 (2017).

[0408] In some embodiments, the small molecule compounds herein may modulate the editing T cells ex vivo by CRISPR to block one or more immune checkpoint receptors to reduce immunosuppression by cancer cells. In some embodiments, T cells are edited ex vivo by CRISPR to knock-out or knock-down an endogenous gene involved in the programmed death-1 (PD-1) signaling pathway, such as PD-1 and PD-L1. In some embodiments, T cells are edited ex vivo by CRISPR to mutate the *Pdcd1* locus or the *CD274* locus. In some embodiments, T cells are edited ex vivo by CRISPR using one or more guide sequences targeting the first exon of PD-1. See Rupp et al., *Scientific Reports* 7:737 (2017); Liu et al., *Cell Research* 27:154-157 (2017).

[0409] In some embodiments, the small molecule compounds herein may modulate the editing T cells ex vivo by CRISPR to eliminate potential alloreactive TCRs to allow allogeneic adoptive transfer. In some embodiments, T cells are edited ex vivo by CRISPR to knock-out or knock-down an endogenous gene encoding a TCR (e.g., an $\alpha\beta$ TCR) to avoid graft-versus-host-disease (GVHD). In some embodiments, T cells are edited ex vivo by CRISPR to mutate the TRAC locus. In some embodiments, T cells are edited ex vivo by CRISPR using one or more guide sequences targeting the first exon of TRAC. See Liu et al., *Cell Research* 27:154-157 (2017). In some embodiments, the method comprises use of CRISPR to knock-in an exogenous gene encoding a CAR or a TCR into the TRAC locus, while simultaneously knocking-out the endogenous TCR (e.g., with a donor sequence encoding a self-cleaving P2A peptide following the CAR cDNA). See Eyquem et al., *Nature* 543:113-117 (2017). In some embodiments, the exogenous gene comprises a promoter-less CAR-encoding or TCR-encoding sequence which is inserted operably downstream of an endogenous TCR promoter.

[0410] In some embodiments, the small molecule compounds herein may modulate the editing T cells ex vivo by CRISPR to knock-out or knock-down an endogenous gene

encoding an HLA-I protein to minimize immunogenicity of the edited T cells. In some embodiments, T cells are edited ex vivo by CRISPR to mutate the beta-2 microglobulin (B2M) locus. In some embodiments, T cells are edited ex vivo by CRISPR using one or more guide sequences targeting the first exon of B2M. See Liu et al., Cell Research 27:154-157 (2017). In some embodiments, the method comprises use of CRISPR to knock-in an exogenous gene encoding a CAR or a TCR into the B2M locus, while simultaneously knocking-out the endogenous B2M (e.g., with a donor sequence encoding a self-cleaving P2A peptide following the CAR cDNA). See Eyquem et al., Nature 543:113-117 (2017). In some embodiments, the exogenous gene comprises a promoter-less CAR-encoding or TCR-encoding sequence which is inserted operably downstream of an endogenous B2M promoter.

[0411] In some embodiments, the small molecule compounds herein may modulate the editing T cells ex vivo by CRISPR to knock-out or knock-down an endogenous gene encoding an antigen targeted by an exogenous CAR or TCR. In some embodiments, the T cells are edited ex vivo by CRISPR to knock-out or knock-down the expression of a tumor antigen selected from human telomerase reverse transcriptase (hTERT), survivin, mouse double minute 2 homolog (MDM2), cytochrome P450 1B 1 (CYP1B), HER2/neu, Wilms' tumor gene 1 (WT1), livin, alphafetoprotein (AFP), carcinoembryonic antigen (CEA), mucin 16 (MUC16), MUC1, prostate-specific membrane antigen (PSMA), p53 or cyclin (DI) (see WO2016/011210). In some embodiments, the T cells are edited ex vivo by CRISPR to knock-out or knock-down the expression of an antigen selected from B cell maturation antigen (BCMA), transmembrane activator and CAML Interactor (TACI), or B-cell activating factor receptor (BAFF-R), CD38, CD138, CS-1, CD33, CD26, CD30, CD53, CD92, CD100, CD148, CD150, CD200, CD261, CD262, or CD362 (see WO2017/011804).

Xenotransplantation

[0412] The small molecule compounds herein may modulate RNA-guided DNA nucleases that are adapted to be used to provide modified tissues for transplantation. For example, RNA-guided DNA nucleases may be used to knockout, knockdown or disrupt selected genes in an animal, such as a transgenic pig (such as the human heme oxygenase-1 transgenic pig line), for example by disrupting expression of genes that encode epitopes recognized by the human immune system, e.g., xenoantigen genes. Candidate porcine genes for disruption may for example include $\alpha(1,3)$ -galactosyltransferase and cytidine monophosphate-N-acetylneuraminic acid hydroxylase genes (see PCT Patent Publication

WO 2014/066505). In addition, genes encoding endogenous retroviruses may be disrupted, for example the genes encoding all porcine endogenous retroviruses (see Yang et al., 2015, Genome-wide inactivation of porcine endogenous retroviruses (PERVs), Science 27 November 2015: Vol. 350 no. 6264 pp. 1101-1104). In addition, RNA-guided DNA nucleases may be used to target a site for integration of additional genes in xenotransplant donor animals, such as a human CD55 gene to improve protection against hyperacute rejection.

General Gene Therapy Considerations

[0413] Examples of disease-associated genes and polynucleotides AMD disease specific information is available from McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, Md.) and National Center for Biotechnology Information, National Library of Medicine (Bethesda, Md.), available on the World Wide Web.

[0414] Mutations in these genes and pathways can result in production of improper proteins or proteins in improper amounts which affect function. Further examples of genes, diseases and proteins are hereby incorporated by reference from US Provisional application 61/736,527 filed December 12, 2012. Such genes, proteins and pathways may be the target polynucleotide of a CRISPR complex of the present invention. Examples of disease-associated genes and polynucleotides are listed in Tables 1 and 2. Examples of signaling biochemical pathway-associated genes and polynucleotides are listed in Table 3.

[0415] Table 1

DISEASE/DISORDERS	GENE(S)
Neoplasia	PTEN; ATM; ATR; EGFR; ERBB2; ERBB3; ERBB4;
	Notch1; Notch2; Notch3; Notch4; AKT; AKT2; AKT3; HIF;
	HIF1a; HIF3a; Met; HRG; Bcl2; PPAR alpha; PPAR
	gamma; WT1 (Wilms Tumor); FGF Receptor Family
	members (5 members: 1, 2, 3, 4, 5); CDKN2a; APC; RB
	(retinoblastoma); MEN1; VHL; BRCA1; BRCA2; AR
	(Androgen Receptor); TSG101; IGF; IGF Receptor; Igf1 (4
	variants); Igf2 (3 variants); Igf 1 Receptor; Igf 2 Receptor;
	Bax; Bcl2; caspases family (9 members:
	1, 2, 3, 4, 6, 7, 8, 9, 12); Kras; Apc
Age-related Macular	Abcr; Ccl2; Cc2; cp (ceruloplasmin); Timp3; cathepsinD;
Degeneration	Vldlr; Ccr2
Schizophrenia	Neuregulin1 (Nrg1); Erb4 (receptor for Neuregulin);
	Complexin1 (Cplx1); Tph1 Tryptophan hydroxylase; Tph2
	Tryptophan hydroxylase 2; Neurexin 1; GSK3; GSK3a;
	GSK3b

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

[00100] Table 2:

Blood and coagulation diseases and disorders	Anemia (CDAN1, CDA1, RPS19, DBA, PKLR, PK1, NT5C3, UMPH1, PSN1, RHAG, RH50A, NRAMP2, SPTB, ALAS2, ANH1, ASB, ABCB7, ABC7, ASAT); Bare lymphocyte syndrome (TAPBP, TPSN, TAP2, ABCB3, PSF2, RING11, MHC2TA, C2TA, RFX5, RFXAP, RFX5), Bleeding disorders (TBXA2R, P2RX1, P2X1); Factor H and factor H-like 1 (HF1, CFH, HUS); Factor V and factor VIII (MCFD2); Factor VII deficiency (F7); Factor X deficiency (F10); Factor XI deficiency (F11); Factor XII deficiency (F12, HAF); Factor XIIIa deficiency (F13A1, F13A); Factor XIIIb deficiency (F13B); Fanconi anemia (FANCA, FACA, FA1, FA, FAA, FAAP95, FAAP90, FLJ34064, FANCB, FANCC, FACC, BRCA2, FANCD1, FANCD2, FANCD, FACD, FAD, FANCE, FACE, FANCF, XRCC9, FANCG, BRIP1, BACH1, FANCI, PHF9, FANCL, FANCM, KIAA1596); Hemophagocytic lymphohistiocytosis disorders (PRF1, HPLH2, UNC13D, MUNC13-4, HPLH3, HLH3, FHL3); Hemophilia A (F8, F8C, HEMA); Hemophilia B (F9, HEMB), Hemorrhagic disorders (PI, ATT, F5); Leukocyte deficiencies and disorders (ITGB2, CD18, LCAMB, LAD, EIF2B1, EIF2BA, EIF2B2,
--	--

	EIF2B3, EIF2B5, LVWM, CACH, CLE, EIF2B4); Sickle cell anemia (HBB); Thalassemia (HBA2, HBB, HBD, LCRB, HBA1).
Cell dysregulation and oncology diseases and disorders	B-cell non-Hodgkin lymphoma (BCL7A, BCL7); Leukemia (TAL1, TCL5, SCL, TAL2, FLT3, NBS1, NBS, ZNFN1A1, IK1, LYF1, HOXD4, HOX4B, BCR, CML, PHL, ALL, ARNT, KRAS2, RASK2, GMPS, AF10, ARHGEF12, LARG, KIAA0382, CALM, CLTH, CEBPA, CEBP, CHIC2, BTL, FLT3, KIT, PBT, LPP, NPM1, NUP214, D9S46E, CAN, CAIN, RUNX1, CBFA2, AML1, WHSC1L1, NSD3, FLT3, AF1Q, NPM1, NUMA1, ZNF145, PLZF, PML, MYL, STAT5B, AF10, CALM, CLTH, ARL11, ARLTS1, P2RX7, P2X7, BCR, CML, PHL, ALL, GRAF, NF1, VRNF, WSS, NFNS, PTPN11, PTP2C, SHP2, NS1, BCL2, CCND1, PRAD1, BCL1, TCRA, GATA1, GF1, ERYF1, NFE1, ABL1, NQO1, DIA4, NMOR1, NUP214, D9S46E, CAN, CAIN).
Inflammation and immune related diseases and disorders	AIDS (KIR3DL1, NKAT3, NKB1, AMB11, KIR3DS1, IFNG, CXCL12, SDF1); Autoimmune lymphoproliferative syndrome (TNFRSF6, APT1, FAS, CD95, ALPS1A); Combined immunodeficiency, (IL2RG, SCIDX1, SCIDX, IMD4); HIV-1 (CCL5, SCYA5, D17S136E, TCP228), HIV susceptibility or infection (IL10, CSIF, CMKBR2, CCR2, CMKBR5, CCCR5 (CCR5)); Immunodeficiencies (CD3E, CD3G, AICDA, AID, HIGM2, TNFRSF5, CD40, UNG, DGU, HIGM4, TNFSF5, CD40LG, HIGM1, IGM, FOXP3, IPEX, AIID, XPID, PIDX, TNFRSF14B, TACI); Inflammation (IL-10, IL-1 (IL-1a, IL-1b), IL-13, IL-17 (IL-17a (CTLA8), IL-17b, IL-17c, IL-17d, IL-17f), IL-23, Cx3cr1, ptpn22, TNFa, NOD2/CARD15 for IBD, IL-6, IL-12 (IL-12a, IL-12b), CTLA4, Cx3cl1); Severe combined immunodeficiencies (SCIDs)(JAK3, JAKL, DCLRE1C, ARTEMIS, SCIDA, RAG1, RAG2, ADA, PTPRC, CD45, LCA, IL7R, CD3D, T3D, IL2RG, SCIDX1, SCIDX, IMD4).
Metabolic, liver, kidney and protein diseases and disorders	Amyloid neuropathy (TTR, PALB); Amyloidosis (APOA1, APP, AAA, CVAP, AD1, GSN, FGA, LYZ, TTR, PALB); Cirrhosis (KRT18, KRT8, CIRH1A, NAIC, TEX292, KIAA1988); Cystic fibrosis (CFTR, ABCC7, CF, MRP7); Glycogen storage diseases (SLC2A2, GLUT2, G6PC, G6PT, G6PT1, GAA, LAMP2, LAMPB, AGL, GDE, GBE1, GYS2, PYGL, PFKM); Hepatic adenoma, 142330 (TCF1, HNF1A, MODY3), Hepatic failure, early onset, and neurologic disorder (SCOD1, SCO1), Hepatic lipase deficiency (LIPC), Hepatoblastoma, cancer and carcinomas (CTNNB1, PDGFRL, PDGRL, PRLTS, AXIN1, AXIN, CTNNB1, TP53, P53, LFS1, IGF2R, MPRI, MET, CASP8, MCH5; Medullary cystic kidney disease (UMOD, HNFJ, FJHN, MCKD2, ADMCKD2); Phenylketonuria (PAH, PKU1, QDPR, DHPR, PTS); Polycystic kidney and hepatic disease (FCYT, PKHD1, ARPKD, PKD1, PKD2, PKD4, PKDTS, PRKCSH, G19P1, PCLD, SEC63).
Muscular / Skeletal diseases and	Becker muscular dystrophy (DMD, BMD, MYF6), Duchenne Muscular Dystrophy (DMD, BMD); Emery-Dreifuss muscular dystrophy

disorders	(LMNA, LMN1, EMD2, FPLD, CMD1A, HGPS, LGMD1B, LMNA, LMN1, EMD2, FPLD, CMD1A); Facioscapulohumeral muscular dystrophy (FSHMD1A, FSHD1A); Muscular dystrophy (FKRP, MDC1C, LGMD2I, LAMA2, LAMM, LARGE, KIAA0609, MDC1D, FCMD, TTID, MYOT, CAPN3, CANP3, DYSF, LGMD2B, SGCG, LGMD2C, DMDA1, SCG3, SGCA, ADL, DAG2, LGMD2D, DMDA2, SGCB, LGMD2E, SGCD, SGD, LGMD2F, CMD1L, TCAP, LGMD2G, CMD1N, TRIM32, HT2A, LGMD2H, FKRP, MDC1C, LGMD2I, TTN, CMD1G, TMD, LGMD2J, POMT1, CAV3, LGMD1C, SEPN1, SELN, RSMD1, PLEC1, PLTN, EBS1); Osteopetrosis (LRP5, BMND1, LRP7, LR3, OPPG, VBCH2, CLCN7, CLC7, OPTA2, OSTM1, GL, TCIRG1, TIRC7, OC116, OPTB1); Muscular atrophy (VAPB, VAPC, ALS8, SMN1, SMA1, SMA2, SMA3, SMA4, BSCL2, SPG17, GARS, SMAD1, CMT2D, HEXB, IGHMBP2, SMUBP2, CATF1, SMARD1).
Neurological and neuronal diseases and disorders	ALS (SOD1, ALS2, STEX, FUS, TARDBP, VEGF (VEGF-a, VEGF-b, VEGF-c); Alzheimer disease (APP, AAA, CVAP, AD1, APOE, AD2, PSEN2, AD4, STM2, APBB2, FE65L1, NOS3, PLAUI, URK, ACE, DCP1, ACE1, MPO, PACIP1, PAXIP1L, PTIP, A2M, BLMH, BMH, PSEN1, AD3); Autism (Mecp2, BZRAP1, MDGA2, Sema5A, Neurexin 1, GLO1, MECP2, RTT, PPMX, MRX16, MRX79, NLGN3, NLGN4, KIAA1260, AUTSX2); Fragile X Syndrome (FMR2, FXR1, FXR2, mGLUR5); Huntington's disease and disease like disorders (HD, IT15, PRNP, PRIP, JPH3, JP3, HDL2, TBP, SCA17); Parkinson disease (NR4A2, NURR1, NOT, TINUR, SNCAIP, TBP, SCA17, SNCA, NACP, PARK1, PARK4, DJ1, PARK7, LRRK2, PARK8, PINK1, PARK6, UCHL1, PARK5, SNCA, NACP, PARK1, PARK4, PRKN, PARK2, PDJ, DBH, NDUFV2); Rett syndrome (MECP2, RTT, PPMX, MRX16, MRX79, CDKL5, STK9, MECP2, RTT, PPMX, MRX16, MRX79, x-Synuclein, DJ-1); Schizophrenia (Neuregulin1 (Nrg1), Erb4 (receptor for Neuregulin), Complexin1 (Cplx1), Tph1 Tryptophan hydroxylase, Tph2, Tryptophan hydroxylase 2, Neurexin 1, GSK3, GSK3a, GSK3b, 5-HTT (Slc6a4), COMT, DRD (Drd1a), SLC6A3, DAOA, DTNBP1, Dao (Dao1)); Secretase Related Disorders (APH-1 (alpha and beta), Presenilin (Psen1), nicastrin, (Ncstn), PEN-2, Nos1, Parp1, Nat1, Nat2); Trinucleotide Repeat Disorders (HTT (Huntington's Dx), SBMA/SMA1/AR (Kennedy's Dx), FXN/X25 (Friedrich's Ataxia), ATX3 (Machado- Joseph's Dx), ATXN1 and ATXN2 (spinocerebellar ataxias), DMPK (myotonic dystrophy), Atrophin-1 and Atn1 (DRPLA Dx), CBP (Creb-BP - global instability), VLDLR (Alzheimer's), Atxn7, Atxn10).
Occular diseases and disorders	Age-related macular degeneration (Abcr, Ccl2, Cc2, cp (ceruloplasmin), Timp3, cathepsinD, Vldlr, Ccr2); Cataract (CRYAA, CRYA1, CRYBB2, CRYB2, PITX3, BFSP2, CP49, CP47, CRYAA, CRYA1, PAX6, AN2, MGDA, CRYBA1, CRYB1, CRYGC, CRYG3, CCL, LIM2, MP19, CRYGD, CRYG4, BFSP2, CP49, CP47, HSF4, CTM, HSF4, CTM, MIP, AQP0, CRYAB, CRYA2, CTPP2, CRYBB1, CRYGD, CRYG4, CRYBB2,

	CRYB2, CRYGC, CRYG3, CCL, CRYAA, CRYA1, GJA8, CX50, CAE1, GJA3, CX46, CZP3, CAE3, CCM1, CAM, KRIT1); Corneal clouding and dystrophy (APOA1, TGFBI, CSD2, CDGG1, CSD, BIGH3, CDG2, TACSTD2, TROP2, M1S1, VSX1, RINX, PPCD, PPD, KTCN, COL8A2, FECD, PPCD2, PIP5K3, CFD); Cornea plana congenital (KERA, CNA2); Glaucoma (MYOC, TIGR, GLC1A, JOAG, GPOA, OPTN, GLC1E, FIP2, HYPL, NRP, CYP1B1, GLC3A, OPA1, NTG, NPG, CYP1B1, GLC3A); Leber congenital amaurosis (CRB1, RP12, CRX, CORD2, CRD, RPGRIP1, LCA6, CORD9, RPE65, RP20, AIPL1, LCA4, GUCY2D, GUC2D, LCA1, CORD6, RDH12, LCA3); Macular dystrophy (ELOVL4, ADMD, STGD2, STGD3, RDS, RP7, PRPH2, PRPH, AVMD, AOFMD, VMD2).
--	--

[0416] Table 3:

CELLULAR FUNCTION	GENES
PI3K/AKT Signaling	PRKCE; ITGAM; ITGA5; IRAK1; PRKAA2; EIF2AK2;
	PTEN; EIF4E; PRKCZ; GRK6; MAPK1; TSC1; PLK1;
	AKT2; IKBKB; PIK3CA; CDK8; CDKN1B; NFKB2; BCL2;
	PIK3CB; PPP2R1A; MAPK8; BCL2L1; MAPK3; TSC2;
	ITGA1; KRAS; EIF4EBP1; RELA; PRKCD; NOS3;
	PRKAA1; MAPK9; CDK2; PPP2CA; PIM1; ITGB7;
	YWHAZ; ILK; TP53; RAF1; IKBKG; RELB; DYRK1A;
	CDKN1A; ITGB1; MAP2K2; JAK1; AKT1; JAK2; PIK3R1;
	CHUK; PDPK1; PPP2R5C; CTNNB1; MAP2K1; NFKB1;
	PAK3; ITGB3; CCND1; GSK3A; FRAP1; SFN; ITGA2;
	TTK; CSNK1A1; BRAF; GSK3B; AKT3; FOXO1; SGK;
	HSP90AA1; RPS6KB1
ERK/MAPK Signaling	PRKCE; ITGAM; ITGA5; HSPB1; IRAK1; PRKAA2;
	EIF2AK2; RAC1; RAP1A; TLN1; EIF4E; ELK1; GRK6;
	MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8; CREB1;
	PRKCI; PTK2; FOS; RPS6KA4; PIK3CB; PPP2R1A;
	PIK3C3; MAPK8; MAPK3; ITGA1; ETS1; KRAS; MYCN;
	EIF4EBP1; PPARG; PRKCD; PRKAA1; MAPK9; SRC;
	CDK2; PPP2CA; PIM1; PIK3C2A; ITGB7; YWHAZ;
	PPP1CC; KSR1; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIK3R1; STAT3; PPP2R5C; MAP2K1;
	PAK3; ITGB3; ESR1; ITGA2; MYC; TTK; CSNK1A1;
	CRKL; BRAF; ATF4; PRKCA; SRF; STAT1; SGK
Glucocorticoid Receptor Signaling	RAC1; TAF4B; EP300; SMAD2; TRAF6; PCAF; ELK1;
	MAPK1; SMAD3; AKT2; IKBKB; NCOR2; UBE2I;
	PIK3CA; CREB1; FOS; HSPA5; NFKB2; BCL2;
	MAP3K14; STAT5B; PIK3CB; PIK3C3; MAPK8; BCL2L1;
	MAPK3; TSC22D3; MAPK10; NRIP1; KRAS; MAPK13;
	RELA; STAT5A; MAPK9; NOS2A; PBX1; NR3C1;
	PIK3C2A; CDKN1C; TRAF2; SERPINE1; NCOA3;
	MAPK14; TNF; RAF1; IKBKG; MAP3K7; CREBBP;
	CDKN1A; MAP2K2; JAK1; IL8; NCOA2; AKT1; JAK2;

	PIK3R1; CHUK; STAT3; MAP2K1; NFKB1; TGFBR1;
	ESR1; SMAD4; CEBPB; JUN; AR; AKT3; CCL2; MMP1;
	STAT1; IL6; HSP90AA1
Axonal Guidance Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; ADAM12;
	IGF1; RAC1; RAP1A; EIF4E; PRKCZ; NRP1; NTRK2;
	ARHGEF7; SMO; ROCK2; MAPK1; PGF; RAC2;
	PTPN11; GNAS; AKT2; PIK3CA; ERBB2; PRKCI; PTK2;
	CFL1; GNAQ; PIK3CB; CXCL12; PIK3C3; WNT11;
	PRKD1; GNB2L1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	PRKCD; PIK3C2A; ITGB7; GLI2; PXN; VASP; RAF1;
	FYN; ITGB1; MAP2K2; PAK4; ADAM17; AKT1; PIK3R1;
	GLI1; WNT5A; ADAM10; MAP2K1; PAK3; ITGB3;
	CDC42; VEGFA; ITGA2; EPHA8; CRKL; RND1; GSK3B;
	AKT3; PRKCA
Ephrin Receptor Signaling	PRKCE; ITGAM; ROCK1; ITGA5; CXCR4; IRAK1;
	PRKAA2; EIF2AK2; RAC1; RAP1A; GRK6; ROCK2;
	MAPK1; PGF; RAC2; PTPN11; GNAS; PLK1; AKT2;
	DOK1; CDK8; CREB1; PTK2; CFL1; GNAQ; MAP3K14;
	CXCL12; MAPK8; GNB2L1; ABL1; MAPK3; ITGA1;
	KRAS; RHOA; PRKCD; PRKAA1; MAPK9; SRC; CDK2;
	PIM1; ITGB7; PXN; RAF1; FYN; DYRK1A; ITGB1;
	MAP2K2; PAK4; AKT1; JAK2; STAT3; ADAM10;
	MAP2K1; PAK3; ITGB3; CDC42; VEGFA; ITGA2;
	EPHA8; TTK; CSNK1A1; CRKL; BRAF; PTPN13; ATF4;
	AKT3; SGK
Actin Cytoskeleton Signaling	ACTN4; PRKCE; ITGAM; ROCK1; ITGA5; IRAK1;
	PRKAA2; EIF2AK2; RAC1; INS; ARHGEF7; GRK6;
	ROCK2; MAPK1; RAC2; PLK1; AKT2; PIK3CA; CDK8;
	PTK2; CFL1; PIK3CB; MYH9; DIAPH1; PIK3C3; MAPK8;
	F2R; MAPK3; SLC9A1; ITGA1; KRAS; RHOA; PRKCD;
	PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A; ITGB7;
	PPP1CC; PXN; VIL2; RAF1; GSN; DYRK1A; ITGB1;
	MAP2K2; PAK4; PIP5K1A; PIK3R1; MAP2K1; PAK3;
	ITGB3; CDC42; APC; ITGA2; TTK; CSNK1A1; CRKL;
	BRAF; VAV3; SGK
Huntington's Disease Signaling	PRKCE; IGF1; EP300; RCOR1; PRKCZ; HDAC4; TGM2;
	MAPK1; CAPNS1; AKT2; EGFR; NCOR2; SP1; CAPN2;
	PIK3CA; HDAC5; CREB1; PRKCI; HSPA5; REST;
	GNAQ; PIK3CB; PIK3C3; MAPK8; IGF1R; PRKD1;
	GNB2L1; BCL2L1; CAPN1; MAPK3; CASP8; HDAC2;
	HDAC7A; PRKCD; HDAC11; MAPK9; HDAC9; PIK3C2A;
	HDAC3; TP53; CASP9; CREBBP; AKT1; PIK3R1;
	PDPK1; CASP1; APAF1; FRAP1; CASP2; JUN; BAX;
	ATF4; AKT3; PRKCA; CLTC; SGK; HDAC6; CASP3

Apoptosis Signaling	PRKCE; ROCK1; BID; IRAK1; PRKAA2; EIF2AK2; BAK1;
	BIRC4; GRK6; MAPK1; CAPNS1; PLK1; AKT2; IKBKB;
	CAPN2; CDK8; FAS; NFKB2; BCL2; MAP3K14; MAPK8;
	BCL2L1; CAPN1; MAPK3; CASP8; KRAS; RELA;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; TP53; TNF;
	RAF1; IKBKG; RELB; CASP9; DYRK1A; MAP2K2;
	CHUK; APAF1; MAP2K1; NFKB1; PAK3; LMNA; CASP2;
	BIRC2; TTK; CSNK1A1; BRAF; BAX; PRKCA; SGK;
	CASP3; BIRC3; PARP1
B Cell Receptor Signaling	RAC1; PTEN; LYN; ELK1; MAPK1; RAC2; PTPN11;
	AKT2; IKBKB; PIK3CA; CREB1; SYK; NFKB2; CAMK2A;
	MAP3K14; PIK3CB; PIK3C3; MAPK8; BCL2L1; ABL1;
	MAPK3; ETS1; KRAS; MAPK13; RELA; PTPN6; MAPK9;
	EGR1; PIK3C2A; BTK; MAPK14; RAF1; IKBKG; RELB;
	MAP3K7; MAP2K2; AKT1; PIK3R1; CHUK; MAP2K1;
	NFKB1; CDC42; GSK3A; FRAP1; BCL6; BCL10; JUN;
	GSK3B; ATF4; AKT3; VAV3; RPS6KB1
Leukocyte Extravasation Signaling	ACTN4; CD44; PRKCE; ITGAM; ROCK1; CXCR4; CYBA;
	RAC1; RAP1A; PRKCZ; ROCK2; RAC2; PTPN11;
	MMP14; PIK3CA; PRKCI; PTK2; PIK3CB; CXCL12;
	PIK3C3; MAPK8; PRKD1; ABL1; MAPK10; CYBB;
	MAPK13; RHOA; PRKCD; MAPK9; SRC; PIK3C2A; BTK;
	MAPK14; NOX1; PXN; VIL2; VASP; ITGB1; MAP2K2;
	CTNND1; PIK3R1; CTNNB1; CLDN1; CDC42; F11R; ITK;
	CRKL; VAV3; CTTN; PRKCA; MMP1; MMP9
Integrin Signaling	ACTN4; ITGAM; ROCK1; ITGA5; RAC1; PTEN; RAP1A;
	TLN1; ARHGEF7; MAPK1; RAC2; CAPNS1; AKT2;
	CAPN2; PIK3CA; PTK2; PIK3CB; PIK3C3; MAPK8;
	CAV1; CAPN1; ABL1; MAPK3; ITGA1; KRAS; RHOA;
	SRC; PIK3C2A; ITGB7; PPP1CC; ILK; PXN; VASP;
	RAF1; FYN; ITGB1; MAP2K2; PAK4; AKT1; PIK3R1;
	TNK2; MAP2K1; PAK3; ITGB3; CDC42; RND3; ITGA2;
	CRKL; BRAF; GSK3B; AKT3
Acute Phase Response Signaling	IRAK1; SOD2; MYD88; TRAF6; ELK1; MAPK1; PTPN11;
	AKT2; IKBKB; PIK3CA; FOS; NFKB2; MAP3K14;
	PIK3CB; MAPK8; RIPK1; MAPK3; IL6ST; KRAS;
	MAPK13; IL6R; RELA; SOCS1; MAPK9; FTL; NR3C1;
	TRAF2; SERPINE1; MAPK14; TNF; RAF1; PDK1;
	IKBKG; RELB; MAP3K7; MAP2K2; AKT1; JAK2; PIK3R1;
	CHUK; STAT3; MAP2K1; NFKB1; FRAP1; CEBPB; JUN;
	AKT3; IL1R1; IL6
PTEN Signaling	ITGAM; ITGA5; RAC1; PTEN; PRKCZ; BCL2L11;
	MAPK1; RAC2; AKT2; EGFR; IKBKB; CBL; PIK3CA;
	CDKN1B; PTK2; NFKB2; BCL2; PIK3CB; BCL2L1;

	MAPK3; ITGA1; KRAS; ITGB7; ILK; PDGFRB; INSR;
	RAF1; IKBKG; CASP9; CDKN1A; ITGB1; MAP2K2;
	AKT1; PIK3R1; CHUK; PDGFRA; PDPK1; MAP2K1;
	NFKB1; ITGB3; CDC42; CCND1; GSK3A; ITGA2;
	GSK3B; AKT3; FOXO1; CASP3; RPS6KB1
p53 Signaling	PTEN; EP300; BBC3; PCAF; FASN; BRCA1; GADD45A;
	BIRC5; AKT2; PIK3CA; CHEK1; TP53INP1; BCL2;
	PIK3CB; PIK3C3; MAPK8; THBS1; ATR; BCL2L1; E2F1;
	PMAIP1; CHEK2; TNFRSF10B; TP73; RB1; HDAC9;
	CDK2; PIK3C2A; MAPK14; TP53; LRDD; CDKN1A;
	HIPK2; AKT1; PIK3R1; RRM2B; APAF1; CTNNB1;
	SIRT1; CCND1; PRKDC; ATM; SFN; CDKN2A; JUN;
	SNAI2; GSK3B; BAX; AKT3
Aryl Hydrocarbon Receptor Signaling	HSPB1; EP300; FASN; TGM2; RXRA; MAPK1; NQO1;
	NCOR2; SP1; ARNT; CDKN1B; FOS; CHEK1;
	SMARCA4; NFKB2; MAPK8; ALDH1A1; ATR; E2F1;
	MAPK3; NRIP1; CHEK2; RELA; TP73; GSTP1; RB1;
	SRC; CDK2; AHR; NFE2L2; NCOA3; TP53; TNF;
	CDKN1A; NCOA2; APAF1; NFKB1; CCND1; ATM; ESR1;
	CDKN2A; MYC; JUN; ESR2; BAX; IL6; CYP1B1;
	HSP90AA1
Xenobiotic Metabolism Signaling	PRKCE; EP300; PRKCZ; RXRA; MAPK1; NQO1;
	NCOR2; PIK3CA; ARNT; PRKCI; NFKB2; CAMK2A;
	PIK3CB; PPP2R1A; PIK3C3; MAPK8; PRKD1;
	ALDH1A1; MAPK3; NRIP1; KRAS; MAPK13; PRKCD;
	GSTP1; MAPK9; NOS2A; ABCB1; AHR; PPP2CA; FTL;
	NFE2L2; PIK3C2A; PPARGC1A; MAPK14; TNF; RAF1;
	CREBBP; MAP2K2; PIK3R1; PPP2R5C; MAP2K1;
	NFKB1; KEAP1; PRKCA; EIF2AK3; IL6; CYP1B1;
	HSP90AA1
SAPK/JNK Signaling	PRKCE; IRAK1; PRKAA2; EIF2AK2; RAC1; ELK1;
	GRK6; MAPK1; GADD45A; RAC2; PLK1; AKT2; PIK3CA;
	FADD; CDK8; PIK3CB; PIK3C3; MAPK8; RIPK1;
	GNB2L1; IRS1; MAPK3; MAPK10; DAXX; KRAS;
	PRKCD; PRKAA1; MAPK9; CDK2; PIM1; PIK3C2A;
	TRAF2; TP53; LCK; MAP3K7; DYRK1A; MAP2K2;
	PIK3R1; MAP2K1; PAK3; CDC42; JUN; TTK; CSNK1A1;
	CRKL; BRAF; SGK
PPAr/RXR Signaling	PRKAA2; EP300; INS; SMAD2; TRAF6; PPARA; FASN;
	RXRA; MAPK1; SMAD3; GNAS; IKBKB; NCOR2;
	ABCA1; GNAQ; NFKB2; MAP3K14; STAT5B; MAPK8;
	IRS1; MAPK3; KRAS; RELA; PRKAA1; PPARGC1A;
	NCOA3; MAPK14; INSR; RAF1; IKBKG; RELB; MAP3K7;
	CREBBP; MAP2K2; JAK2; CHUK; MAP2K1; NFKB1;

	TGFBR1; SMAD4; JUN; IL1R1; PRKCA; IL6; HSP90AA1; ADIPOQ
NF-KB Signaling	IRAK1; EIF2AK2; EP300; INS; MYD88; PRKCZ; TRAF6; TBK1; AKT2; EGFR; IKBKB; PIK3CA; BTRC; NFKB2; MAP3K14; PIK3CB; PIK3C3; MAPK8; RIPK1; HDAC2; KRAS; RELA; PIK3C2A; TRAF2; TLR4; PDGFRB; TNF; INSR; LCK; IKBKG; RELB; MAP3K7; CREBBP; AKT1; PIK3R1; CHUK; PDGFRA; NFKB1; TLR2; BCL10; GSK3B; AKT3; TNFAIP3; IL1R1
Neuregulin Signaling	ERBB4; PRKCE; ITGAM; ITGA5; PTEN; PRKCZ; ELK1; MAPK1; PTPN11; AKT2; EGFR; ERBB2; PRKCI; CDKN1B; STAT5B; PRKD1; MAPK3; ITGA1; KRAS; PRKCD; STAT5A; SRC; ITGB7; RAF1; ITGB1; MAP2K2; ADAM17; AKT1; PIK3R1; PDPK1; MAP2K1; ITGB3; EREG; FRAP1; PSEN1; ITGA2; MYC; NRG1; CRKL; AKT3; PRKCA; HSP90AA1; RPS6KB1
Wnt & Beta catenin Signaling	CD44; EP300; LRP6; DVL3; CSNK1E; GJA1; SMO; AKT2; PIN1; CDH1; BTRC; GNAQ; MARK2; PPP2R1A; WNT11; SRC; DKK1; PPP2CA; SOX6; SFRP2; ILK; LEF1; SOX9; TP53; MAP3K7; CREBBP; TCF7L2; AKT1; PPP2R5C; WNT5A; LRP5; CTNNB1; TGFBR1; CCND1; GSK3A; DVL1; APC; CDKN2A; MYC; CSNK1A1; GSK3B; AKT3; SOX2
Insulin Receptor Signaling	PTEN; INS; EIF4E; PTPN1; PRKCZ; MAPK1; TSC1; PTPN11; AKT2; CBL; PIK3CA; PRKCI; PIK3CB; PIK3C3; MAPK8; IRS1; MAPK3; TSC2; KRAS; EIF4EBP1; SLC2A4; PIK3C2A; PPP1CC; INSR; RAF1; FYN; MAP2K2; JAK1; AKT1; JAK2; PIK3R1; PDPK1; MAP2K1; GSK3A; FRAP1; CRKL; GSK3B; AKT3; FOXO1; SGK; RPS6KB1
IL-6 Signaling	HSPB1; TRAF6; MAPKAPK2; ELK1; MAPK1; PTPN11; IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK3; MAPK10; IL6ST; KRAS; MAPK13; IL6R; RELA; SOCS1; MAPK9; ABCB1; TRAF2; MAPK14; TNF; RAF1; IKBKG; RELB; MAP3K7; MAP2K2; IL8; JAK2; CHUK; STAT3; MAP2K1; NFKB1; CEBPB; JUN; IL1R1; SRF; IL6
Hepatic Cholestasis	PRKCE; IRAK1; INS; MYD88; PRKCZ; TRAF6; PPARA; RXRA; IKBKB; PRKCI; NFKB2; MAP3K14; MAPK8; PRKD1; MAPK10; RELA; PRKCD; MAPK9; ABCB1; TRAF2; TLR4; TNF; INSR; IKBKG; RELB; MAP3K7; IL8; CHUK; NR1H2; TJP2; NFKB1; ESR1; SREBF1; FGFR4; JUN; IL1R1; PRKCA; IL6
IGF-1 Signaling	IGF1; PRKCZ; ELK1; MAPK1; PTPN11; NEDD4; AKT2; PIK3CA; PRKCI; PTK2; FOS; PIK3CB; PIK3C3; MAPK8;

	IGF1R; IRS1; MAPK3; IGFBP7; KRAS; PIK3C2A;
	YWHAZ; PXN; RAF1; CASP9; MAP2K2; AKT1; PIK3R1;
	PDPK1; MAP2K1; IGFBP2; SFN; JUN; CYR61; AKT3;
	FOXO1; SRF; CTGF; RPS6KB1
NRF2-mediated Oxidative	PRKCE; EP300; SOD2; PRKCZ; MAPK1; SQSTM1;
Stress Response	NQO1; PIK3CA; PRKCI; FOS; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; KRAS; PRKCD; GSTP1; MAPK9; FTL;
	NFE2L2; PIK3C2A; MAPK14; RAF1; MAP3K7; CREBBP;
	MAP2K2; AKT1; PIK3R1; MAP2K1; PPIB; JUN; KEAP1;
	GSK3B; ATF4; PRKCA; EIF2AK3; HSP90AA1
Hepatic Fibrosis/Hepatic	EDN1; IGF1; KDR; FLT1; SMAD2; FGFR1; MET; PGF;
Stellate Cell Activation	SMAD3; EGFR; FAS; CSF1; NFKB2; BCL2; MYH9;
	IGF1R; IL6R; RELA; TLR4; PDGFRB; TNF; RELB; IL8;
	PDGFRA; NFKB1; TGFBR1; SMAD4; VEGFA; BAX;
	IL1R1; CCL2; HGF; MMP1; STAT1; IL6; CTGF; MMP9
PPAR Signaling	EP300; INS; TRAF6; PPARA; RXRA; MAPK1; IKBKB;
	NCOR2; FOS; NFKB2; MAP3K14; STAT5B; MAPK3;
	NRIP1; KRAS; PPARG; RELA; STAT5A; TRAF2;
	PPARGC1A; PDGFRB; TNF; INSR; RAF1; IKBKG;
	RELB; MAP3K7; CREBBP; MAP2K2; CHUK; PDGFRA;
	MAP2K1; NFKB1; JUN; IL1R1; HSP90AA1
Fc Epsilon RI Signaling	PRKCE; RAC1; PRKCZ; LYN; MAPK1; RAC2; PTPN11;
	AKT2; PIK3CA; SYK; PRKCI; PIK3CB; PIK3C3; MAPK8;
	PRKD1; MAPK3; MAPK10; KRAS; MAPK13; PRKCD;
	MAPK9; PIK3C2A; BTK; MAPK14; TNF; RAF1; FYN;
	MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1; AKT3;
	VAV3; PRKCA
G-Protein Coupled Receptor Signaling	PRKCE; RAP1A; RGS16; MAPK1; GNAS; AKT2; IKBKB;
	PIK3CA; CREB1; GNAQ; NFKB2; CAMK2A; PIK3CB;
	PIK3C3; MAPK3; KRAS; RELA; SRC; PIK3C2A; RAF1;
	IKBKG; RELB; FYN; MAP2K2; AKT1; PIK3R1; CHUK;
	PDPK1; STAT3; MAP2K1; NFKB1; BRAF; ATF4; AKT3;
	PRKCA
Inositol Phosphate Metabolism	PRKCE; IRAK1; PRKAA2; EIF2AK2; PTEN; GRK6;
	MAPK1; PLK1; AKT2; PIK3CA; CDK8; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PRKCD; PRKAA1; MAPK9; CDK2;
	PIM1; PIK3C2A; DYRK1A; MAP2K2; PIP5K1A; PIK3R1;
	MAP2K1; PAK3; ATM; TTK; CSNK1A1; BRAF; SGK
PDGF Signaling	EIF2AK2; ELK1; ABL2; MAPK1; PIK3CA; FOS; PIK3CB;
	PIK3C3; MAPK8; CAV1; ABL1; MAPK3; KRAS; SRC;
	PIK3C2A; PDGFRB; RAF1; MAP2K2; JAK1; JAK2;
	PIK3R1; PDGFRA; STAT3; SPHK1; MAP2K1; MYC;
	JUN; CRKL; PRKCA; SRF; STAT1; SPHK2
VEGF Signaling	ACTN4; ROCK1; KDR; FLT1; ROCK2; MAPK1; PGF;

	AKT2; PIK3CA; ARNT; PTK2; BCL2; PIK3CB; PIK3C3;
	BCL2L1; MAPK3; KRAS; HIF1A; NOS3; PIK3C2A; PXN;
	RAF1; MAP2K2; ELAVL1; AKT1; PIK3R1; MAP2K1; SFN;
	VEGFA; AKT3; FOXO1; PRKCA
Natural Killer Cell Signaling	PRKCE; RAC1; PRKCZ; MAPK1; RAC2; PTPN11;
	KIR2DL3; AKT2; PIK3CA; SYK; PRKCI; PIK3CB;
	PIK3C3; PRKD1; MAPK3; KRAS; PRKCD; PTPN6;
	PIK3C2A; LCK; RAF1; FYN; MAP2K2; PAK4; AKT1;
	PIK3R1; MAP2K1; PAK3; AKT3; VAV3; PRKCA
Cell Cycle: G1/S	HDAC4; SMAD3; SUV39H1; HDAC5; CDKN1B; BTRC;
Checkpoint Regulation	ATR; ABL1; E2F1; HDAC2; HDAC7A; RB1; HDAC11;
	HDAC9; CDK2; E2F2; HDAC3; TP53; CDKN1A; CCND1;
	E2F4; ATM; RBL2; SMAD4; CDKN2A; MYC; NRG1;
	GSK3B; RBL1; HDAC6
T Cell Receptor Signaling	RAC1; ELK1; MAPK1; IKBKB; CBL; PIK3CA; FOS;
	NFKB2; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	RELA; PIK3C2A; BTK; LCK; RAF1; IKBKG; RELB; FYN;
	MAP2K2; PIK3R1; CHUK; MAP2K1; NFKB1; ITK; BCL10;
	JUN; VAV3
Death Receptor Signaling	CRADD; HSPB1; BID; BIRC4; TBK1; IKBKB; FADD;
	FAS; NFKB2; BCL2; MAP3K14; MAPK8; RIPK1; CASP8;
	DAXX; TNFRSF10B; RELA; TRAF2; TNF; IKBKG; RELB;
	CASP9; CHUK; APAF1; NFKB1; CASP2; BIRC2; CASP3;
	BIRC3
FGF Signaling	RAC1; FGFR1; MET; MAPKAPK2; MAPK1; PTPN11;
	AKT2; PIK3CA; CREB1; PIK3CB; PIK3C3; MAPK8;
	MAPK3; MAPK13; PTPN6; PIK3C2A; MAPK14; RAF1;
	AKT1; PIK3R1; STAT3; MAP2K1; FGFR4; CRKL; ATF4;
	AKT3; PRKCA; HGF
GM-CSF Signaling	LYN; ELK1; MAPK1; PTPN11; AKT2; PIK3CA; CAMK2A;
	STAT5B; PIK3CB; PIK3C3; GNB2L1; BCL2L1; MAPK3;
	ETS1; KRAS; RUNX1; PIM1; PIK3C2A; RAF1; MAP2K2;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; CCND1; AKT3;
	STAT1
Amyotrophic Lateral Sclerosis Signaling	BID; IGF1; RAC1; BIRC4; PGF; CAPNS1; CAPN2;
	PIK3CA; BCL2; PIK3CB; PIK3C3; BCL2L1; CAPN1;
	PIK3C2A; TP53; CASP9; PIK3R1; RAB5A; CASP1;
	APAF1; VEGFA; BIRC2; BAX; AKT3; CASP3; BIRC3
JAK/Stat Signaling	PTPN1; MAPK1; PTPN11; AKT2; PIK3CA; STAT5B;
	PIK3CB; PIK3C3; MAPK3; KRAS; SOCS1; STAT5A;
	PTPN6; PIK3C2A; RAF1; CDKN1A; MAP2K2; JAK1;
	AKT1; JAK2; PIK3R1; STAT3; MAP2K1; FRAP1; AKT3;
	STAT1

Nicotinate and Nicotinamide	PRKCE; IRAK1; PRKAA2; EIF2AK2; GRK6; MAPK1;
Metabolism	PLK1; AKT2; CDK8; MAPK8; MAPK3; PRKCD; PRKAA1;
	PBEF1; MAPK9; CDK2; PIM1; DYRK1A; MAP2K2;
	MAP2K1; PAK3; NT5E; TTK; CSNK1A1; BRAF; SGK
Chemokine Signaling	CXCR4; ROCK2; MAPK1; PTK2; FOS; CFL1; GNAQ;
	CAMK2A; CXCL12; MAPK8; MAPK3; KRAS; MAPK13;
	RHOA; CCR3; SRC; PPP1CC; MAPK14; NOX1; RAF1;
	MAP2K2; MAP2K1; JUN; CCL2; PRKCA
IL-2 Signaling	ELK1; MAPK1; PTPN11; AKT2; PIK3CA; SYK; FOS;
	STAT5B; PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS;
	SOCS1; STAT5A; PIK3C2A; LCK; RAF1; MAP2K2;
	JAK1; AKT1; PIK3R1; MAP2K1; JUN; AKT3
Synaptic Long Term Depression	PRKCE; IGF1; PRKCZ; PRDX6; LYN; MAPK1; GNAS;
	PRKCI; GNAQ; PPP2R1A; IGF1R; PRKD1; MAPK3;
	KRAS; GRN; PRKCD; NOS3; NOS2A; PPP2CA;
	YWHAZ; RAF1; MAP2K2; PPP2R5C; MAP2K1; PRKCA
Estrogen Receptor Signaling	TAF4B; EP300; CARM1; PCAF; MAPK1; NCOR2;
	SMARCA4; MAPK3; NRIP1; KRAS; SRC; NR3C1;
	HDAC3; PPARGC1A; RBM9; NCOA3; RAF1; CREBBP;
	MAP2K2; NCOA2; MAP2K1; PRKDC; ESR1; ESR2
Protein Ubiquitination Pathway	TRAF6; SMURF1; BIRC4; BRCA1; UCHL1; NEDD4;
	CBL; UBE2I; BTRC; HSPA5; USP7; USP10; FBXW7;
	USP9X; STUB1; USP22; B2M; BIRC2; PARK2; USP8;
	USP1; VHL; HSP90AA1; BIRC3
IL-10 Signaling	TRAF6; CCR1; ELK1; IKBKB; SP1; FOS; NFKB2;
	MAP3K14; MAPK8; MAPK13; RELA; MAPK14; TNF;
	IKBKG; RELB; MAP3K7; JAK1; CHUK; STAT3; NFKB1;
	JUN; IL1R1; IL6
VDR/RXR Activation	PRKCE; EP300; PRKCZ; RXRA; GADD45A; HES1;
	NCOR2; SP1; PRKCI; CDKN1B; PRKD1; PRKCD;
	RUNX2; KLF4; YY1; NCOA3; CDKN1A; NCOA2; SPP1;
	LRP5; CEBPB; FOXO1; PRKCA
TGF-beta Signaling	EP300; SMAD2; SMURF1; MAPK1; SMAD3; SMAD1;
	FOS; MAPK8; MAPK3; KRAS; MAPK9; RUNX2;
	SERPINE1; RAF1; MAP3K7; CREBBP; MAP2K2;
	MAP2K1; TGFBR1; SMAD4; JUN; SMAD5
Toll-like Receptor Signaling	IRAK1; EIF2AK2; MYD88; TRAF6; PPARA; ELK1;
	IKBKB; FOS; NFKB2; MAP3K14; MAPK8; MAPK13;
	RELA; TLR4; MAPK14; IKBKG; RELB; MAP3K7; CHUK;
	NFKB1; TLR2; JUN
p38 MAPK Signaling	HSPB1; IRAK1; TRAF6; MAPKAPK2; ELK1; FADD; FAS;
	CREB1; DDIT3; RPS6KA4; DAXX; MAPK13; TRAF2;
	MAPK14; TNF; MAP3K7; TGFBR1; MYC; ATF4; IL1R1;
	SRE; STAT1

Neurotrophin/TRK Signaling	NTRK2; MAPK1; PTPN11; PIK3CA; CREB1; FOS;
	PIK3CB; PIK3C3; MAPK8; MAPK3; KRAS; PIK3C2A;
	RAF1; MAP2K2; AKT1; PIK3R1; PDPK1; MAP2K1;
	CDC42; JUN; ATF4
FXR/RXR Activation	INS; PPARA; FASN; RXRA; AKT2; SDC1; MAPK8;
	APOB; MAPK10; PPARG; MTTP; MAPK9; PPARGC1A;
	TNF; CREBBP; AKT1; SREBF1; FGFR4; AKT3; FOXO1
Synaptic Long Term Potentiation	PRKCE; RAP1A; EP300; PRKCZ; MAPK1; CREB1;
	PRKCI; GNAQ; CAMK2A; PRKD1; MAPK3; KRAS;
	PRKCD; PPP1CC; RAF1; CREBBP; MAP2K2; MAP2K1;
	ATF4; PRKCA
Calcium Signaling	RAP1A; EP300; HDAC4; MAPK1; HDAC5; CREB1;
	CAMK2A; MYH9; MAPK3; HDAC2; HDAC7A; HDAC11;
	HDAC9; HDAC3; CREBBP; CALR; CAMKK2; ATF4;
	HDAC6
EGF Signaling	ELK1; MAPK1; EGFR; PIK3CA; FOS; PIK3CB; PIK3C3;
	MAPK8; MAPK3; PIK3C2A; RAF1; JAK1; PIK3R1;
	STAT3; MAP2K1; JUN; PRKCA; SRF; STAT1
Hypoxia Signaling in the Cardiovascular System	EDN1; PTEN; EP300; NQO1; UBE2I; CREB1; ARNT;
	HIF1A; SLC2A4; NOS3; TP53; LDHA; AKT1; ATM;
	VEGFA; JUN; ATF4; VHL; HSP90AA1
LPS/IL-1 Mediated Inhibition of RXR Function	IRAK1; MYD88; TRAF6; PPARA; RXRA; ABCA1;
	MAPK8; ALDH1A1; GSTP1; MAPK9; ABCB1; TRAF2;
	TLR4; TNF; MAP3K7; NR1H2; SREBF1; JUN; IL1R1
LXR/RXR Activation	FASN; RXRA; NCOR2; ABCA1; NFKB2; IRF3; RELA;
	NOS2A; TLR4; TNF; RELB; LDLR; NR1H2; NFKB1;
	SREBF1; IL1R1; CCL2; IL6; MMP9
Amyloid Processing	PRKCE; CSNK1E; MAPK1; CAPNS1; AKT2; CAPN2;
	CAPN1; MAPK3; MAPK13; MAPT; MAPK14; AKT1;
	PSEN1; CSNK1A1; GSK3B; AKT3; APP
IL-4 Signaling	AKT2; PIK3CA; PIK3CB; PIK3C3; IRS1; KRAS; SOCS1;
	PTPN6; NR3C1; PIK3C2A; JAK1; AKT1; JAK2; PIK3R1;
	FRAP1; AKT3; RPS6KB1
Cell Cycle: G2/M DNA Damage Checkpoint Regulation	EP300; PCAF; BRCA1; GADD45A; PLK1; BTRC;
	CHEK1; ATR; CHEK2; YWHAZ; TP53; CDKN1A;
	PRKDC; ATM; SFN; CDKN2A
Nitric Oxide Signaling in the Cardiovascular System	KDR; FLT1; PGF; AKT2; PIK3CA; PIK3CB; PIK3C3;
	CAV1; PRKCD; NOS3; PIK3C2A; AKT1; PIK3R1;
	VEGFA; AKT3; HSP90AA1
Purine Metabolism	NME2; SMARCA4; MYH9; RRM2; ADAR; EIF2AK4;
	PKM2; ENTPD1; RAD51; RRM2B; TJP2; RAD51C;
	NT5E; POLD1; NME1
cAMP-mediated	RAP1A; MAPK1; GNAS; CREB1; CAMK2A; MAPK3;

Signaling	
	SRC; RAF1; MAP2K2; STAT3; MAP2K1; BRAF; ATF4
Mitochondrial Dysfunction	SOD2; MAPK8; CASP8; MAPK10; MAPK9; CASP9;
	PARK7; PSEN1; PARK2; APP; CASP3
Notch Signaling	HES1; JAG1; NUMB; NOTCH4; ADAM17; NOTCH2;
	PSEN1; NOTCH3; NOTCH1; DLL4
Endoplasmic Reticulum Stress Pathway	HSPA5; MAPK8; XBP1; TRAF2; ATF6; CASP9; ATF4;
	EIF2AK3; CASP3
Pyrimidine Metabolism	NME2; AICDA; RRM2; EIF2AK4; ENTPD1; RRM2B;
	NT5E; POLD1; NME1
Parkinson's Signaling	UCHL1; MAPK8; MAPK13; MAPK14; CASP9; PARK7;
	PARK2; CASP3
Cardiac & Beta Adrenergic Signaling	GNAS; GNAQ; PPP2R1A; GNB2L1; PPP2CA; PPP1CC;
	PPP2R5C
Glycolysis/Gluconeogenesis	HK2; GCK; GPI; ALDH1A1; PKM2; LDHA; HK1
Interferon Signaling	IRF1; SOCS1; JAK1; JAK2; IFITM1; STAT1; IFIT3
Sonic Hedgehog Signaling	ARRB2; SMO; GLI2; DYRK1A; GLI1; GSK3B; DYRK1B
Glycerophospholipid Metabolism	PLD1; GRN; GPAM; YWHAZ; SPHK1; SPHK2
Phospholipid Degradation	PRDX6; PLD1; GRN; YWHAZ; SPHK1; SPHK2
Tryptophan Metabolism	SIAH2; PRMT5; NEDD4; ALDH1A1; CYP1B1; SIAH1
Lysine Degradation	SUV39H1; EHMT2; NSD1; SETD7; PPP2R5C
Nucleotide Excision Repair Pathway	ERCC5; ERCC4; XPA; XPC; ERCC1
Starch and Sucrose Metabolism	UCHL1; HK2; GCK; GPI; HK1
Aminosugars Metabolism	NQO1; HK2; GCK; HK1
Arachidonic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Circadian Rhythm Signaling	CSNK1E; CREB1; ATF4; NR1D1
Coagulation System	BDKRB1; F2R; SERPINE1; F3
Dopamine Receptor Signaling	PPP2R1A; PPP2CA; PPP1CC; PPP2R5C
Glutathione Metabolism	IDH2; GSTP1; ANPEP; IDH1
Glycerolipid Metabolism	ALDH1A1; GPAM; SPHK1; SPHK2
Linoleic Acid Metabolism	PRDX6; GRN; YWHAZ; CYP1B1
Methionine Metabolism	DNMT1; DNMT3B; AHCY; DNMT3A

Pyruvate Metabolism	GLO1; ALDH1A1; PKM2; LDHA
Arginine and Proline Metabolism	ALDH1A1; NOS3; NOS2A
Eicosanoid Signaling	PRDX6; GRN; YWHAZ
Fructose and Mannose Metabolism	HK2; GCK; HK1
Galactose Metabolism	HK2; GCK; HK1
Stilbene, Coumarine and Lignin Biosynthesis	PRDX6; PRDX1; TYR
Antigen Presentation Pathway	CALR; B2M
Biosynthesis of Steroids	NQO1; DHCR7
Butanoate Metabolism	ALDH1A1; NLGN1
Citrate Cycle	IDH2; IDH1
Fatty Acid Metabolism	ALDH1A1; CYP1B1
Glycerophospholipid Metabolism	PRDX6; CHKA
Histidine Metabolism	PRMT5; ALDH1A1
Inositol Metabolism	ERO1L; APEX1
Metabolism of Xenobiotics by Cytochrome p450	GSTP1; CYP1B1
Methane Metabolism	PRDX6; PRDX1
Phenylalanine Metabolism	PRDX6; PRDX1
Propanoate Metabolism	ALDH1A1; LDHA
Selenoamino Acid Metabolism	PRMT5; AHCY
Sphingolipid Metabolism	SPHK1; SPHK2
Aminophosphonate Metabolism	PRMT5
Androgen and Estrogen Metabolism	PRMT5
Ascorbate and Aldarate Metabolism	ALDH1A1
Bile Acid Biosynthesis	ALDH1A1
Cysteine Metabolism	LDHA
Fatty Acid Biosynthesis	FASN
Glutamate Receptor Signaling	GNB2L1
NRF2-mediated Oxidative Stress Response	PRDX1
Pentose Phosphate Pathway	GPI

Pentose and Glucuronate	UHL1
Interconversions	
Retinol Metabolism	ALDH1A1
Riboflavin Metabolism	TYR
Tyrosine Metabolism	PRMT5, TYR
Ubiquinone Biosynthesis	PRMT5
Valine, Leucine and Isoleucine Degradation	ALDH1A1
Glycine, Serine and Threonine Metabolism	CHKA
Lysine Degradation	ALDH1A1
Pain/Taste	TRPM5; TRPA1
Pain	TRPM7; TRPC5; TRPC6; TRPC1; Cnr1; cnr2; Grk2;
	Trpa1; Pomc; Cgrp; Crf; Pka; Era; Nr2b; TRPM5; Prkaca;
	Prkacb; Prkar1a; Prkar2a
Mitochondrial Function	AIF; CytC; SMAC (Diablo); Aifm-1; Aifm-2
Developmental Neurology	BMP-4; Chordin (Chrd); Noggin (Nog); WNT (Wnt2;
	Wnt2b; Wnt3a; Wnt4; Wnt5a; Wnt6; Wnt7b; Wnt8b;
	Wnt9a; Wnt9b; Wnt10a; Wnt10b; Wnt16); beta-catenin;
	Dkk-1; Frizzled related proteins; Otx-2; Gbx2; FGF-8;
	Reelin; Dab1; unc-86 (Pou4f1 or Brn3a); Numb; Reln

[0417] Embodiments of the applications of the small molecule compounds also include those related to knocking out genes, amplifying genes and repairing particular mutations associated with DNA repeat instability and neurological disorders (Robert D. Wells, Tetsuo Ashizawa, Genetic Instabilities and Neurological Diseases, Second Edition, Academic Press, Oct 13, 2011 – Medical). Specific aspects of tandem repeat sequences have been found to be responsible for more than twenty human diseases (New insights into repeat instability: role of RNA•DNA hybrids. McIvor EI, Polak U, Napierala M. RNA Biol. 2010 Sep-Oct;7(5):551-8). The present effector protein systems may be harnessed to correct these defects of genomic instability.

[0418] Several further aspects of the invention relate to correcting defects associated with a wide range of genetic diseases which are further described on the website of the National Institutes of Health under the topic subsection Genetic Disorders (website at health.nih.gov/topic/GeneticDisorders). The genetic brain diseases may include but are not limited to Adrenoleukodystrophy, Agenesis of the Corpus Callosum, Aicardi Syndrome, Alpers' Disease, Alzheimer's Disease, Barth Syndrome, Batten Disease, CADASIL,

Cerebellar Degeneration, Fabry's Disease, Gerstmann-Straussler-Scheinker Disease, Huntington's Disease and other Triplet Repeat Disorders, Leigh's Disease, Lesch-Nyhan Syndrome, Menkes Disease, Mitochondrial Myopathies and NINDS Colpocephaly. These diseases are further described on the website of the National Institutes of Health under the subsection Genetic Brain Disorders.

SCREENING

[0419] Further disclosed herein include methods for screening, identifying, analyzing, and/or evaluating compounds that modulate (e.g., inhibit) RNA-guided nucleases. In some embodiments, such methods may comprise a combination of biochemical and cellular assays. The methods may be performed for screening, identifying, analyzing, and/or evaluating compounds that modulate (e.g., inhibit) any RNA-guided nucleases, such as RNA-guided nucleases in any CRISPR-Cas system. Examples of Cas proteins include those of Class 1 (e.g., Type I, Type III, and Type IV) and Class 2 (e.g., Type II, Type V, and Type VI) Cas proteins, e.g., Cas9, Cas12 (e.g., Cas12a, Cas12b, Cas12c, Cas12d), Cas13 (e.g., Cas13a, Cas13b, Cas13c, Cas13d), CasX, CasY, Cas14, variants thereof (e.g., mutated forms, truncated forms), homologs thereof, and orthologs thereof.

EGFP Disruption Assay

[0420] In some embodiments, a quantitative human cell-based reporter assay that enables rapid quantitation of targeted nuclease activities is used to characterize off-target cleavage of Cas9-based RNA guided endonucleases. In this assay, the activities of nucleases targeted to a single integrated EGFP reporter gene can be quantified by assessing loss of fluorescence signal in human U2OS.EGFP cells caused by inactivating frameshift insertion/deletion (indel) mutations introduced by error prone non-homologous end-joining (NHEJ) repair of nuclease-induced double-stranded breaks (DSBs).

[0421] In one protocol, U2OS.EGFP cells harboring a single integrated copy of an EGFP-PEST fusion gene are cultured (see e.g., Reyon et al., Nat Biotech 30, 460-465 (2012), which is herein incorporated by reference in its entirety). For transfections, 200,000 cells are nucleofected with gRNA expression plasmid and pJDS246 together with 30 ng of a Td-tomato-encoding plasmid using the SE Cell Line 4D-Nucleofector™ X Kit (Lonza) according to the manufacturer's protocol. Cells are analyzed 2 days post-transfection using a BD LSRII flow cytometer. Transfections for optimizing gRNA/Cas9 plasmid concentration are performed in triplicate and all other transfections are performed in duplicate.

[0422] PCR amplification is used for sequence verification of endogenous human genomic sites. PCR reactions are performed using Phusion Hot Start II high-fidelity DNA polymerase (NEB). Loci are amplified using touchdown PCR (98° C., 10 s; 72-62° C., -1° C./cycle, 15 s; 72° C., 30 s] 10 cycles, [98° C., 10 s; 62° C., 15 s; 72° C., 30 s] 25 cycles). Alternatively, PCR for other targets are performed with 35 cycles at a constant annealing temperature of 68° C. or 72° C. and 3% DMSO or 1M betaine, if necessary. PCR products are analyzed on a QIAxCEL capillary electrophoresis system to verify both size and purity. Validated products are treated with ExoSap-IT (Affymetrix) and sequenced by the Sanger method (MGH DNA Sequencing Core) to verify each target site.

[0423] Small molecule compounds of RNA guided endonucleases can also be used to regulate genome editing technologies in other organisms, including invertebrates, plants, and unicellular organisms (e.g., bacteria). Potential uses include regulating gene drives for entomological and agricultural uses. In addition, it is anticipated that Cas9 inhibitors will be valuable probes to understand the role of Cas9 in CRISPR-mediated bacterial immunity (e.g., spacer acquisition) (Nunez et al., *Nature*. 2015 Mar 12;519(7542):193-8; Heler et al., *Nature* 2015, 519, 199-202). Along similar lines, Cas9 inhibitors can be deployed for directed evolution of Cas9. It is hypothesized that Cas9 inhibitors will disrupt bacterial immunity against bacteriophages (or toxic DNA) by interfering with the CRISPR-Cas9-based immune surveillance system in bacteria. Akin to the development of antibiotic resistance, bacteria will be forced to evolve Cas9 protein.

[0424] Nuclease inhibitors of the invention are useful to inhibit the nuclease activity of a CRISPR protein – guide complex. The guide need not be naturally occurring and can comprise, e.g., non-naturally occurring nucleotides, analogs, and/or chemical modifications.

[0425] The present invention also contemplates use of the systems described herein to provide RNA-guided gene drives, for example in systems analogous to gene drives described in PCT Patent Publication WO 2015/105928. Further reference can be found for instance in Esvelt et al. (eLife 2014; 3:e03401; DOI: 10.7554/eLife.03401.001); Webber et al. (PNAS; 2015; 112(34):10565-10567); DeFrancesco (Nature Biotechnology, 2015, 33(10):1019-1021); DiCarlo et al. (Nature Biotechnology, 2015; 33: 1250-1255); Gantz et al. (PNAS; 2015; 112(49):E6736-E6743). Systems of this kind may for example provide methods for altering eukaryotic germline cells, by introducing into the germline cell a nucleic acid sequence encoding an RNA or DNA-guided DNA or RNA nuclease and one or more guide

RNAs or guide DNAs. The guide RNAs/DNAs may be designed to be complementary to one or more target locations on (genomic) DNA or RNA of the germline cell. The nucleic acid sequence encoding the DNA/RNA guided DNA/RNA nuclease and the nucleic acid sequence encoding the guide RNAs/DNAs may be provided on constructs between flanking sequences, with promoters arranged such that the germline cell may express the nuclease and the guides, together with any desired cargo-encoding sequences that are also situated between the flanking sequences. The flanking sequences will typically include a sequence which is identical to a corresponding sequence on a selected target chromosome, so that the flanking sequences work with the components encoded by the construct to facilitate insertion of the foreign nucleic acid construct sequences into RNA or DNA at a target cut site by mechanisms such as homologous recombination, to render the germline cell homozygous for the foreign nucleic acid sequence. In this way, gene-drive systems are capable of introgressing desired cargo genes throughout a breeding population (Gantz et al., 2015, Highly efficient Cas9-mediated gene drive for population modification of the malaria vector mosquito *Anopheles stephensi*, PNAS 2015, published ahead of print November 23, 2015, doi:10.1073/pnas.1521077112; Esvelt et al., 2014, Concerning DNA- or RNA-guided gene drives for the alteration of wild populations eLife 2014;3:e03401). In select embodiments, target sequences may be selected which have few potential off-target sites in a genome. Targeting multiple sites within a target locus, using multiple guide RNAs, may increase the cutting frequency and hinder the evolution of drive resistant alleles. Truncated guide RNAs may reduce off-target cutting. Paired nickases may be used instead of a single nuclease, to further increase specificity. Gene drive constructs may include cargo sequences encoding transcriptional regulators, for example to activate homologous recombination genes and/or repress non-homologous end-joining. Target sites may be chosen within an essential gene, so that non-homologous end-joining events may cause lethality rather than creating a drive-resistant allele. The gene drive constructs can be engineered to function in a range of hosts at a range of temperatures (Cho et al. 2013, Rapid and Tunable Control of Protein Stability in *Caenorhabditis elegans* Using a Small Molecule, PLoS ONE 8(8): e72393. doi:10.1371/journal.pone.0072393).

Modulation gene editing mechanisms

[0426] The small molecule compound herein may be administered to cells or organisms at doses effective to impact gene editing outcomes, e.g., to control the gene editing mechanisms via NHEJ or HDR.

[0427] The activity of NHEJ and HDR DSB repair varies significantly by cell type and cell state. NHEJ is not highly regulated by the cell cycle and is efficient across cell types, allowing for high levels of gene disruption in accessible target cell populations. In contrast, HDR acts primarily during S/G2 phase, and is therefore restricted to cells that are actively dividing, limiting treatments that require precise genome modifications to mitotic cells [Ciccia, A. & Elledge, S.J. *Molecular cell* 40, 179-204 (2010); Chapman, J.R., et al. *Molecular cell* 47, 497-510 (2012)].

[0428] The small molecule compounds may affect the gene editing mechanisms by modulating the function and activity of the RNA-guided nuclease involved in the gene editing. The efficiency of correction via HDR may be controlled by the epigenetic state or sequence of the targeted locus, or the specific repair template configuration (single vs. double stranded, long vs. short homology arms) used [Hacein-Bey-Abina, S., et al. *The New England journal of medicine* 346, 1185-1193 (2002); Gaspar, H.B., et al. *Lancet* 364, 2181-2187 (2004); Beumer, K.J., et al. *G3* (2013)]. The relative activity of NHEJ and HDR machineries in target cells may also affect gene correction efficiency, as these pathways may compete to resolve DSBs [Beumer, K.J., et al. *Proceedings of the National Academy of Sciences of the United States of America* 105, 19821-19826 (2008)]. HDR also imposes a delivery challenge not seen with NHEJ strategies, as it requires the concurrent delivery of nucleases and repair templates. In practice, these constraints have so far led to low levels of HDR in therapeutically relevant cell types. Clinical translation has therefore largely focused on NHEJ strategies to treat disease, although proof-of-concept preclinical HDR treatments have now been described for mouse models of haemophilia B and hereditary tyrosinemia [Li, H., et al. *Nature* 475, 217-221 (2011); Yin, H., et al. *Nature biotechnology* 32, 551-553 (2014)].

Detecting Nuclease Activity

[0429] The methods detecting nuclease activity, for example in Cas9, are based in part on displacement of a weakly held distal non-target strand (Fig. 1A) that can be displaced upon addition of excess complementary single stranded DNA. Systems designed for use in the

nuclease activity methods comprise fluorescently labeling the 5' end of the non-target strand which can be quenched in a cleavage-dependent manner by adding in excess a complementary DNA strand labeled with a 3' quencher. Upon displacement of the unlabeled strand and annealing of the two labeled strands, fluorescence is quenched by a FRET mechanism, providing a proxy measurement for Cas9 activity at the RuvC domain based on the extent of fluorescence loss. This concept can be applied, for example in the Cas12 systems utilizing 3' end of the non-target strand where the Pam is located 5' of the target sequence as with a Cas12a enzyme. Accordingly, this approach can be extrapolated further for Cas proteins based on PAM location and with an understanding of the cleavage activity of the particular Cas protein complexes utilized.

[0430] In other aspects, the invention provides a method of detecting nuclease activity of an RNA-guided nuclease, and a method of identifying an inhibitor of an RNA-guided nuclease. The invention can be used to detect nuclease activity or identify inhibitors of nuclease activity of class 2 CRISPR systems, including Type II, Type V, and Type VI systems. Accordingly, the method can be used to detect nuclease activity or identify inhibitors of Cas9, including but not limited to SpCas9, SaCas9, StCas9 and other Cas9 orthologs. Similarly, the method can be used to detect nuclease activity or identify inhibitors of Cas12, including but not limited to orthologs of Cas12a, Cas12b, and Cas12c. Moreover, the method can be used to detect nuclease activity or identify inhibitors of Cas13, including but not limited to orthologs of Cas13a, Cas13b, Cas13c, and Cas13d.

[0431] In certain embodiments, the invention provides a method of detecting nuclease activity of an RNA guided endonuclease-guide complex comprising: contacting a) an RNA guided endonuclease-guide complex, b) a double-stranded probe comprising a first oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 5' terminus and a second oligonucleotide complementary to the first oligonucleotide, wherein the second oligonucleotide is designed to hybridize to the guide whereby the first oligonucleotide and the second oligonucleotide are cleaved by the endonuclease, and c) a third oligonucleotide comprising a fluorescence quencher at the 3' terminus and designed to hybridize to and displace the cleaved first oligonucleotide from the cleaved second oligonucleotide, and detecting a decrease in fluorescence polarization relative to a reference, thereby detecting nuclease activity of the RNA guided endonuclease-guide complex. The method is generally

applicable for an RNA guided endonuclease system, wherein the PAM is located 3' to the target sequence along the non-target strand (NTS) of the DNA, for instance Cas9.

[0432] In certain embodiments, the invention provides a method of detecting nuclease activity of an RNA guided endonuclease-guide complex comprising: contacting a) an RNA guided endonuclease-guide complex, b) a double-stranded probe comprising a first oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 3' terminus and a second oligonucleotide complementary to the first oligonucleotide, wherein the second oligonucleotide is designed to hybridize to the guide whereby the first oligonucleotide and the second oligonucleotide are cleaved by the endonuclease, and c) a third oligonucleotide comprising a fluorescence quencher at the 5' terminus and designed to hybridize to and displace the cleaved first oligonucleotide from the cleaved second oligonucleotide, and detecting a decrease in fluorescence polarization relative to a reference, thereby detecting nuclease activity of the RNA guided endonuclease-guide complex.

[0433] The third oligonucleotide may also be referred to herein as a displacer oligonucleotide. In certain such embodiments, the displacer oligonucleotide comprises: 5'-ATAGTGAGTCGTATTA/3IAbRQSp/ (SEQ ID NO:3).

[0434] The method is generally applicable for an RNA guided endonuclease system wherein the PAM is located 5' to the target sequence along the non-target strand (NTS) of the DNA, for instance Cas12a (Cpf1).

[0435] In an aspect, the methods can further comprise contacting the elements i) an RNA guided nuclease-guide complex, ii) a double-stranded probe comprising a first oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 5' terminus or the 3' terminus and a second oligonucleotide complementary to the first oligonucleotide, wherein the second oligonucleotide is designed to hybridize to the guide whereby the first oligonucleotide and the second oligonucleotide are cleaved by the endonuclease, and iii) a third oligonucleotide comprising a fluorescence quencher, the fluorescent quencher located at the 3' terminus when the fluorescent reporter is at the 5' terminus of the first oligonucleotide, or at the 5' terminus when the fluorescent reporter is at the 3' terminus of the first oligonucleotide, and designed to hybridize to and displace the cleaved first oligonucleotide from the cleaved second oligonucleotide with an agent. The step of detecting the decrease in the reduction of fluorescence polarization relative to the reference can thereby identify the agent as an inhibitor of the nuclease activity of the RNA guided endonuclease-guide

complex. The quantitation of the reduction of fluorescence polarization can be indicative of the amount of inhibition by an agent, and therefore, its relative level of effectiveness as an inhibitor. Thus, the use of an agent in the detection methods can further allow for screening of inhibitors.

[0436] The invention provides methods of identifying an agent that inhibits the nuclease activity of an RNA guided endonuclease-guide complex. The methods may comprise one or more primary assays. The primary assays may be biochemical assays that assess the binding of the RNA-guided endonuclease with a target DNA. In some embodiments, the primary assay may be a fluorescence Polarization-based Assays. The fluorescence Polarization-based Assay may monitor the change in the fluorescence polarization of the fluorophore-labelled PAM-rich target DNA (e.g., a 12PAM-DNA) upon binding to [Cas9:guideRNA] complex. In this assay, the complexation of [Cas9:guideRNA] to PAM-rich target DNA shows a dose-dependent increase in fluorophore polarization.

[0437] Fluorescence polarization is a useful technique to monitor the interaction between two molecules, including for example, Cas9-gRNA (ribonucleoprotein) complex and target DNA (e.g., 12PAM). The methods may further comprise one or more secondary assays. The secondary assays may be cell-based assays for testing the effect of candidate compounds on RNA-guided endonuclease activity in cells.

[0438] In some embodiments, the secondary assay may be an EGFP disruption assay. In such assay, a quantitative human cell-based reporter assay that enables rapid quantitation of targeted nuclease activities is used to characterize off-target cleavage of Cas9-based RNA guided endonucleases. In this assay, the activities of nucleases targeted to a single integrated EGFP reporter gene can be quantified by assessing loss of fluorescence signal in human U2OS.EGFP cells caused by inactivating frameshift insertion/deletion (indel) mutations introduced by error prone non-homologous end-joining (NHEJ) repair of nuclease-induced double-stranded breaks (DSBs).

[0439] In certain embodiments, the invention provides a method of identifying an agent that inhibits nuclease activity of an RNA guided endonuclease-guide complex comprising: contacting an RNA guided endonuclease-guide complex with an agent in the presence of a double-stranded probe comprising a first oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 5' terminus and a second oligonucleotide complementary to the first oligonucleotide, wherein the second oligonucleotide is designed to hybridize to the

guide whereby the first oligonucleotide and the second oligonucleotide are cleaved by the nuclease, and a third oligonucleotide comprising a fluorescence quencher at the 3' terminus and designed to hybridize to and displace the cleaved first oligonucleotide from the cleaved second oligonucleotide, and detecting a decrease in the reduction of fluorescence polarization relative to a reference, thereby identifying the agent as an inhibitor of the nuclease activity of the RNA guided endonuclease-guide complex.

[0440] The method is particularly useful for identifying an inhibitor of an RNA guided endonuclease system such as Cas9 wherein the PAM is located 3' to the target sequence along the non-target strand (NTS) of the DNA.

[0441] In certain embodiments, the invention provides a method of identifying an agent that inhibits nuclease activity of an RNA guided endonuclease-guide complex comprising: contacting the RNA guided endonuclease-guide complex with an agent in the presence of a double-stranded probe comprising a first oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 3' terminus and a second oligonucleotide complementary to the first oligonucleotide, wherein the second oligonucleotide is designed to hybridize to the guide whereby the first oligonucleotide and the second oligonucleotide are cleaved by the nuclease, and a third oligonucleotide comprising a fluorescence quencher at the 5' terminus and designed to hybridize to and displace the cleaved first oligonucleotide from the cleaved second oligonucleotide, and detecting a decrease in the reduction of fluorescence polarization relative to a reference, thereby identifying the agent as an inhibitor of the nuclease activity of the RNA guided endonuclease-guide complex.

[0442] The method is particularly useful for identifying an inhibitor of an RNA guided endonuclease system such as Cas12a (Cpf1) wherein the PAM is located 5' to the target sequence along the non-target strand (NTS) of the DNA.

Fluorescence Polarization-based Assays

[0443] In one aspect, the invention provides an assay designed to identify inhibitors of CRISPR nuclease activity. The assay monitors the change in the fluorescence polarization of the fluorophore-labelled double-stranded PAM-containing substrate DNA (DS-Fluor) upon binding to a CRISPR:guideRNA complex. Fluorescence polarization is a useful technique to monitor the interaction between two molecules. In this assay, the complexation of CRISPR:guideRNA complex to the double-stranded PAM-containing substrate DNA results in cleavage of the substrate DNA. Upon cleavage, the length of the double-stranded region of

the substrate DNA is reduced. The fluorophore-bearing first strand can then be displaced from the complementary strand by a displacer strand bearing a quencher moiety (Disp-Q). (see, e.g., Fig. 1).

[0444] The FP-based method was shown to be PAM dependent and can be optimized by adjusting, e.g., the DS-Fluor:CRISPR ratio, the DS-Fluor:Disp-Q ratio, incubation time, and other parameters. The method was also shown to be sensitive to inhibition of SpCas9 by AcrIIA4.

[0445] The FP-based method was then performed to screen 50,000 compounds consisting of structurally diverse scaffolds with vast functional variability against SaCas9. DMSO (the solvent) was used as the negative control and ApoSaCas9 nuclease was used as a positive control. Interestingly, a portion of the hit compounds were found to have similarity in their molecular scaffold with variation in the stereochemistry and functionality. Compounds were examined for dose dependence on activity and showed dose dependent Cas9 inhibition activity with low IC₅₀ values. These finding further confirms and validate the primary assay. The FP-based method is generally useful to screen for CRISPR nuclease inhibitor compounds. In particular, the method has also been used to screen for inhibitors of SaCas9 and FnCas9

[0446] According to the invention, the method can be optimized by adjusting absolute and relative reagent concentrations, probe lengths, probe sequences and base-pairing capacity, time course, and other parameters.

[0447] The relative concentration of DS-Fluor to CRISPR protein is adjustable and can be, without limitation, 5:1, 2:1, 1:1, 1:2, 1:5, 1:10, 1:20, 1:50, or from 5:1 to 2:1, or from 2:1 to 1:1, or from 1:1 to 1:2, or from 1:2 to 1:5, or from 1:5 to 1:10, or from 1:10 to 1:20, or from 1:20 to 1:50.

[0448] The relative concentration of DS-Fluor to Disp-Q quencher is adjustable and can be, without limitation, 5:1, 2:1, 1:1, 1:2, 1:5, 1:10, 1:20, 1:50, or from 5:1 to 2:1, or from 2:1 to 1:1, or from 1:1 to 1:2, or from 1:2 to 1:5, or from 1:5 to 1:10, or from 1:10 to 1:20, or from 1:20 to 1:50.

[0449] The reaction time is adjustable and can be, without limitation 5 min, 10 min, 15 min, 20 min, 30 min, 45 min, 60 min, 90 min, 120 min, 150 min, 180 min, 210 min, 240 min, or longer or from 5 to 10 min., from 10 to 15 min., from 15 to 20 min, from 20 to 30 min,

from 30 to 45 min, from 45 to 60 min, from 60 to 90 min, from 90 to 120 min, from 120 to 150 min, from 150 to 180 min, from 180 to 210 min, from 210 to 240 min.

[0450] In various embodiments such as exemplified herein, the double-stranded probe, i.e., the labeled strand of the double-stranded probe comprises an oligonucleotide comprising: 5Alex647N/TAATACGACTCACTATAGGACGCGACCGAAATGGTGAAGGACGGGT-3' (SEQ ID NO:4).

[0451] 5Alex647N indicates linkage of an Alexa Fluor® 647 fluorophore. In other embodiments, other fluorophores can be employed.

[0452] In certain such embodiments, the complementary oligonucleotide comprises: 5'-ACCCGTCCTTCACCATTTTCGGTCGCGTCCTATAGTGAGTCGTATTA-3' (SEQ ID NO:5).

[0453] In certain such embodiments, the displacer oligonucleotide comprises: 5'-ATAGTGAGTCGTATTA/3IAbRQSp/ (SEQ ID NO:6).

[0454] 3IAbRQSp indicates linkage of a RedQ quencher. In other embodiments, other quenchers can be employed, see, e.g. Table 2.

[0455] In the above sequences, the PAM is bold, protospacer italicized, and displacer binding site underlined.

[0456] The following table provides additional fluorophores useful for the invention.

Table 1 - Fluorophore labels for fluorescent hybridization probes

Fluorophore	Alternative Fluorophore	Excitation (nm)	Emission (nm)
FAM		495	515
TET	CAL Fluor Gold 540	525	540
HEX	JOE, VIC, CAL Fluor Orange 560	535	555
Cy3	NED, Quasar 570, Oyster 556	550	570
TMR	CAL Fluor Red 590	555	575
ROX	LC red 610, CAL Fluor Red 610	575	605
Texas red	LC red 610, CAL Fluor Red 610	585	605
LC red 640	CAL Fluor Red 635	625	640

Cy5	LC red 670, Quasar 670, Oyster 645	650	670
LC red 705	Cy5.5	680	710

CAL and Quasar fluorophores are available from Biosearch Technologies; VIC and NED are available from Applied Biosystems; Cy dyes are available from Amersham Biosciences; Oyster fluorophores are available from Integrated DNA Technologies; and LC (Light Cycler) fluorophores are available from Roche Applied Science.

[0457] The following table provides additional quenchers useful for the invention.

Table 2 - Quencher labels for fluorescent hybridization probes

Quencher	Absorption Maximum (nm)
DDQ-I	430
Dabcyl	475
Eclipse	530
Iowa Black FQ	532
BHQ-1	534
QSY-7	571
BHQ-2	580
DDQ-II	630
Iowa Black RQ	645
QSY-21	660
BHQ-3	670

DDQ or Deep Dark Quenchers are available from Eurogentec; Eclipse quenchers are available from Epoch Biosciences; Iowa quenchers are available from Integrated DNA Technologies; BHQ or Black Hole quenchers are available from Biosearch Technologies; and QSY quenchers are available from Molecular Probes.

Spinach Transcription Assay

[0458] In one aspect, the invention provides a transcription assay to detect the activity of an RNA guided endonuclease. In various embodiments, the transcription assay involves expression of a nucleic acid aptamer that binds a molecular fluorophore to generate a

fluorescent signal. Such aptamer-fluorophore combinations are known in the art, including for example, the Spinach aptamer having the sequence

[0459] 5'-GGGAGACGCAACUGAAUGAAAUGGUGAAGGACGGGUCCAGGUGUGGCUGCUUCGGCAGUGCAGCUUGUUGAGUAGAGUGUGAGCUCCGCGUAACUAGUCGCGUCAC-3' (SEQ ID NO:11) and the fluorophore 4-(3,5-difluoro-4-hydroxybenzylidene)-1,2-dimethyl-1*H*-imidazol-5-one (DFHBI) (see, e.g., US2012/0252699 and US2014/0220560, each of which is incorporated herein in their entirety). In the spinach assay, Cas9 can cleave the DNA template and thus inhibits the *in vitro* transcription of the nucleic acid aptamer. Suppression in the transcription results in the reduction of RNA aptamer-fluorophore concentration and hence in the fluorescence signal. *In vitro* transcription reactions may comprise a purified linear DNA template containing a promoter operatively linked to a nucleic acid sequence encoding an RNA aptamer, ribonucleotide triphosphates, a buffer system (e.g., including DTT and magnesium ions, and an appropriate phage RNA polymerase (e.g., T7 polymerase).

SURVEYOR Nuclease Assay

[0460] In various embodiments, SURVEYOR nuclease assay is used to assess genome modification (see e.g., US20150356239, which is herein incorporated by reference in its entirety). In one protocol, 293FT cells are transfected with plasmid DNA. Cells were incubated at 37°C for 72 hours post-transfection prior to genomic DNA extraction. Genomic DNA is extracted using the QuickExtract DNA Extraction Solution (Epicentre) following the manufacturer's protocol. Briefly, pelleted cells are resuspended in QuickExtract solution and incubated at 65°C for 15 minutes and 98°C for 10 minutes.

[0461] The genomic region flanking the CRISPR target site for each gene is PCR amplified, and products are purified using QiaQuick Spin Column (Qiagen) following the manufacturer's protocol. 400ng total of the purified PCR products are mixed with 2µl 10X Taq DNA Polymerase PCR buffer (Enzytsaties) and ultrapure water to a final volume of 20µl, and subjected to a re-annealing process to enable heteroduplex formation: 95°C for 10 min, 95°C to 85°C ramping at - 2°C/s, 85°C to 25°C at - 0.25°C/s, and 25°C hold for 1 minute. After re-annealing, products are treated with SURVEYOR nuclease and SURVEYOR enhancer S (Transgenomics) following the manufacturer's recommended protocol, and analyzed on 4-20% Novex TBE poly-acrylamide gels (Life Technologies). Gels re stained with SYBR Gold DNA stain (Life Technologies) for 30 minutes and imaged

with a Gel Doc gel imaging system (Bio-rad). Quantification is based on relative band intensities.

Test Compounds and Extracts

[0462] In general, small molecule compounds are known in the art or are identified from large libraries of both natural product or synthetic (or semi-synthetic) extracts or chemical libraries or from polypeptide or nucleic acid libraries, according to methods known in the art. Those skilled in the field of drug discovery and development will understand that the precise source of test extracts or compounds is not critical to the screening procedure(s) of the invention. Compounds used in screens may include known compounds (for example, known therapeutics used for other diseases or disorders). Alternatively, virtually any number of unknown chemical extracts or compounds can be screened using the methods described herein. Examples of such extracts or compounds include, but are not limited to, plant-, fungal-, prokaryotic- or animal-based extracts, fermentation broths, and synthetic compounds, as well as modification of existing compounds.

[0463] Numerous methods are also available for generating random or directed synthesis (e.g., semi-synthesis or total synthesis) of any number of chemical compounds, including, but not limited to, saccharide-, lipid-, peptide-, and nucleic acid-based compounds. Synthetic compound libraries are commercially available from Brandon Associates (Merrimack, N.H.) and Aldrich Chemical (Milwaukee, Wis.). Alternatively, chemical compounds to be used as candidate compounds can be synthesized from readily available starting materials using standard synthetic techniques and methodologies known to those of ordinary skill in the art. For example, a library of 8,000 novel small molecules is available, which was created using combinatorial methods of Diversity-Oriented Synthesis (DOS) (Comer et al, Proc Natl Acad Sci U S A 108, 6751 (Apr 26, 2011 ; Lowe et al, J Org Chem 77, 7187 (Sep 7, 2012); Marcaurelle et al, J Am Chem Soc 132, 16962 (Dec 1, 2010))— to investigate chemical compounds not represented in traditional pharmaceutical libraries (Schreiber, S. L. (2000). Science 287, 1964-1969; Schreiber et al, Nat Biotechnol 28, 904 (Sep, 2010), each of which is herein incorporated by reference in their entirety). Synthetic chemistry transformations and protecting group methodologies (protection and deprotection) useful in synthesizing the compounds identified by the methods described herein are known in the art and include, for example, those such as described in R. Larock, Comprehensive Organic Transformations, VCH Publishers (1989); T. W. Greene and P. G. M. Wuts, Protective Groups in Organic

Synthesis, 2nd ed., John Wiley and Sons (1991); L. Fieser and M. Fieser, Fieser and Fieser's Reagents for Organic Synthesis, John Wiley and Sons (1994); and L. Paquette, ed., Encyclopedia of Reagents for Organic Synthesis, John Wiley and Sons (1995), and subsequent editions thereof.

[0464] Alternatively, libraries of natural compounds in the form of bacterial, fungal, plant, and animal extracts are commercially available from a number of sources, including Biotics (Sussex, UK), Xenova (Slough, UK), Harbor Branch Oceanographics Institute (Ft. Pierce, Fla.), and PharmaMar, U.S.A. (Cambridge, Mass.). In addition, natural and synthetically produced libraries are produced, if desired, according to methods known in the art, e.g., by standard extraction and fractionation methods. Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt et al., Proc. Natl. Acad. Sci. U.S.A.90:6909, 1993; Erb et al., Proc. Natl. Acad. Sci. USA 91:11422, 1994; Zuckermann et al., J. Med. Chem.37:2678, 1994; Cho et al., Science 261:1303, 1993; Carrell et al., Angew. Chem. Int. Ed. Engl.33:2059, 1994; Carell et al., Angew. Chem. Int. Ed. Engl. 33:2061, 1994; and Gallop et al., J. Med. Chem.37:1233, 1994. Furthermore, if desired, any library or compound is readily modified using standard chemical, physical, or biochemical methods.

[0465] Libraries of compounds may be presented in solution (e.g., Houghten, Biotechniques 13:412-421, 1992), or on beads (Lam, Nature 354:82-84, 1991), chips (Fodor, Nature 364:555-556, 1993), bacteria (Ladner, U.S. Patent No.5,223,409), spores (Ladner U.S. Patent No.5,223,409), plasmids (Cull et al., Proc Natl Acad Sci USA 89:1865-1869, 1992) or on phage (Scott and Smith, Science 249:386-390, 1990; Devlin, Science 249:404-406, 1990; Cwirla et al. Proc. Natl. Acad. Sci.87:63786382, 1990; Felici, J. Mol. Biol.222:301-310, 1991; Ladner supra.).

[0466] In addition, those skilled in the art of drug discovery and development readily understand that methods for dereplication (e.g., taxonomic dereplication, biological dereplication, and chemical dereplication, or any combination thereof) or the elimination of replicates or repeats of materials already known for their activity should be employed whenever possible.

[0467] When a crude extract is identified as containing a compound of interest, further fractionation of the positive lead extract is necessary to isolate chemical constituents responsible for the observed effect. Thus, the goal of the extraction, fractionation, and

purification process is the careful characterization and identification of a chemical entity within the crude extract that achieves a desired biological effect. Methods of fractionation and purification of such heterogenous extracts are known in the art.

[0468] Small molecules of the invention preferably have a molecular weight below 2,000 daltons, more preferably between 300 and 1,000 daltons, and most preferably between 400 and 700 daltons. It is preferred that these small molecules are organic molecules.

Formulations

[0469] Agents described herein, including analogs thereof, and/or agents discovered to have medicinal value using the methods described herein are useful as Cas nuclease inhibitors. For therapeutic uses, the compositions or agents identified using the methods disclosed herein may be administered systemically, for example, formulated in a pharmaceutically-acceptable buffer such as physiological saline. Preferable routes of administration include, for example, subcutaneous, intravenous, interperitoneally, intramuscular, or intradermal injections that provide continuous, sustained levels of the drug in the patient. Treatment of human patients or other animals will be carried out using a therapeutically effective amount of a therapeutic identified herein in a physiologically-acceptable carrier. Suitable carriers and their formulation are described, for example, in Remington's Pharmaceutical Sciences by E. W. Martin. The amount of the therapeutic agent to be administered varies depending upon the manner of administration, the age and body weight of the patient, and with the clinical symptoms. Generally, amounts will be in the range of those used for other agents used in the treatment of other diseases associated with diabetes.

[0470] The disclosed compounds may be administered alone (e.g., in saline or buffer) or using any delivery vehicles known in the art. For instance, the following delivery vehicles have been described: Cochleates; Emulsomes, ISCOMs; Liposomes; Live bacterial vectors (e.g., Salmonella, Escherichia coli, Bacillus calmatte-guerin, Shigella, Lactobacillus); Live viral vectors (e.g., Vaccinia, adenovirus, Herpes Simplex); Microspheres; Nucleic acid vaccines; Polymers; Polymer rings; Proteosomes; Sodium Fluoride; Transgenic plants; Virosomes; Virus-like particles. Other delivery vehicles are known in the art and some additional examples are provided below.

[0471] The disclosed compounds may be administered by any route known, such as, for example, orally, transdermally, intravenously, cutaneously, subcutaneously, nasally, intramuscularly, intraperitoneally, intracranially, and intracerebroventricularly.

[0472] In certain embodiments, disclosed compounds are administered at dosage levels greater than about 0.001 mg/kg, such as greater than about 0.01 mg/kg or greater than about 0.1 mg/kg. For example, the dosage level may be from about 0.001 mg/kg to about 50 mg/kg such as from about 0.01 mg/kg to about 25 mg/kg, from about 0.1 mg/kg to about 10 mg/kg, or from about 1 mg/kg to about 5 mg/kg of subject body weight per day, one or more times a day, to obtain the desired therapeutic effect. It will also be appreciated that dosages smaller than about 0.001 mg/kg or greater than about 50 mg/kg (for example about 50-100 mg/kg) can also be administered to a subject.

[0473] In one embodiment, the compound is administered once-daily, twice-daily, or three-times daily. In one embodiment, the compound is administered continuously (i.e., every day) or intermittently (e.g., 3-5 days a week). In another embodiment, administration could be on an intermittent schedule.

[0474] Further, administration less frequently than daily, such as, for example, every other day may be chosen. In additional embodiments, administration with at least 2 days between doses may be chosen. By way of example only, dosing may be every third day, bi-weekly or weekly. As another example, a single, acute dose may be administered. Alternatively, compounds can be administered on a non-regular basis e.g., whenever symptoms begin. For any compound described herein the effective amount can be initially determined from animal models.

[0475] Toxicity and efficacy of the compounds can be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., for determining the LD₅₀ (the dose lethal to 50% of the population) and the ED₅₀ (the dose therapeutically effective in 50% of the population). The dose ratio between toxic and therapeutic effects is the therapeutic index and it can be expressed as the ratio LD₅₀/ED₅₀. Compounds that exhibit large therapeutic indices may have a greater effect when practicing the methods as disclosed herein. While compounds that exhibit toxic side effects may be used, care should be taken to design a delivery system that targets such compounds to the site of affected tissue in order to minimize potential damage to uninfected cells and, thereby, reduce side effects.

[0476] Data obtained from the cell culture assays and animal studies can be used in formulating a range of dosage of the compounds disclosed herein for use in humans. The dosage of such agents lies within a range of circulating concentrations that include the ED₅₀ with little or no toxicity. The dosage may vary within this range depending upon the dosage

form employed and the route of administration utilized. For any compound used in the disclosed methods, the effective dose can be estimated initially from cell culture assays. A dose may be formulated in animal models to achieve a circulating plasma concentration range that includes the IC_{50} (i.e., the concentration of the test compound that achieves a half-maximal inhibition of symptoms) as determined in cell culture. Such information can be used to more accurately determine useful doses in humans. Levels in plasma may be measured, for example, by high performance liquid chromatography. In certain embodiments, pharmaceutical compositions may comprise, for example, at least about 0.1% of an active compound. In other embodiments, the active compound may comprise between about 2% to about 75% of the weight of the unit, or between about 25% to about 60%, for example, and any range derivable therein. Multiple doses of the compounds are also contemplated.

[0477] The formulations disclosed herein are administered in pharmaceutically acceptable solutions, which may routinely contain pharmaceutically acceptable concentrations of salt, buffering agents, preservatives, compatible carriers, and optionally other therapeutic ingredients.

[0478] For use in therapy, an effective amount of one or more disclosed compounds can be administered to a subject by any mode that delivers the compound(s) to the desired surface, e.g., mucosal, systemic. Administering the pharmaceutical composition of the present disclosure may be accomplished by any means known to the skilled artisan. Disclosed compounds may be administered orally, transdermally, intravenously, cutaneously, subcutaneously, nasally, intramuscularly, intraperitoneally, intracranially, or intracerebroventricularly.

[0479] For oral administration, one or more compounds can be formulated readily by combining the active compound(s) with pharmaceutically acceptable carriers well known in the art. Such carriers enable the compounds to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions and the like, for oral ingestion by a subject to be treated.

[0480] Pharmaceutical preparations for oral use can be obtained as solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are, in particular, fillers such as sugars, including lactose, sucrose, mannitol, or sorbitol; cellulose preparations such as, for example, maize starch, wheat starch, rice starch, potato starch,

gelatin, gum tragacanth, methyl cellulose, hydroxypropylmethyl-cellulose, sodium carboxymethylcellulose, and/or polyvinylpyrrolidone (PVP). If desired, disintegrating agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, or alginic acid or a salt thereof such as sodium alginate. Optionally the oral formulations may also be formulated in saline or buffers, i.e. EDTA for neutralizing internal acid conditions or may be administered without any carriers.

[0481] Also specifically contemplated are oral dosage forms of one or more disclosed compounds. The compound(s) may be chemically modified so that oral delivery of the derivative is efficacious. Generally, the chemical modification contemplated is the attachment of at least one moiety to the compound itself, where said moiety permits (a) inhibition of proteolysis; and (b) uptake into the blood stream from the stomach or intestine. Also desired is the increase in overall stability of the compound(s) and increase in circulation time in the body. Examples of such moieties include: polyethylene glycol, copolymers of ethylene glycol and propylene glycol, carboxymethyl cellulose, dextran, polyvinyl alcohol, polyvinyl pyrrolidone and polyproline. Other polymers that could be used are poly-1,3-dioxolane and poly-1,3,6-tioxocane. In some aspects for pharmaceutical usage, as indicated above, are polyethylene glycol moieties.

[0482] The location of release may be the stomach, the small intestine (the duodenum, the jejunum, or the ileum), or the large intestine. One skilled in the art has available formulations which will not dissolve in the stomach, yet will release the material in the duodenum or elsewhere in the intestine. In some aspects, the release will avoid the deleterious effects of the stomach environment, either by protection of the compound or by release of the biologically active material beyond the stomach environment, such as in the intestine.

[0483] To ensure full gastric resistance a coating impermeable to at least pH 5.0 is important. Examples of the more common inert ingredients that are used as enteric coatings are cellulose acetate trimellitate (CAT), hydroxypropylmethylcellulose phthalate (HPMCP), HPMCP 50, HPMCP 55, polyvinyl acetate phthalate (PVAP), Eudragit L30D, Aquateric, cellulose acetate phthalate (CAP), Eudragit L, Eudragit S, and Shellac. These coatings may be used as mixed films.

[0484] A coating or mixture of coatings can also be used on tablets, which are not intended for protection against the stomach. This can include sugar coatings, or coatings

which make the tablet easier to swallow. Capsules may consist of a hard shell (such as gelatin) for delivery of dry therapeutic i.e. powder; for liquid forms, a soft gelatin shell may be used. The shell material of cachets could be thick starch or other edible paper. For pills, lozenges, molded tablets or tablet triturates, moist massing techniques can be used.

[0485] The disclosed compounds can be included in the formulation as fine multiparticulates in the form of granules or pellets of particle size about 1 mm. The formulation of the material for capsule administration could also be as a powder, lightly compressed plugs or even as tablets. The compound could be prepared by compression.

[0486] Colorants and flavoring agents may all be included. For example, the compound may be formulated (such as by liposome or microsphere encapsulation) and then further contained within an edible product, such as a refrigerated beverage containing colorants and flavoring agents.

[0487] One may dilute or increase the volume of compound delivered with an inert material. These diluents could include carbohydrates, especially mannitol, α -lactose, anhydrous lactose, cellulose, sucrose, modified dextrans and starch. Certain inorganic salts may be also be used as fillers including calcium triphosphate, magnesium carbonate and sodium chloride. Some commercially available diluents are Fast-Flo, Emdex, STA-Rx 1500, Emcompress and Avicell. Disintegrants may be included in the formulation of the therapeutic into a solid dosage form. Materials used as disintegrates include but are not limited to starch, including the commercial disintegrant based on starch, Explotab. Sodium starch glycolate, Amberlite, sodium carboxymethylcellulose, ultramylopectin, sodium alginate, gelatin, orange peel, acid carboxymethyl cellulose, natural sponge and bentonite may all be used. Another form of the disintegrants is the insoluble cationic exchange resins. Powdered gums may be used as disintegrants and as binders and these can include powdered gums such as agar, Karaya or tragacanth. Alginic acid and its sodium salt are also useful as disintegrants.

[0488] Binders may be used to hold the therapeutic together to form a hard tablet and include materials from natural products such as acacia, tragacanth, starch and gelatin. Others include methyl cellulose (MC), ethyl cellulose (EC) and carboxymethyl cellulose (CMC). Polyvinyl pyrrolidone (PVP) and hydroxypropylmethyl cellulose (HPMC) could both be used in alcoholic solutions to granulate the therapeutic.

[0489] An anti-frictional agent may be included in the formulation of the compound to prevent sticking during the formulation process. Lubricants may be used as a layer between

the compound and the die wall, and these can include but are not limited to; stearic acid including its magnesium and calcium salts, polytetrafluoroethylene (PTFE), liquid paraffin, vegetable oils and waxes. Soluble lubricants may also be used such as sodium lauryl sulfate, magnesium lauryl sulfate, polyethylene glycol of various molecular weights, Carbowax 4000 and 6000. Glidants that might improve the flow properties of the drug during formulation and to aid rearrangement during compression might be added. The glidants may include starch, talc, pyrogenic silica and hydrated silicoaluminate.

[0490] To aid dissolution of the compound into the aqueous environment a surfactant might be added as a wetting agent. Surfactants may include anionic detergents such as sodium lauryl sulfate, dioctyl sodium sulfosuccinate and dioctyl sodium sulfonate. Cationic detergents might be used and could include benzalkonium chloride or benzethonium chloride. The list of potential non-ionic detergents that could be included in the formulation as surfactants are lauromacrogol 400, polyoxyl 40 stearate, polyoxyethylene hydrogenated castor oil 10, 50 and 60, glycerol monostearate, polysorbate 40, 60, 65 and 80, sucrose fatty acid ester, methyl cellulose and carboxymethyl cellulose. These surfactants could be present in the formulation of the compound either alone or as a mixture in different ratios.

[0491] Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a plasticizer, such as glycerol or sorbitol. The push-fit capsules can contain the active ingredients in admixture with filler such as lactose, binders such as starches, and/or lubricants such as talc or magnesium stearate and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid paraffin, or liquid polyethylene glycols. In addition, stabilizers may be added. Microspheres formulated for oral administration may also be used. Such microspheres have been well defined in the art. All formulations for oral administration should be in dosages suitable for such administration.

[0492] For buccal administration, the compositions may take the form of tablets or lozenges formulated in conventional manner.

[0493] For administration by inhalation, the compounds for use according to the present disclosure may be conveniently delivered in the form of an aerosol spray presentation from pressurized packs or a nebulizer, with the use of a suitable propellant, e.g., dichlorodifluoromethane, trichlorofluoromethane, dichlorotetrafluoroethane, carbon dioxide or other suitable gas. In the case of a pressurized aerosol the dosage unit may be determined

by providing a valve to deliver a metered amount. Capsules and cartridges of e.g. gelatin for use in an inhaler or insufflator may be formulated containing a powder mix of the compound and a suitable powder base such as lactose or starch.

[0494] Also contemplated herein is pulmonary delivery of the compounds of the disclosure. The compound is delivered to the lungs of a mammal while inhaling and traverses across the lung epithelial lining to the blood stream using methods well known in the art.

[0495] Contemplated for use in the practice of methods disclosed herein are a wide range of mechanical devices designed for pulmonary delivery of therapeutic products, including but not limited to nebulizers, metered dose inhalers, and powder inhalers, all of which are familiar to those skilled in the art. Some specific examples of commercially available devices suitable for the practice of these methods are the Ultravent nebulizer, manufactured by Mallinckrodt, Inc., St. Louis, Missouri; the Acorn II nebulizer, manufactured by Marquest Medical Products, Englewood, Colorado; the Ventolin metered dose inhaler, manufactured by Glaxo Inc., Research Triangle Park, North Carolina; and the Spinhaler powder inhaler, manufactured by Fisons Corp., Bedford, Massachusetts.

[0496] All such devices require the use of formulations suitable for the dispensing of compound. Typically, each formulation is specific to the type of device employed and may involve the use of an appropriate propellant material, in addition to the usual diluents, and/or carriers useful in therapy. Also, the use of liposomes, microcapsules or microspheres, inclusion complexes, or other types of carriers is contemplated. Chemically modified compound may also be prepared in different formulations depending on the type of chemical modification or the type of device employed. Formulations suitable for use with a nebulizer, either jet or ultrasonic, will typically comprise compound dissolved in water at a concentration of about 0.1 to about 25 mg of biologically active compound per mL of solution. The formulation may also include a buffer and a simple sugar (e.g., for stabilization and regulation of osmotic pressure). The nebulizer formulation may also contain a surfactant, to reduce or prevent surface induced aggregation of the compound caused by atomization of the solution in forming the aerosol.

[0497] Formulations for use with a metered-dose inhaler device will generally comprise a finely divided powder containing the compound suspended in a propellant with the aid of a surfactant. The propellant may be any conventional material employed for this purpose, such as a chlorofluorocarbon, a hydrochlorofluorocarbon, a hydrofluorocarbon, or a hydrocarbon,

including trichlorofluoromethane, dichlorodifluoromethane, dichlorotetrafluoroethanol, and 1,1,1,2-tetrafluoroethane, or combinations thereof. Suitable surfactants include sorbitan trioleate and soya lecithin. Oleic acid may also be useful as a surfactant.

[0498] Formulations for dispensing from a powder inhaler device will comprise a finely divided dry powder containing compound and may also include a bulking agent, such as lactose, sorbitol, sucrose, or mannitol in amounts which facilitate dispersal of the powder from the device, e.g., about 50 to about 90% by weight of the formulation. The compound should most advantageously be prepared in particulate form with an average particle size of less than 10 mm (or microns), such as about 0.5 to about 5 mm, for an effective delivery to the distal lung.

[0499] Nasal delivery of a disclosed compound is also contemplated. Nasal delivery allows the passage of a compound to the blood stream directly after administering the therapeutic product to the nose, without the necessity for deposition of the product in the lung. Formulations for nasal delivery include those with dextran or cyclodextran.

[0500] For nasal administration, a useful device is a small, hard bottle to which a metered dose sprayer is attached. In one embodiment, the metered dose is delivered by drawing the pharmaceutical composition solution into a chamber of defined volume, which chamber has an aperture dimensioned to aerosolize and aerosol formulation by forming a spray when a liquid in the chamber is compressed. The chamber is compressed to administer the pharmaceutical composition. In a specific embodiment, the chamber is a piston arrangement. Such devices are commercially available.

[0501] Alternatively, a plastic squeeze bottle with an aperture or opening dimensioned to aerosolize an aerosol formulation by forming a spray when squeezed is used. The opening is usually found in the top of the bottle, and the top is generally tapered to partially fit in the nasal passages for efficient administration of the aerosol formulation. In some aspects, the nasal inhaler will provide a metered amount of the aerosol formulation, for administration of a measured dose of the drug.

[0502] The compound, when it is desirable to deliver them systemically, may be formulated for parenteral administration by injection, e.g., by bolus injection or continuous infusion. Formulations for injection may be presented in unit dosage form, e.g., in ampoules or in multi-dose containers, with an added preservative. The compositions may take such

forms as suspensions, solutions or emulsions in oily or aqueous vehicles, and may contain formulatory agents such as suspending, stabilizing and/or dispersing agents.

[0503] Pharmaceutical formulations for parenteral administration include aqueous solutions of the active compounds in water-soluble form. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions.

[0504] Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

[0505] Alternatively, the active compounds may be in powder form for constitution with a suitable vehicle, e.g., sterile pyrogen-free water, before use.

[0506] The compounds may also be formulated in rectal or vaginal compositions such as suppositories or retention enemas, e.g., containing conventional suppository bases such as cocoa butter or other glycerides.

[0507] In addition to the formulations described previously, the compounds may also be formulated as a depot preparation. Such long acting formulations may be formulated with suitable polymeric or hydrophobic materials (for example as an emulsion in an acceptable oil) or ion exchange resins, or as sparingly soluble derivatives, for example, as a sparingly soluble salt.

[0508] The pharmaceutical compositions also may comprise suitable solid or gel phase carriers or excipients. Examples of such carriers or excipients include but are not limited to calcium carbonate, calcium phosphate, various sugars, starches, cellulose derivatives, gelatin, and polymers such as polyethylene glycols.

[0509] Suitable liquid or solid pharmaceutical preparation forms are, for example, aqueous or saline solutions for inhalation, microencapsulated, encochleated, coated onto microscopic gold particles, contained in liposomes, nebulized, aerosols, pellets for implantation into the skin, or dried onto a sharp object to be scratched into the skin. The pharmaceutical compositions also include granules, powders, tablets, coated tablets, (micro)capsules, suppositories, syrups, emulsions, suspensions, creams, drops or preparations with protracted release of active compounds, in whose preparation excipients and additives

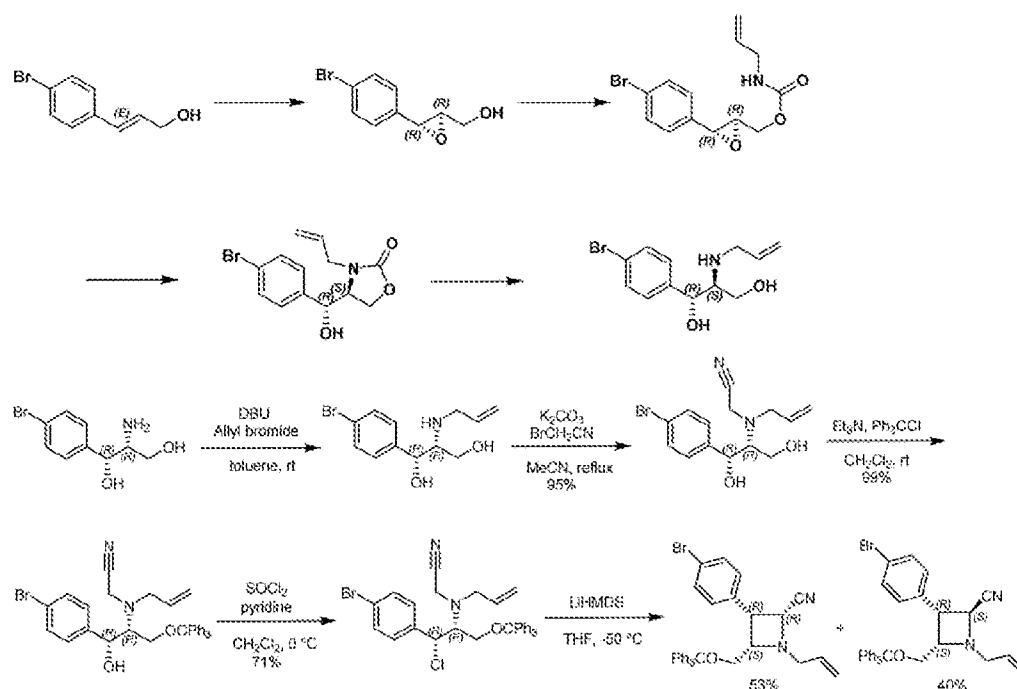
and/or auxiliaries such as disintegrants, binders, coating agents, swelling agents, lubricants, flavorings, sweeteners or solubilizers are customarily used as described above. The pharmaceutical compositions are suitable for use in a variety of drug delivery systems.

[0510] The compounds may be administered per se (neat) or in the form of a pharmaceutically acceptable salt. When used in medicine the salts should be pharmaceutically acceptable, but non-pharmaceutically acceptable salts may conveniently be used to prepare pharmaceutically acceptable salts thereof. Such salts include, but are not limited to, those prepared from the following acids: hydrochloric, hydrobromic, sulphuric, nitric, phosphoric, maleic, acetic, salicylic, p-toluene sulphonic, tartaric, citric, methane sulphonic, formic, malonic, succinic, naphthalene-2-sulphonic, and benzene sulphonic. Also, such salts can be prepared as alkaline metal or alkaline earth salts, such as sodium, potassium or calcium salts of the carboxylic acid group.

[0511] Suitable buffering agents include: acetic acid and a salt (about 1-2% w/v); citric acid and a salt (about 1-3% w/v); boric acid and a salt (about 0.5-2.5% w/v); and phosphoric acid and a salt (about 0.8-2% w/v). Suitable preservatives include benzalkonium chloride (about 0.003-0.03% w/v); chlorobutanol (about 0.3-0.9% w/v); parabens (about 0.01-0.25% w/v) and thimerosal (about 0.004-0.02% w/v).

[0512] The pharmaceutical compositions contain an effective amount of a disclosed compound optionally included in a pharmaceutically acceptable carrier. The term pharmaceutically acceptable carrier means one or more compatible solid or liquid filler, diluents or encapsulating substances which are suitable for administration to a human or other vertebrate animal. The term carrier denotes an organic or inorganic ingredient, natural or synthetic, with which the active ingredient is combined to facilitate the application. The components of the pharmaceutical compositions also are capable of being commingled with the compounds, and with each other, in a manner such that there is no interaction which would substantially impair the desired pharmaceutical efficiency.

[0513] Provided herein are methods of synthesizing disclosed compounds. A compound provided herein can be synthesized using a variety of methods known in the art. The schemes and description below depict an exemplary route for the synthesis of PubChem CID: 54650473 (BRD-K95427608 also referred to herein as BRD-7608) and provides a method of synthesizing compounds with the ring structure of Formula I.



Kits

[0514] The present compositions may be assembled into kits or pharmaceutical systems. The kits can include instructions for the treatment regime, reagents, equipment (test tubes, reaction vessels, needles, syringes, etc.) and standards for calibrating or conducting the treatment. The instructions provided in a kit according to the invention may be directed to suitable operational parameters in the form of a label or a separate insert. Optionally, the kit may further comprise a standard or control information so that the test sample can be compared with the control information standard to determine if whether a consistent result is achieved.

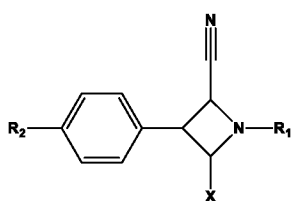
[0515] The container means of the kits will generally include at least one vial, test tube, flask, bottle, or other container means, into which a component may be placed, and preferably, suitably aliquoted. Where there is more than one component in the kit, the kit also will generally contain additional containers into which the additional components may be separately placed. However, various combinations of components may be comprised in a container. The kits of the present invention also will typically include a means for packaging the component containers in close confinement for commercial sale. Such packaging may include injection or blow-molded plastic containers into which the desired component containers are retained.

[0516] The practice of the present invention employs, unless otherwise indicated, conventional techniques of molecular biology (including recombinant techniques), microbiology, cell biology, biochemistry and immunology, which are well within the purview of the skilled artisan. Such techniques are explained fully in the literature, such as, “Molecular Cloning: A Laboratory Manual”, second edition (Sambrook, 1989); “Oligonucleotide Synthesis” (Gait, 1984); “Animal Cell Culture” (Freshney, 1987); “Methods in Enzymology” “Handbook of Experimental Immunology” (Weir, 1996); “Gene Transfer Vectors for Mammalian Cells” (Miller and Calos, 1987); “Current Protocols in Molecular Biology” (Ausubel, 1987); “PCR: The Polymerase Chain Reaction”, (Mullis, 1994); “Current Protocols in Immunology” (Coligan, 1991). These techniques are applicable to the production of the polynucleotides and polypeptides of the invention, and, as such, may be considered in making and practicing the invention. Particularly useful techniques for particular embodiments will be discussed in the sections that follow.

[0517] The following examples are put forth so as to provide those of ordinary skill in the art with a complete disclosure and description of how to make and use the assay, screening, and therapeutic methods of the invention, and are not intended to limit the scope of what the inventors regard as their invention.

[0518] The present application also provides aspects and embodiments as set forth in the following numbered Statements:

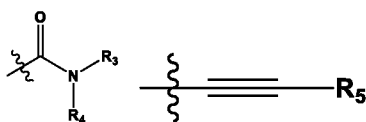
Statement 1. A method of inhibiting the nuclease activity of an RNA guided endonuclease, the method comprising contacting the RNA guided endonuclease-guide complex with a small molecule of Formula I:

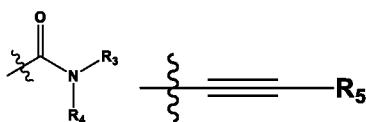


(I)

wherein:

R₁ is selected from hydrogen, COR₆, CHR₇R₈, or SO₂R₉;



R₂ is selected from hydrogen, , or a cyclic structure selected from cycloalkane, cycloalkene, aryl, or heteroaryl, each optionally substituted with one or more groups R₁₀;

R₃ is selected from hydrogen, C₁-C₁₂ alkyl, aryl, alkylaryl, optionally substituted with one or more groups selected from halogen, alkyl, alkoxy, or amine;

R₄ is selected from hydrogen, C₁-C₁₂ alkyl, aryl, alkylaryl, optionally substituted with one or more groups selected from halogen, alkyl, alkoxy, or amine;

R₅ is selected from C₁-C₁₂ alkyl and a cyclic structure selected from cycloalkane, cycloalkene, aryl, or heteroaryl, each optionally substituted with one or more groups R₁₁;

R₆ independently selected at each occurrence from C₁₋₁₂ alkyl, alkylcycloalkane, cycloalkane, aryl, heteroaryl, alkylaryl, alkylheteroaryl, or alkylamine;

R₇ is selected from hydrogen, C₁₋₁₂ alkyl, aryl, alkylaryl, hydroxyl, amine, alkoxy, or halogen;

R₈ is selected from hydrogen, C₁₋₁₂ alkyl, aryl, alkylaryl, hydroxyl, amine, alkoxy, or halogen;

R₉ is selected from C₁₋₁₂ alkyl or a cyclic structure selected from cycloalkane, cycloalkene, aryl, or heteroaryl, each optionally substituted with one or more groups R₁₂;

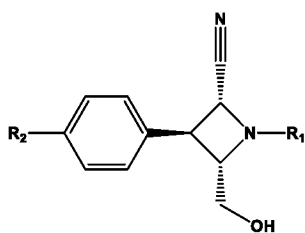
R₁₀ is independently selected at each occurrence from halogen, C₁-C₁₂ alkyl, haloalkyl, or alkoxy;

R₁₁ is independently selected at each occurrence from halogen, C₁-C₁₂ alkyl, haloalkyl, or alkoxy;

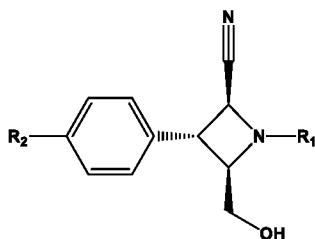
R₁₂ is independently selected at each occurrence from hydrogen, halogen, C₁-C₃ alkyl, haloalkyl, or alkoxy; and

X is selected from =O, hydroxyl, alkylalcohol, amine, or alkylamine.

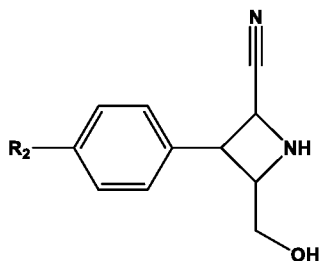
Statement 2. The method of Statement 1, wherein the compound of Formula I has the structure:



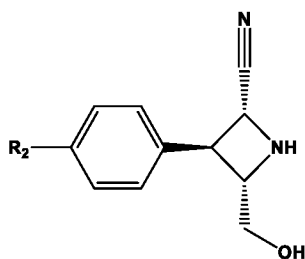
Statement 3. The method of Statement 1, wherein the compound of Formula I has the structure:



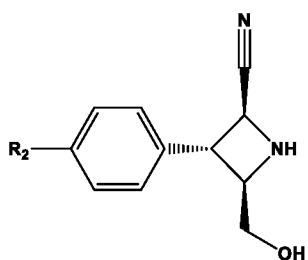
Statement 4. The method of Statement 1, wherein the compound of Formula I has the structure:



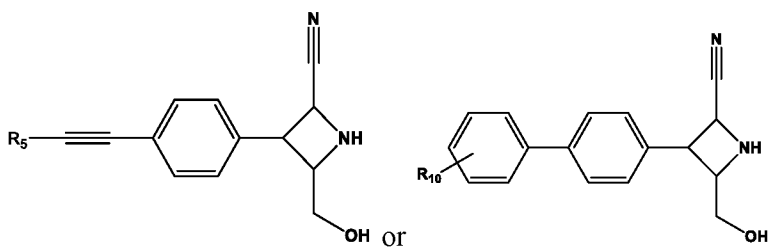
Statement 5. The method of Statement 2, wherein the compound of Formula I has the structure:



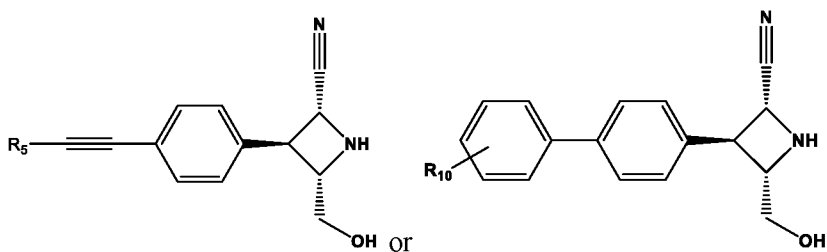
Statement 6. The method of Statement 3, wherein the compound of Formula I has the structure:



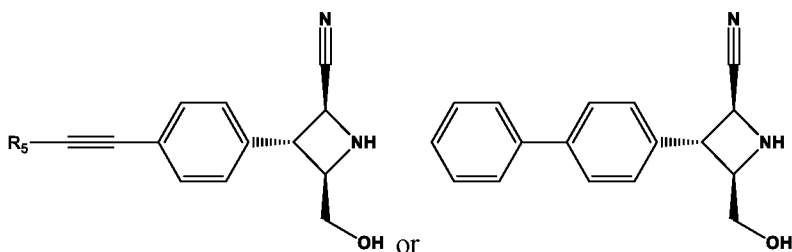
Statement 7. The method of Statement 2, wherein compound of Formula I has the structure:



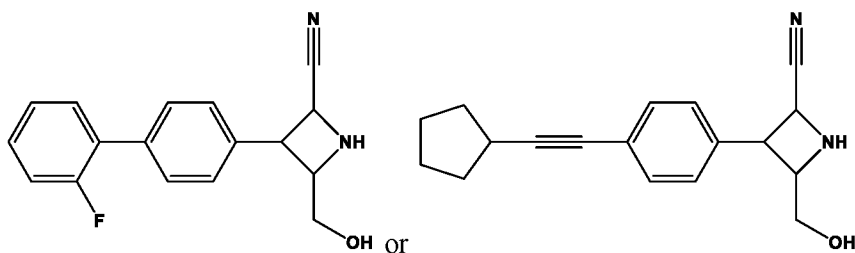
Statement 8. The method of Statement 5, wherein Formula I has the structure:



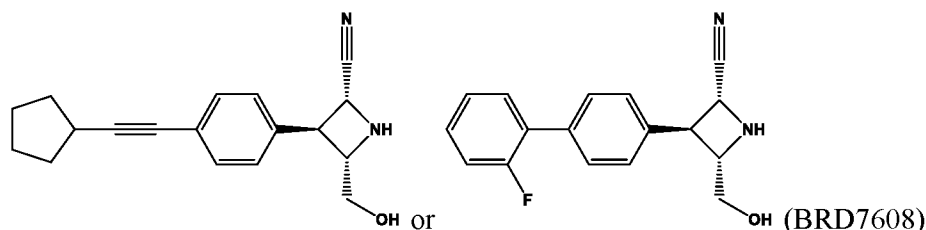
Statement 9. The method of Statement 6, wherein Formula I has the structure:



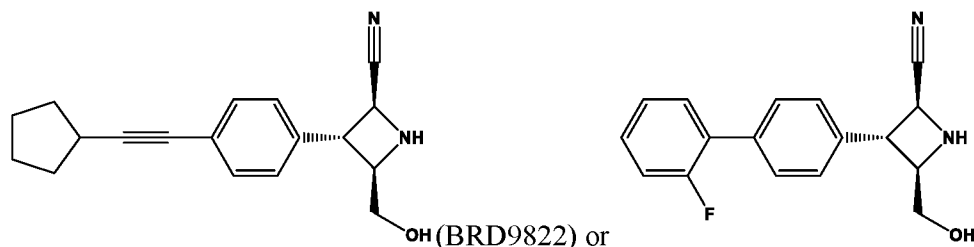
Statement 10. The method of Statement 1, wherein the compound is:



Statement 11. The method of Statement 4, wherein the compound is:



Statement 12. The method of Statement 4, wherein the compound is:



Statement 13. The method of any one of Statements 1 to 12, wherein the method is in vitro or in vivo.

Statement 14. The method of any one of Statements 1 to 3, wherein the method is performed in a cell.

Statement 15. The method of Statement 4, wherein the cell is a prokaryotic or eukaryotic cell.

Statement 16. The method of Statement 15, wherein the prokaryotic cell is a bacterium.

Statement 17. The method of Statement 15, wherein the eukaryotic cell is a human, mammalian, insect, plant, or yeast cell.

Statement 18. The method of any one of Statements 14 to 17, wherein the cell is in an organism.

Statement 19. The method of Statement 18, wherein the organism is a human, mammal, vertebrate, invertebrate, insect, or plant.

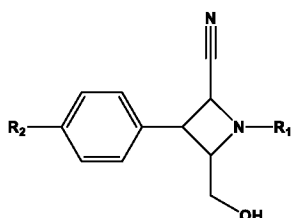
Statement 20. The method of any one of Statements 1 to 19, wherein the RNA-guided nuclease is Cas9.

Statement 21. The method of Statement 20, wherein the RNA-guided nuclease is SpCas9 or SaCas9.

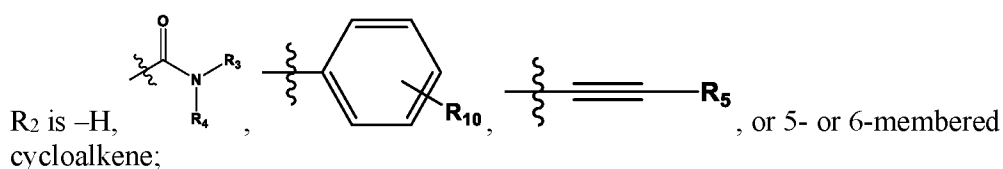
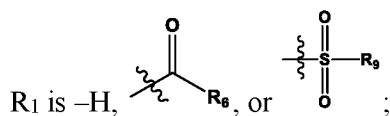
Statement 22. The method of any one of Statements 1 to 21, wherein the RNA-guided nuclease is Cas12a.

Statement 23. The method of Statement 22, wherein the RNA-guided nuclease is FnCas12a.

Statement 24. A compound of Formula I:

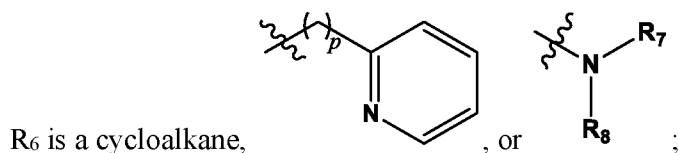
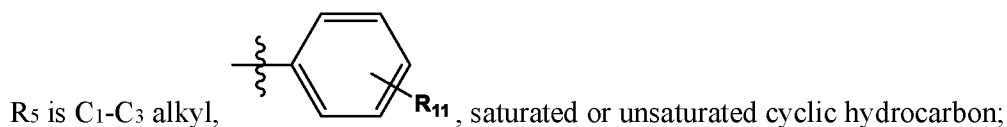


wherein:



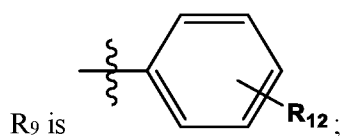
R₃ is -H, or C₁-C₃ alkyl;

R₄ is -H, or C₁-C₃ alkyl;



R₇ is -H, or C₁-C₃ alkyl;

R₈ is -H, or C₁-C₃ alkyl;



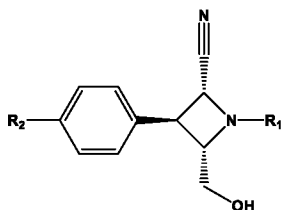
R_{10} is a hydrogen, halide, C_1 - C_3 alkyl, or alkoxy;

R_{11} is a hydrogen, halide, C_1 - C_3 alkyl, or alkoxy;

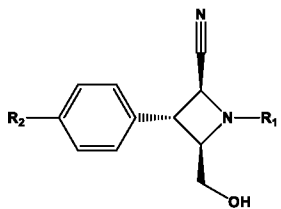
R_{12} is a hydrogen, halide, C_1 - C_3 alkyl, or alkoxy; and

P is 0 or 1.

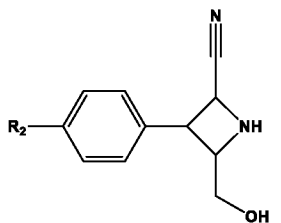
Statement 25. The compound of Statement 24, wherein Formula I is according to Formula I(a):



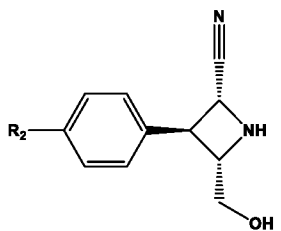
Statement 26. The compound of Statement 24, wherein Formula I is according to Formula Ib:



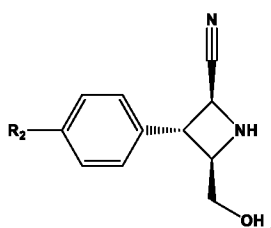
Statement 27. The compound of Statement 24, wherein the compound of Formula I has the structure:



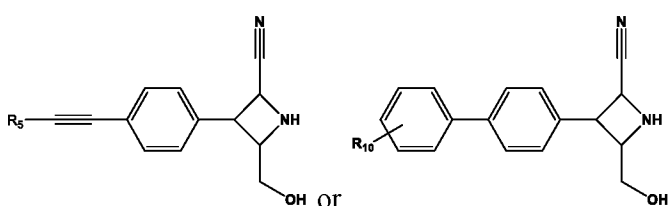
Statement 28. The compound of Statement 25, wherein the compound of Formula Ia has the structure:



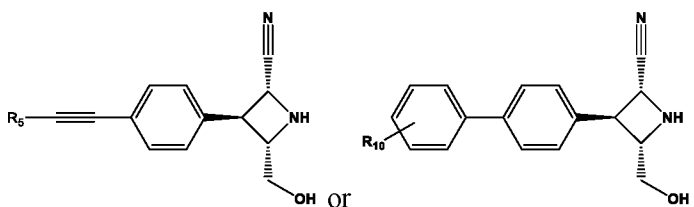
Statement 29. The compound of Statement 26, wherein the compound of Formula Ib has the structure:



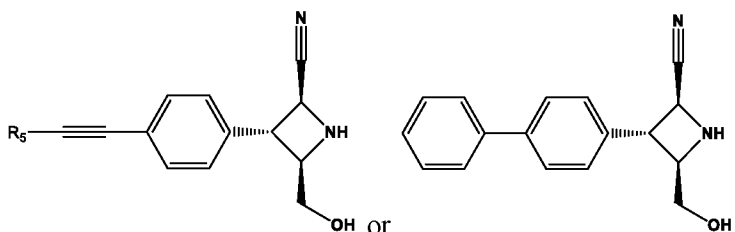
Statement 30. The compound of Statement 27, wherein compound of Formula I has the structure:



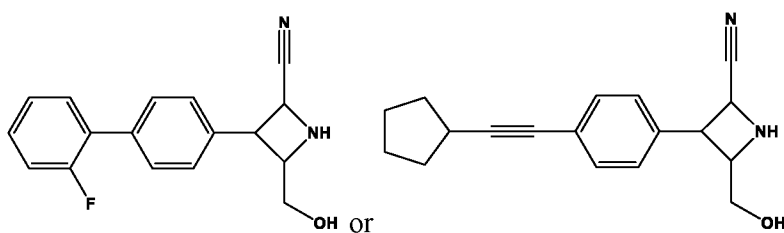
Statement 31. The compound of Statement 28, wherein the compound of Formula Ia has the structure:



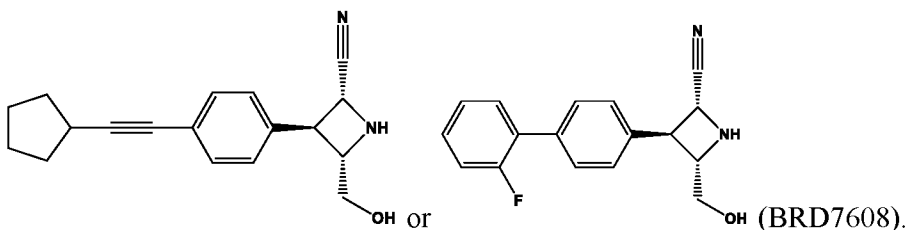
Statement 32. The compound of Statement 29, wherein Formula Ib has the structure:



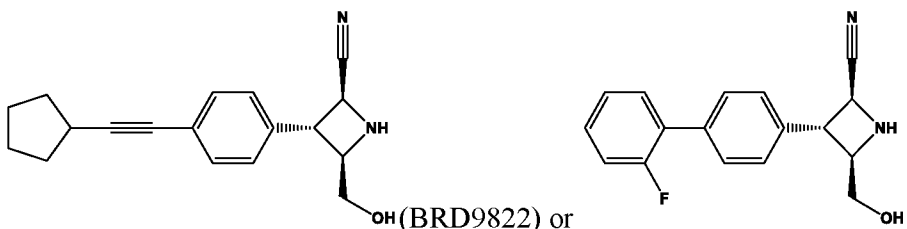
Statement 33. The compound of Statement 30, wherein the compound of Formula I has the structure:



Statement 34. The compound of Statement 31, wherein the compound has the structure:



Statement 35. The compound of Statement 32, wherein the compound has the structure:



Statement 36. A method of detecting endonuclease activity of an RNA guided endonuclease-guide complex comprising:

- a) contacting the following elements i) the RNA guided endonuclease-guide complex, ii) a double-stranded probe comprising a first oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 5' terminus or the 3' terminus and a second oligonucleotide complementary to the first oligonucleotide, wherein the second oligonucleotide is designed to hybridize to the guide whereby the first oligonucleotide and the second oligonucleotide are cleaved by the endonuclease, and iii) a third oligonucleotide comprising a fluorescence quencher, the fluorescent quencher located at the 3' terminus when the fluorescent reporter is at the 5' terminus of the first oligonucleotide, or at the 5' terminus when the fluorescent reporter is at the 3' terminus of the first oligonucleotide, and designed to hybridize to and displace the cleaved first oligonucleotide from the cleaved second oligonucleotide; and

detecting a decrease in fluorescence polarization relative to a reference, thereby detecting nuclease activity of the RNA guided endonuclease-guide complex.

Statement 37. The method of Statement 35, wherein the fluorescent reporter is at the 5' terminus and the RNA-guided nuclease is Cas9.

Statement 38. The method of Statement 36, wherein the RNA-guided nuclease is SpCas9 or SaCas9.

Statement 39. The method of Statement 35, wherein the fluorescent reporter is at the 3' terminus and the RNA-guided endonuclease is Cas12a.

Statement 40. The method of Statement 38, wherein the RNA-guided endonuclease is FnCas12a.

Statement 41. The method of any one of Statements 35 to 39, wherein the nuclease activity is inhibited by an agent.

Statement 42. The method of Statement 35, further comprising contacting the elements of step a) an agent; and wherein the detecting the decrease in the reduction of fluorescence polarization relative to the reference thereby identifies the agent as an inhibitor of the nuclease activity of the RNA guided endonuclease-guide complex.

Statement 43. The method of Statement 41, wherein the wherein the fluorescent reporter is at the 5' terminus and the RNA-guided nuclease is Cas9.

Statement 44. The method of Statement 42, wherein the RNA-guided nuclease is SpCas9 or SaCas9.

Statement 45. The method of Statement **Error! Reference source not found.**, wherein the wherein the fluorescent reporter is at the 3' terminus RNA-guided nuclease is Cas12a.

Statement 46. The method of Statement 45, wherein the RNA-guided nuclease is FnCas12a.

Statement 47. The method of any one of Statements 40 to 45, wherein the agent is a small molecule.

[0519] 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 – Cas Activity Assay

[0520] SpCas9 binds to its DNA substrate with nanomolar affinity, even following double-stranded cleavage. However, of the 4 resulting DNA fragments, it was discovered that the distal non-target strand (Fig. 1A) is weakly held, and can be displaced upon addition of excess complementary single stranded DNA. We envisioned a system wherein by fluorescently labeling the 5' end of the non-target strand, we could quench the fluorescence in a cleavage-dependent manner by adding in excess a complementary DNA strand labeled with a 3' quencher. Upon displacement of the unlabeled strand and annealing of the two labeled strands, fluorescence is quenched by a FRET mechanism, providing a proxy measurement for Cas9 activity at the RuvC domain based on the extent of fluorescence loss. Fig. 1A shows the following components:

[0521] Nucleotide Components: DS-oligo, which in the example is DS-AF647, where AF647 is Alexa Fluor 647 (the fluorophore used in this assay). This is made up of the SS-AF647 (see below) and the unlabeled reverse complement strand. Both of these DNA strands are annealed together to produce a double stranded DNA prior to use in the assay. Stocks of this oligo can be made and stored, for example at a concentration of 1 uM. SS-oligo: This is also referred to as SS-AF647. This is the non-duplexed DNA containing just the fluorophore. Stocks of this oligo from IDT are HPLC-purified and at a concentration of 100 uM. Unlabeled DNA: This is the unlabeled reverse complement strand of the SS-oligo. Stocks of this oligo from IDT are HPLC-purified and can be stored, for example at a concentration of 100 uM. Q-oligo: This is the quencher-containing displacer strand, and is the reverse complement of the fluorophore containing SS-oligo. Depending on the nature of the quencher, it will be referred to by multiple names. For example, FAMQ refers to a Q-oligo containing the Iowa Black® FQ; best for green fluorophores but does work for deep red fluorophores. RedQ refers to a Q-oligo containing the Iowa Black® RQ; best for red – deep red fluorophores. Stocks of this oligo from IDT are HPLC-purified and can be stored at a concentration of 100 uM. Prediluted stocks can be made and stored, for example at concentrations of 5 uM.

[0522] Cas9 Nuclease Components: Cas9 RNP: This is the “active” form of what is typically referred to simply as Cas9. It contains premixed Cas9 and gRNA at a ratio of 1:1.2 and is referred to herein as active Cas9, Cas9+gRNA, Cas9/gRNA, Cas9 RNP, or RNP (ribonucleoprotein). Apo Cas9: This is the “inactive” form of Cas9, as it does not contain any gRNA, and is referred to herein as ApoCas9 or inactive Cas9.

[0523] Materials

[0524] 1. Biological materials:

[0525] a. DNA substrates/quenchers

Table 3 - DNA substrates/quenchers (PAM is bold, protospacer italicized, and displacer binding underlined)		
5' AF647-SpCas9 Substrate-1 Fwd	/5Alex647N/ <u>TAATACGACTCACTATAGGACGCGACC</u> <u>GAAATGGTGAAGGACGGGT</u> (SEQ ID NO:12)	SS-oligo
5' SpCas9 Unlabeled Substrate-1 Rev	ACCCGTCCTTCACCA <u>TTTCGGTCGCGTCCTATAGTG</u> AGTCGTATTA (SEQ ID NO:13)	Unlabeled DNA
3' RedQ-SpCas9 Displacer-1 Rev	ATAGTGAGTCGTATTA/3IAbRQSp/ (SEQ ID NO:14)	Q-oligo

[0526] b. Purified and active SpCas9: MW (His6-MBP-TEV version, kDa): 201.382; Molar Extinction Coefficient (His6-MBP-TEV version, cm⁻¹ M⁻¹ 1000-1): ~188; MW (TEV-cleaved version, kDa): 158.731; Molar Extinction Coefficient (TEV-cleaved version, cm⁻¹ M⁻¹ 1000-1): 120.450 (see expasy and PMC4176945). The protein may be expressed from pMJ806.

[0527] c. Purified and active gRNA (specific for the DS-oligo):SpCas9 gRNA substrate-1

Table 4 – gRNA oligos	
Structure of SpCas9 gRNA	5'- ____ protospacer ____ GTTTTAGAGCTAGAAATAGCAAGTTAAA ATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCG AGTCGGTGCTTTT-3' (SEQ ID NO:15)
SpCas9 sgRNA Substrate-1 Fwd	TAATACGACTCACTATAG <u>CT</u> ATAGGACGCGACCGAAAGT TTTAGAGC TAGAAAT (SEQ ID NO:16)
sgRNA Universal Rev (should be SpCas9 sgRNA)	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACG GACTAGCCT TATTTTAACTTGCTATTTCTAGCTCTAAAC

Universal Rev)	(SEQ ID NO:17)
----------------	----------------

[0528] 2. Buffers: 10x Cas9 activity assay buffer: 200 mM Tris-HCl, pH=7.5, 1 M KCl, 50 mM MgCl₂; nanopure water.

[0529] 3. Screening supplies: 384-well plates: Corning® 384 Well Low Flange Black Flat Bottom Polystyrene NBS™ Microplate, Nonsterile (from VWR or Sigma Aldrich); plastic lids; 500 mL conical tubes (us.vwr.com/store/product/4648440/corning-centrifuge-tubes-sterile-polypropylene); Combi manifold (ICCB); compound library source plates; Envision plate reader (Thermo Fisher) with AF647 protocol, optimized for plate dimension, plate height, and gain (see “envision-optimization.pdf” on ICCB website:

Table 5 – Envision components	
	Filter/mirror name
Top mirror	Cy5
Bottom mirror	N/A
Exc. filter	Cy5 620
Using of excitation filter	Top
2nd exc. filter	N/A
Using of 2nd excitation filter	Top
Ems. filter	Cy5 685
2nd ems. filter	N/A

Example 2 - Small-Scale SDA (n=3)

[0530] Small scale SDAs are useful for testing activity of various assay components, inhibitor re-tests, optimizing, and troubleshooting. Three typical test parameters are: 1) DS-Oligo + Q + Cas9/gRNA (active) – max activity of nuclease, variable fluorescence signal; 2) 2. DS-Oligo + Q + Apo Cas9 (inactive) – maximum fluorescence signal of assay, upper bound; 3) SS-oligo + Q + Apo Cas9 – minimum fluorescence signal of assay, lower bound. By setting the signal from condition 2 as the max (“1”) and condition 3 as the min (“0”), the fraction cleaved can be calculated: (Cas9/gRNA – SS-oligo) / (ApoCas9 – SS-oligo).

Table 6 - Exemplary Assay

Wells and conditions	DS-Oligo	SS-Oligo	Q-Oligo	Nuclease/ sgRNA	Apo nuclease
Wells A, B, C1	1 nM		5 nM	10 nM	
Wells A, B, C2	1 nM		5 nM		10 nM
Wells A, B, C3		1 nM	5 nM		10 nM
Wells A, B, C4			5 nM		10 nM

[0531] Condition C4 provides for determination of autofluorescence of buffer + Q + Apo nuclease (background).

[0532] Samples are typically prepared in Epi tubes (low bind, 1.5 mL) at volumes of 150 uL, and once mixed, distributed to the wells of a 384 well plate 45 uL at a time for n=3 technical replicates. For example 10x stocks of all components may be added as follows:

Table 7 – Exemplary Setup (volumes of 10x stocks)

	DS- Oligo	SS- Oligo	Q-Oligo	Nuclease /sgRNA	Apo nuclease	10x assay buffer	water
Wells A, B, C1	15		15	15		15	90
Wells A, B, C2	25		15		15	15	90
Wells A, B, C3		15	15		15	15	90
Wells A, B, C4			15		15	15	105

[0533] Exemplary parameters stress-tested with SpCas9 SDA: 1) Reproducibility of RNP across multiple days; 2) Sensitivity of oligos to freeze thaw; 3) Effect of DTT on activity; 4) Storage of reagents at RT vs 4°C; 5) Effects of Combi addition versus manual pipetting; 6) Effects of incubating in different plate lot numbers; 7) Effects of incubating in a plate versus epi tube; 8) Effect of quencher concentration (relative to substrate); 9) Effect of RNP dose (relative to substrate); 10) Effect of different colored fluorophores (FAM, Cy5, AF647).

[0534] Guidance and considerations: 1) Most crucial aspect of SpCas9 SDA activity is dose of RNP relative to DS-fluorophore. Controlling and verifying active SpCas9 RNPs and

accurate concentrations are key to a successful assay. DTT is not important; 2) Stability of substrates is not as important, but should be kept at -20°C as a good practice. Repeated freeze thaws are fine, but keep at high concentration and reasonably; 3) FnCpfI is very stable at 4°C in any form; 4) Incubating SDA reagents in the 384-well plate is optimal for best signal. Leaving reagents in tubes for multiple hours can decrease signal, potentially through a DNA-denaturing mechanism or Cas9 crashing out; 5) These protein-specific lessons are not necessarily transferable to other CRISPR-nucleases and should be optimized on a case-by-case basis.

[0535] Procedure:

[0536] Step 1: Prepare the 1 uM DS-AF647 substrate, 1 uM SS-AF647, and 5 uM Q-oligo (scale up as needed, makes 100 uL of 1 uM DS/SS-AF647 and 5 uM of quencher strand)

[0537] 1. Mix 1 uL of 5'-DS-FAM-3' oligo (100 uM), 1 uL of 3'-complementary-5' oligo (100 uM), and 1 uL of 10x assay buffer, and 7 uL of water. Distribute amongst a PCR tubes (max volume 50 uL), and anneal using the "ramp anneal" protocol.

[0538] a. Ramp anneal = 95°C for 5 min, slow cooling to 25°C, takes ~45 min.

[0539] 2. Dilute each 10 uM 1:10 to 1 uM using 1x assay buffer. In this case, use 90 uL of buffer. This is the 1 uM stock of DS-AF647.

[0540] 3. Dilute 1 uL of stock SS-AF647 (100 uM) with 99 uL of 1x assay buffer to get 100 uL of 1 uM SS-AF647.

[0541] 4. Dilute 5 uL of stock RedQ (100 uM) with 95 uL of 1x assay buffer to get 100 uL of 5 uM quencher.

[0542] Step 2: Prepare the Cas9 RNP/ApoCas9:

[0543] From stock SpCas9 and stock gRNAs, dilute both in 1x assay buffer to a final concentration of 1 uM/1.2 uM, respectively. Incubate for 5 min at room temperature, then store on ice if needed. Do the same for ApoCas9. The SpCas9/gRNA RNP is unstable and tends to crash out at higher concentrations, and should be dilute. However, for these assays, it isn't too urgent to quickly dilute to lower concentrations. Never incubate at 37°C.

[0544] Step 3: Prepare 10x stocks of all reagents: Dilute all reagents 1:10 to get 10x stocks. Multiple dilutions may be needed.

[0545] Step 4: Add 10x stocks (according to the above table), mix, and distribute to 384 well plate. The reaction is initiated upon addition of Cas9, so add this last. However, if doing

inhibitor studies, Cas9 should be added first and DNA added last. Add to the 384 well plate 45 uL at a time in sequential order.

[0546] Step 5: Incubate plate at 37°C for 2 hours and then read on Envision AF647 protocol. Envision data can be exported as a .txt, .xlsx, or .mht file.

Example 3 - Screening-Scale SDA (many plates)

Table 8 - Exemplary plate layout (compounds in columns 3-22)

Column	1	2	3-22	23	24
	No DMSO	DMSO	DMSO / Compound	No DMSO	DMSO
	active Cas9, DS, Q	activeCas9, DS, Q	Active Cas9, DS, Q	inactive Cas9, DS, Q	inactive Cas9, DS, Q

[0547] Useful controls: 1) DS-Oligo + Q + Cas9/gRNA (active) + DMSO (Column 1-2, n=32 per plate, lowest signal, max activity); 2) DS-Oligo + Q + Apo Cas9 (inactive) + no compound (Column 23-24, n=32, highest signal, max inhibition).

[0548] Procedure

[0549] Clearly denote what compound plates will be screened for a given experiment and tabulate them in a spreadsheet. If you have already barcoded some 384-well assay plates, assign 2 barcoded plates per single compound plate. Otherwise, manually write down what compound plate is to be used for which 384-well assay plate.

[0550] Stocks of DS-AF647 and RedQ can be made at 1 uM and 5 uM, respectively. For SpCas9, the optimized final concentrations of the assay are as follows: [DS-AF647] = 0.5 nM; [RedQ] = 2.5 nM; [SpCas9/gRNA] = 5/6 nM; Incubation time at 37°C = 2 hours; Expected Z' = 0.68.

[0551] Step 1: Prepare the RNP/RNP 2x screening solution. Premix the required buffer and water amounts together in a 500 mL conical tube (or larger volume reservoir). From stock SpCas9 and stock gRNAs, dilute both in 1x assay buffer to a final concentration of 1 uM/1.2 uM, respectively. Incubate for 1 min at room temperature or on ice, then immediately dilute to 10 nM. Incubate for at least 10 min at room temperature at 10 nM. This step is important, as the SpCas9/gRNA RNP is unstable and tends to crash out at higher concentrations. It is important to follow this step exactly. Do not incubate at 37°C.

[0552] Step 2: Prepare the ApoCas9 2x screening solution. Premix the required buffer and water amounts together in an appropriate reservoir. This uses a smaller volume than in step 1, so don't use a 500 mL conical. From stock SpCas9, dilute in 1x assay buffer to a final concentration of 1 uM. Incubate for 1 min at room temperature, then immediately dilute to 10 nM. Incubate for at least 10 min at room temperature. This step is not as urgent as above, as Apo SpCas9 is very stable.

[0553] Step 3: Addition of Cas9 solutions to 384-well assay plates: This may be done using the Combi. 2x Cas9/sgRNA is added to columns 1-22 first for all plates. Next, the Combi will be washed, primed with 2x Apo Cas9, and added to Columns 23 and 24.

[0554] Step 4: Pinning the compounds into the plates and counterscreen #1: Compounds should be pinned to yield a final concentration of 20 uM (in 50 uL volume). For most compound plates, this is 100 nL of a 10 mM DMSO stock of compound. Check to see if this is the case for the plates you are screening. The pinned plates will be incubated with Cas9/sgRNA and Cas9 alone at room temperature for at least 30 minutes. During this stage, read the plates using the AF647 protocol. This is counterscreen #1. Keep the order of your plates straight, and read in order of reagent addition.

[0555] Step 5: Addition of DS/Q solutions to 384-well assay plates: This may also be done using the Combi. 2x DS/Q solution will be prepared as in the spreadsheet. This solution should be prepared about 10 minutes before you are ready to use, and should be kept shielded from direct light as best as possible. Add this to every column of the plate (1-24). Immediately after addition, read the plates using the AF647 protocol. This is counterscreen #2. Keep the order of your plates straight, and read in order of reagent addition. Once the read is complete, place a plastic lid on the plate

[0556] Step 6: Incubate the plates at 37°C for 2 hours; Incubate the plates spread out in a 37°C incubator for 2 hours. Plates may be stacked on top of each other, but try to minimize this.

[0557] Step 7: Read the primary screen: Remove the lids, and read the plates using the AF647 protocol. This is the primary screen. Once a plate has been read, add the lid again and return it to the incubator.

[0558] Step 8: Read the kinetic counter screen ~ 24 hours later: The next day, you will need to re-read the plates using the AF647 protocol after 24 hours has passed. This is the final counterscreen #3. After this, you can discard the plates.

Example 3 - Optimization

[0559] A PAM-sequence is required for quenching in the SDA as demonstrated in Fig. 3A and B. Exemplary Optimizations are depicted in Figs. 4-8. Fig. 4 exemplifies the determination of the Limit of Detection (LOD) for the SpCas9 DS-AF647 substrate/RedQ displacer strand pair. Fig. 5 exemplifies optimization of the SpCas9/DS-AF647 ratio and time course of the reaction. DS-AF647 and RedQ were fixed at 1 nM and 5 nM. Fig. 6 shows kinetics of the SpCas9 SDA reaction at various ratios of SpCas9/DS-AF647. Fig. 7 exemplifies optimization of the ratio of RedQ quencher strand to DS-AF647, and time course of the reaction. DS-AF647 and SpCas9 were fixed at 1 nM and 5 nM. Fig. 8 shows fine tuning of optimized conditions at the Combi distribution level (n=32, all reagents were dispensed via Combi head to simulate the native screens).

Example 4 – Identification of inhibitors of SpCas9 endonuclease

[0560] The Broad Institute diversity-oriented synthesis (DOS) library was screened for inhibition of SpCas9. A scatter plot is shown in Fig. 11. Fig. 12A shows the hit rate distribution among various chemical classes of the library. The following table lists 56 compounds having 3-sigma activity or greater by Depositor-Supplied Synonyms (See Pubchem Open Chemistry Database).

[0561] Compounds of the azetidine nitrile class were then tested to identify cell-permeable inhibitors (Fig. 12B-C). Fig. 13 shows 20 compounds having 3-sigma activity or greater by both plasmid and RNP eGFP assay.

[0562] Activity and structure activity relationship (SAR) was investigated for compounds of interest. Fig. 14 shows activity of two of the compounds, designated BRD7608 and BRD9822. BRD7608 is deposited in the PubChem Open Chemistry Database as PubChem CID: 54650473. BRD9822 is deposited in the PubChem Open Chemistry Database as PubChem CID: 54651106. Fig. 15 shows SAR data based on the eGFP screen (Fig. 12B-12C).

Example 5 – Identification of inhibitors of SaCas9 endonuclease

[0563] A small molecule library was screened for inhibition of SaCas9. A scatter plot is shown in Fig. 11

Example 6 – Identification of inhibitors of FnCpf1 endonuclease

[0564] The small molecule library was screened for inhibition of FnCpf1. A scatter plot is shown in Fig. 11.

Example 7 – Cellular inhibition of SaCas9.

[0565] SpCas9 mediated knockdown of eGFP is inhibited by BRD7608 ($IC_{50} = 3.9 \mu M$) and by BRD9822 ($IC_{50} = 5.5 \mu M$) (Fig. 16B). eGFP fluorescence is maintained in the presence of eGFP-targeted SpCas9 activity by inhibitors BRD7608 or BRD9822. (Fig. 16C).

Example 8 - Inhibition of gene tagging in cells by CRISPR-mediated homologous recombination.

[0566] CRISPR mediated gene tagging with HiBiT (an 11 amino acid peptide tag detectable by bioluminescence) (Fig. 17A) was inhibited in a dose-dependent manner by BRD7608 and BRD9822 (Fig. 17B).

Example 9 – Structure Activity Relationships

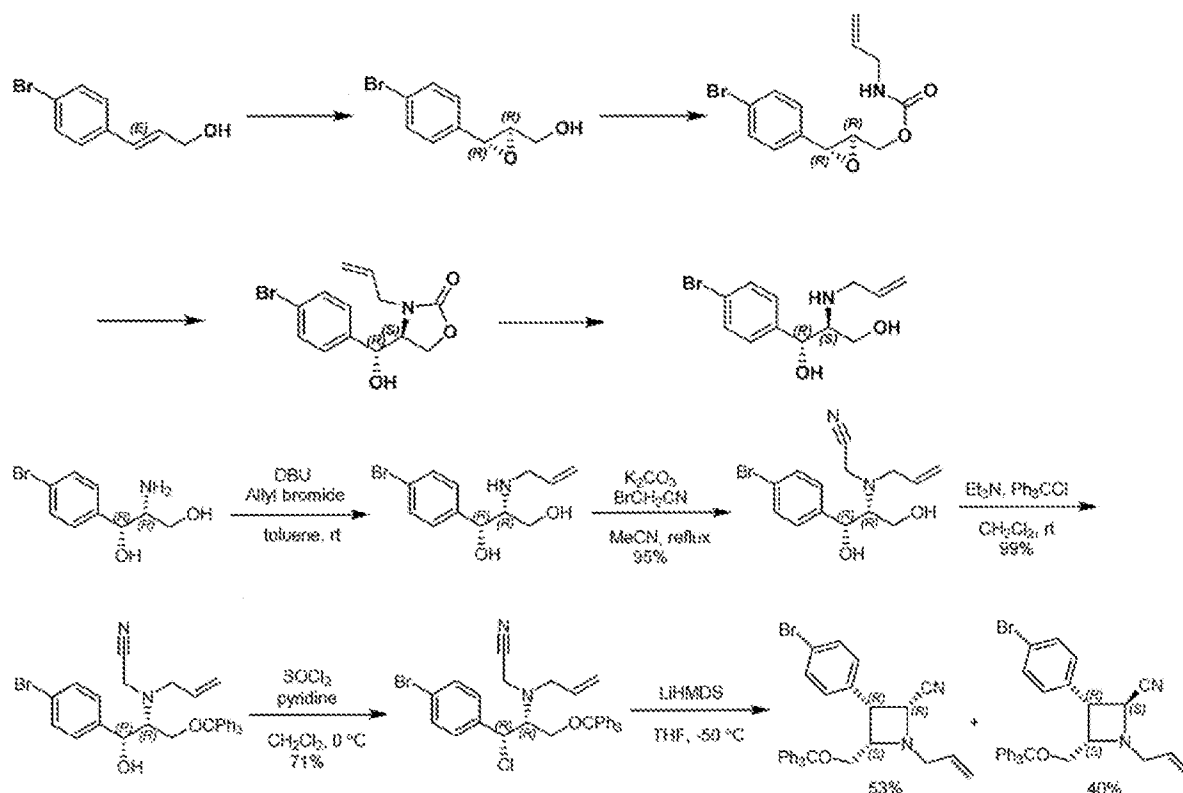
[0567] Structure Activity Relationships of BRD7608 (Fig. 19) and BRD9822 (Fig. 20) and related compounds were investigated.

Example 10 – Broad Spectrum CRISPR inhibition.

[0568] BRD7608 and BRD9822 were determined to be broad-spectrum CRISPR inhibitors. As shown in Fig. 22, the compounds inhibit the different CRISPR proteins at similar concentrations.

Example 11 – Synthesis of Compounds

[0569] The following routine includes synthesis of PubChem CID: 54650473 (BRD-K95427608 also referred to herein as BRD-7608) and is exemplary.

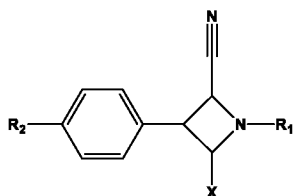


[0570] Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific 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:

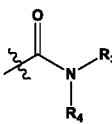
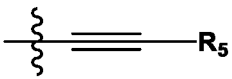
1. A method of inhibiting the nuclease activity of an RNA guided endonuclease, the method comprising contacting the RNA guided endonuclease-guide complex with a small molecule of Formula I:



(I)

wherein:

R₁ is selected from hydrogen, COR₆, CHR₇R₈, or SO₂R₉;

R₂ is selected from hydrogen, , , or a cyclic structure selected from cycloalkane, cycloalkene, aryl, or heteroaryl, each optionally substituted with one or more groups R₁₀;

R₃ is selected from hydrogen, C₁-C₁₂ alkyl, aryl, alkylaryl, optionally substituted with one or more groups selected from halogen, alkyl, alkoxy, or amine;

R₄ is selected from hydrogen, C₁-C₁₂ alkyl, aryl, alkylaryl, optionally substituted with one or more groups selected from halogen, alkyl, alkoxy, or amine;

R₅ is selected from C₁-C₁₂ alkyl and a cyclic structure selected from cycloalkane, cycloalkene, aryl, or heteroaryl, each optionally substituted with one or more groups R₁₁;

R₆ independently selected at each occurrence from C₁₋₁₂ alkyl, alkylcycloalkane, cycloalkane, aryl, heteroaryl, alkylaryl, alkylheteroaryl, or alkylamine;

R₇ is selected from hydrogen, C₁₋₁₂ alkyl, aryl, alkylaryl, hydroxyl, amine, alkoxy, or halogen;

R₈ is selected from hydrogen, C₁₋₁₂ alkyl, aryl, alkylaryl, hydroxyl, amine, alkoxy, or halogen;

R₉ is selected from C₁₋₁₂ alkyl or a cyclic structure selected from cycloalkane, cycloalkene, aryl, or heteroaryl, each optionally substituted with one or more groups R₁₂;

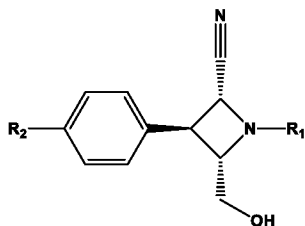
R₁₀ is independently selected at each occurrence from halogen, C₁₋₁₂ alkyl, haloalkyl, or alkoxy;

R₁₁ is independently selected at each occurrence from halogen, C₁₋₁₂ alkyl, haloalkyl, or alkoxy;

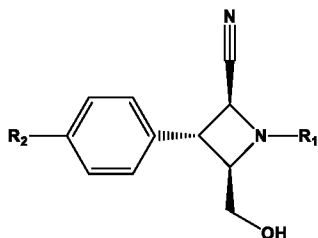
R₁₂ is independently selected at each occurrence from hydrogen, halogen, C₁₋₃ alkyl, haloalkyl, or alkoxy; and

X is selected from =O, hydroxyl, alkylalcohol, amine, or alkylamine.

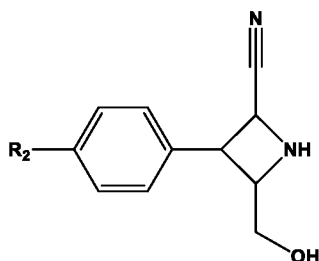
2. The method of claim 1, wherein the compound of Formula I has the structure:



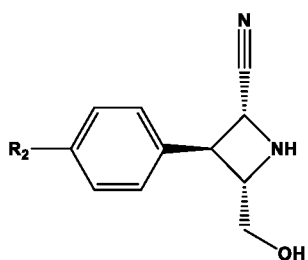
3. The method of claim 1, wherein the compound of Formula I has the structure:



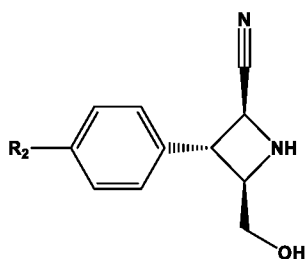
4. The method of claim 1, wherein the compound of Formula I has the structure:



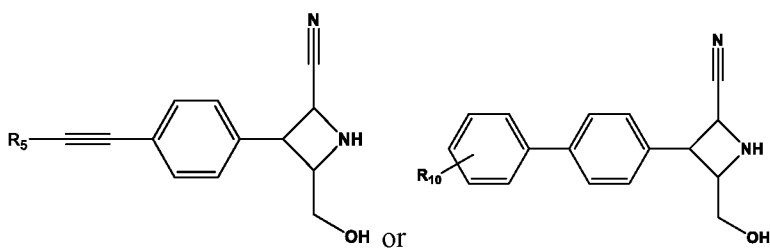
5. The method of claim 2, wherein the compound of Formula I has the structure:



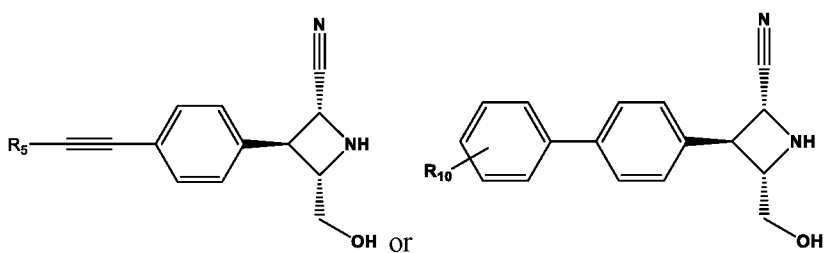
6. The method of claim 3, wherein the compound of Formula I has the structure:



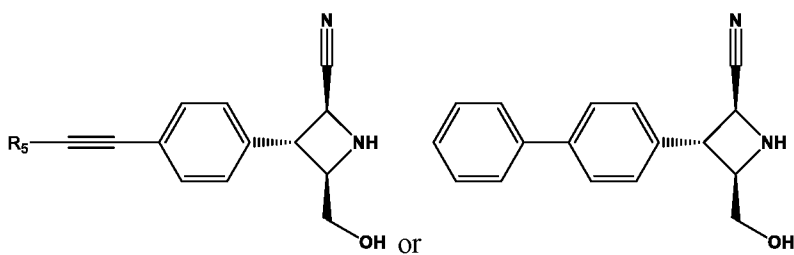
7. The method of claim 2, wherein compound of Formula I has the structure:



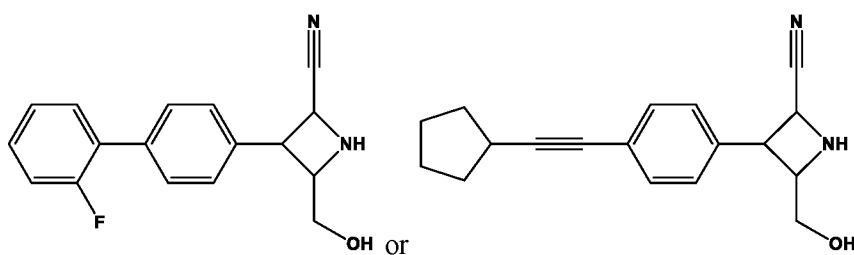
8. The method of claim 5, wherein Formula I has the structure:



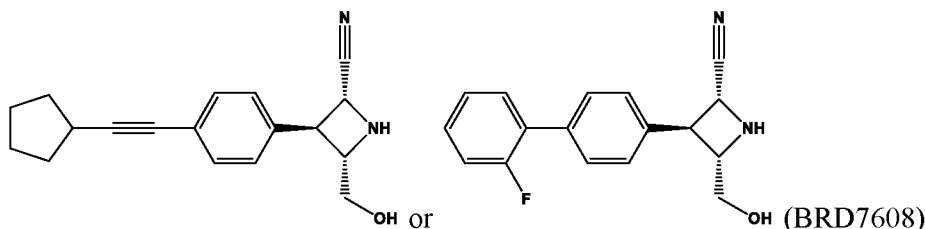
9. The method of claim 6, wherein Formula I has the structure:



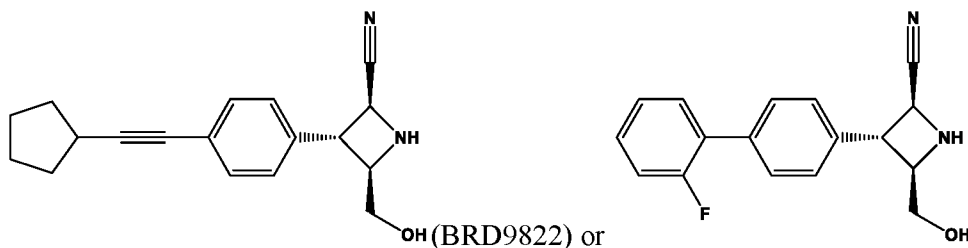
10. The method of claim 1, wherein the compound is:



11. The method of claim 4, wherein the compound is:

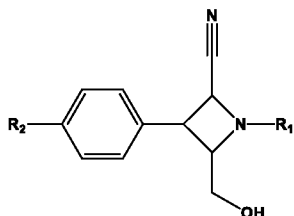


12. The method of claim 4, wherein the compound is:

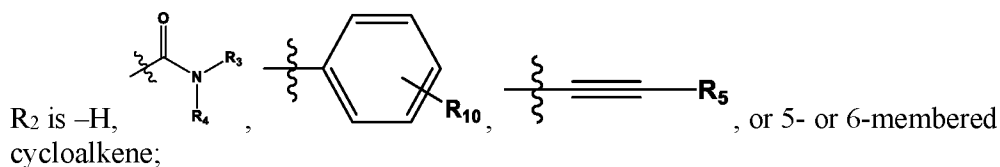
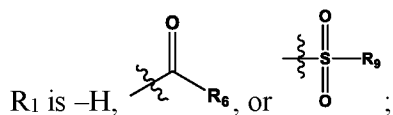


13. The method of any one of claims 1 to 12, wherein the method is in vitro or in vivo.
14. The method of any one of claims 1 to 3, wherein the method is performed in a cell.
15. The method of claim 4, wherein the cell is a prokaryotic or eukaryotic cell.
16. The method of claim 15, wherein the prokaryotic cell is a bacterium.
17. The method of claim 15, wherein the eukaryotic cell is a human, mammalian, insect, plant, or yeast cell.
18. The method of any one of claims 14 to 17, wherein the cell is in an organism.
19. The method of claim 18, wherein the organism is a human, mammal, vertebrate, invertebrate, insect, or plant.
20. The method of any one of claims 1 to 19, wherein the RNA-guided nuclease is Cas9.

21. The method of claim 20, wherein the RNA-guided nuclease is SpCas9 or SaCas9.
22. The method of any one of claims 1 to 21, wherein the RNA-guided nuclease is Cas12a.
23. The method of claim 22, wherein the RNA-guided nuclease is FnCas12a.
24. A compound of Formula I:

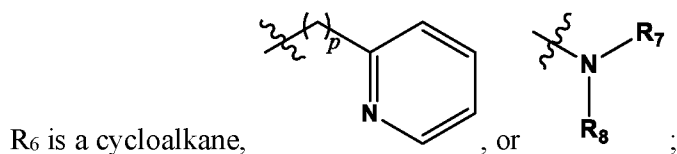
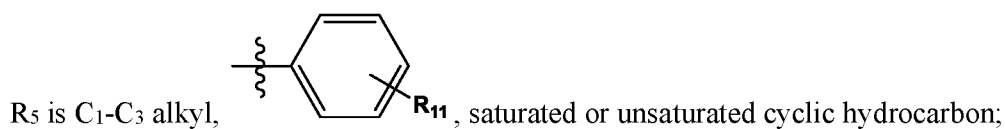


wherein:



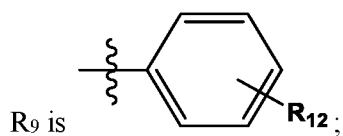
R_3 is $-H$, or C_1 - C_3 alkyl;

R_4 is $-H$, or C_1 - C_3 alkyl;



R_7 is $-H$, or C_1 - C_3 alkyl;

R_8 is $-H$, or C_1 - C_3 alkyl;



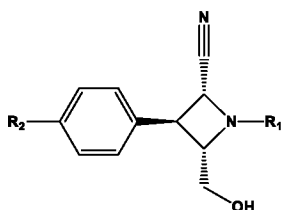
R₁₀ is a hydrogen, halide, C₁-C₃ alkyl, or alkoxy;

R₁₁ is a hydrogen, halide, C₁-C₃ alkyl, or alkoxy;

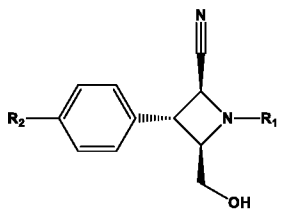
R₁₂ is a hydrogen, halide, C₁-C₃ alkyl, or alkoxy; and

P is 0 or 1.

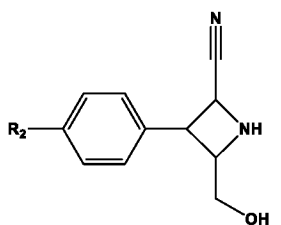
25. The compound of claim 24, wherein Formula I is according to Formula I(a):



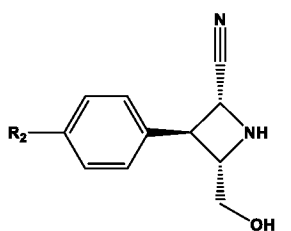
26. The compound of claim 24, wherein Formula I is according to Formula Ib:



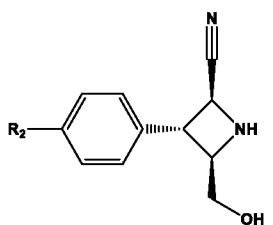
27. The compound of claim 24, wherein the compound of Formula I has the structure:



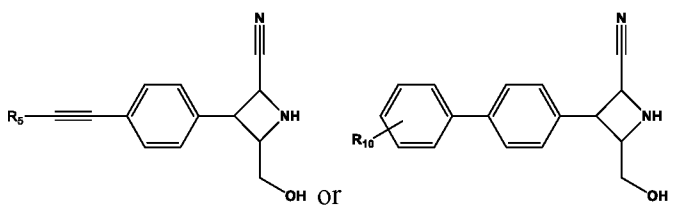
28. The compound of claim 25, wherein the compound of Formula Ia has the structure:



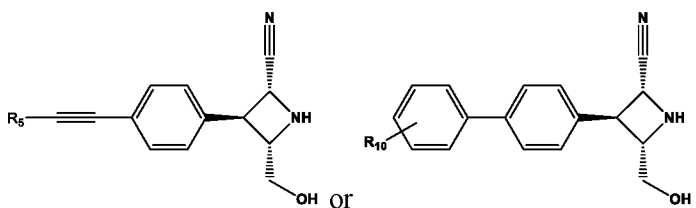
29. The compound of claim 26, wherein the compound of Formula Ib has the structure:



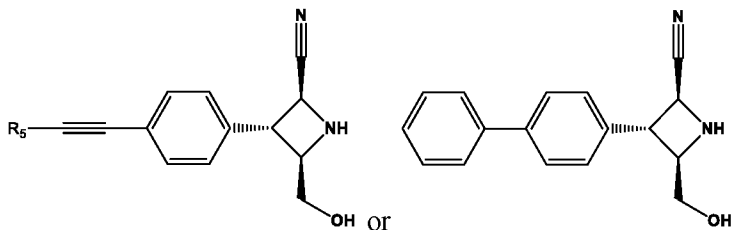
30. The compound of claim 27, wherein compound of Formula I has the structure:



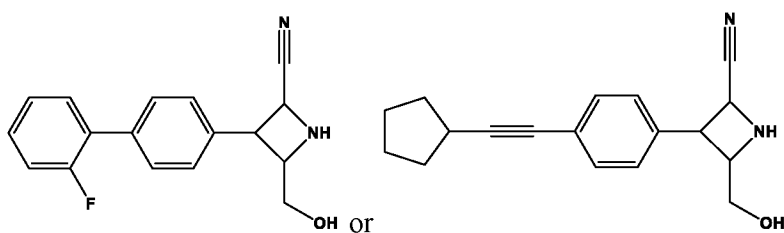
31. The compound of claim 28, wherein the compound of Formula Ia has the structure:



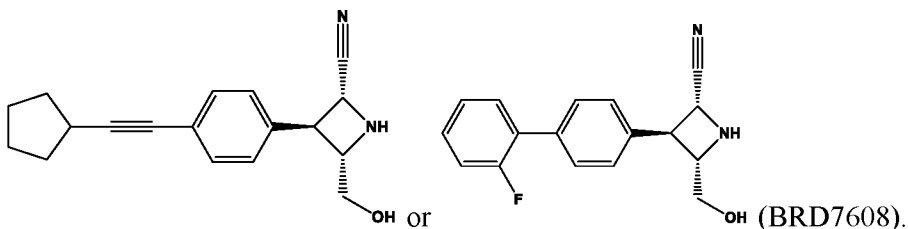
32. The compound of claim 29, wherein Formula Ib has the structure:



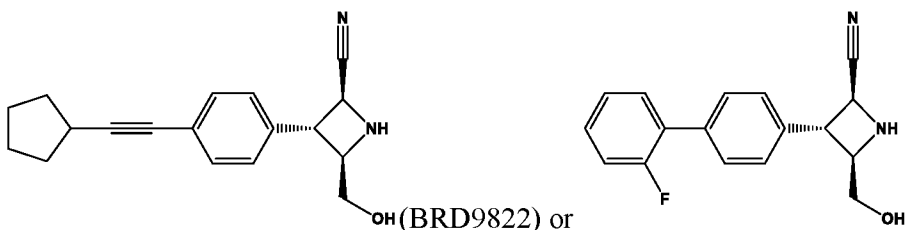
33. The compound of claim 30, wherein the compound of Formula I has the structure:



34. The compound of claim 31, wherein the compound has the structure:



35. The compound of claim 32, wherein the compound has the structure:



36. A method of detecting endonuclease activity of an RNA guided endonuclease-guide complex comprising:

- b) contacting the following elements i) the RNA guided endonuclease-guide complex, ii) a double-stranded probe comprising a first oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 5' terminus or the 3' terminus and a second oligonucleotide complementary to the first oligonucleotide, wherein the second oligonucleotide is designed to hybridize to the guide whereby the first oligonucleotide and the second oligonucleotide are cleaved by the endonuclease, and iii) a third oligonucleotide comprising a fluorescence quencher, the fluorescent quencher located at the 3' terminus when the fluorescent reporter is at the 5' terminus of the first oligonucleotide, or at the 5' terminus when the fluorescent reporter is at the 3' terminus of the first oligonucleotide, and designed to hybridize to and displace the cleaved first oligonucleotide from the cleaved second oligonucleotide; and
- detecting a decrease in fluorescence polarization relative to a reference, thereby detecting nuclease activity of the RNA guided endonuclease-guide complex.

37. The method of claim 35, wherein the fluorescent reporter is at the 5' terminus and the RNA-guided nuclease is Cas9.

38. The method of claim 36, wherein the RNA-guided nuclease is SpCas9 or SaCas9.

39. The method of claim 35, wherein the fluorescent reporter is at the 3' terminus and the RNA-guided endonuclease is Cas12a.

40. The method of claim 38, wherein the RNA-guided endonuclease is FnCas12a.

41. The method of any one of claims 35 to 39, wherein the nuclease activity is inhibited by an agent.

42. The method of claim 35, further comprising
contacting the elements of step a) an agent; and
wherein the detecting the decrease in the reduction of fluorescence polarization relative to the reference thereby identifies the agent as an inhibitor of the nuclease activity of the RNA guided endonuclease-guide complex.

43. The method of claim 42, wherein the fluorescent reporter is at the 5' terminus and the RNA-guided nuclease is Cas9.

44. The method of claim 43, wherein the RNA-guided nuclease is SpCas9 or SaCas9.

45. The method of claim 42, wherein the wherein the fluorescent reporter is at the 3' terminus RNA-guided nuclease is Cas12a.

46. The method of claim 45, wherein the RNA-guided nuclease is FnCas12a.

47. The method of any one of claims 41 to 46, wherein the agent is a small molecule.

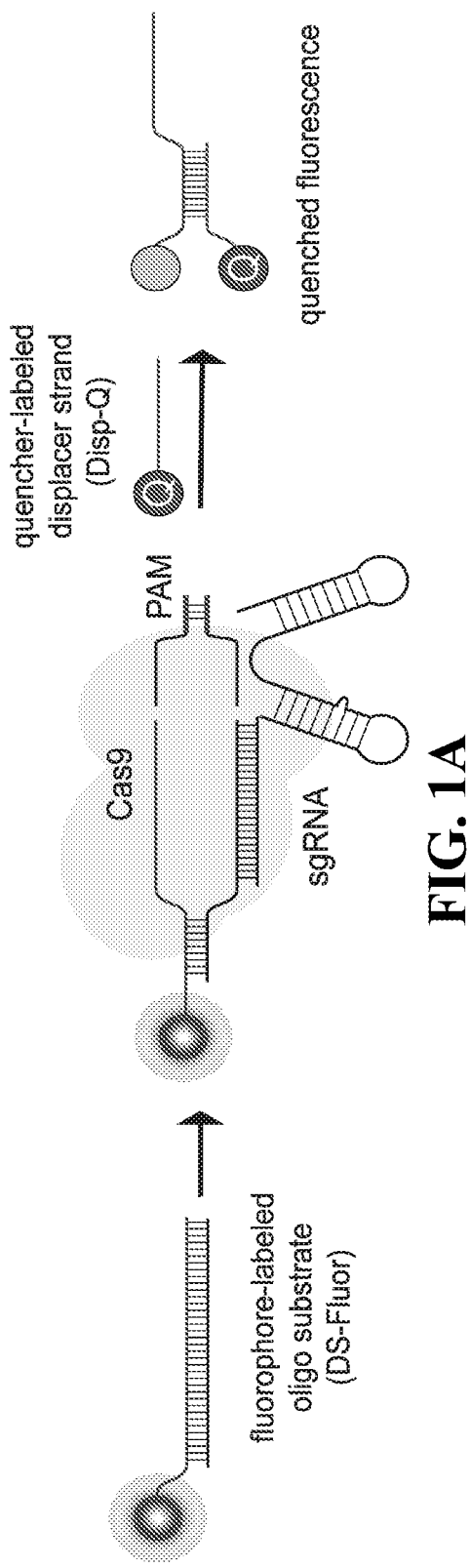


FIG. 1A

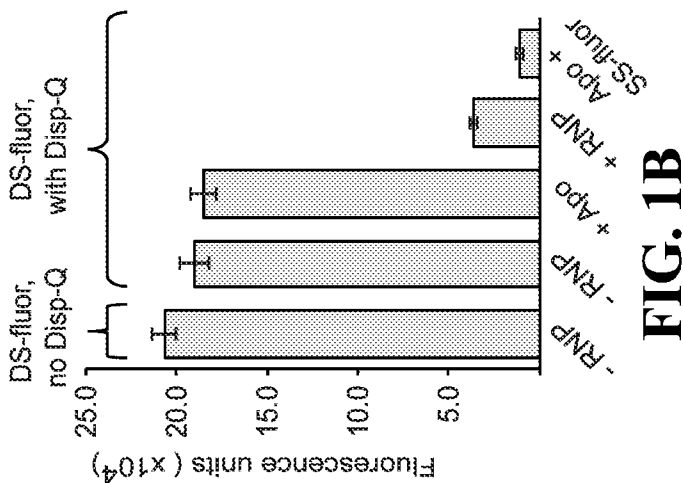


FIG. 1B

2/22

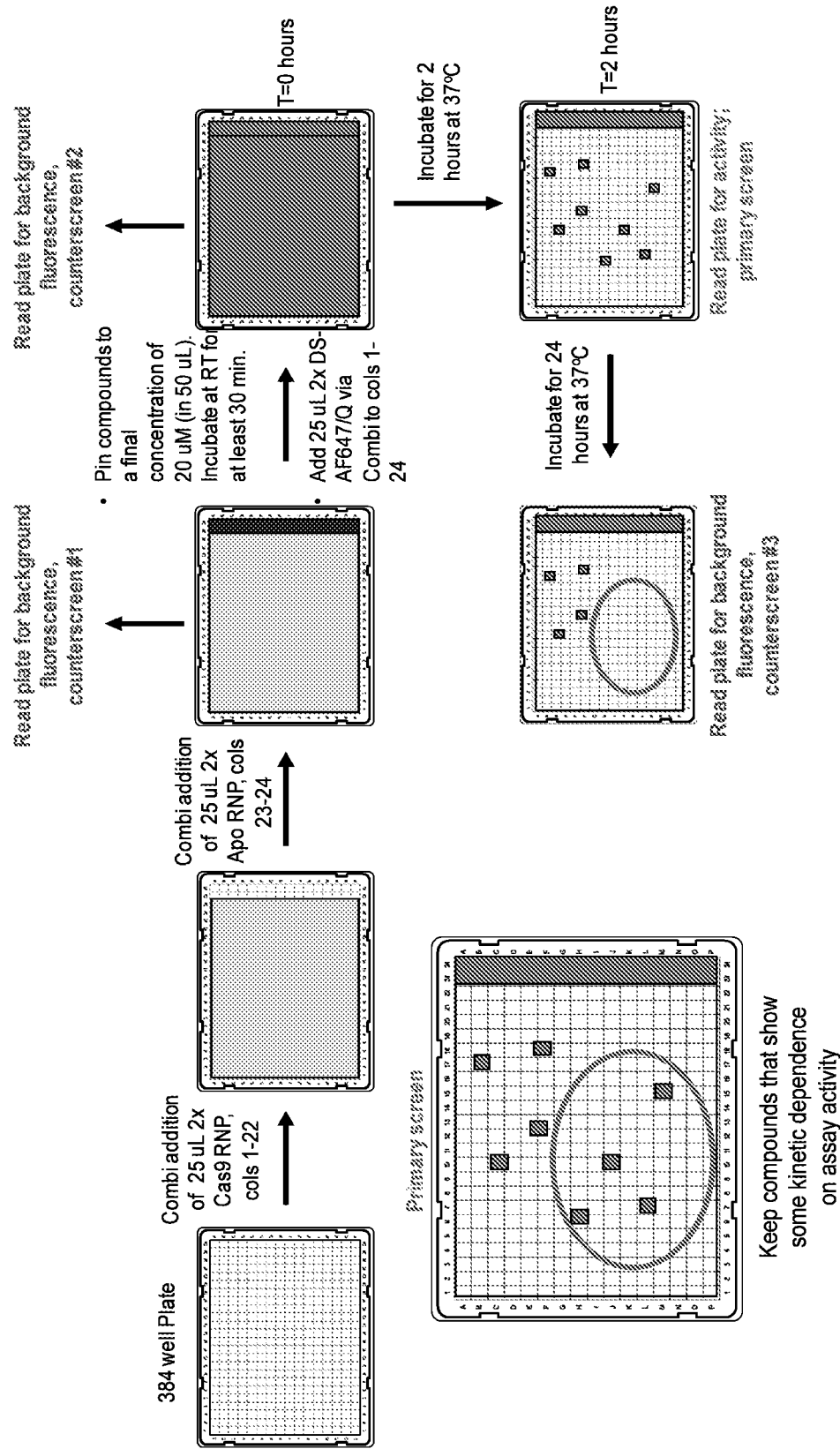


FIG. 2

3/22

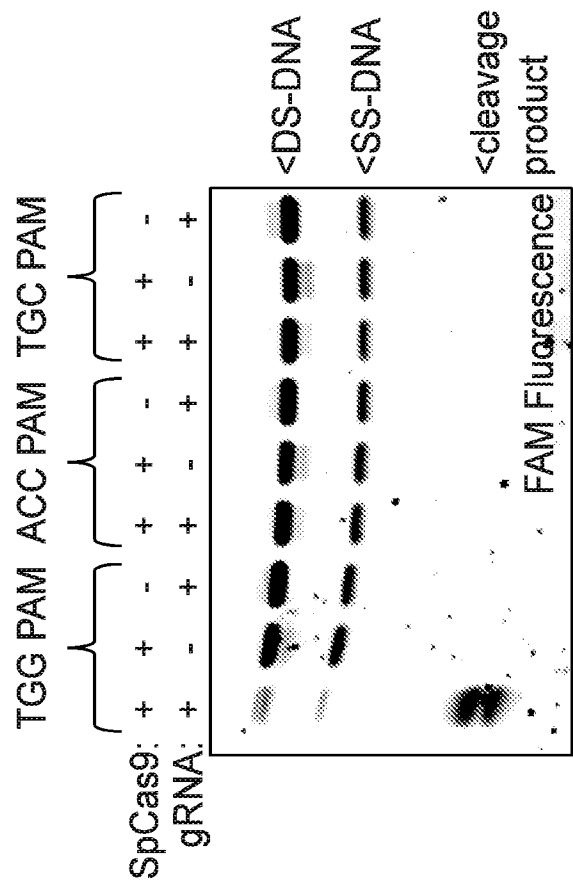


FIG. 3B

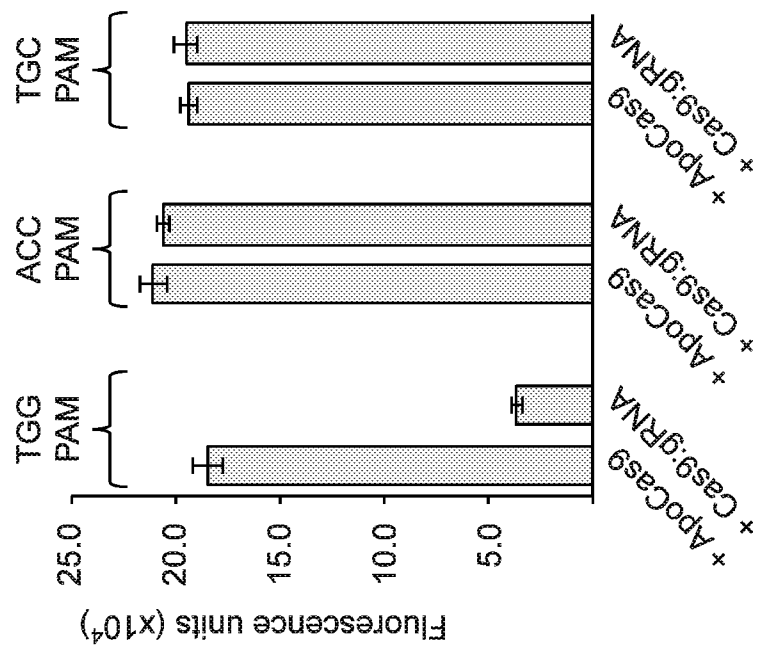
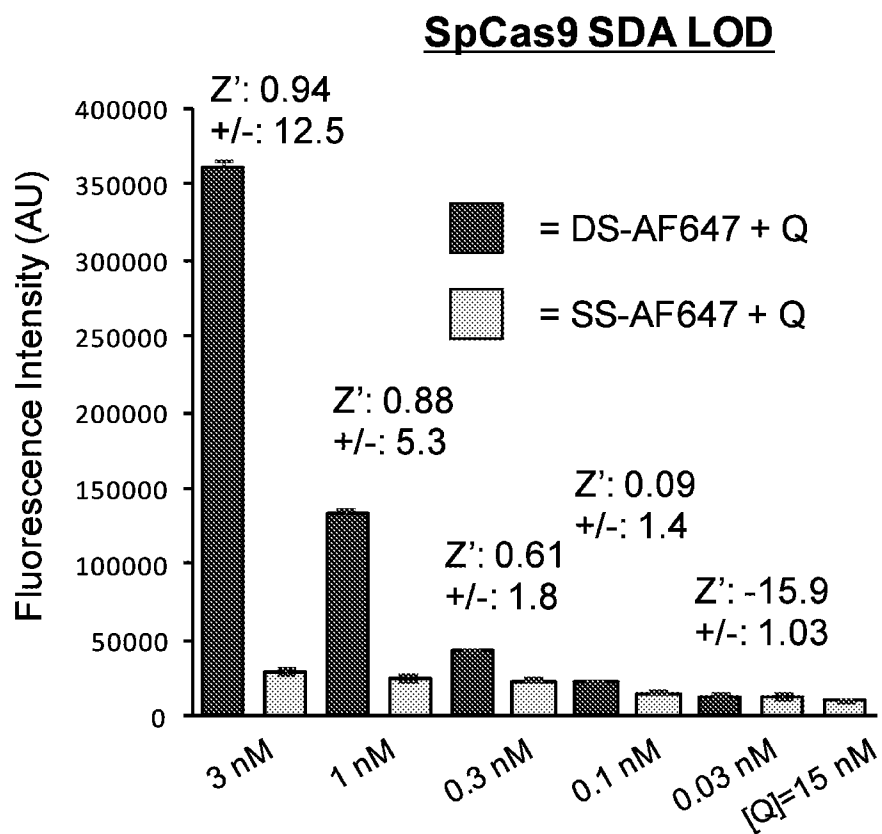


FIG. 3A

4/22

**FIG. 4**

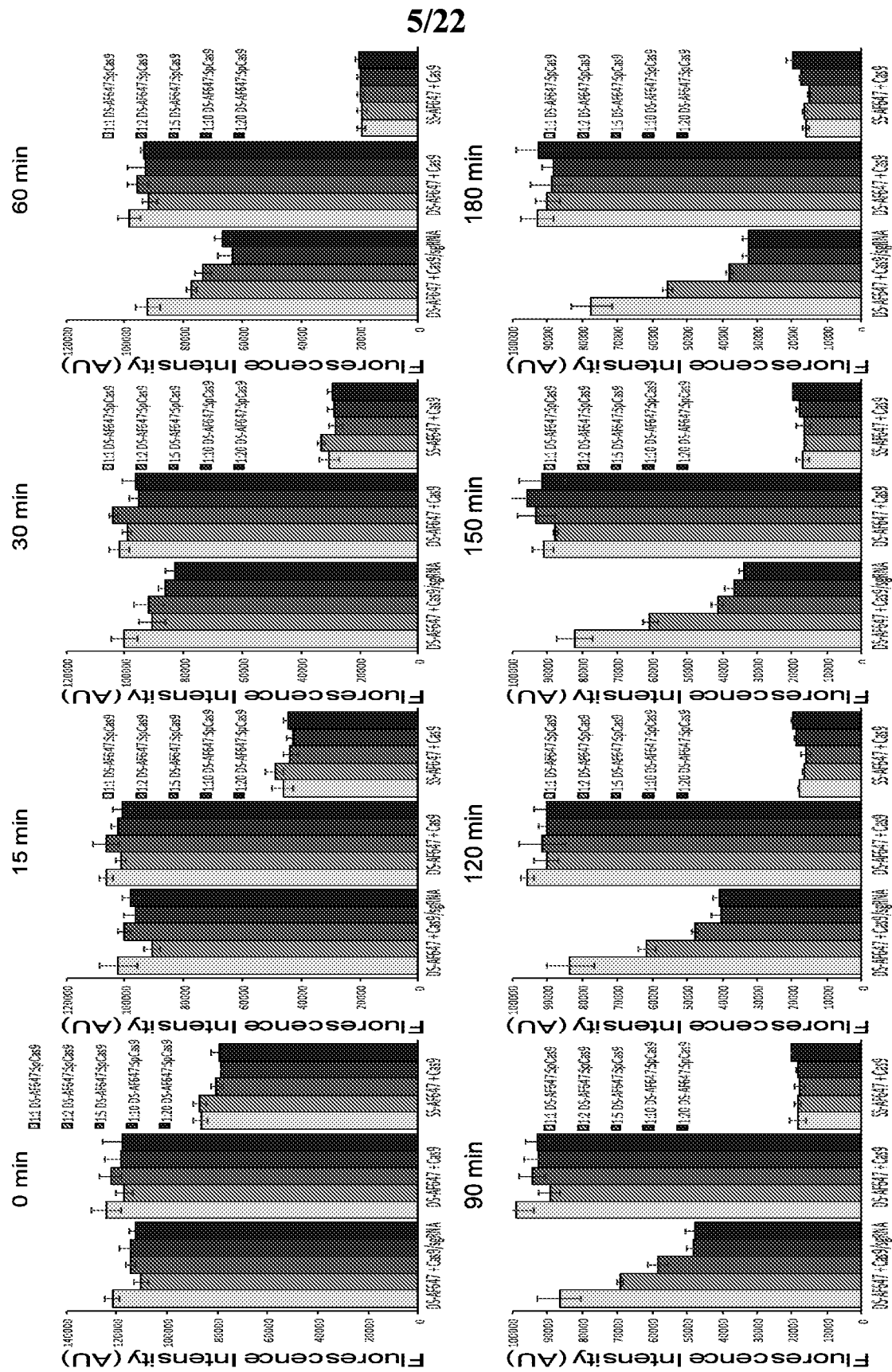


FIG. 5

6/22

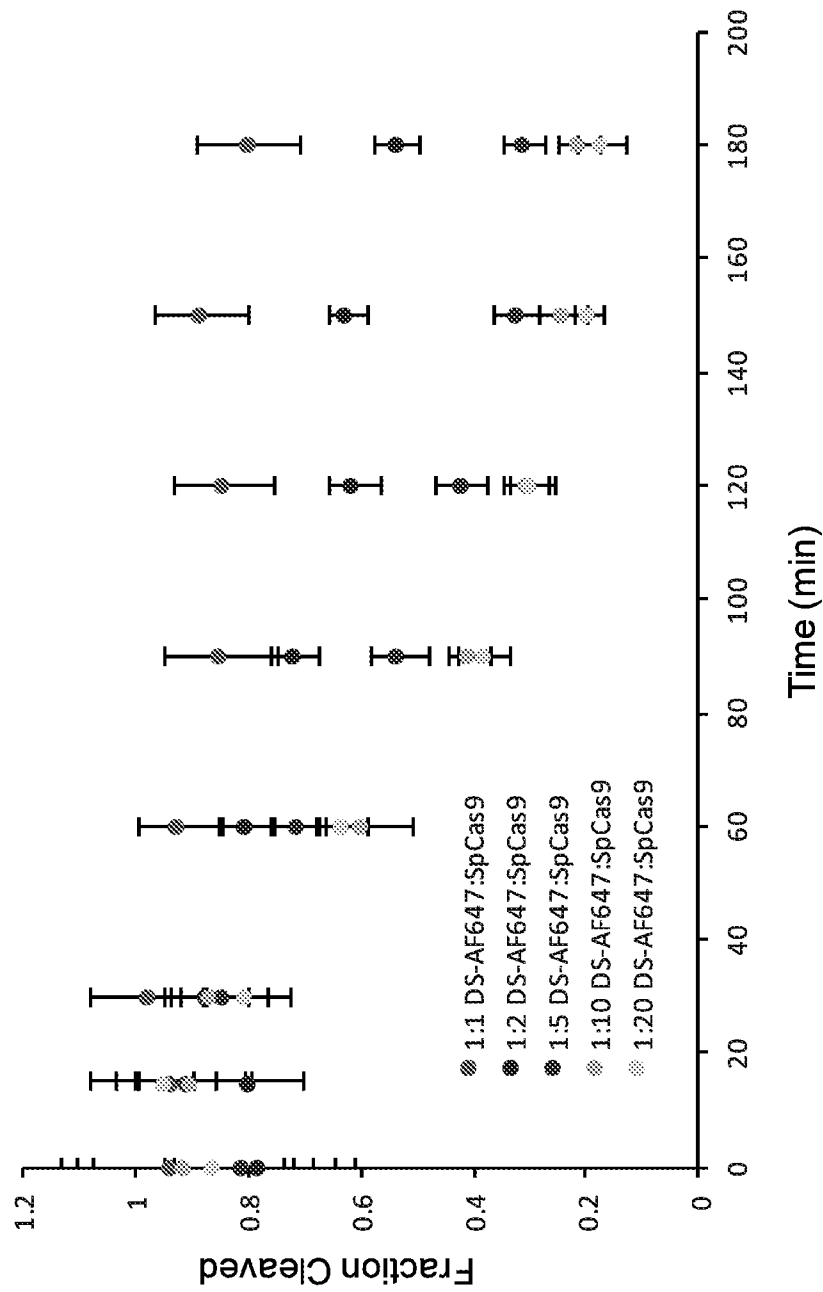


FIG. 6

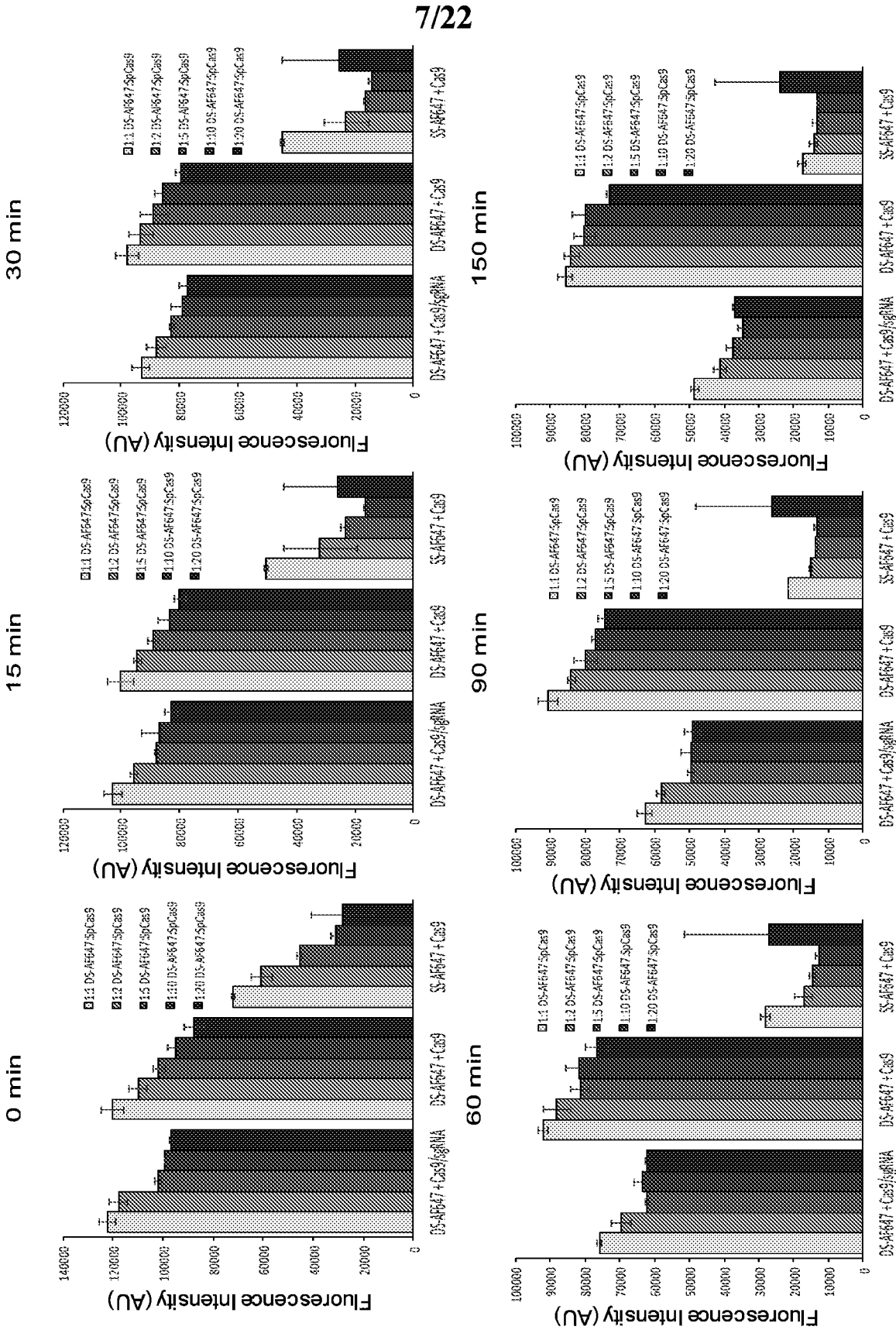


FIG. 7

8/22

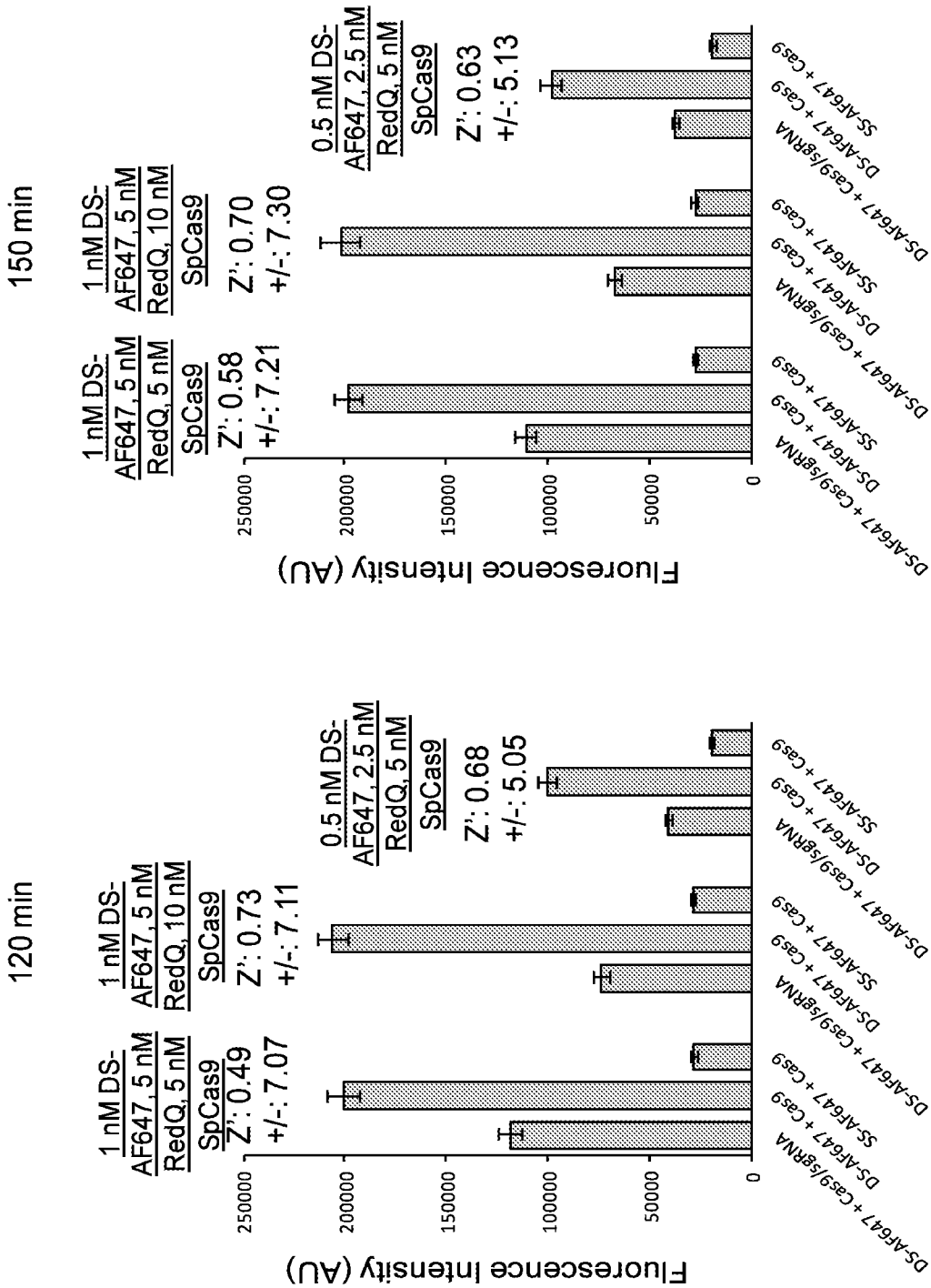


FIG. 8

9/22

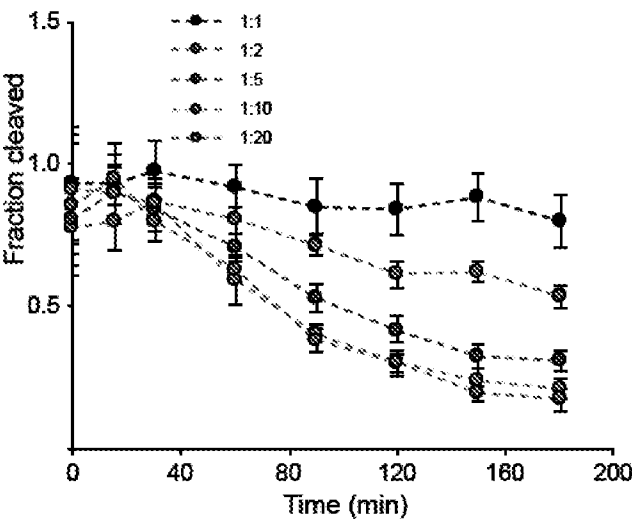


FIG. 9A

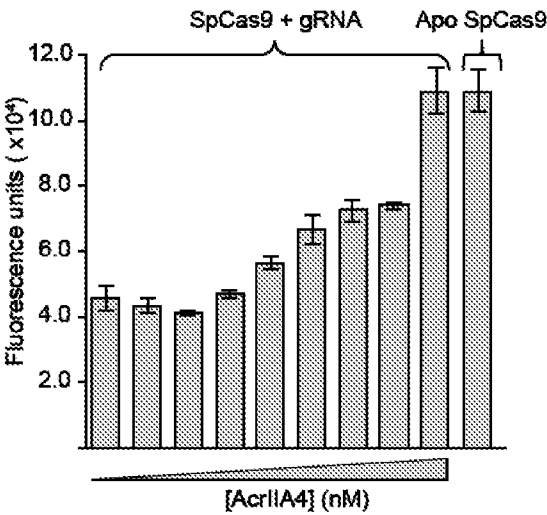


FIG. 9B

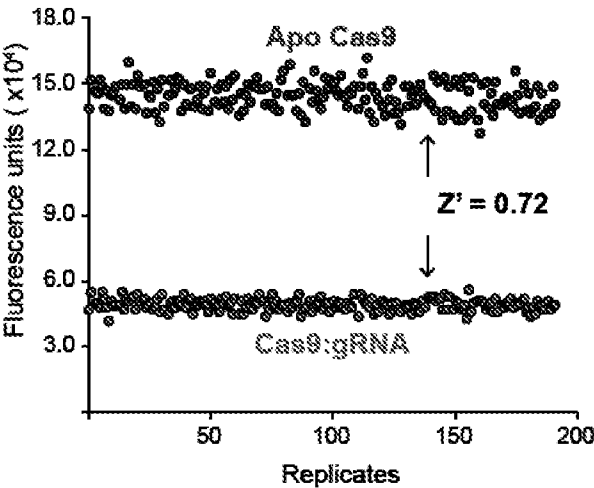


FIG. 9C

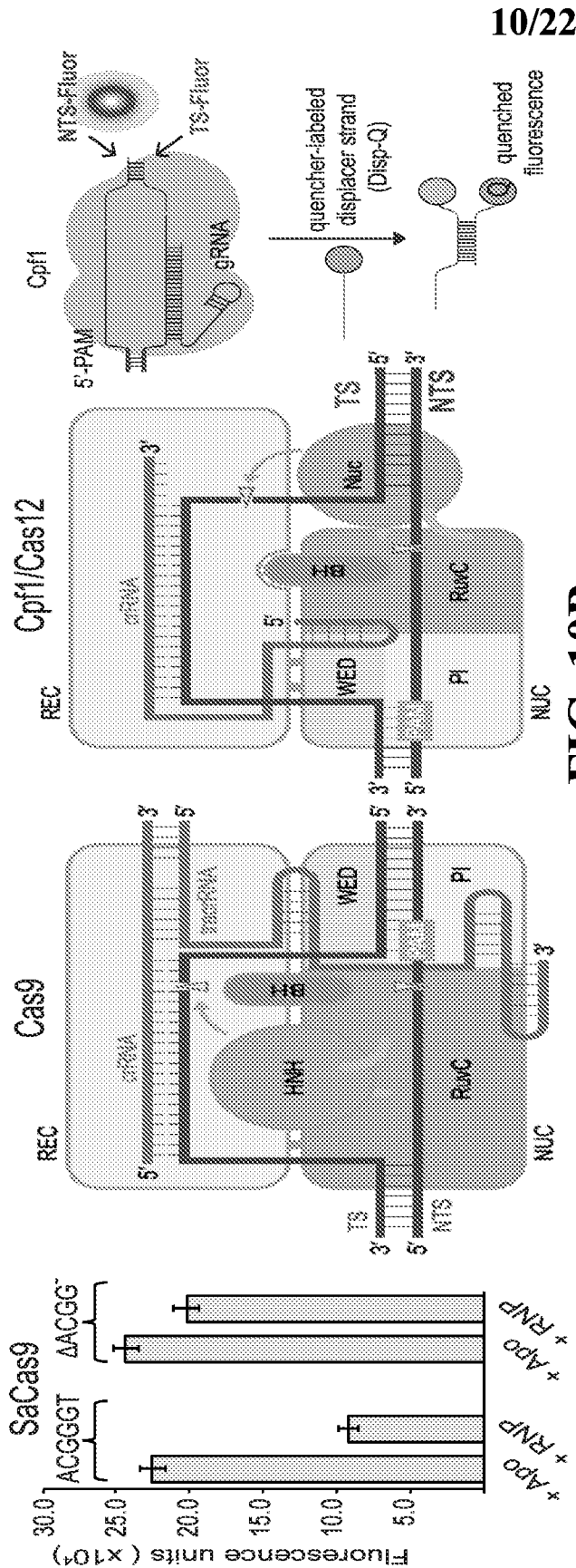


FIG. 10A

FIG. 10B

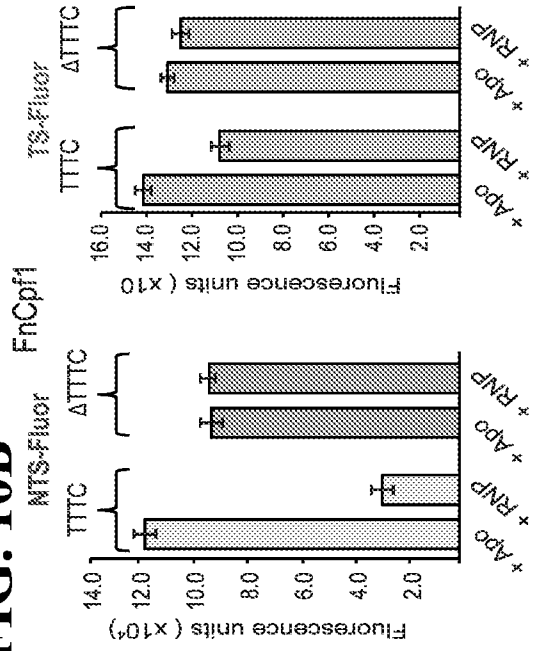


FIG. 10C

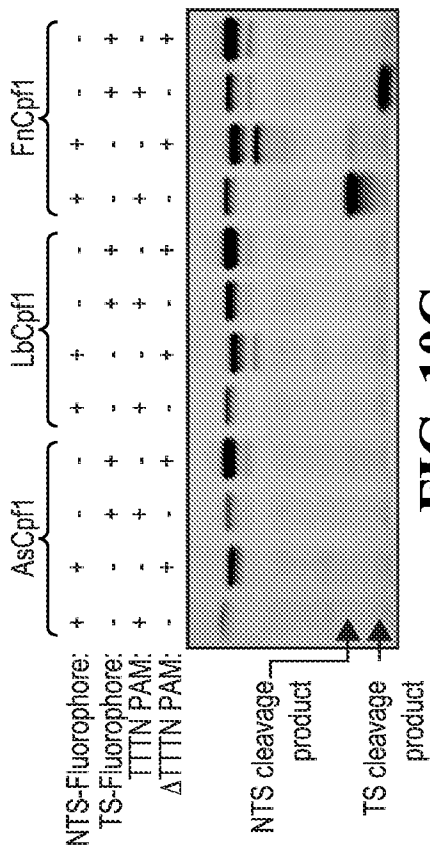


FIG. 10D

11/22

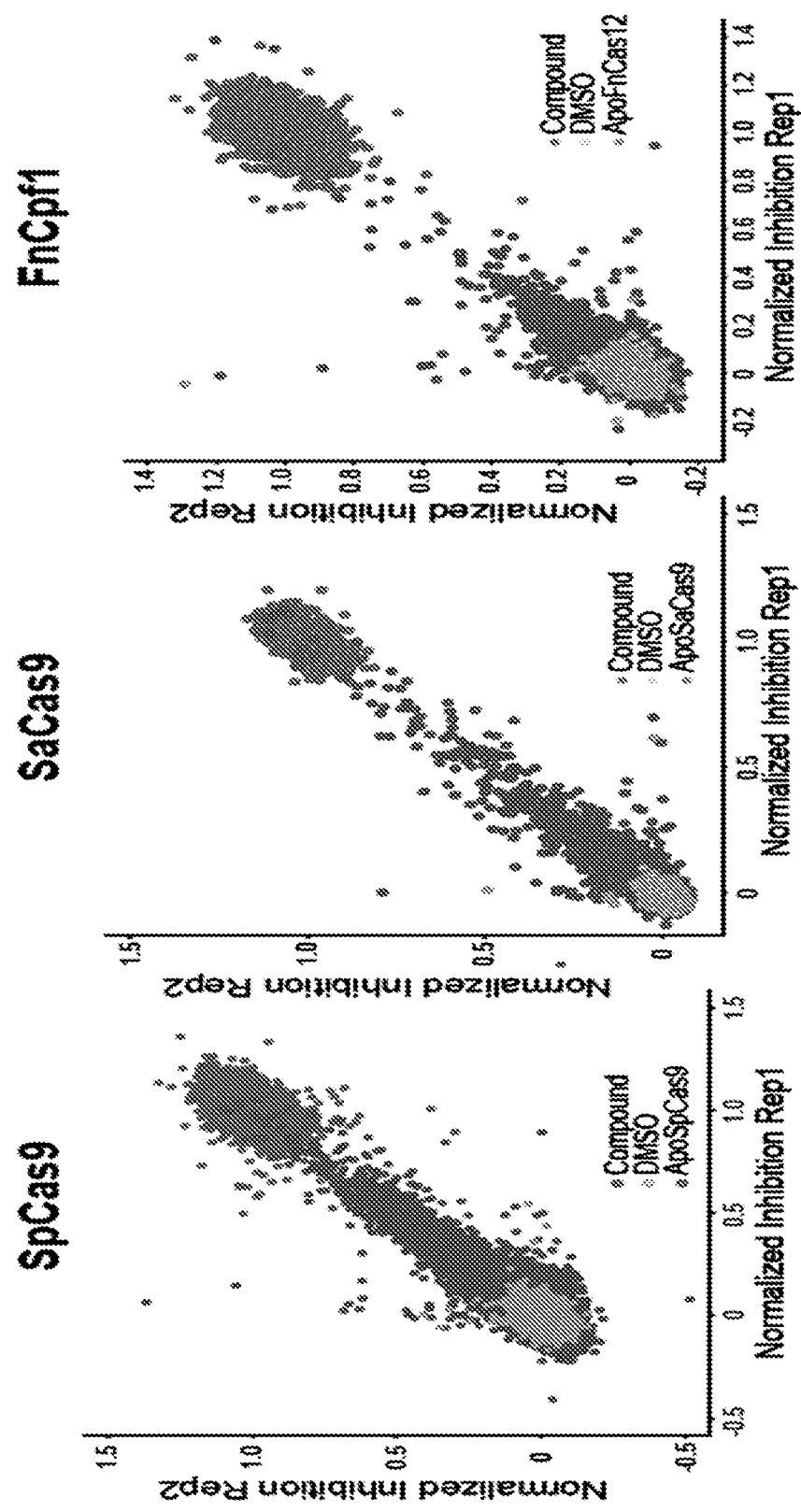
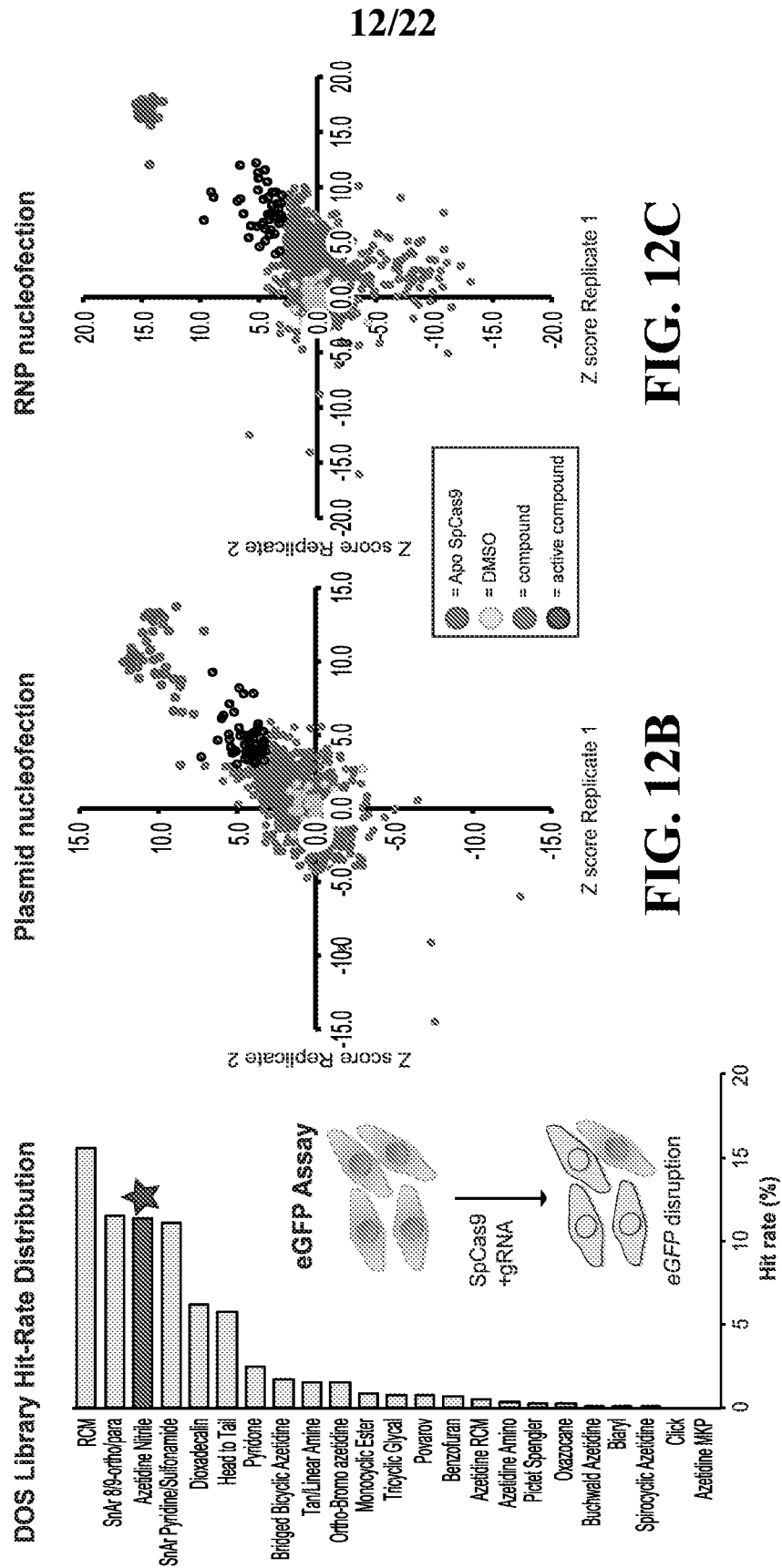
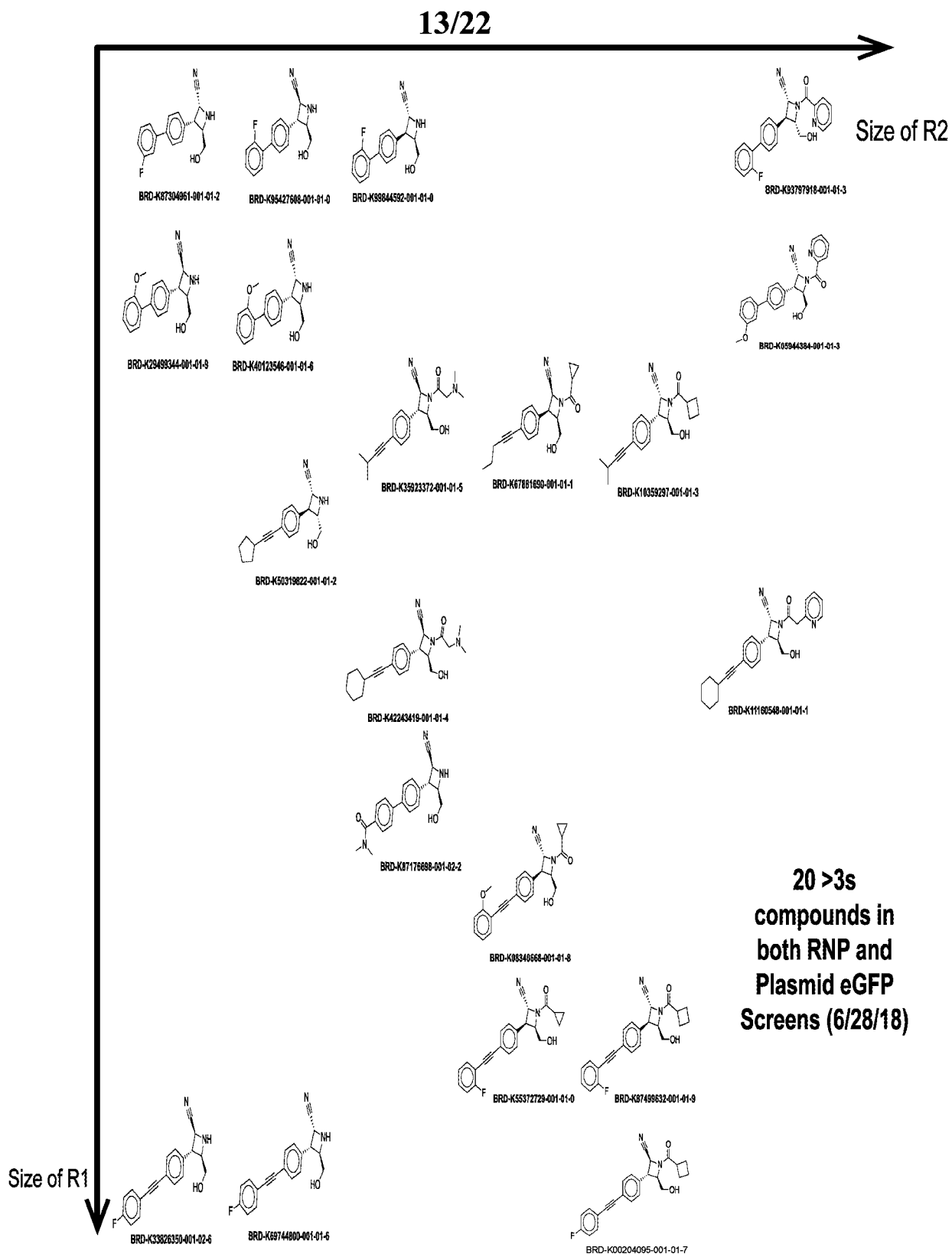


FIG. 11



**FIG. 13**

14/22

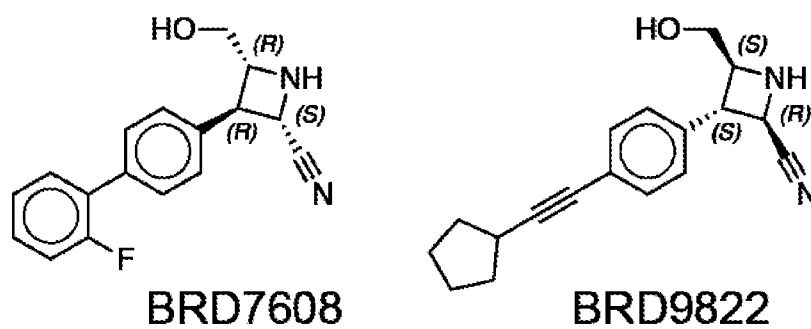


FIG. 14A

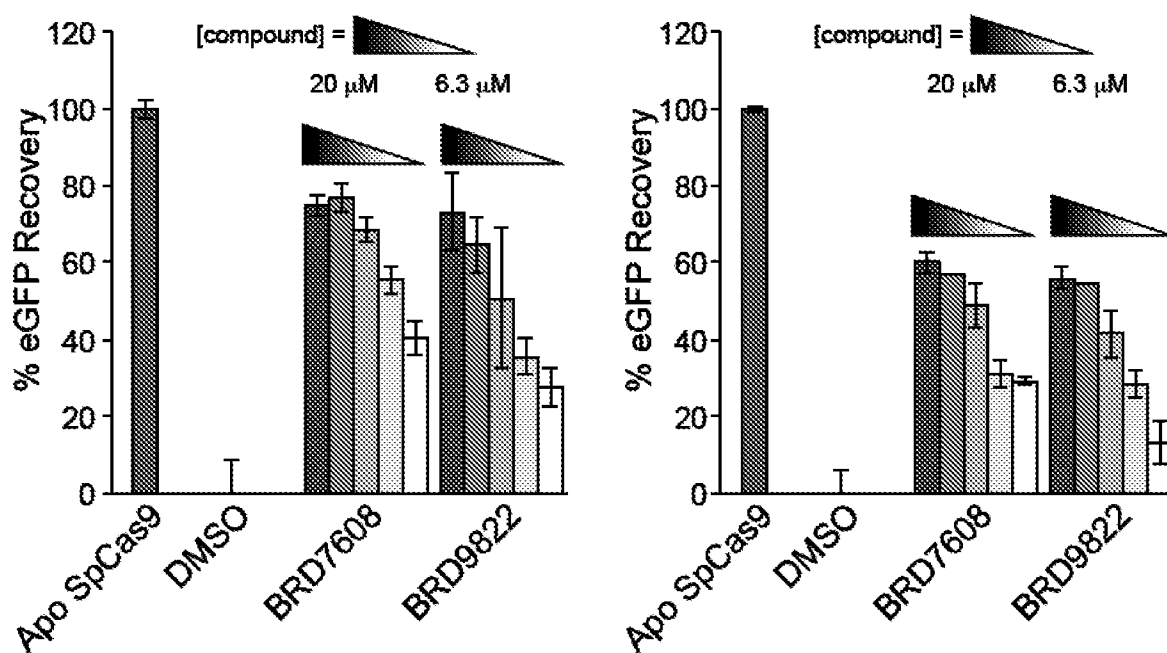
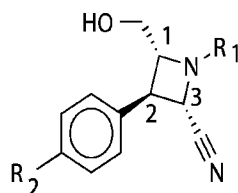


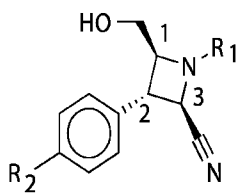
FIG. 14B

15/22

a-d: N-capping R group (R_1)
 1-11: Aryl substituent R group (R_2)



BRD7608

 $R_1 = H$ $R_2 = 5$ 

BRD9822

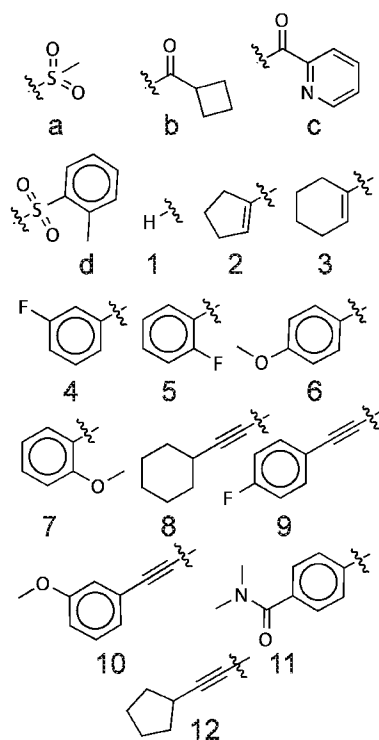
 $R_1 = H$ $R_2 = 12$ 

FIG. 15A

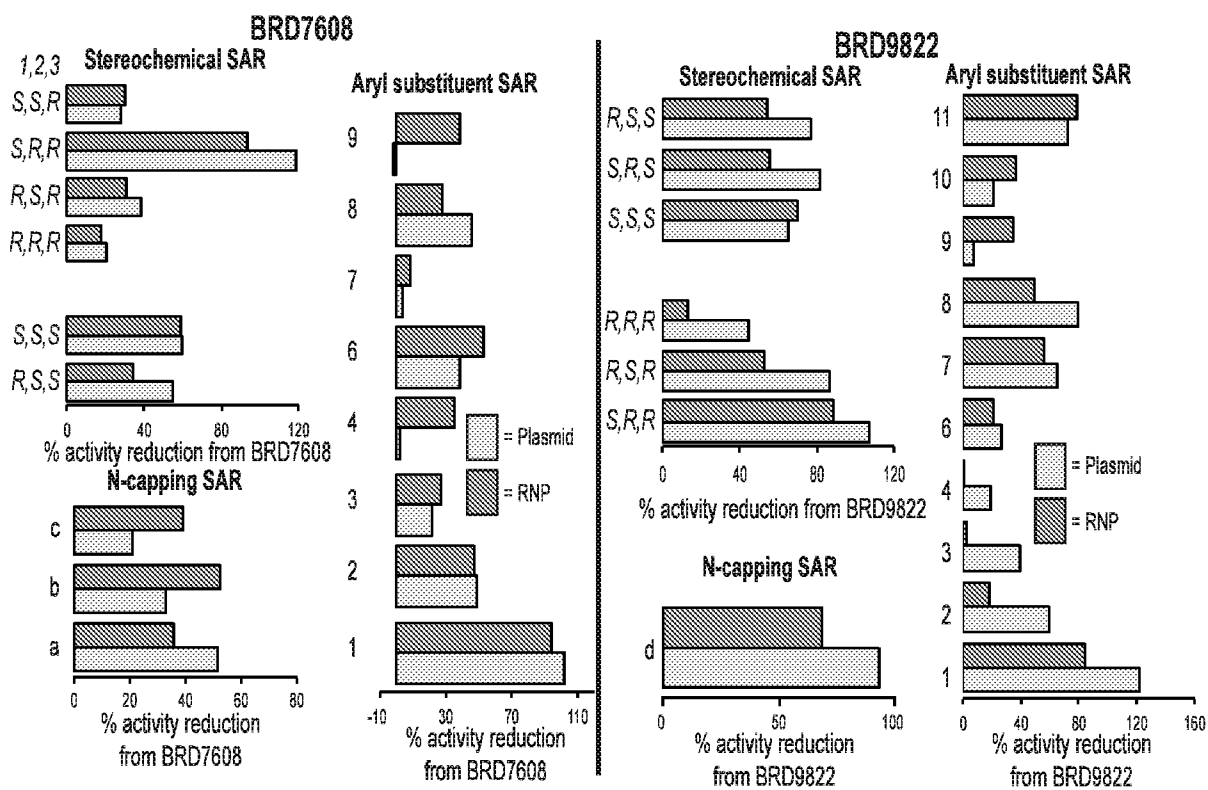


FIG. 15B

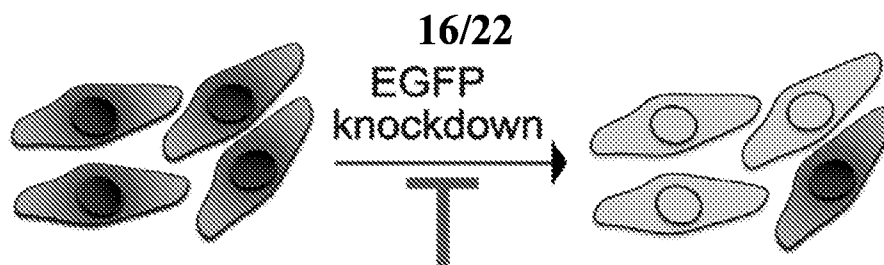


FIG. 16A

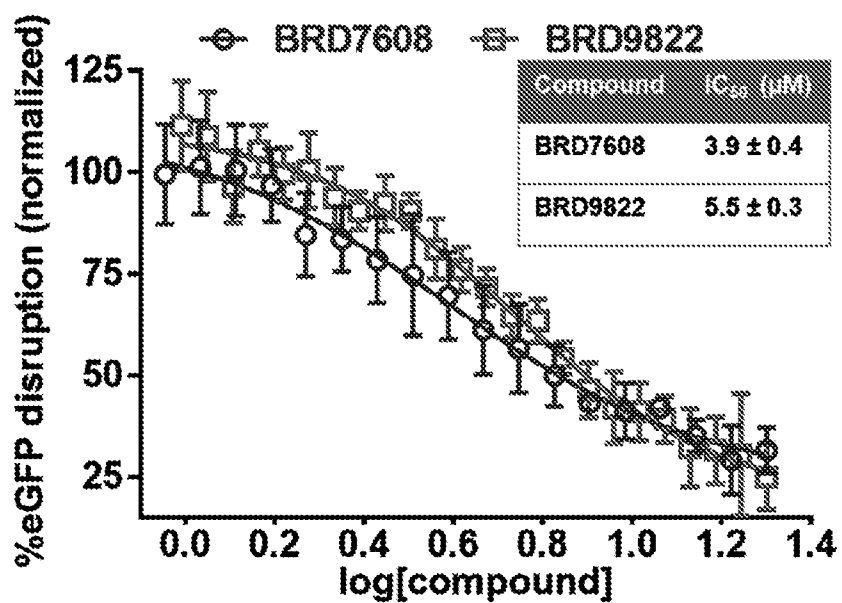


FIG. 16B

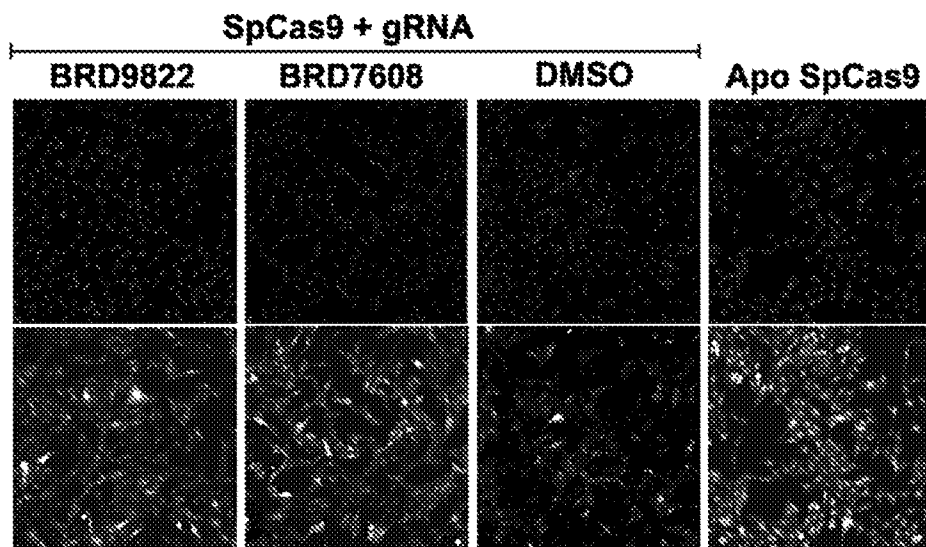


FIG. 16C

17/22

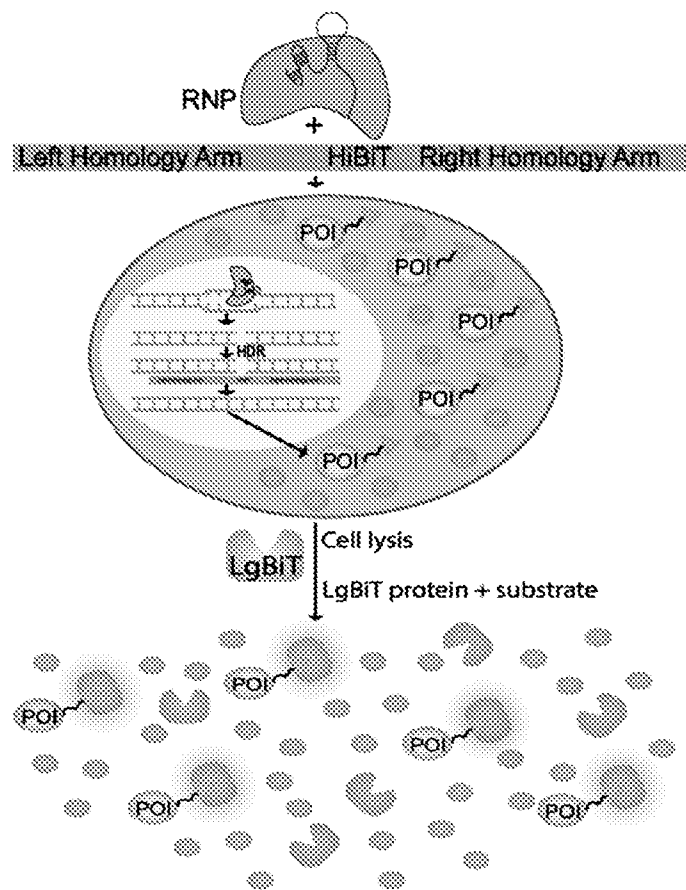


FIG. 17A

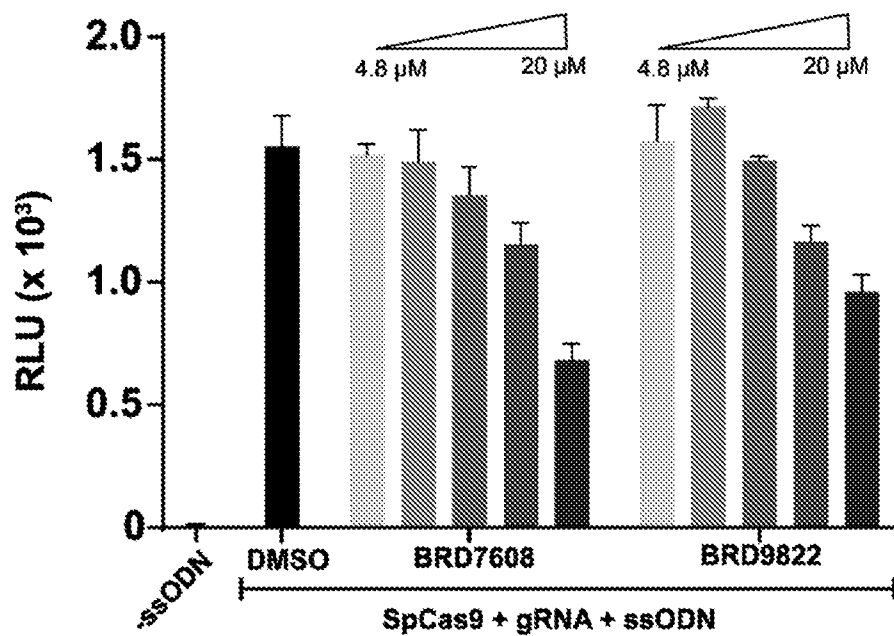


FIG. 17B

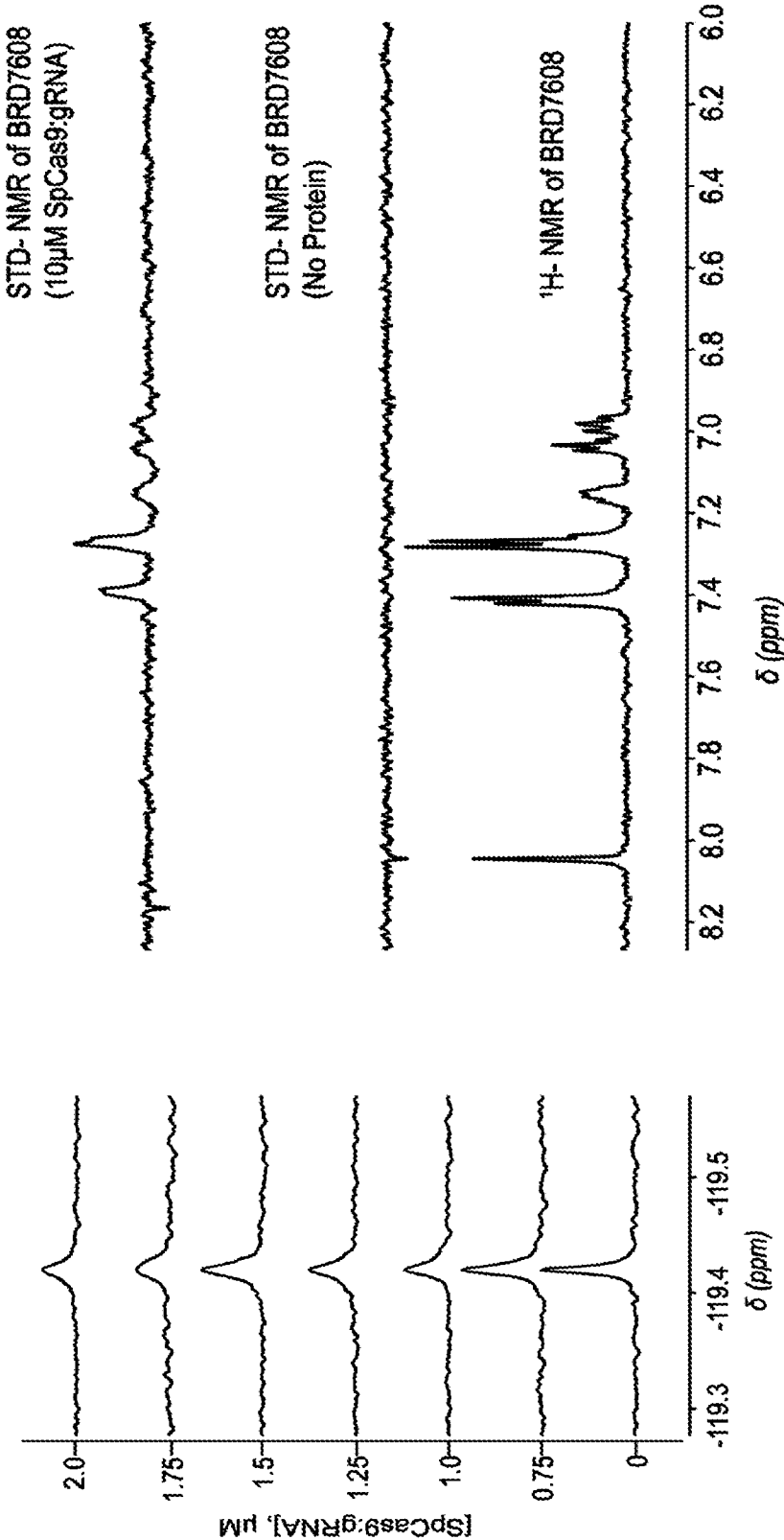


FIG. 18B

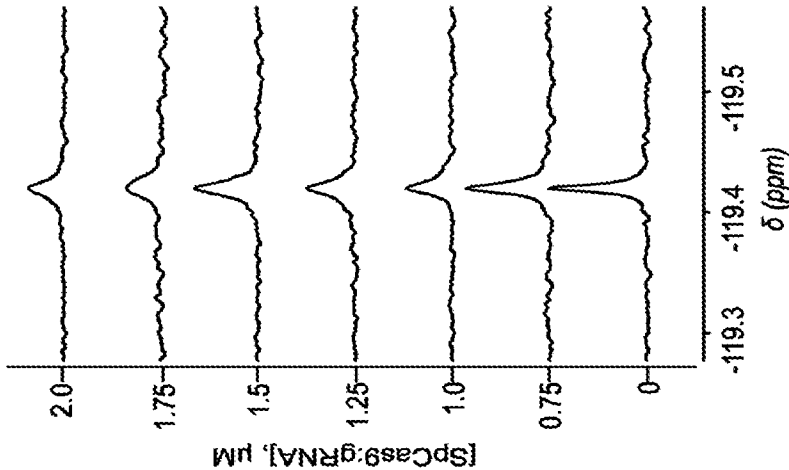


FIG. 18A

19/22

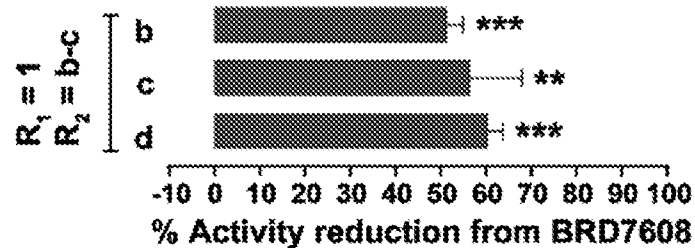
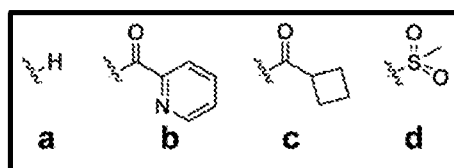
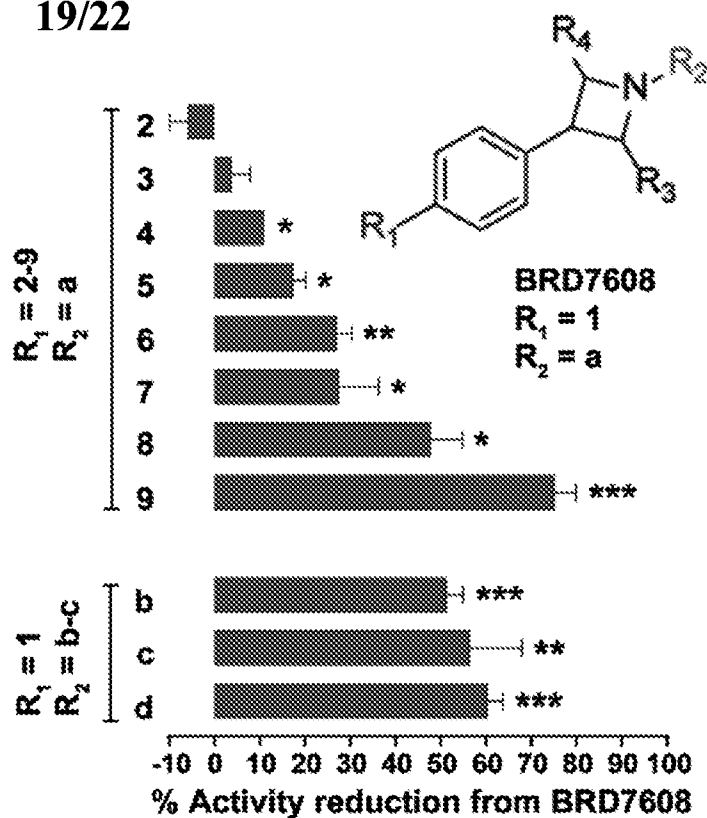
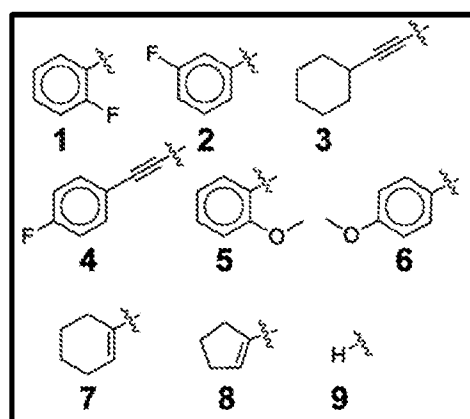


FIG. 19A

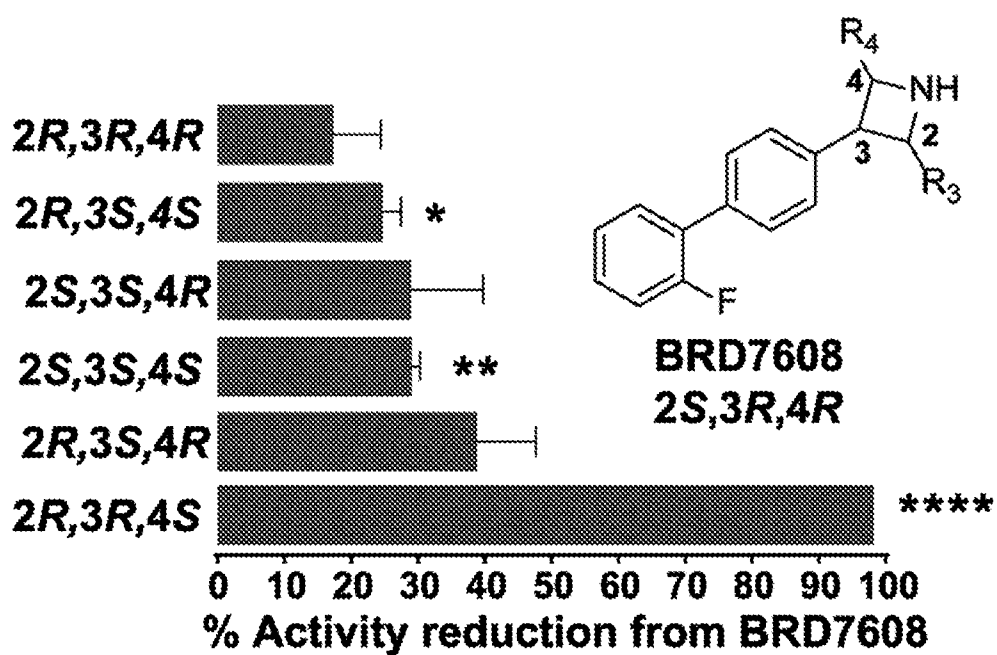


FIG. 19B

20/22

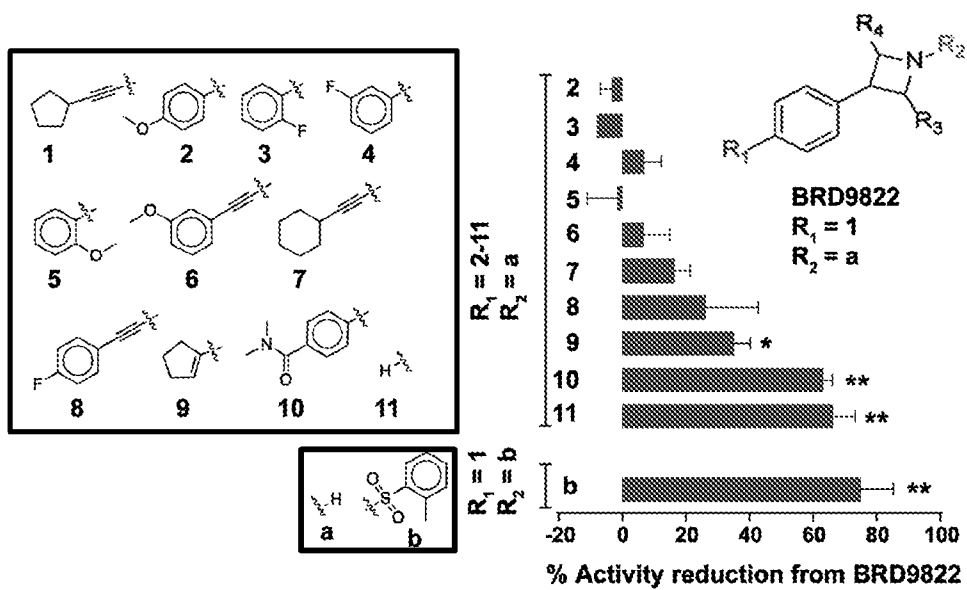


FIG. 20A

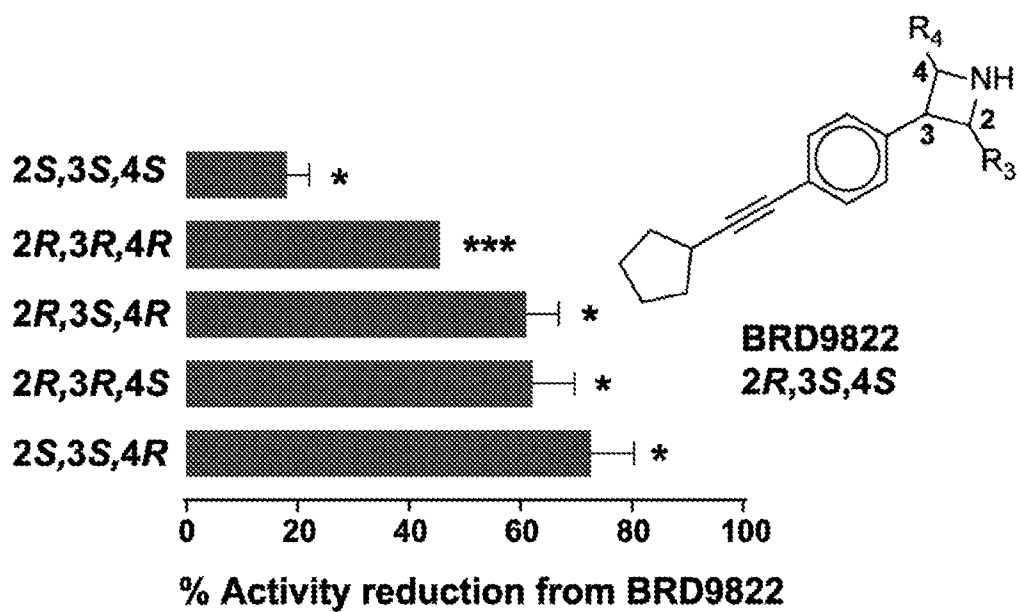


FIG. 20B

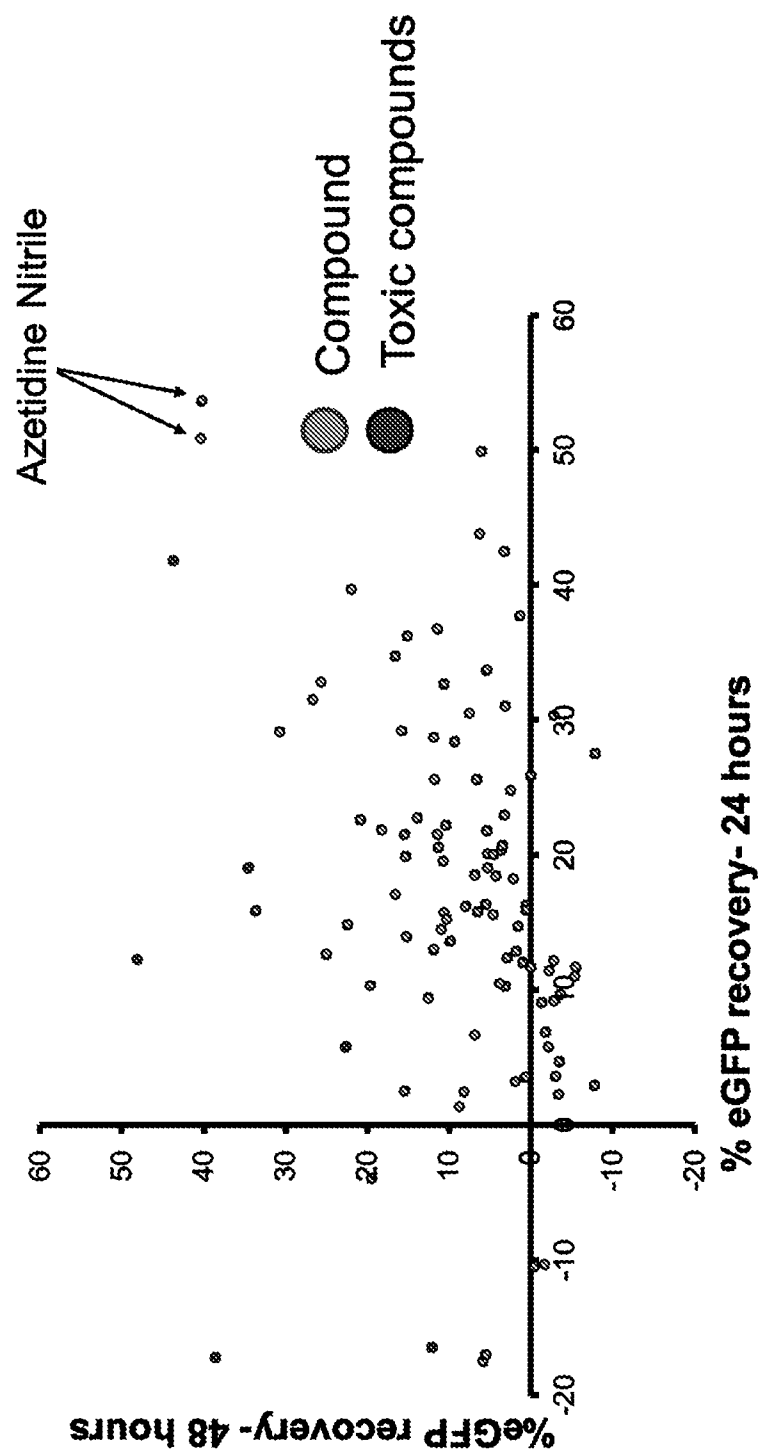


FIG. 21

22/22

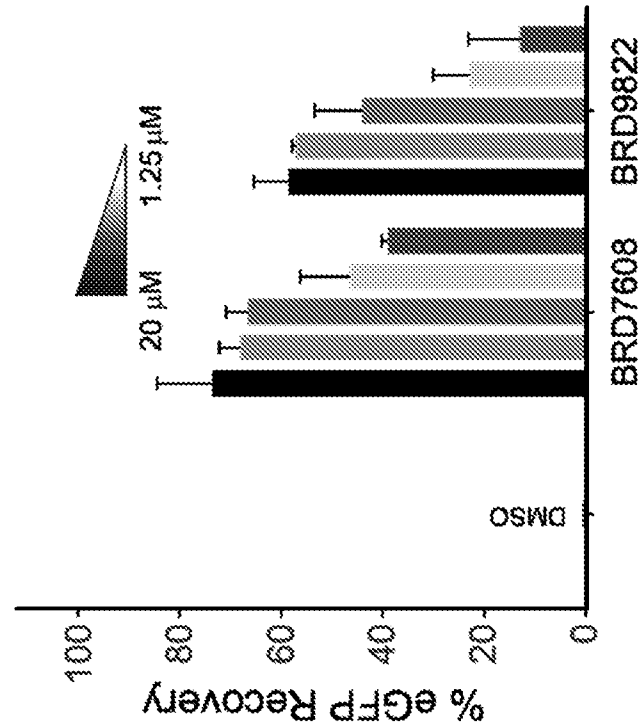


FIG. 22B

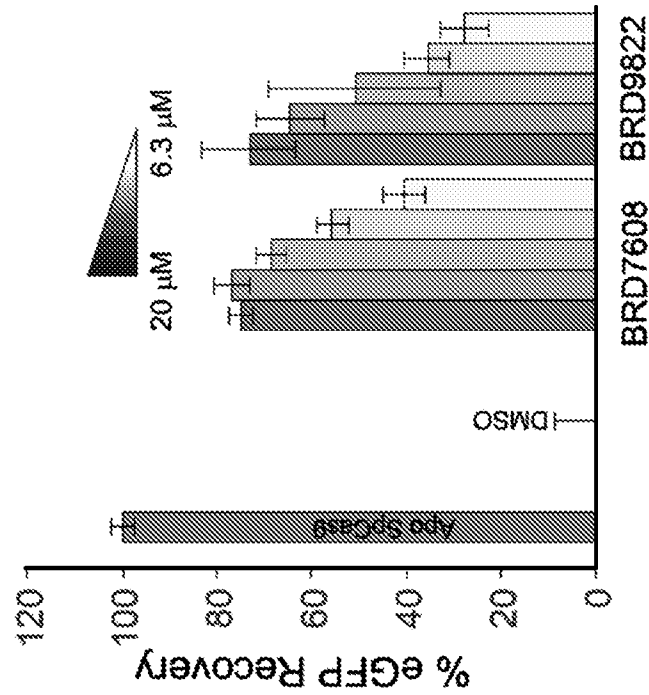


FIG. 22A

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2019/047362

A. CLASSIFICATION OF SUBJECT MATTER
INV. C07D205/04 C12N9/22
ADD.

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

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
C07D C12N

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

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

EPO-Internal, WPI Data, CHEM ABS Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	WO 2018/085288 A1 (HARVARD COLLEGE; THE BRIGHAM AND WOMEN'S HOSPITAL) 11 May 2018 (2018-05-11) the whole document	1-35
X	----- WELLINGTON S ET AL: "A small-molecule allosteric inhibitor of Mycobacterium tuberculosis tryptophan synthase", NATURE CHEMICAL BIOLOGY, vol. 13, no. 9, 3 July 2017 (2017-07-03), pages 943-950, XP055645353, ISSN: 1552-4450, DOI: 10.1038/nchembio.2420 Supplementary data online; compounds 1,3 ----- -/-	24,27



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

28 November 2019

Date of mailing of the international search report

30/01/2020

Name and mailing address of the ISA/

European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040,
Fax: (+31-70) 340-3016

Authorized officer

Elliott, Adrian

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2019/047362

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MAETANI M ET AL: "Discovery of Antimalarial Azetidine-2-carbonitriles That Inhibit P. falciparum Dihydroorotate Dehydrogenase", ACS MEDICINAL CHEMISTRY LETTERS, vol. 8, no. 4, 27 February 2017 (2017-02-27), pages 438-442, XP055645364, ISSN: 1948-5875, DOI: 10.1021/acsmmedchemlett.7b00030 page 440; table 1; compounds 1-6,9,10,13-17,28-30 -----	24
X	WO 2014/028946 A2 (THE BROAD INSTITUTE INC) 20 February 2014 (2014-02-20) page 38 - page 39; compounds 20a, 21a -----	24
X	WO 2016/172631 A2 (HARVARD COLLEGE) 27 October 2016 (2016-10-27) page 209 - page 210; table 4; compounds D1, D2, D7, D9 -----	24
X	KATO N ET AL: "Diversity-oriented synthesis yields novel multistage antimalarial inhibitors", NATURE, vol. 538, no. 7625, 7 September 2016 (2016-09-07), pages 344-349, XP055645378, ISSN: 0028-0836, DOI: 10.1038/nature19804 compounds and their stereoisomers -----	24
X,P	WO 2018/175385 A1 (THE BROAD INSTITUTE INC; HARVARD COLLEGE) 27 September 2018 (2018-09-27) page 119; compound 134 -----	24,27

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US2019/047362

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.:
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:

see additional sheet

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 an 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.:

1-35

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.

FURTHER INFORMATION CONTINUED FROM PCT/ISA/ 210

This International Searching Authority found multiple (groups of) inventions in this international application, as follows:

1. claims: 1-35

A method of inhibiting the activity of an RNA guided endonuclease-guide RNA complex using compounds of formula (I) and the compounds of formula (I).

2. claims: 36-47

A method of detecting endonuclease activity of an RNA guided endonuclease-guide complex comprising: b) contacting the following elements i) the RNA guided endonuclease-guide complex, ii) a double-stranded probe comprising a first oligonucleotide comprising at least one PAM site and a fluorescent reporter at the 5' terminus or the 3' terminus and a second oligonucleotide complementary to the first oligonucleotide, wherein the second oligonucleotide is designed to hybridize to the guide whereby the first oligonucleotide and the second oligonucleotide are cleaved by the endonuclease, and iii) a third oligonucleotide comprising a fluorescence quencher, the fluorescent quencher located at the 3' terminus when the fluorescent reporter is at the 5' terminus of the first oligonucleotide, or at the 5' terminus when the fluorescent reporter is at the 3' terminus of the first oligonucleotide, and designed to hybridize to and displace the cleaved first oligonucleotide from the cleaved second oligonucleotide; and detecting a decrease in fluorescence polarization relative to a reference, thereby detecting nuclease activity of the RNA guided endonuclease-guide complex.

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2019/047362

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2018085288 A1	11-05-2018	US 2019263807 A1	29-08-2019
		WO 2018085288 A1	11-05-2018
WO 2014028946 A2	20-02-2014	US 2015216855 A1	06-08-2015
		WO 2014028946 A2	20-02-2014
WO 2016172631 A2	27-10-2016	US 2018194768 A1	12-07-2018
		WO 2016172631 A2	27-10-2016
WO 2018175385 A1	27-09-2018	AU 2018240031 A1	10-10-2019
		CA 3057087 A1	27-09-2018
		CN 110650738 A	03-01-2020
		EP 3600286 A1	05-02-2020
		KR 20190129968 A	20-11-2019
		SG 11201908691P A	30-10-2019
		WO 2018175385 A1	27-09-2018