



(51) International Patent Classification:

C12N 15/10 (2006.01) C12N 15/63 (2006.01)
C12N 15/62 (2006.01) C12N 9/22 (2006.01)

(21) International Application Number:

PCT/US2017/027126

(22) International Filing Date:

12 April 2017 (12.04.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/322,026 13 April 2016 (13.04.2016) US

(71) Applicant: EDITAS MEDICINE, INC. [US/US]; 11 Hurley Street, Cambridge, MA 02141 (US).

(72) Inventors: COTTA-RAMUSINO, Cecilia; 36 Prince Street, Unit A, Cambridge, MA 02139 (US). JAYARAM, Hariharan; 184 Raymond Street, #4, Cambridge, MA 02140 (US). ZURIS, John, Anthony; 113 Harvard Street, Apt. 7, Cambridge, MA 02139 (US).

(74) Agents: ZACHARAKIS, Maria Laccotripe et al.; Mc-Carter & English, LLP, 265 Franklin Street, Boston, MA 02110 (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, 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).

Declarations under Rule 4.17:

- as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))
- as to the applicant's entitlement to claim the priority of the earlier application (Rule 4.17(iii))

Published:

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

(54) Title: CAS9 FUSION MOLECULES GENE EDITING SYSTEMS, AND METHODS OF USE THEREOF

(57) Abstract: Disclosed herein are enzymatically active Cas9 (eaCas9) fusion molecules, comprising an eaCas9 molecule linked, e.g., covalently or non-covalently, to a template nucleic acid; gene editing systems comprising the eaCas9 fusion molecules, and methods of use thereof.



CAS9 FUSION MOLECULES, GENE EDITING SYSTEMS, AND METHODS OF USE THEREOF

RELATED APPLICATIONS

5 This application claims priority to U.S. Provisional Application No. 62/322,026, filed on April 13, 2016, the entire contents of which are expressly incorporated herein by reference.

SEQUENCE LISTING

10 The instant application contains a Sequence Listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety. Said ASCII copy, created on April 11, 2017, is named EM057PCT1_SL_2017-04-12.txt and is 251 KB in size.

FIELD OF THE INVENTION

15 The invention relates to Cas9 fusion molecules and methods and components for increasing editing of a target nucleic acid sequence by gene correction using an exogenous homologous region, and applications thereof.

BACKGROUND

20 The CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats)/Cas (CRISPR-associated) system evolved in bacteria and archaea as an adaptive immune system to defend against viral attack. Upon exposure to a virus, short segments of viral DNA are integrated into the CRISPR locus. RNA is transcribed from a portion of the CRISPR locus
25 that includes the viral sequence. That RNA, which contains a sequence complimentary to the viral genome, mediates targeting of a Cas9 protein to the sequence in the viral genome. The Cas9 protein cleaves and thereby silences the viral target.

 Recently, the CRISPR/Cas system has attracted widespread interest as a tool for genome editing through the generation of site-specific double strand breaks (DSBs). Current
30 CRISPR/Cas systems that generate site-specific DSBs can be used to edit DNA in eukaryotic cells, *e.g.*, by producing deletions, insertions and/or changes in nucleotide sequence.

 Without wishing to be bound by any theory, it is thought that the mechanism by which an individual DSB is repaired varies depending on whether or not the DNA ends created by the DSB undergo endo- or exonucleolytic processing (also referred to as “end

resection” or “processing”). When no end resection takes place, a DSB is generally repaired by a pathway referred to as classical non-homologous end joining (C-NHEJ). C-NHEJ is considered an “error-prone” pathway inasmuch as it leads in some cases to the formation of small insertions and deletions, though it may also result in perfect repair of DSBs without sequence alterations.

In contrast, if end resection does take place, the ends of a DSB may include one or more overhangs (for example, 3’ overhangs or 5’ overhangs), which can interact with nearby homologous sequences. Again, the mechanism by which the DSB is repaired may vary depending on the extent of processing. When the ends of a DSB undergo relatively limited end resection, the DSB is generally processed by alternative non-homologous end joining (ALT-NHEJ), a class of pathways that includes blunt end-joining (blunt EJ), microhomology mediated end joining (MMEJ), and synthesis dependent micro homology mediated end joining (SD-MMEJ). However, when end resection is extensive, the resulting overhangs may undergo strand invasion of highly homologous sequences (which can be endogenous sequences, for instance from a sister chromatid, or heterologous sequences from an exogenous template), followed by repair of the DSB by a homology-dependent recombination (HDR) pathway.

While a cell could, in theory, repair DNA breaks via any of a number of DNA damage repair pathways, in certain circumstances it is useful or desirable to manipulate the local environment in which the DSB is formed in order to drive a particular mode of repair. For instance, the addition of an exogenous homologous DNA sequence (also referred to as a “donor template” or “template nucleic acid”) to a CRISPR/Cas system may tend to drive repair of DSBs through HDR-based gene correction. However, gene correction strategies that rely on exogenous donor templates are complicated by the potential for interactions between the donor template, the Cas9 and the guide RNA. At the same time, because the donor template is not a naturally occurring part of the CRISPR/Cas complex it may only be present and accessible at a fraction of the DSBs formed by the CRISPR/Cas system, and the desired gene correction may only occur in a fraction of instances.

SUMMARY

This disclosure provides systems, methods and compositions that facilitate gene correction by reconciling the need to localize the donor template at DSBs with the need to prevent interactions between the donor template and the guide RNA or the Cas9. In the various aspects of the disclosure, one or more Cas9 fusion molecules comprising a Cas9

polypeptide linked to a template nucleic acid sequence are utilized to increase the frequency and efficiency of DNA repair of DSBs using gene correction. The Cas9 fusion molecules of the invention comprise Cas9 molecules linked both covalently and non-covalently to template nucleic acids. While not wishing to be bound by theory, it is believed that, by optimizing the length of a linker between the Cas9 polypeptide and the template nucleic acid, hybridization of the template nucleic acid to a gRNA associated with the Cas9 molecule, and/or interactions between the template nucleic acid and DNA binding regions of Cas9, are reduced or even eliminated, while at the same time ensuring that the template nucleic acid is available to participate in HDR, thereby improving the efficiency of gene correction. In some cases, the efficiency of DNA repair via gene correction pathways may be significantly enhanced (*e.g.*, doubled) when the donor template is linked to the Cas9 molecule, as compared to the un-linked molecule. Again, without wishing to be bound by any theory, it is also believed that by linking the donor template to the Cas9, the potential for degradation of the donor template (*e.g.*, during trafficking into the nucleus) is reduced and nuclear localization of the template is improved.

In one aspect, this disclosure relates to compositions and methods for modifying a target nucleic acid in a cell, involving an enzymatically active Cas9 (eaCas9) fusion molecule. For example, the enzymatically active Cas9 (eaCas9) fusion molecule may be an eaCas9 molecule linked to a template nucleic acid.

In one embodiment, the eaCas9 molecule is covalently linked to the template nucleic acid. In certain embodiments, the eaCas9 molecule is covalently linked to the template nucleic acid using a polypeptide linker. In some embodiments, the polypeptide linker has a length sufficient to reduce or prevent hybridization of the template nucleic acid to a gRNA molecule associated with the eaCas9 molecule. In other embodiments, the polypeptide linker is sufficiently long to allow the eaCas9 molecule to bind to a target nucleic acid without steric interference. In other embodiments, the polypeptide linker is sufficiently long to allow the template nucleic acid to interact with the eaCas9 molecule without steric interference.

In certain embodiments, the polypeptide linker is between about 3 and about 100 amino acids in length. In some embodiments, the polypeptide linker is an XTEN linker, *e.g.*, an XTEN linker with the amino acid sequence SGSETPGTSESATPES. In other embodiments, the polypeptide linker is a GGS₉ linker that is 27 amino acids in length. In other embodiments, the polypeptide linker is a GGS₆ linker that is 18 amino acids in length. In another embodiment, the polypeptide linker is a GGS linker that is 3 amino acids in length.

In one embodiment, the eaCas9 molecule is a variant eaCas9 molecule which has been modified to have at least one modification at a surface exposed residue. In some embodiments, the eaCas9 molecule has at least one modification a non-cysteine amino acid residue to a cysteine amino acid residue. In one embodiment, the eaCas9 molecule is a variant eaCas9 molecule which has a surface exposed thiol group.

In one embodiment, a template nucleic acid has a maleimide modification. In another embodiment, a variant eaCas9 molecule, for example, an eaCas9 molecule which has a surface exposed thiol group, is linked to the template nucleic acid having a maleimide modification using thiol coupling.

In one embodiment, an eaCas9 molecule which has a surface exposed thiol group, is linked to the template nucleic acid having an acrydite modification using thiol coupling.

In one embodiment, an eaCas9 molecule has been modified to have a succinimidyl-6-hydrazino-nicotinamide modification. In another embodiment, the succinimidyl-6-hydrazino-nicotinamide modified eaCas9 (*i.e.*, the S-HyNic eaCas9 molecule) molecule is linked a template nucleic acid having a 4Fb modification, using amine coupling.

In one embodiment, an eaCas9 molecule has been modified to have a HaloTag®, a SNAP-tag®, a CLIP-tag®, a ACP-tag®, or a MCP-tag® linked to the eaCas9 molecule. For example, in certain embodiments, an eaCas9-HaloTag molecule is lined to a haloalkane modified template nucleic acid, *e.g.*, using an S_N2 reaction.

In one embodiment, the eaCas9 molecule is covalently linked to the template nucleic acid using a synthetic linker. In certain embodiments, the eaCas9 molecule is a variant eaCas9 molecule which has been modified to have a *N*-[ε-Maleimidocaproic acid] hydrazide (EMCH) and a 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) coupling agent. In other embodiments, the template nucleic acid has a carboxy group, such that the variant eaCas9 molecule is linked to the template nucleic acid by conjugation of the carboxy group on the template nucleic acid to a primary amine of a hydrazine group on the variant Cas9 molecule which has been modified to have a *N*-[ε-Maleimidocaproic acid] hydrazide (EMCH) and a 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) coupling agent.

In one embodiment, the eaCas9 molecule is non-covalently linked to the template nucleic acid. In certain embodiments, the eaCas9 molecule is covalently linked to a first ligand, the template nucleic acid is covalently linked to a second ligand, and the first ligand and the second ligand are non-covalently linked to a ligand acceptor molecule. For example, in some embodiments, the first and second ligands are biotin. In some embodiments, the ligand acceptor molecule is streptavidin. In another embodiment, eaCas9 molecule non-

covalently linked to the template nucleic acid, includes a linker between the eaCas9 molecule and the first ligand. In some embodiments, the linker is sufficiently long to allow the eaCas9 molecule to bind to a target nucleic acid. For example, the linker may be sufficiently long to allow the template nucleic acid to interact with the eaCas9 molecule without steric

5 interference.

In one embodiment, the eaCas9 molecule is covalently linked to a polypeptide, and the polypeptide is non-covalently bound to the template nucleic acid. In some embodiments, the polypeptide is a nucleic acid binding protein. For example, the nucleic acid binding protein may be any one of Rad52, Rad52-yeast, RPA-4 subunit, BRCA2, Rad51, Rad51B, Rad51C, XRCC2, XRCC3, RecA, RadA, HNRNPA1, UP1 Filament of HNRNPA1, NABP2 (SSB1), NABP1 (SSB2), and UHRF1.

In one embodiment, a eaCas9 fusion molecule includes a template nucleic acid having a double stranded nucleic acid sequence or a single stranded nucleic acid sequence.

In one embodiment, a eaCas9 fusion molecule includes a wild-type Cas9 molecule

15 In one embodiment, a eaCas9 fusion molecule includes a Cas9 nickase molecule.

In one embodiment, a eaCas9 fusion molecule includes a split Cas9 molecule or an inducible Cas9 molecule.

In one embodiment, a gene editing system includes at least one eaCas9 fusion molecule having an eaCas9 molecule linked to a template nucleic acid and at least one gRNA molecule. In some embodiments, the gene editing system, which includes at least one gRNA molecule and a eaCas9 fusion molecule, are designed to associate with a target nucleic acid and generate a double strand break on the target nucleic acid. In some embodiments, the double strand break is repaired by at least one DNA repair pathway, thereby producing a modified target nucleic acid.

25 In one embodiment, a gene editing system includes a first eaCas9 fusion molecule, which includes a first Cas9 nickase molecule linked to a first template nucleic acid, a first gRNA molecule, a second eaCas9 fusion molecule, which includes a second Cas9 nickase molecule linked to a second template nucleic acid; and a second gRNA molecule.

In one embodiment, a gene editing system includes a first eaCas9 fusion molecule, which includes a first Cas9 nickase molecule linked to a template nucleic acid and a first gRNA molecule.

In one embodiment, a gene editing system includes a first eaCas9 fusion molecule, which includes a first Cas9 nickase molecule linked to a template nucleic acid, a first gRNA

molecule, a second eaCas9 fusion, which includes a second Cas9 nickase molecule linked to a template nucleic acid and a first gRNA molecule.

In one embodiment, a gene editing system includes a first eaCas9 fusion molecule, which includes a first Cas9 nickase molecule linked to a first template nucleic acid, a first gRNA molecule, a second eaCas9 fusion molecule, which includes a second Cas9 nickase molecule linked to a second template nucleic acid, and a second gRNA molecule. In some embodiments,

the first gRNA molecule and the first eaCas9 fusion molecule are designed to associate with a target nucleic acid and generate a first single strand break on a first strand of the target nucleic acid, and the second gRNA molecule and the second eaCas9 fusion molecule are designed to associate with the target nucleic acid and generate a second single strand break on a second strand of the target nucleic acid, forming a double strand break in the target nucleic acid having a first overhang and a second overhang, such that the double strand break is repaired by at least one DNA repair pathway, thereby producing a modified target nucleic acid.

In one embodiment, each Cas9 nickase molecule has N-terminal RuvC-like domain cleavage activity, but no HNH-like domain cleavage activity. In other embodiments, each Cas9 nickase molecule comprises an amino acid mutation at an amino acid position corresponding to amino acid position N863 of *Streptococcus pyogenes* Cas9.

In one embodiment, each Cas9 nickase molecule has HNH-like domain cleavage activity but no N-terminal RuvC-like domain cleavage activity. In other embodiments, each Cas9 nickase molecule comprises an amino acid mutation at an amino acid position corresponding to amino acid position D10 of *Streptococcus pyogenes* Cas9.

In one embodiment, a gene editing system includes an enzymatically inactive Cas9 (eiCas9) molecule

In one embodiment, a cell includes a gene editing system as disclosed herein.

In one embodiment, a pharmaceutical composition includes a gene editing system as disclosed herein.

In one embodiment, a method of modifying a target nucleic acid in a cell includes contacting a cell with a gRNA molecule and an eaCas9 fusion molecule, which includes an eaCas9 molecule linked to a template nucleic acid. In some embodiments, the gRNA molecule and the eaCas9 fusion molecule associate with the target nucleic acid and generate a double strand break in the target nucleic acid, such that the double strand break is repaired by gene correction using the template nucleic acid of the eaCas9 fusion molecule.

In one embodiment, a method of modifying a target nucleic acid in a cell includes contacting a cell with a first gRNA molecule, a first eaCas9 molecule, a second gRNA molecule; and a second eaCas9 molecule; such that at least one of the first and second eaCas9 molecule is linked to a template nucleic acid, further such that the first gRNA molecule and the first eaCas9 molecule associate with the target nucleic acid and generate a first single strand cleavage event on a first strand of the target nucleic acid. In some embodiments, the second gRNA molecule and the second eaCas9 molecule associate with the target nucleic acid and generate a second single strand cleavage event on a second strand of the target nucleic acid, such that a double strand break having a first overhang and a second overhang is formed, and such that the first overhang and the second overhang in the target nucleic acid are repaired by gene correction using the template nucleic acid. In one embodiment, the first eaCas9 molecule is linked to the template nucleic acid. In another embodiment, both the first eaCas9 molecule and the second eaCas9 molecule are linked to the template nucleic acid.

In one embodiment, each eaCas9 has N-terminal RuvC-like domain cleavage activity, but no HNH-like domain cleavage activity. In another embodiment, each eaCas9 has an amino acid mutation at an amino acid position corresponding to amino acid position N863 of *Streptococcus pyogenes* Cas9.

In one embodiment, each eaCas9 has HNH-like domain cleavage activity but no N-terminal RuvC-like domain cleavage activity. In another embodiment, each eaCas9 has an amino acid mutation at an amino acid position corresponding to amino acid position D10 of *Streptococcus pyogenes* Cas9.

In one embodiment, the cell is a mammalian cell. In other embodiments, the cell is a human cell.

In one embodiment, a cell is altered by any of the methods disclosed or described herein.

In one embodiment, a pharmaceutical composition includes a cell altered by any of the methods disclosed or described herein.

In one embodiment, a nucleic acid molecule encodes at least one eaCas9 fusion molecule, which includes an eaCas9 molecule and a polypeptide. In some embodiments, the polypeptide is a nucleic acid binding protein. In some embodiments, the nucleic acid binding protein may be any one of Rad52, Rad52-yeast, RPA-4 subunit, BRCA2, Rad51, Rad51B, Rad51C, XRCC2, XRCC3, RecA, RadA, HNRNPA1, UP1 Filament of HNRNPA1, NABP2 (SSB1), NABP1 (SSB2), or UHRF1.

In one embodiment, a vector includes any of the nucleic acid molecules disclosed or described herein.

In one embodiment, a method of modifying a target nucleic acid in a cell, includes contacting the cell with a gRNA molecule and an eaCas9 fusion molecule, which includes an
5 eaCas9 molecule linked to a template nucleic acid, such that the gRNA molecule and the eaCas9 fusion molecule associate with the target nucleic acid and generate a double strand break in the target nucleic acid. In some embodiments, a first overhang and a second overhang in the target nucleic acid are repaired by gene correction using the template nucleic acid in the eaCas9 fusion molecule, thereby modifying the target nucleic acid in the cell.

10 In one embodiment, a method of modifying a target nucleic acid in a cell includes contacting the cell with a first gRNA molecule, a first eaCas9 fusion molecule, which includes a first Cas9 nickase molecule linked to a first template nucleic acid, a second gRNA molecule, and a second eaCas9 fusion molecule, which includes a second Cas9 nickase molecule linked to a second template nucleic acid. In some embodiments, the first gRNA
15 molecule and the first eaCas9 fusion molecule associate with the target nucleic acid and generate a first single strand cleavage event on a first strand of the target nucleic acid, and the second gRNA molecule and the second eaCas9 fusion molecule associate with the target nucleic acid and generate a second single strand cleavage event on a second strand of the target nucleic acid, forming a double strand break having a first overhang and a second
20 overhang. In certain embodiments, the first overhang and the second overhang in the target nucleic acid are repaired by gene correction using the first and second template nucleic acid, thereby modifying the target nucleic acid in the cell.

Headings, including numeric and alphabetical headings and subheadings, are for organization and presentation and are not intended to be limiting.

25 Other features and advantages of the invention will be apparent from the detailed description, drawings, and from the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 depicts a scheme of generating a Cas9 fusion molecule by covalent conjugation
30 of a single- or multi-cysteine variant Cas9 protein to a template nucleic acid that contains a 5'-maleimide-modification.

Fig. 2 depicts the results of an electrophoretic mobility shift assay of the products from reacting a 5'-maleimide-modified template nucleic acid with a cysteine-variant Cas9

protein molecule as described in Example 1. The data indicate the generation of Cas9 fusion molecules containing at least one covalently attached template nucleic acid.

Fig. 3 depicts a Cas9 fusion molecule formed by covalent conjugation of a Cas9-HaloTag protein molecule to a template nucleic acid. The single stranded oligoDNA (ssDNA) is covalently attached to the HaloTag component of the Cas9-HaloTag protein molecule.

Fig. 4A depicts results of the pre-conjugation and pre-annealed methods for conjugating a Cas9-HaloTag protein fusion to template sequence using standard PAGE analysis for various Cas9 protein fusions provided in Table 9.

Fig. 4B depicts results of the pre-conjugation and pre-annealed methods for conjugating a Cas9-HaloTag protein fusion to template sequence using standard PAGE analysis for the HGC Cas9 protein fusion provided in Table 9.

Fig. 5 depicts the normalized percentages of HDR editing efficiency using a nucleofection assay in U2OS cells. The results compare results obtained using a Cas9 protein fusion provided in Table 9 conjugated to a template nucleic acid with a Cas9-HaloTag protein molecule that is not conjugated to a template nucleic acid. The results demonstrate that the Cas9 fusion molecules increase HDR efficiency when conjugated to template nucleic acid compared to reactions performed with unconjugated template nucleic acid.

Fig. 6 depicts a scheme of generating a Cas9 fusion molecule by covalent conjugation of a single- or multi-cysteine variant Cas9 protein molecule to a template nucleic acid that contains an acrydite-modified template nucleic acid.

Fig. 7 depicts a scheme generating a Cas9 fusion molecule by covalent conjugation of a N-[ε-Maleimidocaproic acid] hydrazide (EMCH)-modified Cas9 protein molecule with a 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)-modified template nucleic acid.

Fig. 8 depicts a scheme generating a Cas9 fusion molecule by non-covalent conjugation of a Cas9 protein molecule, covalently linked to biotin, and a template nucleic acid, covalently linked to biotin, via the interaction of the biotin moiety of the Cas9 protein molecule and the biotin moiety of the template nucleic acid with streptavidin.

DETAILED DESCRIPTION

In order that the invention is understood, certain terms are herein defined.

Definitions

“Alt-HDR” or “alternative HDR,” or alternative homology-directed repair, as used herein, refers to the process of repairing DNA damage using a homologous nucleic acid (*e.g.*,

an endogenous homologous sequence, *e.g.*, a sister chromatid, or an exogenous nucleic acid, *e.g.*, a template nucleic acid). Alt-HDR is distinct from canonical HDR in that the process utilizes different pathways from canonical HDR, and can be inhibited by the canonical HDR mediators, RAD51 and BRCA2. Also, alt-HDR uses a single-stranded or nicked homologous nucleic acid for repair of the break.

“ALT-NHEJ” or “alternative NHEJ”, or alternative non-homologous end joining, as used herein, is a type of alternative end joining repair process, and utilizes a different pathway than that of canonical NHEJ. In alternative NHEJ, a small degree of resection occurs at the break ends on both sides of the break to reveal single-stranded overhangs.

Ligation or annealing of the overhangs results in the deletion of sequence. ALT-NHEJ is a category that includes microhomology-mediated end joining (MMEJ), blunt end joining (EJ), and synthesis-dependent microhomology-mediated end joining (SD-MMEJ). In MMEJ, microhomologies, or short spans of homologous sequences, *e.g.*, 5 nucleotides or more, on the single-strand are aligned to guide repair, and leads to the deletion of sequence between the microhomologies.

“Amino acids” as used herein encompasses the canonical amino acids as well as analogs thereof. “Canonical HDR,” or canonical homology-directed repair, as used herein, refers to the process of repairing DNA damage using a homologous nucleic acid (*e.g.*, an endogenous homologous sequence, *e.g.*, a sister chromatid, or an exogenous nucleic acid, *e.g.*, a template nucleic acid). Canonical HDR typically acts when there has been significant resection at the double-strand break, forming at least one single stranded portion of DNA. In a normal cell, HDR typically involves a series of steps such as recognition of the break, stabilization of the break, resection, stabilization of single stranded DNA, formation of a DNA crossover intermediate, resolution of the crossover intermediate, and ligation. The process requires RAD51 and BRCA2, and the homologous nucleic acid is typically double-stranded.

“Canonical NHEJ”, or canonical non-homologous end joining, as used herein, refers to the process of repairing double-strand breaks in which the break ends are directly ligated. This process does not require a homologous nucleic acid to guide the repair, and can result in deletion or insertion of one or more nucleotides. This process requires the Ku heterodimer (Ku70/Ku80), the catalytic subunit of DNA-PK (DN-PKcs), and/or DNA ligase XRCC4/LIG4. Unless indicated otherwise, the term “HDR” as used herein encompasses canonical HDR and alt-HDR.

A "Cas9 molecule," as used herein, refers to a Cas9 polypeptide or a nucleic acid encoding a Cas9 polypeptide. A "Cas9 polypeptide" is a polypeptide that can interact with a gRNA molecule and, in concert with the gRNA molecule, localize to a site comprising a target domain and, in certain embodiments, a PAM sequence. Cas9 molecules include both naturally occurring Cas9 molecules and Cas9 molecules and engineered, altered, or modified Cas9 molecules or Cas9 polypeptides that differ, *e.g.*, by at least one amino acid residue, from a reference sequence, *e.g.*, the most similar naturally occurring Cas9 molecule. (The terms altered, engineered or modified, as used in this context, refer merely to a difference from a reference or naturally occurring sequence, and impose no specific process or origin limitations.) A Cas9 molecule may be a Cas9 polypeptide or a nucleic acid encoding a Cas9 polypeptide. A Cas9 molecule may be a nuclease (an enzyme that cleaves both strands of a double-stranded nucleic acid), a nickase (an enzyme that cleaves one strand of a double-stranded nucleic acid), or an enzymatically inactive (or dead) Cas9 molecule. A Cas9 molecule having nuclease or nickase activity is referred to as an "enzymatically active Cas9 molecule" (an "eaCas9" molecule). A Cas9 molecule lacking the ability to cleave target nucleic acid is referred to as an "enzymatically inactive Cas9 molecule" (an "eiCas9" molecule). A Cas9 molecule may also be a split Cas9 molecule or an inducible Cas9 molecule, as described in more detail below.

In certain embodiments, a Cas9 molecule meets one or both of the following criteria: it has at least 20, 30, 40, 50, 55, 60, 65, 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% homology with, or it differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 35, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350 or 400, amino acid residues from, the amino acid sequence of a reference sequences, *e.g.*, naturally occurring Cas9 molecule.

In certain embodiments, a Cas9 molecule meets one or both of the following criteria: it has at least 20, 30, 40, 50, 55, 60, 65, 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% homology with, or it differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 25, 30, 35, 40, 35, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 100, 150, 200, 250, 300, 350 or 400, amino acid residues from, the amino acid sequence of a reference sequences, *e.g.*, naturally-occurring Cas9 molecule.

In certain embodiments, each domain of the Cas9 molecule (*e.g.*, the domains named herein) will, independently have: at least 20, 30, 40, 50, 55, 60, 65, 70, 75, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% homology with such a domain described herein. In certain embodiments at least 1, 2, 3, 4, 5, of 6 domains will

have, independently, at least 50, 60, 70, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100% homology with a corresponding domain, while any remaining domains will be absent, or have less homology to their corresponding naturally occurring domains.

5 In certain embodiments, the Cas9 molecule is a *S. pyogenes* Cas9 variant. In certain embodiments, the Cas9 variant is the EQR variant. In certain embodiments, the Cas9 variant is the VRER variant. In certain embodiments, the eiCas9 molecule is a *S. pyogenes* Cas9 variant. In certain embodiments, the Cas9 variant is the EQR variant. In certain
10 embodiments, the Cas9 variant is the VRER variant. In certain embodiments, a Cas9 system comprises a Cas9 molecule, *e.g.*, a Cas9 molecule described herein, *e.g.*, the Cas9 EQR variant or the Cas9 VRER variant.

 In certain embodiments, the Cas9 molecule is a *S. aureus* Cas9 variant. In certain
15 embodiments, the Cas9 variant is the KKH (E782K/N968K/R1015H) variant (see, *e.g.*, Kleinstiver 2015, the entire contents of which are expressly incorporated herein by
20 reference). In certain embodiments, the Cas9 variant is the E782K/K929R/R1015H variant (see, *e.g.*, Kleinstiver 2015). In certain embodiments, the Cas9 variant is the E782K/K929R/N968K/R1015H variant (see, *e.g.*, Kleinstiver 2015). In certain embodiments the Cas9 variant comprises one or more mutations in one of the following residues: E782, K929, N968, R1015. In certain embodiments the Cas9 variant comprises one or more of the
25 following mutations: E782K, K929R, N968K, R1015H and R1015Q (see, *e.g.*, Kleinstiver 2015). In certain embodiments, a Cas9 system comprises a Cas9 molecule, *e.g.*, a Cas9 molecule described herein, *e.g.*, the Cas9 KKH variant.

 A “Cas9 fusion molecule”, “Cas9 fusion protein”, or “Cas9 fusion”, as used herein,
25 refers to a chimeric protein comprising a Cas9 molecule, *e.g.*, Cas9 protein or Cas9 polypeptide, or a fragment thereof, linked to a template nucleic acid. In some embodiments, the template nucleic acid is a nucleic acid, *e.g.*, DNA or RNA. In some embodiments, the template nucleic acid is single-stranded or double-stranded. In some embodiments the template nucleic acid is circular nucleic acid, while in other embodiments the template nucleic acid is linear nucleic acid. In certain embodiments, the Cas9 fusion molecule
30 comprises a Cas9 molecule covalently linked to a template nucleic acid. In other embodiments, the Cas9 fusion molecule comprises a Cas9 molecule non-covalently linked to a template nucleic acid. In certain embodiments, a Cas9 fusion molecule is linked to more than one template nucleic acid.

As used herein, the term “Cas9 system” or “gene editing system” refers to a system capable of altering a target nucleic acid by one of many DNA repair pathways. In certain embodiments, the Cas9 system described herein promotes repair of a target nucleic acid via an HDR pathway. In some embodiments, a Cas9 system comprises a gRNA and a Cas9 molecule. In other embodiments, a Cas9 system comprises a gRNA and a Cas9 fusion molecule. In some embodiments, a Cas9 system further comprises a second gRNA. In yet another embodiment, a Cas9 system comprises a gRNA, a Cas9 molecule, and a second gRNA. In yet other embodiment, a Cas9 system comprises a gRNA, a Cas9 fusion molecule, and a second gRNA. In some embodiments, a Cas9 system comprises a gRNA, two Cas9 molecules, and a second gRNA. In some embodiments, a Cas9 system comprises a gRNA, two Cas9 fusion molecules, and a second gRNA. In some embodiments, a Cas9 system comprises a first gRNA, a second gRNA, a first Cas9 molecule, and a second Cas9 molecule. In other embodiments, a Cas9 system comprises a first gRNA, a second gRNA, a first Cas9 fusion molecule, and a second Cas9 fusion molecule. In some embodiments, a Cas9 system further comprises a template nucleic acid. In other embodiments, a Cas9 system further comprises a template nucleic acid provided by the Cas9 fusion molecule.

As used herein, the term “cleavage event” refers to a break in a nucleic acid molecule. A cleavage event may be a single-strand cleavage event, or a double-strand cleavage event. A single-strand cleavage event may result in a 5’ overhang or a 3’ overhang. A double-stranded cleavage event may result in blunt ends, two 5’ overhangs, or two 3’ overhangs.

A disorder “caused by” a mutation, as used herein, refers to a disorder that is made more likely or severe by the presence of the mutation, compared to a subject that does not have the mutation. The mutation need not be the only cause of a disorder, *i.e.*, the disorder can still be caused by the mutation even if other causes, such as environmental factors or lifestyle factors, contribute causally to the disorder. In embodiments, the disorder is caused by the mutation if the mutation is a medically recognized risk factor for developing the disorder, and/or if a study has found that the mutation contributes causally to development of the disorder.

The term “covalent”, as used herein, refers to a form of chemical bonding characterized by the sharing of one or more pairs of electrons between two components, producing a mutual attraction that holds the two components together. The sharing of the one or more pairs of electrons between two components may either be direct (*e.g.*, via reactive groups on the surface the two components, *e.g.*, a Cas9 polypeptide and a template nucleic acid) or indirect (via a linker molecule).

“Derived from”, as used herein, refers to the source or origin of a molecular entity, *e.g.*, a nucleic acid or protein. The source of a molecular entity may be naturally-occurring, recombinant, unpurified, or a purified molecular entity. For example, a polypeptide that is derived from a second polypeptide comprises an amino acid sequence that is identical or substantially similar, *e.g.*, is more than 50% homologous to, the amino acid sequence of the second protein. The derived molecular entity, *e.g.*, a nucleic acid or protein, can comprise one or more modifications, *e.g.*, one or more amino acid or nucleotide changes.

“Domain,” as used herein, is used to describe a segment of, or a portion of a protein or nucleic acid. Unless otherwise indicated, a domain is not required to have any specific functional property.

Calculations of homology or sequence identity between two sequences (the terms are used interchangeably herein) are performed as follows. The sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The optimal alignment is determined as the best score using the GAP program in the GCG software package with a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frame shift gap penalty of 5. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences.

As used herein, the term “endogenous” gene, “endogenous” nucleic acid, or “endogenous” homologous region refers to a native gene, nucleic acid, or region of a gene, which is in its natural location in the genome, *e.g.*, chromosome or plasmid, of a cell. In contrast, the term “exogenous” gene or “exogenous” nucleic acid refers to a gene, nucleic acid, or region of a gene which is not native within a cell, but which is introduced into the cell during the methods of the invention. An exogenous gene or exogenous nucleic acid may be homologous to, or identical to, an endogenous gene or an endogenous nucleic acid.

As used herein, the term “endogenous homologous region” refers to an endogenous template nucleic acid sequence which is homologous to at least a portion of a target gene, and which can be used in conjunction with a Cas9 molecule and a gRNA molecule to modify, *e.g.*, correct, a sequence of the target gene. In one embodiment, the endogenous homologous

region is DNA. In another embodiment, the endogenous homologous region is double stranded DNA. In another embodiment, the endogenous homologous region is single stranded DNA. In one embodiment, the endogenous homologous region is at least 70%, 75%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% homologous to at least a portion of the target gene.

As used herein, the term “enzymatically inactive Cas9” (“eiCas9”) or eiCas9 polypeptide refers to Cas9 molecules having no, or no substantial, cleavage activity. For example, an eiCas9 molecule or eiCas9 polypeptide can lack cleavage activity or have substantially less, *e.g.*, less than 20, 10, 5, 1 or 0.1 % of the cleavage activity of a reference Cas9 molecule or eiCas9 polypeptide, as measured by an assay described herein.

In one embodiment, a Cas9 molecule is an eiCas9 molecule comprising one or more differences in a RuvC domain and/or in an HNH domain as compared to a reference Cas9 molecule, and the eiCas9 molecule does not cleave a nucleic acid, or cleaves with significantly less efficiency than does wild type, *e.g.*, when compared with wild type in a cleavage assay, *e.g.*, as described herein, cuts with less than 50, 25, 10, or 1% of a reference Cas9 molecule, as measured by an assay described herein. The reference Cas9 molecule can be a naturally occurring unmodified Cas9 molecule, *e.g.*, a naturally occurring Cas9 molecule such as a Cas9 molecule of *S. pyogenes*, *S. thermophilus*, *S. aureus*, *C. jejuni* or *N. meningitidis*. In one embodiment, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology. In one embodiment, the eiCas9 molecule lacks substantial cleavage activity associated with a RuvC domain and cleavage activity associated with an HNH domain.

Whether or not a particular sequence, *e.g.*, a substitution, may affect one or more activity, such as targeting activity, cleavage activity, etc., can be evaluated or predicted, *e.g.*, by evaluating whether the mutation is conservative. In one embodiment, a “non-essential” amino acid residue, as used in the context of a Cas9 molecule, is a residue that can be altered from the wild-type sequence of a Cas9 molecule, *e.g.*, a naturally occurring Cas9 molecule, *e.g.*, an eiCas9 molecule, without abolishing or more preferably, without substantially altering a Cas9 activity (*e.g.*, cleavage activity), whereas changing an “essential” amino acid residue results in a substantial loss of activity (*e.g.*, cleavage activity).

Although an enzymatically inactive (eiCas9) Cas9 molecule itself can block transcription when recruited to early regions in the coding sequence, more robust repression can be achieved by fusing a transcriptional repression domain (for example KRAB, SID or ERD) to the Cas9 and recruiting it to the target knockdown position, *e.g.*, within 1000 bp of

sequence 3' of the start codon or within 500 bp of a promoter region 5' of the start codon of a gene. It is likely that targeting DNaseI hypersensitive sites (DHSs) of the promoter may yield more efficient gene repression or activation because these regions are more likely to be accessible to the Cas9 protein and are also more likely to harbor sites for endogenous transcription factors. Especially for gene repression, it is contemplated herein that blocking the binding site of an endogenous transcription factor would aid in downregulating gene expression. In one embodiment, one or more eiCas9 molecules may be used to block binding of one or more endogenous transcription factors. In another embodiment, an eiCas9 molecule can be fused to a chromatin modifying protein. Altering chromatin status can result in decreased expression of the target gene. One or more eiCas9 molecules fused to one or more chromatin modifying proteins may be used to alter chromatin status.

As used herein, "error-prone" repair refers to a DNA repair process that has a higher tendency to introduce mutations into the site being repaired. For instance, alt-NHEJ and SSA are error-prone pathways; C-NHEJ is also error prone because it sometimes leads to the creation of a small degree of alteration of the site (even though in some instances C-NHEJ results in error-free repair); and HR, alt-HR, and SSA in the case of a single-strand oligo donor are not error-prone. As used herein, the term "gRNA molecule" or "gRNA" refers to a guide RNA which is capable of targeting a Cas9 molecule to a target nucleic acid. In one embodiment, the term "gRNA molecule" refers to a guide ribonucleic acid. In another embodiment, the term "gRNA molecule" refers to a nucleic acid encoding a gRNA. In one embodiment, a gRNA molecule is non-naturally occurring. In one embodiment, a gRNA molecule is a synthetic gRNA molecule.

"Governing gRNA molecule," as used herein, refers to a gRNA molecule that comprises a targeting domain that is complementary to a target domain on a nucleic acid that comprises a sequence that encodes a component of the CRISPR/Cas system that is introduced into a cell or subject. A governing gRNA does not target an endogenous cell or subject sequence. In one embodiment, a governing gRNA molecule comprises a targeting domain that is complementary with a target sequence on: (a) a nucleic acid that encodes a Cas9 molecule; (b) a nucleic acid that encodes a gRNA molecule which comprises a targeting domain that targets a target gene (a target gene gRNA); or on more than one nucleic acid that encodes a CRISPR/Cas component, *e.g.*, both (a) and (b). In one embodiment, a nucleic acid molecule that encodes a CRISPR/Cas component, *e.g.*, that encodes a Cas9 molecule or a target gene gRNA molecule, comprises more than one target domain that is complementary with a governing gRNA targeting domain. While not wishing to be bound by theory, it is

believed that a governing gRNA molecule complexes with a Cas9 molecule and results in Cas9 mediated inactivation of the targeted nucleic acid, *e.g.*, by cleavage or by binding to the nucleic acid, and results in cessation or reduction of the production of a CRISPR/Cas system component. In one embodiment, the Cas9 molecule forms two complexes: a complex

5 comprising a Cas9 molecule with a target gene gRNA molecule, which complex will alter the target gene; and a complex comprising a Cas9 molecule with a governing gRNA molecule, which complex will act to prevent further production of a CRISPR/Cas system component, *e.g.*, a Cas9 molecule or a target gene gRNA molecule. In one embodiment, a governing gRNA molecule/Cas9 molecule complex binds to or promotes cleavage of a control region

10 sequence, *e.g.*, a promoter, operably linked to a sequence that encodes a Cas9 molecule, a sequence that encodes a transcribed region, an exon, or an intron, for the Cas9 molecule. In one embodiment, a governing gRNA molecule/Cas9 molecule complex binds to or promotes cleavage of a control region sequence, *e.g.*, a promoter, operably linked to a gRNA molecule, or a sequence that encodes the gRNA molecule. In one embodiment, the governing gRNA

15 molecule, *e.g.*, a Cas9-targeting governing gRNA molecule, or a target gene gRNA-targeting governing gRNA molecule, limits the effect of the Cas9 molecule/target gene gRNA molecule complex-mediated gene targeting. In one embodiment, a governing gRNA places temporal, level of expression, or other limits, on activity of the Cas9 molecule/target gene gRNA molecule complex. In one embodiment, a governing gRNA reduces off-target or other

20 unwanted activity. In one embodiment, a governing gRNA molecule inhibits, *e.g.*, entirely or substantially entirely inhibits, the production of a component of the Cas9 system and thereby limits, or governs, its activity.

“HDR”, or homology-directed repair, as used herein, refers to the process of repairing DNA damage using a homologous nucleic acid (*e.g.*, an endogenous nucleic acid, *e.g.*, a

25 sister chromatid, or an exogenous nucleic acid, *e.g.*, a template nucleic acid). HDR typically occurs when there has been significant resection at a double-strand break, forming at least one single stranded portion of DNA. HDR is a category that includes, for example, single-strand annealing (SSA), homologous recombination (HR), single strand template repair (SST-R), and a third, not yet fully characterized alternative homologous recombination (alt-HR)

30 DNA repair pathway. In some embodiments, HDR includes gene conversion and gene correction. In some embodiments, the term HDR does not encompass canonical NHEJ (C-NHEJ). In some embodiments, the term HDR does not encompass alternative non-homologous end joining (Alt-NHEJ) (*e.g.*, blunt end-joining (blunt EJ), (micro homology

mediated end joining (MMEJ), and synthesis dependent microhomology-mediated end joining (SD-MMEJ)).

The terms "homology" or "identity," as used interchangeably herein, refer to sequence identity between two amino acid sequences or two nucleic acid sequences, with identity being a more strict comparison. The phrases "percent identity or homology" and "% identity or homology" refer to the percentage of sequence identity found in a comparison of two or more amino acid sequences or nucleic acid sequences. Two or more sequences can be anywhere from 0-100% identical, or any value there between. Identity can be determined by comparing a position in each sequence that can be aligned for purposes of comparison to a reference sequence. When a position in the compared sequence is occupied by the same nucleotide base or amino acid, then the molecules are identical at that position. A degree of identity of amino acid sequences is a function of the number of identical amino acids at positions shared by the amino acid sequences. A degree of identity between nucleic acid sequences is a function of the number of identical or matching nucleotides at positions shared by the nucleic acid sequences. A degree of homology of amino acid sequences is a function of the number of amino acids at positions shared by the polypeptide sequences.

Calculations of homology or sequence identity between two sequences (the terms are used interchangeably herein) are performed as follows. The sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). The optimal alignment is determined as the best score using the GAP program in the GCG software package with a Blossum 62 scoring matrix with a gap penalty of 12, a gap extend penalty of 4, and a frame shift gap penalty of 5. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences.

"Gene conversion", as used herein, refers to the process of repairing DNA damage by homology directed recombination (HDR) using an endogenous nucleic acid, *e.g.*, a sister chromatid or a plasmid, as a template nucleic acid. Without being bound by theory, in some embodiments, BRCA1, BRCA2 and/or RAD51 are believed to be involved in gene conversion. In some embodiments, the endogenous nucleic acid is a nucleic acid sequence

having homology, *e.g.*, significant homology, with a fragment of DNA proximal to the site of the DNA lesion or mutation. In some embodiments, the template is not an exogenous nucleic acid.

“Gene correction”, as used herein, refers to the process of repairing DNA damage by homology directed recombination using an exogenous nucleic acid, *e.g.*, a donor template nucleic acid. In some embodiments, the exogenous nucleic acid is single-stranded. In some embodiments, the exogenous nucleic acid is double-stranded. In one embodiment, the donor template nucleic acid is a circular nucleic acid sequence. In another embodiment, the donor template nucleic acid is a linear nucleic acid sequence.

“Homologous recombination” or “HR” refers to a type of HDR DNA-repair which typically acts occurs when there has been significant resection at the double-strand break, forming at least one single stranded portion of DNA. In a normal cell, HR typically involves a series of steps such as recognition of the break, stabilization of the break, resection, stabilization of single stranded DNA, formation of a DNA crossover intermediate, resolution of the crossover intermediate, and ligation. The process requires RAD51 and BRCA2, and the homologous nucleic acid is typically double-stranded. In some embodiments, homologous recombination includes gene conversion.

“Ligand acceptor molecule,” as used herein, refers to a substance or molecule that specifically interacts non-covalently with at least one ligand. In one embodiment, the ligand acceptor molecule is streptavidin.

The term “ligand,” as used herein, refers to a substance or molecule that specifically interacts with another substance or molecule (*e.g.*, a ligand acceptor molecule). In one embodiment, the ligand is biotin. In some embodiments, the ligand is a high-affinity ligand (*e.g.*, a ligand that has high affinity for its receptor).

The term “linked” or “linkage” as used herein means an interaction between molecules or parts of molecules. Two molecules that are linked may be covalently linked or non-covalently linked.

The term “linker”, as used herein, refers to a molecule which facilitates an interaction between molecules or parts of molecules. In one embodiment, a linker is a polypeptide linker. In another embodiment, a linker is a nucleic acid linker.

The term “peptide linker” or “polypeptide linker” as used herein means a peptide or polypeptide comprising two or more amino acids residues joined by peptide bonds. Such peptide or polypeptide linkers are well known in the art. Linkers comprise naturally occurring and/or non-naturally occurring peptides or polypeptides.

“Modulator,” as used herein, refers to an entity, *e.g.*, a compound, that can alter the activity (*e.g.*, enzymatic activity, transcriptional activity, or translational activity), amount, distribution, or structure of a subject molecule or genetic sequence. In one embodiment, modulation comprises cleavage, *e.g.*, breaking of a covalent or non-covalent bond, or the forming of a covalent or non-covalent bond, *e.g.*, the attachment of a moiety, to the subject molecule. In one embodiment, a modulator alters the, three dimensional, secondary, tertiary, or quaternary structure, of a subject molecule. A modulator can increase, decrease, initiate, or eliminate a subject activity.

As used herein, the term “mutation” refers to a change in the sequence of a nucleic acid as compared to a wild-type sequence of the nucleic acid, resulting a variant form of the nucleic acid. A mutation in a nucleic acid may be caused by the alteration of a single base pair in the nucleic acid, or the insertion, deletion, or rearrangement of larger sections of the nucleic acid. A mutation in a gene may result in variants of the protein encoded by the gene which are associated with genetic disorders.

The term “non-covalent bond” refers to a variety of interactions between molecules or parts of molecules that are not covalent in nature, which provide force to hold the molecules or parts of molecules together usually in a specific orientation or conformation. Such non-covalent interactions include *inter alia* ionic bonds, hydrophobic interactions, hydrogen bonds, Van-der-Waals forces, and dipole-dipole bonds.

“Non-homologous end joining” or “NHEJ,” as used herein, refers to ligation mediated repair and/or non-template mediated repair including canonical NHEJ (cNHEJ), alternative NHEJ (altNHEJ), microhomology-mediated end joining (MMEJ), single-strand annealing (SSA), and synthesis-dependent microhomology-mediated end joining (SD-MMEJ). Unless indicate otherwise, “NHEJ” as used herein encompasses canonical NHEJ, alt-NHEJ, MMEJ, SSA and SD-MMEJ.

“Polypeptide,” as used herein, refers to a polymer of amino acids.

The term “protein”, as used herein, is intended to refer to a biomolecule comprised of amino acids arranged in the form of a polypeptide. A protein may be a full-length protein, or a fragment thereof.

As used herein, the term “processing,” with respect to overhangs, refers to either the endonucleolytic processing or the exonucleolytic processing of a break in a nucleic acid molecule. In one embodiment, processing of a 5’ overhang in a nucleic acid molecule may result in a 3’ overhang. In another embodiment, processing of a 3’ overhang in a nucleic acid molecule may result in a 5’ overhang.

A “reference molecule,” as used herein, refers to a molecule to which a modified or candidate molecule is compared. For example, a reference Cas9 molecule refers to a Cas9 molecule to which a modified or candidate Cas9 molecule is compared. The modified or candidate molecule may be compared to the reference molecule on the basis of sequence (e.g., the modified or candidate may have X% sequence identity or homology with the reference molecule) or activity (e.g., the modified or candidate molecule may have X% of the activity of the reference molecule). For example, where the reference molecule is a Cas9 molecule, a modified or candidate may be characterized as having no more than 10% of the nuclease activity of the reference Cas9 molecule. Examples of reference Cas9 molecules include naturally occurring unmodified Cas9 molecules, e.g., a naturally occurring Cas9 molecule from *S. pyogenes*, *S. aureus*, *S. thermophilus* or *N. meningitidis*. In certain embodiments, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology with the modified or candidate Cas9 molecule to which it is being compared. In certain embodiments, the reference Cas9 molecule is a parental molecule having a naturally occurring or known sequence on which a mutation has been made to arrive at the modified or candidate Cas9 molecule.

“Replacement,” or “replaced,” as used herein with reference to a modification of a molecule does not require a process limitation but merely indicates that the replacement entity is present.

“Resection”, as used herein, refers to exonuclease-mediated digestion of one strand of a double-stranded DNA molecule, which results in a single-stranded overhang. Resection may occur, e.g., on one or both sides of a double-stranded break. Resection can be measured by, for instance, extracting genomic DNA, digesting it with an enzyme that selectively degrades dsDNA, and performing quantitative PCR using primers spanning the DSB site, e.g., as described herein.

“SSA” or “Single-strand Annealing”, as used herein, refers to the process where RAD52 as opposed to RAD51 in the HR pathways, binds to the single stranded portion of DNA and promotes annealing of the two single stranded DNA segments at repetitive regions. Once RAD52 binds XFP/ERCC1 removes DNA flaps to make the DNA more suitable for ligation.

“Subject,” as used herein, may mean either a human or non-human animal. The term includes, but is not limited to, mammals (e.g., humans, other primates, pigs, rodents (e.g., mice and rats or hamsters), rabbits, guinea pigs, cows, horses, cats, dogs, sheep, and goats). In one embodiment, the subject is a human. In another embodiment, the subject is poultry.

In another embodiment, the subject is piscine. In certain embodiments, the subject is a human, and in certain of these embodiments the human is an infant, child, young adult, or adult.

As used herein, the terms “target nucleic acid” or “target gene” refer to a nucleic acid which is being targeted for alteration, *e.g.*, by gene correction, by a Cas9 system described herein. In certain embodiments, a target nucleic acid comprises one gene. In certain embodiments, a target nucleic acid may comprise one or more genes, *e.g.*, two genes, three genes, four genes, or five genes. In one embodiment, a target nucleic acid may comprise a promoter region, or control region, of a gene. In one embodiment, a target nucleic acid may comprise an intron of a gene. In another embodiment, a target nucleic acid may comprise an exon of a gene. In one embodiment, a target nucleic acid may comprise a coding region of gene. In one embodiment, a target nucleic acid may comprise a non-coding region of a gene.

“Target position” as used herein, refers to a site on a target nucleic acid that is modified by a Cas9 molecule-dependent or a Cas9 fusion molecule-dependent process. For example, the target position can be modified by a Cas9 molecule-mediated cleavage (or a Cas9 fusion molecule-mediated cleavage) of the target nucleic acid and template nucleic acid directed modification, *e.g.*, correction, of the target position. In one embodiment, a target position can be a site between two nucleotides, *e.g.*, adjacent nucleotides, on the target nucleic acid into which one or more nucleotides is added based on homology with a template nucleic acid. The target position may comprise one or more nucleotides that are altered, *e.g.*, corrected, based on homology with a template nucleic acid. In another embodiment, the target position may comprise one or more nucleotides that are deleted based on homology with a template nucleic acid. In one embodiment, the target position is within a “target sequence” (*e.g.*, the sequence to which the gRNA binds). In one embodiment, a target position is upstream or downstream of a target sequence (*e.g.*, the sequence to which the gRNA binds).

“Target region,” “target domain,” or “target sequence,” as used herein, is a nucleic acid sequence that comprises a target position and at least one nucleotide position outside the target position. In certain embodiments, the target position is flanked by sequences of the target position region, *i.e.*, the target position is disposed in the target position region such that there are target position region sequences both 5’ and 3’ to the target position. In certain embodiments, the target position region provides sufficient sequences on each side (*i.e.*, 5’ and 3’) of the target position to allow gene correction of the target position, wherein the gene

correction uses an exogenous sequence homologous with the target position region as a template.

A “template nucleic acid,” “exogenous homologous region,” “donor nucleic acid,” “exogenous template,” or “donor template” as that term is used herein, refers to a nucleic acid sequence which can be used in conjunction with a Cas9 molecule (or a Cas9 fusion molecule) and a gRNA molecule and services as a guide for altering the structure of a target position. In some embodiments, the template nucleic acid is homologous to at least a portion of a target gene, and which can be used in conjunction with a Cas9 molecule and a gRNA molecule to modify, *e.g.*, correct, a sequence of the target gene. In one embodiment, the target nucleic acid is modified to have the some or all of the sequence of the template nucleic acid, typically at or near cleavage site(s). In some embodiments, the template nucleic acid is a nucleic acid, *e.g.*, DNA or RNA. In one embodiment, the template nucleic acid is single stranded. In an alternate embodiment, the template nucleic acid is double stranded. In one embodiment, the template nucleic acid is DNA, *e.g.*, double stranded DNA. In an alternate embodiment, the template nucleic acid is single stranded DNA. In one embodiment, the template nucleic acid is encoded on the same vector backbone, *e.g.*, AAV genome, plasmid DNA, as the Cas9 and gRNA. In one embodiment, the template nucleic acid is excised from a vector backbone *in vivo*, *e.g.*, it is flanked by gRNA recognition sequences. In one embodiment, the template nucleic acid comprises endogenous genomic sequence. In one embodiment, the template nucleic acid is an RNA. In some embodiments the template nucleic acid is circular nucleic acid. In other embodiments, the template nucleic acid is linear nucleic acid.

In one embodiment, the template nucleic acid alters the structure of the target position by participating in a homology directed repair event, *e.g.*, a gene correction event. In one embodiment, the template nucleic acid alters the sequence of the target position. In one embodiment, the template nucleic acid results in the incorporation of a modified, or non-naturally occurring base into the target nucleic acid.

Typically, the template sequence undergoes a breakage mediated or catalyzed recombination with the target sequence. In one embodiment, the template nucleic acid includes sequence that corresponds to a site on the target sequence that is cleaved by an eaCas9 mediated cleavage event. In one embodiment, the template nucleic acid includes sequence that corresponds to both, a first site on the target sequence that is cleaved in a first Cas9 mediated event, and a second site on the target sequence that is cleaved in a second Cas9 mediated event.

In one embodiment, the template nucleic acid can include sequence which results in an alteration in the coding sequence of a translated sequence, *e.g.*, one which results in the substitution of one amino acid for another in a protein product, *e.g.*, transforming a mutant allele into a wild type allele, transforming a wild type allele into a mutant allele, and/or
 5 introducing a stop codon, insertion of an amino acid residue, deletion of an amino acid residue, or a nonsense mutation.

In other embodiments, the template nucleic acid can include sequence which results in an alteration in a non-coding sequence, *e.g.*, an alteration in an exon or in a 5' or 3' non-translated or non-transcribed region. Such alterations include an alteration in a control
 10 element, *e.g.*, a promoter, enhancer, and an alteration in a cis-acting or trans-acting control element.

A template nucleic acid having homology with a target position in a gene, *e.g.*, a gene described herein, can be used to alter the structure of a target sequence. The template sequence can be used to alter an unwanted structure, *e.g.*, an unwanted or mutant nucleotide.

15 A template nucleic acid typically comprises the following components:

[5' homology arm]-[replacement sequence]-[3' homology arm].

The homology arms provide for recombination into the chromosome, thus replacing the undesired element, *e.g.*, a mutation or signature, with a replacement sequence, *e.g.*, the desired, or corrected sequence. In one embodiment, the homology arms flank the most distal
 20 cleavage sites.

In one embodiment, the 3' end of the 5' homology arm is the position next to the 5' end of the replacement sequence. In one embodiment, the 5' homology arm can extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, or 5000 nucleotides 5' from the 5' end of the replacement sequence.

25 In one embodiment, the 5' end of the 3' homology arm is the position next to the 3' end of the replacement sequence. In one embodiment, the 3' homology arm can extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, or 5000 nucleotides 3' from the 3' end of the replacement sequence.

In one embodiment, to correct a mutation, the homology arms, *e.g.*, the 5' and 3' homology arms, may each comprise about 1000 base pairs (bp) of sequence flanking the most distal gRNAs (*e.g.*, 1000bp of sequence on either side of the mutation).

It is contemplated herein that one or both homology arms may be shortened to avoid including certain sequence repeat elements, *e.g.*, Alu repeats or LINE elements. For example, a 5' homology arm may be shortened to avoid a sequence repeat element. In other

embodiments, a 3' homology arm may be shortened to avoid a sequence repeat element. In some embodiments, both the 5' and the 3' homology arms may be shortened to avoid including certain sequence repeat elements.

It is contemplated herein that template nucleic acids for correcting a mutation may be designed for use as a single-stranded oligonucleotide, *e.g.*, a single-stranded oligodeoxynucleotide (ssODN). When using a ssODN, 5' and 3' homology arms may range up to about 200 base pairs (bp) in length, *e.g.*, at least 25, 50, 75, 100, 125, 150, 175, or 200 bp in length. Longer homology arms are also contemplated for ssODNs as improvements in oligonucleotide synthesis continue to be made. In some embodiments, a longer homology arm is made by a method other than chemical synthesis, *e.g.*, by denaturing a long double stranded nucleic acid and purifying one of the strands, *e.g.*, by affinity for a strand-specific sequence anchored to a solid substrate.

While not wishing to be bound by theory, in some embodiments HDR proceeds more efficiently when the template nucleic acid has extended homology 5' to a nick, (*i.e.*, in the 5' direction of the nicked strand). A nick, as referred to herein, refers to a single strand break in a nucleic acid. Accordingly, in some embodiments, the template nucleic acid has a longer homology arm and a shorter homology arm, wherein the longer homology arm can anneal 5' of the nick. In some embodiments, the arm that can anneal 5' to the nick is at least 25, 50, 75, 100, 125, 150, 175, or 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, or 5000 nucleotides from the nick or the 5' or 3' end of the replacement sequence. In some embodiments, the arm that can anneal 5' to the nick is at least 10%, 20%, 30%, 40%, or 50% longer than the arm that can anneal 3' to the nick. In some embodiments, the arm that can anneal 5' to the nick is at least 2x, 3x, 4x, or 5x longer than the arm that can anneal 3' to the nick. Depending on whether a ssDNA template can anneal to the intact strand or the nicked strand, the homology arm that anneals 5' to the nick may be at the 5' end of the ssDNA template or the 3' end of the ssDNA template, respectively.

Similarly, in some embodiments, the template nucleic acid has a 5' homology arm, a replacement sequence, and a 3' homology arm, such that the template nucleic acid has extended homology to the 5' of the nick. For example, the 5' homology arm and 3' homology arm may be substantially the same length, but the replacement sequence may extend farther 5' of the nick than 3' of the nick. In some embodiments, the replacement sequence extends at least 10%, 20%, 30%, 40%, 50%, 2x, 3x, 4x, or 5x further to the 5' end of the nick than the 3' end of the nick.

While not wishing to be bound by theory, in some embodiments HDR proceeds more efficiently when the template nucleic acid is centered on the nick. Accordingly, in some embodiments, the template nucleic acid has two homology arms that are essentially the same size. For instance, the first homology arm of a template nucleic acid may have a length that is within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% of the second homology arm of the template nucleic acid.

Similarly, in some embodiments, the template nucleic acid has a 5' homology arm, a replacement sequence, and a 3' homology arm, such that the template nucleic acid extends substantially the same distance on either side of the nick. For example, the homology arms may have different lengths, but the replacement sequence may be selected to compensate for this. For example, the replacement sequence may extend further 5' from the nick than it does 3' of the nick, but the homology arm 5' of the nick is shorter than the homology arm 3' of the nick, to compensate. The converse is also possible, *e.g.*, that the replacement sequence may extend further 3' from the nick than it does 5' of the nick, but the homology arm 3' of the nick is shorter than the homology arm 5' of the nick, to compensate.

A "variant Cas9 molecule," as used herein refers to a Cas9 molecule with at least one modification, *e.g.*, a mutation or chemical modification to at least one amino acid residue of the wild-type Cas9 molecule.

Exemplary arrangements of linear nucleic acid template systems

In one embodiment, the template nucleic acid is double stranded. In one embodiment, the template nucleic acid is single stranded. In one embodiment, the nucleic acid template system comprises a single stranded portion and a double stranded portion. In one embodiment, the template nucleic acid comprises about 50 to 100, *e.g.*, 55 to 95, 60 to 90, 65 to 85, or 70 to 80, base pairs, homology on either side of the nick and/or replacement sequence. In one embodiment, the template nucleic acid comprises about 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 base pairs homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequences.

In one embodiment, the template nucleic acid comprises about 150 to 200, *e.g.*, 155 to 195, 160 to 190, 165 to 185, or 170 to 180, base pairs homology 3' of the nick and/or replacement sequence. In one embodiment, the template nucleic acid comprises about 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, or 200 base pairs homology 3' of the nick or replacement sequence. In one embodiment, the template nucleic acid comprises less than

about 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, or 10 base pairs homology 5' of the nick or replacement sequence.

In one embodiment, the template nucleic acid comprises about 150 to 200, *e.g.*, 155 to 195, 160 to 190, 165 to 185, or 170 to 180, base pairs homology 5' of the nick and/or replacement sequence. In one embodiment, the template nucleic acid comprises about 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, or 200 base pairs homology 5' of the nick or replacement sequence. In one embodiment, the template nucleic acid comprises less than about 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, or 10 base pairs homology 3' of the nick or replacement sequence.

Exemplary Template Nucleic Acids

In one embodiment, the template nucleic acid is a single stranded nucleic acid. In another embodiment, the template nucleic acid is a double stranded nucleic acid. In some embodiments, the template nucleic acid comprises a nucleotide sequence, *e.g.*, of one or more nucleotides, that will be added to or will create a change in, or correct the sequence of the target nucleic acid to a desired sequence. In other embodiments, the template nucleic acid comprises a nucleotide sequence that may be used to modify the target position. In other embodiments, the template nucleic acid comprises a nucleotide sequence, *e.g.*, of one or more nucleotides, that corresponds to wild type sequence of the target nucleic acid, *e.g.*, of the target position.

The template nucleic acid may comprise a replacement sequence. A replacement sequence, as the term is used herein, refers to a sequence which will serve as the template for making the desired change, or correction, in the target nucleic acid. The replacement sequence is homologous, but not identical to, the target nucleic acid. In some embodiments, the template nucleic acid comprises a 5' homology arm. In other embodiments, the template nucleic acid comprises a 3' homology arm.

In embodiments, the template nucleic acid is linear double stranded DNA. The length may be, *e.g.*, about 150-200 base pairs, *e.g.*, about 150, 160, 170, 180, 190, or 200 base pairs. The length may be, *e.g.*, at least 150, 160, 170, 180, 190, or 200 base pairs. In some embodiments, the length is no greater than 150, 160, 170, 180, 190, or 200 base pairs. In some embodiments, a double stranded template nucleic acid has a length of about 160 base pairs, *e.g.*, about 155-165, 150-170, 140-180, 130-190, 120-200, 110-210, 100-220, 90-230, or 80-240 base pairs.

The template nucleic acid can be linear single stranded DNA. In embodiments, the template nucleic acid is (i) linear single stranded DNA that can anneal to the nicked strand of the target nucleic acid, (ii) linear single stranded DNA that can anneal to the intact strand of the target nucleic acid, (iii) linear single stranded DNA that can anneal to the transcribed strand of the target nucleic acid, (iv) linear single stranded DNA that can anneal to the non-transcribed strand of the target nucleic acid, or more than one of the preceding. The length may be, *e.g.*, about 150-200 nucleotides, *e.g.*, about 150, 160, 170, 180, 190, or 200 nucleotides. The length may be, *e.g.*, at least 150, 160, 170, 180, 190, or 200 nucleotides. In some embodiments, the length is no greater than 150, 160, 170, 180, 190, or 200 nucleotides. In some embodiments, a single stranded template nucleic acid has a length of about 160 nucleotides, *e.g.*, about 155-165, 150-170, 140-180, 130-190, 120-200, 110-210, 100-220, 90-230, or 80-240 nucleotides.

In some embodiments, the template nucleic acid is circular double stranded DNA, *e.g.*, a plasmid. In some embodiments, the template nucleic acid comprises about 500 to 1000 base pairs of homology on either side of the replacement sequence and/or the nick. In some embodiments, the template nucleic acid comprises about 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 base pairs of homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence. In some embodiments, the template nucleic acid comprises at least 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 base pairs of homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence. In some embodiments, the template nucleic acid comprises no more than 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 base pairs of homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence.

In some embodiments, the template nucleic acid is an adenovirus vector, *e.g.*, an AAV vector, *e.g.*, a ssDNA molecule of a length and sequence that allows it to be packaged in an AAV capsid. The vector may be, *e.g.*, less than 5 kb and may contain an ITR sequence that promotes packaging into the capsid. The vector may be integration-deficient. In some embodiments, the template nucleic acid comprises about 150 to 1000 nucleotides of homology on either side of the replacement sequence and/or the nick. In some embodiments, the template nucleic acid comprises about 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence. In some

embodiments, the template nucleic acid comprises at least 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence. In some embodiments, the template nucleic acid comprises at most 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence.

In some embodiments, the template nucleic acid is a lentiviral vector, *e.g.*, an IDLV (integration deficiency lentivirus). In some embodiments, the template nucleic acid comprises about 500 to 1000 base pairs of homology on either side of the replacement sequence and/or the nick. In some embodiments, the template nucleic acid comprises about 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 base pairs of homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence. In some embodiments, the template nucleic acid comprises at least 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 base pairs of homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence. In some embodiments, the template nucleic acid comprises no more than 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 base pairs of homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence.

In one embodiment, the template nucleic acid comprises one or more mutations, *e.g.*, silent mutations, that prevent Cas9 from recognizing and cleaving the template nucleic acid. The template nucleic acid may comprise, *e.g.*, at least 1, 2, 3, 4, 5, 10, 20, or 30 silent mutations relative to the corresponding sequence in the genome of the cell to be altered. In some embodiments, the template nucleic acid comprises at most 2, 3, 4, 5, 10, 20, 30, or 50 silent mutations relative to the corresponding sequence in the genome of the cell to be altered. In one embodiment, the cDNA comprises one or more mutations, *e.g.*, silent mutations that prevent Cas9 from recognizing and cleaving the template nucleic acid. The template nucleic acid may comprise, *e.g.*, at least 1, 2, 3, 4, 5, 10, 20, or 30 silent mutations relative to the corresponding sequence in the genome of the cell to be altered. In some embodiments, the template nucleic acid comprises at most 2, 3, 4, 5, 10, 20, 30, or 50 silent mutations relative to the corresponding sequence in the genome of the cell to be altered.

In one embodiment, the template nucleic acid alters the structure of the target position by participating in a homology directed repair event. In one embodiment, the template

nucleic acid alters the sequence of the target position. In one embodiment, the template nucleic acid results in the incorporation of a modified, or non-naturally occurring base into the target nucleic acid.

Typically, the template sequence undergoes a breakage mediated or catalyzed recombination with the target sequence. In one embodiment, the template nucleic acid includes sequence that corresponds to a site on the target sequence that is cleaved by an Cas9 mediated cleavage event. In one embodiment, the template nucleic acid includes sequence that corresponds to both, a first site on the target sequence that is cleaved in a first Cas9 mediated event, and a second site on the target sequence that is cleaved in a second Cas9 mediated event.

In one embodiment, the template nucleic acid can include sequence which results in an alteration in the coding sequence of a translated sequence, *e.g.*, one which results in the substitution of one amino acid for another in a protein product, *e.g.*, transforming a mutant allele into a wild type allele, transforming a wild type allele into a mutant allele, and/or introducing a stop codon, insertion of an amino acid residue, deletion of an amino acid residue, or a nonsense mutation.

In other embodiments, the template nucleic acid can include sequence which results in an alteration in a non-coding sequence, *e.g.*, an alteration in an exon or in a 5' or 3' non-translated or non-transcribed region. Such alterations include an alteration in a control element, *e.g.*, a promoter, enhancer, and an alteration in a cis-acting or trans-acting control element.

A template nucleic acid having homology with a target position can be used to alter the structure of a target sequence. The template sequence can be used to alter an unwanted structure, *e.g.*, an unwanted or mutant nucleotide.

Table 1 below provides exemplary template nucleic acids. In one embodiment, the template nucleic acid includes the 5' homology arm and the 3' homology arm of a row from **Table 1**. In another embodiment, a 5' homology arm from the first column can be combined with a 3' homology arm from **Table 1**. In each embodiment, a combination of the 5' and 3' homology arms include a replacement sequence.

Table 1

Length of the 5' homology arm (the number of nucleotides)	Replacement Sequence	Length of the 3' homology arm (the number of nucleotides)
10 or more		10 or more
20 or more		20 or more
50 or more		50 or more
100 or more		100 or more
150 or more		150 or more
200 or more		200 or more
250 or more		250 or more
300 or more		300 or more
350 or more		350 or more
400 or more		400 or more
450 or more		450 or more
500 or more		500 or more
550 or more		550 or more
600 or more		600 or more
650 or more		650 or more
700 or more		700 or more
750 or more		750 or more
800 or more		800 or more
850 or more		850 or more
900 or more		900 or more
1000 or more		1000 or more
1100 or more		1100 or more
1200 or more		1200 or more
1300 or more		1300 or more
1400 or more		1400 or more
1500 or more		1500 or more
1600 or more		1600 or more
1700 or more		1700 or more
1800 or more		1800 or more
1900 or more		1900 or more
1200 or more		1200 or more
At least 50 but not long enough to include a repeated element.		At least 50 but not long enough to include a repeated element.
At least 100 but not long enough to include a repeated element.		At least 100 but not long enough to include a repeated element.
At least 150 but not long enough to include a repeated element.		At least 150 but not long enough to include a repeated element.
5 to 100 nucleotides		5 to 100 nucleotides
10 to 150 nucleotides		10 to 150 nucleotides
20 to 150 nucleotides		20 to 150 nucleotides

“Treat,” “treating” and “treatment,” as used herein, mean the treatment of a disease in a mammal, *e.g.*, in a human, including (a) inhibiting the disease, *i.e.*, arresting or preventing its development or progression; (b) relieving the disease, *i.e.*, causing regression of the disease state; and (c) relieving one or more symptoms of the disease; and (d) curing the disease.

“Prevent,” “preventing” and “prevention,” as used herein, means the prevention of a disease in a mammal, *e.g.*, in a human, including (a) avoiding or precluding the disease; (b) affecting the predisposition toward the disease (c) preventing or delaying the onset of at least one symptom of the disease.

An “up-regulator”, as used herein, refers to an agent that directly increases the activity of a specified biological pathway. Directly increasing the activity of the pathway refers to (i) the up-regulator binding to a component of that pathway (*e.g.*, a protein that acts in the pathway or an mRNA encoding that protein) and increasing the level or activity of that component, *e.g.*, by increasing the concentration or specific activity of that component, or (ii) the up-regulator is an added amount of a component that is ordinarily present in the pathway at a given level, *e.g.*, an overexpressed protein. An up-regulator may, *e.g.*, speed up one of the steps of that pathway or increase the level or activity of a component in that pathway. An up-regulator may be, *e.g.*, a protein in the pathway, *e.g.*, one may overexpress a protein that is ordinarily in the pathway to increase the overall activity of the pathway. The pathway may be, *e.g.*, a DNA damage repair pathway, for example, HDR, *e.g.*, gene correction. In one embodiment, the increased level or activity is compared to what would be seen in the absence of the up-regulator.

“Wild type”, as used herein, refers to a gene or polypeptide which has the characteristics, *e.g.*, the nucleotide or amino acid sequence, of a gene or polypeptide from a naturally-occurring source. The term “wild type” typically includes the most frequent observation of a particular gene or polypeptide in a population of organisms found in nature.

“X” as used herein in the context of an amino acid sequence, refers to any amino acid (*e.g.*, any of the twenty natural amino acids) unless otherwise specified.

I. Cas9 Fusion Molecules

Various types of Cas9 fusion molecules or Cas9 fusion polypeptides are disclosed herein. In some embodiments, a Cas9 fusion molecule is a chimeric protein comprising a Cas9 protein or a Cas9 polypeptide, or a fragment thereof, covalently linked to at least one template nucleic acid. In other embodiments, a Cas9 fusion molecule is a chimeric protein

comprising a Cas9 protein or a Cas9 polypeptide, or a fragment thereof, non-covalently linked to at least one template nucleic acid. Exemplary Cas9 fusion molecules are provided below.

5 Cas9-Template Nucleic Acid Covalent Fusions

Cas9 molecules of the invention can be directly linked to a template nucleic acid by various forms of covalent attachment, each of which are described in the subsections, below.

Linkers to Connect Cas9 Molecules to a Template Nucleic Acid

10 In one embodiment, a linker covalently connects a Cas9 molecule to a template nucleic acid to form a Cas9 fusion molecule. A linker may be a short peptide sequence that connects a protein domain and a nucleic acid (*e.g.*, DNA or RNA). Linkers may be composed of flexible residues like glycine and serine so that the adjacent protein domains are free to move relative to one another. In certain embodiments, the linker has sufficient length
15 and flexibility to allow a Cas9 molecule to bind to a target nucleic acid, *e.g.*, so that the binding event is not sterically prohibited.

In one embodiment, a linker is attached to the C-terminus of a Cas9 molecule. Alternatively, a linker is attached to the N-terminus of a Cas9 molecule. In another embodiment, the linker is attached to a position other than the C-terminus or the N-terminus
20 of a Cas9 molecule, *e.g.*, an internal residue of the Cas9 molecule.

In one embodiment, a linker is attached to the C-terminus of the template nucleic acid. Alternatively, the linker is attached to the N-terminus of the template nucleic acid.

In one embodiment, the linker is attached to the 5' end of the template nucleic acid. In another embodiment, the linker is attached to the 3' end of the template nucleic acid. In
25 yet another embodiment, the linker is attached to a residue other than the 5'-end or the 3'-end of the donor nucleic acid (*e.g.*, an internal nucleic acid residue).

In some embodiments, the linker from about 3 to 100 amino acids in length. The linker may be, *e.g.*, 3-10, 6-10, 10-15, 15-20, 20-30, 30-40, 40-50, 50-60, 60-70, 70-80, 80-90 or 90-100 amino acids in length. The linker may be, *e.g.*, at least 3, 6, 7, 8, 9, 10, 15, 20, 30,
30 40, 50, 60, 70, 80 or 90 amino acids in length. In other embodiments, the linker is, *e.g.*, at most 7, 8, 9, 10, 15, 20, 30, 40, 50, 60, 70, 80, 90 or 100 amino acids in length. Ranges comprising any combination of these endpoints are also envisioned.

In some embodiments, the linker is encoded by a nucleic acid sequence comprising about 9 to about 300 nucleotides or base pairs. The nucleic acid may be, *e.g.*, 9-300, 9-210,

9-99, 9-45 nucleotides in length. The linker may be, *e.g.*, at least 9, 21, 45, 99, or 210 nucleotides in length. In some embodiments, the linker is, *e.g.*, at most 9, 18, 21, 45, 99, or 210 nucleotides in length. Ranges comprising any combination of these endpoints are also envisioned.

- 5 In some embodiments, the linker comprises glycine and serine residues. In some embodiments the linker consists of glycine and serine residues. For instance, the linker may comprise one of more modules such as GGS, GSGS, GGGS, GGGGS or GGSG. In some embodiments, the linker comprises a plurality of modules comprising glycine and serine, *e.g.*, at least 2, 3, 4, 5, 10, or 15 of these modules, and/or at most 3, 4, 5, 10, 15, or 20 of these
- 10 modules, or any combination of these endpoints. In some embodiments, each module in the linker has the same sequence, and in other embodiments, at least two modules in a linker have different sequences from each other.

In some embodiments, the linker is an XTEN linker or a variation of an XTEN linker such as SGSETPGTSESA, SGSETPGTSESATPES, or SGSETPGTSESATPEGGSGGS.

- 15 Additional information on the XTEN linker may be found in Schellenberger *et al.* (2009), NATURE BIOTECHNOLOGY 27: 1186-1190, the entire contents of which are incorporated herein by reference.

Exemplary linker modules are given in **Table 2**:

Linker	SEQ ID NO
GGS	206
GSGS	207
GGGS	208
GGGGS	209
GGSG	210
SGSETPGTSESA	211
SGSETPGTSESATPES	212
SGSETPGTSESATPEGGSGGS	213
GGSGGSGGSGGSGGSGGSGGSGGSGGSGGS	214

- 20 Additional exemplary linker modules are given in **Table 3**:

Name	Description	Length (nt)
BBa_J176131	PLrigid	60
BBa_J18920	2aa GS linker	6
BBa_J18921	6aa [GS] _x linker	18

Name	Description	Length (nt)
BBa_J18922	10aa [GS]x linker	30
BBa_K105012	10 aa flexible protein domain linker	30
BBa_K133132	8 aa protein domain linker	24
BBa_K1486003	flexible linker 2x (GGGS)	24
BBa_K1486004	flexible linker 2x (GGGGS)	30
BBa_K1486037	linker	39
BBa_K157009	Split fluorophore linker; Freiburg standard	51
BBa_K157013	15 aa flexible glycine-serine protein domain linker; Freiburg standard	45
BBa_K243004	Short Linker (Gly-Gly-Ser-Gly)	12
BBa_K243005	Middle Linker (Gly-Gly-Ser-Gly)x2	24
BBa_K243006	Long Linker (Gly-Gly-Ser-Gly)x3	36
BBa_K243029	GSAT Linker	108
BBa_K243030	SEG	108
BBa_K404300	SEG-Linker	108
BBa_K404301	GSAT-Linker	108
BBa_K404303	Z-EGFR-1907_Short-Linker	192
BBa_K404304	Z-EGFR-1907_Middle-Linker	204
BBa_K404305	Z-EGFR-1907_Long-Linker	216
BBa_K404306	Z-EGFR-1907_SEG-Linker	288
BBa_K416001	(Gly4Ser)3 Flexible Peptide Linker	45
BBa_K648005	Short Fusion Protein Linker: GGSG with standard 25 prefix/suffix	12
BBa_K648006	Long 10AA Fusion Protein Linker with Standard 25 Prefix/Suffix	30
BBa_K648007	Medium 6AA Fusion Protein Linker: GGSGGS with Standard 25 Prefix/Suffix	18

In another embodiment, linkers can comprise a direct bond or an atom such as, *e.g.*, an oxygen (O) or sulfur (S), a unit such as -NR- wherein R is hydrogen or alkyl, -C(O)-, -C(O)O-, -C(O)NH-, SO, SO₂, -SO₂NH- or a chain of atoms, such as substituted or unsubstituted alkyl, substituted or unsubstituted alkenyl, substituted or unsubstituted alkynyl, arylalkyl, heteroarylalkyl. In some embodiments, one or more methylenes in the chain of atoms can be replaced with one or more of O, S, S(O), SO₂, -SO₂NH-, -NR-, -NR₂, -C(O)-, -C(O)O-, -C(O)NH-, a cleavable linking group, substituted or unsubstituted aryl, substituted or unsubstituted heteroaryl, and substituted or unsubstituted heterocyclic.

In some embodiments, the template nucleic acid is attached to the Cas9 molecule through a linker that is itself stable under physiological conditions, such as an alkylene chain, and does not result in release of the donor nucleic acid sequence from the Cas9 molecule for at least 2, 3, 4, 5, 10, 15, 24 or 48 hours or for at least 1, 2, 3, 4, 5 or 10 days when administered to a subject. In some embodiments, the template nucleic acid and the Cas9

molecule comprise residues of a functional groups through which reaction and linkage of the donor nucleic acid sequence to the Cas9 molecule was achieved. In some embodiments, the functional groups, which may be the same or different, terminal or internal, of the donor nucleic acid sequence or Cas9 molecule comprise an amino, acid, imidazole, hydroxyl, thio, acyl halide, $-\text{HC}=\text{CH}-$, $-\text{C}\equiv\text{C}-$ group, or derivative thereof. In some embodiments, the linker comprises a hydrocarbylene group wherein one or more methylene groups is optionally replaced by a group Y (provided that none of the Y groups are adjacent to each other), wherein each Y, independently for each occurrence, is selected from, substituted or unsubstituted aryl, heteroaryl, cycloalkyl, heterocycloalkyl, or $-\text{O}-$, $\text{C}(=\text{X})$ (wherein X is NR_1 , O or S), $-\text{NR}_1-$, $-\text{NR}_1\text{C}(\text{O})-$, $-\text{C}(\text{O})\text{NR}_1-$, $-\text{S}(\text{O})_n-$, $-\text{NR}_1\text{S}(\text{O})_n-$, $-\text{S}(\text{O})_n\text{NR}_1-$, $-\text{NR}_1\text{C}(\text{O})-\text{NR}_1-$; and R_1 , independently for each occurrence, represents H or a lower alkyl and wherein n is 0, 1, or 2.

In some embodiments, the linker comprises an alkylene moiety or a heteroalkylene moiety (*e.g.*, an alkylene glycol moiety such as ethylene glycol). In some embodiments, a linker comprises a poly-L-glutamic acid, polylactic acid, poly(ethyleneimine), an oligosaccharide, an amino acid (*e.g.*, glycine), an amino acid chain, or any other suitable linkage. The linker groups can be biologically inactive, such as a PEG, polyglycolic acid, or polylactic acid chain. In certain embodiments, the linker group represents a derivatized or non-derivatized amino acid (*e.g.*, glycine).

The Cas9 molecule attached to the linker may be any Cas9 molecule described herein, *e.g.*, a nickase Cas9 molecule, or a Cas9 molecule capable of making a double stranded break.

Direct Cross-linking using a Maleimide-modified Template Nucleic Acid

In one embodiment, a maleimide-modified template nucleic acid is cross-linked to a Cas9 molecule to form a Cas9 fusion molecule.

In one embodiment, template nucleic acid is prepared such that some or all of the sequence contain a maleimide-modification. In some embodiments, template nucleic acid is prepared such that some or all of the sequence maleimide-modified at its 5'-end. In one embodiment, the template nucleic acid is single stranded. In an alternate embodiment, the template nucleic acid is double stranded. In one embodiment, the template nucleic acid is DNA, *e.g.*, double stranded DNA. In an alternate embodiment, the template nucleic acid is single stranded DNA.

In one embodiment, the Cas9 molecule is a wild-type molecule. In other embodiments, the Cas9 molecule is not a wild-type molecule. In certain embodiments, the Cas9 molecule has at least one modification. In another embodiment, the Cas9 molecule has at least one modification at a surface-exposed amino acid residue. In other embodiments, the Cas9 molecule has at least one mutation. In other embodiments, the Cas9 molecule has at least one mutation that results in a change from a non-cysteine amino acid residue to a cysteine amino acid residue. In other embodiments, the at least one mutation is a mutation forms a cysteine-variant Cas9 molecule (*e.g.*, a Cas9 molecule having at least one cysteine residue). In certain embodiments, the cysteine-variant Cas9 molecule has more than one cysteine residue.

In other embodiments, the Cas9 fusion molecule is formed by nature of the formation of a covalent bond between the cysteine-variant Cas9 molecule (*e.g.*, a Cas9 molecule with at least one surface exposed cysteine residue) and the maleimide-modified template nucleic acid. Without being bound by theory, in some embodiments, it is believed that the at least one surface exposed thiol groups on the cysteine-variant Cas9 molecule is reactive with the α,β -unsaturated carbonyl component of the maleimide-modified template nucleic acid.

The Cas9 molecule attached to the maleimide-modified template nucleic acid may be any Cas9 molecule described herein, *e.g.*, a nickase Cas9 molecule or a Cas9 molecule capable of making a double stranded break.

Direct Cross-linking via Formation of a Bis-aryl Hydrazine-based Conjugate

In one embodiment, a 5'-4-Formylbenzamide (4FB)-modified template nucleic acid is cross-linked to a 6-hydrazino-nicotinamide (HyNic)-modified Cas9 molecule to form a Cas9 fusion molecule.

In one embodiment, a template nucleic acid is synthesized to have the some or all of the sequence 4FB-modified. In some embodiments, the template nucleic acid is prepared such that the some or all of the sequence 4FB-modified at its 5'-end. In one embodiment, the template nucleic acid is single stranded. In an alternate embodiment, the template nucleic acid is double stranded. In one embodiment, the template nucleic acid is DNA, *e.g.*, double stranded DNA. In an alternate embodiment, the template nucleic acid is single stranded DNA.

In one embodiment, the Cas9 molecule is a wild-type molecule. In other embodiments, the Cas9 molecule is not a wild-type molecule. In certain embodiments, the Cas9 molecule has at least one modification. In another embodiment, the Cas9 molecule has

at least one modification at a surface-exposed amino acid residue. In another embodiment, the Cas9 molecule has at least one modification at a surface-exposed amino acid residue that is characterized as possessing a primary amine group. In some embodiments, the at least one modification forms a 6-hydrazino-nicotinamide (HyNic)-modified Cas9 molecule. In certain
5 embodiments, the variant Cas9 molecule has more than one 6-hydrazino-nicotinamide (HyNic) modification.

In other embodiments, the Cas9 fusion molecule is formed by nature of the formation of a covalent bond between the 6-hydrazino-nicotinamide (HyNic)-modified Cas9 molecule and the 4FB-modified template nucleic acid. Without being bound by theory, in some
10 embodiments, it is believed that a succinimidyl ester functionality of S-HyNic readily reacts with an amine moiety (*e.g.*, a primary amine) on the surface of a Cas9 molecule to form a 6-hydrazino-nicotinamide (HyNic)-modified Cas9 molecule. Without being bound by theory, in some embodiments, it is believed that the at least one surface exposed hydrazino-nicotinamide (HyNic) moiety on the 6-hydrazino-nicotinamide (HyNic)-modified Cas9
15 molecule is reactive with the formylbenzamide (4FB) moiety of the 4FB-modified template nucleic acid to form a stable bis-aryl hydrazine conjugate.

The Cas9 molecule attached to the maleimide-modified template nucleic acid may be any Cas9 molecule described herein, *e.g.*, a nickase Cas9 molecule or a Cas9 molecule capable of making a double stranded break.

Direct Cross-linking using a N-terminal and/or C-terminal Cas9 Fusion Molecule

Nucleic acid sequences encoding a Cas9 fusion molecule are also provided herein.

In one embodiment, a nucleic acid encoding a Cas9 fusion molecule can be a synthetic nucleic acid sequence. For example, the synthetic nucleic acid molecule can further
25 be chemically modified, *e.g.*, as described below.

In one embodiment, a vector can comprise a nucleic acid sequence that encodes a Cas9 fusion molecule. In another embodiment, a vector can comprise a sequence encoding a polypeptide tag (a HaloTag® molecule, a SNAP-tag®, a CLIP-tag®, a ACP-tag® or a MCP-tag®), fused, *e.g.*, to a Cas9 fusion molecule nucleic acid sequence. For example, a vector
30 can comprise a sequence encoding a polypeptide tag (HaloTag® molecule, a SNAP-tag®, a CLIP-tag®, a ACP-tag® or a MCP-tag®) fused to the sequence encoding a Cas9 molecule. In some embodiments, the sequence encoding a polypeptide tag is fused to the N-terminus of the sequence encoding the Cas9 molecule. In other embodiments, the sequence encoding a polypeptide tag is fused to the C-terminus of the sequence encoding the Cas9 molecule. In

certain embodiments, the vector comprises a sequence encoding a linker between the sequence encoding a polypeptide tag and the sequence encoding a Cas9 molecule.

In one embodiment, a Cas9 fusion molecule comprises a polypeptide tag (*e.g.*, a HaloTag® molecule, a SNAP-tag®, a CLIP-tag®, a ACP-tag® or a MCP-tag®) fused to a Cas9 molecule. In certain embodiments, a Cas9 fusion molecule comprises a HaloTag fused to the N-terminus of the Cas9 molecule. In another embodiment, a Cas9 fusion molecule comprises a HaloTag fused to the C-terminus of the Cas9 molecule. In certain embodiments, a Cas9 fusion molecule comprises a HaloTag fused to a linker sequence (*e.g.*, an XTEN linker, a GGS9 linker, a GGS6 linker, or a GGS linker) fused to the N-terminus of the Cas9 molecule. In another embodiment, a Cas9 fusion molecule comprises a HaloTag fused to a linker sequence (*e.g.*, an XTEN linker, a GGS9 linker, a GGS6 linker, or a GGS linker) fused to the C-terminus of the Cas9 molecule.

In one embodiment, the polypeptide tag is a molecule comprising at least one modification. In another embodiment, the polypeptide tag has at least one modification at a surface-exposed amino acid residue. In other embodiments, the polypeptide tag has at least one mutation. In some embodiments, the at least one mutation comprises a H272F mutation (*e.g.*, a HaloTag-variant).

In one embodiment, a template nucleic acid is synthesized to have the some or all of the sequence modified to contain a primary halogen (*e.g.*, a haloalkane-modified template nucleic acid). In some embodiments, a template nucleic acid is synthesized to have the some or all of the sequence modified to contain a primary halogen at its 5'-end. In some embodiments, a template nucleic acid is synthesized to have the some or all of the sequence modified to contain a primary halogen at its 3'-end. In other embodiments, a template nucleic acid is synthesized to have the some or all of the sequence modified to contain a primary halogen at an internal position of the template nucleic acid. In one embodiment, the template nucleic acid is single stranded. In an alternate embodiment, the template nucleic acid is double stranded. In one embodiment, the template nucleic acid is DNA, *e.g.*, double stranded DNA. In an alternate embodiment, the template nucleic acid is single stranded DNA.

In other embodiments, the Cas9 fusion molecule is formed by nature of the formation of a covalent bond between the polypeptide tag (*e.g.*, a HaloTag-variant) of the Cas9 fusion molecule and the primary halogen of the template nucleic acid. Without being bound by theory, in some embodiments, it is believed that certain active site residues of the HaloTag-

variant readily reacts via a nucleophilic attack with the at least one primary halogen on the template nucleic acid to form a stable conjugate.

The Cas9 fusion molecule described herein may be any Cas9 molecule described herein, *e.g.*, a Cas9 nickase molecule, or a Cas9 molecule capable of making a double stranded break.

Direct Cross-linking using an Acrydite-modified Template Nucleic Acid

In one embodiment, an acrydite-modified template nucleic acid is cross-linked to a Cas9 molecule to form a Cas9 fusion molecule.

In one embodiment, a template nucleic acid is prepared such that some or all of the sequence is acrydite-modified. In some embodiments, template nucleic acid is prepared such that some or all of the sequence is acrydite-modified at its 5'-end. In some embodiments, template nucleic acid is prepared such that some or all of the sequence is acrydite-modified at its 3'-end. In some embodiments, template nucleic acid is prepared such that some or all of the sequence is acrydite-modified at an internal position of the template nucleic acid. In one embodiment, the template nucleic acid is single stranded. In an alternate embodiment, the template nucleic acid is double stranded. In one embodiment, the template nucleic acid is DNA, *e.g.*, double stranded DNA. In an alternate embodiment, the template nucleic acid is single stranded DNA.

In one embodiment, the Cas9 molecule is a wild-type molecule. In other embodiments, the Cas9 molecule is not a wild-type molecule. In certain embodiments, the Cas9 molecule has at least one modification. In another embodiment, the Cas9 molecule has at least one modification at a surface-exposed amino acid residue. In other embodiments, the Cas9 molecule has at least one mutation. In other embodiments, the Cas9 molecule has at least one mutation that results in a change from a non-cysteine amino acid residue to a cysteine amino acid residue. In some embodiments, the at least one mutation is a mutation that forms a cysteine-variant Cas9 molecule (*e.g.*, a Cas9 molecule having at least one cysteine residue). In certain embodiments, the cysteine-variant Cas9 molecule has more than one cysteine residue.

In other embodiments, the Cas9 fusion molecule is formed by nature of the formation of a covalent bond between the cysteine-variant Cas9 molecule (*e.g.*, a Cas9 molecule with at least one surface exposed cysteine residue) and the acrydite-modified template nucleic acid. Without being bound by theory, in some embodiments, it is believed that the at least one surface exposed thiol group (*e.g.*, a surface exposed cysteine) on the cysteine-variant Cas9

molecule is reactive with the acrylic acid moiety of the acrydite-modified template nucleic acid.

The Cas9 molecule attached to the acrydite-modified template nucleic acid may be any Cas9 molecule described herein, *e.g.*, a Cas9 nickase molecule, or a Cas9 molecule

5 capable of making a double stranded break.

Direct Cross-linking using Heterobifunctional Crosslinkers EMCH/EDC

In one embodiment, a carboxy-modified template nucleic acid is cross-linked to a *N*-[ϵ -Maleimidocaproic acid] hydrazide (EMCH)-modified Cas9 molecule via 1-ethyl-3-(3-
10 dimethylaminopropyl) carbodiimide (EDC)-based coupling to form a Cas9 fusion molecule.

In one embodiment, a template nucleic acid is prepared such that some or all of the sequence is carboxy-modified (*e.g.*, a template nucleic acid with an exposed carboxyl group). In some embodiments, a template nucleic acid is prepared such that some or all of the sequence is carboxy-modified at its 5'-end. In some embodiments, a template nucleic acid is
15 prepared so that some or all of the sequence carboxy-modified at its 3'-end. In some embodiments, a template nucleic acid is prepared such that some or all of the sequence carboxy-modified at an internal position of the template nucleic acid. In one embodiment, the template nucleic acid is single stranded. In an alternate embodiment, the template nucleic acid is double stranded. In one embodiment, the template nucleic acid is DNA, *e.g.*, double
20 stranded DNA. In an alternate embodiment, the template nucleic acid is single stranded DNA.

In one embodiment, a carboxy-modified template nucleic acid is coupled to 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) to form an activated carboxy-modified template nucleic acid (*e.g.*, a carboxy-modified template nucleic acid coupled to EDC)
25 comprising an *O*-acylisourea ester.

In one embodiment, the Cas9 molecule is a wild-type molecule. In other embodiments, the Cas9 molecule is not a wild-type molecule. In certain embodiments, the Cas9 molecule has at least one modification. In another embodiment, the Cas9 molecule has at least one modification at a surface-exposed amino acid residue. In other embodiments, the
30 Cas9 molecule has at least one mutation. In another embodiment, the Cas9 molecule has at least one mutation at a surface-exposed amino acid residue that contains a thiol group. In some embodiments, the at least one mutation is a mutation that forms a cysteine variant Cas9 molecule (*e.g.*, a Cas9 molecule having at least one cysteine residue). In other embodiments, the Cas9 molecule has at least one mutation that results in a change from a non-cysteine

amino acid residue to a cysteine amino acid residue. In certain embodiments, the variant Cas9 molecule has more than one cysteine residue. In some embodiments, the at least one modification forms a *N*-[ϵ -Maleimidocaproic acid] hydrazide (EMCH)-variant Cas9 molecule. In certain embodiments, the *N*-[ϵ -Maleimidocaproic acid] hydrazide (EMCH)-variant Cas9 molecule has more than one EMCH modification.

In other embodiments, the Cas9 fusion molecule is formed by nature of the formation of a covalent bond between the EMCH-variant Cas9 molecule and the activated carboxy-modified template nucleic acid (*e.g.*, a carboxy-modified template nucleic acid coupled to EDC). Without being bound by theory, in some embodiments, it is believed that a primary amine component of the EMCH-variant Cas9 molecule readily reacts with the *O*-acylisourea ester of the activated carboxy-modified template nucleic acid (*e.g.*, a carboxy-modified template nucleic acid coupled to EDC) to form a stable conjugate.

The Cas9 molecule attached to the maleimide-modified template nucleic acid may be any Cas9 molecule described herein, *e.g.*, a Cas9 nickase molecule, or a Cas9 molecule capable of making a double stranded break.

Cas9-Donor Template Non-covalent Fusions

Cas9 fusion molecules described herein may also be linked to a template nucleic acid by various forms of non-covalent attachment. Non-covalent interactions generally include hydrogen bonds, ionic bonds, van der Waals interactions, and hydrophobic interactions. Non-covalent linkages are described in more detail in the subsections, below.

Non-covalent Attachment using Biotin and Streptavidin to Form Cas9 Fusion Molecules

In one embodiment, a Cas9 molecule is non-covalently attached to a template nucleic acid using biotin and streptavidin to form a Cas9 fusion molecule.

In one embodiment, a Cas9 molecule is covalently linked to a first ligand. In some embodiments, the first ligand is biotin. In certain embodiments, the Cas9 molecule comprises a linker between the Cas9 molecule and the first ligand. In other embodiments, the linker between the Cas9 molecule and the first ligand is sufficiently long to allow the Cas9 molecule to bind to a target nucleic acid and the exogenous donor template sequence without steric interference. In one embodiment, the linker comprises a polypeptide. In other embodiments, the polypeptide comprises serine, glycine, or glycine and serine. In other embodiments, the polypeptide comprises a XTEN-based linker. In some embodiments, the

length of the linker varies from at least 3 amino acid residues to at least 60 amino acids in length. In one embodiment, the first ligand is an affinity ligand (*e.g.*, a high affinity ligand).

In one embodiment, a template nucleic acid is covalently linked to a second ligand. In some embodiments, the second ligand is biotin. In one embodiment, the second ligand is
 5 an affinity ligand (*e.g.*, a high affinity ligand).

In another embodiment, a ligand acceptor protein is bound non-covalently and directly to both the first and second affinity ligand (*e.g.*, a high affinity ligand) at distinct ligand binding sites on a ligand acceptor protein. In certain embodiments, the ligand acceptor protein is streptavidin. Without being bound by theory, in some embodiments, it is believed
 10 that by nature of the ability of the ligand acceptor protein binding to a first and a second affinity ligand (*e.g.*, a high affinity ligand), a Cas9 molecule is non-covalently linked to a template nucleic acid forming a Cas9 fusion molecule.

The Cas9 molecule non-covalently attached to a template nucleic acid may be any Cas9 molecule described herein, *e.g.*, a Cas9 nickase molecule, or a Cas9 molecule capable
 15 of making a double stranded break.

Non-covalent Attachment using Nucleic Acid Binding Proteins to Form Cas9 Fusion Molecules

In one embodiment, a Cas9 molecule is non-covalently attached to a template nucleic acid using a nucleic acid binding protein to form a Cas9 fusion molecule. For example, an
 20 eaCas9 molecule may be covalently linked to a polypeptide, *e.g.*, a nucleic acid binding protein wherein the polypeptide is non-covalently bound to the template nucleic acid.

Nucleic acid binding proteins are well known to one of ordinary skill in the art. For example, nucleic acid binding proteins include, but are not limited to, Rad52, Rad52-yeast,
 25 RPA-4 subunit, BRCA2, Rad51, Rad51B, Rad51C, XRCC2, XRCC3, RecA, RadA, HNRNPA1, UP1 Filament of HNRNPA1, NABP2 (SSB1), NABP1 (SSB2), and UHRF1.

The Cas9 molecule non-covalently attached to a template nucleic acid may be any Cas9 molecule described herein, *e.g.*, a Cas9 nickase molecule, or a Cas9 molecule capable
 30 of making a double stranded break.

Guide RNA (gRNA) Molecules

A gRNA molecule, as that term is used herein, refers to a nucleic acid that promotes the specific targeting or homing of a gRNA molecule/Cas9 molecule complex to a target nucleic acid. gRNA molecules can be unimolecular (having a single RNA molecule) (*e.g.*,

chimeric or modular (comprising more than one, and typically two, separate RNA molecules). The gRNA molecules provided herein comprise a targeting domain comprising, consisting of, or consisting essentially of a nucleic acid sequence fully or partially complementary to a target domain. In certain embodiments, the gRNA molecule further comprises one or more additional domains, including for example a first complementarity domain, a linking domain, a second complementarity domain, a proximal domain, a tail domain, and a 5' extension domain. Each of these domains is discussed in detail below. Additional details on gRNAs are provided in Section I entitled "gRNA molecules" of PCT Application WO 2015/048577, the entire contents of which are expressly incorporated herein by reference. In certain embodiments, one or more of the domains in the gRNA molecule comprises an amino acid sequence identical to or sharing sequence homology with a naturally occurring sequence, *e.g.*, from *S. pyogenes*, *S. aureus*, or *S. thermophilus*.

In certain embodiments, a unimolecular, or chimeric, gRNA comprises, preferably from 5' to 3':

- a targeting domain complementary to a target domain in a target gene;
- a first complementarity domain;
- a linking domain;
- a second complementarity domain (which is complementary to the first complementarity domain);
- a proximal domain; and
- optionally, a tail domain.

In certain embodiments, a modular gRNA comprises:

- a first strand comprising, preferably from 5' to 3':
 - a targeting domain (which is complementary to a target domain in the target gene); and
 - a first complementarity domain; and
- a second strand, comprising, preferably from 5' to 3':
 - optionally, a 5' extension domain;
 - a second complementarity domain;
 - a proximal domain; and
 - optionally, a tail domain.

Each of these domains are described in more detail, below.

Targeting Domain

The targeting domain (sometimes referred to alternatively as the guide sequence or complementarity region) comprises, consists of, or consists essentially of a nucleic acid sequence that is complementary or partially complementary to a target nucleic acid sequence, *e.g.*, a target nucleic acid sequence in a target gene. The nucleic acid sequence in a target gene to which all or a portion of the targeting domain is complementary or partially complementary is referred to herein as the target domain. In certain embodiments, the target domain comprises a target position within the target gene. In other embodiments, a target position lies outside (*i.e.*, upstream or downstream of) the target domain. In certain embodiments, the target domain is located entirely within a target gene, *e.g.*, in a coding region, an intron, or an exon. In other embodiments, all or part of the target domain is located outside of a target gene, *e.g.*, in a control region or in a non-coding region.

Methods for selecting targeting domains are known in the art (see, *e.g.*, Fu 2014; Sternberg 2014).

The strand of the target nucleic acid comprising the target domain is referred to herein as the “complementary strand” because it is complementary to the targeting domain sequence. Since the targeting domain is part of a gRNA molecule, it comprises the base uracil (U) rather than thymine (T); conversely, any DNA molecule encoding the gRNA molecule will comprise thymine rather than uracil. In a targeting domain/target domain pair, the uracil bases in the targeting domain will pair with the adenine bases in the target domain. In certain embodiments, the degree of complementarity between the targeting domain and target domain is sufficient to allow targeting of a Cas9 molecule to the target nucleic acid.

In certain embodiments, the targeting domain comprises a core domain and an optional secondary domain. In certain of these embodiments, the core domain is located 3' to the secondary domain, and in certain of these embodiments the core domain is located at or near the 3' end of the targeting domain. In certain of these embodiments, the core domain consists of or consists essentially of about 8 to about 13 nucleotides at the 3' end of the targeting domain. In certain embodiments, only the core domain is complementary or partially complementary to the corresponding portion of the target domain, and in certain of these embodiments the core domain is fully complementary to the corresponding portion of the target domain. In other embodiments, the secondary domain is also complementary or partially complementary to a portion of the target domain. In certain embodiments, the core domain is complementary or partially complementary to a core domain target in the target domain, while the secondary domain is complementary or partially complementary to a

secondary domain target in the target domain. In certain embodiments, the core domain and secondary domain have the same degree of complementarity with their respective corresponding portions of the target domain. In other embodiments, the degree of complementarity between the core domain and its target and the degree of complementarity between the secondary domain and its target may differ. In certain of these embodiments, the core domain may have a higher degree of complementarity for its target than the secondary domain, whereas in other embodiments the secondary domain may have a higher degree of complementarity than the core domain.

In certain embodiments, the targeting domain and/or the core domain within the targeting domain is 3 to 100, 5 to 100, 10 to 100, or 20 to 100 nucleotides in length, and in certain of these embodiments the targeting domain or core domain is 3 to 15, 3 to 20, 5 to 20, 10 to 20, 15 to 20, 5 to 50, 10 to 50, or 20 to 50 nucleotides in length. In certain embodiments, the targeting domain and/or the core domain within the targeting domain is 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length. In certain embodiments, the targeting domain and/or the core domain within the targeting domain is 6 +/-2, 7 +/-2, 8 +/-2, 9 +/-2, 10 +/-2, 10 +/-4, 10 +/-5, 11 +/-2, 12 +/-2, 13 +/-2, 14 +/-2, 15 +/-2, or 16 +/-2, 20 +/-5, 30 +/-5, 40 +/-5, 50 +/-5, 60 +/-5, 70 +/-5, 80 +/-5, 90 +/-5, or 100 +/-5 nucleotides in length.

In certain embodiments wherein the targeting domain includes a core domain, the core domain is 3 to 20 nucleotides in length, and in certain of these embodiments the core domain 5 to 15 or 8 to 13 nucleotides in length. In certain embodiments wherein the targeting domain includes a secondary domain, the secondary domain is 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14 or 15 nucleotides in length. In certain embodiments wherein the targeting domain comprises a core domain that is 8 to 13 nucleotides in length, the targeting domain is 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, or 16 nucleotides in length, and the secondary domain is 13 to 18, 12 to 17, 11 to 16, 10 to 15, 9 to 14, 8 to 13, 7 to 12, 6 to 11, 5 to 10, 4 to 9, or 3 to 8 nucleotides in length, respectively.

In certain embodiments, the targeting domain is fully complementary to the target domain. Likewise, where the targeting domain comprises a core domain and/or a secondary domain, in certain embodiments one or both of the core domain and the secondary domain are fully complementary to the corresponding portions of the target domain. In other embodiments, the targeting domain is partially complementary to the target domain, and in certain of these embodiments where the targeting domain comprises a core domain and/or a secondary domain, one or both of the core domain and the secondary domain are partially

complementary to the corresponding portions of the target domain. In certain of these embodiments, the nucleic acid sequence of the targeting domain, or the core domain or targeting domain within the targeting domain, is at least 80%, 85%, 90%, or 95%

complementary to the target domain or to the corresponding portion of the target domain. In

5 certain embodiments, the targeting domain and/or the core or secondary domains within the targeting domain include one or more nucleotides that are not complementary with the target domain or a portion thereof, and in certain of these embodiments the targeting domain and/or the core or secondary domains within the targeting domain include 1, 2, 3, 4, 5, 6, 7, or 8 nucleotides that are not complementary with the target domain. In certain embodiments, the

10 core domain includes 1, 2, 3, 4, or 5 nucleotides that are not complementary with the corresponding portion of the target domain. In certain embodiments wherein the targeting domain includes one or more nucleotides that are not complementary with the target domain, one or more of said non-complementary nucleotides are located within five nucleotides of the 5' or 3' end of the targeting domain. In certain of these embodiments, the targeting domain

15 includes 1, 2, 3, 4, or 5 nucleotides within five nucleotides of its 5' end, 3' end, or both its 5' and 3' ends that are not complementary to the target domain. In certain embodiments wherein the targeting domain includes two or more nucleotides that are not complementary to the target domain, two or more of said non-complementary nucleotides are adjacent to one another, and in certain of these embodiments the two or more consecutive non-

20 complementary nucleotides are located within five nucleotides of the 5' or 3' end of the targeting domain. In other embodiments, the two or more consecutive non-complementary nucleotides are both located more than five nucleotides from the 5' and 3' ends of the targeting domain.

In one embodiment, the gRNA molecule, *e.g.*, a gRNA molecule comprising a
25 targeting domain, which is complementary with the target gene, is a modular gRNA molecule. In another embodiment, the gRNA molecule is a unimolecular or chimeric gRNA molecule.

In one embodiment, the nucleic acid encodes a gRNA molecule, *e.g.*, the first gRNA molecule, comprising a targeting domain comprising a sequence that is the same as, or differs
30 by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence described herein. In one embodiment, the nucleic acid encodes a gRNA molecule comprising a targeting domain described herein.

In certain embodiments, the targeting domain comprises 16 nucleotides. In certain embodiments, the targeting domain comprises 17 nucleotides. In certain embodiments, the

targeting domain comprises 18 nucleotides. In certain embodiments, the targeting domain comprises 19 nucleotides. In certain embodiments, the targeting domain comprises 20 nucleotides. In certain embodiments, the targeting domain comprises 21 nucleotides. In certain embodiments, the targeting domain comprises 22 nucleotides. In certain
5 embodiments, the targeting domain comprises 23 nucleotides. In certain embodiments, the targeting domain comprises 24 nucleotides. In certain embodiments, the targeting domain comprises 25 nucleotides. In certain embodiments, the targeting domain comprises 26 nucleotides.

10 In certain embodiments, the targeting domain which is complementary with the target gene is 16 nucleotides or more in length. In certain embodiments, the targeting domain is 16 nucleotides in length. In certain embodiments, the targeting domain is 17 nucleotides in length. In another embodiment, the targeting domain is 18 nucleotides in length. In still another embodiment, the targeting domain is 19 nucleotides in length. In still another
15 embodiment, the targeting domain is 20 nucleotides in length. In still another embodiment, the targeting domain is 21 nucleotides in length. In still another embodiment, the targeting domain is 22 nucleotides in length. In still another embodiment, the targeting domain is 23 nucleotides in length. In still another embodiment, the targeting domain is 24 nucleotides in length. In still another embodiment, the targeting domain is 25 nucleotides in length. In still another embodiment, the targeting domain is 26 nucleotides in length.

20 In one embodiment, a nucleic acid encodes a modular gRNA molecule, *e.g.*, one or more nucleic acids encode a modular gRNA molecule. In another embodiment, a nucleic acid encodes a chimeric gRNA molecule. The nucleic acid may encode a gRNA molecule, *e.g.*, the first gRNA molecule, comprising a targeting domain comprising 16 nucleotides or more in length. In one embodiment, the nucleic acid encodes a gRNA molecule, *e.g.*, the first
25 gRNA molecule, comprising a targeting domain that is 16 nucleotides in length. In another embodiment, the nucleic acid encodes a gRNA molecule, *e.g.*, the first gRNA molecule, comprising a targeting domain that is 17 nucleotides in length. In still another embodiment, the nucleic acid encodes a gRNA molecule, *e.g.*, the first gRNA molecule, comprising a targeting domain that is 18 nucleotides in length. In still another embodiment, the nucleic
30 acid encodes a gRNA molecule, *e.g.*, the first gRNA molecule, comprising a targeting domain that is 19 nucleotides in length. In still another embodiment, the nucleic acid encodes a gRNA molecule, *e.g.*, the first gRNA molecule, comprising a targeting domain that is 20 nucleotides in length. In still another embodiment, the nucleic acid encodes a gRNA molecule, *e.g.*, the first gRNA molecule, comprising a targeting domain that is 21 nucleotides

in length. In still another embodiment, the nucleic acid encodes a gRNA molecule, *e.g.*, the first gRNA molecule, comprising a targeting domain that is 22 nucleotides in length. In still another embodiment, the nucleic acid encodes a gRNA molecule, *e.g.*, the first gRNA molecule, comprising a targeting domain that is 23 nucleotides in length. In still another embodiment, the nucleic acid encodes a gRNA molecule, *e.g.*, the first gRNA molecule, comprising a targeting domain that is 24 nucleotides in length. In still another embodiment, the nucleic acid encodes a gRNA molecule, *e.g.*, the first gRNA molecule, comprising a targeting domain that is 25 nucleotides in length. In still another embodiment, the nucleic acid encodes a gRNA molecule, *e.g.*, the first gRNA molecule, comprising a targeting domain that is 26 nucleotides in length.

In certain embodiments, the targeting domain, core domain, and/or secondary domain do not comprise any modifications. In other embodiments, the targeting domain, core domain, and/or secondary domain, or one or more nucleotides therein, have a modification, including but not limited to the modifications set forth below. In certain embodiments, one or more nucleotides of the targeting domain, core domain, and/or secondary domain may comprise a 2' modification (*e.g.*, a modification at the 2' position on ribose), *e.g.*, a 2-acetylation, *e.g.*, a 2' methylation. In certain embodiments, the backbone of the targeting domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the targeting domain, core domain, and/or secondary domain render the targeting domain and/or the gRNA comprising the targeting domain less susceptible to degradation or more bio-compatible, *e.g.*, less immunogenic. In certain embodiments, the targeting domain and/or the core or secondary domains include 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments the targeting domain and/or core or secondary domains include 1, 2, 3, or 4 modifications within five nucleotides of their respective 5' ends and/or 1, 2, 3, or 4 modifications within five nucleotides of their respective 3' ends. In certain embodiments, the targeting domain and/or the core or secondary domains comprise modifications at two or more consecutive nucleotides.

In certain embodiments wherein the targeting domain includes core and secondary domains, the core and secondary domains contain the same number of modifications. In certain of these embodiments, both domains are free of modifications. In other embodiments, the core domain includes more modifications than the secondary domain, or vice versa.

In certain embodiments, modifications to one or more nucleotides in the targeting domain, including in the core or secondary domains, are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification using a system

as set forth below. gRNAs having a candidate targeting domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated using a system as set forth below. The candidate targeting domain can be placed, either alone or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target, and evaluated.

In certain embodiments, all of the modified nucleotides are complementary to and capable of hybridizing to corresponding nucleotides present in the target domain. In another embodiment, 1, 2, 3, 4, 5, 6, 7 or 8 or more modified nucleotides are not complementary to or capable of hybridizing to corresponding nucleotides present in the target domain.

First and second complementarity domains

The first and second complementarity (sometimes referred to alternatively as the crRNA-derived hairpin sequence and tracrRNA-derived hairpin sequences, respectively) domains are fully or partially complementary to one another. In certain embodiments, the degree of complementarity is sufficient for the two domains to form a duplexed region under at least some physiological conditions. In certain embodiments, the degree of complementarity between the first and second complementarity domains, together with other properties of the gRNA, is sufficient to allow targeting of a Cas9 molecule to a target nucleic acid.

In certain embodiments the first and/or second complementarity domain includes one or more nucleotides that lack complementarity with the corresponding complementarity domain. In certain embodiments, the first and/or second complementarity domain includes 1, 2, 3, 4, 5, or 6 nucleotides that do not complement with the corresponding complementarity domain. For example, the second complementarity domain may contain 1, 2, 3, 4, 5, or 6 nucleotides that do not pair with corresponding nucleotides in the first complementarity domain. In certain embodiments, the nucleotides on the first or second complementarity domain that do not complement with the corresponding complementarity domain loop out from the duplex formed between the first and second complementarity domains. In certain of these embodiments, the unpaired loop-out is located on the second complementarity domain, and in certain of these embodiments the unpaired region begins 1, 2, 3, 4, 5, or 6 nucleotides from the 5' end of the second complementarity domain.

In certain embodiments, the first complementarity domain is 5 to 30, 5 to 25, 7 to 25, 5 to 24, 5 to 23, 7 to 22, 5 to 22, 5 to 21, 5 to 20, 7 to 18, 7 to 15, 9 to 16, or 10 to 14 nucleotides in length, and in certain of these embodiments the first complementarity domain

is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides in length. In certain embodiments, the second complementarity domain is 5 to 27, 7 to 27, 7 to 25, 5 to 24, 5 to 23, 5 to 22, 5 to 21, 7 to 20, 5 to 20, 7 to 18, 7 to 17, 9 to 16, or 10 to 14 nucleotides in length, and in certain of these embodiments the second complementarity domain is 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length. In certain embodiments, the first and second complementarity domains are each independently 6 +/-2, 7 +/-2, 8 +/-2, 9 +/-2, 10 +/-2, 11 +/-2, 12 +/-2, 13 +/-2, 14 +/-2, 15 +/-2, 16 +/-2, 17 +/-2, 18 +/-2, 19 +/-2, or 20 +/-2, 21 +/-2, 22 +/-2, 23 +/-2, or 24 +/-2 nucleotides in length. In certain embodiments, the second complementarity domain is longer than the first complementarity domain, *e.g.*, 2, 3, 4, 5, or 6 nucleotides longer.

In certain embodiments, the first and/or second complementarity domains each independently comprise three subdomains, which, in the 5' to 3' direction are: a 5' subdomain, a central subdomain, and a 3' subdomain. In certain embodiments, the 5' subdomain and 3' subdomain of the first complementarity domain are fully or partially complementary to the 3' subdomain and 5' subdomain, respectively, of the second complementarity domain.

In certain embodiments, the 5' subdomain of the first complementarity domain is 4 to 9 nucleotides in length, and in certain of these embodiments the 5' domain is 4, 5, 6, 7, 8, or 9 nucleotides in length. In certain embodiments, the 5' subdomain of the second complementarity domain is 3 to 25, 4 to 22, 4 to 18, or 4 to 10 nucleotides in length, and in certain of these embodiments the 5' domain is 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides in length. In certain embodiments, the central subdomain of the first complementarity domain is 1, 2, or 3 nucleotides in length. In certain embodiments, the central subdomain of the second complementarity domain is 1, 2, 3, 4, or 5 nucleotides in length. In certain embodiments, the 3' subdomain of the first complementarity domain is 3 to 25, 4 to 22, 4 to 18, or 4 to 10 nucleotides in length, and in certain of these embodiments the 3' subdomain is 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 nucleotides in length. In certain embodiments, the 3' subdomain of the second complementarity domain is 4 to 9, *e.g.*, 4, 5, 6, 7, 8 or 9 nucleotides in length.

The first and/or second complementarity domains can share homology with, or be derived from, naturally occurring or reference first and/or second complementarity domain. In certain of these embodiments, the first and/or second complementarity domains have at least 50%, 60%, 70%, 80%, 85%, 90%, or 95% homology with, or differ by no more than 1, 2, 3, 4, 5, or 6 nucleotides from, the naturally occurring or reference first and/or second

complementarity domain. In certain of these embodiments, the first and/or second complementarity domains may have at least 50%, 60%, 70%, 80%, 85%, 90%, or 95% homology with homology with a first and/or second complementarity domain from *S. pyogenes* or *S. aureus*.

5 In certain embodiments, the first and/or second complementarity domains do not comprise any modifications. In other embodiments, the first and/or second complementarity domains or one or more nucleotides therein have a modification, including but not limited to a modification set forth below. In certain embodiments, one or more nucleotides of the first and/or second complementarity domain may comprise a 2' modification (*e.g.*, a modification
10 at the 2' position on ribose), *e.g.*, a 2-acetylation, *e.g.*, a 2' methylation. In certain embodiments, the backbone of the targeting domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the first and/or second complementarity domain render the first and/or second complementarity domain and/or the gRNA comprising the first and/or second complementarity less susceptible
15 to degradation or more bio-compatible, *e.g.*, less immunogenic. In certain embodiments, the first and/or second complementarity domains each independently include 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments the first and/or second complementarity domains each independently include 1, 2, 3, or 4 modifications within five nucleotides of their respective 5' ends, 3' ends, or both their 5' and 3' ends. In other
20 embodiments, the first and/or second complementarity domains each independently contain no modifications within five nucleotides of their respective 5' ends, 3' ends, or both their 5' and 3' ends. In certain embodiments, one or both of the first and second complementarity domains comprise modifications at two or more consecutive nucleotides.

In certain embodiments, modifications to one or more nucleotides in the first and/or
25 second complementarity domains are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in a system as set forth below. gRNAs having a candidate first or second complementarity domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated in a system as set forth below. The candidate complementarity domain can be placed, either alone or
30 with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target, and evaluated.

In certain embodiments, the duplexed region formed by the first and second complementarity domains is, for example, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, or 22 bp in length, excluding any looped out or unpaired nucleotides.

In certain embodiments, the first and second complementarity domains, when duplexed, comprise 11 paired nucleotides (see, for *e.g.*, gRNA of SEQ ID NO:5). In certain embodiments, the first and second complementarity domains, when duplexed, comprise 15 paired nucleotides (see, *e.g.*, gRNA of SEQ ID NO:27). In certain embodiments, the first and second complementarity domains, when duplexed, comprise 16 paired nucleotides (see, *e.g.*, gRNA of SEQ ID NO:28). In certain embodiments, the first and second complementarity domains, when duplexed, comprise 21 paired nucleotides (see, *e.g.*, gRNA of SEQ ID NO:29).

In certain embodiments, one or more nucleotides are exchanged between the first and second complementarity domains to remove poly-U tracts. For example, nucleotides 23 and 48 or nucleotides 26 and 45 of the gRNA of SEQ ID NO:5 may be exchanged to generate the gRNA of SEQ ID NOs:30 or 31, respectively. Similarly, nucleotides 23 and 39 of the gRNA of SEQ ID NO:29 may be exchanged with nucleotides 50 and 68 to generate the gRNA of SEQ ID NO:32.

Linking domain

The linking domain is disposed between and serves to link the first and second complementarity domains in a unimolecular or chimeric gRNA. In certain embodiments, part of the linking domain is from a crRNA-derived region, and another part is from a tracrRNA-derived region.

In certain embodiments, the linking domain links the first and second complementarity domains covalently. In certain of these embodiments, the linking domain consists of or comprises a covalent bond. In other embodiments, the linking domain links the first and second complementarity domains non-covalently. In certain embodiments, the linking domain is ten or fewer nucleotides in length, *e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides. In other embodiments, the linking domain is greater than 10 nucleotides in length, *e.g.*, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, or 25 or more nucleotides. In certain embodiments, the linking domain is 2 to 50, 2 to 40, 2 to 30, 2 to 20, 2 to 10, 2 to 5, 10 to 100, 10 to 90, 10 to 80, 10 to 70, 10 to 60, 10 to 50, 10 to 40, 10 to 30, 10 to 20, 10 to 15, 20 to 100, 20 to 90, 20 to 80, 20 to 70, 20 to 60, 20 to 50, 20 to 40, 20 to 30, or 20 to 25 nucleotides in length. In certain embodiments, the linking domain is 10 +/-5, 20 +/-5, 20 +/-10, 30 +/-5, 30 +/-10, 40 +/-5, 40 +/-10, 50 +/-5, 50 +/-10, 60 +/-5, 60 +/-10, 70 +/-5, 70 +/-10, 80 +/-5, 80 +/-10, 90 +/-5, 90 +/-10, 100 +/-5, or 100 +/-10 nucleotides in length.

In certain embodiments, the linking domain shares homology with, or is derived from, a naturally occurring sequence, *e.g.*, the sequence of a tracrRNA that is 5' to the second complementarity domain. In certain embodiments, the linking domain has at least 50%, 60%, 70%, 80%, 90%, or 95% homology with or differs by no more than 1, 2, 3, 4, 5, or 6

5 nucleotides from a linking domain disclosed herein.

In certain embodiments, the linking domain does not comprise any modifications. In other embodiments, the linking domain or one or more nucleotides therein have a modification, including but not limited to the modifications set forth below. In certain embodiments, one or more nucleotides of the linking domain may comprise a 2' modification (*e.g.*, a modification at the 2' position on ribose), *e.g.*, a 2-acetylation, *e.g.*, a 2' methylation. In certain embodiments, the backbone of the linking domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the linking domain render the linking domain and/or the gRNA comprising the linking domain less susceptible to degradation or more bio-compatible, *e.g.*, less immunogenic. In certain

10 modifications at the 2' position on ribose), *e.g.*, a 2-acetylation, *e.g.*, a 2' methylation. In certain embodiments, the backbone of the linking domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the linking domain render the linking domain and/or the gRNA comprising the linking domain less susceptible to degradation or more bio-compatible, *e.g.*, less immunogenic. In certain

15 modifications at the 2' position on ribose), *e.g.*, a 2-acetylation, *e.g.*, a 2' methylation. In certain embodiments, the backbone of the linking domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the linking domain render the linking domain and/or the gRNA comprising the linking domain less susceptible to degradation or more bio-compatible, *e.g.*, less immunogenic. In certain

embodiments, the linking domain includes 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments the linking domain includes 1, 2, 3, or 4 modifications within five nucleotides of its 5' and/or 3' end. In certain embodiments, the linking domain comprises modifications at two or more consecutive nucleotides.

In certain embodiments, modifications to one or more nucleotides in the linking domain are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in a system as set forth below. gRNAs having a candidate linking domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated in a system as set forth below. The candidate linking domain can be placed, either alone or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target, and evaluated.

20 domain are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in a system as set forth below. gRNAs having a candidate linking domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated in a system as set forth below. The candidate linking domain can be placed, either alone or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target, and evaluated.

25 molecule/Cas9 molecule system known to be functional with a selected target, and evaluated.

In certain embodiments, the linking domain comprises a duplexed region, typically adjacent to or within 1, 2, or 3 nucleotides of the 3' end of the first complementarity domain and/or the 5' end of the second complementarity domain. In certain of these embodiments, the duplexed region of the linking region is 10+/-5, 15+/-5, 20+/-5, 20+/-10, or 30+/-5 bp in length. In certain embodiments, the duplexed region of the linking domain is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 bp in length. In certain embodiments, the sequences forming the duplexed region of the linking domain are fully complementarity. In other embodiments, one or both of the sequences forming the duplexed region contain one or more nucleotides

30 length. In certain embodiments, the duplexed region of the linking domain is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, or 15 bp in length. In certain embodiments, the sequences forming the duplexed region of the linking domain are fully complementarity. In other embodiments, one or both of the sequences forming the duplexed region contain one or more nucleotides

(*e.g.*, 1, 2, 3, 4, 5, 6, 7, or 8 nucleotides) that are not complementary with the other duplex sequence.

5' extension domain

5 In certain embodiments, a modular gRNA as disclosed herein comprises a 5' extension domain, *i.e.*, one or more additional nucleotides 5' to the second complementarity domain. In certain embodiments, the 5' extension domain is 2 to 10 or more, 2 to 9, 2 to 8, 2 to 7, 2 to 6, 2 to 5, or 2 to 4 nucleotides in length, and in certain of these embodiments the 5' extension domain is 2, 3, 4, 5, 6, 7, 8, 9, or 10 or more nucleotides in length.

10 In certain embodiments, the 5' extension domain nucleotides do not comprise modifications, *e.g.*, modifications of the type provided below. However, in certain embodiments, the 5' extension domain comprises one or more modifications, *e.g.*, modifications that it render it less susceptible to degradation or more bio-compatible, *e.g.*, less immunogenic. By way of example, the backbone of the 5' extension domain can be
15 modified with a phosphorothioate, or other modification(s) as set forth below. In certain embodiments, a nucleotide of the 5' extension domain can comprise a 2' modification (*e.g.*, a modification at the 2' position on ribose), *e.g.*, a 2-acetylation, *e.g.*, a 2' methylation, or other modification(s) as set forth below.

In certain embodiments, the 5' extension domain can comprise as many as 1, 2, 3, 4,
20 5, 6, 7, or 8 modifications. In certain embodiments, the 5' extension domain comprises as many as 1, 2, 3, or 4 modifications within 5 nucleotides of its 5' end, *e.g.*, in a modular gRNA molecule. In certain embodiments, the 5' extension domain comprises as many as 1, 2, 3, or 4 modifications within 5 nucleotides of its 3' end, *e.g.*, in a modular gRNA molecule.

In certain embodiments, the 5' extension domain comprises modifications at two
25 consecutive nucleotides, *e.g.*, two consecutive nucleotides that are within 5 nucleotides of the 5' end of the 5' extension domain, within 5 nucleotides of the 3' end of the 5' extension domain, or more than 5 nucleotides away from one or both ends of the 5' extension domain. In certain embodiments, no two consecutive nucleotides are modified within 5 nucleotides of the 5' end of the 5' extension domain, within 5 nucleotides of the 3' end of the 5' extension
30 domain, or within a region that is more than 5 nucleotides away from one or both ends of the 5' extension domain. In certain embodiments, no nucleotide is modified within 5 nucleotides of the 5' end of the 5' extension domain, within 5 nucleotides of the 3' end of the 5' extension domain, or within a region that is more than 5 nucleotides away from one or both ends of the 5' extension domain.

Modifications in the 5' extension domain can be selected so as to not interfere with gRNA molecule efficacy, which can be evaluated by testing a candidate modification in a system as set forth below. gRNAs having a candidate 5' extension domain having a selected length, sequence, degree of complementarity, or degree of modification, can be evaluated in a system as set forth below. The candidate 5' extension domain can be placed, either alone, or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target and evaluated.

In certain embodiments, the 5' extension domain has at least 60, 70, 80, 85, 90, or 95% homology with, or differs by no more than 1, 2, 3, 4, 5, or 6 nucleotides from, a reference 5' extension domain, *e.g.*, a naturally occurring, *e.g.*, an *S. pyogenes*, *S. aureus*, or *S. thermophilus*, 5' extension domain, or a 5' extension domain described herein.

Proximal domain

In certain embodiments, the proximal domain is 5 to 20 or more nucleotides in length, *e.g.*, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length. In certain of these embodiments, the proximal domain is 6 +/-2, 7 +/-2, 8 +/-2, 9 +/-2, 10 +/-2, 11 +/-2, 12 +/-2, 13 +/-2, 14 +/-2, 14 +/-2, 16 +/-2, 17 +/-2, 18 +/-2, 19 +/-2, or 20 +/-2 nucleotides in length. In certain embodiments, the proximal domain is 5 to 20, 7, to 18, 9 to 16, or 10 to 14 nucleotides in length.

In certain embodiments, the proximal domain can share homology with or be derived from a naturally occurring proximal domain. In certain of these embodiments, the proximal domain has at least 50%, 60%, 70%, 80%, 85%, 90%, or 95% homology with or differs by no more than 1, 2, 3, 4, 5, or 6 nucleotides from a proximal domain disclosed herein, *e.g.*, an *S. pyogenes*, *S. aureus*, or *S. thermophilus* proximal domain.

In certain embodiments, the proximal domain does not comprise any modifications. In other embodiments, the proximal domain or one or more nucleotides therein have a modification, including but not limited to the modifications set forth in herein. In certain embodiments, one or more nucleotides of the proximal domain may comprise a 2' modification (*e.g.*, a modification at the 2' position on ribose), *e.g.*, a 2-acetylation, *e.g.*, a 2' methylation. In certain embodiments, the backbone of the proximal domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the proximal domain render the proximal domain and/or the gRNA comprising the proximal domain less susceptible to degradation or more bio-compatible, *e.g.*, less immunogenic. In certain embodiments, the proximal domain includes 1, 2, 3, 4, 5, 6, 7, or 8

or more modifications, and in certain of these embodiments the proximal domain includes 1, 2, 3, or 4 modifications within five nucleotides of its 5' and/or 3' end. In certain embodiments, the proximal domain comprises modifications at two or more consecutive nucleotides.

5 In certain embodiments, modifications to one or more nucleotides in the proximal domain are selected to not interfere with targeting efficacy, which can be evaluated by testing a candidate modification in a system as set forth below. gRNAs having a candidate proximal domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated in a system as set forth below. The candidate proximal domain
10 can be placed, either alone or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target, and evaluated.

Tail domain

A broad spectrum of tail domains are suitable for use in the gRNA molecules
15 disclosed herein.

In certain embodiments, the tail domain is absent. In other embodiments, the tail domain is 1 to 100 or more nucleotides in length, *e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, or 100 nucleotides in length. In certain embodiments, the tail domain is 1 to 5, 1 to 10, 1 to 15, 1 to 20, 1 to 50, 10 to 100, 20 to 100, 10 to 90, 20 to 90, 10 to 80, 20 to
20 80, 10 to 70, 20 to 70, 10 to 60, 20 to 60, 10 to 50, 20 to 50, 10 to 40, 20 to 40, 10 to 30, 20 to 30, 20 to 25, 10 to 20, or 10 to 15 nucleotides in length. In certain embodiments, the tail domain is 5 +/-5, 10 +/-5, 20+/-10, 20+/-5, 25+/-10, 30+/-10, 30+/-5, 40+/-10, 40+/-5, 50+/-10, 50+/-5, 60+/-10, 60+/-5, 70+/-10, 70+/-5, 80+/-10, 80+/-5, 90+/-10, 90+/-5, 100+/-10, or 100+/-5 nucleotides in length,

25 In certain embodiments, the tail domain can share homology with or be derived from a naturally occurring tail domain or the 5' end of a naturally occurring tail domain. In certain of these embodiments, the proximal domain has at least 50%, 60%, 70%, 80%, 85%, 90%, or 95% homology with or differs by no more than 1, 2, 3, 4, 5, or 6 nucleotides from a naturally occurring tail domain disclosed herein, *e.g.*, an *S. pyogenes*, *S. aureus*, or *S. thermophilus* tail
30 domain.

In certain embodiments, the tail domain includes sequences that are complementary to each other and which, under at least some physiological conditions, form a duplexed region. In certain of these embodiments, the tail domain comprises a tail duplex domain which can form a tail duplexed region. In certain embodiments, the tail duplexed region is 3, 4, 5, 6, 7,

8, 9, 10, 11, or 12 bp in length. In certain embodiments, the tail domain comprises a single stranded domain 3' to the tail duplex domain that does not form a duplex. In certain of these embodiments, the single stranded domain is 3 to 10 nucleotides in length, *e.g.*, 3, 4, 5, 6, 7, 8, 9, 10, or 4 to 6 nucleotides in length.

5 In certain embodiments, the tail domain does not comprise any modifications. In other embodiments, the tail domain or one or more nucleotides therein have a modification, including but not limited to the modifications set forth herein. In certain embodiments, one or more nucleotides of the tail domain may comprise a 2' modification (*e.g.*, a modification at the 2' position on ribose), *e.g.*, a 2-acetylation, *e.g.*, a 2' methylation. In certain
10 embodiments, the backbone of the tail domain can be modified with a phosphorothioate. In certain embodiments, modifications to one or more nucleotides of the tail domain render the tail domain and/or the gRNA comprising the tail domain less susceptible to degradation or more bio-compatible, *e.g.*, less immunogenic. In certain embodiments, the tail domain includes 1, 2, 3, 4, 5, 6, 7, or 8 or more modifications, and in certain of these embodiments
15 the tail domain includes 1, 2, 3, or 4 modifications within five nucleotides of its 5' and/or 3' end. In certain embodiments, the tail domain comprises modifications at two or more consecutive nucleotides.

In certain embodiments, modifications to one or more nucleotides in the tail domain are selected to not interfere with targeting efficacy, which can be evaluated by testing a
20 candidate modification as set forth below. gRNAs having a candidate tail domain having a selected length, sequence, degree of complementarity, or degree of modification can be evaluated using a system as set forth below. The candidate tail domain can be placed, either alone or with one or more other candidate changes in a gRNA molecule/Cas9 molecule system known to be functional with a selected target, and evaluated.

25 In certain embodiments, the tail domain includes nucleotides at the 3' end that are related to the method of *in vitro* or *in vivo* transcription. When a T7 promoter is used for *in vitro* transcription of the gRNA, these nucleotides may be any nucleotides present before the 3' end of the DNA template. When a U6 promoter is used for *in vivo* transcription, these nucleotides may be the sequence UUUUUU. When an H1 promoter is used for transcription,
30 these nucleotides may be the sequence UUUU. When alternate pol-III promoters are used, these nucleotides may be various numbers of uracil bases depending on, *e.g.*, the termination signal of the pol-III promoter, or they may include alternate bases.

In certain embodiments, the proximal and tail domain taken together comprise, consist of, or consist essentially of the sequence set forth in SEQ ID NOs: 33, 34, 35, 36, or 38.

5 Exemplary unimolecular/chimeric gRNAs

In certain embodiments, a gRNA as disclosed herein has the structure: 5' [targeting domain]-[first complementarity domain]-[linking domain]-[second complementarity domain]-[proximal domain]-[tail domain]-3', wherein:

10 the targeting domain comprises a core domain and optionally a secondary domain, and is 10 to 50 nucleotides in length;

the first complementarity domain is 5 to 25 nucleotides in length and, in certain embodiments has at least 50, 60, 70, 80, 85, 90, or 95% homology with a reference first complementarity domain disclosed herein;

the linking domain is 1 to 5 nucleotides in length;

15 the second complementarity domain is 5 to 27 nucleotides in length and, in certain embodiments has at least 50, 60, 70, 80, 85, 90, or 95% homology with a reference second complementarity domain disclosed herein;

the proximal domain is 5 to 20 nucleotides in length and, in certain embodiments has at least 50, 60, 70, 80, 85, 90, or 95% homology with a reference proximal domain disclosed
20 herein; and

the tail domain is absent or a nucleotide sequence is 1 to 50 nucleotides in length and, in certain embodiments has at least 50, 60, 70, 80, 85, 90, or 95% homology with a reference tail domain disclosed herein.

In certain embodiments, a unimolecular gRNA as disclosed herein comprises,
25 preferably from 5' to 3':

a targeting domain, *e.g.*, comprising 10-50 nucleotides;

a first complementarity domain, *e.g.*, comprising 15, 16, 17, 18, 19, 20, 21, 22,
23, 24, 25, or 26 nucleotides;

a linking domain;

30 a second complementarity domain;

a proximal domain; and

a tail domain,

wherein,

- (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides;
- (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain; or
- 5 (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the sequence from (a), (b), and/or (c) has at least 50%, 60%, 10 70%, 75%, 80%, 85%, 90%, 95%, or 99% homology with the corresponding sequence of a naturally occurring gRNA, or with a gRNA described herein.

In certain embodiments, the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 15 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that are complementary to the corresponding nucleotides of the first complementarity domain.

In certain embodiments, the targeting domain consists of, consists essentially of, or 20 comprises 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides (*e.g.*, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 consecutive nucleotides) complementary or partially complementary to the target domain or a portion thereof, *e.g.*, the targeting domain is 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length. In certain of these embodiments, the targeting domain is complementary to the target domain over the entire length of the targeting domain, the 25 entire length of the target domain, or both.

In certain embodiments, a unimolecular or chimeric gRNA molecule disclosed herein (comprising a targeting domain, a first complementary domain, a linking domain, a second complementary domain, a proximal domain and, optionally, a tail domain) comprises the amino acid sequence set forth in SEQ ID NO:45, wherein the targeting domain is listed as 30 N's (residues 1-20) but may range in length from 16 to 26 nucleotides, and wherein the final six residues (residues 97-102) represent a termination signal for the U6 promoter but may be absent or fewer in number. In certain embodiments, the unimolecular, or chimeric, gRNA molecule is a *S. pyogenes* gRNA molecule.

In certain embodiments, a unimolecular or chimeric gRNA molecule disclosed herein (comprising a targeting domain, a first complementary domain, a linking domain, a second complementary domain, a proximal domain and, optionally, a tail domain) comprises the amino acid sequence set forth in SEQ ID NO:40, wherein the targeting domain is listed as 20
 5 Ns (residues 1-20) but may range in length from 16 to 26 nucleotides, and wherein the final six residues (residues 97-102) represent a termination signal for the U6 promoter but may be absent or fewer in number. In certain embodiments, the unimolecular or chimeric gRNA molecule is an *S. aureus* gRNA molecule.

10 Exemplary modular gRNAs

In certain embodiments, a modular gRNA disclosed herein comprises:

- a first strand comprising, preferably from 5' to 3';
 - a targeting domain, *e.g.*, comprising 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides;
 - 15 a first complementarity domain; and
- a second strand, comprising, preferably from 5' to 3':
 - optionally a 5' extension domain;
 - a second complementarity domain;
 - a proximal domain; and
 - 20 a tail domain,

wherein:

- (a) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides;
- (b) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 25 3' to the last nucleotide of the second complementarity domain; or
- (c) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

30 In certain embodiments, the sequence from (a), (b), or (c), has at least 60, 75, 80, 85, 90, 95, or 99% homology with the corresponding sequence of a naturally occurring gRNA, or with a gRNA described herein.

In certain embodiments, the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is

5 complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides (*e.g.*, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides in length.

10 In certain embodiments, the targeting domain consists of, consists essentially of, or comprises 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 nucleotides (*e.g.*, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, or 26 consecutive nucleotides) complementary to the target domain or a portion thereof. In certain of these embodiments, the targeting domain is complementary to the target domain over the entire length of the targeting domain, the entire length of the target
15 domain, or both.

In certain embodiments, the targeting domain comprises, has, or consists of, 16 nucleotides (*e.g.*, 16 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 16 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or
20 53 nucleotides.

In certain embodiments, the targeting domain comprises, has, or consists of, 16 nucleotides (*e.g.*, 16 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 16 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second
25 complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 16 nucleotides (*e.g.*, 16 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 16 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second
30 complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain has, or consists of, 17 nucleotides (*e.g.*, 17 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the

targeting domain is 17 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, the targeting domain has, or consists of, 17 nucleotides (*e.g.*, 17 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the
5 targeting domain is 17 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, the targeting domain has, or consists of, 17 nucleotides (*e.g.*, 17 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the
10 targeting domain is 17 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain has, or consists of, 18 nucleotides (*e.g.*,
15 18 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 18 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, the targeting domain has, or consists of, 18 nucleotides (*e.g.*, 18 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the
20 targeting domain is 18 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, the targeting domain has, or consists of, 18 nucleotides (*e.g.*, 18 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the
25 targeting domain is 18 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 19
30 nucleotides (*e.g.*, 19 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 19 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, the targeting domain comprises, has, or consists of, 19 nucleotides (*e.g.*, 19 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 19 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 19 nucleotides (*e.g.*, 19 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 19 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 20 nucleotides (*e.g.*, 20 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 20 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, the targeting domain comprises, has, or consists of, 20 nucleotides (*e.g.*, 20 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 20 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 20 nucleotides (*e.g.*, 20 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 20 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 21 nucleotides (*e.g.*, 21 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 21 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, the targeting domain comprises, has, or consists of, 21 nucleotides (*e.g.*, 21 consecutive nucleotides) having complementarity with the target

domain, *e.g.*, the targeting domain is 21 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 21 nucleotides (*e.g.*, 21 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 21 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 22 nucleotides (*e.g.*, 22 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 22 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, the targeting domain comprises, has, or consists of, 22 nucleotides (*e.g.*, 22 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 22 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 22 nucleotides (*e.g.*, 22 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 22 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 23 nucleotides (*e.g.*, 23 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 23 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, the targeting domain comprises, has, or consists of, 23 nucleotides (*e.g.*, 23 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 23 nucleotides in length; and there are at least 15, 18,

20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 23 nucleotides (*e.g.*, 23 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 23 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 24 nucleotides (*e.g.*, 24 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 24 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, the targeting domain comprises, has, or consists of, 24 nucleotides (*e.g.*, 24 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 24 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 24 nucleotides (*e.g.*, 24 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 24 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 25 nucleotides (*e.g.*, 25 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 25 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, the targeting domain comprises, has, or consists of, 25 nucleotides (*e.g.*, 25 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 25 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 25 nucleotides (*e.g.*, 25 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 25 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 26 nucleotides (*e.g.*, 26 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 26 nucleotides in length; and the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides.

In certain embodiments, the targeting domain comprises, has, or consists of, 26 nucleotides (*e.g.*, 26 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 26 nucleotides in length; and there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain.

In certain embodiments, the targeting domain comprises, has, or consists of, 26 nucleotides (*e.g.*, 26 consecutive nucleotides) having complementarity with the target domain, *e.g.*, the targeting domain is 26 nucleotides in length; and there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain.

Methods for Designing gRNA Molecules

Methods for selecting, designing, and validating targeting domains for use in the gRNAs described herein are provided. Exemplary targeting domains for incorporation into gRNAs are also provided herein.

Methods for selection and validation of target sequences as well as off-target analyses have been described (see, *e.g.*, Mali 2013; Hsu 2013; Fu 2014; Heigwer 2014; Bae 2014; and Xiao 2014). For example, a software tool can be used to optimize the choice of potential targeting domains corresponding to a user's target sequence, *e.g.*, to minimize total off-target activity across the genome. Off-target activity may be other than cleavage. For each possible targeting domain choice using *S. pyogenes* Cas9, the tool can identify all off-target sequences (preceding either NAG or NGG PAMs) across the genome that contain up to certain number

(*e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10) of mismatched base-pairs. The cleavage efficiency at each off-target sequence can be predicted, *e.g.*, using an experimentally-derived weighting scheme. Each possible targeting domain is then ranked according to its total predicted off-target cleavage; the top-ranked targeting domains represent those that are likely to have the greatest on-target cleavage and the least off-target cleavage. Other functions, *e.g.*, automated reagent design for CRISPR construction, primer design for the on-target Surveyor assay, and primer design for high-throughput detection and quantification of off-target cleavage via next-gen sequencing, can also be included in the tool. Candidate targeting domains and gRNAs comprising those targeting domains can be functionally evaluated by using methods known in the art and/or as set forth herein.

As a non-limiting example, targeting domains for use in gRNAs for use with *S. pyogenes*, and *S. aureus* Cas9s were identified using a DNA sequence searching algorithm. 17-mer and 20-mer targeting domains were designed for *S. pyogenes* targets, while 18-mer, 19-mer, 20-mer, 21-mer, 22-mer, 23-mer, and 24-mer targeting domains were designed for *S. aureus* targets. gRNA design was carried out using a custom gRNA design software based on the public tool cas-offinder (Bae 2014). This software scores guides after calculating their genome-wide off-target propensity. Typically matches ranging from perfect matches to 7 mismatches are considered for guides ranging in length from 17 to 24. Once the off-target sites are computationally-determined, an aggregate score is calculated for each guide and summarized in a tabular output using a web-interface. In addition to identifying potential target sites adjacent to PAM sequences, the software also identifies all PAM adjacent sequences that differ by 1, 2, 3 or more than 3 nucleotides from the selected target sites. Genomic DNA sequences for a target gene may be obtained from the UCSC Genome browser and sequences screened for repeat elements using the publically available RepeatMasker program. RepeatMasker searches input DNA sequences for repeated elements and regions of low complexity. The output is a detailed annotation of the repeats present in a given query sequence.

Following identification, targeting domains were ranked into tiers based on their distance to the target site, their orthogonality and presence of a 5' G (based on identification of close matches in the human genome containing a relevant PAM *e.g.*, NGG PAM for *S. pyogenes*, NNGRRT or NNGRRV PAM for *S. aureus*). Orthogonality refers to the number of sequences in the human genome that contain a minimum number of mismatches to the target sequence. A "high level of orthogonality" or "good orthogonality" may, for example, refer to 20-mer targeting domains that have no identical sequences in the human genome besides the

intended target, nor any sequences that contain one or two mismatches in the target sequence. Targeting domains with good orthogonality are selected to minimize off-target DNA cleavage.

Targeting domains were identified for both single-gRNA nuclease cleavage and for a dual-gRNA paired “nickase” strategy. Criteria for selecting targeting domains and the determination of which targeting domains can be incorporated into a gRNA and used for the dual-gRNA paired “nickase” strategy is based on two considerations:

1. gRNA pairs should be oriented on the DNA such that PAMs are facing out and cutting with the D10A Cas9 nickase will result in 5' overhangs.
2. An assumption that cleaving with dual nickase pairs will result in deletion of the entire intervening sequence at a reasonable frequency. However, cleaving with dual nickase pairs can also result in indel mutations at the site of only one of the gRNA molecules. Candidate pair members can be tested for how efficiently they remove the entire sequence versus causing indel mutations at the target site of one gRNA molecule.

Other gRNA Design Strategy

In certain embodiments, two or more (*e.g.*, three or four) gRNA molecules are used with one Cas9 molecule. In another embodiment, when two or more (*e.g.*, three or four) gRNAs are used with two or more Cas9 molecules, at least one Cas9 molecule is from a different species than the other Cas9 molecule(s). For example, when two gRNA molecules are used with two Cas9 molecules, one Cas9 molecule can be from one species and the other Cas9 molecule can be from a different species. Both Cas9 species are used to generate a single or double-strand break, as desired.

In certain embodiments, dual targeting is used to create two nicks on opposite DNA strands by using Cas9 nickases (*e.g.*, a *S. pyogenes* Cas9 nickase) with two targeting domains that are complementary to opposite DNA strands, *e.g.*, a gRNA molecule comprising any minus strand targeting domain may be paired any gRNA molecule comprising a plus strand targeting domain provided that the two gRNAs are oriented on the DNA such that PAMs face outward and the distance between the 5' ends of the gRNAs is 0-50bp. When selecting gRNA molecules for use in a nickase pair, one gRNA molecule targets a domain in the complementary strand and the second gRNA molecule targets a domain in the non-complementary strand, *e.g.*, a gRNA comprising any minus strand targeting domain may be paired any gRNA molecule comprising a plus strand targeting domain targeting the same

target position. In certain embodiments, two 20-mer gRNAs are used to target two Cas9 nucleases (e.g., two *S. pyogenes* Cas9 nucleases) or two Cas9 nickases (e.g., two *S. pyogenes* Cas9 nickases). Any of the targeting domains described herein can be used with a Cas9 molecule that generates a single-strand break (i.e., a *S. pyogenes* or *S. aureus* Cas9 nickase) or with a Cas9 molecule that generates a double-strand break (i.e., *S. pyogenes* or *S. aureus* Cas9 nuclease).

gRNA molecules, as described herein, may comprise from 5' to 3': a targeting domain (comprising a "core domain", and optionally a "secondary domain"); a first complementarity domain; a linking domain; a second complementarity domain; a proximal domain; and a tail domain. In one embodiment, the proximal domain and tail domain are taken together as a single domain.

In one embodiment, a gRNA molecule comprises a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 20 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

In another embodiment, a gRNA molecule comprises a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 25 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

In another embodiment, a gRNA molecule comprises a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 30 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

In another embodiment, a gRNA molecule comprises a linking domain of no more than 25 nucleotides in length; a proximal and tail domain, that taken together, are at least 40 nucleotides in length; and a targeting domain equal to or greater than 16, 17, 18, 19, 20, 21, 22, 23, 24, 25 or 26 nucleotides in length.

When two gRNAs are designed for use with two Cas9 molecules, the two Cas9 molecules may be from different species. Both Cas9 species may be used to generate a single or double strand break, as desired.

It is contemplated herein that any upstream gRNA described herein may be paired with any downstream gRNA described herein. When an upstream gRNA designed for use with one species of Cas9 molecule is paired with a downstream gRNA designed for use from

a different species of Cas9 molecule, both Cas9 species are used to generate a single or double-strand break, as desired.

Cas9 Molecules

5 Cas9 molecules of a variety of species can be used in the methods and compositions described herein. While *S. pyogenes* and *S. aureus* Cas9 molecules are the subject of much of the disclosure herein, Cas9 molecules of, derived from, or based on the Cas9 proteins of other species listed herein can be used as well. These include, for example, Cas9 molecules from *Acidovorax avenae*, *Actinobacillus pleuropneumoniae*, *Actinobacillus succinogenes*,
 10 *Actinobacillus suis*, *Actinomyces sp.*, *cycliphilus denitrificans*, *Aminomonas paucivorans*, *Bacillus cereus*, *Bacillus smithii*, *Bacillus thuringiensis*, *Bacteroides sp.*, *Blastopirellula marina*, *Bradyrhizobium sp.*, *Brevibacillus laterosporus*, *Campylobacter coli*, *Campylobacter jejuni*, *Campylobacter lari*, *Candidatus Puniceispirillum*, *Clostridium cellulolyticum*, *Clostridium perfringens*, *Corynebacterium accolens*, *Corynebacterium diphtheria*,
 15 *Corynebacterium matruchotii*, *Dinoroseobacter shibae*, *Eubacterium dolichum*, *gamma proteobacterium*, *Gluconacetobacter diazotrophicus*, *Haemophilus parainfluenzae*, *Haemophilus sputorum*, *Helicobacter canadensis*, *Helicobacter cinaedi*, *Helicobacter mustelae*, *Ilyobacter polytropus*, *Kingella kingae*, *Lactobacillus crispatus*, *Listeria ivanovii*, *Listeria monocytogenes*, *Listeriaceae bacterium*, *Methylocystis sp.*, *Methylosinus*
 20 *trichosporium*, *Mobiluncus mulieris*, *Neisseria bacilliformis*, *Neisseria cinerea*, *Neisseria flavescens*, *Neisseria lactamica*, *Neisseria meningitidis*, *Neisseria sp.*, *Neisseria wadsworthii*, *Nitrosomonas sp.*, *Parvibaculum lavamentivorans*, *Pasteurella multocida*, *Phascolarctobacterium succinatutens*, *Ralstonia syzygii*, *Rhodopseudomonas palustris*, *Rhodovulum sp.*, *Simonsiella muelleri*, *Sphingomonas sp.*, *Sporolactobacillus vineae*,
 25 *Staphylococcus lugdunensis*, *Streptococcus sp.*, *Subdoligranulum sp.*, *Tistrella mobilis*, *Treponema sp.*, or *Verminephrobacter eiseniae*. In some embodiments, the Cas9 molecule is a split Cas9 molecule or an inducible Cas9 molecule, as described in more detail in WO15/089427 and WO14/018423, the entire contents of each of which are expressly incorporated herein by reference.

30

Cas9 Domains

Crystal structures have been determined for two different naturally occurring bacterial Cas9 molecules (Jinek *et al.* 2014) and for *S. pyogenes* Cas9 with a guide RNA (*e.g.*, a synthetic fusion of crRNA and tracrRNA) (Nishimasu *et al.* 2014; and Anders 2014).

A naturally-occurring Cas9 molecule comprises two lobes: a recognition (REC) lobe and a nuclease (NUC) lobe; each of which further comprise domains described herein. The domain nomenclature and the numbering of the amino acid residues encompassed by each domain used throughout this disclosure is as described previously in (Nishimasu 2014). The numbering of the amino acid residues is with reference to Cas9 from *S. pyogenes*.

The REC lobe comprises the arginine-rich bridge helix (BH), the REC1 domain, and the REC2 domain. The REC lobe does not share structural similarity with other known proteins, indicating that it is a Cas9-specific functional domain. The BH domain is a long α helix and arginine rich region and comprises amino acids 60-93 of the sequence of *S.*

pyogenes Cas9. The REC1 domain is important for recognition of the repeat:anti-repeat duplex, *e.g.*, of a gRNA or a tracrRNA, and is therefore critical for Cas9 activity by recognizing the target sequence. The REC1 domain comprises two REC1 motifs at amino acids 94 to 179 and 308 to 717 of the sequence of *S. pyogenes* Cas9. These two REC1 domains, though separated by the REC2 domain in the linear primary structure, assemble in the tertiary structure to form the REC1 domain. The REC2 domain, or parts thereof, may also play a role in the recognition of the repeat:anti-repeat duplex. The REC2 domain comprises amino acids 180-307 of the sequence of *S. pyogenes* Cas9.

The NUC lobe comprises the RuvC domain, the HNH domain, and the PAM-interacting (PI) domain. The RuvC domain shares structural similarity to retroviral integrase superfamily members and cleaves a single strand, *e.g.*, the non-complementary strand of the target nucleic acid molecule. The RuvC domain is assembled from the three split RuvC motifs (RuvC I, RuvCII, and RuvCIII, which are often commonly referred to in the art as RuvCI domain, or N-terminal RuvC domain, RuvCII domain, and RuvCIII domain) at amino acids 1-59, 718-769, and 909-1098, respectively, of the sequence of *S. pyogenes* Cas9.

Similar to the REC1 domain, the three RuvC motifs are linearly separated by other domains in the primary structure, however in the tertiary structure, the three RuvC motifs assemble and form the RuvC domain. The HNH domain shares structural similarity with HNH endonucleases, and cleaves a single strand, *e.g.*, the complementary strand of the target nucleic acid molecule. The HNH domain lies between the RuvC II-III motifs and comprises amino acids 775-908 of the sequence of *S. pyogenes* Cas9. The PI domain interacts with the PAM of the target nucleic acid molecule, and comprises amino acids 1099-1368 of the sequence of *S. pyogenes* Cas9.

RuvC-like domain and an HNH-like domain

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain and a RuvC-like domain and in certain of these embodiments cleavage activity is dependent on the RuvC-like domain and the HNH-like domain. A Cas9 molecule or Cas9 polypeptide can comprise one or more of a RuvC-like domain and an HNH-like domain. In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises a RuvC-like domain, *e.g.*, a RuvC-like domain described below, and/or an HNH-like domain, *e.g.*, an HNH-like domain described below.

RuvC-like domains

In certain embodiments, a RuvC-like domain cleaves, a single strand, *e.g.*, the non-complementary strand of the target nucleic acid molecule. The Cas9 molecule or Cas9 polypeptide can include more than one RuvC-like domain (*e.g.*, one, two, three or more RuvC-like domains). In certain embodiments, a RuvC-like domain is at least 5, 6, 7, 8 amino acids in length but not more than 20, 19, 18, 17, 16 or 15 amino acids in length. In certain embodiments, the Cas9 molecule or Cas9 polypeptide comprises an N-terminal RuvC-like domain of about 10 to 20 amino acids, *e.g.*, about 15 amino acids in length.

N-terminal RuvC-like domains

Some naturally occurring Cas9 molecules comprise more than one RuvC-like domain with cleavage being dependent on the N-terminal RuvC-like domain. Accordingly, a Cas9 molecule or Cas9 polypeptide can comprise an N-terminal RuvC-like domain. Exemplary N-terminal RuvC-like domains are described below.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an N-terminal RuvC-like domain comprising an amino acid sequence of Formula I:

D-X₁-G-X₂-X₃-X₄-X₅-G-X₆-X₇-X₈-X₉ (SEQ ID NO: 8),

wherein,

X₁ is selected from I, V, M, L and T (*e.g.*, selected from I, V, and L);

X₂ is selected from T, I, V, S, N, Y, E and L (*e.g.*, selected from T, V, and I);

X₃ is selected from N, S, G, A, D, T, R, M and F (*e.g.*, A or N);

X₄ is selected from S, Y, N and F (*e.g.*, S);

X₅ is selected from V, I, L, C, T and F (*e.g.*, selected from V, I and L);

X₆ is selected from W, F, V, Y, S and L (*e.g.*, W);

X₇ is selected from A, S, C, V and G (*e.g.*, selected from A and S);

X₈ is selected from V, I, L, A, M and H (*e.g.*, selected from V, I, M and L); and

X₉ is selected from any amino acid or is absent (*e.g.*, selected from T, V, I, L, Δ, F, S, A, Y, M and R, or, *e.g.*, selected from T, V, I, L and Δ).

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of
5 SEQ ID NO:8, by as many as 1 but no more than 2, 3, 4, or 5 residues.

In certain embodiments, the N-terminal RuvC-like domain is cleavage competent.

In other embodiments, the N-terminal RuvC-like domain is cleavage incompetent.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an N-terminal RuvC-like domain comprising an amino acid sequence of Formula II:

10 D-X₁-G-X₂-X₃-S-X₅-G-X₆-X₇-X₈-X₉, (SEQ ID NO: 9),

wherein

X₁ is selected from I, V, M, L and T (*e.g.*, selected from I, V, and L);

X₂ is selected from T, I, V, S, N, Y, E and L (*e.g.*, selected from T, V, and I);

X₃ is selected from N, S, G, A, D, T, R, M and F (*e.g.*, A or N);

15 X₅ is selected from V, I, L, C, T and F (*e.g.*, selected from V, I and L);

X₆ is selected from W, F, V, Y, S and L (*e.g.*, W);

X₇ is selected from A, S, C, V and G (*e.g.*, selected from A and S);

X₈ is selected from V, I, L, A, M and H (*e.g.*, selected from V, I, M and L); and

20 X₉ is selected from any amino acid or is absent (*e.g.*, selected from T, V, I, L, Δ, F, S, A, Y, M and R or selected from *e.g.*, T, V, I, L and Δ).

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of
SEQ ID NO:9 by as many as 1 but not more than 2, 3, 4, or 5 residues.

In certain embodiments, the N-terminal RuvC-like domain comprises an amino acid
sequence of Formula III:

25 D-I-G-X₂-X₃-S-V-G-W-A-X₈-X₉ (SEQ ID NO: 10),

wherein

X₂ is selected from T, I, V, S, N, Y, E and L (*e.g.*, selected from T, V, and I);

X₃ is selected from N, S, G, A, D, T, R, M and F (*e.g.*, A or N);

X₈ is selected from V, I, L, A, M and H (*e.g.*, selected from V, I, M and L); and

30 X₉ is selected from any amino acid or is absent (*e.g.*, selected from T, V, I, L, Δ, F, S, A, Y, M and R or selected from *e.g.*, T, V, I, L and Δ).

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of
SEQ ID NO:10 by as many as 1 but not more than, 2, 3, 4, or 5 residues.

In certain embodiments, the N-terminal RuvC-like domain comprises an amino acid sequence of Formula IV:

D-I-G-T-N-S-V-G-W-A-V-X (SEQ ID NO: 11),

wherein

5 X is a non-polar alkyl amino acid or a hydroxyl amino acid, *e.g.*, X is selected from V, I, L and T.

In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of SEQ ID NO:11 by as many as 1 but not more than, 2, 3, 4, or 5 residues.

10 In certain embodiments, the N-terminal RuvC-like domain differs from a sequence of an N-terminal RuvC like domain disclosed herein, *e.g.*, in any one of SEQ ID Nos: 54-103, as many as 1 but no more than 2, 3, 4, or 5 residues. In certain embodiments, 1, 2, 3 or all of the highly conserved residues of SEQ ID Nos: 54-103 are present.

In certain embodiment, the N-terminal RuvC-like domain differs from a sequence of an N-terminal RuvC-like domain disclosed herein, *e.g.*, in any one of SEQ ID Nos: 104-177,
15 as many as 1 but no more than 2, 3, 4, or 5 residues. In certain embodiments, 1, 2, or all of the highly conserved residues identified of SEQ ID Nos: 104-177 are present.

Additional RuvC-like domains

In addition to the N-terminal RuvC-like domain, the Cas9 molecule or Cas9
20 polypeptide can comprise one or more additional RuvC-like domains. In certain embodiments, the Cas9 molecule or Cas9 polypeptide can comprise two additional RuvC-like domains. Preferably, the additional RuvC-like domain is at least 5 amino acids in length and, *e.g.*, less than 15 amino acids in length, *e.g.*, 5 to 10 amino acids in length, *e.g.*, 8 amino acids in length.

25 An additional RuvC-like domain can comprise an amino acid sequence of Formula V: I-X₁-X₂-E-X₃-A-R-E (SEQ ID NO:12), wherein

X₁ is V or H;

X₂ is I, L or V (*e.g.*, I or V); and

X₃ is M or T.

30 In certain embodiments, the additional RuvC-like domain comprises an amino acid sequence of Formula VI:

I-V-X₂-E-M-A-R-E (SEQ ID NO:13), wherein

X₂ is I, L or V (*e.g.*, I or V).

An additional RuvC-like domain can comprise an amino acid sequence of Formula VII:

H-H-A-X₁-D-A-X₂-X₃ (SEQ ID NO: 14), wherein

X₁ is H or L;

5 X₂ is R or V; and

X₃ is E or V.

In certain embodiments, the additional RuvC-like domain comprises the amino acid sequence: H-H-A-H-D-A-Y-L (SEQ ID NO:15).

10 In certain embodiments, the additional RuvC-like domain differs from a sequence of SEQ ID NOs: 12-15 by as many as 1 but not more than 2, 3, 4, or 5 residues.

In certain embodiment, the sequence flanking the N-terminal RuvC-like domain has the amino acid sequence of Formula VIII:

K-X₁'-Y-X₂'-X₃'-X₄'-Z-T-D-X₉'-Y, (SEQ ID NO: 16).

wherein

15 X₁' is selected from K and P;

X₂' is selected from V, L, I, and F (*e.g.*, V, I and L);

X₃' is selected from G, A and S (*e.g.*, G);

X₄' is selected from L, I, V and F (*e.g.*, L);

X₉' is selected from D, E, N and Q; and

20 Z is an N-terminal RuvC-like domain, *e.g.*, as described above, *e.g.*, having 5 to 20 amino acids.

HNH-like domains

In certain embodiments, an HNH-like domain cleaves a single stranded
25 complementary domain, *e.g.*, a complementary strand of a double stranded nucleic acid molecule. In certain embodiments, an HNH-like domain is at least 15, 20, or 25 amino acids in length but not more than 40, 35, or 30 amino acids in length, *e.g.*, 20 to 35 amino acids in length, *e.g.*, 25 to 30 amino acids in length. Exemplary HNH-like domains are described below.

30 In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain having an amino acid sequence of Formula IX:

X₁-X₂-X₃-H-X₄-X₅-P-X₆-X₇-X₈-X₉-X₁₀-X₁₁-X₁₂-X₁₃-X₁₄-X₁₅-N-X₁₆-X₁₇-X₁₈-X₁₉-X₂₀-X₂₁-
X₂₂-X₂₃-N (SEQ ID NO: 17), wherein

X₁ is selected from D, E, Q and N (*e.g.*, D and E);

X₂ is selected from L, I, R, Q, V, M and K;

X₃ is selected from D and E;

X₄ is selected from I, V, T, A and L (*e.g.*, A, I and V);

X₅ is selected from V, Y, I, L, F and W (*e.g.*, V, I and L);

5 X₆ is selected from Q, H, R, K, Y, I, L, F and W;

X₇ is selected from S, A, D, T and K (*e.g.*, S and A);

X₈ is selected from F, L, V, K, Y, M, I, R, A, E, D and Q (*e.g.*, F);

X₉ is selected from L, R, T, I, V, S, C, Y, K, F and G;

X₁₀ is selected from K, Q, Y, T, F, L, W, M, A, E, G, and S;

10 X₁₁ is selected from D, S, N, R, L and T (*e.g.*, D);

X₁₂ is selected from D, N and S;

X₁₃ is selected from S, A, T, G and R (*e.g.*, S);

X₁₄ is selected from I, L, F, S, R, Y, Q, W, D, K and H (*e.g.*, I, L and F);

X₁₅ is selected from D, S, I, N, E, A, H, F, L, Q, M, G, Y and V;

15 X₁₆ is selected from K, L, R, M, T and F (*e.g.*, L, R and K);

X₁₇ is selected from V, L, I, A and T;

X₁₈ is selected from L, I, V and A (*e.g.*, L and I);

X₁₉ is selected from T, V, C, E, S and A (*e.g.*, T and V);

X₂₀ is selected from R, F, T, W, E, L, N, C, K, V, S, Q, I, Y, H and A;

20 X₂₁ is selected from S, P, R, K, N, A, H, Q, G and L;

X₂₂ is selected from D, G, T, N, S, K, A, I, E, L, Q, R and Y; and

X₂₃ is selected from K, V, A, E, Y, I, C, L, S, T, G, K, M, D and F.

In certain embodiments, a HNH-like domain differs from a sequence of SEQ ID NO:

17 by at least one but not more than, 2, 3, 4, or 5 residues.

25 In certain embodiments, the HNH-like domain is cleavage competent.

In other embodiments, the HNH-like domain is cleavage incompetent.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain comprising an amino acid sequence of Formula X:

X₁-X₂-X₃-H-X₄-X₅-P-X₆-S-X₈-X₉-X₁₀-D-D-S-X₁₄-X₁₅-N-K-V-L-X₁₉-X₂₀-X₂₁-X₂₂-

30 X₂₃-N (SEQ ID NO: 18),

wherein

X₁ is selected from D and E;

X₂ is selected from L, I, R, Q, V, M and K;

X₃ is selected from D and E;

X₄ is selected from I, V, T, A and L (*e.g.*, A, I and V);

X₅ is selected from V, Y, I, L, F and W (*e.g.*, V, I and L);

X₆ is selected from Q, H, R, K, Y, I, L, F and W;

X₈ is selected from F, L, V, K, Y, M, I, R, A, E, D and Q (*e.g.*, F);

5 X₉ is selected from L, R, T, I, V, S, C, Y, K, F and G;

X₁₀ is selected from K, Q, Y, T, F, L, W, M, A, E, G, and S;

X₁₄ is selected from I, L, F, S, R, Y, Q, W, D, K and H (*e.g.*, I, L and F);

X₁₅ is selected from D, S, I, N, E, A, H, F, L, Q, M, G, Y and V;

X₁₉ is selected from T, V, C, E, S and A (*e.g.*, T and V);

10 X₂₀ is selected from R, F, T, W, E, L, N, C, K, V, S, Q, I, Y, H and A;

X₂₁ is selected from S, P, R, K, N, A, H, Q, G and L;

X₂₂ is selected from D, G, T, N, S, K, A, I, E, L, Q, R and Y; and

X₂₃ is selected from K, V, A, E, Y, I, C, L, S, T, G, K, M, D and F.

In certain embodiments, the HNH-like domain differs from a sequence of SEQ ID

15 NO: 18 by 1, 2, 3, 4, or 5 residues.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain comprising an amino acid sequence of Formula XI:

X₁-V-X₃-H-I-V-P-X₆-S-X₈-X₉-X₁₀-D-D-S-X₁₄-X₁₅-N-K-V-L-T-X₂₀-X₂₁-X₂₂-X₂₃-N

(SEQ ID NO:19),

20 wherein

X₁ is selected from D and E;

X₃ is selected from D and E;

X₆ is selected from Q, H, R, K, Y, I, L and W;

X₈ is selected from F, L, V, K, Y, M, I, R, A, E, D and Q (*e.g.*, F);

25 X₉ is selected from L, R, T, I, V, S, C, Y, K, F and G;

X₁₀ is selected from K, Q, Y, T, F, L, W, M, A, E, G, and S;

X₁₄ is selected from I, L, F, S, R, Y, Q, W, D, K and H (*e.g.*, I, L and F);

X₁₅ is selected from D, S, I, N, E, A, H, F, L, Q, M, G, Y and V;

X₂₀ is selected from R, F, T, W, E, L, N, C, K, V, S, Q, I, Y, H and A;

30 X₂₁ is selected from S, P, R, K, N, A, H, Q, G and L;

X₂₂ is selected from D, G, T, N, S, K, A, I, E, L, Q, R and Y; and

X₂₃ is selected from K, V, A, E, Y, I, C, L, S, T, G, K, M, D and F.

In certain embodiments, the HNH-like domain differs from a sequence of SEQ ID NO: 19 by 1, 2, 3, 4, or 5 residues.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an HNH-like domain having an amino acid sequence of Formula XII:

D-X₂-D-H-I-X₅-P-Q-X₇-F-X₉-X₁₀-D-X₁₂-S-I-D-N-X₁₆-V-L-X₁₉-X₂₀-S-X₂₂-X₂₃-N (SEQ ID NO:20),

5 wherein

X₂ is selected from I and V;

X₅ is selected from I and V;

X₇ is selected from A and S;

X₉ is selected from I and L;

10 X₁₀ is selected from K and T;

X₁₂ is selected from D and N;

X₁₆ is selected from R, K and L;

X₁₉ is selected from T and V;

X₂₀ is selected from S and R;

15 X₂₂ is selected from K, D and A; and

X₂₃ is selected from E, K, G and N (*e.g.*, the Cas9 molecule or Cas9 polypeptide can comprise an HNH-like domain as described herein).

In certain embodiments, the HNH-like domain differs from a sequence of SEQ ID NO: 20 by as many as 1 but not more than 2, 3, 4, or 5 residues.

20 In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises the amino acid sequence of formula XIII:

L-Y-Y-L-Q-N-G-X₁'-D-M-Y-X₂'-X₃'-X₄'-X₅'-L-D-I-X₆'-X₇'-L-S-X₈'-Y-Z-N-R-X₉'-K-X₁₀'-D-X₁₁'-V-P (SEQ ID NO: 21),

wherein

25 X₁' is selected from K and R;

X₂' is selected from V and T;

X₃' is selected from G and D;

X₄' is selected from E, Q and D;

X₅' is selected from E and D;

30 X₆' is selected from D, N and H;

X₇' is selected from Y, R and N;

X₈' is selected from Q, D and N;

X₉' is selected from G and E;

X₁₀' is selected from S and G;

X₁₁' is selected from D and N; and

Z is an HNH-like domain, *e.g.*, as described above.

In certain embodiment, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence that differs from a sequence of SEQ ID NO:21 by as many as 1 but not more than 2, 3, 4, or 5 residues.

In certain embodiments, the HNH-like domain differs from a sequence of an HNH-like domain disclosed herein by as many as 1 but not more than 2, 3, 4, or 5 residues. In certain embodiments, 1 or both of the highly conserved residues of are present.

In certain embodiments, the HNH-like domain differs from a sequence of an HNH-like domain disclosed herein by as many as 1 but not more than 2, 3, 4, or 5 residues. In certain embodiments, 1, 2, all 3 of the highly conserved residues are present.

Inducible Cas9 Molecules and Gene Editing Systems

In some embodiments, the Cas9 fusion molecule comprises an inducible Cas9 molecule, as described in more detail in WO15/089427 and WO14/018423, the entire contents of each of which are expressly incorporated herein by reference. Inducible Cas9 molecules are summarized briefly, below.

In one aspect, disclosed herein is a non-naturally occurring or engineered gene editing system, comprising a Cas9 molecule, which may comprise at least one switch, wherein the activity of said gene editing system is controlled by contact with at least one inducer energy source as to the switch. In an embodiment, the control as to the at least one switch or the activity of the gene editing system may be activated, enhanced, terminated or repressed. The contact with the at least one inducer energy source may result in a first effect and a second effect. The first effect may be one or more of nuclear import, nuclear export, recruitment of a secondary component (such as an effector molecule), conformational change (of protein, DNA or RNA), cleavage, release of cargo (such as a caged molecule or a co-factor), association or dissociation. The second effect may be one or more of activation, enhancement, termination or repression of the control as to the at least one switch or the activity of the gene editing system. In one embodiment, the first effect and the second effect may occur in a cascade.

In one embodiment, the Cas9 molecule may further comprise at least one nuclear localization signal (NLS), nuclear export signal (NES), functional domain, flexible linker, mutation, deletion, alteration or truncation. The one or more of the NLS, the NES or the functional domain may be conditionally activated or inactivated. In another embodiment, the

mutation may be one or more of a mutation in a transcription factor homology region, a mutation in a DNA binding domain (such as mutating basic residues of a basic helix loop helix), a mutation in an endogenous NLS or a mutation in an endogenous NES. The disclosure comprehends that the inducer energy source may be heat, ultrasound,

5 electromagnetic energy or chemical. In a preferred embodiment of the invention, the inducer energy source may be an antibiotic, a small molecule, a hormone, a hormone derivative, a steroid or a steroid derivative. In a more preferred embodiment, the inducer energy source maybe abscisic acid (ABA), doxycycline (DOX), cumate, rapamycin, 4-hydroxytamoxifen (4OHT), estrogen or ecdysone. The disclosure also provides that the at least one switch may
10 be selected from the group consisting of antibiotic based inducible systems, electromagnetic energy based inducible systems, small molecule based inducible systems, nuclear receptor based inducible systems and hormone based inducible systems. In a more preferred embodiment, the at least one switch may be selected from the group consisting of tetracycline (Tet)/DOX inducible systems, light inducible systems, ABA inducible systems, cumate
15 repressor/operator systems, 4OHT/estrogen inducible systems, ecdysone-based inducible systems and FKBP12/FRAP (FKBP12-rapamycin complex) inducible systems.

The at least one functional domain may be selected from the group consisting of: transposase domain, integrase domain, recombinase domain, resolvase domain, invertase domain, protease domain, DNA methyltransferase domain, DNA hydroxymethylase domain,
20 DNA demethylase domain, histone acetylase domain, histone deacetylases domain, nuclease domain, repressor domain, activator domain, nuclear-localization signal domains, transcription-regulatory protein (or transcription complex recruiting) domain, cellular uptake activity associated domain, nucleic acid binding domain, antibody presentation domain, histone modifying enzymes, recruiter of histone modifying enzymes; inhibitor of histone
25 modifying enzymes, histone methyltransferase, histone demethylase, histone kinase, histone phosphatase, histone ribosylase, histone deribosylase, histone ubiquitinase, histone deubiquitinase, histone biotinase or histone tail protease.

Specifically, the disclosure provides for systems or methods as described herein, wherein the gene editing system may comprise a vector system comprising: a) a first
30 regulatory element operably linked to a gene editing system guide RNA that targets a locus of interest, b) a second regulatory inducible element operably linked to a Cas9 fusion protein, wherein components (a) and (b) may be located on same or different vectors of the system, wherein the guide RNA targets DNA of the locus of interest, wherein the Cas9 fusion protein and the guide RNA do not naturally occur together. In a preferred embodiment of the

invention, the Cas9 fusion protein comprises an inducible Cas9 enzyme. The invention also provides for the vector being a AAV or a lentivirus.

Split Cas9 Molecules and Gene Editing Systems

5 In some embodiments, the Cas9 fusion molecule comprises a split Cas9 molecule, as described in more detail in WO15/089427 and WO14/018423, the entire contents of each of which are expressly incorporated herein by reference. Split Cas9 molecules are summarized briefly, below.

10 In an aspect, disclosed herein is a non-naturally occurring or engineered inducible CRISPR enzyme, *e.g.*, Cas9 enzyme, comprising: a first CRISPR enzyme fusion construct attached to a first half of an inducible dimer and a second CRISPR enzyme fusion construct attached to a second half of the inducible dimer, wherein the first CRISPR enzyme fusion construct is operably linked to one or more nuclear localization signals, wherein the second CRISPR enzyme fusion construct is operably linked to one or more nuclear export signals,
15 wherein contact with an inducer energy source brings the first and second halves of the inducible dimer together, wherein bringing the first and second halves of the inducible dimer together allows the first and second CRISPR enzyme fusion constructs to constitute a functional gene editing system.

20 In another aspect, in the inducible gene editing system, the inducible dimer is or comprises or consists essentially of or consists of an inducible heterodimer. In an aspect, in inducible gene editing system, the first half or a first portion or a first fragment of the inducible heterodimer is or comprises or consists of or consists essentially of an FKBP, optionally FKBP 12. In an aspect, in the inducible gene editing system, the second half or a second portion or a second fragment of the inducible heterodimer is or comprises or consists
25 of or consists essentially of FRB. In one aspect, in the inducible gene editing system, the arrangement of the first CRISPR enzyme fusion construct is or comprises or consists of or consists essentially of N' terminal Cas9 part- FRB - NES. In another aspect, in the inducible gene editing system, the arrangement of the first CRISPR enzyme fusion construct is or comprises or consists of or consists essentially of NES-N' terminal Cas9 part- FRB - NES. In
30 one aspect in the inducible gene editing system, the arrangement of the second CRISPR enzyme fusion construct is or comprises or consists essentially of or consists of C terminal Cas9 part-FKBP-NLS. In another aspect, in the inducible gene editing system, the arrangement of the second CRISPR enzyme fusion construct is or comprises or consists of or consists essentially of NLS-C terminal Cas9 part-FKBP-NLS. In an aspect, in inducible gene

editing system there can be a linker that separates the Cas9 part from the half or portion or fragment of the inducible dimer. In an aspect, in the inducible gene editing system, the inducer energy source is or comprises or consists essentially of or consists of rapamycin. In an aspect, in inducible gene editing system, the inducible dimer is an inducible homodimer.

5 In an aspect, in inducible gene editing system, the CRISPR enzyme is Cas9, *e.g.*, SpCas9 or SaCas9. In an aspect in an gene editing system, the Cas9 is split into two parts at any one of the following split points, according or with reference to SpCas9: a split position between 202A/203S; a split position between 255F/256D; a split position between 310E/311I; a split position between 534R/535 ; a split position between 572E/573C; a split position between
10 713S/714G; a split position between 1003L/104E; a split position between 1 G54G/1 Q55E; a split position between 11 14N/1115S; a split position between 1152K/1153S; a split position between 1245K/1246G; or a split between 1098 and 1099. In an aspect, in the inducible gene editing system, one or more functional domains are associated with one or both parts of the Cas9 enzyme, *e.g.*, the functional domains optionally including a transcriptional activator, a
15 transcriptional or a nuclease such as a *fok* I nuclease. In an aspect, in the inducible gene editing system, the functional gene editing system binds to the target sequence and the enzyme is a deadCas9, optionally having a diminished nuclease activity of at least 97%, or 100% (or no more than 3% and advantageously 0%) nuclease activity) as compared with the CRISPR enzyme not having the at least one mutation. In an aspect, in the inducible gene
20 editing system, the deadCas9 (CRISPR enzyme) comprises two or more mutations wherein two or more of DIG, E762, H840, N854, N863, or D986 according to SpCas9 protein or any corresponding ortholog or N580 according to SaCas9 protein are mutated, or the CRISPR enzyme comprises at least one mutation, *e.g.*, wherein at least H840 is mutated. The disclosure further provides, a polynucleotide encoding the inducible gene editing system as
25 herein discussed.

Also disclosed herein is a vector for delivery of the first CRISPR enzyme fusion construct, attached to a first half or portion or fragment of an inducible dimer and operably linked to one or more nuclear localization signals, according as herein discussed. In an aspect, disclosed herein is a vector for delivery of the second CRISPR enzyme fusion
30 construct, attached to a second half or portion or fragment of an inducible dimer and operably linked to one or more nuclear export signals.

Cas9 Activities

In certain embodiments, the Cas9 molecule or Cas9 polypeptide is capable of cleaving a target nucleic acid molecule. Typically wild-type Cas9 molecules cleave both strands of a target nucleic acid molecule. Cas9 molecules and Cas9 polypeptides can be engineered to
 5 alter nuclease cleavage (or other properties), *e.g.*, to provide a Cas9 molecule or Cas9 polypeptide which is a nickase, or which lacks the ability to cleave target nucleic acid. A Cas9 molecule or Cas9 polypeptide that is capable of cleaving a target nucleic acid molecule is referred to herein as an eaCas9 (an enzymatically active Cas9) molecule or eaCas9 polypeptide.

10 In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide comprises one or more of the following enzymatic activities:

a nickase activity, *i.e.*, the ability to cleave a single strand, *e.g.*, the non-complementary strand or the complementary strand, of a nucleic acid molecule;

a double stranded nuclease activity, *i.e.*, the ability to cleave both strands of a double
 15 stranded nucleic acid and create a double stranded break, which in one embodiment is the presence of two nickase activities;

an endonuclease activity;

an exonuclease activity; and

a helicase activity, *i.e.*, the ability to unwind the helical structure of a double stranded
 20 nucleic acid.

In certain embodiments, an enzymatically active Cas9 or eaCas9 molecule or eaCas9 polypeptide cleaves both DNA strands and results in a double stranded break. In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide cleaves only one strand, *e.g.*, the strand to which the gRNA hybridizes to, or the strand complementary to the strand the gRNA
 25 hybridizes with. In one embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with an HNH domain. In one embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with a RuvC domain. In one embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with an HNH domain and cleavage activity associated with a RuvC domain. In
 30 one embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an active, or cleavage competent, HNH domain and an inactive, or cleavage incompetent, RuvC domain. In one embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an inactive, or cleavage incompetent, HNH domain and an active, or cleavage competent, RuvC domain.

Some Cas9 molecules or Cas9 polypeptides have the ability to interact with a gRNA molecule, and in conjunction with the gRNA molecule localize to a core target domain, but are incapable of cleaving the target nucleic acid, or incapable of cleaving at efficient rates. Cas9 molecules having no, or no substantial, cleavage activity are referred to herein as an eiCas9 molecule or eiCas9 polypeptide. For example, an eiCas9 molecule or eiCas9 polypeptide can lack cleavage activity or have substantially less, *e.g.*, less than 20, 10, 5, 1 or 0.1 % of the cleavage activity of a reference Cas9 molecule or eiCas9 polypeptide, as measured by an assay described herein.

Targeting and PAMs

A Cas9 molecule or Cas9 polypeptide that can interact with a gRNA molecule and, in concert with the gRNA molecule, localizes to a site which comprises a target domain, and in certain embodiments, a PAM sequence.

In certain embodiments, the ability of an eaCas9 molecule or eaCas9 polypeptide to interact with and cleave a target nucleic acid is PAM sequence dependent. A PAM sequence is a sequence in the target nucleic acid. In one embodiment, cleavage of the target nucleic acid occurs upstream from the PAM sequence. eaCas9 molecules from different bacterial species can recognize different sequence motifs (*e.g.*, PAM sequences). In one embodiment, an eaCas9 molecule of *S. pyogenes* recognizes the sequence motif NGG and directs cleavage of a target nucleic acid sequence 1 to 10, *e.g.*, 3 to 5, bp upstream from that sequence (see, *e.g.*, Mali 2013). In one embodiment, an eaCas9 molecule of *S. thermophilus* recognizes the sequence motif NGGNG and/or NNAGAAW (W = A or T) and directs cleavage of a target nucleic acid sequence 1 to 10, *e.g.*, 3 to 5, bp upstream from these sequences (see, *e.g.*, Horvath 2010; Deveau 2008). In one embodiment, an eaCas9 molecule of *S. mutans* recognizes the sequence motif NGG and/or NAAR (R = A or G) and directs cleavage of a target nucleic acid sequence 1 to 10, *e.g.*, 3 to 5, bp upstream from this sequence (see, *e.g.*, Deveau 2008). In one embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRR (R = A or G) and directs cleavage of a target nucleic acid sequence 1 to 10, *e.g.*, 3 to 5, bp upstream from that sequence. In one embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRRN (R = A or G) and directs cleavage of a target nucleic acid sequence 1 to 10, *e.g.*, 3 to 5, bp upstream from that sequence. In one embodiment, an eaCas9 molecule of *S. aureus* recognizes the sequence motif NNGRRT (R = A or G) and directs cleavage of a target nucleic acid sequence 1 to 10, *e.g.*, 3 to 5, base pairs upstream from that sequence. In one embodiment, an eaCas9 molecule

of *S. aureus* recognizes the sequence motif NNGRRV (R = A or G) and directs cleavage of a target nucleic acid sequence 1 to 10, *e.g.*, 3 to 5, bp upstream from that sequence. The ability of a Cas9 molecule to recognize a PAM sequence can be determined, *e.g.*, using a transformation assay as described in Jinek 2012. In the aforementioned embodiments, N can be any nucleotide residue, *e.g.*, any of A, G, C, or T.

As is discussed herein, Cas9 molecules can be engineered to alter the PAM specificity of the Cas9 molecule.

Exemplary naturally occurring Cas9 molecules have been described previously (see, *e.g.*, Chylinski 2013). Such Cas9 molecules include Cas9 molecules of a cluster 1 bacterial family, cluster 2 bacterial family, cluster 3 bacterial family, cluster 4 bacterial family, cluster 5 bacterial family, cluster 6 bacterial family, a cluster 7 bacterial family, a cluster 8 bacterial family, a cluster 9 bacterial family, a cluster 10 bacterial family, a cluster 11 bacterial family, a cluster 12 bacterial family, a cluster 13 bacterial family, a cluster 14 bacterial family, a cluster 15 bacterial family, a cluster 16 bacterial family, a cluster 17 bacterial family, a cluster 18 bacterial family, a cluster 19 bacterial family, a cluster 20 bacterial family, a cluster 21 bacterial family, a cluster 22 bacterial family, a cluster 23 bacterial family, a cluster 24 bacterial family, a cluster 25 bacterial family, a cluster 26 bacterial family, a cluster 27 bacterial family, a cluster 28 bacterial family, a cluster 29 bacterial family, a cluster 30 bacterial family, a cluster 31 bacterial family, a cluster 32 bacterial family, a cluster 33 bacterial family, a cluster 34 bacterial family, a cluster 35 bacterial family, a cluster 36 bacterial family, a cluster 37 bacterial family, a cluster 38 bacterial family, a cluster 39 bacterial family, a cluster 40 bacterial family, a cluster 41 bacterial family, a cluster 42 bacterial family, a cluster 43 bacterial family, a cluster 44 bacterial family, a cluster 45 bacterial family, a cluster 46 bacterial family, a cluster 47 bacterial family, a cluster 48 bacterial family, a cluster 49 bacterial family, a cluster 50 bacterial family, a cluster 51 bacterial family, a cluster 52 bacterial family, a cluster 53 bacterial family, a cluster 54 bacterial family, a cluster 55 bacterial family, a cluster 56 bacterial family, a cluster 57 bacterial family, a cluster 58 bacterial family, a cluster 59 bacterial family, a cluster 60 bacterial family, a cluster 61 bacterial family, a cluster 62 bacterial family, a cluster 63 bacterial family, a cluster 64 bacterial family, a cluster 65 bacterial family, a cluster 66 bacterial family, a cluster 67 bacterial family, a cluster 68 bacterial family, a cluster 69 bacterial family, a cluster 70 bacterial family, a cluster 71 bacterial family, a cluster 72 bacterial family, a cluster 73 bacterial family, a cluster 74 bacterial family, a

cluster 75 bacterial family, a cluster 76 bacterial family, a cluster 77 bacterial family, or a cluster 78 bacterial family.

Exemplary naturally occurring Cas9 molecules include a Cas9 molecule of a cluster 1 bacterial family. Examples include a Cas9 molecule of: *S. aureus*, *S. pyogenes* (e.g., strain SF370, MGAS10270, MGAS10750, MGAS2096, MGAS315, MGAS5005, MGAS6180, MGAS9429, NZ131 and SSI-1), *S. thermophilus* (e.g., strain LMD-9), *S. pseudoporcinus* (e.g., strain SPIN 20026), *S. mutans* (e.g., strain UA159, NN2025), *S. macacae* (e.g., strain NCTC11558), *S. gallolyticus* (e.g., strain UCN34, ATCC BAA-2069), *S. equines* (e.g., strain ATCC 9812, MGCS 124), *S. dysdalactiae* (e.g., strain GGS 124), *S. bovis* (e.g., strain ATCC 700338), *S. anginosus* (e.g., strain F0211), *S. agalactiae* (e.g., strain NEM316, A909), *Listeria monocytogenes* (e.g., strain F6854), *Listeria innocua* (*L. innocua*, e.g., strain Clip11262), *Enterococcus italicus* (e.g., strain DSM 15952), or *Enterococcus faecium* (e.g., strain 1,231,408).

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence: having 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% homology with; differs at no more than, 2, 5, 10, 15, 20, 30, or 40% of the amino acid residues when compared with; differs by at least 1, 2, 5, 10 or 20 amino acids, but by no more than 100, 80, 70, 60, 50, 40 or 30 amino acids from; or is identical to any Cas9 molecule sequence described herein, or to a naturally occurring Cas9 molecule sequence, e.g., a Cas9 molecule from a species listed herein (e.g., SEQ ID NO:1-4 or described in Chylinski 2013 or Hou 2013). In one embodiment, the Cas9 molecule or Cas9 polypeptide comprises one or more of the following activities: a nickase activity; a double stranded cleavage activity (e.g., an endonuclease and/or exonuclease activity); a helicase activity; or the ability, together with a gRNA molecule, to localize to a target nucleic acid.

A comparison of the sequence of a number of Cas9 molecules indicate that certain regions are conserved. These are identified below as:

region 1 (residues 1 to 180, or in the case of region 1, residues 120 to 180)

region 2 (residues 360 to 480);

region 3 (residues 660 to 720);

region 4 (residues 817 to 900); and

region 5 (residues 900 to 960).

In one embodiment, a Cas9 molecule or Cas9 polypeptide comprises regions 1-5, together with sufficient additional Cas9 molecule sequence to provide a biologically active molecule, e.g., a Cas9 molecule having at least one activity described herein. In one

embodiment, each of regions 1-5, independently, have 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% homology with the corresponding residues of a Cas9 molecule or Cas9 polypeptide described herein, *e.g.*, a sequence from SEQ ID Nos: 1-4.

In one embodiment, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 1:

having 50%, 60%, 70%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% homology with amino acids 1-180 of the amino acid sequence of Cas9 of *S. pyogenes*;

differs by at least 1, 2, 5, 10 or 20 amino acids but by no more than 90, 80, 70, 60, 50, 40 or 30 amino acids from amino acids 1-180 of the amino acid sequence of Cas9 of *S.*

pyogenes, *S. thermophilus*, *S. mutans*, or *Listeria innocua*; or

is identical to amino acids 1-180 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua*.

In one embodiment, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 1':

having 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% homology with amino acids 120-180 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*;

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20 or 10 amino acids from amino acids 120-180 of the amino acid sequence of Cas9 of *S. pyogenes*, *S.*

thermophilus, *S. mutans*, or *L. innocua* ; or

is identical to amino acids 120-180 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua*.

In one embodiment, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 2:

having 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% homology with amino acids 360-480 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua*;

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20 or 10 amino acids from amino acids 360-480 of the amino acid sequence of Cas9 of *S. pyogenes*, *S.*

thermophilus, *S. mutans*, or *L. innocua*; or

is identical to amino acids 360-480 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua*.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 3:

having 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with amino acids 660-720 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*;

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20 or 10 amino acids from amino acids 660-720 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*; or

is identical to amino acids 660-720 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans* or *L. innocua*.

In one embodiment, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 4:

having 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with amino acids 817-900 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua*;

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20 or 10 amino acids from amino acids 817-900 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua*; or

is identical to amino acids 817-900 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua*.

In one embodiment, a Cas9 molecule or Cas9 polypeptide comprises an amino acid sequence referred to as region 5:

having 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, or 99% homology with amino acids 900-960 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua*;

differs by at least 1, 2, or 5 amino acids but by no more than 35, 30, 25, 20 or 10 amino acids from amino acids 900-960 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua*; or

is identical to amino acids 900-960 of the amino acid sequence of Cas9 of *S. pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua*.

Engineered Or Altered Cas9 Molecules And Cas9 Polypeptides

Cas9 molecules and Cas9 polypeptides described herein can possess any of a number of properties, including: nickase activity, nuclease activity (*e.g.*, endonuclease and/or exonuclease activity); helicase activity; the ability to associate functionally with a gRNA molecule; and the ability to target (or localize to) a site on a nucleic acid (*e.g.*, PAM

recognition and specificity). In certain embodiments, a Cas9 molecule or Cas9 polypeptide can include all or a subset of these properties. In a typical embodiment, a Cas9 molecule or Cas9 polypeptide has the ability to interact with a gRNA molecule and, in concert with the gRNA molecule, localize to a site in a nucleic acid. Other activities, *e.g.*, PAM specificity, cleavage activity, or helicase activity can vary more widely in Cas9 molecules and Cas9 polypeptides.

Cas9 molecules include engineered Cas9 molecules and engineered Cas9 polypeptides (engineered, as used in this context, means merely that the Cas9 molecule or Cas9 polypeptide differs from a reference sequences, and implies no process or origin limitation). An engineered Cas9 molecule or Cas9 polypeptide can comprise altered enzymatic properties, *e.g.*, altered nuclease activity, (as compared with a naturally occurring or other reference Cas9 molecule) or altered helicase activity. As discussed herein, an engineered Cas9 molecule or Cas9 polypeptide can have nickase activity (as opposed to double-strand nuclease activity). In one embodiment an engineered Cas9 molecule or Cas9 polypeptide can have an alteration that alters its size, *e.g.*, a deletion of amino acid sequence that reduces its size, *e.g.*, without significant effect on one or more, or any Cas9 activity. In one embodiment, an engineered Cas9 molecule or Cas9 polypeptide can comprise an alteration that affects PAM recognition. For example, an engineered Cas9 molecule can be altered to recognize a PAM sequence other than that recognized by the endogenous wild-type PI domain. In one embodiment a Cas9 molecule or Cas9 polypeptide can differ in sequence from a naturally occurring Cas9 molecule but not have significant alteration in one or more Cas9 activities.

Cas9 molecules or Cas9 polypeptides with desired properties can be made in a number of ways, *e.g.*, by alteration of a parental, *e.g.*, naturally occurring, Cas9 molecules or Cas9 polypeptides, to provide an altered Cas9 molecule or Cas9 polypeptide having a desired property. For example, one or more mutations or differences relative to a parental Cas9 molecule, *e.g.*, a naturally occurring or engineered Cas9 molecule, can be introduced. Such mutations and differences comprise: substitutions (*e.g.*, conservative substitutions or substitutions of non-essential amino acids); insertions; or deletions. In one embodiment, a Cas9 molecule or Cas9 polypeptide can comprises one or more mutations or differences, *e.g.*, at least 1, 2, 3, 4, 5, 10, 15, 20, 30, 40 or 50 mutations but less than 200, 100, or 80 mutations relative to a reference, *e.g.*, a parental, Cas9 molecule.

In certain embodiments, a mutation or mutations do not have a substantial effect on a Cas9 activity, *e.g.*, a Cas9 activity described herein. In other embodiments, a mutation or mutations have a substantial effect on a Cas9 activity, *e.g.*, a Cas9 activity described herein.

5 Non-Cleaving and Modified-Cleavage Cas9 Molecules and Cas9 Polypeptides

In one embodiment, a Cas9 molecule or Cas9 polypeptide comprises a cleavage property that differs from naturally occurring Cas9 molecules, *e.g.*, that differs from the naturally occurring Cas9 molecule having the closest homology. For example, a Cas9 molecule or Cas9 polypeptide can differ from naturally occurring Cas9 molecules, *e.g.*, a
 10 Cas9 molecule of *S. pyogenes*, as follows: its ability to modulate, *e.g.*, decreased or increased, cleavage of a double stranded nucleic acid (endonuclease and/or exonuclease activity), *e.g.*, as compared to a naturally occurring Cas9 molecule (*e.g.*, a Cas9 molecule of *S. pyogenes*); its ability to modulate, *e.g.*, decreased or increased, cleavage of a single-strand of a nucleic acid, *e.g.*, a non-complementary strand of a nucleic acid molecule or a complementary strand
 15 of a nucleic acid molecule (nickase activity), *e.g.*, as compared to a naturally occurring Cas9 molecule (*e.g.*, a Cas9 molecule of *S. pyogenes*); or the ability to cleave a nucleic acid molecule, *e.g.*, a double stranded or single stranded nucleic acid molecule, can be eliminated.

In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide comprises one or more of the following activities: cleavage activity associated with an N-terminal RuvC-like
 20 domain; cleavage activity associated with an HNH-like domain; cleavage activity associated with an HNH-like domain and cleavage activity associated with an N-terminal RuvC-like domain.

In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide comprises an active, or cleavage competent, HNH-like domain (*e.g.*, an HNH-like domain described
 25 herein) and an inactive, or cleavage incompetent, N-terminal RuvC-like domain. An exemplary inactive, or cleavage incompetent N-terminal RuvC-like domain can have a mutation of an aspartic acid in an N-terminal RuvC-like domain, *e.g.*, an aspartic acid at position 10 of SEQ ID NO:2, *e.g.*, can be substituted with an alanine. In one embodiment, the eaCas9 molecule or eaCas9 polypeptide differs from wild-type in the N-terminal RuvC-
 30 like domain and does not cleave the target nucleic acid, or cleaves with significantly less efficiency, *e.g.*, less than 20, 10, 5, 1 or .1 % of the cleavage activity of a reference Cas9 molecule, *e.g.*, as measured by an assay described herein. The reference Cas9 molecule can be a naturally occurring unmodified Cas9 molecule, *e.g.*, a naturally occurring Cas9 molecule such as a Cas9 molecule of *S. pyogenes*, *S. aureus*, or *S. thermophilus*. In one embodiment,

the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology.

In certain embodiments, an eaCas9 molecule or eaCas9 polypeptide comprises an inactive, or cleavage incompetent, HNH domain and an active, or cleavage competent, N-terminal RuvC-like domain (*e.g.*, a RuvC-like domain described herein). Exemplary inactive, or cleavage incompetent HNH-like domains can have a mutation at one or more of: a histidine in an HNH-like domain, for example, at position 856 of the *S. pyogenes* Cas9 sequence (SEQ ID NO:2), *e.g.*, can be substituted with an alanine; and one or more asparagines in an HNH-like domain, for example, at position 870 and/or 879 of the *S.*
 5 *pyogenes* Cas9 sequence (SEQ ID NO:2) *e.g.*, can be substituted with an alanine. In one embodiment, the eaCas9 differs from wild-type in the HNH-like domain and does not cleave the target nucleic acid, or cleaves with significantly less efficiency, *e.g.*, less than 20, 10, 5, 1 or 0.1% of the cleavage activity of a reference Cas9 molecule, *e.g.*, as measured by an assay described herein. The reference Cas9 molecule can be a naturally occurring unmodified Cas9
 10 molecule, *e.g.*, a naturally occurring Cas9 molecule such as a Cas9 molecule of *S. pyogenes*, *S. aureus*, or *S. thermophilus*. In one embodiment, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology.

In certain embodiments, exemplary Cas9 activities comprise one or more of PAM specificity, cleavage activity, and helicase activity. A mutation(s) can be present, *e.g.*, in: one
 20 or more RuvC domains, *e.g.*, an N-terminal RuvC domain; an HNH domain; a region outside the RuvC domains and the HNH domain. In one embodiment, a mutation(s) is present in a RuvC domain. In one embodiment, a mutation(s) is present in an HNH domain. In one embodiment, mutations are present in both a RuvC domain and an HNH domain.

Exemplary mutations that may be made in the RuvC domain or HNH domain with reference to the *S. pyogenes* Cas9 sequence include: D10A, E762A, H840A, N854A, N863A and/or D986A. Exemplary mutations that may be made in the RuvC domain with reference to the *S. aureus* Cas9 sequence include N580A.

In one embodiment, a Cas9 molecule is an eiCas9 molecule comprising one or more differences in a RuvC domain and/or in an HNH domain as compared to a reference Cas9
 30 molecule, and the eiCas9 molecule does not cleave a nucleic acid, or cleaves with significantly less efficiency than does wild type, *e.g.*, when compared with wild type in a cleavage assay, *e.g.*, as described herein, cuts with less than 50, 25, 10, or 1% of a reference Cas9 molecule, as measured by an assay described herein.

Whether or not a particular sequence, *e.g.*, a substitution, may affect one or more activity, such as targeting activity, cleavage activity, etc., can be evaluated or predicted, *e.g.*, by evaluating whether the mutation is conservative. In one embodiment, a “non-essential” amino acid residue, as used in the context of a Cas9 molecule, is a residue that can be altered
 5 from the wild-type sequence of a Cas9 molecule, *e.g.*, a naturally occurring Cas9 molecule, *e.g.*, an eaCas9 molecule, without abolishing or more preferably, without substantially altering a Cas9 activity (*e.g.*, cleavage activity), whereas changing an “essential” amino acid residue results in a substantial loss of activity (*e.g.*, cleavage activity).

In one embodiment, a Cas9 molecule comprises a cleavage property that differs from
 10 naturally occurring Cas9 molecules, *e.g.*, that differs from the naturally occurring Cas9 molecule having the closest homology. For example, a Cas9 molecule can differ from naturally occurring Cas9 molecules, *e.g.*, a Cas9 molecule of *S. aureus* or *S. pyogenes* as follows: its ability to modulate, *e.g.*, decreased or increased, cleavage of a double stranded break (endonuclease and/or exonuclease activity), *e.g.*, as compared to a naturally occurring
 15 Cas9 molecule (*e.g.*, a Cas9 molecule of *S. aureus* or *S. pyogenes*); its ability to modulate, *e.g.*, decreased or increased, cleavage of a single-strand of a nucleic acid, *e.g.*, a non-complimentary strand of a nucleic acid molecule or a complementary strand of a nucleic acid molecule (nickase activity), *e.g.*, as compared to a naturally occurring Cas9 molecule (*e.g.*, a Cas9 molecule of *S. aureus* or *S. pyogenes*); or the ability to cleave a nucleic acid molecule,
 20 *e.g.*, a double stranded or single stranded nucleic acid molecule, can be eliminated. In certain embodiments, the nickase is *S. aureus* Cas9-derived nickase comprising the sequence of SEQ ID NO: 214 (D10A) or SEQ ID NO: 215 (N580A) (Friedland 2015).

In certain embodiments, the altered Cas9 molecule is an eaCas9 molecule comprising one or more of the following activities: cleavage activity associated with a RuvC domain;
 25 cleavage activity associated with an HNH domain; cleavage activity associated with an HNH domain and cleavage activity associated with a RuvC domain.

In one embodiment, the altered Cas9 molecule is an eiCas9 molecule which does not cleave a nucleic acid molecule (either double stranded or single stranded nucleic acid molecules) or cleaves a nucleic acid molecule with significantly less efficiency, *e.g.*, less than
 30 20, 10, 5, 1 or 0.1% of the cleavage activity of a reference Cas9 molecule, *e.g.*, as measured by an assay described herein. The reference Cas9 molecule can be a naturally occurring unmodified Cas9 molecule, *e.g.*, a naturally occurring Cas9 molecule such as a Cas9 molecule of *S. pyogenes*, *S. thermophilus*, *S. aureus*, *C. jejuni* or *N. meningitidis*. In one embodiment, the reference Cas9 molecule is the naturally occurring Cas9 molecule having

the closest sequence identity or homology. In one embodiment, the eiCas9 molecule lacks substantial cleavage activity associated with a RuvC domain and cleavage activity associated with an HNH domain.

In certain embodiments, the altered Cas9 molecule or Cas9 polypeptide, *e.g.*, an eaCas9 molecule or eaCas9 polypeptide, can be a fusion, *e.g.*, of two or more different Cas9 molecules, *e.g.*, of two or more naturally occurring Cas9 molecules of different species. For example, a fragment of a naturally occurring Cas9 molecule of one species can be fused to a fragment of a Cas9 molecule of a second species. As an example, a fragment of a Cas9 molecule of *S. pyogenes* comprising an N-terminal RuvC-like domain can be fused to a fragment of Cas9 molecule of a species other than *S. pyogenes* (*e.g.*, *S. thermophilus*) comprising an HNH-like domain.

Cas9 with Altered or No PAM Recognition

Naturally-occurring Cas9 molecules can recognize specific PAM sequences, for example the PAM recognition sequences described above for, *e.g.*, *S. pyogenes*, *S. thermophilus*, *S. mutans*, and *S. aureus*.

In certain embodiments, a Cas9 molecule or Cas9 polypeptide has the same PAM specificities as a naturally occurring Cas9 molecule. In other embodiments, a Cas9 molecule or Cas9 polypeptide has a PAM specificity not associated with a naturally occurring Cas9 molecule, or a PAM specificity not associated with the naturally occurring Cas9 molecule to which it has the closest sequence homology. For example, a naturally occurring Cas9 molecule can be altered, *e.g.*, to alter PAM recognition, *e.g.*, to alter the PAM sequence that the Cas9 molecule or Cas9 polypeptide recognizes in order to decrease off-target sites and/or improve specificity; or eliminate a PAM recognition requirement. In certain embodiments, a Cas9 molecule or Cas9 polypeptide can be altered, *e.g.*, to increase length of PAM recognition sequence and/or improve Cas9 specificity to high level of identity (*e.g.*, 98%, 99% or 100% match between gRNA and a PAM sequence), *e.g.*, to decrease off-target sites and/or increase specificity. In certain embodiments, the length of the PAM recognition sequence is at least 4, 5, 6, 7, 8, 9, 10 or 15 amino acids in length. In one embodiment, the Cas9 specificity requires at least 90%, 95%, 96%, 97%, 98%, 99% or more homology between the gRNA and the PAM sequence. Cas9 molecules or Cas9 polypeptides that recognize different PAM sequences and/or have reduced off-target activity can be generated using directed evolution. Exemplary methods and systems that can be used for directed

evolution of Cas9 molecules are described (see, *e.g.*, Esvelt 2011). Candidate Cas9 molecules can be evaluated, *e.g.*, by methods described below.

Size-Optimized Cas9 Molecules

5 Engineered Cas9 molecules and engineered Cas9 polypeptides described herein include a Cas9 molecule or Cas9 polypeptide comprising a deletion that reduces the size of the molecule while still retaining desired Cas9 properties, *e.g.*, essentially native conformation, Cas9 nuclease activity, and/or target nucleic acid molecule recognition. Provided herein are Cas9 molecules or Cas9 polypeptides comprising one or more deletions
10 and optionally one or more linkers, wherein a linker is disposed between the amino acid residues that flank the deletion. Methods for identifying suitable deletions in a reference Cas9 molecule, methods for generating Cas9 molecules with a deletion and a linker, and methods for using such Cas9 molecules will be apparent to one of ordinary skill in the art upon review of this document.

15 A Cas9 molecule, *e.g.*, a *S. aureus* or *S. pyogenes* Cas9 molecule, having a deletion is smaller, *e.g.*, has reduced number of amino acids, than the corresponding naturally-occurring Cas9 molecule. The smaller size of the Cas9 molecules allows increased flexibility for delivery methods, and thereby increases utility for genome-editing. A Cas9 molecule can comprise one or more deletions that do not substantially affect or decrease the activity of the
20 resultant Cas9 molecules described herein. Activities that are retained in the Cas9 molecules comprising a deletion as described herein include one or more of the following:

 a nickase activity, *i.e.*, the ability to cleave a single strand, *e.g.*, the non-complementary strand or the complementary strand, of a nucleic acid molecule; a double stranded nuclease activity, *i.e.*, the ability to cleave both strands of a double stranded nucleic
25 acid and create a double stranded break, which in one embodiment is the presence of two nickase activities; an endonuclease activity; an exonuclease activity; a helicase activity, *i.e.*, the ability to unwind the helical structure of a double stranded nucleic acid; and recognition activity of a nucleic acid molecule, *e.g.*, a target nucleic acid or a gRNA molecule.

 Activity of the Cas9 molecules described herein can be assessed using the activity
30 assays described herein or in the art.

Identifying regions suitable for deletion

Suitable regions of Cas9 molecules for deletion can be identified by a variety of methods. Naturally-occurring orthologous Cas9 molecules from various bacterial species can

be modeled onto the crystal structure of *S. pyogenes* Cas9 (Nishimasu 2014) to examine the level of conservation across the selected Cas9 orthologs with respect to the three-dimensional conformation of the protein. Less conserved or unconserved regions that are spatially located distant from regions involved in Cas9 activity, *e.g.*, interface with the target nucleic acid molecule and/or gRNA, represent regions or domains are candidates for deletion without substantially affecting or decreasing Cas9 activity.

Nucleic Acids Encoding Cas9 Molecules

Nucleic acids encoding the Cas9 molecules or Cas9 polypeptides, *e.g.*, an eaCas9 molecule or eaCas9 polypeptides are provided herein. Exemplary nucleic acids encoding Cas9 molecules or Cas9 polypeptides have been described previously (see, *e.g.*, Cong 2013; Wang 2013; Mali 2013; Jinek 2012).

In one embodiment, a nucleic acid encoding a Cas9 molecule or Cas9 polypeptide can be a synthetic nucleic acid sequence. For example, the synthetic nucleic acid molecule can be chemically modified, *e.g.*, as described herein. In one embodiment, the Cas9 mRNA has one or more (*e.g.*, all of the following properties: it is capped, polyadenylated, substituted with 5-methylcytidine and/or pseudouridine).

In addition, or alternatively, the synthetic nucleic acid sequence can be codon optimized, *e.g.*, at least one non-common codon or less-common codon has been replaced by a common codon. For example, the synthetic nucleic acid can direct the synthesis of an optimized messenger mRNA, *e.g.*, optimized for expression in a mammalian expression system, *e.g.*, described herein.

In addition, or alternatively, a nucleic acid encoding a Cas9 molecule or Cas9 polypeptide may comprise a nuclear localization sequence (NLS). Nuclear localization sequences are known in the art.

An exemplary codon optimized nucleic acid sequence encoding a Cas9 molecule of *S. pyogenes* is set forth in SEQ ID NO: 22. The corresponding amino acid sequence of an *S. pyogenes* Cas9 molecule is set forth in SEQ ID NO: 23.

Exemplary codon optimized nucleic acid sequence encoding a Cas9 molecule of *S. aureus* is set forth in SEQ ID NO: 26, 39, 213 and 214.

If any of the above Cas9 sequences are fused with a peptide or polypeptide at the C-terminus, it is understood that the stop codon will be removed.

Other Cas Molecules and Cas Polypeptides

Various types of Cas molecules or Cas polypeptides can be used to practice the inventions disclosed herein. In some embodiments, Cas molecules of Type II Cas systems are used. In other embodiments, Cas molecules of other Cas systems are used. For example, Type I or Type III Cas molecules may be used. Exemplary Cas molecules (and Cas systems) have been described previously (see, *e.g.*, Haft 2005; Makarova 2011). Exemplary Cas molecules (and Cas systems) are also shown in **Table 4**.

Table 4: Cas Systems					
Gene name[‡]	System type or subtype	Name from Haft 2005[§]	Structure of encoded protein (PDB accessions)[¶]	Families (and superfamily) of encoded protein^{***}	Representatives
<i>cas1</i>	• Type I • Type II • Type III	<i>cas1</i>	3GOD, 3LFX and 2YZS	COG1518	SERP2463, SPy1047 and <i>ygbT</i>
<i>cas2</i>	• Type I • Type II • Type III	<i>cas2</i>	2IVY, 2I8E and 3EXC	COG1343 and COG3512	SERP2462, SPy1048, SPy1723 (N-terminal domain) and <i>ygbF</i>
<i>cas3'</i>	• Type I ^{‡‡}	<i>cas3</i>	NA	COG1203	APE1232 and <i>ygcB</i>
<i>cas3''</i>	• Subtype I-A • Subtype I-B	NA	NA	COG2254	APE1231 and BH0336
<i>cas4</i>	• Subtype I-A • Subtype I-B • Subtype I-C • Subtype I-D • Subtype II-B	<i>cas4</i> and <i>csa1</i>	NA	COG1468	APE1239 and BH0340
<i>cas5</i>	• Subtype I-A • Subtype I-B • Subtype I-C • Subtype I-E	<i>cas5a</i> , <i>cas5d</i> , <i>cas5e</i> , <i>cas5h</i> , <i>cas5p</i> , <i>cas5t</i> and <i>cmx5</i>	3KG4	COG1688 (RAMP)	APE1234, BH0337, <i>devS</i> and <i>ygcI</i>
<i>cas6</i>	• Subtype I-A • Subtype I-B • Subtype I-D • Subtype III-A • Subtype III-B	<i>cas6</i> and <i>cmx6</i>	3I4H	COG1583 and COG5551 (RAMP)	PF1131 and slr7014
<i>cas6e</i>	• Subtype I-E	<i>cse3</i>	1WJ9	(RAMP)	<i>ygcH</i>
<i>cas6f</i>	• Subtype I-F	<i>csy4</i>	2XLJ	(RAMP)	y1727
<i>cas7</i>	• Subtype I-A • Subtype I-B • Subtype I-C • Subtype I-E	<i>csa2</i> , <i>csd2</i> , <i>cse4</i> , <i>csh2</i> , <i>csp1</i> and <i>cst2</i>	NA	COG1857 and COG3649 (RAMP)	<i>devR</i> and <i>ygcJ</i>
<i>cas8a1</i>	• Subtype I-A ^{‡‡}	<i>cmx1</i> , <i>cst1</i> , <i>csx8</i> , <i>csx13</i> and CXXC-CXXC	NA	BH0338-like	LA3191 ^{§§} and PG2018 ^{§§}

Table 4: Cas Systems

Gene name [‡]	System type or subtype	Name from Haft 2005 [§]	Structure of encoded protein (PDB accessions) [¶]	Families (and superfamily) of encoded protein ^{***}	Representatives
<i>cas8a2</i>	• Subtype I-A ^{††}	<i>csa4</i> and <i>csx9</i>	NA	PH0918	AF0070, AF1873, MJ0385, PF0637, PH0918 and SSO1401
<i>cas8b</i>	• Subtype I-B ^{††}	<i>csh1</i> and TM1802	NA	BH0338-like	MTH1090 and TM1802
<i>cas8c</i>	• Subtype I-C ^{††}	<i>csd1</i> and <i>csp2</i>	NA	BH0338-like	BH0338
<i>cas9</i>	• Type II ^{††}	<i>csn1</i> and <i>csx12</i>	NA	COG3513	FTN_0757 and SPy1046
<i>cas10</i>	• Type III ^{††}	<i>cmr2</i> , <i>csm1</i> and <i>csx11</i>	NA	COG1353	MTH326, Rv2823c ^{§§} and TM1794 ^{§§}
<i>cas10d</i>	• Subtype I-D ^{††}	<i>csc3</i>	NA	COG1353	slr7011
<i>csy1</i>	• Subtype I-F ^{††}	<i>csy1</i>	NA	y1724-like	y1724
<i>csy2</i>	• Subtype I-F	<i>csy2</i>	NA	(RAMP)	y1725
<i>csy3</i>	• Subtype I-F	<i>csy3</i>	NA	(RAMP)	y1726
<i>cse1</i>	• Subtype I-E ^{††}	<i>cse1</i>	NA	YgcL-like	<i>ygcL</i>
<i>cse2</i>	• Subtype I-E	<i>cse2</i>	2ZCA	YgcK-like	<i>ygcK</i>
<i>csc1</i>	• Subtype I-D	<i>csc1</i>	NA	alr1563-like (RAMP)	alr1563
<i>csc2</i>	• Subtype I-D	<i>csc1</i> and <i>csc2</i>	NA	COG1337 (RAMP)	slr7012
<i>csa5</i>	• Subtype I-A	<i>csa5</i>	NA	AF1870	AF1870, MJ0380, PF0643 and SSO1398
<i>csn2</i>	• Subtype II-A	<i>csn2</i>	NA	SPy1049-like	SPy1049
<i>csm2</i>	• Subtype III-A ^{††}	<i>csm2</i>	NA	COG1421	MTH1081 and SERP2460
<i>csm3</i>	• Subtype III-A	<i>csc2</i> and <i>csm3</i>	NA	COG1337 (RAMP)	MTH1080 and SERP2459
<i>csm4</i>	• Subtype III-A	<i>csm4</i>	NA	COG1567 (RAMP)	MTH1079 and SERP2458
<i>csm5</i>	• Subtype III-A	<i>csm5</i>	NA	COG1332 (RAMP)	MTH1078 and SERP2457
<i>csm6</i>	• Subtype III-A	APE2256 and <i>csm6</i>	2WTE	COG1517	APE2256 and SSO1445
<i>cmr1</i>	• Subtype III-B	<i>cmr1</i>	NA	COG1367 (RAMP)	PF1130
<i>cmr3</i>	• Subtype III-	<i>cmr3</i>	NA	COG1769	PF1128

Table 4: Cas Systems

Gene name [‡]	System type or subtype	Name from Haft 2005 [§]	Structure of encoded protein (PDB accessions) [¶]	Families (and superfamily) of encoded protein ^{***}	Representatives
	B			(RAMP)	
<i>cmr4</i>	• Subtype III-B	<i>cmr4</i>	NA	COG1336 (RAMP)	PF1126
<i>cmr5</i>	• Subtype III-B ^{††}	<i>cmr5</i>	2ZOP and 2OEB	COG3337	MTH324 and PF1125
<i>cmr6</i>	• Subtype III-B	<i>cmr6</i>	NA	COG1604 (RAMP)	PF1124
<i>csb1</i>	• Subtype I-U	GSU0053	NA	(RAMP)	Balac_1306 and GSU0053
<i>csb2</i>	• Subtype I-U ^{§§}	NA	NA	(RAMP)	Balac_1305 and GSU0054
<i>csb3</i>	• Subtype I-U	NA	NA	(RAMP)	Balac_1303 ^{§§}
<i>csx17</i>	• Subtype I-U	NA	NA	NA	Btus_2683
<i>csx14</i>	• Subtype I-U	NA	NA	NA	GSU0052
<i>csx10</i>	• Subtype I-U	<i>csx10</i>	NA	(RAMP)	Caur_2274
<i>csx16</i>	• Subtype III-U	VVA1548	NA	NA	VVA1548
<i>csaX</i>	• Subtype III-U	<i>csaX</i>	NA	NA	SSO1438
<i>csx3</i>	• Subtype III-U	<i>csx3</i>	NA	NA	AF1864
<i>csx1</i>	• Subtype III-U	<i>csa3</i> , <i>csx1</i> , <i>csx2</i> , DXTHG, NE0113 and TIGR02710	1XMX and 2I71	COG1517 and COG4006	MJ1666, NE0113, PF1127 and TM1812
<i>csx15</i>	• Unknown	NA	NA	TTE2665	TTE2665
<i>csf1</i>	• Type U	<i>csf1</i>	NA	NA	AFE_1038
<i>csf2</i>	• Type U	<i>csf2</i>	NA	(RAMP)	AFE_1039
<i>csf3</i>	• Type U	<i>csf3</i>	NA	(RAMP)	AFE_1040
<i>csf4</i>	• Type U	<i>csf4</i>	NA	NA	AFE_1037

Functional Analysis of Candidate Molecules

Candidate Cas9 molecules, candidate gRNA molecules, candidate Cas9 molecule/gRNA molecule complexes, can be evaluated by art-known methods or as described herein. For example, exemplary methods for evaluating the endonuclease activity of Cas9 molecule have been described previously (Jinek 2012).

5

Binding and Cleavage Assay: Testing the endonuclease activity of Cas9 molecule

The ability of a Cas9 molecule/gRNA molecule complex to bind to and cleave a target nucleic acid can be evaluated in a plasmid cleavage assay. In this assay, synthetic or *in vitro*-transcribed gRNA molecule is pre-annealed prior to the reaction by heating to 95°C and slowly cooling down to room temperature. Native or restriction digest-linearized plasmid DNA (300 ng (~8 nM)) is incubated for 60 min at 37°C with purified Cas9 protein molecule (50-500 nM) and gRNA (50-500 nM, 1:1) in a Cas9 plasmid cleavage buffer (20 mM HEPES pH 7.5, 150 mM KCl, 0.5 mM DTT, 0.1 mM EDTA) with or without 10 mM MgCl₂. The reactions are stopped with 5X DNA loading buffer (30% glycerol, 1.2% SDS, 250 mM EDTA), resolved by a 0.8 or 1% agarose gel electrophoresis and visualized by ethidium bromide staining. The resulting cleavage products indicate whether the Cas9 molecule cleaves both DNA strands, or only one of the two strands. For example, linear DNA products indicate the cleavage of both DNA strands. Nicked open circular products indicate that only one of the two strands is cleaved.

Alternatively, the ability of a Cas9 molecule/gRNA molecule complex to bind to and cleave a target nucleic acid can be evaluated in an oligonucleotide DNA cleavage assay. In this assay, DNA oligonucleotides (10 pmol) are radiolabeled by incubating with 5 units T4 polynucleotide kinase and ~3–6 pmol (~20–40 mCi) [γ -³²P]-ATP in 1X T4 polynucleotide kinase reaction buffer at 37°C for 30 min, in a 50 μ L reaction. After heat inactivation (65°C for 20 min), reactions are purified through a column to remove unincorporated label. Duplex substrates (100 nM) are generated by annealing labeled oligonucleotides with equimolar amounts of unlabeled complementary oligonucleotide at 95°C for 3 min, followed by slow cooling to room temperature. For cleavage assays, gRNA molecules are annealed by heating to 95°C for 30 s, followed by slow cooling to room temperature. Cas9 (500 nM final concentration) is pre-incubated with the annealed gRNA molecules (500 nM) in cleavage assay buffer (20 mM HEPES pH 7.5, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT, 5% glycerol) in a total volume of 9 μ L. Reactions are initiated by the addition of 1 μ L target DNA (10 nM) and incubated for 1 h at 37°C. Reactions are quenched by the addition of 20 μ L of loading dye (5 mM EDTA, 0.025% SDS, 5% glycerol in formamide) and heated to 95°C for 5 min. Cleavage products are resolved on 12% denaturing polyacrylamide gels containing 7 M urea and visualized by phosphorimaging. The resulting cleavage products indicate that whether the complementary strand, the non-complementary strand, or both, are cleaved.

One or both of these assays can be used to evaluate the suitability of a candidate gRNA molecule or candidate Cas9 molecule.

Binding Assay: Testing the binding of Cas9 molecule to target DNA

Exemplary methods for evaluating the binding of Cas9 molecule to target DNA have been described previously (Jinek 2012).

For example, in an electrophoretic mobility shift assay, target DNA duplexes are formed by mixing of each strand (10 nmol) in deionized water, heating to 95°C for 3 min and slow cooling to room temperature. All DNAs are purified on 8% native gels containing 1X TBE. DNA bands are visualized by UV shadowing, excised, and eluted by soaking gel pieces in DEPC-treated H₂O. Eluted DNA is ethanol precipitated and dissolved in DEPC-treated H₂O. DNA samples are 5' end labeled with [γ -³²P]-ATP using T4 polynucleotide kinase for 30 min at 37°C. Polynucleotide kinase is heat denatured at 65°C for 20 min, and unincorporated radiolabel is removed using a column. Binding assays are performed in buffer containing 20 mM HEPES pH 7.5, 100 mM KCl, 5 mM MgCl₂, 1 mM DTT and 10% glycerol in a total volume of 10 μ L. Cas9 protein molecule is programmed with equimolar amounts of pre-annealed gRNA molecule and titrated from 100 pM to 1 μ M. Radiolabeled DNA is added to a final concentration of 20 pM. Samples are incubated for 1 h at 37°C and resolved at 4°C on an 8% native polyacrylamide gel containing 1X TBE and 5 mM MgCl₂. Gels are dried and DNA visualized by phosphorimaging.

Differential Scanning Fluorimetry (DSF)

The thermostability of Cas9 molecule-gRNA ribonucleoprotein (RNP) complexes can be measured via DSF. This technique measures the thermostability of a protein, which can increase under favorable conditions such as the addition of a binding RNA molecule, *e.g.*, a gRNA.

The assay is performed using two different protocols, one to test the best stoichiometric ratio of gRNA:Cas9 protein and another to determine the best solution conditions for RNP formation.

To determine the best solution to form RNP complexes, a 2 μ M solution of Cas9 in water + 10x SYPRO Orange® (Life Technologies cat#S-6650) and dispensed into a 384 well plate. An equimolar amount of gRNA diluted in solutions with varied pH and salt is then added. After incubating at room temperature for 10 min. and brief centrifugation to remove any bubbles, a Bio-Rad CFX384™ Real-Time System C1000 Touch™ Thermal Cycler with the Bio-Rad CFX Manager software is used to run a gradient from 20°C to 90°C with a 1°C increase in temperature every 10 seconds.

The second assay consists of mixing various concentrations of gRNA with 2 μ M Cas 9 in optimal buffer from the assay above and incubating at RT for 10 min in a 384 well plate.

An equal volume of optimal buffer + 10x SYPRO Orange® (Life Technologies cat#S-6650) is added and the plate sealed with Microseal® B adhesive (MSB-1001). Following brief centrifugation to remove any bubbles, a Bio-Rad CFX384™ Real-Time System C1000 Touch™ Thermal Cycler with the Bio-Rad CFX Manager software is used to run a gradient from 20°C to 90°C with a 1°C increase in temperature every 10 seconds.

Resection Assay: Testing a Cas9 to promote resection

The ability of a Cas9 to promote resection can be evaluated by measuring the levels of single stranded DNA at specific double strand break sites in human cells using quantitative methods (as described in Zhou 2014). In this assay, a cell line is delivered, *e.g.*, by transfection, a candidate Cas9 or a candidate Cas9 fusion protein. The cells are cultured for a sufficient amount of time to allow nuclease activity and resection to occur. Genomic DNA is carefully extracted using a method in which cells are embedded in low-gelling point agar that protects the DNA from shearing and damage during extraction. The genomic DNA is digested with a restriction enzyme that selectively cuts double-stranded DNA. Primers for quantitative PCR that span up to 5 kb of the double strand break site are designed. The results from the PCR reaction show the levels of single strand DNA detected at each of the primer positions. Thus, the length and the level of resection promoted by the candidate Cas9 or Cas9 fusion protein can be determined from this assay.

Other qualitative assays for identifying the occurrence of resection include the detection of proteins or protein complexes that bind to single-stranded DNA after resection has occurred, *e.g.*, RPA foci, Rad51 foci, or BrDU detection by immunofluorescence. Antibodies for RPA protein and Rad51 are known in the art.

Genome Editing Approaches

Mutations in a target gene may be corrected using one of the approaches discussed herein. In one embodiment, a mutation in a target gene is corrected by homology directed repair (HDR) using an exogenously provided template nucleic acid, referred to herein as “gene correction”. In another embodiment, a mutation in a target gene is corrected by homology directed repair without using an exogenously provided template nucleic acid, referred to herein as gene correction.

HDR Repair and Template Nucleic Acids

In certain embodiments of the methods provided herein, HDR-mediated sequence alteration is used to alter and/or correct (*e.g.*, repair or edit) the sequence of one or more

nucleotides in a genome (*e.g.*, a point mutation in a target gene). While not wishing to be bound by theory, it is believed that HDR-mediated alteration of a target sequence within a target gene occurs by HDR with an exogenously provided donor template or template nucleic acid in a process referred to herein as gene correction. For example, the donor template or template nucleic acid provides for alteration of the target sequence. It is contemplated that a plasmid donor can be used as a template for homologous recombination. It is further contemplated that a single stranded donor template can be used as a template for alteration of the target sequence by alternate methods of HDR (*e.g.*, single-strand annealing) between the target sequence and the donor template. Donor template-effected alteration of a target sequence depends on cleavage by a Cas9 molecule. Cleavage by Cas9 can comprise a double-strand break or two single-strand breaks.

In one embodiment, the target position or target position regions has at least 50%, 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% homology with an endogenous homologous sequence.

In one embodiment, the target position region, except for the target position, differs by 1, 2, 3, 4, 5, 10, 25, 50, 100 or fewer, nucleotides with an endogenous homologous sequence.

In one embodiment, the target position region has at least 50%, 60%, 70%, 80%, 90%, 92%, 94%, 96%, 98%, or 99% homology with an endogenous homologous sequence over at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 750, 1,000, 2500, 5000, or 10000 nucleotides.

In one embodiment, the target position region, except for the target position, differs by 1, 2, 3, 4, 5, 10, 25, 50, 100 or fewer, nucleotides with an endogenous homologous sequence over at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 750, 1,000, 2500, 5000, or 10000 nucleotides.

In one embodiment, the endogenous homologous sequence comprises a domain, *e.g.*, a catalytic domain, a domain that binds a target, a structural domain, found in the gene that comprises the target position.

In certain embodiments of the methods provided herein, HDR-mediated alteration is used to alter a single nucleotide in a target sequence. These embodiments may utilize either one double-strand break or two single-strand breaks. In certain embodiments, a single nucleotide alteration is incorporated using (1) one double-strand break, (2) two single-strand breaks, (3) two double-strand breaks with a break occurring on each side of the target position, (4) one double-strand break and two single-strand breaks with the double-strand

break and two single-strand breaks occurring on each side of the target position (5) four single-strand breaks with a pair of single stranded breaks occurring on each side of the target position, or (6) one single-strand break.

5 In certain embodiments wherein a single-stranded template nucleic acid is used, the target position can be altered by alternative HDR.

Donor template-effected alteration of a target position depends on cleavage by a Cas9 molecule. Cleavage by Cas9 can comprise a nick, a double-strand break, or two single-strand breaks, *e.g.*, one on each strand of the target nucleic acid. After introduction of the breaks on the target nucleic acid, resection occurs at the break ends resulting in single stranded
10 overhanging DNA regions.

In canonical HDR, a double-stranded donor template is introduced, comprising homologous sequence to the target nucleic acid that will either be directly incorporated into the target nucleic acid or used as a template to change the sequence of the target nucleic acid. After resection at the break, repair can progress by different pathways, *e.g.*, by the double
15 Holliday junction model (or double-strand break repair, DSB, pathway) or the synthesis-dependent strand annealing (SDSA) pathway. In the double Holliday junction model, strand invasion by the two single stranded overhangs of the target nucleic acid to the homologous sequences in the donor template occurs, resulting in the formation of an intermediate with two Holliday junctions. The junctions migrate as new DNA is synthesized from the ends of
20 the invading strand to fill the gap resulting from the resection. The end of the newly synthesized DNA is ligated to the resected end, and the junctions are resolved, resulting in the alteration of the target nucleic acid, *e.g.*, incorporation of the altered sequence of the donor template at the corresponding target position. Crossover with the donor template may occur upon resolution of the junctions. In the SDSA pathway, only one single stranded
25 overhang invades the donor template and new DNA is synthesized from the end of the invading strand to fill the gap resulting from resection. The newly synthesized DNA then anneals to the remaining single stranded overhang, new DNA is synthesized to fill in the gap, and the strands are ligated to produce the altered DNA duplex.

In alternative HDR, a single-strand donor template, *e.g.*, template nucleic acid, is
30 introduced. A nick, single-strand break, or double-strand break at the target nucleic acid, for altering a desired target position, is mediated by a Cas9 molecule, *e.g.*, described herein, and resection at the break occurs to reveal single stranded overhangs. Incorporation of the sequence of the template nucleic acid to correct or alter the target position of the target nucleic acid typically occurs by the SDSA pathway, as described above.

Additional details on template nucleic acids are provided in Section IV entitled “Template nucleic acids” in International Application PCT/US2014/057905, now published as WO2015/048577, the entire contents of which are expressly incorporated herein by reference.

5 In certain embodiments, double-strand cleavage is effected by a Cas9 molecule having cleavage activity associated with an HNH-like domain and cleavage activity associated with a RuvC-like domain, *e.g.*, an N-terminal RuvC-like domain, *e.g.*, a wild type Cas9. Such embodiments require only a single gRNA molecule.

10 In certain embodiments, one single-strand break, or nick, is effected by a Cas9 molecule having nickase activity, *e.g.*, a Cas9 nickase as described herein. A nicked target nucleic acid can be a substrate for alt-HDR.

15 In other embodiments, two single-strand breaks, or nicks, are effected by a Cas9 molecule having nickase activity, *e.g.*, cleavage activity associated with an HNH-like domain or cleavage activity associated with an N-terminal RuvC-like domain. Such embodiments usually require two gRNAs, one for placement of each single-strand break. In one embodiment, the Cas9 molecule having nickase activity cleaves the strand to which the gRNA hybridizes, but not the strand that is complementary to the strand to which the gRNA hybridizes. In one embodiment, the Cas9 molecule having nickase activity does not cleave the strand to which the gRNA hybridizes, but rather cleaves the strand that is complementary
20 to the strand to which the gRNA hybridizes.

In certain embodiments, the nickase has HNH activity, *e.g.*, a Cas9 molecule having the RuvC activity inactivated, *e.g.*, a Cas9 molecule having a mutation at D10, *e.g.*, the D10A mutation. D10A inactivates RuvC; therefore, the Cas9 nickase has (only) HNH activity and will cut on the strand to which the gRNA hybridizes (*e.g.*, the complementary strand, which
25 does not have the NGG PAM on it). In other embodiments, a Cas9 molecule having an H840, *e.g.*, an H840A, mutation can be used as a nickase. H840A inactivates HNH; therefore, the Cas9 nickase has (only) RuvC activity and cuts on the non-complementary strand (*e.g.*, the strand that has the NGG PAM and whose sequence is identical to the gRNA). In other embodiments, a Cas9 molecule having an N863 mutation, *e.g.*, the N863A mutation,
30 mutation can be used as a nickase. N863A inactivates HNH therefore the Cas9 nickase has (only) RuvC activity and cuts on the non-complementary strand (the strand that has the NGG PAM and whose sequence is identical to the gRNA).

In certain embodiments, in which a nickase and two gRNAs are used to position two single-strand nicks, one nick is on the + strand and one nick is on the – strand of the target

nucleic acid. The PAMs can be outwardly facing or inwardly facing. The gRNAs can be selected such that the gRNAs are separated by, from about 0-50, 0-100, or 0-200 nucleotides. In one embodiment, there is no overlap between the target sequences that are complementary to the targeting domains of the two gRNAs. In one embodiment, the gRNAs do not overlap
 5 and are separated by as much as 50, 100, or 200 nucleotides. In one embodiment, the use of two gRNAs can increase specificity, *e.g.*, by decreasing off-target binding (Ran 2013).

In certain embodiments, a single nick can be used to induce HDR, *e.g.*, alt-HDR. It is contemplated herein that a single nick can be used to increase the ratio of HR to NHEJ at a given cleavage site. In certain embodiments, a single-strand break is formed in the strand of
 10 the target nucleic acid to which the targeting domain of said gRNA is complementary. In certain embodiments, a single-strand break is formed in the strand of the target nucleic acid other than the strand to which the targeting domain of said gRNA is complementary.

Placement of double-strand or single-strand breaks relative to the target position

15 A double-strand break or single-strand break in one of the strands should be sufficiently close to target position that an alteration is produced in the desired region, *e.g.*, correction of a mutation occurs. In certain embodiments, the distance is not more than 50, 100, 200, 300, 350 or 400 nucleotides. While not wishing to be bound by theory, in certain embodiments, it is believed that the break should be sufficiently close to target position such
 20 that the target position is within the region that is subject to exonuclease-mediated removal during end resection. If the distance between the target position and a break is too great, the sequence desired to be altered may not be included in the end resection and, therefore, may not be altered, as donor sequence, either exogenously provided donor sequence or endogenous genomic donor sequence, in some embodiments is only used to alter sequence
 25 within the end resection region.

In certain embodiments, the gRNA targeting domain is configured such that a cleavage event, *e.g.*, a double-strand or single-strand break, is positioned within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150 or 200 nucleotides of the region desired to be altered, *e.g.*, a mutation. The break, *e.g.*, a double-strand or single-strand break,
 30 can be positioned upstream or downstream of the region desired to be altered, *e.g.*, a mutation. In some embodiments, a break is positioned within the region desired to be altered, *e.g.*, within a region defined by at least two mutant nucleotides. In some embodiments, a break is positioned immediately adjacent to the region desired to be altered, *e.g.*, immediately upstream or downstream of a mutation.

In certain embodiments, a single-strand break is accompanied by an additional single-strand break, positioned by a second gRNA molecule, as discussed below. For example, the targeting domains bind configured such that a cleavage event, *e.g.*, the two single-strand breaks, are positioned within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150 or 200 nucleotides of a target position. In one embodiment, the first and second gRNA molecules are configured such that when guiding a Cas9 nickase, a single-strand break is accompanied by an additional single-strand break, positioned by a second gRNA, sufficiently close to one another to result in alteration of the desired region. In one embodiment, the first and second gRNA molecules are configured such that a single-strand break positioned by said second gRNA is within 10, 20, 30, 40, or 50 nucleotides of the break positioned by said first gRNA molecule, *e.g.*, when the Cas9 is a nickase. In one embodiment, the two gRNA molecules are configured to position cuts at the same position, or within a few nucleotides of one another, on different strands, *e.g.*, essentially mimicking a double-strand break.

In certain embodiments, in which a gRNA (unimolecular (or chimeric) or modular gRNA) and Cas9 nuclease induce a double-strand break for the purpose of inducing HDR-mediated alteration, the cleavage site is between 0-200 bp (*e.g.*, 0-175, 0 to 150, 0 to 125, 0 to 100, 0 to 75, 0 to 50, 0 to 25, 25 to 200, 25 to 175, 25 to 150, 25 to 125, 25 to 100, 25 to 75, 25 to 50, 50 to 200, 50 to 175, 50 to 150, 50 to 125, 50 to 100, 50 to 75, 75 to 200, 75 to 175, 75 to 150, 75 to 125, 75 to 100 bp) away from the target position. In certain embodiments, the cleavage site is between 0-100 bp (*e.g.*, 0 to 75, 0 to 50, 0 to 25, 25 to 100, 25 to 75, 25 to 50, 50 to 100, 50 to 75 or 75 to 100 bp) away from the target position.

In certain embodiments, one can promote HDR by using nickases to generate a break with overhangs. While not wishing to be bound by theory, the single stranded nature of the overhangs can enhance the cell's likelihood of repairing the break by HDR as opposed to, *e.g.*, NHEJ. Specifically, in certain embodiments, HDR is promoted by selecting a first gRNA that targets a first nickase to a first target sequence, and a second gRNA that targets a second nickase to a second target sequence which is on the opposite DNA strand from the first target sequence and offset from the first nick.

In certain embodiment, the targeting domain of a gRNA molecule is configured to position a cleavage event sufficiently far from a preselected nucleotide, *e.g.*, the nucleotide of a coding region, such that the nucleotide is not altered. In certain embodiments, the targeting domain of a gRNA molecule is configured to position an intronic cleavage event sufficiently far from an intron/exon border, or naturally occurring splice signal, to avoid alteration of the

exonic sequence or unwanted splicing events. The gRNA molecule may be a first, second, third and/or fourth gRNA molecule, as described herein.

Placement of a first break and a second break relative to each other

5 In certain embodiments, a double-strand break can be accompanied by an additional double-strand break, positioned by a second gRNA molecule, as is discussed below.

In certain embodiments, a double-strand break can be accompanied by two additional single-strand breaks, positioned by a second gRNA molecule and a third gRNA molecule.

10 In certain embodiments, a first and second single-strand breaks can be accompanied by two additional single-strand breaks positioned by a third gRNA molecule and a fourth gRNA molecule.

When two or more gRNAs are used to position two or more cleavage events, *e.g.*, double-strand or single-strand breaks, in a target nucleic acid, it is contemplated that the two or more cleavage events may be made by the same or different Cas9 proteins. For example, 15 when two gRNAs are used to position two double stranded breaks, a single Cas9 nuclease may be used to create both double stranded breaks. When two or more gRNAs are used to position two or more single stranded breaks (nicks), a single Cas9 nickase may be used to create the two or more nicks. When two or more gRNAs are used to position at least one double stranded break and at least one single stranded break, two Cas9 proteins may be used, 20 *e.g.*, one Cas9 nuclease and one Cas9 nickase. It is contemplated that when two or more Cas9 proteins are used that the two or more Cas9 proteins may be delivered sequentially to control specificity of a double stranded versus a single stranded break at the desired position in the target nucleic acid.

25 In some embodiments, the targeting domain of the first gRNA molecule and the targeting domain of the second gRNA molecules are complementary to opposite strands of the target nucleic acid molecule. In some embodiments, the gRNA molecule and the second gRNA molecule are configured such that the PAMs are oriented outward. In some embodiments, the gRNA molecule and the second gRNA molecule are configured such that the PAMs are oriented inward.

30 In certain embodiments, two gRNA are selected to direct Cas9-mediated cleavage at two positions that are a preselected distance from each other. In certain embodiments, the two points of cleavage are on opposite strands of the target nucleic acid. In some embodiments, the two cleavage points form a blunt ended break, and in other embodiments, they are offset so that the DNA ends comprise one or two overhangs (*e.g.*, one or more 5'

overhangs and/or one or more 3' overhangs). In some embodiments, each cleavage event is a nick. In some embodiments, the nicks are close enough together that they form a break that is recognized by the double stranded break machinery (as opposed to being recognized by, *e.g.*, the SSBr machinery). In certain embodiments, the nicks are far enough apart that they

5 create an overhang that is a substrate for HDR, *i.e.*, the placement of the breaks mimics a DNA substrate that has experienced some resection. For instance, in some embodiments the nicks are spaced to create an overhang that is a substrate for processive resection. In some embodiments, the two breaks are spaced within 25-65 nucleotides of each other. The two breaks may be, *e.g.*, about 25, 30, 35, 40, 45, 50, 55, 60 or 65 nucleotides of each other. The

10 two breaks may be, *e.g.*, at least about 25, 30, 35, 40, 45, 50, 55, 60 or 65 nucleotides of each other. The two breaks may be, *e.g.*, at most about 30, 35, 40, 45, 50, 55, 60 or 65 nucleotides of each other. In embodiments, the two breaks are about 25-30, 30-35, 35-40, 40-45, 45-50, 50-55, 55-60, or 60-65 nucleotides of each other.

In some embodiments, the break that mimics a resected break comprises a 3'

15 overhang (*e.g.*, generated by a DSB and a nick, where the nick leaves a 3' overhang), a 5' overhang (*e.g.*, generated by a DSB and a nick, where the nick leaves a 5' overhang), a 3' and a 5' overhang (*e.g.*, generated by three cuts), two 3' overhangs (*e.g.*, generated by two nicks that are offset from each other), or two 5' overhangs (*e.g.*, generated by two nicks that are offset from each other).

20 In certain embodiments, in which two gRNAs (independently, unimolecular (or chimeric) or modular gRNA) complexing with Cas9 nickases induce two single-strand breaks for the purpose of inducing HDR-mediated alteration (*e.g.*, correction), the closer nick is between 0-200 bp (*e.g.*, 0-175, 0 to 150, 0 to 125, 0 to 100, 0 to 75, 0 to 50, 0 to 25, 25 to 200, 25 to 175, 25 to 150, 25 to 125, 25 to 100, 25 to 75, 25 to 50, 50 to 200, 50 to 175, 50 to

25 150, 50 to 125, 50 to 100, 50 to 75, 75 to 200, 75 to 175, 75 to 150, 75 to 125, or 75 to 100 bp) away from the target position and the two nicks will ideally be within 25-65 bp of each other (*e.g.*, 25 to 50, 25 to 45, 25 to 40, 25 to 35, 25 to 30, 30 to 55, 30 to 50, 30 to 45, 30 to 40, 30 to 35, 35 to 55, 35 to 50, 35 to 45, 35 to 40, 40 to 55, 40 to 50, 40 to 45 bp, 45 to 50 bp, 50 to 55 bp, 55 to 60 bp, or 60 to 65 bp) and no more than 100 bp away from each other

30 (*e.g.*, no more than 90, 80, 70, 60, 50, 40, 30, 20, 10, or 5 bp away from each other). In certain embodiments, the cleavage site is between 0-100 bp (*e.g.*, 0 to 75, 0 to 50, 0 to 25, 25 to 100, 25 to 75, 25 to 50, 50 to 100, 50 to 75, or 75 to 100 bp) away from the target position.

In some embodiments, two gRNAs, *e.g.*, independently, unimolecular (or chimeric) or modular gRNA, are configured to position a double-strand break on both sides of a target

position. In other embodiments, three gRNAs, *e.g.*, independently, unimolecular (or chimeric) or modular gRNA, are configured to position a double-strand break (*i.e.*, one gRNA complexes with a Cas9 nuclease) and two single-strand breaks or paired single stranded breaks (*i.e.*, two gRNAs complex with Cas9 nickases) on either side of the target position. In other embodiments, four gRNAs, *e.g.*, independently, unimolecular (or chimeric) or modular gRNA, are configured to generate two pairs of single stranded breaks (*i.e.*, two pairs of two gRNA molecules complex with Cas9 nickases) on either side of the target position. The double-strand break(s) or the closer of the two single-strand nicks in a pair will ideally be within 0-500 bp of the target position (*e.g.*, no more than 450, 400, 350, 300, 250, 200, 150, 100, 50 or 25 bp from the target position). When nickases are used, the two nicks in a pair are, in certain embodiments, within 25-65 bp of each other (*e.g.*, between 25 to 55, 25 to 50, 25 to 45, 25 to 40, 25 to 35, 25 to 30, 50 to 55, 45 to 55, 40 to 55, 35 to 55, 30 to 55, 30 to 50, 35 to 50, 40 to 50, 45 to 50, 35 to 45, 40 to 45 bp, 45 to 50 bp, 50 to 55 bp, 55 to 60 bp, or 60 to 65 bp) and no more than 100 bp away from each other (*e.g.*, no more than 90, 80, 70, 60, 50, 40, 30, 20, or 10 bp).

When two gRNAs are used to target Cas9 molecules to breaks, different combinations of Cas9 molecules are envisioned. In some embodiments, a first gRNA is used to target a first Cas9 molecule to a first target position, and a second gRNA is used to target a second Cas9 molecule to a second target position. In some embodiments, the first Cas9 molecule creates a nick on the first strand of the target nucleic acid, and the second Cas9 molecule creates a nick on the opposite strand, resulting in a double stranded break (*e.g.*, a blunt ended cut or a cut with overhangs).

Different combinations of nickases can be chosen to target one single stranded break to one strand and a second single stranded break to the opposite strand. When choosing a combination, one can take into account that there are nickases having one active RuvC-like domain, and nickases having one active HNH domain. In certain embodiments, a RuvC-like domain cleaves the non-complementary strand of the target nucleic acid molecule. In certain embodiments, an HNH-like domain cleaves a single stranded complementary domain, *e.g.*, a complementary strand of a double stranded nucleic acid molecule. Generally, if both Cas9 molecules have the same active domain (*e.g.*, both have an active RuvC domain or both have an active HNH domain), one will choose two gRNAs that bind to opposite strands of the target. In more detail, in some embodiments, a first gRNA is complementary with a first strand of the target nucleic acid and binds a nickase having an active RuvC-like domain and causes that nickase to cleave the strand that is non-complementary to that first gRNA, *i.e.*, a

second strand of the target nucleic acid; and a second gRNA is complementary with a second strand of the target nucleic acid and binds a nickase having an active RuvC-like domain and causes that nickase to cleave the strand that is non-complementary to that second gRNA, *i.e.*, the first strand of the target nucleic acid. Conversely, In some embodiments, a first gRNA is
5 complementary with a first strand of the target nucleic acid and binds a nickase having an active HNH domain and causes that nickase to cleave the strand that is complementary to that first gRNA, *i.e.*, a first strand of the target nucleic acid; and a second gRNA is complementary with a second strand of the target nucleic acid and binds a nickase having an active HNH domain and causes that nickase to cleave the strand that is complementary to that
10 second gRNA, *i.e.*, the second strand of the target nucleic acid. In another arrangement, if one Cas9 molecule has an active RuvC-like domain and the other Cas9 molecule has an active HNH domain, the gRNAs for both Cas9 molecules can be complementary to the same strand of the target nucleic acid, so that the Cas9 molecule with the active RuvC-like domain will cleave the non-complementary strand and the Cas9 molecule with the HNH domain will
15 cleave the complementary strand, resulting in a double stranded break.

In one embodiment, the cleavage event comprises one or more breaks, *e.g.*, one or more single-strand breaks, one or more double-strand breaks, or a combination thereof.

In one embodiment, the cleavage event comprises any one of the following: (a) one single-strand break; (b) two single-strand breaks; (c) three single-strand breaks; (d) four
20 single-strand breaks; (e) one double-strand break; (f) two double-strand breaks; (g) one single-strand break and one double-strand break; (h) two single-strand breaks and one double-strand break; or (i) any combination thereof.

In one embodiment, the gRNA molecule and the second gRNA molecule position a cleavage event on each strand of a target nucleic acid.

25 In one embodiment, the cleavage event flanks the target position, and wherein the terminus (created by the cleavage event) closest to the target position, for each cleavage event, is a 5' terminus, *e.g.*, resulting in a 5' overhang.

While not wishing to be bound by theory, it is believed that, in one embodiment, the sequence exposed by a cleavage event (*e.g.*, a single-strand cleavage event) mediated by a
30 gRNA molecule and a Cas9 fusion molecule (*e.g.*, a Cas9 nickase, *e.g.*, a Cas9 molecule containing D10A or N863A mutation) may affect (*e.g.*, increase or decrease) gene correction efficiency. For example, the sequence exposed by the cleavage event can include a 5' overhang, a 3' overhang, a product of the nucleolytic processing of a 5' overhang, a product of the nucleolytic processing of a 3' overhang, or any combination thereof. In one

embodiment, the exposed sequence comprises or consists of a 5' overhang. In another embodiment, the exposed sequence comprises or consists of a 3' overhang. In one embodiment, the exposed sequence comprises or consists of a product of the nucleolytic processing of a 5' overhang. In another embodiment, the exposed sequence comprises or consists of a product of the nucleolytic processing of a 3' overhang.

In one embodiment, the 5' overhang is between 1 and 20000 nucleotides, 5 and 20000 nucleotides, 10 and 20000 nucleotides, 20 and 20000 nucleotides, 30 and 20000 nucleotides, between 35 and 20000 nucleotides, between 40 and 20000 nucleotides, between 50 and 20000 nucleotides, between 1000 and 10000 nucleotides, or between 500 and 5000 nucleotides in length, *e.g.*, between 1 and 100 nucleotides, between 1 and 50 nucleotides, between 1 and 25 nucleotides, between 40 and 60 nucleotides, between 40 and 55 nucleotides, or between 45 and 50 nucleotides in length, *e.g.*, at least about 1, 5, 10, 20, 30, 35, 40, 45, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, or 15000 nucleotides in length. The sequence exposed by the Cas9 fusion molecule/gRNA molecule mediated cleavage event can constitute a substrate used for homology search in gene correction.

In one embodiment, the exposed sequence differs by 1, 2, 3, 4, 5, 10, 25, 50, 100 or fewer, nucleotides with an endogenous homologous sequence. In one embodiment, the exposed sequence has at least 50%, 60%, 70%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% homology with an endogenous homologous sequence over at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 750, 1000, 2500, 5000, or 10000 nucleotides. In one embodiment, the exposed sequence differs by 1, 2, 3, 4, 5, 10, 25, 50, 100 or fewer, nucleotides with an endogenous homologous sequence over at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 750, 1000, 2500, 5000, or 10000 nucleotides.

In one embodiment, the cleavage event flanks the target position, and the terminus (created by a cleavage event) closest to the target position, for each cleavage event, is a 3' terminus, *e.g.*, resulting a 3' overhang.

In one embodiment, the 3' overhang is between 1 and 20000 nucleotides, 5 and 20000 nucleotides, 10 and 20000 nucleotides, 20 and 20000 nucleotides, between 30 and 20000 nucleotides, between 35 and 20000 nucleotides, between 40 and 20000 nucleotides, between 50 and 20000 nucleotides, between 1000 and 10000 nucleotides, or between 500 and 5000 nucleotides in length, *e.g.*, between 1 and 100 nucleotides, between 1 and 50 nucleotides, between 1 and 25 nucleotides, between 40 and 60 nucleotides, between 40 and 55

nucleotides, or between 45 and 50 nucleotides in length, *e.g.*, at least about 30, 35, 40, 45, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, or 15000 nucleotides in length.

In one embodiment, the distance between the cleavage event and the target position is between 10 and 10000 nucleotides in length, *e.g.*, between 50 and 5000 nucleotides, between 100 and 1000 nucleotides, between 200 and 800 nucleotides, between 400 and 600 nucleotides, between 100 and 500 nucleotides, or between 500 and 1000 nucleotides in length, *e.g.*, at least about 20, 30, 40, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000 nucleotides in length.

In one embodiment, the cleavage event comprises a single-strand break, and wherein the distance between the single-strand break and the target position is between 10 and 10000 nucleotides in length, *e.g.*, between 50 and 5000 nucleotides, between 100 and 1000 nucleotides, between 200 and 800 nucleotides, between 400 and 600 nucleotides, between 100 and 500 nucleotides, or between 500 and 1000 nucleotides in length, *e.g.*, at least about 20, 30, 40, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000 nucleotides in length.

In one embodiment, the cleavage event comprises two, three or four single-strand breaks, and wherein the distance between each of the single-strand breaks and the target position is between 10 and 10000 nucleotides in length, *e.g.*, between 50 and 5000 nucleotides, between 100 and 1000 nucleotides, between 200 and 800 nucleotides, between 400 and 600 nucleotides, between 100 and 500 nucleotides, or between 500 and 1000 nucleotides in length, *e.g.*, at least about 20, 30, 40, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000 nucleotides in length.

In one embodiment, the cleavage event comprises a double-strand break, and wherein the distance between the double-strand break and the target position is between 10 and 10000 nucleotides in length, *e.g.*, between 50 and 5000 nucleotides, between 100 and 1000 nucleotides, between 200 and 800 nucleotides, between 400 and 600 nucleotides, between 100 and 500 nucleotides, or between 500 and 1000 nucleotides in length, *e.g.*, at least about 20, 30, 40, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000 nucleotides in length.

In one embodiment, the cleavage event comprises two double-strand breaks, and wherein the distance between each of the double-strand breaks and the target position is between 10 and 10000 nucleotides in length, *e.g.*, between 50 and 5000 nucleotides, between 100 and 1000 nucleotides, between 200 and 800 nucleotides, between 400 and 600

nucleotides, between 100 and 500 nucleotides, or between 500 and 1000 nucleotides in length, *e.g.*, at least about 20, 30, 40, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000 nucleotides in length.

In one embodiment, the cleavage event comprises a single-strand break and a double-strand break, wherein the distance between the single-strand break and the target position is between 10 and 10000 nucleotides in length, *e.g.*, between 50 and 5000 nucleotides or between 100 and 1000 nucleotides in length, *e.g.*, about 20, 30, 40, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000 nucleotides in length, and

wherein the distance between the double-strand break and the target position is between 10 and 10000 nucleotides in length, *e.g.*, between 50 and 5000 nucleotides or between 100 and 1000 nucleotides in length, *e.g.*, at least about 20, 30, 40, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000 nucleotides in length.

In one embodiment, the cleavage event comprises two single-strand breaks and a double-strand break,

wherein the distance between each of the single-strand breaks and the target position is between 10 and 10000 nucleotides in length, *e.g.*, between 50 and 5000 nucleotides or between 100 and 1000 nucleotides in length, *e.g.*, about 20, 30, 40, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000 nucleotides in length, and

wherein the distance between the double-strand break and the target position is between 10 and 10000 nucleotides in length, *e.g.*, between 50 and 5000 nucleotides or between 100 and 1000 nucleotides in length, *e.g.*, at least about 20, 30, 40, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, or 10000 nucleotides in length.

In one embodiment, the cleavage event comprises two or more single-strand breaks, two or more double-strand breaks, or two single-strand breaks and one double-strand breaks,

wherein the distance between any of the two breaks that are present on the same strand is between 30 and 20000 nucleotides, 40 and 20000 nucleotides, or 50 and 20000 nucleotides in length, *e.g.*, between 1000 and 10000 nucleotides or between 500 and 5000 nucleotides in length, *e.g.*, between 40 and 60 nucleotides, between 40 and 55 nucleotides, or between 45 and 50 nucleotides in length, *e.g.*, at least about 30, 35, 40, 45, 50, 75, 100, 200,

300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, or 15000 nucleotides in length.

In one embodiment, the cleavage event comprises two or more single-strand breaks, two or more double-strand breaks, or two single-strand breaks and one double-strand breaks,

wherein the distance between at least two breaks that are present on different strands is between 30 and 20000 nucleotides, 40 and 20000 nucleotides, or 50 and 20000 nucleotides in length, *e.g.*, between 1000 and 10000 nucleotides or between 500 and 5000 nucleotides in length, *e.g.*, between 40 and 60 nucleotides, between 40 and 55 nucleotides, or between 45 and 50 nucleotides in length, *e.g.*, at least about 30, 35, 40, 45, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, or 15000 nucleotides in length.

In one embodiment, the cleavage event comprises two single-strand breaks, wherein the distance between the two single breaks is between 30 and 20000 nucleotides, 40 and 20000 nucleotides, or 50 and 20000 nucleotides in length, *e.g.*, between 1000 and 10000 nucleotides or between 500 and 5000 nucleotides in length, *e.g.*, between 40 and 60 nucleotides, between 40 and 55 nucleotides, or between 45 and 50 nucleotides in length, *e.g.*, at least about 30, 35, 40, 45, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, or 15000 nucleotides in length. In one embodiment, the single-strand breaks are present on different strands. In another embodiment, the single-strand breaks are present on the same strand. In one embodiment, the cleavage event further comprises one or more (*e.g.*, two) of single-strand break, double-strand break, or both.

In one embodiment, the Cas9 molecule comprises HNH-like domain cleavage activity but has no, or no significant, N-terminal RuvC-like domain cleavage activity. In one embodiment, the eaCas9 molecule is an HNH-like domain nickase, *e.g.*, the Cas9 molecule comprises a mutation at D10, *e.g.*, D10A. In another embodiment, the eaCas9 molecule comprises N-terminal RuvC-like domain cleavage activity but has no, or no significant, HNH-like domain cleavage activity. In one embodiment, the Cas9 molecule is an N-terminal RuvC-like domain nickase, *e.g.*, the eaCas9 molecule comprises a mutation at N863, *e.g.*, N863A.

In one embodiment, the first gRNA molecule positions a cleavage event on a strand that does not bind to the first gRNA molecule.

In one embodiment, the second gRNA molecule positions a cleavage event on a strand that does not bind to the second gRNA molecule.

In one embodiment, the first gRNA molecule positions a cleavage event on a strand that does not bind to the first gRNA and the second gRNA molecule positions a cleavage event on a strand that does not bind to the second gRNA molecule, and wherein the gRNA molecule and the second gRNA molecule bind to different strands, *e.g.*, resulting in a 3'

5 overhang on each strand.

In one embodiment, the first gRNA molecule positions a cleavage event 5' to the target position on the first strand. In one embodiment, the second gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand. In one embodiment, the second gRNA molecule positions a cleavage event 5' to the target position on the second strand. In one embodiment, the first gRNA molecule positions a cleavage event 5' to the target position on the first strand, and wherein the second gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand. In one embodiment, the first gRNA molecule positions a cleavage event 3' to the target position on the first strand.

15 In one embodiment, the second gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand. In one embodiment, the first gRNA molecule positions a cleavage event 3' to the target position on the first strand, and wherein the second gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand. In one
20 embodiment, the first gRNA molecule positions a cleavage event 5' to the target position on the first strand, and wherein the second gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 5' overhang. In one embodiment, the first gRNA molecule positions a cleavage event 3' to the target position on the first strand, and wherein the second gRNA molecule
25 positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 5' overhang. In one embodiment, the first gRNA molecule positions a cleavage event 5' to the target position on the first strand, and wherein the second gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 3'
30 overhang. In one embodiment, the first gRNA molecule positions a cleavage event 3' to the target position on the first strand, and wherein the second gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 3' overhang. In one embodiment, the target position

comprises a mutation. In one embodiment, the mutation is associated with a disease phenotype.

In one embodiment, the first gRNA molecule positions a cleavage event on a strand that binds to the gRNA molecule.

5 In one embodiment, the second gRNA molecule positions a cleavage event on a strand that binds to the second gRNA molecule.

In one embodiment, the first gRNA molecule positions a cleavage event on a strand that binds to the gRNA and the second gRNA molecule positions a cleavage event on a strand that binds to the second gRNA molecule, and wherein the first gRNA molecule and the
10 second gRNA molecule bind to different strands, *e.g.*, resulting in a 5' overhang on each strand.

In one embodiment, the gRNA molecule, together with the Cas9 molecule (*e.g.*, a nickase), positions a cleavage event on a strand (*e.g.*, a first strand or a second strand),

In one embodiment, the gRNA molecule positions a cleavage event 5' to the target
15 position on the first strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with a D10A mutation), *e.g.*, to place a single-strand cleavage event sufficiently close to the target position (*e.g.*, within 10000, 9000, 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1000, 800, 600, 500, 400, 300, 200,
100, 75, 50, 40, 30, 20, 10, 5, or 1 bp to the target position).

20 In one embodiment, the gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with a D10A mutation), *e.g.*, to place a single-strand cleavage event sufficiently close to the target position (*e.g.*, within 10000, 9000, 8000, 7000, 6000,
25 5000, 4000, 3000, 2000, 1000, 800, 600, 500, 400, 300, 200, 100, 75, 50, 40, 30, 20, 10, 5, or 1 bp to the target position).

In one embodiment, the gRNA molecule positions a cleavage event 3' to the target position on the first strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with a D10A mutation), *e.g.*,
30 to place a single-strand cleavage event sufficiently close to the target position (*e.g.*, within 10000, 9000, 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1000, 800, 600, 500, 400, 300, 200, 100, 75, 50, 40, 30, 20, 10, 5, or 1 bp to the target position).

In one embodiment, the gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand. This

embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with a D10A mutation), *e.g.*, to place a single-strand cleavage event sufficiently close to the target position (*e.g.*, within 10000, 9000, 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1000, 800, 600, 500, 400, 300, 200, 100, 75, 50, 40, 30, 20, 10, 5, or 1 bp to the target position).

In one embodiment, the gRNA molecule positions a cleavage event 5' to the target position on the first strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with an N863A mutation), *e.g.*, to place a single-strand cleavage event sufficiently close to the target position (*e.g.*, within 10000, 9000, 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1000, 800, 600, 500, 400, 300, 200, 100, 75, 50, 40, 30, 20, 10, 5, or 1 bp to the target position).

In one embodiment, the gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with an N863A mutation), *e.g.*, to place a single-strand cleavage event sufficiently close to the target position (*e.g.*, within 10000, 9000, 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1000, 800, 600, 500, 400, 300, 200, 100, 75, 50, 40, 30, 20, 10, 5, or 1 bp to the target position).

In one embodiment, the gRNA molecule positions a cleavage event 3' to the target position on the first strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with an N863A mutation), *e.g.*, to place a single-strand cleavage event sufficiently close to the target position (*e.g.*, within 10000, 9000, 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1000, 800, 600, 500, 400, 300, 200, 100, 75, 50, 40, 30, 20, 10, 5, or 1 bp to the target position).

In one embodiment, the gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with an N863A mutation), *e.g.*, to place a single-strand cleavage event sufficiently close to the target position (*e.g.*, within 10000, 9000, 8000, 7000, 6000, 5000, 4000, 3000, 2000, 1000, 800, 600, 500, 400, 300, 200, 100, 75, 50, 40, 30, 20, 10, 5, or 1 bp to the target position).

In one embodiment the gRNA molecule, together with the Cas9 molecule (*e.g.*, a nickase), positions a cleavage event on the strand that binds to the gRNA molecule; and the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the

strand that binds to the second gRNA molecule, wherein the gRNA molecule and the second gRNA molecule bind to different strands, the gRNA molecule positions a cleavage event 5' to the target position on the first strand, and the second gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with a D10A mutation), *e.g.*, to place single-strand cleavage events on each side of the target position.

In one embodiment, the cleavage event positioned by the gRNA molecule and the cleavage event positioned by the second gRNA molecule are separated by 10 to 10000, 10 to 5000, 10 to 2500, 10 to 1000, 10 to 750, 10 to 500, 10 to 400, 10 to 300, 10 to 200, 10 to 100, 10 to 75, 10 to 50, or 10 to 25 base pairs.

In one embodiment:

the gRNA molecule, together with the Cas9 molecule (a nickase), positions a cleavage event on the strand other than the strand that binds to the gRNA molecule; and

the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand other than the strand that binds to the second gRNA molecule, wherein:

the gRNA molecule and the second gRNA molecule bind to different strands,

the gRNA molecule positions a cleavage event 5' to the target position on the first strand, and

the second gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with an N863A mutation), *e.g.*, to place single-strand cleavage events on each side of the target position.

In one embodiment, the cleavage event positioned by the gRNA molecule and the cleavage event positioned by the second gRNA molecule are separated by 10 to 10000, 10 to 5000, 10 to 2500, 10 to 1000, 10 to 750, 10 to 500, 10 to 400, 10 to 300, 10 to 200, 10 to 100, 10 to 75, 10 to 50, or 10 to 25 base pairs.

In one embodiment:

the gRNA molecule, together with the Cas9 molecule (a nickase), positions a cleavage event on the strand that binds to the gRNA molecule; and

the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand that binds to the second gRNA molecule,

wherein:

the gRNA molecule and the second gRNA molecule bind to different strands,

the gRNA molecule positions a cleavage event 3' to the target position on the first strand, and

5 the second gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with a D10A mutation), *e.g.*, to place single-strand cleavage events on each side of the target position.

10 In one embodiment, the cleavage event positioned by the gRNA molecule and the cleavage event positioned by the second gRNA molecule are separated by 10 to 10000, 10 to 5000, 10 to 2500, 10 to 1000, 10 to 750, 10 to 500, 10 to 400, 10 to 300, 10 to 200, 10 to 100, 10 to 75, 10 to 50, or 10 to 25 base pairs.

In one embodiment:

15 the gRNA molecule, together with the Cas9 molecule (a nickase), positions a cleavage event on the strand other than the strand that binds to the gRNA molecule; and

the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand other than the strand that binds to the second gRNA molecule,

wherein:

20 the gRNA molecule and the second gRNA molecule bind to different strands, the gRNA molecule positions a cleavage event 3' to the target position on the first strand, and

the second gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand. This embodiment
25 allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with a N863A mutation), *e.g.*, to place single-strand cleavage events on each side of the target position.

In one embodiment, the cleavage event positioned by the gRNA molecule and the cleavage event positioned by the second gRNA molecule are separated by 10 to 10000, 10 to
30 5000, 10 to 2500, 10 to 1000, 10 to 750, 10 to 500, 10 to 400, 10 to 300, 10 to 200, 10 to 100, 10 to 75, 10 to 50, or 10 to 25 base pairs.

In one embodiment:

the gRNA molecule, together with the Cas9 molecule (*e.g.*, a nickase), positions a cleavage event on the strand that binds to the gRNA molecule; and

the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand that binds to the second gRNA molecule,

wherein:

the gRNA molecule and the second gRNA molecule bind to different strands,

5 the gRNA molecule positions a cleavage event 5' to the target position on the first strand, and

the second gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 5' overhang. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with a D10A mutation), *e.g.*, to place single-strand cleavage events on one side of the target position, *e.g.*, to produce a 5' overhang.

In one embodiment, the cleavage event positioned by the gRNA molecule and the cleavage event positioned by the second gRNA molecule are separated by 10 to 10000, 10 to 5000, 10 to 2500, 10 to 1000, 10 to 750, 10 to 500, 10 to 400, 10 to 300, 10 to 200, 10 to 100, 15 10 to 75, 10 to 50, or 10 to 25 base pairs.

In one embodiment:

the gRNA molecule, together with the Cas9 molecule (a nickase), positions a cleavage event on the strand other than the strand that binds to the gRNA molecule; and

the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand other than the strand that binds to the second gRNA molecule,

wherein:

the gRNA molecule and the second gRNA molecule bind to different strands,

the gRNA molecule positions a cleavage event 5' to the target position on the first strand, and

25 the second gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 5' overhang. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with an N863A mutation), *e.g.*, to place single-strand cleavage events on each side of the target position, *e.g.*, to produce a 5' overhang.

In one embodiment, the cleavage event positioned by the gRNA molecule and the cleavage event positioned by the second gRNA molecule are separated by 10 to 10000, 10 to 5000, 10 to 2500, 10 to 1000, 10 to 750, 10 to 500, 10 to 400, 10 to 300, 10 to 200, 10 to 100, 10 to 75, 10 to 50, or 10 to 25 base pairs.

In one embodiment:

the gRNA molecule, together with the Cas9 molecule (*e.g.*, a nickase), positions a cleavage event on the strand that binds to the gRNA molecule; and

the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand that binds to the second gRNA molecule,
 wherein:

the gRNA molecule and the second gRNA molecule bind to different strands,

the gRNA molecule positions a cleavage event 3' to the target position on the first strand, and

the second gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 5' overhang. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with a D10A mutation), *e.g.*, to place single-strand cleavage events on one side of the target position, *e.g.*, to produce a 5' overhang.

In one embodiment, the cleavage event positioned by the gRNA molecule and the cleavage event positioned by the second gRNA molecule are separated by 10 to 10000, 10 to 5000, 10 to 2500, 10 to 1000, 10 to 750, 10 to 500, 10 to 400, 10 to 300, 10 to 200, 10 to 100, 10 to 75, 10 to 50, or 10 to 25 base pairs.

In one embodiment:

the gRNA molecule, together with the Cas9 molecule (a nickase), positions a cleavage event on the strand other than the strand that binds to the gRNA molecule; and

the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand other than the strand that binds to the second gRNA molecule,

wherein:

the gRNA molecule and the second gRNA molecule bind to different strands,

the gRNA molecule positions a cleavage event 3' to the target position on the first strand, and

the second gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 5' overhang. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with an N863A mutation), *e.g.*, to place single-strand cleavage events on each side of the target position, *e.g.*, to produce a 5' overhang.

In one embodiment, the cleavage event positioned by the gRNA molecule and the cleavage event positioned by the second gRNA molecule are separated by 10 to 10000, 10 to 5000, 10 to 2500, 10 to 1000, 10 to 750, 10 to 500, 10 to 400, 10 to 300, 10 to 200, 10 to 100, 10 to 75, 10 to 50, or 10 to 25 base pairs.

5 In one embodiment:

the gRNA molecule, together with the Cas9 molecule (*e.g.*, a nickase), positions a cleavage event on the strand that binds to the gRNA molecule; and

the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand that binds to the second gRNA molecule,

10 wherein:

the gRNA molecule and the second gRNA molecule bind to different strands,

the gRNA molecule positions a cleavage event 5' to the target position on the first strand, and

the second gRNA molecule positions a cleavage event 5' to the target position
15 (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 3' overhang. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with a D10A mutation), *e.g.*, to place single-strand cleavage events on one side of the target position, *e.g.*, to produce a 3' overhang.

In one embodiment, the cleavage event positioned by the gRNA molecule and the
20 cleavage event positioned by the second gRNA molecule are separated by 10 to 10000, 10 to 5000, 10 to 2500, 10 to 1000, 10 to 750, 10 to 500, 10 to 400, 10 to 300, 10 to 200, 10 to 100, 10 to 75, 10 to 50, or 10 to 25 base pairs.

In one embodiment:

the gRNA molecule, together with the Cas9 molecule (a nickase), positions a cleavage
25 event on the strand other than the strand that binds to the gRNA molecule; and

the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand other than the strand that binds to the second gRNA molecule,

wherein:

the gRNA molecule and the second gRNA molecule bind to different strands,

30 the gRNA molecule positions a cleavage event 5' to the target position on the first strand, and

the second gRNA molecule positions a cleavage event 5' to the target position
(relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 3' overhang. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9

molecule that is a nickase (*e.g.*, a Cas9 molecule with an N863A mutation), *e.g.*, to place single-strand cleavage events on each side of the target position, *e.g.*, to produce a 3' overhang.

In one embodiment, the cleavage event positioned by the gRNA molecule and the cleavage event positioned by the second gRNA molecule are separated by 10 to 10000, 10 to 5000, 10 to 2500, 10 to 1000, 10 to 750, 10 to 500, 10 to 400, 10 to 300, 10 to 200, 10 to 100, 10 to 75, 10 to 50, or 10 to 25 base pairs.

In one embodiment:

the gRNA molecule, together with the Cas9 molecule (*e.g.*, a nickase), positions a cleavage event on the strand that binds to the gRNA molecule; and

the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand that binds to the second gRNA molecule,

wherein:

the gRNA molecule and the second gRNA molecule bind to different strands,

the gRNA molecule positions a cleavage event 3' to the target position on the first strand, and

the second gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 3' overhang. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with a D10A mutation), *e.g.*, to place single-strand cleavage events on one side of the target position, *e.g.*, to produce a 3' overhang.

In one embodiment, the cleavage event positioned by the gRNA molecule and the cleavage event positioned by the second gRNA molecule are separated by 10 to 10000, 10 to 5000, 10 to 2500, 10 to 1000, 10 to 750, 10 to 500, 10 to 400, 10 to 300, 10 to 200, 10 to 100, 10 to 75, 10 to 50, or 10 to 25 base pairs.

In one embodiment:

the gRNA molecule, together with the Cas9 molecule (a nickase), positions a cleavage event on the strand other than the strand that binds to the gRNA molecule; and

the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand other than the strand that binds to the second gRNA molecule,

wherein:

the gRNA molecule and the second gRNA molecule bind to different strands,

the gRNA molecule positions a cleavage event 3' to the target position on the first strand, and

the second gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 3' overhang. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with an N863A mutation), *e.g.*, to place
 5 single-strand cleavage events on each side of the target position, *e.g.*, to produce a 3' overhang.

In one embodiment, the cleavage event positioned by the gRNA molecule and the cleavage event positioned by the second gRNA molecule are separated by 10 to 10000, 10 to 5000, 10 to 2500, 10 to 1000, 10 to 750, 10 to 500, 10 to 400, 10 to 300, 10 to 200, 10 to 100,
 10 10 to 75, 10 to 50, or 10 to 25 base pairs.

Homology arms of the donor template

A homology arm should extend at least as far as the region in which end resection may occur, *e.g.*, in order to allow the resected single stranded overhang to find a
 15 complementary region within the donor template. The overall length could be limited by parameters such as plasmid size or viral packaging limits. In one embodiment, a homology arm does not extend into repeated elements, *e.g.*, Alu repeats or LINE repeats.

Exemplary homology arm lengths include at least 50, 100, 250, 500, 750, 1000, 2000, 3000, 4000, or 5000 nucleotides. In some embodiments, the homology arm length is 50-100,
 20 100-250, 250-500, 500-750, 750-1000, 1000-2000, 2000-3000, 3000-4000, or 4000-5000 nucleotides.

Target position, as used herein, refers to a site on a target nucleic acid (*e.g.*, the chromosome) that is modified by a Cas9 molecule-dependent process. For example, the target position can be a modified Cas9 molecule cleavage of the target nucleic acid and
 25 template nucleic acid directed modification, *e.g.*, correction, of the target position. In one embodiment, a target position can be a site between two nucleotides, *e.g.*, adjacent nucleotides, on the target nucleic acid into which one or more nucleotides is added. The target position may comprise one or more nucleotides that are altered, *e.g.*, corrected, by a template nucleic acid. In one embodiment, the target position is within a target sequence
 30 (*e.g.*, the sequence to which the gRNA binds). In one embodiment, a target position is upstream or downstream of a target sequence (*e.g.*, the sequence to which the gRNA binds).

A template nucleic acid, as that term is used herein, refers to a nucleic acid sequence which can be used in conjunction with a Cas9 molecule and a gRNA molecule to alter the structure of a target position. In certain embodiments, the target nucleic acid is modified to

have the some or all of the sequence of the template nucleic acid, typically at or near cleavage site(s). In one embodiment, the template nucleic acid is single stranded. In certain embodiments, the template nucleic acid is double stranded. In certain embodiments, the template nucleic acid is DNA, *e.g.*, double stranded DNA. In other embodiments, the template nucleic acid is single stranded DNA. In certain embodiments, the template nucleic acid is encoded on the same vector backbone, *e.g.*, AAV genome, plasmid DNA, as the Cas9 and gRNA. In one embodiment, the template nucleic acid is excised from a vector backbone *in vivo*, *e.g.*, it is flanked by gRNA recognition sequences. In certain embodiments, the template nucleic acid comprises endogenous genomic sequence. In certain embodiments, a template nucleic acid is a template nucleic acid covalently linked to the Cas9 molecule. In certain embodiments, a template nucleic acid is a template nucleic acid non-covalently linked to the Cas9 molecule.

In certain embodiments, the template nucleic acid alters the structure of the target position by participating in an HDR event. In certain embodiments, the template nucleic acid alters the sequence of the target position. In certain embodiments, the template nucleic acid results in the incorporation of a modified, or non-naturally occurring base into the target nucleic acid.

Typically, the template sequence undergoes a breakage mediated or catalyzed recombination with the target sequence. In certain embodiments, the template nucleic acid includes sequence that corresponds to a site on the target sequence that is cleaved by an eaCas9 mediated cleavage event. In certain embodiments, the template nucleic acid includes sequence that corresponds to both, a first site on the target sequence that is cleaved in a first Cas9 mediated event, and a second site on the target sequence that is cleaved in a second Cas9 mediated event.

In one embodiment, the template nucleic acid can include sequence which results in an alteration in the coding sequence of a translated sequence, *e.g.*, one which results in the substitution of one amino acid for another in a protein product, *e.g.*, transforming a mutant allele into a wild type allele, transforming a wild type allele into a mutant allele, and/or introducing a stop codon, insertion of an amino acid residue, deletion of an amino acid residue, or a nonsense mutation.

In other embodiments, the template nucleic acid can include sequence which results in an alteration in a non-coding sequence, *e.g.*, an alteration in an exon or in a 5' or 3' non-translated or non-transcribed region. Such alterations include an alteration in a control

element, *e.g.*, a promoter, enhancer, and an alteration in a cis-acting or trans-acting control element.

A template nucleic acid having homology with a target position in the target gene can be used to alter the structure of a target sequence (*e.g.*, to correct a mutation present in a target position of an endogenous target gene). The template sequence can be used to alter an unwanted structure, *e.g.*, an unwanted or mutant nucleotide.

A template nucleic acid typically comprises the following components:

[5' homology arm]-[replacement sequence]-[3' homology arm].

The homology arms provide for recombination into the chromosome, thus replacing the undesired element, *e.g.*, a mutation or signature, with the replacement sequence. In certain embodiments, the homology arms flank the most distal cleavage sites.

In certain embodiments, the 3' end of the 5' homology arm is the position next to the 5' end of the replacement sequence. In one embodiment, the 5' homology arm can extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, or 5000 nucleotides 5' from the 5' end of the replacement sequence.

In certain embodiments, the 5' end of the 3' homology arm is the position next to the 3' end of the replacement sequence. In certain embodiments, the 3' homology arm can extend at least 10, 20, 30, 40, 50, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, or 5000 nucleotides 3' from the 3' end of the replacement sequence.

In certain embodiments, to alter one or more nucleotides at a target position (*e.g.*, to correct a mutation), the homology arms, *e.g.*, the 5' and 3' homology arms, may each comprise about 1000 bp of sequence flanking the most distal gRNAs (*e.g.*, 1000 bp of sequence on either side of the target position (*e.g.*, the mutation)).

It is contemplated herein that one or both homology arms may be shortened to avoid including certain sequence repeat elements, *e.g.*, Alu repeats or LINE elements. For example, a 5' homology arm may be shortened to avoid a sequence repeat element. In other embodiments, a 3' homology arm may be shortened to avoid a sequence repeat element. In some embodiments, both the 5' and the 3' homology arms may be shortened to avoid including certain sequence repeat elements.

It is contemplated herein that template nucleic acids for altering the sequence (*e.g.*, correcting a mutation) of a target position may be designed for use as a single-stranded oligonucleotide, *e.g.*, a single-stranded oligodeoxynucleotide (ssODN). When using a ssODN, 5' and 3' homology arms may range up to about 200 bp in length, *e.g.*, at least 25, 50, 75, 100, 125, 150, 175, or 200 bp in length. Longer homology arms are also

contemplated for ssODNs as improvements in oligonucleotide synthesis continue to be made. In some embodiments, a longer homology arm is made by a method other than chemical synthesis, *e.g.*, by denaturing a long double stranded nucleic acid and purifying one of the strands, *e.g.*, by affinity for a strand-specific sequence anchored to a solid substrate.

5 While not wishing to be bound by theory, in certain embodiments alt-HDR proceeds more efficiently when the template nucleic acid has extended homology 5' to the nick (*i.e.*, in the 5' direction of the nicked strand). Accordingly, in some embodiments, the template nucleic acid has a longer homology arm and a shorter homology arm, wherein the longer homology arm can anneal 5' of the nick. In some embodiments, the arm that can anneal 5' to the nick is at least 25, 50, 75, 100, 125, 150, 175, or 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, 2000, 3000, 4000, or 5000 nucleotides from the nick or the 5' or 3' end of the replacement sequence. In some embodiments, the arm that can anneal 5' to the nick is at least 10%, 20%, 30%, 40%, or 50% longer than the arm that can anneal 3' to the nick. In some embodiments, the arm that can anneal 5' to the nick is at least 2x, 3x, 4x, or 5x longer than the arm that can anneal 3' to the nick. Depending on whether a ssDNA template can anneal to the intact strand or the nicked strand, the homology arm that anneals 5' to the nick may be at the 5' end of the ssDNA template or the 3' end of the ssDNA template, respectively.

Similarly, in some embodiments, the template nucleic acid has a 5' homology arm, a replacement sequence, and a 3' homology arm, such that the template nucleic acid has extended homology to the 5' of the nick. For example, the 5' homology arm and 3' homology arm may be substantially the same length, but the replacement sequence may extend farther 5' of the nick than 3' of the nick. In some embodiments, the replacement sequence extends at least 10%, 20%, 30%, 40%, 50%, 2x, 3x, 4x, or 5x further to the 5' end of the nick than the 3' end of the nick.

While not wishing to be bound by theory, In some embodiments, alt-HDR proceeds more efficiently when the template nucleic acid is centered on the nick. Accordingly, in some embodiments, the template nucleic acid has two homology arms that are essentially the same size. For instance, the first homology arm of a template nucleic acid may have a length that is within 10%, 9%, 8%, 7%, 6%, 5%, 4%, 3%, 2%, or 1% of the second homology arm of the template nucleic acid.

Similarly, in some embodiments, the template nucleic acid has a 5' homology arm, a replacement sequence, and a 3' homology arm, such that the template nucleic acid extends substantially the same distance on either side of the nick. For example, the homology arms

may have different lengths, but the replacement sequence may be selected to compensate for this. For example, the replacement sequence may extend further 5' from the nick than it does 3' of the nick, but the homology arm 5' of the nick is shorter than the homology arm 3' of the nick, to compensate. The converse is also possible, *e.g.*, that the replacement sequence may extend further 3' from the nick than it does 5' of the nick, but the homology arm 3' of the nick is shorter than the homology arm 5' of the nick, to compensate.

Exemplary template nucleic acids

In a preferred embodiment, and in order to increase DNA repair via gene correction, the template nucleic acid is linked to the Cas9 molecule as part of a Cas9 fusion molecule. In certain embodiments, the template nucleic acid is double stranded. In other embodiments, the template nucleic acid is single stranded. In certain embodiments, the template nucleic acid comprises a single stranded portion and a double stranded portion. In certain embodiments, the template nucleic acid comprises about 50 to 100, *e.g.*, 55 to 95, 60 to 90, 65 to 85, or 70 to 80 bp, homology on either side of the nick and/or replacement sequence. In certain embodiments, the template nucleic acid comprises about 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, or 100 bp homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequences.

In certain embodiments, the template nucleic acid comprises about 150 to 200 bp, *e.g.*, 155 to 195, 160 to 190, 165 to 185, or 170 to 180 bp, homology 3' of the nick and/or replacement sequence. In certain embodiments, the template nucleic acid comprises about 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, or 200 bp homology 3' of the nick or replacement sequence. In some embodiments, the template nucleic acid comprises less than about 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, or 10 bp homology 5' of the nick or replacement sequence.

In certain embodiments, the template nucleic acid comprises about 150 to 200 bp, *e.g.*, 155 to 195, 160 to 190, 165 to 185, or 170 to 180 bp, homology 5' of the nick and/or replacement sequence. In certain embodiments, the template nucleic acid comprises about 150, 155, 160, 165, 170, 175, 180, 185, 190, 195, or 200 bp homology 5' of the nick or replacement sequence. In certain embodiments, the template nucleic acid comprises less than about 100, 90, 80, 70, 60, 50, 40, 30, 20, 15, or 10 bp homology 3' of the nick or replacement sequence.

In certain embodiments, the template nucleic acid comprises a nucleotide sequence, *e.g.*, of one or more nucleotides, that will be added to or will template a change in the target

nucleic acid. In other embodiments, the template nucleic acid comprises a nucleotide sequence that may be used to modify the target position. In other embodiments, the template nucleic acid comprises a nucleotide sequence, *e.g.*, of one or more nucleotides, that corresponds to wild type sequence of the target nucleic acid, *e.g.*, of the target position.

5 The template nucleic acid may comprise a replacement sequence. In some embodiments, the template nucleic acid comprises a 5' homology arm. In some embodiments, the template nucleic acid comprises a 3' homology arm.

In certain embodiments, the template nucleic acid is linear double stranded DNA.

The length may be, *e.g.*, about 150-200 bp, *e.g.*, about 150, 160, 170, 180, 190, or 200 bp.

10 The length may be, *e.g.*, at least 150, 160, 170, 180, 190, or 200 bp. In some embodiments, the length is no greater than 150, 160, 170, 180, 190, or 200 bp. In some embodiments, a double stranded template nucleic acid has a length of about 160 bp, *e.g.*, about 155-165, 150-170, 140-180, 130-190, 120-200, 110-210, 100-220, 90-230, or 80-240 bp.

The template nucleic acid can be linear single stranded DNA. In certain

15 embodiments, the template nucleic acid is (i) linear single stranded DNA that can anneal to the nicked strand of the target nucleic acid, (ii) linear single stranded DNA that can anneal to the intact strand of the target nucleic acid, (iii) linear single stranded DNA that can anneal to the plus strand of the target nucleic acid, (iv) linear single stranded DNA that can anneal to the minus strand of the target nucleic acid, or more than one of the preceding. The length
20 may be, *e.g.*, about 150-200 nucleotides, *e.g.*, about 150, 160, 170, 180, 190, or 200 nucleotides. The length may be, *e.g.*, at least 150, 160, 170, 180, 190, or 200 nucleotides. In some embodiments, the length is no greater than 150, 160, 170, 180, 190, or 200 nucleotides. In some embodiments, a single stranded template nucleic acid has a length of about 160
25 nucleotides, *e.g.*, about 155-165, 150-170, 140-180, 130-190, 120-200, 110-210, 100-220, 90-230, or 80-240 nucleotides.

In some embodiments, the template nucleic acid is circular double stranded DNA, *e.g.*, a plasmid. In some embodiments, the template nucleic acid comprises about 500 to 1000 bp of homology on either side of the replacement sequence and/or the nick. In some
30 embodiments, the template nucleic acid comprises about 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 bp of homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence. In some
embodiments, the template nucleic acid comprises at least 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 bp of homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence. In some

embodiments, the template nucleic acid comprises no more than 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 bp of homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence.

In certain embodiments, one or both homology arms may be shortened to avoid including certain sequence repeat elements, *e.g.*, Alu repeats, LINE elements. For example, a 5' homology arm may be shortened to avoid a sequence repeat element, while a 3' homology arm may be shortened to avoid a sequence repeat element. In some embodiments, both the 5' and the 3' homology arms may be shortened to avoid including certain sequence repeat elements.

In some embodiments, the Cas9 fusion molecule, comprising the template nucleic acid, is in an adenovirus vector, *e.g.*, an AAV vector, *e.g.*, a ssDNA molecule of a length and sequence that allows it to be packaged in an AAV capsid. The vector may be, *e.g.*, less than 5 kb and may contain an ITR sequence that promotes packaging into the capsid. The vector may be integration-deficient. In some embodiments, the template nucleic acid comprises about 150 to 1000 nucleotides of homology on either side of the replacement sequence and/or the nick. In some embodiments, the template nucleic acid comprises about 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence. In some embodiments, the template nucleic acid comprises at least 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence. In some embodiments, the template nucleic acid comprises at most 100, 150, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 nucleotides 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence.

In some embodiments, the Cas9 fusion molecule, comprising the template nucleic acid, is in a lentiviral vector, *e.g.*, an IDLV (integration deficiency lentivirus). In some embodiments, the template nucleic acid comprises about 500 to 1000 base pairs of homology on either side of the replacement sequence and/or the nick. In some embodiments, the template nucleic acid comprises about 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 bp of homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence. In some embodiments, the template nucleic acid comprises at least 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 bp of homology 5' of the nick or replacement sequence, 3' of the nick or replacement

sequence, or both 5' and 3' of the nick or replacement sequence. In some embodiments, the template nucleic acid comprises no more than 300, 400, 500, 600, 700, 800, 900, 1000, 1500, or 2000 bp of homology 5' of the nick or replacement sequence, 3' of the nick or replacement sequence, or both 5' and 3' of the nick or replacement sequence.

5 In certain embodiments, the template nucleic acid alters the structure of the target position by participating in an HDR event. In some embodiments, the template nucleic acid alters the sequence of the target position. In some embodiments, the template nucleic acid results in the incorporation of a modified, or non-naturally occurring nucleotide base into the target nucleic acid.

10 Typically, the template sequence undergoes a breakage mediated or catalyzed recombination with the target sequence. In some embodiments, the template nucleic acid includes sequence that corresponds to a site on the target sequence that is cleaved by an eaCas9 mediated cleavage event. In some embodiments, the template nucleic acid includes sequence that corresponds to both, a first site on the target sequence that is cleaved in a first
15 Cas9 mediated event, and a second site on the target sequence that is cleaved in a second Cas9 mediated event.

 In some embodiments, the template nucleic acid can include sequence which results in an alteration in the coding sequence of a translated sequence, *e.g.*, one which results in the substitution of one amino acid for another in a protein product, *e.g.*, transforming a mutant
20 allele into a wild type allele, transforming a wild type allele into a mutant allele, and/or introduction of a stop codon, insertion of an amino acid residue, deletion of an amino acid residue, or a nonsense mutation.

 In some embodiments, the template nucleic acid can include sequence which results in an alteration in a non-coding sequence, *e.g.*, an alteration in an exon or in a 5' or 3' non-
25 translated or non-transcribed region. Such alterations include an alteration in a control element, *e.g.*, a promoter or enhancer, or an alteration in a cis-acting or trans-acting control element.

 In some embodiments, a template nucleic acid having homology with a target position can be used to alter the structure of a target sequence. The template nucleic acid sequence
30 can be used to alter an unwanted structure, *e.g.*, an unwanted or mutant nucleotide.

 In some embodiments, shorter homology arms, *e.g.*, 5' and/or 3' homology arms may be used. In certain embodiments, the length of the 5' homology arm is about 5 to about 100 nucleotides. In some embodiments, the length of the 5' homology arm is about 10 to about 150 nucleotides. In some embodiments, the length of the 5' homology arm is about 20 to

about 150 nucleotides. In certain embodiments, the length of the 5' homology arm is about 10, 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1100, 1200, or more nucleotides in length.

In certain embodiments, the length of the 3' homology arm is about 5 to about 100 nucleotides. In some embodiments, the length of the 3' homology arm is about 10 to about 150 nucleotides. In some embodiments, the length of the 3' homology arm is about 20 to about 150 nucleotides. In certain embodiments, the length of the 3' homology arm is about 10, 20, 50, 100, 150, 200, 250, 300, 350, 400, 450, 500, 550, 600, 650, 700, 750, 800, 850, 900, 950, 1000, 1100, 1200, or more nucleotides in length.

It is contemplated herein that one or both homology arms may be shortened to avoid including certain sequence repeat elements, *e.g.*, Alu repeats, LINE elements. For example, a 5' homology arm may be shortened to avoid a sequence repeat element. In one embodiment, a 3' homology arm may be shortened to avoid a sequence repeat element. In one embodiment, both the 5' and the 3' homology arms may be shortened to avoid including certain sequence repeat elements. In some embodiments, the length of the 5' homology arm is at least 50 nucleotides in length, but not long enough to include a repeated element. In some embodiments, the length of the 5' homology arm is at least 100 nucleotides in length, but not long enough to include a repeated element. In some embodiments, the length of the 5' homology arm is at least 150 nucleotides in length, but not long enough to include a repeated element. In some embodiments, the length of the 3' homology arm is at least 50 nucleotides in length, but not long enough to include a repeated element. In some embodiments, the length of the 3' homology arm is at least 100 nucleotides in length, but not long enough to include a repeated element. In some embodiments, the length of the 3' homology arm is at least 150 nucleotides in length, but not long enough to include a repeated element.

It is contemplated herein that template nucleic acids for correcting a mutation may be designed for use as a single-stranded oligonucleotide (ssODN), *e.g.*, a single-stranded oligodeoxynucleotide. When using a ssODN, 5' and 3' homology arms may range up to about 200 bp in length, *e.g.*, at least 25, 50, 75, 100, 125, 150, 175, or 200 bp in length.

Longer homology arms are also contemplated for ssODNs as improvements in oligonucleotide synthesis continue to be made.

Silent mutations in the template nucleic acid

It is contemplated herein that Cas9 could potentially cleave donor constructs either prior to or following homology directed repair (*e.g.*, homologous recombination), resulting in a possible non-homologous-end-joining event and further DNA sequence mutation at the chromosomal locus of interest. Therefore, to avoid cleavage of the donor sequence before and/or after Cas9-mediated homology directed repair, in some embodiments, alternate versions of the donor sequence may be used where silent mutations are introduced. These silent mutations may disrupt Cas9 binding and cleavage, but not disrupt the amino acid sequence of the repaired gene. For example, mutations may include those made to a donor sequence to repair the target gene, the mutant form of which can cause disease.

Increasing Gene Correction

In certain embodiments of the methods provided herein, the frequency of preferred repair outcomes generated using a Cas9 fusion molecule described herein may be increased as compared to the frequency of preferred repair outcomes with a Cas9 fusion molecule and a template nucleic acid which are not fused. In some embodiments, the frequency of gene correction resulting from a Cas9 fusion molecule induced-lesion in a target position of a target cell overexpressing a gene correction pathway component is increased at least about 1-fold, at least about 2-fold, at least about 3-fold, at least about 4-fold, at least about 5-fold, at least about 6-fold, at least about 7-fold, at least about 8-fold, at least about 9-fold, at least about 10-fold, or more, as compared to the frequency of gene correction resulting from a Cas9 molecule and a target nucleic acid which are not fused in a target position.

In some embodiments, the frequency of gene correction resulting from a Cas9 fusion molecule induced-lesion in a target position of a target cell overexpressing a gene correction pathway component is increased at least 5% (*e.g.*, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at least about 85%, at least about 90%, at least about 95%, at least about 100%, at least about 150%, at least about 200%, at least about 300%, at least about 400%, at least about 500%, at least about 600%, at least about 700%, at least about 800%, at least about 900%, or more.

NHEJ Approaches for Gene Targeting

In certain embodiments of the methods provided herein, NHEJ-mediated deletion is used to delete all or part of a target gene. As described herein, nuclease-induced NHEJ can also be used to remove (*e.g.*, delete) sequences in a gene of interest.

While not wishing to be bound by theory, it is believed that, in certain embodiments, the genomic alterations associated with the methods described herein rely on nuclease-induced NHEJ and the error-prone nature of the NHEJ repair pathway. NHEJ repairs a double-strand break in the DNA by joining together the two ends; however, generally, the original sequence is restored only if two compatible ends, exactly as they were formed by the double-strand break, are perfectly ligated. The DNA ends of the double-strand break are frequently the subject of enzymatic processing, resulting in the addition or removal of nucleotides, *e.g.*, resection, at one or both strands, prior to rejoining of the ends. This results in the presence of insertion and/or deletion (indel) mutations in the DNA sequence at the site of the NHEJ repair. Two-thirds of these mutations typically alter the reading frame and, therefore, produce a non-functional protein. Additionally, mutations that maintain the reading frame, but which insert or delete a significant amount of sequence, can destroy functionality of the protein. This is locus dependent as mutations in critical functional domains are likely less tolerable than mutations in non-critical regions of the protein.

The indel mutations generated by NHEJ are unpredictable in nature; however, at a given break site certain indel sequences are favored and are over represented in the population, likely due to small regions of microhomology. The lengths of deletions can vary widely; most commonly in the 1-50 bp range, but they can easily reach greater than 100-200 bp. In some embodiments, the deletion is at least about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, 40, 47, 50, 75, 100, 200, 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, 15000, 20000, 25000, 30000, 40000, 50000, 60000, 70000, 80000, 90000, 100000, 200000, 300000, 400000, 500000, 600000, 700000, 800000, 900000, 1000000 or more nucleotides in length. Insertions tend to be shorter and often include short duplications of the sequence immediately surrounding the break site. However, it is possible to obtain large insertions, and in these cases, the inserted sequence has often been traced to other regions of the genome or to plasmid DNA present in the cells.

Because NHEJ is a mutagenic process, it can also be used to delete small sequence motifs as long as the generation of a specific final sequence is not required. If a double-strand break is targeted near to a short target sequence, the deletion mutations caused by the

NHEJ repair often span, and therefore remove, the unwanted nucleotides. For the deletion of larger DNA segments, introducing two double-strand breaks, one on each side of the sequence, can result in NHEJ between the ends with removal of the entire intervening sequence. Both of these approaches can be used to delete specific DNA sequences; however, the error-prone nature of NHEJ may still produce indel mutations at the site of repair.

Both double-strand cleaving Cas9 molecules and single strand, or nickase, Cas9 molecules can be used in the methods and compositions described herein to generate NHEJ-mediated indels. NHEJ-mediated indels targeted to the gene, *e.g.*, a coding region, *e.g.*, an early coding region of a gene of interest can be used to knockout (*i.e.*, eliminate expression of) a gene of interest. For example, early coding region of a gene of interest includes sequence immediately following a transcription start site, within a first exon of the coding sequence, or within 500 bp of the transcription start site (*e.g.*, less than 500, 450, 400, 350, 300, 250, 200, 150, 100 or 50 bp).

Placement of double-strand or single-strand breaks relative to the target position

In certain embodiments, in which a gRNA and Cas9 nuclease generate a double-strand break for the purpose of inducing NHEJ-mediated indels, a gRNA, *e.g.*, a unimolecular (or chimeric) or modular gRNA molecule, is configured to position one double-strand break in close proximity to a nucleotide of the target position. In one embodiment, the cleavage site is between 0-30 bp away from the target position (*e.g.*, less than 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2 or 1 bp from the target position).

In certain embodiments, in which two gRNAs complexing with Cas9 nickases induce two single-strand breaks for the purpose of inducing NHEJ-mediated indels, two gRNAs, *e.g.*, independently, unimolecular (or chimeric) or modular gRNA, are configured to position two single-strand breaks to provide for NHEJ repair a nucleotide of the target position. In certain embodiments, the gRNAs are configured to position cuts at the same position, or within a few nucleotides of one another, on different strands, essentially mimicking a double-strand break. In certain embodiments, the closer nick is between 0-30 bp away from the target position (*e.g.*, less than 30, 25, 20, 15, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1 bp from the target position), and the two nicks are within 25-55 bp of each other (*e.g.*, between 25 to 50, 25 to 45, 25 to 40, 25 to 35, 25 to 30, 50 to 55, 45 to 55, 40 to 55, 35 to 55, 30 to 55, 30 to 50, 35 to 50, 40 to 50, 45 to 50, 35 to 45, or 40 to 45 bp) and no more than 100 bp away from each other (*e.g.*, no more than 90, 80, 70, 60, 50, 40, 30, 20, or 10 bp). In certain embodiments, the gRNAs are configured to place a single-strand break on either side of a nucleotide of the target position.

Both double-strand cleaving eaCas9 molecules and single strand, or nickase, eaCas9 molecules can be used in the methods and compositions described herein to generate breaks both sides of a target position. Double-strand or paired single-strand breaks may be generated on both sides of a target position to remove the nucleic acid sequence between the two cuts (*e.g.*, the region between the two breaks is deleted). In certain embodiments, two gRNAs, *e.g.*, independently, unimolecular (or chimeric) or modular gRNA, are configured to position a double-strand break on both sides of a target position. In other embodiments, three gRNAs, *e.g.*, independently, unimolecular (or chimeric) or modular gRNA, are configured to position a double-strand break (*i.e.*, one gRNA complexes with a Cas9 nuclease) and two single-strand breaks or paired single-strand breaks (*i.e.*, two gRNAs complex with Cas9 nickases) on either side of the target position. In certain embodiments, four gRNAs, *e.g.*, independently, unimolecular (or chimeric) or modular gRNA, are configured to generate two pairs of single-strand breaks (*i.e.*, two pairs of two gRNAs complex with Cas9 nickases) on either side of the target position. The double-strand break(s) or the closer of the two single-strand nicks in a pair will ideally be within 0-500 bp of the target position (*e.g.*, no more than 450, 400, 350, 300, 250, 200, 150, 100, 50, or 25 bp from the target position). When nickases are used, the two nicks in a pair are within 25-55 bp of each other (*e.g.*, between 25 to 50, 25 to 45, 25 to 40, 25 to 35, 25 to 30, 50 to 55, 45 to 55, 40 to 55, 35 to 55, 30 to 55, 30 to 50, 35 to 50, 40 to 50, 45 to 50, 35 to 45, or 40 to 45 bp) and no more than 100 bp away from each other (*e.g.*, no more than 90, 80, 70, 60, 50, 40, 30, 20, or 10 bp).

Targeted Knockdown

Unlike CRISPR/Cas-mediated gene knockout, which permanently eliminates expression by mutating the gene at the DNA level, CRISPR/Cas knockdown allows for temporary reduction of gene expression through the use of artificial transcription factors. Mutating key residues in both DNA cleavage domains of the Cas9 molecule (*e.g.*, the D10A and H840A mutations) results in the generation of a catalytically inactive Cas9 (referred to herein as “eiCas9”, which is also known as dead Cas9 or dCas9) molecule. An eiCas9 complexes with a gRNA and localizes to the DNA sequence specified by that gRNA’s targeting domain, however, it does not cleave the target DNA. Fusion of the eiCas9 to an effector domain, *e.g.*, a transcription repression domain, enables recruitment of the effector to any DNA site specified by the gRNA. Although an eiCas9 itself can block transcription when recruited to early regions in the coding sequence, more robust repression can be achieved by fusing a transcriptional repression domain (for example KRAB, SID or ERD) to

the eiCas9, referred to herein as a “Cas9-repressor”, and recruiting the transcriptional repression domain to the target knockdown position, *e.g.*, within 1000 bp of sequence 3’ of the start codon or within 500 bp of a promoter region 5’ of the start codon of a gene. It is likely that targeting DNase I hypersensitive sites (DHSs) of the promoter may yield more efficient gene repression or activation because these regions are more likely to be accessible to the eiCas9 and are also more likely to harbor sites for endogenous transcription factors. Especially for gene repression, it is contemplated herein that blocking the binding site of an endogenous transcription factor would aid in downregulating gene expression. In certain embodiments, one or more eiCas9 molecules may be used to block binding of one or more endogenous transcription factors. In some embodiments, an eiCas9 molecule can be fused to a chromatin modifying protein. Altering chromatin status can result in decreased expression of the target gene. One or more eiCas9 molecules fused to one or more chromatin modifying proteins may be used to alter chromatin status.

In one embodiment, a gRNA molecule can be targeted to a known transcription response elements (*e.g.*, promoters, enhancers, etc.), a known upstream activating sequences (UAS), and/or sequences of unknown or known function that are suspected of being able to control expression of the target DNA.

CRISPR/Cas-mediated gene knockdown can be used to reduce expression of an unwanted allele or transcript. Contemplated herein are scenarios wherein permanent destruction of the gene is not ideal. In these scenarios, site-specific repression may be used to temporarily reduce or eliminate expression. It is also contemplated herein that the off-target effects of a Cas9-repressor may be less severe than those of a Cas9-nuclease as a nuclease can cleave any DNA sequence and cause mutations whereas a Cas9-repressor may only have an effect if it targets the promoter region of an actively transcribed gene. However, while nuclease-mediated knockout is permanent, repression may only persist as long as the Cas9-repressor is present in the cells. Once the repressor is no longer present, it is likely that endogenous transcription factors and gene regulatory elements would restore expression to its natural state.

Single-Strand Annealing

Single-strand annealing (SSA) is another DNA repair process that repairs a double-strand break between two repeat sequences present in a target nucleic acid. Repeat sequences utilized by the SSA pathway are generally greater than 30 nucleotides in length. Resection at

the break ends occurs to reveal repeat sequences on both strands of the target nucleic acid. After resection, single-strand overhangs containing the repeat sequences are coated with RPA protein to prevent the repeats sequences from inappropriate annealing, *e.g.*, to themselves. RAD52 binds to and each of the repeat sequences on the overhangs and aligns the sequences to enable the annealing of the complementary repeat sequences. After annealing, the single-strand flaps of the overhangs are cleaved. New DNA synthesis fills in any gaps, and ligation restores the DNA duplex. As a result of the processing, the DNA sequence between the two repeats is deleted. The length of the deletion can depend on many factors including the location of the two repeats utilized, and the pathway or processivity of the resection.

In contrast to HDR pathways, SSA does not require a template nucleic acid to alter or correct a target nucleic acid sequence. Instead, the complementary repeat sequence is utilized.

Other DNA Repair Pathways

SSBR (single-strand break repair)

Single-stranded breaks (SSB) in the genome are repaired by the SSBR pathway, which is a distinct mechanism from the DSB repair mechanisms discussed above. The SSBR pathway has four major stages: SSB detection, DNA end processing, DNA gap filling, and DNA ligation. A more detailed explanation is given in Caldecott 2008, and a summary is given here.

In the first stage, when a SSB forms, PARP1 and/or PARP2 recognize the break and recruit repair machinery. The binding and activity of PARP1 at DNA breaks is transient and it seems to accelerate SSBr by promoting the focal accumulation or stability of SSBr protein complexes at the lesion. Arguably the most important of these SSBr proteins is XRCC1, which functions as a molecular scaffold that interacts with, stabilizes, and stimulates multiple enzymatic components of the SSBr process including the protein responsible for cleaning the DNA 3' and 5' ends. For instance, XRCC1 interacts with several proteins (DNA polymerase beta, PNK, and three nucleases, APE1, APTX, and APLF) that promote end processing. APE1 has endonuclease activity. APLF exhibits endonuclease and 3' to 5' exonuclease activities. APTX has endonuclease and 3' to 5' exonuclease activity.

This end processing is an important stage of SSBR since the 3'- and/or 5'-termini of most, if not all, SSBs are damaged. End processing generally involves restoring a damaged 3'-end to a hydroxylated state and and/or a damaged 5' end to a phosphate moiety, so that the ends become ligation-competent. Enzymes that can process damaged 3' termini include

PNKP, APE1, and TDP1. Enzymes that can process damaged 5' termini include PNKP, DNA polymerase beta, and APTX. LIG3 (DNA ligase III) can also participate in end processing. Once the ends are cleaned, gap filling can occur.

At the DNA gap filling stage, the proteins typically present are PARP1, DNA polymerase beta, XRCC1, FEN1 (flap endonuclease 1), DNA polymerase delta/epsilon, PCNA, and LIG1. There are two ways of gap filling, the short patch repair and the long patch repair. Short patch repair involves the insertion of a single nucleotide that is missing. At some SSBs, "gap filling" might continue displacing two or more nucleotides (displacement of up to 12 bases have been reported). FEN1 is an endonuclease that removes the displaced 5'-residues. Multiple DNA polymerases, including Pol β , are involved in the repair of SSBs, with the choice of DNA polymerase influenced by the source and type of SSB.

In the fourth stage, a DNA ligase such as LIG1 (Ligase I) or LIG3 (Ligase III) catalyzes joining of the ends. Short patch repair uses Ligase III and long patch repair uses Ligase I.

Sometimes, SSBR is replication-coupled. This pathway can involve one or more of CtIP, MRN, ERCC1, and FEN1. Additional factors that may promote SSBR include: aPARP, PARP1, PARP2, PARG, XRCC1, DNA polymerase β , DNA polymerase delta, DNA polymerase epsilon, PCNA, LIG1, PNK, PNKP, APE1, APTX, APLF, TDP1, LIG3, FEN1, CtIP, MRN, and ERCC1.

MMR (mismatch repair)

Cells contain three excision repair pathways: MMR, BER, and NER. The excision repair pathways have a common feature in that they typically recognize a lesion on one strand of the DNA, then exo/endonucleases remove the lesion and leave a 1-30 nucleotide gap that is sub-sequentially filled in by DNA polymerase and finally sealed with ligase. A more complete picture is given in Li 2008, and a summary is provided here.

Mismatch repair (MMR) operates on mispaired DNA bases.

The MSH2/6 or MSH2/3 complexes both have ATPase activity that plays an important role in mismatch recognition and the initiation of repair. MSH2/6 preferentially recognizes base-base mismatches and identifies mispairs of 1 or 2 nucleotides, while MSH2/3 preferentially recognizes larger ID mispairs.

hMLH1 heterodimerizes with hPMS2 to form hMutL α which possesses an ATPase activity and is important for multiple steps of MMR. It possesses a PCNA/replication factor C (RFC)-dependent endonuclease activity which plays an important role in 3' nick-directed

MMR involving EXO1 (EXO1 is a participant in both HR and MMR). It regulates termination of mismatch-provoked excision. Ligase I is the relevant ligase for this pathway. Additional factors that may promote MMR include: EXO1, MSH2, MSH3, MSH6, MLH1, PMS2, MLH3, DNA Pol delta, RPA, HMGB1, RFC, and DNA ligase I.

Base excision repair (BER)

The base excision repair (BER) pathway is active throughout the cell cycle; it is responsible primarily for removing small, non-helix-distorting base lesions from the genome. In contrast, the related Nucleotide Excision Repair pathway (discussed in the next section) repairs bulky helix-distorting lesions. A more detailed explanation is given in Caldecott 2008, and a summary is given here.

Upon DNA base damage, base excision repair (BER) is initiated and the process can be simplified into five major steps: (a) removal of the damaged DNA base; (b) incision of the subsequent a basic site; (c) clean-up of the DNA ends; (d) insertion of the desired nucleotide into the repair gap; and (e) ligation of the remaining nick in the DNA backbone. These last steps are similar to the SSBR.

In the first step, a damage-specific DNA glycosylase excises the damaged base through cleavage of the N-glycosidic bond linking the base to the sugar phosphate backbone. Then AP endonuclease-1 (APE1) or bifunctional DNA glycosylases with an associated lyase activity incises the phosphodiester backbone to create a DNA single-strand break (SSB). The third step of BER involves cleaning-up of the DNA ends. The fourth step in BER is conducted by Pol β that adds a new complementary nucleotide into the repair gap and in the final step XRCC1/Ligase III seals the remaining nick in the DNA backbone. This completes the short-patch BER pathway in which the majority (~80%) of damaged DNA bases are repaired. However, if the 5'-ends in step 3 are resistant to end processing activity, following one nucleotide insertion by Pol β there is then a polymerase switch to the replicative DNA polymerases, Pol δ/ϵ , which then add ~2–8 more nucleotides into the DNA repair gap. This creates a 5'-flap structure, which is recognized and excised by flap endonuclease-1 (FEN-1) in association with the processivity factor proliferating cell nuclear antigen (PCNA). DNA ligase I then seals the remaining nick in the DNA backbone and completes long-patch BER. Additional factors that may promote the BER pathway include: DNA glycosylase, APE1, Pol β , Pol delta, Pol epsilon, XRCC1, Ligase III, FEN-1, PCNA, RECQL4, WRN, MYH, PNKP, and APTX.

Nucleotide excision repair (NER)

Nucleotide excision repair (NER) is an important excision mechanism that removes bulky helix-distorting lesions from DNA. Additional details about NER are given in Marteijn *et al.* 2014, and a summary is given here. NER a broad pathway encompassing two smaller pathways: global genomic NER (GG-NER) and transcription coupled repair NER (TC-NER). GG-NER and TC-NER use different factors for recognizing DNA damage. However, they utilize the same machinery for lesion incision, repair, and ligation.

Once damage is recognized, the cell removes a short single-stranded DNA segment that contains the lesion. Endonucleases XPF/ERCC1 and XPG (encoded by ERCC5) remove the lesion by cutting the damaged strand on either side of the lesion, resulting in a single-strand gap of 22–30 nucleotides. Next, the cell performs DNA gap filling synthesis and ligation. Involved in this process are: PCNA, RFC, DNA Pol δ , DNA Pol ϵ or DNA Pol κ , and DNA ligase I or XRCC1/Ligase III. Replicating cells tend to use DNA pol ϵ and DNA ligase I, while non-replicating cells tend to use DNA Pol δ , DNA Pol κ , and the XRCC1/Ligase III complex to perform the ligation step.

NER can involve the following factors: XPA-G, POLH, XPF, ERCC1, XPA-G, and LIG1. Transcription-coupled NER (TC-NER) can involve the following factors: CSA, CSB, XPB, XPD, XPG, ERCC1, and TTDA. Additional factors that may promote the NER repair pathway include XPA-G, POLH, XPF, ERCC1, XPA-G, LIG1, CSA, CSB, XPA, XPB, XPC, XPD, XPF, XPG, TTDA, UVSSA, USP7, CETN2, RAD23B, UV-DDB, CAK subcomplex, RPA, and PCNA.

Interstrand Crosslink (ICL)

A dedicated pathway called the ICL repair pathway repairs interstrand crosslinks. Interstrand crosslinks, or covalent crosslinks between bases in different DNA strand, can occur during replication or transcription. ICL repair involves the coordination of multiple repair processes, in particular, nucleolytic activity, translesion synthesis (TLS), and HDR. Nucleases are recruited to excise the ICL on either side of the crosslinked bases, while TLS and HDR are coordinated to repair the cut strands. ICL repair can involve the following factors: endonucleases, *e.g.*, XPF and RAD51C, endonucleases such as RAD51, translesion polymerases, *e.g.*, DNA polymerase zeta and Rev1, and the Fanconi anemia (FA) proteins, *e.g.*, FancJ.

Other pathways

Several other DNA repair pathways exist in mammals.

Translesion synthesis (TLS) is a pathway for repairing a single stranded break left after a defective replication event and involves translesion polymerases, *e.g.*, DNA pol ζ and Rev1.

Error-free postreplication repair (PRR) is another pathway for repairing a single stranded break left after a defective replication event.

Examples of gRNAs in Genome Editing Methods

gRNA molecules as described herein can be used with Cas9 molecules, or Cas9 fusion molecules, that generate a double-strand break or a single-strand break to alter the sequence of a target nucleic acid, *e.g.*, a target position or target genetic signature. gRNA molecules useful in these methods are described below.

In certain embodiments, the gRNA, *e.g.*, a chimeric gRNA, is configured such that it comprises one or more of the following properties;

a) it can position, *e.g.*, when targeting a Cas9 molecule, or Cas9 fusion molecule, that makes double-strand breaks, a double-strand break (i) within 50, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of a target position, or (ii) sufficiently close that the target position is within the region of end resection;

b) it has a targeting domain of at least 16 nucleotides, *e.g.*, a targeting domain of (i) 16, (ii) 17, (iii) 18, (iv) 19, (v) 20, (vi) 21, (vii) 22, (viii) 23, (ix) 24, (x) 25, or (xi) 26 nucleotides; and

(c)(i) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides, *e.g.*, at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides from a naturally occurring *S. pyogenes* or *S. aureus*, tail and proximal domain, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(ii) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain, *e.g.*, at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides from the corresponding sequence of a naturally occurring *S. pyogenes* or *S. aureus* gRNA, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(iii) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain, *e.g.*, at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides from the corresponding sequence of a naturally

occurring *S. pyogenes* or *S. aureus* gRNA, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(iv) the tail domain is at least 10, 15, 20, 25, 30, 35 or 40 nucleotides in length, *e.g.*, it comprises at least 10, 15, 20, 25, 30, 35 or 40 nucleotides from a naturally occurring *S. pyogenes* or *S. aureus* tail domain, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom; or

(c)(v) the tail domain comprises 15, 20, 25, 30, 35, 40 nucleotides or all of the corresponding portions of a naturally occurring tail domain, *e.g.*, a naturally occurring *S. pyogenes* or *S. aureus* tail domain.

In certain embodiments, the gRNA is configured such that it comprises properties a and b(i); a and b(ii); a and b(iii); a and b(iv); a and b(v); a and b(vi); a and b(vii); a and b(viii); a and b(ix); a and b(x); a and b(xi); a and c; a, b, and c; a(i), b(i), and c(i); a(i), b(i), and c(ii); a(i), b(ii), and c(i); a(i), b(ii), and c(ii); a(i), b(iii), and c(i); a(i), b(iii), and c(ii); a(i), b(iv), and c(i); a(i), b(iv), and c(ii); a(i), b(v), and c(i); a(i), b(v), and c(ii); a(i), b(vi), and c(i); a(i), b(vi), and c(ii); a(i), b(vii), and c(i); a(i), b(vii), and c(ii); a(i), b(viii), and c(i); a(i), b(viii), and c(ii); a(i), b(ix), and c(i); a(i), b(ix), and c(ii); a(i), b(x), and c(i); a(i), b(x), and c(ii); a(i), b(xi), or c(i); a(i), b(xi), and c(ii).

In certain embodiments, the gRNA, *e.g.*, a chimeric gRNA, is configured such that it comprises one or more of the following properties:

(a) one or both of the gRNAs can position, *e.g.*, when targeting a Cas9 molecule, or Cas9 fusion molecule, that makes single-strand breaks, a single-strand break within (i) 50, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of a target position, or (ii) sufficiently close that the target position is within the region of end resection;

(b) one or both have a targeting domain of at least 16 nucleotides, *e.g.*, a targeting domain of (i) 16, (ii) 17, (iii) 18, (iv) 19, (v) 20, (vi) 21, (vii) 22, (viii) 23, (ix) 24, (x) 25, or (xi) 26 nucleotides; and

(c)(i) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides, *e.g.*, at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides from a naturally occurring *S. pyogenes* or *S. aureus* tail and proximal domain, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(ii) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain, *e.g.*, at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides from the corresponding sequence of a naturally

occurring *S. pyogenes*, or *S. aureus* gRNA, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(iii) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain, *e.g.*, at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides from the corresponding sequence of a naturally occurring *S. pyogenes* or *S. aureus* gRNA, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(iv) the tail domain is at least 10, 15, 20, 25, 30, 35 or 40 nucleotides in length, *e.g.*, it comprises at least 10, 15, 20, 25, 30, 35 or 40 nucleotides from a naturally occurring *S. pyogenes*, or *S. aureus* tail domain, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom; or

(c)(v) the tail domain comprises 15, 20, 25, 30, 35, 40 nucleotides or all of the corresponding portions of a naturally occurring tail domain, *e.g.*, a naturally occurring *S. pyogenes* or *S. aureus* tail domain.

In certain embodiments, the gRNA is configured such that it comprises properties: a and b(i); a and b(ii); a and b(iii); a and b(iv); a and b(v); a and b(vi); a and b(vii); a and b(viii); a and b(ix); a and b(x); a and b(xi); a and c; a, b, and c; a(i), b(i), and c(i); a(i), b(i), and c(ii); a(i), b(ii), and c(i); a(i), b(ii), and c(ii); a(i), b(iii), and c(i); a(i), b(iii), and c(ii); a(i), b(iv), and c(i); a(i), b(iv), and c(ii); a(i), b(v), and c(i); a(i), b(v), and c(ii); a(i), b(vi), and c(i); a(i), b(vi), and c(ii); a(i), b(vii), and c(i); a(i), b(vii), and c(ii); a(i), b(viii), and c(i); a(i), b(viii), and c(ii); a(i), b(ix), and c(i); a(i), b(ix), and c(ii); a(i), b(x), and c(i); a(i), b(x), and c(ii); a(i), b(xi), and c(i); or a(i), b(xi), and c(ii).

In certain embodiments, the gRNA is used with a Cas9 nickase molecule having HNH activity, *e.g.*, a Cas9 molecule, or a Cas9 fusion molecule, having the RuvC activity inactivated, *e.g.*, a Cas9 molecule having a mutation at D10, *e.g.*, the D10A mutation.

In one embodiment, the gRNA is used with a Cas9 nickase molecule having RuvC activity, *e.g.*, a Cas9 molecule, or a Cas9 fusion molecule, having the HNH activity inactivated, *e.g.*, a Cas9 molecule having a mutation at 840, *e.g.*, the H840A.

In one embodiment, the gRNAs are used with a Cas9 nickase molecule having RuvC activity, *e.g.*, a Cas9 molecule, or a Cas9 fusion molecule, having the HNH activity inactivated, *e.g.*, a Cas9 molecule having a mutation at N863, *e.g.*, the N863A mutation.

In embodiment, a pair of gRNAs, *e.g.*, a pair of chimeric gRNAs, comprising a first and a second gRNA, is configured such that they comprises one or more of the following properties:

a) one or both of the gRNA molecules can position, *e.g.*, when targeting a Cas9 molecule, or a Cas9 fusion molecule, that makes single-strand breaks, a single-strand break within (i) 50, 100, 150, 200, 250, 300, 350, 400, 450, or 500 nucleotides of a target position, or (ii) sufficiently close that the target position is within the region of end resection;

b) one or both have a targeting domain of at least 16 nucleotides, *e.g.*, a targeting domain of (i) 16, (ii) 17, (iii) 18, (iv) 19, (v) 20, (vi) 21, (vii) 22, (viii) 23, (ix) 24, (x) 25, or (xi) 26 nucleotides;

(c)(i) the proximal and tail domain, when taken together, comprise at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides, *e.g.*, at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides from a naturally occurring *S. pyogenes* or *S. aureus* tail and proximal domain, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(ii) there are at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides 3' to the last nucleotide of the second complementarity domain, *e.g.*, at least 15, 18, 20, 25, 30, 31, 35, 40, 45, 49, 50, or 53 nucleotides from the corresponding sequence of a naturally occurring *S. pyogenes* or *S. aureus* gRNA, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(iii) there are at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides 3' to the last nucleotide of the second complementarity domain that is complementary to its corresponding nucleotide of the first complementarity domain, *e.g.*, at least 16, 19, 21, 26, 31, 32, 36, 41, 46, 50, 51, or 54 nucleotides from the corresponding sequence of a naturally occurring *S. pyogenes* or *S. aureus* gRNA, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom;

(c)(iv) the tail domain is at least 10, 15, 20, 25, 30, 35 or 40 nucleotides in length, *e.g.*, it comprises at least 10, 15, 20, 25, 30, 35 or 40 nucleotides from a naturally occurring *S. pyogenes* or *S. aureus* tail domain; or, or a sequence that differs by no more than 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides therefrom; or

(c)(v) the tail domain comprises 15, 20, 25, 30, 35, or 40 nucleotides or all of the corresponding portions of a naturally occurring tail domain, *e.g.*, a naturally occurring *S. pyogenes* or *S. aureus* tail domain;

(d) the gRNAs are configured such that, when hybridized to target nucleic acid, they are separated by 0-50, 0-100, 0-200, at least 10, at least 20, at least 30 or at least 50 nucleotides;

(e) the breaks made by the first gRNA and second gRNA are on different strands; and

5 (f) the PAMs are facing outwards.

In certain embodiments, one or both of the gRNAs is configured such that it comprises properties a and b(i); a and b(ii); a and b(iii); a and b(iv); a and b(v); a and b(vi); a and b(vii); a and b(viii); a and b(ix); a and b(x); a and b(xi); a and c; a, b, and c; a(i), b(i), and c(i); a(i), b(i), and c(ii); a(i), b(i), c, and d; a(i), b(i), c, and e; a(i), b(i), c, d, and e; a(i), b(ii), and c(i); a(i), b(ii), and c(ii); a(i), b(ii), c, and d; a(i), b(ii), c, and e; a(i), b(ii), c, d, and e; a(i), b(iii), and c(i); a(i), b(iii), and c(ii); a(i), b(iii), c, and d; a(i), b(iii), c, and e; a(i), b(iii), c, d, and e; a(i), b(iv), and c(i); a(i), b(iv), and c(ii); a(i), b(iv), c, and d; a(i), b(iv), c, and e; a(i), b(iv), c, d, and e; a(i), b(v), and c(i); a(i), b(v), and c(ii); a(i), b(v), c, and d; a(i), b(v), c, and e; a(i), b(v), c, d, and e; a(i), b(vi), and c(i); a(i), b(vi), and c(ii); a(i), b(vi), c, and d; a(i), b(vi), c, and e; a(i), b(vi), c, d, and e; a(i), b(vii), and c(i); a(i), b(vii), and c(ii); a(i), b(vii), c, and d; a(i), b(vii), c, and e; a(i), b(vii), c, d, and e; a(i), b(viii), and c(i); a(i), b(viii), and c(ii); a(i), b(viii), c, and d; a(i), b(viii), c, and e; a(i), b(viii), c, d, and e; a(i), b(ix), and c(i); a(i), b(ix), and c(ii); a(i), b(ix), c, and d; a(i), b(ix), c, and e; a(i), b(ix), c, d, and e; a(i), b(x), and c(i); a(i), b(x), and c(ii); a(i), b(x), c, and d; a(i), b(x), c, and e; a(i), b(x), c, d, and e; a(i), b(xi), and c(i); a(i), b(xi), and c(ii); a(i), b(xi), c, and d; a(i), b(xi), c, and e; or a(i), b(xi), c, d, and e.

In certain embodiments, the gRNAs are used with a Cas9 nickase molecule having HNH activity, *e.g.*, a Cas9 molecule, or a Cas9 fusion molecule, having the RuvC activity inactivated, *e.g.*, a Cas9 molecule having a mutation at D10, *e.g.*, the D10A mutation.

25 In certain embodiments, the gRNAs are used with a Cas9 nickase molecule having RuvC activity, *e.g.*, a Cas9 molecule, or a Cas9 fusion molecule, having the HNH activity inactivated, *e.g.*, a Cas9 molecule having a mutation at H840, *e.g.*, the H840A mutation.

In certain embodiments, the gRNAs are used with a Cas9 nickase molecule having RuvC activity, *e.g.*, a Cas9 molecule, or a Cas9 fusion molecule, having the HNH activity inactivated, *e.g.*, a Cas9 molecule having a mutation at N863, *e.g.*, the N863A mutation.

30

VI. Target Cells

Cas9 fusion molecules and gRNA molecules, *e.g.*, a Cas9 (fusion) molecule/gRNA molecule complex, can be used to manipulate a cell, *e.g.*, to edit a target nucleic acid, in a wide variety of cells. Additional details on types of cells that can be manipulated may be found in the section entitled “VIIA. TARGETS: CELLS” of PCT Application WO 2015/048577, the entire contents of which are expressly incorporated herein by reference.

In certain embodiments, a cell is manipulated by editing (*e.g.*, introducing a mutation in) a target gene as described herein. In one embodiment, a cell, or a population of cells, is manipulated by editing one or more non-coding sequences, *e.g.*, an alteration in an intron or in a 5' or 3' non-translated or non-transcribed region. In one embodiment, a cell, or a population of cells, is manipulated by editing the sequence of a control element, *e.g.*, a promoter, enhancer, or a cis-acting or trans-acting control element. In one embodiment, a cell, or a population of cells, is manipulated by editing one or more coding sequences, *e.g.*, an alteration in an exon. In some embodiments, a cell, or a population of cells, is manipulated *in vitro*. In other embodiments, a cell, or a population of cells, is manipulated *ex vivo*. In some embodiments, a cell, or a population of cells, is manipulated *in vivo*. In some embodiments, the expression of one or more target genes (*e.g.*, one or more target genes described herein) is modulated, *e.g.*, *in vivo*. In other embodiments, the expression of one or more target genes (*e.g.*, one or more target genes described herein) is modulated, *e.g.*, *ex vivo*. In other embodiments, the expression of one or more target genes (*e.g.*, one or more target genes described herein) is modulated, *e.g.*, *in vitro*.

In one embodiment, a cell, or a population of cells, is manipulated by editing (*e.g.*, inducing a mutation in) the target gene, *e.g.*, as described herein. In one embodiment, the expression of the target gene is modulated, *e.g.*, *in vivo*. In another embodiment, the expression of the target gene is modulated, *e.g.*, *ex vivo*.

The Cas9 (or Cas9 fusion molecule) and gRNA molecules described herein can be delivered to a target cell. In certain embodiments, the target cell is an erythroid cell, *e.g.*, an erythroblast. In certain embodiments, erythroid cells are preferentially targeted, *e.g.*, at least about 90%, 95%, 96%, 97%, 98%, 99%, or 100% of the targeted cells are erythroid cells. For example, in the case of *in vivo* delivery, erythroid cells are preferentially targeted, and if cells are treated *ex vivo* and returned to the subject, erythroid cells are preferentially modified. In certain embodiments, the target cell is a circulating blood cell, *e.g.*, a reticulocyte, megakaryocyte erythroid progenitor (MEP) cell, myeloid progenitor cell (CMP/GMP), lymphoid progenitor (LP) cell, hematopoietic stem/progenitor cell (HSC), or endothelial cell

(EC). In certain embodiments, the target cell is a bone marrow cell (*e.g.*, a reticulocyte, an erythroid cell (*e.g.*, erythroblast), an MEP cell, myeloid progenitor cell (CMP/GMP), LP cell, erythroid progenitor (EP) cell, HSC, multipotent progenitor (MPP) cell, endothelial cell (EC), hemogenic endothelial (HE) cell, or mesenchymal stem cell). In certain embodiments, the target cell is a myeloid progenitor cell (*e.g.*, a common myeloid progenitor (CMP) cell or granulocyte macrophage progenitor (GMP) cell). In certain embodiments, the target cell is a lymphoid progenitor cell, *e.g.*, a common lymphoid progenitor (CLP) cell. In certain embodiments, the target cell is an erythroid progenitor cell (*e.g.*, an MEP cell). In certain embodiments, the target cell is a hematopoietic stem/progenitor cell (*e.g.*, a long term HSC (LT-HSC), short term HSC (ST-HSC), MPP cell, or lineage restricted progenitor (LRP) cell). In certain embodiments, the target cell is a CD34⁺ cell, CD34⁺CD90⁺ cell, CD34⁺CD38⁻ cell, CD34⁺CD90⁺CD49f⁺CD38⁻CD45RA⁻ cell, CD105⁺ cell, CD31⁺, or CD133⁺ cell, or a CD34⁺CD90⁺ CD133⁺ cell. In certain embodiments, the target cell is an umbilical cord blood CD34⁺ HSPC, umbilical cord venous endothelial cell, umbilical cord arterial endothelial cell, amniotic fluid CD34⁺ cell, amniotic fluid endothelial cell, placental endothelial cell, or placental hematopoietic CD34⁺ cell. In certain embodiments, the target cell is a mobilized peripheral blood hematopoietic CD34⁺ cell (after the patient is treated with a mobilization agent, *e.g.*, G-CSF or Plerixafor). In certain embodiments, the target cell is a peripheral blood endothelial cell.

In certain embodiments, a target cell is manipulated *ex vivo* by editing (*e.g.*, inducing a mutation in) the target gene and/or modulating the expression of the target gene, then the target cell is administered to the subject. Sources of target cells for *ex vivo* manipulation may include, for example, the subject's blood, cord blood, or marrow. Other sources of target cells for *ex vivo* manipulation may include, for example, heterologous donor blood, cord blood, or bone marrow.

In certain embodiments, an erythrocyte is removed from a subject, manipulated *ex vivo* as described above, and the erythrocyte is returned to the subject. In other embodiments, a hematopoietic stem cell is removed from a subject, manipulated *ex vivo* as described above, and the hematopoietic stem cell is returned to the subject. In certain embodiments, an erythroid progenitor cell is removed from a subject, manipulated *ex vivo* as described above, and the erythroid progenitor cell is returned to the subject. In certain embodiments, an myeloid progenitor cell is removed from a subject, manipulated *ex vivo* as described above, and the myeloid progenitor cell is returned to the subject. In certain embodiments, a hematopoietic stem/progenitor cell (HSC) is removed from a subject, manipulated *ex vivo* as

described above, and returned to the subject. In certain embodiments, a CD34⁺ HSC is removed from a subject, manipulated *ex vivo* as described above, and returned to the subject.

In certain embodiments wherein modified HSCs generated *ex vivo* are administered to a subject without myeloblastic pre-conditioning. In other embodiments, the modified HSCs are administered after mild myeloblastic conditioning such that, followed engraftment, some of the hematopoietic cells are derived from the modified HSCs. In still other embodiments, the modified HSCs are administered after full myeloblastic conditioning such that, following engraftment, 100% of the hematopoietic cells are derived from the modified HSCs.

A suitable cell can also include a stem cell such as, by way of example, an embryonic stem cell, induced pluripotent stem cell, hematopoietic stem cell, or hemogenic endothelial (HE) cell (precursor to both hematopoietic stem cells and endothelial cells). In certain embodiments, the cell is an induced pluripotent stem (iPS) cell or a cell derived from an iPS cell, *e.g.*, an iPS cell generated from the subject, modified using methods disclosed herein and differentiated into a clinically relevant cell such as *e.g.*, an erythrocyte. In one embodiment, AAV is used to transduce the target cells, *e.g.*, the target cells described herein.

Cells produced by the methods described herein may be used immediately. Alternatively, the cells may be frozen (*e.g.*, in liquid nitrogen) and stored for later use. The cells will usually be frozen in 10% dimethylsulfoxide (DMSO), 50% serum, 40% buffered medium, or some other such solution as is commonly used in the art to preserve cells at such freezing temperature and thawed in such a manner as commonly known in the art for thawing frozen cultured cells. Cells may also be thermostabilized for prolonged storage at 4°C.

Delivery, Formulations and Routes of Administration

The components, *e.g.*, a Cas9 fusion molecule and at least one gRNA molecule (*e.g.*, a Cas9 fusion molecule/gRNA molecule complex), can be delivered, formulated, or administered, in a variety of forms, see, *e.g.*, **Tables 5-6**. In certain embodiments, the sequence(s) encoding the two or more (*e.g.*, 2, 3, 4, or more) different gRNA molecules are present on the same nucleic acid molecule, *e.g.*, an AAV vector. When a gRNA component is delivered encoded in DNA, the DNA will typically include a control region, *e.g.*, comprising a promoter, to effect expression. In one embodiment, the promoter is a constitutive promoter. In another embodiment, the promoter is a tissue specific promoter. Useful promoters for gRNAs include T7, H1, EF-1a, U6, U1, and tRNA promoters. Promoters with similar or dissimilar strengths can be selected to tune the expression of components. Sequences encoding the Cas9 molecule of a Cas9 fusion molecule can comprise

a nuclear localization signal (NLS), *e.g.*, an SV40 NLS. In one embodiment, the sequence encoding the Cas9 molecule of a Cas9 fusion molecule comprises at least two nuclear localization signals. In one embodiment a promoter for the Cas9 molecule of a Cas9 fusion molecule or a gRNA molecule can be, independently, inducible, tissue specific, or cell

5 specific.

Table 5 provides examples of how the components can be formulated, delivered, or administered.

Table 5

Elements			
Cas9 Fusion Molecule(s)	gRNA Molecule(s)	Optional Donor Template Nucleic Acid	Comments
Protein	DNA	DNA	In this embodiment, a Cas9 fusion molecule, typically an eaCas9 fusion molecule, is provided as a protein, covalently or non-covalently lined to the template nucleic acid, and a gRNA molecule is transcribed from DNA.
Protein	RNA	DNA	In this embodiment, a Cas9 fusion molecule, typically an eaCas9 fusion molecule, is provided as a protein, covalently or non-covalently lined to the template nucleic acid, and a gRNA molecule is provided as transcribed or synthesized RNA

10 **Table 6** summarizes various delivery methods for the components of a Cas system, *e.g.*, the Cas9 molecule component and the gRNA molecule component, as described herein.

Table 6

Delivery Vector/Mode		Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Physical (<i>e.g.</i>, electroporation, particle gun, calcium phosphate transfection, cell compression or squeezing)		YES	Transient	NO	Nucleic Acids and Proteins
<i>Viral</i>	Retrovirus	NO	Stable	YES	RNA

	Lentivirus	YES	Stable	YES/NO with modifications	RNA
	Adenovirus	YES	Transient	NO	DNA
	Adeno-Associated Virus (AAV)	YES	Stable	NO	DNA
	Vaccinia Virus	YES	Very Transient	NO	DNA
	Herpes Simplex Virus	YES	Stable	NO	DNA
Non-Viral	Cationic Liposomes	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
	Polymeric Nanoparticles	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
Biological Non-Viral Delivery Vehicles	Attenuated Bacteria	YES	Transient	NO	Nucleic Acids
	Engineered Bacteriophages	YES	Transient	NO	Nucleic Acids
	Mammalian Virus-like Particles	YES	Transient	NO	Nucleic Acids
	Biological liposomes: Erythrocyte Ghosts and Exosomes	YES	Transient	NO	Nucleic Acids

DNA-based Delivery of one or more gRNA molecules

Nucleic acids encoding gRNA molecules can be administered to subjects or delivered into cells by art-known methods or as described herein. For example, gRNA-encoding DNA
5 can be delivered by, *e.g.*, vectors (*e.g.*, viral or non-viral vectors), non-vector based methods (*e.g.*, using naked DNA or DNA complexes), or a combination thereof.

Nucleic acids encoding gRNA molecules can be conjugated to molecules (*e.g.*, N-acetylgalactosamine) promoting uptake by the target cells (*e.g.*, erythrocytes, HSCs).

In some embodiments, the gRNA-encoding DNA is delivered by a vector (*e.g.*, viral
10 vector/virus or plasmid).

Vectors can comprise a sequence that encodes a gRNA molecule. One or more regulatory/control elements, *e.g.*, promoters, enhancers, introns, polyadenylation signals, Kozak consensus sequences, internal ribosome entry sites (IRES), can be included in the vectors. In some embodiments, the promoter is recognized by RNA polymerase II (*e.g.*, a

CMV promoter). In other embodiments, the promoter is recognized by RNA polymerase III (*e.g.*, a U6 promoter). In some embodiments, the promoter is a regulated promoter (*e.g.*, inducible promoter). In other embodiment, the promoter is a constitutive promoter. In some embodiments, the promoter is a tissue specific promoter. In other embodiments, the promoter is a viral promoter. In some embodiments, the promoter is a non-viral promoter.

In some embodiments, the vector is a viral vector (*e.g.*, for generation of recombinant viruses). In some embodiments, the virus is a DNA virus (*e.g.*, dsDNA or ssDNA virus). In other embodiments, the virus is an RNA virus (*e.g.*, an ssRNA virus). In some embodiments, the virus infects dividing cells. In other embodiments, the virus infects non-dividing cells.

Exemplary viral vectors/viruses include, *e.g.*, retroviruses, lentiviruses, adenovirus, adeno-associated virus (AAV), vaccinia viruses, poxviruses, and herpes simplex viruses.

In some embodiments, the virus infects both dividing and non-dividing cells. In some embodiments, the virus can integrate into the host genome. In some embodiments, the virus is engineered to have reduced immunity, *e.g.*, in human. In some embodiments, the virus is replication-competent. In other embodiments, the virus is replication-defective, *e.g.*, having one or more coding regions for the genes necessary for additional rounds of virion replication and/or packaging replaced with other genes or deleted. In some embodiments, the virus causes transient expression of the gRNA molecule. In other embodiments, the virus causes long-lasting, *e.g.*, at least 1 week, 2 weeks, 1 month, 2 months, 3 months, 6 months, 9 months, 1 year, 2 years, or permanent expression, of the gRNA molecule. The packaging capacity of the viruses may vary, *e.g.*, from at least about 4 kb to at least about 30 kb, *e.g.*, at least about 5 kb, 10 kb, 15 kb, 20 kb, 25 kb, 30 kb, 35 kb, 40 kb, 45 kb, or 50 kb.

In one embodiment, the viral vector recognizes a specific cell type or tissue. For example, the viral vector can be pseudotyped with a different/alternative viral envelope glycoprotein; engineered with a cell type-specific receptor (*e.g.*, genetic modification(s) of one or more viral envelope glycoproteins to incorporate a targeting ligand such as a peptide ligand, a single chain antibody, or a growth factor); and/or engineered to have a molecular bridge with dual specificities with one end recognizing a viral glycoprotein and the other end recognizing a moiety of the target cell surface (*e.g.*, a ligand-receptor, monoclonal antibody, avidin-biotin and chemical conjugation).

In some embodiments, the gRNA-encoding nucleic acid sequence is delivered by a recombinant retrovirus. In some embodiments, the retrovirus (*e.g.*, Moloney murine leukemia virus) comprises a reverse transcriptase, *e.g.*, that allows integration into the host genome. In some embodiments, the retrovirus is replication-competent. In other

embodiments, the retrovirus is replication-defective, *e.g.*, having one of more coding regions for the genes necessary for additional rounds of virion replication and packaging replaced with other genes, or deleted.

5 In one embodiment, the gRNA-encoding nucleic acid sequence is delivered by a recombinant lentivirus. For example, the lentivirus is replication-defective, *e.g.*, does not comprise one or more genes required for viral replication.

In some embodiments, the gRNA-encoding nucleic acid sequence is delivered by a recombinant adenovirus. In some embodiments, the adenovirus is engineered to have reduced immunity in human.

10 In some embodiments, the gRNA-encoding nucleic acid sequence is delivered by a recombinant AAV. In some embodiments, the AAV does not incorporate its genome into that of a host cell, *e.g.*, a target cell as describe herein. In some embodiments, the AAV can incorporate its genome into that of a host cell. In some embodiments, the AAV is a self-complementary adeno-associated virus (scAAV), *e.g.*, a scAAV that packages both strands
15 which anneal together to form double stranded DNA.

In one embodiment, an AAV capsid that can be used in the methods described herein is a capsid sequence from serotype AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV.rh8, AAV.rh10, AAV.rh32/33, AAV.rh43, AAV.rh64R1, or AAV7m8.

20 In one embodiment, the gRNA-encoding DNA is delivered in a re-engineered AAV capsid, *e.g.*, with 50% or greater, *e.g.*, 60% or greater, 70% or greater, 80% or greater, 90% or greater, or 95% or greater, sequence homology with a capsid sequence from serotypes AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV.rh8, AAV.rh10, AAV.rh32/33, AAV.rh43, or AAV.rh64R1.

25 In one embodiment, the gRNA-encoding DNA is delivered by a chimeric AAV capsid. Exemplary chimeric AAV capsids include, but are not limited to, AAV9i1, AAV2i8, AAV-DJ, AAV2G9, AAV2i8G9, or AAV8G9.

In one embodiment, the AAV is a self-complementary adeno-associated virus (scAAV), *e.g.*, a scAAV that packages both strands which anneal together to form double stranded DNA.

30 In some embodiments, the gRNA-encoding DNA is delivered by a hybrid virus, *e.g.*, a hybrid of one or more of the viruses described herein. In one embodiment, the hybrid virus is hybrid of an AAV (*e.g.*, of any AAV serotype), with a Bocavirus, B19 virus, porcine AAV, goose AAV, feline AAV, canine AAV, or MVM.

A packaging cell is used to form a virus particle that is capable of infecting a target cell. Exemplary packaging cells include 293 cells, which can package adenovirus, and ψ 2 or PA317 cells, which can package retrovirus. A viral vector used in gene therapy is usually generated by a producer cell line that packages a nucleic acid vector into a viral particle. The vector typically contains the minimal viral sequences required for packaging and subsequent integration into a host or target cell (if applicable). For example, an AAV vector used in gene therapy typically only possesses inverted terminal repeat (ITR) sequences from the AAV genome which are required for packaging and gene expression in the host or target cell. The missing viral functions can be supplied in *trans* by the packaging cell line and/or plasmid containing E2A, E4, and VA genes from adenovirus, and plasmid encoding *Rep* and *Cap* genes from AAV, as described in "Triple Transfection Protocol." Henceforth, the viral DNA is packaged in a cell line, which contains a helper plasmid encoding the other AAV genes, namely *rep* and *cap*, but lacking ITR sequences. In certain embodiments, the viral DNA is packaged in a producer cell line, which contains E1A and/or E1B genes from adenovirus. The cell line is also infected with adenovirus as a helper. The helper virus (*e.g.*, adenovirus or HSV) or helper plasmid promotes replication of the AAV vector and expression of AAV genes from the helper plasmid with ITRs. The helper plasmid is not packaged in significant amounts due to a lack of ITR sequences. Contamination with adenovirus can be reduced by, *e.g.*, heat treatment to which adenovirus is more sensitive than AAV.

In certain embodiments, the viral vector is capable of cell type and/or tissue type recognition. For example, the viral vector can be pseudotyped with a different/alternative viral envelope glycoprotein; engineered with a cell type-specific receptor (*e.g.*, genetic modification of the viral envelope glycoproteins to incorporate targeting ligands such as a peptide ligand, single chain antibody, or growth factor); and/or engineered to have a molecular bridge with dual specificities with one end recognizing a viral glycoprotein and the other end recognizing a moiety of the target cell surface (*e.g.*, ligand-receptor, monoclonal antibody, avidin-biotin and chemical conjugation).

In certain embodiments, the viral vector achieves cell type specific expression. For example, a tissue-specific promoter can be constructed to restrict expression of the transgene (gRNA) to only the target cell. The specificity of the vector can also be mediated by microRNA-dependent control of transgene expression. In one embodiment, the viral vector has increased efficiency of fusion of the viral vector and a target cell membrane. For example, a fusion protein such as fusion-competent hemagglutinin (HA) can be incorporated to increase viral uptake into cells. In one embodiment, the viral vector has the ability of nuclear

localization. For example, a virus that requires the breakdown of the nuclear envelope (during cell division) and therefore will not infect a non-dividing cell can be altered to incorporate a nuclear localization peptide in the matrix protein of the virus thereby enabling the transduction of non-proliferating cells.

5 In some embodiments, the gRNA-encoding DNA is delivered by a non-vector based method (*e.g.*, using naked DNA or DNA complexes). For example, the DNA can be delivered, *e.g.*, by organically modified silica or silicate (Ormosil), electroporation, transient cell compression or squeezing (see, *e.g.*, Lee 2012), gene gun, sonoporation, magnetofection, lipid-mediated transfection, dendrimers, inorganic nanoparticles, calcium phosphates, or a
10 combination thereof.

In one embodiment, delivery via electroporation comprises mixing the cells with the gRNA-encoding DNA in a cartridge, chamber or cuvette and applying one or more electrical impulses of defined duration and amplitude. In one embodiment, delivery via electroporation is performed using a system in which cells are mixed with the gRNA-encoding DNA in a
15 vessel connected to a device (*e.g.*, a pump) which feeds the mixture into a cartridge, chamber or cuvette wherein one or more electrical impulses of defined duration and amplitude are applied, after which the cells are delivered to a second vessel.

In some embodiments, the gRNA-encoding DNA is delivered by a combination of a vector and a non-vector based method. For example, virosomes combine liposomes with an
20 inactivated virus (*e.g.*, HIV or influenza virus), which can result in more efficient gene transfer, *e.g.*, in respiratory epithelial cells than either viral or liposomal methods alone.

As described above, a nucleic acid may comprise a sequence encoding a gRNA molecule comprising a targeting domain that is complementary with a desired target domain. In one embodiment, the nucleic acid molecule is an AAV vector. Exemplary AAV vectors
25 that may be used in any of the described compositions and methods include an AAV2 vector, a modified AAV2 vector, an AAV3 vector, a modified AAV3 vector, an AAV6 vector, a modified AAV6 vector, an AAV8 vector and an AAV9 vector. In yet another embodiment, the nucleic acid may further comprise a sequence that encodes a second, third and/or fourth gRNA molecule as described herein. Each of the sequence encoding a gRNA molecule
30 comprising a targeting domain that is complementary with a desired target domain and the sequence that encodes a second, third and/or fourth gRNA molecule may be present on the same nucleic acid molecule, *e.g.*, the same vector, *e.g.*, the same viral vector, *e.g.*, the same adeno-associated virus (AAV) vector. In one embodiment, the nucleic acid molecule is an AAV vector.

In another embodiment, the sequence encoding a gRNA molecule comprising a targeting domain that is complementary with a desired target domain and the sequence that encodes a second, third and/or fourth gRNA molecule are on different vectors. For example, the sequence encoding a gRNA molecule comprising a targeting domain that is

5 complementary with a desired target domain may be present on a first nucleic acid molecule, *e.g.*, a first vector, *e.g.*, a first viral vector, *e.g.*, a first AAV vector; and the sequence that encodes a second, third and/or fourth gRNA molecule may be present on a second nucleic acid molecule, *e.g.*, a second vector, *e.g.*, a second vector, *e.g.*, a second AAV vector. In one embodiment, the first and second nucleic acid molecules are AAV vectors.

10 In another embodiment, when a third and/or fourth gRNA molecule are present, each of the sequence encoding a gRNA molecule comprising a targeting domain that is complementary with a desired target domain and the sequence that encodes a second, third and/or fourth gRNA molecule may be present on the same nucleic acid molecule, *e.g.*, the same vector, *e.g.*, the same viral vector, *e.g.*, an AAV vector. In one embodiment, the nucleic acid molecule is an AAV vector. In an alternate embodiment, each of the sequence encoding a gRNA molecule comprising a targeting domain that is complementary with a desired target domain and the sequence that encodes a second, third and/or fourth gRNA molecule may be present on the different nucleic acid molecules, *e.g.*, different vectors, *e.g.*, the different viral vectors, *e.g.*, different AAV vectors. In further embodiments, each of the sequence encoding a gRNA molecule comprising a targeting domain that is complementary with a desired target domain and the sequence that encodes a second, third and/or fourth gRNA molecule may be present on more than one nucleic acid molecule, but fewer than five nucleic acid molecules, *e.g.*, AAV vectors.

25 The nucleic acids described herein may comprise a promoter operably linked to the sequence that encodes the gRNA molecule of the sequence encoding a gRNA molecule comprising a targeting domain that is complementary with a desired target domain, *e.g.*, a promoter described herein. The nucleic acid may further comprise a second promoter operably linked to the sequence that encodes the second, third and/or fourth gRNA molecule, *e.g.*, a promoter described herein. The promoter and second promoter differ from one another. In one embodiment, the promoter and second promoter are the same.

30 In certain embodiments, the delivery vehicle is a non-viral vector, and in certain of these embodiments the non-viral vector is an inorganic nanoparticle. Exemplary inorganic nanoparticles include, *e.g.*, magnetic nanoparticles (*e.g.*, Fe₃MnO₂) or silica. The outer surface of the nanoparticle can be conjugated with a positively charged polymer (*e.g.*,

polyethylenimine, polylysine, polyserine) which allows for attachment (*e.g.*, conjugation or entrapment) of payload. In one embodiment, the non-viral vector is an organic nanoparticle (*e.g.*, entrapment of the payload inside the nanoparticle). Exemplary organic nanoparticles include, *e.g.*, SNALP liposomes that contain cationic lipids together with neutral helper lipids which are coated with polyethylene glycol (PEG) and protamine and nucleic acid complex coated with lipid coating.

Exemplary lipids for gene transfer are shown below in **Table 7**.

Table 7. Lipids Used for Gene Transfer

Lipid	Abbreviation	Feature
1,2-Dioleoyl-sn-glycero-3-phosphatidylcholine	DOPC	Helper
1,2-Dioleoyl-sn-glycero-3-phosphatidylethanolamine	DOPE	Helper
Cholesterol		Helper
N-[1-(2,3-Dioleoyloxy)propyl]N,N,N-trimethylammonium chloride	DOTMA	Cationic
1,2-Dioleoyloxy-3-trimethylammonium-propane	DOTAP	Cationic
Diocetadecylamidoglycylspermine	DOGS	Cationic
N-(3-Aminopropyl)-N,N-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide	GAP-DLRIE	Cationic
Cetyltrimethylammonium bromide	CTAB	Cationic
6-Lauroxyhexyl ornithinate	LHON	Cationic
1-(2,3-Dioleoyloxypropyl)-2,4,6-trimethylpyridinium	2Oc	Cationic
2,3-Dioleoyloxy-N-[2(spermincarboxamido-ethyl)-N,N-dimethyl-1-propanaminium trifluoroacetate	DOSPA	Cationic
1,2-Dioleoyl-3-trimethylammonium-propane	DOPA	Cationic
N-(2-Hydroxyethyl)-N,N-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide	MDRIE	Cationic
Dimyristooxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3β-[N-(N',N'-Dimethylaminoethane)-carbonyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Dioleoyl-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic
Dimethyloctadecylammonium bromide	DDAB	Cationic
Diocetadecylamidoglycylspermidin	DSL	Cationic
rac-[(2,3-Dioctadecyloxypropyl)(2-hydroxyethyl)]-dimethylammonium chloride	CLIP-1	Cationic
rac-[2(2,3-Dihexadecyloxypropyl-oxymethyloxy)ethyl]trimethylammonium bromide	CLIP-6	Cationic
Ethyl dimyristoylphosphatidylcholine	EDMPC	Cationic
1,2-Distearoyloxy-N,N-dimethyl-3-aminopropane	DSDMA	Cationic
1,2-Dimyristoyl-trimethylammonium propane	DMTAP	Cationic
O,O'-Dimyristyl-N-lysyl aspartate	DMKE	Cationic
1,2-Distearoyl-sn-glycero-3-ethylphosphocholine	DSEPC	Cationic
N-Palmitoyl D-erythro-sphingosyl carbamoyl-spermine	CCS	Cationic
N-t-Butyl-N0-tetradecyl-3-tetradecylaminopropionamidine	diC14-amidine	Cationic

Lipid	Abbreviation	Feature
Octadecenolyoxy[ethyl-2-heptadecenyl-3 hydroxyethyl] imidazolinium chloride	DOTIM	Cationic
N1-Cholesteryloxycarbonyl-3,7-diazanonane-1,9-diamine	CDAN	Cationic
2-(3-[Bis(3-amino-propyl)-amino]propylamino)-N-ditetradecylcarbamoyle-ethyl-acetamide	RPR209120	Cationic
1,2-dilinoleyloxy-3- dimethylaminopropane	DLinDMA	Cationic
2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]- dioxolane	DLin-KC2-DMA	Cationic
dilinoleyl- methyl-4-dimethylaminobutyrate	DLin-MC3-DMA	Cationic

Exemplary polymers for gene transfer are shown below in **Table 8**.

Table 8. Polymers Used for Gene Transfer

Polymer	Abbreviation
Poly(ethylene)glycol	PEG
Polyethylenimine	PEI
Dithiobis(succinimidylpropionate)	DSP
Dimethyl-3,3'-dithiobispropionimide	DTBP
Poly(ethylene imine) biscarbamate	PEIC
Poly(L-lysine)	PLL
Histidine modified PLL	
Poly(<i>N</i> -vinylpyrrolidone)	PVP
Poly(propylenimine)	PPI
Poly(amidoamine)	PAMAM
Poly(amido ethylenimine)	SS-PAEI
Triethylenetetramine	TETA
Poly(β -aminoester)	
Poly(4-hydroxy-L-proline ester)	PHP
Poly(allylamine)	
Poly(α -[4-aminobutyl]-L-glycolic acid)	PAGA
Poly(D,L-lactic-co-glycolic acid)	PLGA
Poly(<i>N</i> -ethyl-4-vinylpyridinium bromide)	
Poly(phosphazene)s	PPZ
Poly(phosphoester)s	PPE
Poly(phosphoramidate)s	PPA
Poly(<i>N</i> -2-hydroxypropylmethacrylamide)	pHPMA
Poly (2-(dimethylamino)ethyl methacrylate)	pDMAEMA
Poly(2-aminoethyl propylene phosphate)	PPE-EA
Chitosan	
Galactosylated chitosan	
<i>N</i> -Dodacylated chitosan	
Histone	
Collagen	
Dextran-spermine	D-SPM

In one embodiment, the vehicle has targeting modifications to increase target cell uptake of nanoparticles and liposomes, *e.g.*, cell specific antigens, monoclonal antibodies, single chain antibodies, aptamers, polymers, sugars (*e.g.*, N-acetylgalactosamine (GalNAc)), and cell penetrating peptides. In one embodiment, the vehicle uses fusogenic and endosome-destabilizing peptides/polymers. In one embodiment, the vehicle undergoes acid-triggered conformational changes (*e.g.*, to accelerate endosomal escape of the cargo). In one embodiment, a stimuli-cleavable polymer is used, *e.g.*, for release in a cellular compartment. For example, disulfide-based cationic polymers that are cleaved in the reducing cellular environment can be used.

In one embodiment, the delivery vehicle is a biological non-viral delivery vehicle. In one embodiment, the vehicle is an attenuated bacterium (*e.g.*, naturally or artificially engineered to be invasive but attenuated to prevent pathogenesis and expressing the transgene (*e.g.*, *Listeria monocytogenes*, certain *Salmonella strains*, *Bifidobacterium longum*, and modified *Escherichia coli*), bacteria having nutritional and tissue-specific tropism to target specific tissues, bacteria having modified surface proteins to alter target tissue specificity). In one embodiment, the vehicle is a genetically modified bacteriophage (*e.g.*, engineered phages having large packaging capacity, less immunogenic, containing mammalian plasmid maintenance sequences and having incorporated targeting ligands). In one embodiment, the vehicle is a mammalian virus-like particle. For example, modified viral particles can be generated (*e.g.*, by purification of the “empty” particles followed by *ex vivo* assembly of the virus with the desired cargo). The vehicle can also be engineered to incorporate targeting ligands to alter target tissue specificity. In one embodiment, the vehicle is a biological liposome. For example, the biological liposome is a phospholipid-based particle derived from human cells (*e.g.*, erythrocyte ghosts, which are red blood cells broken down into spherical structures derived from the subject (*e.g.*, tissue targeting can be achieved by attachment of various tissue or cell-specific ligands), or secretory exosomes –subject (*i.e.*, patient) derived membrane-bound nanovesicle (30 -100 nm) of endocytic origin (*e.g.*, can be produced from various cell types and can therefore be taken up by cells without the need of for targeting ligands).

In one embodiment, one or more nucleic acid molecules (*e.g.*, DNA molecules) other than the components of a Cas system, *e.g.*, the Cas9 fusion molecule component and/or the gRNA molecule component described herein, are delivered. In one embodiment, the nucleic acid molecule is delivered at the same time as one or more of the gRNA molecule(s) are delivered. In one embodiment, the nucleic acid molecule is delivered before or after (*e.g.*,

less than about 30 minutes, 1 hour, 2 hours, 3 hours, 6 hours, 9 hours, 12 hours, 1 day, 2 days, 3 days, 1 week, 2 weeks, or 4 weeks) one or more of the gRNA molecule(s) are delivered. In one embodiment, the nucleic acid molecule is delivered by a different means than one or more of the gRNA molecule(s) are delivered. The nucleic acid molecule can be delivered by any of the delivery methods described herein. For example, the nucleic acid molecule can be delivered by a viral vector, *e.g.*, an integration-deficient lentivirus, and the gRNA molecule component can be delivered by electroporation, *e.g.*, such that the toxicity caused by nucleic acids (*e.g.*, DNAs) can be reduced. In one embodiment, the nucleic acid molecule encodes a therapeutic protein, *e.g.*, a protein described herein. In one embodiment, the nucleic acid molecule encodes an RNA molecule, *e.g.*, an RNA molecule described herein.

Delivery of RNA encoding a gRNA molecule

RNA encoding gRNA molecules can be delivered into cells, *e.g.*, target cells described herein, by art-known methods or as described herein. For example, gRNA-encoding RNA can be delivered, *e.g.*, by microinjection, electroporation, transient cell compression or squeezing (see, *e.g.*, Lee 2012), lipid-mediated transfection, peptide-mediated delivery, or a combination thereof. gRNA-encoding RNA can be conjugated to molecules promoting uptake by the target cells (*e.g.*, target cells described herein).

In one embodiment, delivery via electroporation comprises mixing the cells with the RNA encoding gRNA molecules in a cartridge, chamber or cuvette and applying one or more electrical impulses of defined duration and amplitude. In one embodiment, delivery via electroporation is performed using a system in which cells are mixed with the RNA encoding gRNA molecules in a vessel connected to a device (*e.g.*, a pump) which feeds the mixture into a cartridge, chamber or cuvette wherein one or more electrical impulses of defined duration and amplitude are applied, after which the cells are delivered to a second vessel. gRNA-encoding RNA can be conjugated to molecules to promote uptake by the target cells (*e.g.*, target cells described herein).

Delivery of Cas9 Polypeptides and Cas9 Fusion Molecules

Cas9 molecules and Cas9 fusion molecules can be delivered into cells by art-known methods or as described herein. For example, protein molecules can be delivered, *e.g.*, by microinjection, electroporation, transient cell compression or squeezing (see, *e.g.*, Lee 2012), lipid-mediated transfection, peptide-mediated delivery, or a combination thereof. Delivery can be accompanied by DNA encoding a gRNA or by a gRNA. Cas9 proteins and Cas9

fusion molecules can be conjugated to molecules promoting uptake by the target cells (*e.g.*, target cells described herein).

In one embodiment, delivery via electroporation comprises mixing the cells with the Cas9 fusion molecules and/or gRNA molecules in a cartridge, chamber or cuvette and
5 applying one or more electrical impulses of defined duration and amplitude. In one embodiment, delivery via electroporation is performed using a system in which cells are mixed with the Cas9 fusion molecules and/or gRNA molecules in a vessel connected to a device (*e.g.*, a pump) which feeds the mixture into a cartridge, chamber or cuvette wherein one or more electrical impulses of defined duration and amplitude are applied, after which the
10 cells are delivered to a second vessel. gRNA-encoding RNA can be conjugated to molecules to promote uptake by the target cells (*e.g.*, target cells described herein).

Route of Administration

Systemic modes of administration include oral and parenteral routes. Parenteral routes
15 include, by way of example, intravenous, intramarrow, intrarterial, intramuscular, intradermal, subcutaneous, intranasal, inhalation, and intraperitoneal routes. Components administered systemically may be modified or formulated to target, *e.g.*, HSCs, hematopoietic stem/progenitor cells, or erythroid progenitors or precursor cells.

Local modes of administration include, by way of example, intramarrow injection into
20 the trabecular bone or intrafemoral injection into the marrow space, and infusion into the portal vein. In one embodiment, significantly smaller amounts of the components (compared with systemic approaches) may exert an effect when administered locally (for example, directly into the bone marrow) compared to when administered systemically (for example, intravenously). Local modes of administration can reduce or eliminate the incidence of
25 potentially toxic side effects that may occur when therapeutically effective amounts of a component are administered systemically.

Administration may be provided as a periodic bolus (*e.g.*, intravenously) or as continuous infusion from an internal reservoir or from an external reservoir (for example, from an intravenous bag or implantable pump). Components may be administered locally, for
30 example, by continuous release from a sustained release drug delivery device.

In addition, components may be formulated to permit release over a prolonged period of time. A release system can include a matrix of a biodegradable material or a material which releases the incorporated components by diffusion. The components can be homogeneously or heterogeneously distributed within the release system. A variety of release

systems may be useful, however, the choice of the appropriate system will depend upon rate of release required by a particular application. Both non-degradable and degradable release systems can be used. Suitable release systems include polymers and polymeric matrices, non-polymeric matrices, or inorganic and organic excipients and diluents such as, but not limited to, calcium carbonate and sugar (for example, trehalose). Release systems may be natural or synthetic. However, synthetic release systems are preferred because generally they are more reliable, more reproducible and produce more defined release profiles. The release system material can be selected so that components having different molecular weights are released by diffusion through or degradation of the material.

Representative synthetic, biodegradable polymers include, for example: polyamides such as poly(amino acids) and poly(peptides); polyesters such as poly(lactic acid), poly(glycolic acid), poly(lactic-co-glycolic acid), and poly(caprolactone); poly(anhydrides); polyorthoesters; polycarbonates; and chemical derivatives thereof (substitutions, additions of chemical groups, for example, alkyl, alkylene, hydroxylations, oxidations, and other modifications routinely made by those skilled in the art), copolymers and mixtures thereof. Representative synthetic, non-degradable polymers include, for example: polyethers such as poly(ethylene oxide), poly(ethylene glycol), and poly(tetramethylene oxide); vinyl polymers- polyacrylates and polymethacrylates such as methyl, ethyl, other alkyl, hydroxyethyl methacrylate, acrylic and methacrylic acids, and others such as poly(vinyl alcohol), poly(vinyl pyrrolidone), and poly(vinyl acetate); poly(urethanes); cellulose and its derivatives such as alkyl, hydroxyalkyl, ethers, esters, nitrocellulose, and various cellulose acetates; polysiloxanes; and any chemical derivatives thereof (substitutions, additions of chemical groups, for example, alkyl, alkylene, hydroxylations, oxidations, and other modifications routinely made by those skilled in the art), copolymers and mixtures thereof.

Poly(lactide-co-glycolide) microsphere can also be used for injection. Typically the microspheres are composed of a polymer of lactic acid and glycolic acid, which are structured to form hollow spheres. The spheres can be approximately 15-30 microns in diameter and can be loaded with components described herein.

Bi-Modal or Differential Delivery of Components

Separate delivery of the components of a Cas system, *e.g.*, the Cas9 fusion molecule component and the gRNA molecule component, and more particularly, delivery of the components by differing modes, can enhance performance, *e.g.*, by improving tissue specificity and safety.

In one embodiment, the Cas9 fusion molecule and the gRNA molecule are delivered by different modes, or as sometimes referred to herein as differential modes. Different or differential modes, as used herein, refer modes of delivery that confer different pharmacodynamic or pharmacokinetic properties on the subject component molecule, *e.g.*, a Cas9 fusion molecule, gRNA molecule, or payload. For example, the modes of delivery can result in different tissue distribution, different half-life, or different temporal distribution, *e.g.*, in a selected compartment, tissue, or organ.

Some modes of delivery, *e.g.*, delivery by a nucleic acid vector that persists in a cell, or in progeny of a cell, *e.g.*, by autonomous replication or insertion into cellular nucleic acid, result in more persistent expression of and presence of a component. Examples include viral, *e.g.*, AAV or lentivirus, delivery.

By way of example, the components, *e.g.*, a Cas9 fusion molecule and a gRNA molecule, can be delivered by modes that differ in terms of resulting half-life or persistence of the delivered component the body, or in a particular compartment, tissue or organ. In one embodiment, a gRNA molecule can be delivered by such modes. The Cas9 fusion molecule component can be delivered by a mode which results in less persistence or less exposure to the body or a particular compartment or tissue or organ.

More generally, in one embodiment, a first mode of delivery is used to deliver a first component and a second mode of delivery is used to deliver a second component. The first mode of delivery confers a first pharmacodynamic or pharmacokinetic property. The first pharmacodynamic property can be, *e.g.*, distribution, persistence, or exposure, of the component, or of a nucleic acid that encodes the component, in the body, a compartment, tissue or organ. The second mode of delivery confers a second pharmacodynamic or pharmacokinetic property. The second pharmacodynamic property can be, *e.g.*, distribution, persistence, or exposure, of the component, or of a nucleic acid that encodes the component, in the body, a compartment, tissue or organ.

In certain embodiments, the first pharmacodynamic or pharmacokinetic property, *e.g.*, distribution, persistence or exposure, is more limited than the second pharmacodynamic or pharmacokinetic property.

In certain embodiments, the first mode of delivery is selected to optimize, *e.g.*, minimize, a pharmacodynamic or pharmacokinetic property, *e.g.*, distribution, persistence or exposure.

In certain embodiments, the second mode of delivery is selected to optimize, *e.g.*, maximize, a pharmacodynamic or pharmacokinetic property, *e.g.*, distribution, persistence or exposure.

5 In one embodiment, the first mode of delivery comprises the use of a relatively persistent element, *e.g.*, a nucleic acid, *e.g.*, a plasmid or viral vector, *e.g.*, an AAV or lentivirus. As such vectors are relatively persistent product transcribed from them would be relatively persistent.

In certain embodiments, the second mode of delivery comprises a relatively transient element, *e.g.*, an RNA or protein.

10 In certain embodiments, the first component comprises gRNA molecule, and the delivery mode is relatively persistent, *e.g.*, the gRNA is transcribed from a plasmid or viral vector, *e.g.*, an AAV or lentivirus. Transcription of these genes would be of little physiological consequence because the genes do not encode for a protein product, and the gRNAs are incapable of acting in isolation. The second component, a Cas9 fusion molecule,
15 is delivered in a transient manner, for example as protein, ensuring that the full Cas9 molecule/gRNA molecule complex is only present and active for a short period of time.

Furthermore, the components can be delivered in different molecular form or with different delivery vectors that complement one another to enhance safety and tissue specificity.

20 Use of differential delivery modes can enhance performance, safety and/or efficacy, *e.g.*, the likelihood of an eventual off-target modification can be reduced. Delivery of immunogenic components, *e.g.*, Cas9 fusion molecules, by less persistent modes can reduce immunogenicity, as peptides from the bacterially-derived Cas enzyme are displayed on the surface of the cell by MHC molecules. A two-part delivery system can alleviate these
25 drawbacks.

Differential delivery modes can be used to deliver components to different, but overlapping target regions. The formation active complex is minimized outside the overlap of the target regions. Thus, in one embodiment, a first component, *e.g.*, a gRNA molecule is delivered by a first delivery mode that results in a first spatial, *e.g.*, tissue, distribution. A
30 second component, *e.g.*, a Cas9 fusion molecule is delivered by a second delivery mode that results in a second spatial, *e.g.*, tissue, distribution. In one embodiment, the first mode comprises a first element selected from a liposome, nanoparticle, *e.g.*, polymeric nanoparticle, and a nucleic acid, *e.g.*, viral vector. The second mode comprises a second element selected from the group. In one embodiment, the first mode of delivery comprises a

first targeting element, *e.g.*, a cell specific receptor or an antibody, and the second mode of delivery does not include that element. In certain embodiments, the second mode of delivery comprises a second targeting element, *e.g.*, a second cell specific receptor or second antibody.

When the Cas9 fusion molecule is delivered in a liposome, or polymeric nanoparticle,
5 there is the potential for delivery to and therapeutic activity in multiple tissues, when it may be desirable to only target a single tissue. A two-part delivery system can resolve this challenge and enhance tissue specificity. If the gRNA molecule and the Cas9 fusion molecule are packaged in separated delivery vehicles with distinct but overlapping tissue tropism, the fully functional complex is only be formed in the tissue that is targeted by both
10 vectors.

Disclosed herein are methods of altering a cell, *e.g.*, altering the structure, *e.g.*, altering the sequence, of a target nucleic acid of a cell, comprising contacting said cell with:
(a) a gRNA molecule that targets the target gene, *e.g.*, a gRNA molecule as described herein;
(b) a Cas9 fusion molecule, *e.g.*, a Cas9 fusion molecule as described herein; and optionally,
15 (c) a second, third and/or fourth gRNA that targets the target gene, *e.g.*, a gRNA molecule; as described herein. In one embodiment, the method comprises contacting said cell with (a) and (b). In one embodiment, the method comprises contacting said cell with (a), (b), and (c). The targeting domain of the gRNA molecule of (a) and optionally (c) may be selected from a targeting domain sequence described herein.

20 In one embodiment, the method comprises contacting a cell from a subject suffering from or likely to develop a disease. The cell may be from a subject having a mutation at a target position in a target gene. In one embodiment, the cell being contacted in the disclosed method is an erythroid cell. The contacting may be performed *ex vivo* and the contacted cell may be returned to the subject's body after the contacting step. In another embodiment, the
25 contacting step may be performed *in vivo*. In one embodiment, the method of altering a cell as described herein comprises acquiring knowledge of the sequence at a target position in said cell, prior to the contacting step. Acquiring knowledge of the sequence at a target position in the cell may be by sequencing the target gene, or a portion of the target gene. In one embodiment, the contacting step of the method comprises contacting the cell with a
30 nucleic acid, *e.g.*, a vector, *e.g.*, an AAV vector, that expresses at least one of (a), (b), and (c). In one embodiment, the contacting step of the method comprises contacting the cell with a nucleic acid, *e.g.*, a vector, *e.g.*, an AAV vector, that expresses each of (a), (b), and (c). In another embodiment, the contacting step of the method comprises delivering to the cell a Cas9 fusion molecule of (b) and a nucleic acid which encodes a gRNA molecule (a) and

optionally, a second gRNA molecule (c)(i) (and further optionally, a third gRNA molecule (c)(iv) and/or fourth gRNA molecule (c)(iii).

In one embodiment, contacting comprises contacting the cell with a nucleic acid, *e.g.*, a vector, *e.g.*, an AAV vector, *e.g.*, an AAV2 vector, a modified AAV2 vector, an AAV3
5 vector, a modified AAV3 vector, an AAV6 vector, a modified AAV6 vector, an AAV8 vector or an AAV9 vector.

In one embodiment, contacting comprises delivering to the cell a Cas9 fusion molecule of (b), as a protein, and a nucleic acid which encodes (a) and optionally a second, third and/or fourth gRNA molecule of (c).

10 In one embodiment, contacting comprises delivering to the cell a Cas9 fusion molecule of (b), as a protein, said gRNA molecule of (a), as an RNA, and optionally said second, third and/or fourth gRNA molecule of (c), as an RNA.

When the method comprises correcting the mutation at a target position by HDR, a Cas9 fusion molecule of (b), at least one gRNA molecule, *e.g.*, a gRNA molecule of (a) are
15 included in the contacting step. In one embodiment, a cell of the subject is contacted *ex vivo* with (a), (b), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In another embodiment, said cell is returned to the subject's body. In one embodiment, a cell of the subject is contacted is *in vivo* with (a), (b), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In one embodiment, the cell of the subject is
20 contacted *in vivo* by intravenous delivery of (a), (b), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In one embodiment, the cell of the subject is contacted *in vivo* by intramuscular delivery of (a), (b), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In one embodiment, the cell of the subject is contacted *in vivo* by subcutaneous delivery of (a), (b), and optionally (c)(i), further optionally
25 (c)(ii), and still further optionally (c)(iii). In one embodiment, the cell of the subject is contacted *in vivo* by intra-bone marrow (IBM) delivery of (a), (b), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

In one embodiment, contacting comprises contacting the subject with a nucleic acid, *e.g.*, a vector, *e.g.*, an AAV vector, described herein, *e.g.*, a nucleic acid that encodes at least
30 one of (a), (b), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

In one embodiment, contacting comprises delivering to said subject said Cas9 fusion molecule of (b), as a protein, and a nucleic acid which encodes (a), and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

In one embodiment, contacting comprises delivering to the subject the Cas9 fusion molecule of (b), as a protein, the gRNA molecule of (a), as an RNA, and optionally the second, third and/or fourth gRNA molecule of (c), as an RNA.

5 In one embodiment, a cell of the subject is contacted *ex vivo* with (a), (b) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In one embodiment, said cell is returned to the subject's body.

In one embodiment, a cell of the subject is contacted is *in vivo* with (a), (b) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In one embodiment, the cell of the subject is contacted *in vivo* by intravenous delivery of (a), (b) and
10 optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In one embodiment, the cell of the subject is contacted *in vivo* by intramuscular delivery of (a), (b) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In one embodiment, the cell of the subject is contacted *in vivo* by subcutaneous delivery of (a), (b) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In one
15 embodiment, the cell of the subject is contacted *in vivo* by intra-bone marrow (IBM) delivery of (a), (b) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

In one embodiment, contacting comprises contacting the subject with a nucleic acid, *e.g.*, a vector, *e.g.*, an AAV vector, described herein, *e.g.*, a nucleic acid that encodes at least one of (a), (b), and optionally (c)(i), further optionally (c)(ii), and still further optionally
20 (c)(iii).

In one embodiment, contacting comprises delivering to said subject said Cas9 fusion molecule of (b), as a protein, and a nucleic acid which encodes (a) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

In one embodiment, contacting comprises delivering to the subject the Cas9 fusion
25 molecule of (b), as a protein, the gRNA molecule of (a), as an RNA, and optionally the second, third and/or fourth gRNA molecule of (c), as an RNA.

In one embodiment, disclosed herein are kits comprising compositions of the invention and instructions for use.

30 *Ex vivo* delivery

In some embodiments, components described in **Table 5** are introduced into cells which are then introduced into the subject. Methods of introducing the components can include, *e.g.*, any of the delivery methods described in **Table 6**.

Modified Nucleosides, Nucleotides, and Nucleic Acids

Modified nucleosides and modified nucleotides can be present in nucleic acids, *e.g.*, particularly gRNA molecule, but also other forms of RNA, *e.g.*, mRNA, RNAi, or siRNA. As described herein, “nucleoside” is defined as a compound containing a five-carbon sugar molecule (a pentose or ribose) or derivative thereof, and an organic base, purine or pyrimidine, or a derivative thereof. As described herein, “nucleotide” is defined as a nucleoside further comprising a phosphate group.

Modified nucleosides and nucleotides can include one or more of:

(i) alteration, *e.g.*, replacement, of one or both of the non-linking phosphate oxygens and/or of one or more of the linking phosphate oxygens in the phosphodiester backbone linkage;

(ii) alteration, *e.g.*, replacement, of a constituent of the ribose sugar, *e.g.*, of the 2' hydroxyl on the ribose sugar;

(iii) wholesale replacement of the phosphate moiety with “dephospho” linkers;

(iv) modification or replacement of a naturally occurring nucleobase;

(v) replacement or modification of the ribose-phosphate backbone;

(vi) modification of the 3' end or 5' end of the oligonucleotide, *e.g.*, removal, modification or replacement of a terminal phosphate group or conjugation of a moiety; and
(vii) modification of the sugar.

The modifications listed above can be combined to provide modified nucleosides and nucleotides that can have two, three, four, or more modifications. For example, a modified nucleoside or nucleotide can have a modified sugar and a modified nucleobase. In one embodiment, every base of a gRNA is modified, *e.g.*, all bases have a modified phosphate group, *e.g.*, all are phosphorothioate groups. In one embodiment, all, or substantially all, of the phosphate groups of a unimolecular (or chimeric) or modular gRNA molecule are replaced with phosphorothioate groups.

In one embodiment, modified nucleotides, *e.g.*, nucleotides having modifications as described herein, can be incorporated into a nucleic acid, *e.g.*, a “modified nucleic acid.” In one embodiment, the modified nucleic acids comprise one, two, three or more modified nucleotides. In one embodiment, at least 5% (*e.g.*, at least about 5%, at least about 10%, at least about 15%, at least about 20%, at least about 25%, at least about 30%, at least about 35%, at least about 40%, at least about 45%, at least about 50%, at least about 55%, at least about 60%, at least about 65%, at least about 70%, at least about 75%, at least about 80%, at

least about 85%, at least about 90%, at least about 95%, or about 100%) of the positions in a modified nucleic acid are a modified nucleotides.

Unmodified nucleic acids can be prone to degradation by, *e.g.*, cellular nucleases. For example, nucleases can hydrolyze nucleic acid phosphodiester bonds. Accordingly, in one aspect the modified nucleic acids described herein can contain one or more modified nucleosides or nucleotides, *e.g.*, to introduce stability toward nucleases.

In one embodiment, the modified nucleosides, modified nucleotides, and modified nucleic acids described herein can exhibit a reduced innate immune response when introduced into a population of cells, both *in vivo* and *ex vivo*. The term “innate immune response” includes a cellular response to exogenous nucleic acids, including single stranded nucleic acids, generally of viral or bacterial origin, which involves the induction of cytokine expression and release, particularly the interferons, and cell death. In one embodiment, the modified nucleosides, modified nucleotides, and modified nucleic acids described herein can disrupt binding of a major groove interacting partner with the nucleic acid. In one embodiment, the modified nucleosides, modified nucleotides, and modified nucleic acids described herein can exhibit a reduced innate immune response when introduced into a population of cells, both *in vivo* and *ex vivo*, and also disrupt binding of a major groove interacting partner with the nucleic acid.

miRNA binding sites

microRNAs (or miRNAs) are naturally occurring cellular 19-25 nucleotide long noncoding RNAs. They bind to nucleic acid molecules having an appropriate miRNA binding site, *e.g.*, in the 3' UTR of an mRNA, and down-regulate gene expression. While not wishing to be bound by theory it is believed that this down regulation occurs either by reducing nucleic acid molecule stability or by inhibiting translation. An RNA species disclosed herein, *e.g.*, an mRNA encoding Cas9 can comprise an miRNA binding site, *e.g.*, in its 3'UTR. The miRNA binding site can be selected to promote down regulation of expression in a selected cell type. By way of example, the incorporation of a binding site for miR-122, a microRNA abundant in liver, can inhibit the expression of the gene of interest in the liver.

EXAMPLES

The following Examples are merely illustrative and are not intended to limit the scope or content of the invention in any way.

5 Example 1: Constructing single- and multi-cysteine variant Cas9 proteins

- To generate single- and multi-cysteine variant Cas9 proteins, established molecular biology techniques and recombinant DNA procedures known to the ordinarily skilled artisan are used. A nucleotide sequence encoding the protein sequence of *Streptococcus pyogenes* Cas9 (SpCas9) and *Staphylococcus aureus* Cas9 (SaCas9) are modified using site-directed
- 10 mutagenesis based, in part, on the published crystal structures of SpCas9 and SaCas9, as described in, *e.g.*, in Anders *et al.*, 2014 Nature 513(7519): 569-73; Nishimau *et al.* Cell, 162(5): 1113-26, to produce single- and multi-cysteine variant Cas9 proteins. For example, using site-directed mutagenesis, the protein sequence of SpCas9 is first mutated to replace native cysteine residues at positions 80 and 574 with serine (*i.e.*, C80S and C574S).
- 15 Similarly, using site-directed mutagenesis, the protein sequence of SaCas9 is first mutated to replace native cysteine residues at positions 237, 534 and 946 with serine (*i.e.*, C237S, C534 and C946S). It is known that these conservative point mutations do not abolish Cas9 activity; see, *e.g.*, Nishimau *et al.* Cell, 162(5): 1113-26. Site-directed mutagenesis is then performed to generate single- and multi-cysteine variant Cas9 proteins. Exemplary cysteine variant
- 20 Cas9 proteins are provided below:

Plasmid	Mutation(s)	Species	Description
pJZ001	C80S C574S	Pyogenes	SpCas9 no cys variant
pJZ002	C80S	Pyogenes	SpCas9 cys variant
pJZ003	C80S C574S D147C	Pyogenes	SpCas9 cys variant
pJZ004	C80S_C574S S204C	Pyogenes	SpCas9 cys variant
pJZ005	C80S C574S Q228C	Pyogenes	SpCas9 cys variant
pJZ006	C80S_C574S N235C	Pyogenes	SpCas9 cys variant
pJZ007	C80S_C574S D257C	Pyogenes	SpCas9 cys variant
pJZ008	C80S_C574S D284C	Pyogenes	SpCas9 cys variant
pJZ009	C80S C574S T313C	Pyogenes	SpCas9 cys variant
pJZ010	C80S C574S D326C	Pyogenes	SpCas9 cys variant
pJZ011	C80S C574S D384C	Pyogenes	SpCas9 cys variant
pJZ012	C80S C574S D428C	Pyogenes	SpCas9 cys variant
pJZ013	C80S_C574S N504C	Pyogenes	SpCas9 cys variant
pJZ014	C80S C574S R535C	Pyogenes	SpCas9 cys variant
pJZ015	C80S C574S L551C	Pyogenes	SpCas9 cys variant
pJZ016	C80S C574S N556C	Pyogenes	SpCas9 cys variant

Plasmid	Mutation(s)	Species	Description
pJZ017	C80S C574S K558C	Pyogenes	SpCas9 cys variant
pJZ018	C80S C574S E566C	Pyogenes	SpCas9 cys variant
pJZ019	C80S_C574S D567C	Pyogenes	SpCas9 cys variant
pJZ020	C80S C574S T605C	Pyogenes	SpCas9 cys variant
pJZ021	C80S C574S T638C	Pyogenes	SpCas9 cys variant
pJZ022	C80S C574S A640C	Pyogenes	SpCas9 cys variant
pJZ023	C80S C574S Q674C	Pyogenes	SpCas9 cys variant
pJZ024	C80S C574S E945C	Pyogenes	SpCas9 cys variant
pJZ025	C80S C574S N946C	Pyogenes	SpCas9 cys variant
pJZ026	C80S C574S L1004C	Pyogenes	SpCas9 cys variant
pJZ027	C80S C574S T1065C	Pyogenes	SpCas9 cys variant
pJZ028	C80S C574S K1076C	Pyogenes	SpCas9 cys variant
pJZ029	C80S C574S D1117C	Pyogenes	SpCas9 cys variant
pJZ030	C80S C574S S1154C	Pyogenes	SpCas9 cys variant
pJZ031	C80S C574S D1328C	Pyogenes	SpCas9 cys variant
pJZ037	C237S C534S C946A	Aureus	SaCas9 no cys variant
pJZ038	D147C D384C N556C A640C E945C	Pyogenes	SpCas9 multi-cys variant
pJZ039	Q192C S431C E745C D795C D849C	Aureus	SaCas9 multi-cys variant
pJZ040	C237S C534S C946A Q192C	Aureus	SaCas9 cys variant
pJZ041	C237S C534S C946A S431C	Aureus	SaCas9 cys variant
pJZ042	C237S C534S C946A E745C	Aureus	SaCas9 cys variant
pJZ043	C237S C534S C946A D795C	Aureus	SaCas9 cys variant
pJZ044	C237S C534S C946A D849C	Aureus	SaCas9 cys variant

For expression in bacteria, cultured cells, or animal tissues, the nucleotide sequence encoding the single- and multi-cysteine variant Cas9 proteins is operably linked to one or more transcriptional control elements, *e.g.*, promoter and/or enhancer elements, which enable expression in the relevant bacteria, cultured cells, or animal tissue. The single- and multi-cysteine variant Cas9 proteins can be purified from the bacteria, cultured cells, or animal tissue using established biochemical techniques. To generate mRNA encoding the single- and multi-cysteine variant Cas9 proteins, the nucleotide sequence encoding the single- and multi-cysteine variant Cas9 proteins is operably linked to a promoter, *e.g.*, a bacteriophage promoter, *e.g.*, a T7 RNA polymerase promoter enabling *in vitro* transcription of mRNA encoding the single- and multi-cysteine variant Cas9 proteins.

The single- and multi-cysteine variant Cas9 proteins allow for site-specific bioconjugation to other molecular species, as described further herein.

Example 2: Generation of Cas9 fusion molecules by covalent attachment of single- and multi-cysteine variant Cas9 proteins to exogenous donor template sequence using 5'-maleimide-modified exogenous donor template sequence

To generate Cas9 fusion molecules, the single- and multi-cysteine variant Cas9 proteins described in Example 1 are covalently attached to template nucleic acid using a 5'-maleimide-modified exogenous donor template sequence (Figure 1). Generally, the template nucleic acid is prepared such that the 5'-end of the template nucleic acid is modified to include maleimide modification at its 5'-end. The maleimide modification may have a one, or two or more carbon spacer arm between the 5'-end of the template nucleic acid and the maleimide modification for purposes of reducing steric interactions between the maleimide modification and the template nucleic acid. The 5'-maleimide-modified template nucleic acid is then incubated in the presence of a single- and/or multi-cysteine variant Cas9 protein molecule described in Example 1, at 4 °C overnight or at room temperature for 4 hours in buffer containing TCEP (*tris*(2-carboxyethyl)phosphine) to keep free cysteines reduced and prevent disulfide bonds from forming, according to standard procedures known to the ordinarily skilled artisan, to covalently attach the 5'-maleimide-modified template nucleic acid to at least one thiol group (*e.g.*, a surface exposed thiol group) from the single- and/or multi-cysteine variant Cas9 protein molecule described in Example 1, to generate a Cas9 fusion molecule (Figure 1).

To test the efficiency of the Cas9 fusion molecule prepared as described in this Example, the reaction products are analyzed using an electrophoretic mobility shift assay using standard procedures known to the ordinary skilled artisan. Results provided in Figure 2 indicate the generation of Cas9 fusion molecule as described in this Example.

Example 3: Generation of Cas9 fusion molecules by covalent attachment of a Cas9 protein to template nucleic acid via HyNic-4FB Conjugation

To generate Cas9 fusion molecules, a HyNic-modified Cas9 protein is covalently attached to a 5'-4-Formylbenzamide (4FB)-modified template nucleic acid. The template nucleic acid is prepared such that the 5'-end of the template nucleic acid is modified to include a 5'-4-formylbenzamide (4FB) moiety at its 5'-end (TriLink BioTechnologies). The chemistry that couples the Cas9 protein to the 4FB-modified template nucleic acid takes advantage of free primary amine groups that are present naturally (or are present as a consequence of site-directed mutagenesis) on the protein surface. Specifically, a Cas9

protein is incubated in the presence of succinimidyl-6-hydrazino-nicotinamide (S-HyNic) (Solulink), which reacts with at least one primary amine on the Cas9 protein to form a 6-hydrazino-nicotinamide (HyNic)-modified Cas9 molecule. The 6-hydrazino-nicotinamide (HyNic)-modified Cas9 protein molecule is then incubated with the 5'-(4FB)-modified template nucleic acid to generate Cas9 fusion molecules. Assays to test the efficiency of the Cas9 fusion molecule prepared as described in this Example can be performed by analyzing the reaction products via an electrophoretic mobility shift assay using standard procedures known to the ordinary skilled artisan.

10 Example 4: Generation of Cas9 fusion molecules by Reaction of a Cas9-HaloTag protein
 molecule with Haloalkane-modified Template nucleic acid

To generate Cas9 fusion molecules, a Cas9-HaloTag protein molecule is created by expressing a nucleic acid construct encoding a HaloTag, optionally fused to a nucleic acid sequence encoding a linker sequence, fused either to the N-terminus or C-terminus of a nucleic acid sequence encoding a Cas9 protein molecule. The nucleic acid construct encoding a HaloTag, is a variant nucleic acid sequence that encodes for a mutant HaloTag comprising a H272F mutation. The HaloTag variant protein (*i.e.*, the H272F HaloTag protein) facilitates the formation of a covalent bond between the HaloTag variant and a haloalkane-modified nucleic acid, *e.g.*, a bromoalkane-modified template nucleic acid.

Generally, for expression in bacteria, cultured cells, or animal tissues, the nucleotide sequence encoding the Cas9 protein fusion (*i.e.*, a Cas9 protein molecule fused to a HaloTag variant) is operably linked to one or more transcriptional control elements, *e.g.*, promoter and/or enhancer elements, which enable expression in the relevant bacteria, cultured cells, or animal tissue. The Cas9 protein fusion can be purified from the bacteria, cultured cells, or animal tissue using established biochemical techniques. Exemplary Cas9 protein fusions (*i.e.*, a Cas9 protein molecule fused to a HaloTag variant) are provided below:

Table 9

Cas9 protein fusion	Linker sequence (AA)	Abbreviation
HaloTag-XTEN linker-Cas9	SGSETPGTSESATPES	HXC
HaloTag-GGS9 linker-Cas9	GGSGGSGGSGGSGGSGGSGGSGGSGGS	HGC
Cas9-XTEN linker-HaloTag	SGSETPGTSESATPES	CXH
Cas9-GGS9 linker-HaloTag	GGSGGSGGSGGSGGSGGSGGSGGSGGS	CGH

The effect of covalent attachment of the template sequence to Cas9 on the frequency of gene correction (HDR) was examined. Generally, the Cas9 protein fusion molecules of Table 9 were prepared as follows.

Cloning: The Cas9 gene and the HaloTag gene were generated by gene synthesis and cloned into a bacterial expression vector (using standard molecular cloning techniques) with one of the linker sequences provided in Table 9 (*i.e.*, the XTEN or the GGS9 linker) in the arrangements set forth in Table 9. This resulted in 4 unique nucleic acid sequences for expressing the Cas9 protein fusion molecules of Table 9. All four constructs comprise an N-terminal His-tag, with sequence His₆, and a nuclear localization signal (NLS), with sequence PKKKRKV. In alternative embodiments, the His₆-tag and the NLS can be cloned to the C-terminus of the Cas9 protein fusion molecules. In alternative embodiments, the Cas9 gene sequence from either *S. Pyogenes* (*spcas9*) or from *S. aureus* (*sacas9*) may also be used.

Pre-conjugation of template sequence to a Cas9-HaloTag protein fusion: Cas9-HaloTag protein fusions (*i.e.*, HXC, HGC, CXH, and CGH, see Table 9) were each incubated with a bromohexyl (BrH)-labeled 30 nucleotide (30 mer) nucleic acid sequence for 1 hour at 37°C, to form a Cas9-HaloTag protein fusion that is conjugated to the 30 nucleotide nucleic acid sequence (Figures 4A and 4B). A 149 nucleotide nucleic acid sequence (149 mer) was subsequently annealed to the nucleic acid portion of the Cas9-HaloTag protein fusion that is conjugated to the 30 nucleotide nucleic acid sequence by incubating the 149 mer with the Cas9-HaloTag protein fusion that is conjugated to the 30 nucleotide nucleic acid sequence in the presence of a DNA splint, followed by addition of a T4 DNA ligase, incubating the reaction at room temperature for 1 hour (Figure 4). The reaction resulted in a Cas9-HaloTag fusion molecule covalently attached to a 179 nucleotide template sequence (Figure 4). A CC8 guide molecule was subsequently added to the reaction mixture, which was then allowed to incubate for 15 minutes prior to nucleofecting into U2OS cells, according to standard procedures known to one of skill in the art. Four days after nucleofection, genomic DNA was extracted and a target sequence locus was amplified and analyzed by next-generation sequencing and scored for %HDR (see **Fig. 5**).

Conjugation of pre-annealed template sequence to a Cas9-HaloTag protein fusion: The 149 mer was pre-annealed to a DNA splint. The BrH-labeled 30 mer was subsequently ligated to the pre-annealed 149mer - DNA splint complex using T4 DNA ligase. The ligation reaction was performed at room temperature for 1 hour, resulting in a full length template sequence of 179 nucleotides (*i.e.*, 179 mer). The 179 mer was added to a reaction with a Cas9-HaloTag protein fusion (*i.e.*, HXC, HGC, CXH, or CGH) and incubated for 1

hour at 37°C, resulting in a Cas9-HaloTag fusion molecule covalently attached to a 179 nucleotide template sequence (Figure 4). The CC8 guide was then added to this mixture and incubated for 15 minutes prior to nucleofecting into U2OS cells, according to standard procedures known to one of skill in the art. Four days after nucleofection, genomic DNA was extracted and a target sequence locus was amplified and analyzed by next-generation sequencing and scored for %HDR (Figure 5).

Results: HDR efficiency can be assayed using a standard nucleofection assay in U2OS cells by comparing the results using the Cas9 fusion molecules (*i.e.*, Cas9-HaloTag protein fusion – template nucleic acid conjugates) in comparison the Cas9-HaloTag protein fusion without a conjugated template nucleic acid. The results in Figure 5 demonstrate that the Cas9 fusion molecules (*i.e.*, Cas9-HaloTag protein fusion – template nucleic acid conjugates) increased the HDR efficiency compared to reactions performed with unconjugated template nucleic acid. Specifically, the results demonstrated a higher rate of HDR as detected by sequencing when the cells were nucleofected with the Cas9 fusion molecule HGC using the pre-conjugation method for conjugating the template nucleic acid to the Cas9-HaloTag protein fusion. Thus, the recruitment of the template nucleic acid to the Cas9 via covalent attachment resulted in a 4% increase in HDR gene correction (Figure 5). The results also demonstrated that the conjugation method using a pre-annealed full length 179 mer does not yield significant conjugation products to the Halo-Cas9 protein fusions tested. Successful conjugation was only observed via the “pre-conjugation” method, by first conjugating the BrH-labeled 30 mer to the Halo-Cas9 protein fusion (**Fig. 4A**) followed by subsequent splint ligation using T4 DNA ligase and the 149 mer.

Example 5: Generation of Cas9 fusion molecules by covalent attachment of single- and multi-cysteine variant Cas9 proteins to template nucleic acid using acrydite-modified template nucleic acid

To generate Cas9 fusion molecules, the single- and multi-cysteine variant Cas9 proteins described in Example 1 are covalently attached to template nucleic acid using an acrydite-modified template nucleic acid. The template nucleic acid is prepared such that the 5'-end of the template nucleic acid is modified to include an acrydite moiety at its 5'-end. The template nucleic acid may also be synthesized such that the 3'-end of the template nucleic acid is modified to include an acrydite moiety either at its 3'-end. The template nucleic acid may also be synthesized such that an internal nucleic acid residue of the template nucleic acid is modified to include an acrydite moiety either at an internal nucleic acid

residue. The acrydite-modified template nucleic acid is then incubated in the presence of a single- and/or multi-cysteine variant Cas9 protein described in Example 1, according to standard procedures known to the ordinarily skilled artisan, to covalently attach the acrydite-modified template nucleic acid to at least one thiol group (*e.g.*, a surface exposed thiol group) from the single- and/or multi-cysteine variant Cas9 protein described in Example 1, to generate Cas9 fusion molecules (Figure 6). Assays to test the efficiency of the Cas9 fusion molecule prepared as described in this Example can be performed by analyzing the reaction products via an electrophoretic mobility shift assay using standard procedures known to the ordinary skilled artisan.

Example 6: Generation of Cas9 fusion molecules by covalent attachment of a Cas9 protein molecule to template nucleic acid via EMCH and EDC Coupling Agents

To generate Cas9 fusion molecules, a *N*-[ϵ -Maleimidocaproic acid] hydrazide (EMCH)-modified Cas9 protein molecule is covalently attached to a 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)-modified template nucleic acid. The template nucleic acid is prepared such that the 5'-end of the template nucleic acid is modified to include a carboxyl moiety (*i.e.*, -COOH), therefore resulting in a carboxy-modified template nucleic acid. Alternatively, the template nucleic acid may be modified such that the template nucleic acid includes a carboxyl moiety (*i.e.*, -COOH) at its 3'-end or at an internal nucleic acid residue. The carboxy-modified template nucleic acid is then incubated with 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC), according to standard procedures known to the ordinarily skilled artisan, to covalently attach EDC to the carboxy-modified template nucleic acid (Figure 7), thereby forming a 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)-modified template nucleic acid.

The *N*-[ϵ -Maleimidocaproic acid] hydrazide (EMCH)-modified Cas9 protein molecule is prepared as described herein. The chemistry that couples the Cas9 protein molecule to EMCH takes advantage of a surface-exposed thiol group on the Cas9 protein molecule (*e.g.*, a cysteine residue) which is capable of forming a covalent bond with EMCH. The Cas9 protein molecule may be a wild-type Cas9 protein molecule. Alternatively, the Cas9 protein molecule may be a single- or multi-cysteine variant Cas9 protein molecule as described in Example 1. Specifically, the Cas9 protein molecule is incubated in the presence of EMCH, according to standard procedures known to the ordinarily skilled artisan, to form a *N*-[ϵ -Maleimidocaproic acid] hydrazide (EMCH)-modified Cas9 protein molecule.

The chemistry that couples the *N*-[ϵ -Maleimidocaproic acid] hydrazide (EMCH)-modified Cas9 protein molecule to the 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)-modified exogenous donor template sequence takes advantage of the primary amine group of the *N*-[ϵ -Maleimidocaproic acid] hydrazide (EMCH) modified Cas9 protein molecule that can form a covalent bond with the 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)-modified template nucleic acid. Specifically, a *N*-[ϵ -Maleimidocaproic acid] hydrazide (EMCH)-modified Cas9 protein molecule is incubated with the 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)-modified template nucleic acid, according to standard procedures known to the ordinarily skilled artisan, to generate a Cas9 fusion molecule (see also, *Immunogenicity and protective efficacy of Bacillus anthracis poly-gamma-D-glutamic acid capsule covalently coupled to a protein carrier using a novel triazine-based conjugation strategy*. Joyce J, Cook J, Chabot D, Hepler R, Shoop W, Xu Q, Stambaugh T, Aste-Amezaga M, Wang S, Indrawati L, Bruner M, Friedlander A, Keller P, Caulfield M J Biol Chem 2006; (281):8 4831-4843). Assays to test the efficiency of the Cas9 fusion molecule prepared as described in this Example can be performed by analyzing the reaction products via an electrophoretic mobility shift assay using standard procedures known to the ordinary skilled artisan.

Example 7: Generation of Cas9 fusion molecules via Non-covalent Coupling Agents

To generate Cas9 fusion molecules, a Cas9 protein molecule, covalently linked to biotin, and a template nucleic acid, covalently linked to biotin, are non-covalently attached via the interaction of the biotin moiety of the Cas9 protein molecule and the biotin moiety of the template nucleic acid each to a unique single monomer of tetrameric streptavidin. Using standard procedures known to the ordinarily skilled artisan, a Cas9 protein molecule is covalently attached to biotin via a polypeptide linker. The polypeptide linker may be one of the linkers disclosed herein (*e.g.*, an XTEN linker, a GGC3 linker, a GGS6 linker, or a GGS9 linker). Generally, the polypeptide linker is sufficiently long to allow the Cas9 protein molecule to interact with the template nucleic acid without steric interference. The template nucleic acid, which is also covalently linked to a biotin molecule, is incubated with the Cas9-Biotin protein molecule in the presence of streptavidin, according to standard procedures known to the ordinarily skilled artisan, thereby forming a Cas9 fusion molecule (Figure 8). Assays to test the efficiency of the Cas9 fusion molecule prepared as described in this Example can be performed by analyzing the reaction products via an electrophoretic mobility shift assay using standard procedures known to the ordinary skilled artisan.

Incorporation by Reference

All publications, patents, and patent applications mentioned herein are hereby incorporated by reference in their entirety as if each individual publication, patent or patent application was specifically and individually indicated to be incorporated by reference. In case of conflict, the present application, including any definitions herein, will control.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described herein. Such equivalents are intended to be encompassed by the following claims.

References

- Anders *et al. Nature* 513(7519):569-573 (2014)
- Bae *et al. Bioinformatics* 30(10):1473-1475 (2014)
- Caldecott *Nat. Rev. Genet.* 9(8):619-631 (2008)
- Cong *et al. Science* 399(6121):819-823 (2013)
- Chylinski *et al. RNA Biol.* 10(5):726-737 (2013)
- Deveau *et al. J. Bacteriol.* 190(4): 1390-1400 (2008)
- Esvelt *et al. Nature* 472(7344): 499-503 (2011)
- Friedland *et al. Genome Biol.* 16:257 (2015)
- Fu *et al. Nat. Biotechnol.* 32:279-284 (2014)
- Haft *et al. PLoS Computational Biology* 1(6): e60 (2005)
- Heigwer *et al. Nat. Methods* 11(2):122-3 (2014)
- Horvath *et al. Science* 327(5962): 167-170 (2010)
- Hsu *et al. Nat. Biotechnol.* 31(9): 827-32 (2013)
- Jinek *et al. Science* 337(6096):816-821 (2012)
- Jinek *et al. Science* 343(6176):1247997 (2014)
- Kleinstiver *et al. Nat. Biotechnol.* 33(12):1293-8 (2015)
- Lee *et al. Nano Lett.* 12(12):6322-6327 (2012)
- Li *Cell. Res.* 18(1):85-98 (2008)
- Makarova *et al. Nature Review Microbiology* 9:467-477 (2011)
- Mali *et al. Science* 399(6121): 823-826 (2013)
- Martelijn *et al. Nat. Rev. Mol. Cell. Biol.* 15(7):465-481 (2014)

- Mayle *et al. Science* 349: 742-47 (2015)
- Neelsen and Lopes *Nat. Rev. Mol. Cell. Biol.* 16: 207-20 (2015)
- Nishimasu *et al. Cell* 156(5):935-949 (2014)
- Ran *et al. Cell* 154(6): 1380-1389 (2013)
- 5 Saleh-Gohari *et al. Mol. Cell. Biol.* 25: 7158-69 (2005)
- Schlacher *et al. Cancer Cell* 22: 106-16 (2012)
- Sternberg *et al. Nature* 507(7490):62-67 (2014)
- Wang *et al. Cell* 153(4):910-918 (2013)
- Xiao A. *et al. Bioinformatics* 30 (8):1180-1182 (2014)
- 10 Zellweger *et al. J. Cell. Biol.* 205: 563-79 (2015)
- Zhou *et al., Nucleic Acids Res.* 42(3):e19 (2014)

CLAIMS

What is claimed is:

1. An enzymatically active Cas9 (eaCas9) fusion molecule, comprising an eaCas9 molecule linked to a template nucleic acid.
2. The eaCas9 fusion molecule of claim 1, wherein the eaCas9 molecule is covalently linked to the template nucleic acid.
3. The eaCas9 fusion molecule of claim 2, wherein the eaCas9 molecule is covalently linked to the template nucleic acid using a polypeptide linker.
4. The eaCas9 fusion molecule of claim 3, wherein the polypeptide linker has a length sufficient to reduce or prevent hybridization of the template nucleic acid to a gRNA molecule associated with the eaCas9 molecule.
5. The eaCas9 fusion molecule of claim 3 or claim 4, wherein the polypeptide linker is sufficiently long to allow the eaCas9 molecule to bind to a target nucleic acid without steric interference.
6. The eaCas9 fusion molecule of claim 3 or claim 4, wherein the polypeptide linker is between about 3 and about 100 amino acids in length.
7. The eaCas9 fusion molecule of any one of claims 3-6, wherein the polypeptide linker is an XTEN linker.
8. The eaCas9 fusion molecule of any one of claims 3-6, wherein the polypeptide linker is a GGS₉ linker that is 27 amino acids in length.
9. The eaCas9 fusion molecule of any one of claims 3-6, wherein the polypeptide linker is a GGS linker that is 3 amino acids in length.

10. The eaCas9 fusion molecule of any one of claims 2-9, wherein the eaCas9 molecule is a variant eaCas9 molecule which has been modified to comprise at least one modification at a surface exposed residue.
11. The eaCas9 fusion molecule of claim 10, wherein the at least one modification is a modification of a non-cysteine amino acid residue to a cysteine residue.
12. The eaCas9 fusion molecule of claim 11, wherein the variant eaCas9 molecule comprises a surface exposed thiol group.
13. The eaCas9 fusion molecule of claim 11 or 12, wherein the template nucleic acid comprises a maleimide modification, and wherein the variant eaCas9 molecule is linked to the template nucleic acid using thiol coupling.
14. The eaCas9 fusion molecule of claim 11 or 12, wherein the template nucleic acid comprises an acrydite modification, and wherein the variant eaCas9 molecule is linked to the template nucleic acid using thiol coupling.
15. The eaCas9 fusion molecule of claim 10, wherein the variant eaCas9 molecule is a succinimidyl-6-hydrazino-nicotinamide (S-HyNic) eaCas9 molecule.
16. The eaCas9 fusion molecule of claim 15, wherein the template nucleic acid comprises a 4Fb modification, and wherein the S-HyNic eaCas9 molecule is linked to the template nucleic acid using amine coupling.
17. The eaCas9 fusion molecule of claim 10, wherein the variant eaCas9 molecule comprises a HaloTag®, a SNAP-tag®, a CLIP-tag®, a ACP-tag®, or a MCP-tag® linked to the eaCas9 molecule.
18. The eaCas9 fusion molecule of claim 17, wherein the template nucleic acid comprises a haloalkane modification, and wherein the variant eaCas9 molecule is linked to the template nucleic acid using an S_N2 reaction.

19. The eaCas9 fusion molecule of claim 2, wherein the eaCas9 molecule is covalently linked to the template nucleic acid using a synthetic linker.
20. The eaCas9 fusion molecule of claim 19, wherein the eaCas9 molecule is a variant eaCas9 molecule which has been modified to comprise a *N*-[ϵ -Maleimidocaproic acid] hydrazide (EMCH) and a 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide (EDC) coupling agent.
21. The eaCas9 fusion molecule of claim 20, wherein the template nucleic acid comprises a carboxy group, and wherein the variant eaCas9 molecule is linked to the template nucleic acid by conjugation of the carboxy group on the template nucleic acid to a primary amine of a hydrazine group on the variant Cas9 molecule.
22. The eaCas9 fusion molecule of claim 1, wherein the eaCas9 molecule is non-covalently linked to the template nucleic acid.
23. The eaCas9 fusion molecule of claim 22, wherein
- (a) the eaCas9 molecule is covalently linked to a first ligand,
 - (b) the template nucleic acid is covalently linked to a second ligand, and
 - (c) the first ligand and the second ligand are non-covalently linked to a ligand acceptor molecule.
24. The eaCas9 fusion molecule of claim 23, wherein the first ligand is biotin, and wherein the second ligand is biotin.
25. The eaCas9 fusion molecule of claim 23 or claim 24, wherein the ligand acceptor molecule is streptavidin.
26. The eaCas9 fusion molecule of any one of claims 23-25, further comprising a linker between the eaCas9 molecule and the first ligand.
27. The eaCas9 fusion molecule of claim 26, wherein the linker is sufficiently long to allow the eaCas9 molecule to bind to a target nucleic acid.

28. The eaCas9 fusion molecule of claim 26 or claim 27, wherein the linker is sufficiently long to allow the template nucleic acid to interact with the eaCas9 molecule without steric interference.
29. The eaCas9 fusion molecule of claim 22, wherein the eaCas9 molecule is covalently linked to a polypeptide, and wherein the polypeptide is non-covalently bound to the template nucleic acid.
30. The eaCas9 fusion molecule of claim 29, wherein the polypeptide is a nucleic acid binding protein.
31. The eaCas9 fusion molecule of claim 30, wherein the nucleic acid binding protein is selected from the group consisting of Rad52, Rad52-yeast, RPA-4 subunit, BRCA2, Rad51, Rad51B, Rad51C, XRCC2, XRCC3, RecA, RadA, HNRNPA1, UP1 Filament of HNRNPA1, NABP2 (SSB1), NABP1 (SSB2), and UHRF1.
32. The eaCas9 fusion molecule of any one of claims 1-31 wherein the template nucleic acid comprises a double stranded nucleic acid sequence or a single stranded nucleic acid sequence.
33. The eaCas9 fusion molecule of any one of claims 1-32, wherein the eaCas9 molecule is a wild-type Cas9 molecule.
34. The eaCas9 fusion molecule of any one of claims 1-32, wherein the eaCas9 molecule is a Cas9 nickase molecule.
35. The eaCas9 fusion molecule of any one of claims 1-32, wherein the eaCas9 molecule is a split Cas9 molecule or an inducible Cas9 molecule.
36. A gene editing system, comprising
at least one eaCas9 fusion molecule comprising an eaCas9 molecule linked to a template nucleic acid, and
at least one gRNA molecule.

37. The gene editing system of claim 36, wherein the at least one gRNA molecule and the eaCas9 fusion molecule are designed to associate with a target nucleic acid and generate a double strand break on the target nucleic acid, wherein the double strand break is repaired by at least one DNA repair pathway, thereby producing a modified target nucleic acid.
38. The gene editing system of claim 36, comprising
a first eaCas9 fusion molecule, wherein the first eaCas9 fusion molecule comprises a first Cas9 nickase molecule linked to the template nucleic acid; and
a first gRNA molecule.
39. The gene editing system of claim 36, comprising
a first eaCas9 fusion molecule, wherein the first eaCas9 fusion molecule comprises a first Cas9 nickase molecule linked to the template nucleic acid;
a first gRNA molecule;
a second eaCas9 fusion molecule, wherein the second eaCas9 fusion molecule comprises a second Cas9 nickase molecule linked to the template nucleic acid; and
a first gRNA molecule.
40. The gene editing system of claim 36, comprising
a first eaCas9 fusion molecule, wherein the first eaCas9 fusion molecule comprises a first Cas9 nickase molecule linked to a first template nucleic acid;
a first gRNA molecule;
a second eaCas9 fusion molecule, wherein the second eaCas9 fusion molecule comprises a second Cas9 nickase molecule linked to a second template nucleic acid; and
a second gRNA molecule.
41. The gene editing system of claim 40,
wherein the first gRNA molecule and the first eaCas9 fusion molecule are designed to associate with a target nucleic acid and generate a first single strand break on a first strand of the target nucleic acid;
wherein the second gRNA molecule and the second eaCas9 fusion molecule are designed to associate with the target nucleic acid and generate a second single strand break on a second strand of the target nucleic acid, thereby forming a double strand break in the target nucleic acid having a first overhang and a second overhang; and

wherein the double strand break is repaired by at least one DNA repair pathway, thereby producing a modified target nucleic acid.

42. The gene editing system of claim 40, wherein each Cas9 nickase molecule has N-terminal RuvC-like domain cleavage activity but no HNH-like domain cleavage activity.

43. The gene editing system of claim 42, wherein each Cas9 nickase molecule comprises an amino acid mutation at an amino acid position corresponding to amino acid position N863 of *Streptococcus pyogenes* Cas9.

44. The gene editing system of claim 40, wherein each Cas9 nickase molecule has HNH-like domain cleavage activity but no N-terminal RuvC-like domain cleavage activity.

45. The gene editing system of claim 44, wherein each Cas9 nickase molecule comprises an amino acid mutation at an amino acid position corresponding to amino acid position D10 of *Streptococcus pyogenes* Cas9.

46. The gene editing system of any one of claims 36-45, further comprising an enzymatically inactive Cas9 (eiCas9) molecule.

47. A cell comprising a gene editing system of any of claims 36-46.

48. A pharmaceutical composition comprising the gene editing system of any one of claims 36-46.

49. A method of modifying a target nucleic acid in a cell, the method comprising:
contacting the cell with a gRNA molecule and an eaCas9 fusion molecule comprising an eaCas9 molecule linked to a template nucleic acid;

wherein the gRNA molecule and the eaCas9 fusion molecule associate with the target nucleic acid and generate a double strand break in the target nucleic acid; and

wherein the double strand break is repaired by gene correction using the template nucleic acid of the eaCas9 fusion molecule.

50. A method of modifying a target nucleic acid in a cell, the method comprising:
contacting the cell with a first gRNA molecule; a first eaCas9 molecule; a second gRNA molecule; and a second eaCas9 molecule;
wherein at least one of the first and second eaCas9 molecule is linked to a template nucleic acid,
wherein the first gRNA molecule and the first eaCas9 molecule associate with the target nucleic acid and generate a first single strand cleavage event on a first strand of the target nucleic acid;
wherein the second gRNA molecule and the second eaCas9 molecule associate with the target nucleic acid and generate a second single strand cleavage event on a second strand of the target nucleic acid, thereby forming a double strand break having a first overhang and a second overhang; and
wherein the first overhang and the second overhang in the target nucleic acid are repaired by gene correction using the template nucleic acid.
51. The method of claim 50, wherein each eaCas9 has N-terminal RuvC-like domain cleavage activity but no HNH-like domain cleavage activity.
52. The method of claim 51, wherein each eaCas9 comprises an amino acid mutation at an amino acid position corresponding to amino acid position N863 of *Streptococcus pyogenes* Cas9.
53. The method of claim 50, wherein each eaCas9 has HNH-like domain cleavage activity but no N-terminal RuvC-like domain cleavage activity.
54. The method of claim 53, wherein each eaCas9 comprises an amino acid mutation at an amino acid position corresponding to amino acid position D10 of *Streptococcus pyogenes* Cas9.
55. The method of claim 50, wherein the first eaCas9 molecule is linked to the template nucleic acid.
56. The method of claim 50, wherein both the first eaCas9 molecule and the second eaCas9 molecule are linked to the template nucleic acid,

57. The method of any one of claims 49-56, wherein the cell is a mammalian cell.
58. The method of claim 57, wherein the cell is a human cell.
59. A cell altered by the method of any one of claims 49-58.
60. A pharmaceutical composition comprising the cell of claim 59.
61. A nucleic acid molecule encoding at least one eaCas9 fusion molecule, comprising an eaCas9 molecule and a polypeptide.
62. The nucleic acid molecule of claim 61, wherein the polypeptide is a nucleic acid binding protein.
63. The nucleic acid molecule of claim 62, wherein the nucleic acid binding protein is selected from the group consisting of: Rad52, Rad52-yeast, RPA-4 subunit, BRCA2, Rad51, Rad51B, Rad51C, XRCC2, XRCC3, RecA, RadA, HNRNPA1, UP1 Filament of HNRNPA1, NABP2 (SSB1), NABP1 (SSB2), and UHRF1.
64. A vector comprising the nucleic acid molecule of claim 63.

Figure 1

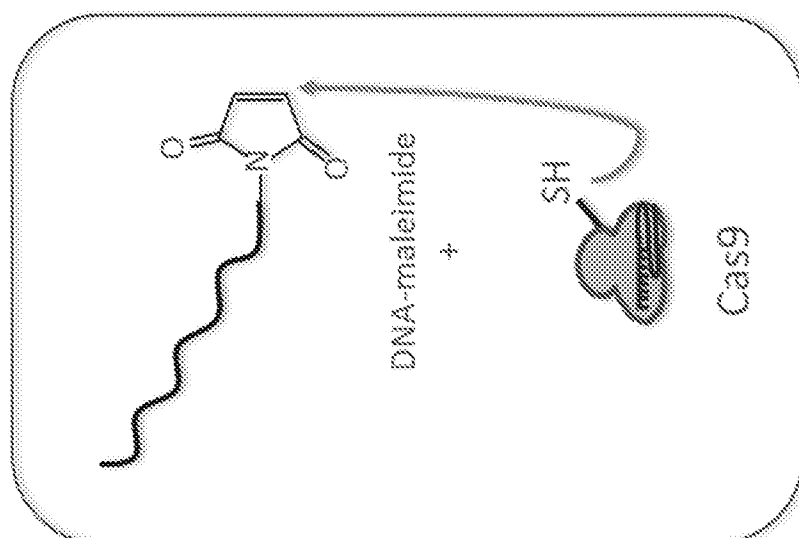


Figure 2

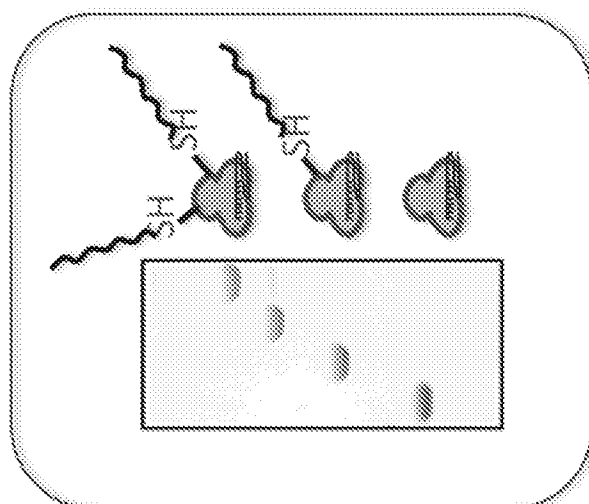


Figure 3

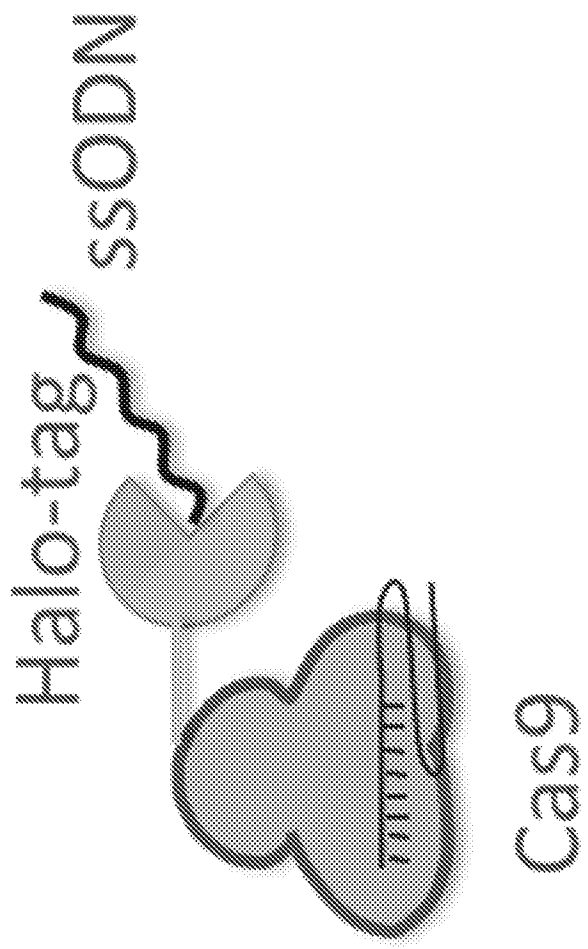


Figure 4A

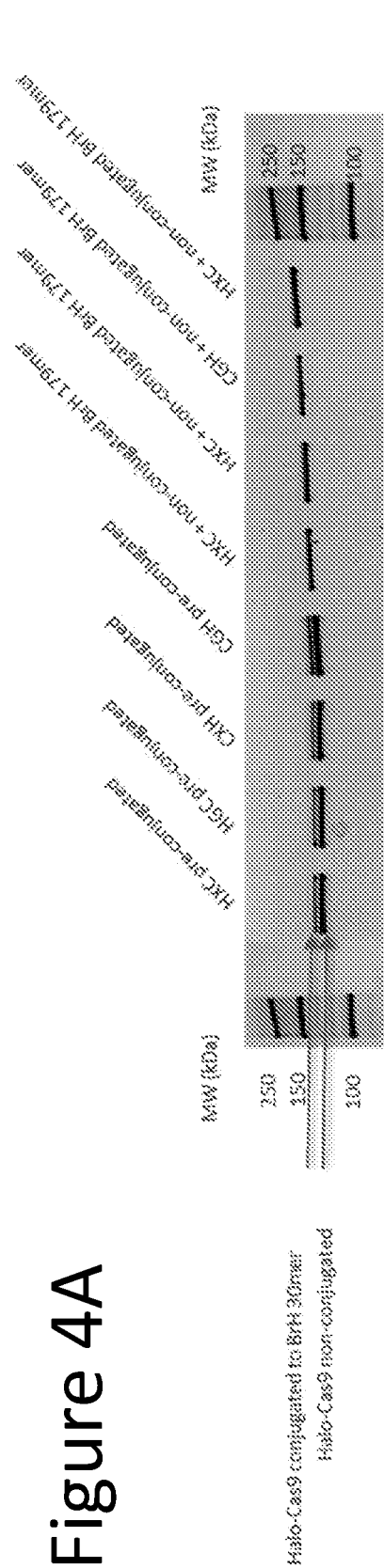


Figure 4B

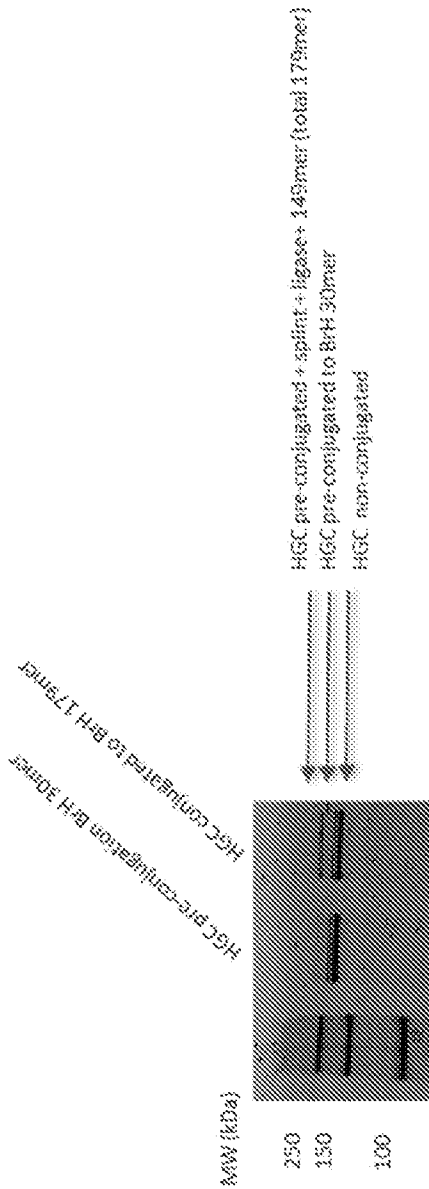
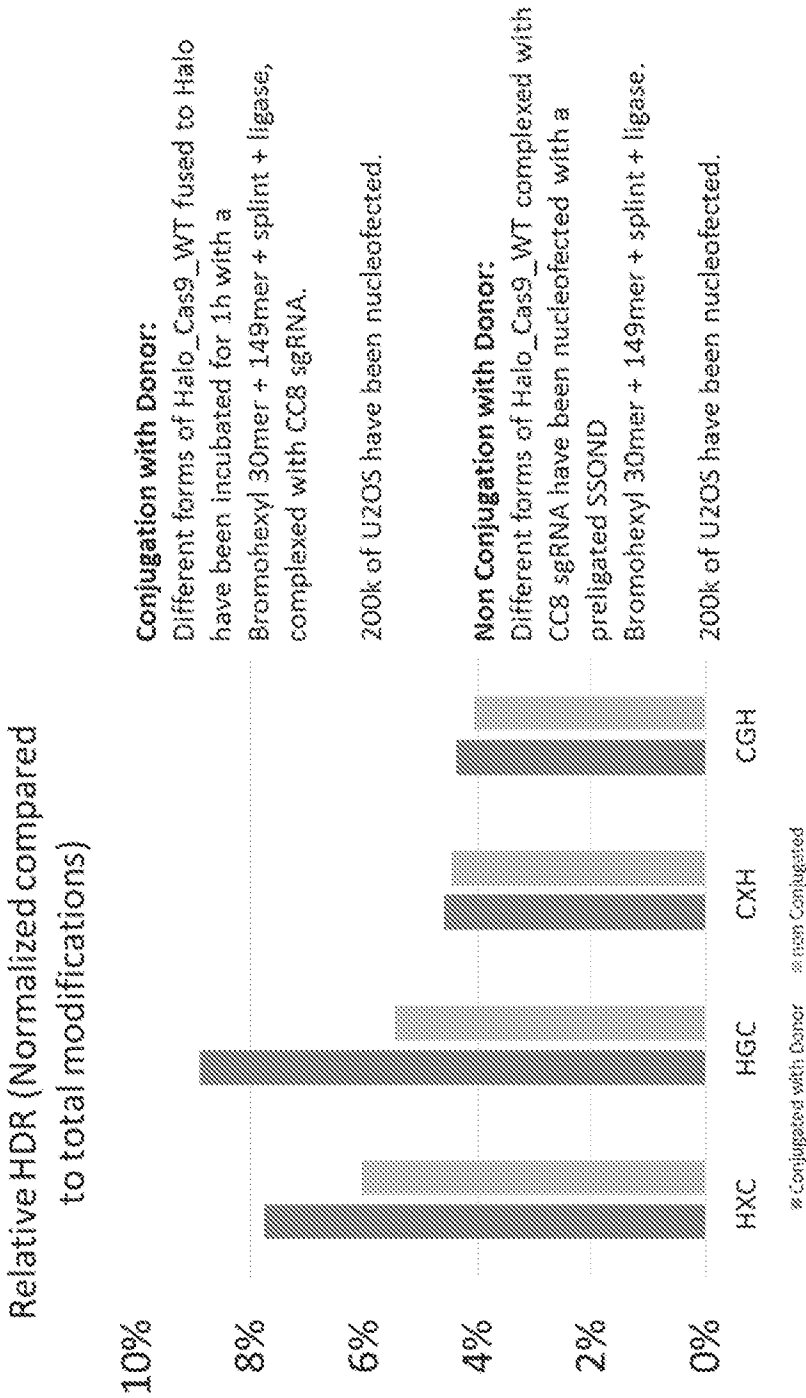


Figure 5



Conclusion: By recruiting the SSODN to the Cas9 in a covalent way we are able to achieve 4% more of HDR modifications

Figure 6

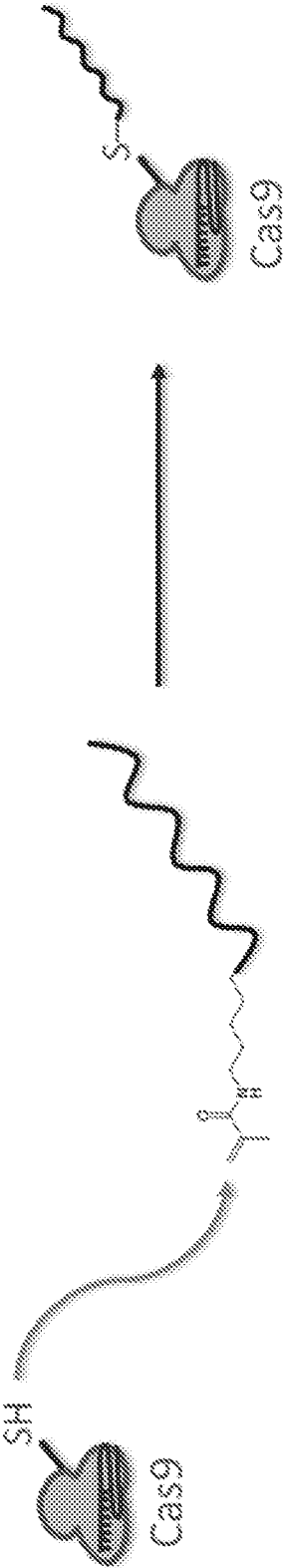


Figure 7

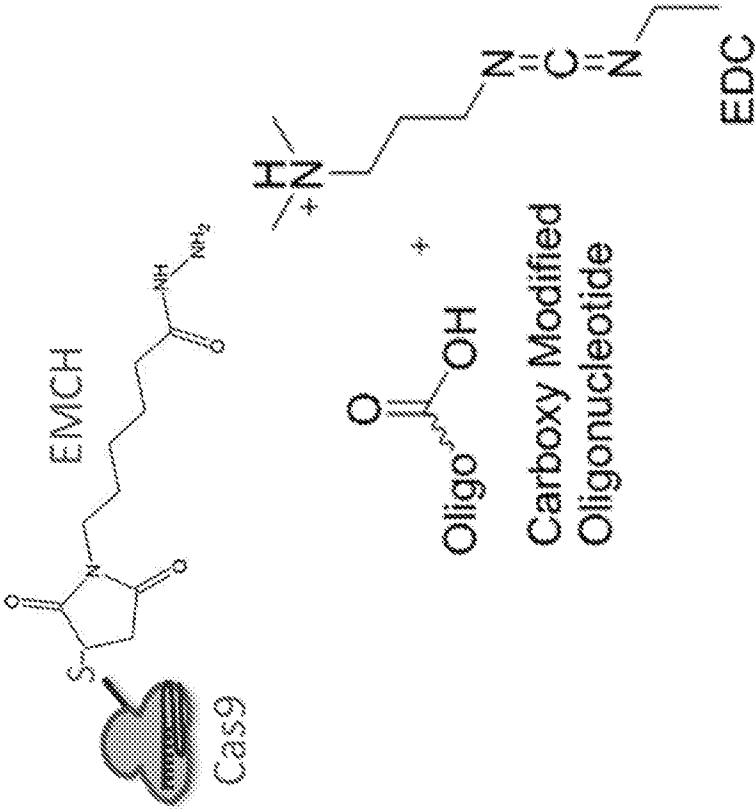
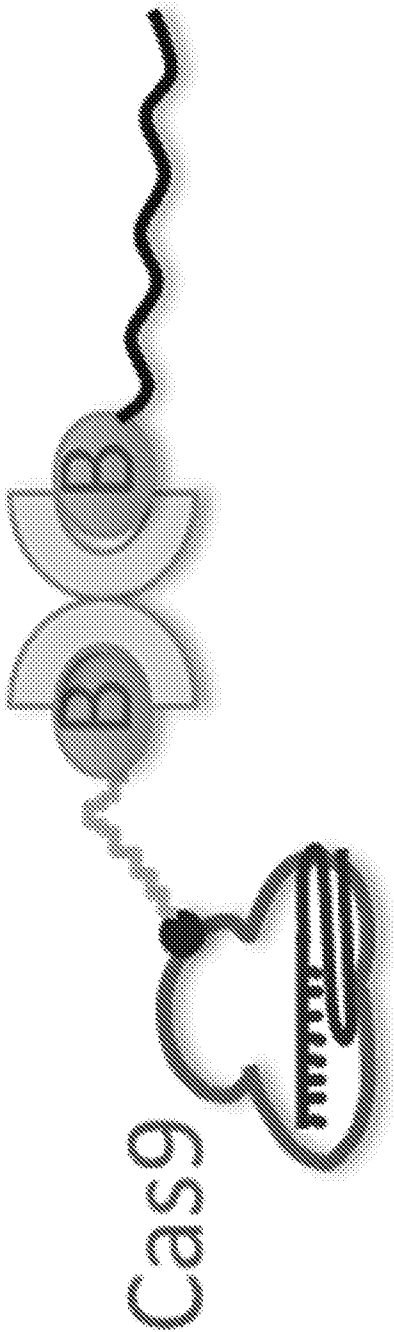


Figure 8



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US2017/027126

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

1. With regard to any nucleotide and/or amino acid sequence disclosed in the international application, the international search was carried out on the basis of a sequence listing:
 - a. ☒ forming part of the international application as filed:
 - ☒ in the form of an Annex C/ST.25 text file.
 - ☐ on paper or in the form of an image file.
 - b. ☐ furnished together with the international application under PCT Rule 13~~ter~~.1(a) for the purposes of international search only in the form of an Annex C/ST.25 text file.
 - c. ☐ furnished subsequent to the international filing date for the purposes of international search only:
 - ☐ in the form of an Annex C/ST.25 text file (Rule 13~~ter~~.1(a)).
 - ☐ on paper or in the form of an image file (Rule 13~~ter~~.1(b) and Administrative Instructions, Section 713).
2. ☐ In addition, in the case that more than one version or copy of a sequence listing has been filed or furnished, the required statements that the information in the subsequent or additional copies is identical to that forming part of the application as filed or does not go beyond the application as filed, as appropriate, were furnished.
3. Additional comments:

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/027126

A. CLASSIFICATION OF SUBJECT MATTER

INV. C12N15/10 C12N15/62 C12N15/63 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)

C12N C07K

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, BIOSIS, CHEM ABS Data, EMBASE, INSPEC, WPI Data

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 2016/054326 A1 (GEN HOSPITAL CORP [US]) 7 April 2016 (2016-04-07)	1-5, 36, 47-50, 59-64
Y	claims 1-31 figure 1	1-64
X	----- CEM KUSCU ET AL: "Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease", NATURE BIOTECHNOLOGY, vol. 32, no. 7, 18 May 2014 (2014-05-18), pages 677-683, XP055382577, US ISSN: 1087-0156, DOI: 10.1038/nbt.2916 the whole document ----- -/--	1-5, 36, 47-50, 59, 60



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

22 June 2017

Date of mailing of the international search report

04/07/2017

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

Bayer, Martin

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2017/027126

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	MEHMET FATIH BOLUKBASI ET AL: "DNA-binding-domain fusions enhance the targeting range and precision of Cas9", NATURE METHODS, vol. 12, no. 12, 19 October 2015 (2015-10-19), pages 1150-1156, XP055382765, ISSN: 1548-7091, DOI: 10.1038/nmeth.3624	1-5,36, 47-50, 59-64
Y	the whole document -----	1-64
X	US 2015/071898 A1 (LIU DAVID R [US] ET AL) 12 March 2015 (2015-03-12)	1-5,36, 47-50, 59-64
Y	claims 1-27 examples 1-4 figure 1 -----	1-64
X,P	KLAUS N. LOVENDAHL ET AL: "Sequence-Directed Covalent Protein-DNA Linkages in a Single Step Using HUH-Tags", JOURNAL OF THE AMERICAN CHEMICAL SOCIETY, vol. 139, no. 20, 16 May 2017 (2017-05-16) , pages 7030-7035, XP055381102, US ISSN: 0002-7863, DOI: 10.1021/jacs.7b02572 the whole document -----	1-5,36, 47-50, 59-64

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2017/027126

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 2016054326 A1	07-04-2016	AU 2015324935 A1 CA 2963080 A1 WO 2016054326 A1	20-04-2017 07-04-2016 07-04-2016
US 2015071898 A1	12-03-2015	AU 2014315070 A1 CA 2923418 A1 EP 3041497 A2 JP 2016537008 A KR 20160050069 A US 2015071898 A1 US 2015071899 A1 US 2016215276 A1 WO 2015035162 A2	17-03-2016 12-03-2015 13-07-2016 01-12-2016 10-05-2016 12-03-2015 12-03-2015 28-07-2016 12-03-2015