



(43) International Publication Date
29 September 2016 (29.09.2016)

(51) International Patent Classification:
C12N 15/10 (2006.01)

(21) International Application Number:
PCT/US2016/024307

(22) International Filing Date:
25 March 2016 (25.03.2016)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/138,948 26 March 2015 (26.03.2015) US
62/182,416 19 June 2015 (19.06.2015) US
62/194,078 17 July 2015 (17.07.2015) US
62/220,660 18 September 2015 (18.09.2015) US
62/232,683 25 September 2015 (25.09.2015) US
62/232,675 25 September 2015 (25.09.2015) US
PCT/US2015/059782
9 November 2015 (09.11.2015) US

(71) Applicant: EDITAS MEDICINE, INC. [US/US]; 300
Third Street, First Floor, Cambridge, MA 02142 (US).

(72) Inventors: COTTA-RAMUSINO, Cecilia; 55 Prince
Street, Cambridge, MA 02139 (US). GORI, Jennifer,
Leah; 20 Spaulding Street, Unit 1, Jamaica Plain, MA
02130 (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, DK, DM,
DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT,
HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR,
KZ, LA, LC, LK, LR, LS, 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).

Published:

- without international search report and to be republished
upon receipt of that report (Rule 48.2(g))
- with sequence listing part of description (Rule 5.2(a))

(54) Title: CRISPR/CAS-MEDIATED GENE CONVERSION

(57) Abstract: CRISPR/CAS-related compositions and methods for altering a cell or treating a disease, for example, by gene conversion, are disclosed.



WO 2016/154579 A2

CRISPR/CAS-MEDIATED GENE CONVERSION

RELATED APPLICATIONS

This application claims priority to U.S. Provisional Patent Application No. 62/138,948, filed on March 26, 2015, and to U.S. Provisional Patent Application No. 62/182,416, filed June 19, 2015, to U.S. Provisional Patent Application No. 62/194,078, filed on July 17, 2015, to U.S. Provisional Patent Application No. 62/220,660, filed on September 18, 2015, to U.S. Provisional Patent Application No. 62/232,675, filed on September 25, 2015, to U.S. Provisional Patent Application No. 62/232,683, filed on September 25, 2015, and to International Application No. PCT/US2015/059782, filed on November 9, 2015, the entire contents of each of which are expressly incorporated herein by reference.

SEQUENCE LISTING

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 March 24, 2016, is named 2016-03-25_126454-00420_ST25.txt and is 1.77 megabytes in size.

FIELD OF THE INVENTION

The invention relates to CRISPR/CAS-related methods and components for increasing editing of a target nucleic acid sequence by gene conversion using an endogenous homologous region, and applications thereof.

BACKGROUND

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 that includes the viral sequence. That RNA, which contains 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 been adapted for genome editing in eukaryotic cells. The introduction of site-specific double strand breaks (DSBs) enables target nucleic acid alteration. After the formation of a DNA double-stranded break (DSB), the major

decision point affecting DNA repair pathway choice is whether or not the DNA ends are endo- and exonucleolytically processed in a process referred to as end resection. When no end resection takes place, the repair pathway engaged to repair the DSB is referred to as classical non-homologous end joining (C-NHEJ). The C-NHEJ repair pathway leads to either perfect repair of the DSBs, in which case the locus is restored without sequence alterations, or to the formation of small insertions and deletions.

In contrast, if the end resection machinery processes the DSB, a 3' overhang is exposed, which engages in homology search. A not yet completely characterized class of pathways that can engage the repair of DSBs after resection is initiated is referred to as alternative non-homologous end joining (ALT-NHEJ). Examples of pathways that are categorized as ALT-NHEJ include blunt end-joining (blunt EJ) and microhomology mediated end joining (MMEJ) leading to deletions, as well as synthesis dependent micro homology mediated end joining (SD-MMEJ), leading to the formation of insertions.

When the end resection is extensive, the exposed 3' overhang can undergo strand invasion of highly homologous sequences, followed by repair of the DSB by a homology-dependent recombination (HDR) pathway, which encompasses gene correction and gene conversion. Gene correction refers to the process of repairing DNA damage by HDR using an exogenous nucleic acid, *e.g.*, an exogenous donor template nucleic acid. Gene conversion is a process whereby a DNA break leads to repair of the sequence in the vicinity of the break by an endogenous homologous sequence, referred to as an endogenous donor template sequence. An endogenous donor template sequence can be derived from a non-sister chromatid sequence or from a heterologous sequence that has a high degree of homology with the sequence undergoing repair. The DNA sequence that is converted can comprise a small portion of a gene (*e.g.*, fewer than 100 bp) or an entire gene sequence (*e.g.*, greater than 5 kb). Gene conversion is a common form of DNA double-strand break repair in both non-mammalian and mammalian, *e.g.*, human, cells.

While a cell could, in theory, repair DNA breaks via any of a number of DNA damage repair pathways, in certain circumstances it is particularly useful to provide an environment more favorable for repair of a break by gene conversion using an endogenous homologous region as a donor template. However, there remains a need to improve the efficiency of gene conversion-mediated modification in order to broaden the applicability and efficiency of genome editing by CRISPR/Cas systems.

SUMMARY

The methods and compositions described herein surprisingly increase DNA repair via gene conversion pathways to alter or modify, *e.g.*, repair, sequences, *e.g.*, genes, such as disease causing genes, using an endogenous homologous region which functions as a donor template, following a Cas9 molecule-mediated cleavage event. In one embodiment, the endogenous homologous region is, for example, a sequence of a non-sister chromatid or a heterologous sequence from a different gene that has a high degree of homology with the sequence undergoing repair. In one embodiment, introduction of two 5' single-strand breaks by a targeted Cas9 nickase increases the rate of DNA repair via gene conversion.

In one aspect, this disclosure relates to methods and compositions for modifying a target gene in a mammalian cell, involving contacting the cell with first and second Cas9 nickases and first and second (non-naturally occurring) guide RNAs (gRNAs). The first Cas9 nickase and the first gRNA generate a first single strand break on a first strand of the target gene, while the second Cas9 nickase and the second gRNA generate a single strand break on a second strand of the target gene, thereby forming within the target gene a double strand break having two 5' overhangs. The double strand break having two 5' overhangs is then repaired by gene conversion using an endogenous homologous region as a template sequence, thereby modifying the target gene to comprise a sequence that is identical to the endogenous homologous region.

In one embodiment, the endogenous homologous region is located on a chromosome on which the target gene is located. For example, the endogenous homologous region may be located in a region of a gene that is homologous to the portion of the target gene to be altered. In some embodiments, the endogenous homologous region is located within a gene cluster comprising the target gene. In one embodiment, the target gene is an *HBB* gene, and the endogenous homologous region is a region of an *HBD* gene. Alternatively or additionally, the target gene to be modified can comprise a sequence (for instance, a point mutation, insertion or deletion) linked to a disease, while the endogenous homologous region sequence does not include such a sequence (for instance, a point mutation, insertion or deletion) linked to a disease. Methods according to this disclosure may be used to alter the genotype of a cell, for instance, to treat a disease, such as sickle-cell anemia, beta-thalassemia, spinal muscular atrophy and/or chronic granulomatous disease.

In another aspect, disclosed herein is a method of modifying an endogenous target gene in a cell, the method comprising contacting the cell with a first gRNA molecule, a first enzymatically active Cas9 (eaCas9) molecule, a second gRNA molecule, and a second

eaCas9 molecule; wherein the first gRNA molecule and the first eaCas9 molecule associate with the target gene and generate a first single strand cleavage event on a first strand of the target gene; wherein the second gRNA molecule and the second eaCas9 molecule associate with the target gene and produce a second single strand cleavage event on a second strand of the target gene, 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 gene are repaired using an endogenous homologous region, thereby modifying the endogenous target gene in the cell.

In one embodiment, after repair of the first overhang and the second overhang, the target gene comprises the sequence of the endogenous homologous region.

In one embodiment, the cell is not contacted with an exogenous nucleic acid homologous to the target gene. In one embodiment, the cell is not contacted with an exogenous template nucleic acid.

In one embodiment, the first overhang and the second overhang are repaired by homology directed repair (HDR). In one embodiment, the homology directed repair (HDR) is gene conversion.

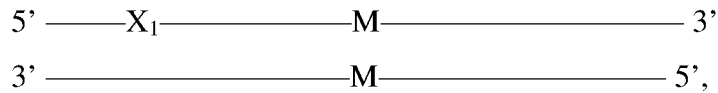
In one embodiment, the first overhang is a 5' overhang. In another embodiment, the second overhang is a 5' overhang. In one embodiment, the first overhang is a 5' overhang, and the second overhang is a 5' overhang.

In one embodiment, the first overhang and the second overhang undergo processing, exposing a first 3' overhang and a second 3' overhang. In one embodiment, the processing is endonucleolytic processing. In another embodiment, the processing is exonucleolytic processing.

In one embodiment, the method is used to correct a mutation in the target gene, wherein the mutation is located between the first single strand break and the second single strand break. In one embodiment, the method is used to correct a mutation in the target gene, wherein the mutation is located within fewer than 50 nucleotides of the first single strand break. In one embodiment, the method is used to correct a mutation in the target gene, wherein the mutation is located within fewer than 50 nucleotides of the second single strand break.

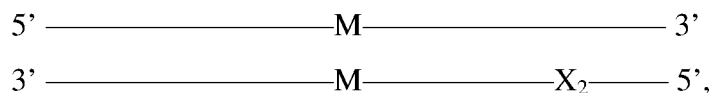
In one embodiment, the method is used to correct a mutation in the endogenous target gene, and wherein the first gRNA molecule positions the first single strand cleavage event 5' to the mutation on the first strand of the target gene. In one embodiment, the method is used to correct a mutation in the endogenous target gene, and wherein the first gRNA molecule

positions the first single strand cleavage event 5' to the mutation on a first strand of the target gene, as shown in the diagram below:



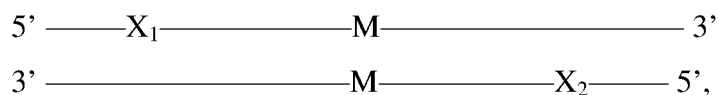
5 wherein X_1 is the first single strand cleavage event and M is mutation in the target gene.

In one embodiment, the method is used to correct a mutation in the endogenous target gene, and wherein the second gRNA molecule positions the second single strand cleavage event 5' to the mutation on a second strand of the target gene. In one embodiment, the method is used to correct a mutation in the endogenous target gene, and wherein the second gRNA molecule positions the second single strand cleavage event 5' to the mutation on a second strand of the target gene, as shown in the diagram below:



wherein X_2 is the second single strand cleavage event and M is the mutation in the target gene.

In one embodiment, the method is used to correct a mutation in the endogenous target gene, and wherein the first gRNA molecule positions the first single strand cleavage event 5' to the mutation on a first strand of the target gene, and wherein the second gRNA molecule positions the second single strand cleavage event 5' to the mutation on a second strand of the target gene. In one embodiment, the method is used to correct a mutation in the endogenous target gene, and wherein the first gRNA molecule positions the first single strand cleavage event 5' to the mutation on a first strand of the target gene, and wherein the second gRNA molecule positions the second single strand cleavage event 5' to the mutation on a second strand of the target gene, as shown in the diagram below:



wherein X_1 is first single strand cleavage event, X_2 is the second single strand cleavage event, and M is the mutation in the target gene.

In one embodiment, the mutation is a point mutation (GAG to GTG) in the sense strand of the *HBB* gene, which results in the substitution of valine for glutamic acid at amino acid position 6 in exon 1 of the *HBB* gene.

In one embodiment, the cell is a population of cells.

In one embodiment, the first single strand cleavage event results in a first 5' overhang, and wherein the second single strand cleavage event results in a second 5' overhang.

In one embodiment, the endogenous target gene has at least 80% sequence homology with the endogenous homologous region. In one embodiment, the endogenous target gene has at least 85% sequence homology with the endogenous homologous region. In one embodiment, the endogenous target gene has at least 92% sequence homology with the endogenous homologous region. In one embodiment, the endogenous target gene has at least 95% sequence homology with the endogenous homologous region. In one embodiment, the endogenous target gene has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence homology with the endogenous homologous region.

In one embodiment, the method is used to correct a mutation in the endogenous target gene, and wherein the endogenous homologous region comprises a domain which is homologous to a domain in the endogenous target gene but which does not comprise the mutation.

In one embodiment, the endogenous homologous region comprises a control region which is homologous to a control region of the target gene, a coding region which is homologous to a coding region of the target gene, a non-coding region which is homologous to a non-coding region of the target gene, an intron which is homologous to an intron of the target gene, or an exon which is homologous to an exon of the target gene.

In one embodiment, the first eaCas9 molecule is a first nickase molecule and the second eaCas9 molecule is a second nickase molecule. In one embodiment, each nickase molecule forms a single strand break in the target gene. In one embodiment, the first eaCas9 molecule and the second eaCas9 molecule are the same species of eaCas9 molecule.

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

In one embodiment, the target gene is an *HBB* gene. In one embodiment, the method is used to correct a mutation in the *HBB* target gene, wherein the mutation in the *HBB* target gene causes sickle cell disease or beta-thalassemia. In one embodiment, the mutation in the *HBB* target gene is a substitution of valine for glutamic acid at amino acid position 6 in exon 1 of the *HBB* gene. In one embodiment, the endogenous homologous region is in an endogenous *HBD* gene.

In one embodiment, the first gRNA molecule is a gRNA molecule comprising a sequence set forth as any one of SEQ ID NOs: 387-485, 6803-6871, or 16010-16256, and the second gRNA molecule is a gRNA molecule comprising a sequence set forth as any one of SEQ ID NOs: 387-485, 6803-6871, or 16010-16256. In one embodiment, the first gRNA molecule is a gRNA molecule comprising a sequence of SEQ ID NO:387, and the second gRNA molecule is a gRNA molecule comprising a sequence of SEQ ID NO:16318. In one embodiment, the first gRNA molecule is a gRNA molecule comprising a sequence of SEQ ID NO:387, and the second gRNA molecule is a gRNA molecule comprising a sequence of SEQ ID NO:388.

In one embodiment, the target gene is an *SMN1* gene. In one embodiment, the method is used to correct a mutation in the *SMN1* target gene, and wherein the mutation in the *SMN1* gene causes spinal muscular atrophy. In one embodiment, the endogenous homologous region is in an *SMN2* gene.

In one embodiment, the target gene is an *NCF1* (p47-PHOX) gene. In one embodiment, the method is used to correct a mutation in the *NCF1* target gene, and wherein the mutation in the *NCF1* gene causes chronic granulomatous disease. In one embodiment, the endogenous homologous region is in a p47-PHOX pseudogene.

In one embodiment, the cell is a population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 12% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 15% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 20% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 25% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 30% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 35% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 40% of the cells in the population of cells.

In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 45% of the cells in the population of cells.

In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 12% to about 45% of the cells in the population of cells.

In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 15% to about 45% of the cells in the population of cells.

5 In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 20% to about 45% of the cells in the population of cells.

In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 25% to about 45% of the cells in the population of cells.

10 In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 12% to about 40% of the cells in the population of cells.

In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 12% to about 30% of the cells in the population of cells.

In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 12% to about 25% of the cells in the population of cells.

15 In one embodiment, the cell is a population of hematopoietic stem cells (HSCs). In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 4% of the cells in the population of HSCs. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 5% of the cells in the population of HSCs. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 6% of the cells in the population of HSCs. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in about 4% to about 6% of the cells in the population of HSCs.

25 In one embodiment, the first overhang and the second overhang in the target gene are repaired by non-homologous end joining (NHEJ) in less than 40% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by non-homologous end joining (NHEJ) in less than 30% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by non-homologous end joining (NHEJ) in less than 20% of the cells in the population of cells. In one embodiment, the NHEJ results in a deletion in the target gene. In one embodiment, the NHEJ results in an insertion in the target gene.

30 In one embodiment, the cell is from a subject suffering from a disease. In one embodiment, the disease is a blood disease, an immune disease, a neurological disease, a cancer, or an infectious disease. In one embodiment, the disease is beta thalassemia, sickle

cell disease, spinal muscular atrophy, or chronic granulomatous disease. In one embodiment, the cell is a mammalian cell, *e.g.*, a human cell.

In one embodiment, the cell is a blood cell, a neuronal cell, an immune cell, a muscle cell, a stem cell, a progenitor cell, or a diseased cell. In one embodiment, the cell is a T-cell.

5 In one embodiment, the method further comprises contacting the cells with CD3/CD28 immunomagnetic beads. In one embodiment, contacting the cells with the immunomagnetic beads increases the level of gene conversion in the cell.

In one embodiment, the contacting step is performed *ex vivo*. In one embodiment, the contacted cell is returned to the body of the subject.

10 In one embodiment, the contacting is performed *in vivo*.

In one embodiment, the method further comprises sequencing the target gene in the cell and sequencing the endogenous homologous region in the cell before the contacting step. In one embodiment, a portion of the target gene is sequenced. In one embodiment, the method further comprises sequencing the target gene in the cell after the contacting step. In
15 one embodiment, a portion of the target gene is sequenced.

In another aspect, disclosed herein is a method of treating a disease in a subject having a mutation in an endogenous target gene, the method comprising contacting a cell from the subject with a first gRNA molecule, a first enzymatically active Cas9 (eaCas9) molecule, a second gRNA molecule, and a second eaCas9 molecule; wherein the first gRNA
20 molecule and the first eaCas9 molecule associate with the target gene and generate a first single strand cleavage event on a first strand of the target gene; wherein the second gRNA molecule and the second eaCas9 molecule associate with the target gene and generate a second single strand cleavage event on a second strand of the target gene, thereby forming a double strand break having a first overhang and a second overhang; and wherein the first
25 overhang and the second overhang in the target gene are repaired using an endogenous homologous region which does not comprise the mutation, correcting the mutation in the endogenous target gene in the cell, thereby treating the disease in the subject having the mutation in the endogenous gene.

In one embodiment, the cell is not contacted with an exogenous nucleic acid
30 homologous to the target gene.

In one embodiment, the cell is a population of cells.

In one embodiment, the endogenous target gene has at least 80% sequence homology with the endogenous homologous region. In one embodiment, the endogenous target gene has at least 85% sequence homology with the endogenous homologous region. In one

embodiment, the endogenous target gene has at least 92% sequence homology with the endogenous homologous region. In one embodiment, the endogenous target gene has at least 95% sequence homology with the endogenous homologous region. In one embodiment, the endogenous target gene has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%,
 5 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence homology with the endogenous homologous region.

In one embodiment, the endogenous homologous region comprises a domain which is homologous to a domain in the target gene but which does not comprise the mutation. In one embodiment, the endogenous homologous region comprises a control region which is
 10 homologous to a control region of the target gene, a coding region which is homologous to a coding region of the target gene, a non-coding region which is homologous to a non-coding region of the target gene, an intron which is homologous to an intron of the target gene, or an exon which is homologous to an exon of the target gene.

In one embodiment, the first eaCas9 molecule and the second eaCas9 molecule are
 15 each nickase molecules. In one embodiment, the first eaCas9 molecule and the second eaCas9 molecule are the same species of nickase molecule. In one embodiment, the eaCas9 nickase molecule comprises a mutation at an amino acid position corresponding to amino acid position D10 of *Streptococcus pyogenes* Cas9.

In one embodiment, the target gene is an *HBB* gene. In one embodiment, the method
 20 is used to correct a mutation in the *HBB* target gene, wherein the mutation in the *HBB* target gene causes sickle cell disease or beta-thalassemia. In one embodiment, the mutation in the *HBB* target gene is a substitution of valine for glutamic acid at amino acid position 6 in exon 1 of the *HBB* gene. In one embodiment, the endogenous homologous region is from an endogenous *HBD* gene.

In one embodiment, the first gRNA molecule is a gRNA molecule comprising a
 25 sequence set forth as any one of SEQ ID NOs: 387-485, 6803-6871, or 16010-16256, and the second gRNA molecule is a gRNA molecule comprising a sequence set forth as any one of SEQ ID NOs: 387-485, 6803-6871, or 16010-16256. In one embodiment, the first gRNA molecule is a gRNA molecule comprising a sequence of SEQ ID NO:387, and the second
 30 gRNA molecule is a gRNA molecule comprising a sequence of SEQ ID NO:16318.

In one embodiment, the target gene is an *SMN1* gene. In one embodiment, the method is used to correct a mutation in the *SMN1* target gene, and wherein the mutation in the *SMN1* gene causes spinal muscular atrophy. In one embodiment, the endogenous homologous region is in an *SMN2* gene.

In one embodiment, the target gene is an *NCF1* (p47-PHOX) gene. In one embodiment, the method is used to correct a mutation in the *NCF1* target gene, and wherein the mutation in the *NCF1* gene causes chronic granulomatous disease. In one embodiment, the endogenous homologous region is in a p47-PHOX pseudogene.

5 In one embodiment, the cell is a blood cell, a neuronal cell, an immune cell, a muscle cell, a stem cell, a progenitor cell, or a diseased cell. In one embodiment, the cell is a mammalian cell, *e.g.*, a human cell.

In one embodiment, the contacting step is performed *ex vivo*. In one embodiment, the contacted cell is returned to the body of the subject. In one embodiment, the contacting is
10 performed *in vivo*.

In one embodiment, the cell is a population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 12% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 15% of the
15 cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 20% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 25% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are
20 repaired by gene conversion in at least 30% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 35% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 40% of the cells in the population of cells.

25 In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 45% of the cells in the population of cells.

In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 12% to about 45% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are
30 repaired by gene conversion about 15% to about 45% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 20% to about 45% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 25% to about 45% of the cells in the population of cells.

In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 12% to about 40% of the cells in the population of cells.

In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 12% to about 30% of the cells in the population of cells.

5 In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion about 12% to about 25% of the cells in the population of cells.

In one embodiment, the cell is a population of hematopoietic stem cells (HSCs). In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 4% of the cells in the population of HSCs. In one embodiment,
10 the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 5% of the cells in the population of HSCs. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in at least 6% of the cells in the population of HSCs. In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion in about 4% to about 6% of the
15 cells in the population of HSCs.

In one embodiment, the first overhang and the second overhang in the target gene are repaired by non-homologous end joining (NHEJ) in less than 40% of the cells in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by non-homologous end joining (NHEJ) in less than 30% of the cells
20 in the population of cells. In one embodiment, the first overhang and the second overhang in the target gene are repaired by non-homologous end joining (NHEJ) in less than 20% of the cells in the population of cells. In one embodiment, the NHEJ results in a deletion in the target gene. In one embodiment, the NHEJ results in an insertion in the target gene.

In one aspect, disclosed herein is a cell, *e.g.*, a mammalian cell, such as a human cell,
25 altered by any of the methods disclosed herein.

In one aspect, disclosed herein is a pharmaceutical composition comprising a cell, *e.g.*, a mammalian cell, such as a human cell, altered by any of the methods disclosed herein.

In another aspect, disclosed herein is a composition comprising: a first non-naturally occurring gRNA molecule, a first non-naturally occurring enzymatically active Cas9
30 (eaCas9) molecule, a second non-naturally occurring gRNA molecule, and a second non-naturally occurring eaCas9 molecule; wherein the first gRNA molecule and the first eaCas9 molecule are designed to associate with a target gene and produce a first single strand cleavage event on a first strand of the target gene; wherein the second gRNA molecule and the second eaCas9 molecule are designed to associate with the target gene and produce a

second single strand cleavage event on a second strand of the target gene, thereby forming a double strand break having a first overhang and a second overhang; and wherein the first gRNA molecule, the first eaCas9 molecule, the second gRNA molecule and the second eaCas9 molecule are designed such that the first overhang and the second overhang in the target gene are repaired by gene conversion.

In one embodiment, the first overhang and the second overhang in the target gene are repaired by gene conversion using an endogenous homologous region.

In one embodiment, the composition does not comprise an exogenous nucleic acid homologous to the target gene.

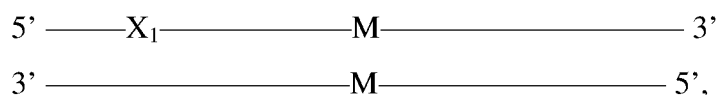
In one embodiment, the first overhang is a 5' overhang, and the second overhang is a 5' overhang.

In one embodiment, the target gene is an *HBB* gene. In another embodiment, the endogenous homologous region is in an *HBD* gene.

In one embodiment, the method is used to correct a mutation in the target gene, wherein the mutation is located between the first single strand break and the second single strand break. In one embodiment, the method is used to correct a mutation in the target gene, wherein the mutation is located within fewer than 50 nucleotides of the first single strand break. In one embodiment, the method is used to correct a mutation in the target gene, wherein the mutation is located within fewer than 50 nucleotides of the second single strand break.

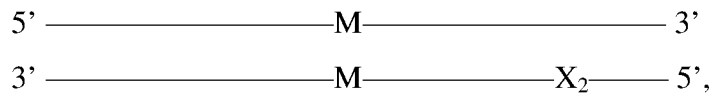
In one embodiment, the first gRNA molecule is designed to position the first single strand cleavage event 5' to a mutation on a first strand of the *HBB* target gene. In one embodiment, the first gRNA molecule is designed to position the first single strand cleavage event 5' to a mutation on a first strand of the *HBB* target gene, as shown in the diagram

below:



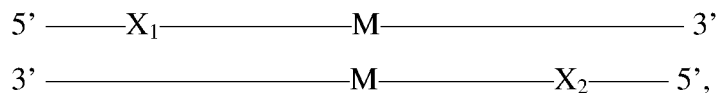
wherein X_1 is the first single strand cleavage event and M is the mutation in the *HBB* target gene.

In one embodiment, the second gRNA molecule is designed to position the second single strand cleavage event 5' to a mutation on a second strand of the *HBB* target gene. In one embodiment, the second gRNA molecule is designed to position the second single strand cleavage event 5' to a mutation on a second strand of the *HBB* target gene, as shown in the diagram below:



wherein X_2 is the second single strand cleavage event and M is the mutation in the *HBB* target gene.

5 In one embodiment, the first gRNA molecule is designed to position the first single strand cleavage event 5' to a mutation on the first strand of the *HBB* target gene, and wherein the second gRNA molecule is designed to position the second single strand cleavage event 5' to the mutation on the second strand of the *HBB* target gene. In one embodiment, the first gRNA molecule is designed to position the first single strand cleavage event 5' to a mutation
 10 on the first strand of the *HBB* target gene, and wherein the second gRNA molecule is designed to position the second single strand cleavage event 5' to the mutation on the second strand of the *HBB* target gene, as shown in the diagram below:



15 wherein X_1 is first single strand cleavage event, X_2 is the second single strand cleavage event, and M is the mutation in the *HBB* target gene.

In one embodiment, the first single strand cleavage event results in a first 5' overhang, and wherein the second single strand cleavage event results in a second 5' overhang.

In one embodiment, the endogenous target gene has at least 80% sequence homology
 20 with the endogenous homologous region. In one embodiment, the endogenous target gene has at least 85% sequence homology with the endogenous homologous region. In one embodiment, the endogenous target gene has at least 92% sequence homology with the endogenous homologous region. In one embodiment, the endogenous target gene has at least 95% sequence homology with the endogenous homologous region. In one embodiment, the
 25 endogenous target gene has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence homology with the endogenous homologous region.

In one embodiment, the first eaCas9 molecule is a first nickase molecule and the second eaCas9 molecule is a second nickase molecule. In one embodiment, each nickase
 30 molecule forms a single strand break in the target gene. In one embodiment, the first eaCas9 nickase molecule and the second eaCas9 nickase molecule are the same species of nickase molecule.

In one embodiment, the eaCas9 nickase molecule comprises HNH-like domain cleavage activity but has no N-terminal RuvC-like domain cleavage activity. In one

embodiment, the eaCas9 molecule is an HNH-like domain nickase. In one embodiment, the eaCas9 nickase molecule comprises a mutation at an amino acid position corresponding to amino acid position D10 of *Streptococcus pyogenes* Cas9.

In one embodiment, the first gRNA molecule is a gRNA molecule comprising a sequence set forth as any one of SEQ ID NOs: 387-485, 6803-6871, or 16010-16256, and the second gRNA molecule is a gRNA molecule comprising a sequence set forth as any one of SEQ ID NOs: 387-485, 6803-6871, or 16010-16256. In one embodiment, the first gRNA molecule is a gRNA molecule comprising a sequence of SEQ ID NO:387, and the second gRNA molecule is a gRNA molecule comprising a sequence of SEQ ID NO:16318.

In one embodiment, the composition is for use in a medicament. In another embodiment, the composition is for use in the treatment of spinal muscular atrophy. In one embodiment, the composition is for use in the treatment of chronic granulomatous disease. In one embodiment, the composition is for use in the treatment of sickle cell disease. In one embodiment, the composition is for use in the treatment of beta thalassemia.

In another aspect, disclosed herein is a method of modifying a target region of a target gene in a mammalian cell comprising generating, within the cell, a first single strand break on a first strand of the target gene and second single strand break on a second strand of the target gene, thereby forming a double strand break with a first 5' overhang and a second 5' overhang; and wherein the double strand break is repaired by gene conversion, thereby modifying the target region of the target gene in the mammalian cell.

In one embodiment, the target region of the target gene is located (a) between the first single strand break and the second single strand break, (b) within fewer than 50 nucleotides of the first single strand break, or (c) within fewer than 50 nucleotides of the second single strand break.

In another aspect, disclosed herein is a method of modifying a target region of a target gene in a mammalian cell comprising generating, within the cell, a first single strand break on a first strand of the target gene and second single strand break on a second strand of the target gene, thereby forming a double strand break with a first 5' overhang and a second 5' overhang; wherein the target region of the target gene is located (a) between the first single strand break and the second single strand break, (b) within fewer than 50 nucleotides of the first single strand break, or (c) within fewer than 50 nucleotides of the second single strand break; and wherein the double strand break is repaired by gene conversion, thereby modifying the target region of the target gene in the mammalian cell.

In one embodiment, the step of forming the first single strand break and the second single strand break comprises contacting the cell with at least one eaCas9 molecule, a first gRNA molecule, and a second gRNA molecule. In one embodiment, the first gRNA molecule and the at least one eaCas9 molecule associate with the target gene and produce the first single strand break, and the second gRNA molecule and the at least one eaCas9 molecule associate with the target gene and produce the second single strand break.

In one embodiment, the double strand break is repaired by gene conversion using an endogenous homologous region.

In one embodiment, the target gene is an *HBB* gene, and wherein the endogenous homologous region is a region of an *HBD* gene. In one embodiment, the first gRNA molecule comprises a targeting sequence comprising SEQ ID NO:387 and the second gRNA molecule comprises a targeting sequence comprising SEQ ID NO:16318.

In one embodiment, the at least one eaCas9 molecule is at least one eaCas9 nickase molecule. In one embodiment, the at least one eaCas9 nickase molecule is at least one HNH-type nickase molecule.

In one embodiment, the endogenous homologous region of the endogenous gene is located within a gene cluster comprising the target gene.

In another embodiment, disclosed herein is a method of increasing the percentage of cells in a population of cells that modify a target region of a target gene by gene conversion using an endogenous homologous region, the method comprising contacting the population of cells with a first gRNA molecule, a first enzymatically active Cas9 (eaCas9) molecule, a second gRNA molecule, and a second eaCas9 molecule; wherein the first gRNA molecule and the first eaCas9 molecule associate with the target gene and generate a first single strand cleavage event on a first strand of the target gene; wherein the second gRNA molecule and the second eaCas9 molecule associate with the target gene and generate a second single strand cleavage event on a second strand of the target gene, thereby forming a double strand break with a first 5' overhang and a second 5' overhang; and wherein the first 5' overhang and the second 5' overhang in the target gene are repaired by gene conversion using the endogenous homologous region, thereby increasing the percentage of cells in the population of cells that modify the target region of the target gene by gene conversion using the endogenous homologous region.

In one embodiment, the first 5' overhang and the second 5' overhang in the target gene are repaired by gene conversion in about 12% to about 45%, about 15% to about 30%, or about 15% to about 40% of the cells in the population of cells. In one embodiment, the

cells are HEK293 cells. In one embodiment, the cells are K562 cells. In one embodiment, the cells are U2OS cells. In another embodiment, the cells are blood mononuclear cells, *e.g.*, CD3+ T lymphocyte cells.

5 In one embodiment, the first 5' overhang and the second 5' overhang in the target gene are repaired by gene conversion in about 4% to about 6% of cells in the population of cells, wherein the cells are HSCs, *e.g.*, CD34+ cells.

10 In one embodiment, the percentage of cells in the population of cells that modify the target region of the target gene by gene conversion using the endogenous homologous region is increased, as compared to a percentage of cells in a population of cells which would modify the target region of the target gene by gene conversion using the endogenous homologous region in the absence of the contacting.

15 In one embodiment, the percentage of cells in the population of cells that modify the target region of the target gene by gene conversion using the endogenous homologous region is increased, as compared to a percentage of cells in a population of cells which would modify the target region of the target gene by gene conversion using the endogenous homologous region after contacting the population of cells with a first gRNA molecule, a first enzymatically active Cas9 (eaCas9) molecule, a second gRNA molecule, and a second eaCas9 molecule which would generate a first 3' overhang and a second 3' overhang in the target gene.

20 In one embodiment, the percentage of cells in the population of cells that modify the target region of the target gene by gene conversion using the endogenous homologous region is increased, as compared to a percentage of cells in a population of cells which would modify the target region of the target gene by gene conversion using the endogenous homologous region after contacting the population of cells with a first gRNA molecule, a first enzymatically active Cas9 (eaCas9) molecule, a second gRNA molecule, and a second
25 eaCas9 molecule which would produce a double strand blunt end break in the target gene.

30 In one embodiment, the percentage of cells in the population of cells that modify the target region of the target gene by gene conversion using the endogenous homologous region is increased about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, 100%, two-fold, three-fold, four-fold, or five-fold. In one embodiment, the percentage of cells in the population of cells that modify the target region of the target gene by gene conversion using the endogenous homologous region is increased two-fold. In one embodiment, the percentage of cells in the population of cells that modify the target region of the target gene by gene conversion using the endogenous homologous region is increased three-fold. In one

embodiment, the percentage of cells in the population of cells that modify the target region of the target gene by gene conversion using the endogenous homologous region is increased four-fold.

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

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 the pair 8/15 of gRNAs surrounding the sickle mutation in combination with a Cas9 nickase (D10A or N863A). The nickases are shown as the grey ovals.

Fig. 2 depicts the percentages of total editing event after a wild type Cas9 or a Cas9 nickase (D10A or N863A). A repetition of at least three independent experiments for each condition is shown.

Fig. 3A depicts the frequency of deletions a wild type Cas9 or a Cas9 nickase (D10A or N863A). A representation of at least 3 independent experiments for each condition is shown.

Fig. 3B depicts the frequency distribution of the length of deletions using a wild type Cas9 and gRNA 8 (similar results have been obtained with gRNA 15).

Fig. 3C depicts the frequency distribution of the length of deletions using a Cas9 nickase (D10A) with gRNAs 8/15 (similar results have been obtained using Cas9 N863A).

Fig. 4A depicts the frequency of gene conversion using a wild type Cas9 or a Cas9 nickase (D10A or N863A).

Fig. 4B shows a scheme representing the region of similarity between the *HBB* and *HBD* loci.

Fig. 5 depicts the frequency of different lengths of *HBD* sequences that were incorporated into the *HBB* locus.

Fig. 6A depicts the frequency of insertions using a wild type Cas9 or a Cas9 nickase (D10A or N863A). A representation of at least three independent experiments for each condition is shown.

Fig. 6B depicts examples of common reads observed in U2OS cells electroporated with plasmid encoding Cas9 N863 and gRNA pair 8/15. The *HBB* reference is shown on the top.

Fig. 7A depicts the overall modification frequency at the *HBB* locus, separated into deletions, insertions, and gene conversion events. The different repair outcomes after WT (wild type), Cas9 D10 nickase or Cas9 N863A nickase activity in U2OS cells were measured by PCT amplification of individual amplification products. Error bars were derived from 5 independent experiments. Total number of sequences analyzed: n=569 for WT Cas9, n=718 for Cas9 D10A nickase, n=556 for Cas9 N863A nickase

Fig. 7B depicts the deletion size distribution for WT, Cas9 D10A nickase, or Cas9 N863A nickase. Total number of sequences analyzed: n=258 for WT Cas9, n=241 for Cas9 D10A nickase, n=103 for Cas9 N863A nickase. Individual deletion size data from each Cas9 variant was scored from Sanger sequencing.

Fig. 8A depicts the overall modification frequency separated into deletions, insertions and gene conversion events for either WT Cas9- (WT), Cas9 D10A nickase- (D10A), or Cas9 N863A nickase-nucleofected (N863A) population of U2OS cells (bulk) or individual single U2OS cell clones. For WT Cas9, 58 sequences were analyzed for populations of cells (bulk) and 92 individual clones were analyzed. For Cas9 D10A nickase, 62 sequences were analyzed for populations of cells (bulk) and 81 individual clones were analyzed. For Cas9 N863A nickase, 73 sequences were analyzed for populations of cells (bulk) and 96 individual clones were analyzed.

Fig. 8B depicts the overall modification frequency separated into deletions, insertions, and gene conversion events for either WT Cas9- (WT), Cas9 D10A nickase- (D10A), or Cas9 N863A nickase-nucleofected (N863A) K562 cells. Number of sequences analyzed: n=91 for WT Cas9, n=85 for Cas9 D10A nickase, n=131 for Cas9 N863A nickase.

Fig. 9 graphically depicts the positions of nucleotide mismatches (thin lines) and deletions (thick line) between the *HBD* and *HBB* genes and a histogram of the relative gene conversion frequency plotted as a function of the position on the *Hbb* locus.

Fig. 10A depicts western blot analyses of the knockdown efficiency of nucleofection of U2OS cells with siRNAs directed against firefly luciferase (FF, control), BRCA2, or RAD51 genes, as well as gel-loading controls. For BRCA2 knockdown, the gel loading control was β -actin. For RAD51 knockdown, the gel loading control was vinculin.

Fig. 10B depicts the overall modification frequency at the *HBB* locus separated into deletions, insertions, and gene conversion as determined by Sanger sequencing. U2OS cells nucleofected with siRNA against firefly luciferase (FF, control), BRCA2, or RAD51 genes, in addition to the Cas9 D10A nickase and gRNA pair 8/15. p-values for the differences in gene conversion frequency were calculated using the two-tailed Student's t-test. Four

independent experiments were performed. Total number of sequences analyzed: n=476 for FF control siRNA treated cells, n=407 for BRCA2 siRNA treated cells, n=389 for RAD51 siRNA treated cells.

Fig. 10C depicts the overall modification frequency at the *HBB* locus resolved for deletions, insertions, and gene conversion as determined by Sanger sequencing. U2OS cells nucleofected with siRNA against firefly luciferase (FF, control), BRCA2, or RAD51 genes, in addition to the Cas9 D10A nickase and gRNA pair 8/15, in the presence of 50 pM of donor template. p-values for the differences in gene conversion frequency were calculated using the two-tailed Student's t-test. Total number of sequence from at least four independent experiments. Total number of sequences analyzed: n=770 for FF control siRNA treated cells, n=652 for BRCA2 siRNA treated cells, n=389 for RAD51 siRNA treated cells.

Fig. 11A depicts the raw number and percentage of insertions detected in individual cells nucleofected with the Cas9 D10A nickase or Cas9 N863A nickase and gRNA pair 8/15 that were either derived from the overhang, or that contain the full overhang repetition.

Overhangs were defined as sequences that had at least a consecutive stretch of 9 nucleotides of the predicted overhang sequence present. P-value was calculated using Fisher's exact test.

Fig. 11B depicts histogram plots of insertion length for either the Cas9 D10A nickase- or Cas9 N863A nickase-induced lesions. Individual insertion size data from each Cas9 variant was scored from Sanger sequencing data of U2OS cells treated with either Cas9 10A nickase or Cas9 N863A nickase and gRNA pair 8/15. Difference in length between Cas9 D10A nickase- and Cas9 N863A nickase-induced lesions was significant ($p=1.274 \times 10^{-12}$; permutation test). Number of insertions plotted: n=75 for Cas9 D10A nickase, and n=325 for Cas9 N863A nickase.

Fig. 12A depicts histogram plots showing how frequently and the position in the overhang of 4-mers found within insertion sequences aligned to different positions along the overhang sequence in U2OS cells treated with either Cas9 10A nickase or Cas9 N863A nickase and gRNA pair 8/15. Parts of insertion sequences containing the 4-mer sequence could not be uniquely mapped to a single position within the overhang.

Fig. 12B depicts exemplary insertion sequences of the full overhang resulting from Cas9 D10A nickase or Cas9 N863A nickase-induced lesions. Cas9 D10 nickase-induced insertions did not show overlaps, while Cas9 N863A nickase-induced insertions showed overlaps, indicating microhomology usage. Positioning of each gRNA (gRNA8 or gRNA15) shown as indicated. Position of the PAM indicated in black.

Fig. 13A is a schematic representation of the donor template.

Fig. 13B depicts the frequency of HDR using a wild type Cas9 or a Cas9 nickase (D10A or N863A).

Fig. 13C depicts different forms of donors and their contribution to HDR.

Fig. 14 depicts the overall modification frequency at the *HBB* locus separated into repair outcomes (*i.e.*, deletions, insertions, gene conversion, and gene correction) for either WT Cas9- (WT), Cas9 D10A nickase- (D10A), or Cas9 N863A nickase-nucleofected (N863A) U2OS cells in the presence of 50 pM ssODN donor template. Repair outcomes were determined by PCR amplification of the *HBB* locus, followed by Sanger sequencing of individual amplification products. p-value for the difference in gene correction frequency was calculated using the two-tailed Student's t-test. Total number of sequences plotted from at least three independent experiments: n=200 for WT Cas9, n=448 for Cas9 D10A nickase, and n=422 for Cas9 N863A nickase.

Fig. 15 depicts the distribution of the editing events on the presence of wild type Cas9 with gRNA 8, gRNA 15 and with the Cas9 nickase (D10A) with gRNA 8/15 pair.

Fig. 16A depicts the overall modification frequency of U2OS cells nucleofected with either WT Cas9, Cas9 D10 nickase, or Cas9 N863A nickase and gRNA8. Sequencing was performed using Illumina MiSeq and modifications were quantified computationally using CRISPResso.

Fig. 16B depicts a 15% urea polyacrylamide gel electrophoresis displaying the cutting efficiency of various amounts of Cas9 D10A nickase and gRNA8 or Cas9 N863A nickase and gRNA8 complexes with double stranded DNA. Cleavage products are indicated.

Fig. 17A depicts the frequency of deletions, insertions, gene conversion, and gene correction observed in U2OS cells nucleofected with either WT Cas9, Cas9 D10A nickase, or Cas9 N863A nickase, gRNA8, donor template and siRNAs against firefly luciferase (FF, control), BRCA2, or RAD51. Sequencing was performed using MiSeq and modifications were scored by computationally using CRISPResso.

Fig. 18A depicts a bar graph of fold change in the rates of total editing events and gene correction for U2OS cells nucleofected with either WT Cas9, Cas9 D10A nickase, or Cas9 N863A nickase, gRNA8, donor template and siRNAs against RAD51.

Fig. 18B depicts histograms of the length of deletions induced in U2OS cells nucleofected with either WT Cas9, or Cas9 D10A nickase, gRNA8, donor template and siRNAs against firefly luciferase (FF, control) or RAD51.

Fig. 18C depicts a model of the possible role of HR in the processing of nicks that escaped single strand break repair in the presence or absence of RAD51.

Fig. 19 depicts genome editing of the *HBB* locus in bone marrow leukemia K562 hematopoietic cells after electroporation of Cas9 protein complexed to HBB gRNAs 8 and 15 (RNP) or Cas9 mRNA co-delivered with HBB gRNAs 8 and 15 (RNA). “gRNA 8” has the targeting domain sequence GUAACGGCAGACUUCUCCUC (SEQ ID NO: 388) and
 5 “gRNA 15” has the targeting domain sequence AAGGUGAACGUGGAUGAAGU (SEQ ID NO: 387).

Fig. 20 is a schematic of the *HBB* locus indicating the position of the different gRNA molecules used in the study.

Fig. 21 depicts the overall editing frequencies for the Cas9 D10A and Cas9 N863A
 10 nickases in terms of overhang length and composition.

Fig. 22 depicts the insertion frequencies for the Cas9 D10A and Cas9 N863A nickases in terms of overhang length and composition.

Fig. 23 depicts the percentages of insertions derived from the overhang sequence for the Cas9 D10A (left bars) and Cas9 N863 (right bars) nickases.

Fig. 24 depicts the lengths of insertions for the Cas9 D10A (left bars) and Cas9 N863
 15 (right bars) nickases.

Fig. 25 depicts the effect of overhang length and composition on gene conversion induced by the Cas9 D10A and Cas9 N863 nickases.

Fig. 26 depicts the effect of overhang length and composition on gene correction
 20 induced by the Cas9 D10A and Cas9 N863A nickases.

Figs. 27A and 27B depict an analysis of stability of D10A RNP *in vitro* and *ex vivo* in adult CD34+ HSCs. **Fig. 27A** depicts Differential Scanning Fluorimetry Shift Assay after complexing D10A protein with the indicated *HBB* gRNAs added at 1:1 molar ratio gRNA:RNP. **Fig. 27B** depicts detection of Cas9 protein in cell lysates 72 hours after HSCs
 25 were electroporated with D10A RNP or D10A mRNA with gRNAs HBB-8 and HBB-15. The electroporation program (P2 or P3) used is indicated at the top of the image. 2X gRNA: 10µg each gRNA was co-delivered with D10A mRNA (vs. 5µg of each gRNA).

Figs. 28A, 28B, and 28C depict adult CD34+ HSCs maintain stem cell phenotype after electroporation with D10A nickase RNP and HBB targeting gRNA pair. **Fig 28A**
 30 depicts gene edited adult CD34+ cells maintain expression of stem cell markers CD34 and CD133 at 72 hours after electroporation. **Fig 28B** depicts absolute live (7-AAD-AnnexinV) CD34+ cell number at indicated time points relative to electroporation of D10A RNP HBB gRNA pair. **Fig 28C** depicts gene edited adult CD34+ cells maintain hematopoietic colony forming cell (CFC) activity and multipotency. E: erythroid, G: granulocyte, M: macrophage,

GM: granulocyte-macrophage, GEMM: granulocyte-erythrocyte-macrophage-monocyte CFCs.

Figs. 29A, 29B, and 29C depict D10A nickase RNP co-delivered with HBB targeted gRNA pair supports gene editing and HDR in adult CD34⁺ HSCs. **Fig. 29A** depicts percentage of gene editing events as detected by T7E1 endonuclease assay analysis. **Fig. 29B** depicts DNA sequence analysis of the *HBB* locus. The subtypes of gene editing events (insertions, deletions, indels, and gene conversion events) are indicated. RNP* refers to use of electroporation program 3. **Fig. 29C** depicts percentages of types of editing events detected in the gDNA from the cells electroporated with the conditions shown in **Fig. 29B**. Data are shown as a percentage of all gene editing events.

Fig. 30 depicts flow cytometry analysis of β -hemoglobin expression in the erythroid progeny differentiated from D10A nickase RNP gene-edited adult CD34⁺ HSCs. CFU-E colonies (far left) differentiated from D10A RNP HBB gRNA electroporated CD34⁺ cells were dissociated, fixed, permeabilized, and stained for β -hemoglobin expression. The gene editing frequencies detect in the parental CD34⁺ cell population are indicated above the histograms for the indicated samples. The percentage of β -hemoglobin expression in each colony was determined by flow cytometry and is indicated at the top right of each histogram.

Figs. 31A and 31B depict that cord blood CD34⁺ HSCs maintained stem cell phenotype after electroporation with D10A nickase RNP and HBB targeting gRNA pair. **Fig. 31A** depicts gene edited CB CD34⁺ cells maintain viability after electroporation. *Right:* Absolute live (7-AAD⁻AnnexinV⁻) CD34⁺ cell number at indicated time points relative to electroporation of D10A RNP *HBB* gRNA pair. **Fig. 31B** depicts that gene edited CB CD34⁺ cells maintained hematopoietic colony forming cell (CFC) activity and multipotency. E: erythroid, G: granulocyte, M: macrophage, GM: granulocyte-macrophage, GEMM: granulocyte-erythrocyte-macrophage-monocyte CFCs. The amounts of D10A RNP delivered per million cells (5 or 10 μ g) and the 2-hour recovery temperature (parentheses) after electroporation of the parental CB CD34⁺ cells are indicated.

Figs. 32A, 32B, and 32C depict D10A RNP co-delivered with HBB targeted gRNA pair supported gene editing and HDR in CB CD34⁺ HSCs. **Fig. 32A** depicts percentage of gene editing events as detected by T7E1 endonuclease assay analysis. **Fig. 32B** depicts DNA sequence analysis of the *HBB* locus. The subtypes of gene editing events (insertions, deletions, indels, and gene conversion events) are indicated as a fraction of the total sequencing reads. **Fig. 32C** depicts subtypes of gene editing events expressed as relative

percentage to the total number gene editing events detected. The amounts of D10A RNP delivered per million cells (5 or 10 μ g) and the 2-hour recovery temperature (parentheses) after electroporation of the parental CB CD34⁺ cells are indicated.

Figs. 33A, 33B, and 33C depict directed differentiation of gene-edited CB CD34⁺ HSCs into erythroblasts. Flow cytometry analysis of day 18 erythroblasts differentiated from gene edited CB CD34⁺ cells. **Fig. 33A** depicts CD71 (transferrin receptor and CD235 (Glycophorin A). **Fig. 33B** depicts fetal hemoglobin (g-hemoglobin). **Fig. 33C** depicts loss of CD45 and dsDNA through enucleation as indicated by the absence of dsDNA (negative for dsDNA binding dye DRAQ5). Note that, unlike adult CD34⁺ cells, CB CD34⁺ cells differentiated into fetal-like erythroblasts that express fetal g-hemoglobin (not adult b-hemoglobin).

Figs. 34A and 34B depict cord blood CD34⁺ HSCs maintained stem cell phenotype after electroporation with Cas9 variant RNPs and HBB targeting gRNA pair. **Fig. 34A** depicts gene edited CB CD34⁺ cells maintain viability after electroporation with WT Cas9 endotoxin-free (EF WT) Cas9, N863A nickase, or D10A nickase co-delivered with HBB gRNA pair. Absolute live (7-AAD⁻AnnexinV⁻) CD34⁺ cell number at indicated time points relative to electroporation. **Fig. 34B** depicts gene edited CB CD34⁺ cells maintained hematopoietic colony forming cell (CFC) activity and multipotency. E: erythroid, G: granulocyte, M: macrophage, GM: granulocyte-macrophage, GEMM: granulocyte-erythrocyte-macrophage-monocyte CFCs. The amounts of RNP delivered per million cells (10 μ g) and the 2-hour recovery temperature (parentheses) after electroporation of the parental CB CD34⁺ cells are indicated.

Figs. 35A and 35B depict comparison of gene editing at the *HBB* locus in CB CD34⁺ cells mediated by WT and nickase Cas9 variant RNPs. **Fig. 35A** depicts T7E1 analysis of the percentage of indels detected 72 hours after electroporation at the targeted site in the *HBB* locus after electroporation of WT Cas9, Endotoxin-free WT Cas9 (EF-WT), N863A nickase, and D10A nickase RNPs, each co-delivered with HBB-8 and HBB-15 gRNA pair. **Fig. 35B** depicts Western blot analysis showing detection of Cas9 variants in cell lysates of CB CD34⁺ cells at the indicated time points after electroporation. The amounts of RNP delivered per million cells (10 μ g) and the 2-hour recovery temperature (parentheses) after electroporation of the parental CB CD34⁺ cells are indicated.

Figs. 36A and 36B depict comparison of HDR and NHEJ events detected at the *HBB* locus after gene editing with WT Cas9 and D10A nickase in CB CD34⁺ HSCs. **Fig. 36A** depicts percentage of gene editing events (72 hours after electroporation) detected by DNA

sequencing analysis and shown as a percentage of the total sequence reads. Cells received RNP (WT Cas9, Endotoxin-free WT Cas9 [EF-WT], N863A and D10A nickases) with HBB-8 and HBB-15 gRNA pair. **Fig. 36B** depicts percentages of types of editing events detected in the gDNA from the cells electroporated with the conditions shown in **Fig. 36A**. Data are shown as a percentage of all gene editing events.

Fig. 37 depicts that down-regulation of Pol Theta in the context of the D10A Cas9 does not affect the gene editing profile. In contrast, Pol Theta down-regulation in the context of a mutant Cas9 (such as the N863A Cas9 nuclease) leads to a strong reduction of the insertion frequency and an increase in the gene conversion rate. Cells were treated with either a control siRNA or an siRNA directed against Pol Theta. The N863A enzyme is directed to make two nicks at the target site resulting in a 3' overhang. In the absence of Pol Theta, the rates of insertion are reduced and gene conversion is increased because the Alt-NHEJ pathway (Pol Theta Dependant) has been downregulated.

Fig. 38 depicts that the gene conversion frequency is highly impaired in the absence of BRCA2 and RAD51. D10A induced gene conversion is dependent on BRCA2 and RAD51. Cells were treated with siRNA against a negative control, BRCA2, and RAD51 in conjunction with D10A and gRNA 8 and 15.

Fig. 39 depicts a model of the genetic requirements of the gene conversion pathway in the context of different Cas9 nucleases (wild type, D10A or N863A) using BRCA2 and RAD51.

Fig. 40A, 40B and **40C** depict that Cas9 D10A nickase RNP co-delivered with a HBB targeted gRNA pair supported gene editing in T lymphocytes expanded from peripheral blood mononuclear cells (MNCs) from a sickle cell disease (SCD) patient. **Fig. 40A** depicts the expansion/activation of the T lymphocyte fraction from bulk MNCs from a SCD patient using CD3/CD28 immunomagnetic beads at day 3 (left panel) and day 7 (right panel) after activation. **Fig. 40B** depicts the total viability of expanded T lymphocytes electroporated with Cas9 D10A nickase complexed with a HBB targeted gRNA pair (HBB-8-sickle and HBB-15), and co-electroporated with a single stranded oligonucleotide donor, following reactivation CD3/CD28 immunomagnetic beads, no beads control (*i.e.*, no exposure CD3/CD28 immunomagnetic beads) or negative control. **Fig. 40C** depicts gene editing frequency at the *HBB* locus of expanded T lymphocytes electroporated with Cas9 D10A nickase complexed with a HBB targeted gRNA pair (HBB-8-sickle and HBB-15), and co-electroporated with a single stranded oligonucleotide donor, following reactivation

CD3/CD28 immunomagnetic beads, no beads control (*i.e.*, no exposure CD3/CD28 immunomagnetic beads) or negative control.

DETAILED DESCRIPTION

5 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
10 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
15 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),
20 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
25 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
30 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).

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.

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 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 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 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 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.

As used herein, the term “Cas9 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 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 some

embodiments, a Cas9 system comprises a gRNA, two Cas9 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 some embodiments, a Cas9 system further comprises a template nucleic acid.

5 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.

10 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
15 disorder, and/or if a study has found that the mutation contributes causally to development of the disorder.

 “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
20 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
25 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
30 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, “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

5 sequence. In an 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 the *HBB* 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 an embodiment, a nucleic acid molecule

10 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

15 nucleic acid, and results in cessation or reduction of the production of a CRISPR/Cas system component. In an embodiment, the Cas9 molecule forms two complexes: a complex comprising a Cas9 molecule with a target gene gRNA molecule, which complex will alter the *HBB* 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,

20 *e.g.*, a Cas9 molecule or a target gene gRNA molecule. In an 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 sequence that encodes a Cas9 molecule, a sequence that encodes a transcribed region, an exon, or an intron, for the Cas9 molecule. In an embodiment, a governing gRNA molecule/Cas9 molecule complex binds to or promotes

25 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 an embodiment, the governing gRNA 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 an embodiment, a governing gRNA places

30 temporal, level of expression, or other limits, on activity of the Cas9 molecule/target gene gRNA molecule complex. In an embodiment, a governing gRNA reduces off-target or other unwanted activity. In an 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 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) 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
5 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
10 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.

15 “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.

“Homologous recombination” or “HR” refers to a type of HDR DNA-repair which
20 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
25 the homologous nucleic acid is typically double-stranded. In some embodiments, homologous recombination includes gene conversion.

“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 an embodiment,
30 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 an 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. For example, a mutation (*e.g.*, GAG → GTG) results in the substitution of valine for glutamic acid at amino acid position 6 in exon 1 of the *HBB* gene. This mutation in the *HBB* gene is associated with beta thalassemia and sickle cell disease.

“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.

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.

“SCD target point position,” as used herein, refers to a target position in the *HBB* gene, typically a single nucleotide, which, if mutated, can result in a protein having a mutant amino acid and give rise to SCD. In an embodiment, the SCD target position is the target position at which a change can give rise to an E6 mutant protein, *e.g.*, a protein having an E6V substitution.

“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 an 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 conversion, 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. In some embodiments, the target gene is an *HBB* gene. In other embodiments, the target gene is an *SMN1* gene. In some embodiments, the target gene is an *NCF1* (p47-PHOX) gene.

“Target position” as used herein, refers to a site on a target nucleic acid that is modified by a Cas9 molecule-dependent process. For example, the target position can be modified by a Cas9 molecule-mediated cleavage of the target nucleic acid and template nucleic acid directed modification, *e.g.*, correction, of the target position. In an 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 an embodiment, the target position is within a “target sequence” (*e.g.*, the sequence to which the gRNA binds). In an 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 conversion of the target position, wherein the gene conversion uses an endogenous sequence homologous with the target position region as a template.

A “template nucleic acid,” as the 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 an 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 an embodiment, the template nucleic acid is single stranded. In an alternate embodiment, the template nucleic acid is double stranded. In an 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 an embodiment, the template nucleic acid is RNA, *e.g.*, double stranded RNA or single stranded RNA. In an embodiment, the template nucleic acid is encoded on the same vector backbone, *e.g.*, AAV genome, plasmid DNA, as the Cas9 and gRNA. In an embodiment, the template nucleic acid is excised from a vector backbone
 5 *in vivo*, *e.g.*, it is flanked by gRNA recognition sequences. In one embodiment, the template DNA is in an ILDV. In one embodiment, the template nucleic acid is an exogenous nucleic acid sequence. In another embodiment, the template nucleic acid sequence is an endogenous nucleic acid sequence, *e.g.*, an endogenous homologous region. In one embodiment, the template nucleic acid is a single stranded oligonucleotide corresponding to a plus strand of a
 10 nucleic acid sequence. In another embodiment, the template nucleic acid is a single stranded oligonucleotide corresponding to a minus strand of a nucleic acid sequence.

“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
 15 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
 20 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
 25 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
 30 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 conversion. In an 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.

Sickle Cell Disease and Methods of Repairing Mutation(s) in the *HBB* Gene

Sickle Cell Disease (SCD), also known as Sickle Cell Anemia (SCA), is a common inherited hematologic disease which affects 80,000 - 90,000 people in the United States. It is common in people of African descent and in Hispanic-Americans, with the prevalence of SCD being 1 in 500 and 1 in 1,000, respectively.

SCD is caused by a mutation in the *beta-globin (HBB)* gene. *HBB* is located on chromosome 11 within the *HBB* gene cluster, which includes genes encoding the delta globin chain, A gamma chain, G gamma chain. The alpha-globin gene is located on chromosome 16. A point mutation (*e.g.*, GAG → GTG) results in the substitution of valine for glutamic acid at amino acid position 6 in exon 1 of the *HBB* gene. Beta hemoglobin chains with this mutation are expressed as HbS. The disease is inherited in an autosomal recessive manner, so that only patients with two HbS alleles have SCD. Subjects who have sickle cell trait (are heterozygous for HbS) only display a phenotype if they are severely dehydrated or oxygen deprived.

Normal adult hemoglobin (Hb) is composed of a tetramer made from two alpha-globin chains and two beta-globin chains. In SCD, the valine at position 6 of the beta-chain is hydrophobic and causes a change in conformation of the beta-globin protein when it is not bound to oxygen. HbS is more likely to polymerize and leads to the characteristic sickle shaped red blood cells (RBCs) found in SCD.

Sickle shape RBCs cause multiple manifestations of disease, which include, *e.g.*, anemia, sickle cell crises, vaso-occlusive crises, aplastic crises and acute chest syndrome. The disease has various manifestations, *e.g.*, vaso-occlusive crisis, splenic sequestration crisis and anemia. Subjects may also suffer from acute chest crisis and infarcts of extremities, end organs and central nervous system. Current treatments for SCD include, *e.g.*, hydration, transfusion, analgesics, the use of hydroxyurea, supplementation with folic acid, penicillin prophylaxis during childhood, and bone marrow transplants. However, there remains a need

for additional methods and compositions that can be used to cure, not just treat, sickle cell disease and other genetic diseases.

One approach to treat or prevent SCD is to repair (*i.e.*, correct) one or more mutations in the *HBB* gene. In this approach, mutant *HBB* allele(s) are corrected and restored to wild type state. While not wishing to be bound by theory, it is believed that correction of the glutamic acid to valine substitution at amino acid 6 in the beta-globin gene restores wild type beta-globin production within erythroid cells. The instant disclosure provides methods for modifying target genes, such as *HBB*, for repair by gene conversion using an endogenous homologous region, such as a region of the *HBD* gene. Specifically, the instant disclosure increases the rates of repair by gene conversion as compared to previously disclosed technologies. The methods described herein can be performed in all cell types, including mammalian and human cells. Specifically with respect to SCD, beta-globin is expressed in cells of erythroid cell lineage. Thus, in one embodiment, an erythroid cell is targeted.

While much of the disclosure herein is presented in the context of the mutation in the *HBB* gene that gives rise to an E6 mutant protein (*e.g.*, E6V mutant protein), the methods and compositions herein are broadly applicable to any mutation, *e.g.*, a point mutation or a deletion, in any gene, including promoters and control regions of a gene, that gives rise to any disease.

In an embodiment, one *HBB* allele is repaired in the subject. In another embodiment, both *HBB* alleles are repaired in the subject. In either situation, the subject can be cured of disease. As the disease only displays a phenotype when both alleles are mutated, repair of a single allele is adequate for a cure.

In one aspect, methods and compositions discussed herein, provide for the correction of the underlying genetic cause of SCD, *e.g.*, the correction of a mutation at a target position in the *HBB* gene, *e.g.*, correction of a mutation at amino acid position 6, *e.g.*, an E6V substitution in the *HBB* gene.

In an embodiment, the method provides for the correction of a mutation at a target position in the *HBB* gene, *e.g.*, correction of a mutation at amino acid position 6, *e.g.*, an E6V substitution in the *HBB* gene. As described herein, in one embodiment, the method comprises the introduction of one or more breaks (*e.g.*, single-strand breaks or double-strand breaks) sufficiently close to (*e.g.*, either 5' or 3' to) the target position in the *HBB* gene, *e.g.*, E6V.

In an embodiment, the targeting domain of the gRNA molecule is configured to provide a cleavage event, *e.g.*, a double-strand break or a single-strand break, sufficiently close to (*e.g.*, either 5' or 3' to) the target position in the *HBB* gene, *e.g.*, E6V to allow

correction, *e.g.*, an alteration in the *HBB* gene, *e.g.*, an alteration associated with HDR. In an embodiment, the 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, 200, 250, 300, 350, 400, 450 or 500 nucleotides of a the target position in the *HBB* gene, *e.g.*, E6V. The break, *e.g.*, a double-strand or single-strand break, can be positioned upstream or downstream of the target position in the *HBB* gene, *e.g.*, E6V.

In an embodiment, a second, third and/or fourth gRNA molecule is configured to provide a cleavage event, *e.g.*, a double-strand break or a single-strand break, sufficiently close to (*e.g.*, either 5' or 3' to) the target position in the *HBB* gene, *e.g.*, E6V to allow correction, *e.g.*, an alteration associated with HDR in the *HBB* gene. In an embodiment, the 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, 200, 250, 300, 350, 400, 450 or 500 nucleotides of a the target position in the *HBB* gene, *e.g.*, E6V. The break, *e.g.*, a double-strand or single-strand break, can be positioned upstream or downstream of the target position in the *HBB* gene, *e.g.*, E6V.

In an embodiment, a single-strand break is accompanied by an additional single-strand break, positioned by a second, third and/or fourth 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, 200, 250, 300, 350, 400, 450 or 500 nucleotides of the target position in the *HBB* gene, *e.g.*, E6V. In an 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 molecule, sufficiently close to one another to result in an alteration of the target position in the *HBB* gene, *e.g.*, E6V. In an 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 an 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 an embodiment, a double-strand break can be accompanied by an additional double-strand break, positioned by a second, third and/or fourth gRNA molecule, as is discussed below. For example, the targeting domain of a first gRNA molecule is configured such that a double-strand break is positioned upstream of the target position in the *HBB* gene,

e.g., E6V, *e.g.*, within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450 or 500 nucleotides of the target position; and the targeting domain of a second gRNA molecule is configured such that a double-strand break is positioned downstream the target position in the *HBB* gene, *e.g.*, E6V, *e.g.*, within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450 or 500 nucleotides of the target position.

In an embodiment, a double-strand break can be accompanied by two additional single-strand breaks, positioned by a second gRNA molecule and a third gRNA molecule. For example, the targeting domain of a first gRNA molecule is configured such that a double-strand break is positioned upstream of the target position in the *HBB* gene, *e.g.*, E6V, *e.g.*, within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450 or 500 nucleotides of the target position; and the targeting domains of a second and third gRNA molecule are configured such that two single-strand breaks are positioned downstream of the target position in the *HBB* gene, *e.g.*, E6V, *e.g.*, within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450 or 500 nucleotides of the target position. In an embodiment, the targeting domain of the first, second and third gRNA molecules are configured such that a cleavage event, *e.g.*, a double-strand or single-strand break, is positioned, independently for each of the gRNA molecules.

In an embodiment, 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. For example, the targeting domain of a first and second gRNA molecule are configured such that two single-strand breaks are positioned upstream of the target position in the *HBB* gene, *e.g.*, E6V, *e.g.*, within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450 or 500 nucleotides of the target position in the *HBB* gene, *e.g.*, E6V; and the targeting domains of a third and fourth gRNA molecule are configured such that two single-strand breaks are positioned downstream of the target position in the *HBB* gene, *e.g.*, E6V, *e.g.*, within 1, 2, 3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 150, 200, 250, 300, 350, 400, 450 or 500 nucleotides of the target position in the *HBB* gene, *e.g.*, E6V.

In an embodiment, a mutation in the *HBB* gene, *e.g.*, E6V is corrected using an exogenously provided template nucleic acid, *e.g.*, by HDR. In another embodiment, a mutation in the *HBB* gene, *e.g.*, E6V is corrected without using an exogenously provided template nucleic acid, *e.g.*, by HDR. In an embodiment, alteration of the target sequence occurs with an endogenous genomic donor sequence, *e.g.*, by HDR. In an embodiment, the

endogenous genomic donor sequence comprises one or more nucleotides derived from the *HBD* gene. In an embodiment, a mutation in the *HBB* gene, *e.g.*, E6V is corrected by an endogenous genomic donor sequence (*e.g.*, an *HBD* gene). In an embodiment, an eaCas9 molecule, *e.g.*, an eaCas9 molecule described herein, is used. In an embodiment, the eaCas9 molecule comprises HNH-like domain cleavage activity but has no, or no significant, N-terminal RuvC-like domain cleavage activity. In an embodiment, the eaCas9 molecule is an HNH-like domain nickase. In an embodiment, the eaCas9 molecule comprises a mutation at D10 (*e.g.*, D10A). In an embodiment, the eaCas9 molecule comprises N-terminal RuvC-like domain cleavage activity but has no, or no significant, HNH-like domain cleavage activity. In an embodiment, the eaCas9 molecule is an N-terminal RuvC-like domain nickase. In an embodiment, the eaCas9 molecule comprises a mutation at H840 (*e.g.*, H840A) or N863 (*e.g.*, N863A).

Methods to Treat or Prevent Sickle Cell Disease (SCD)

Disclosed herein are the approaches to treat or prevent SCD, using the compositions and methods described herein.

One approach to treat or prevent disease, *e.g.*, SCD, is to repair (*i.e.*, correct) one or more mutations in a target gene, *e.g.*, a *HBB* gene, *e.g.*, by gene conversion, using an endogenous homologous region. In this approach, mutant *HBB* allele(s) are corrected and restored to wild type state. While not wishing to be bound by theory, it is believed that correction of the glutamic acid to valine substitution at amino acid 6 in the beta-globin gene restores wild type beta-globin production within erythroid cells. The method described herein can be performed in all cell types. Beta-globin is expressed in cells of erythroid cell lineage. In an embodiment, an erythroid cell is targeted.

In an embodiment, one *HBB* allele is repaired in the subject. In another embodiment, both *HBB* alleles are repaired in the subject. In either situation, the subjects can be cured of disease. As the disease only displays a phenotype when both alleles are mutated, repair of a single allele is adequate for a cure.

In an embodiment, the method comprises initiating treatment of a subject prior to disease onset.

In an embodiment, the method comprises initiating treatment of a subject after disease onset.

In an embodiment, the method comprises initiating treatment of a subject well after disease onset, *e.g.*, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 12, 16, 24, 36, 48 or more months after onset of disease. While not wishing to be bound by theory it is believed that this treatment may be effective if subjects present well into the course of illness.

5 In an embodiment, the method comprises initiating treatment of a subject in an advanced stage of disease.

Overall, initiation of treatment for subjects at all stages of disease is expected to prevent negative consequences of disease and be of benefit to subjects.

10 In an embodiment, the method comprises initiating treatment of a subject prior to disease expression. In an embodiment, the method comprises initiating treatment of a subject in an early stage of disease, *e.g.*, when a subject has tested positive for beta-thalassemia mutations but has no signs or symptoms associated with beta-thalassemia major, minor or intermedia.

15 In an embodiment, the method comprises initiating treatment of a subject at the appearance of microcytic anemia, *e.g.*, in an infant, child, adult or young adult.

In an embodiment, the method comprises initiating treatment of a subject who is transfusion-dependent.

In an embodiment, the method comprises initiating treatment of a subject who has tested positive for a mutation in a beta globin gene.

20 In an embodiment, the method comprises initiating treatment at the appearance of any one or more of the following findings associated or consistent with beta-thalassemia major or beta-thalassemia minor: anemia, diarrhea, fever, failure to thrive, frontal bossing, broken long bones, hepatomegaly, splenomegaly, thrombosis, pulmonary embolus, stroke, leg ulcer, cardiomyopathy, cardiac arrhythmia, and evidence of extramedullary erythropoiesis.

25 In an embodiment, a cell is treated, *e.g.*, *ex vivo*. In an embodiment, an *ex vivo* treated cell is returned to a subject.

In an embodiment, allogenic or autologous bone marrow or erythroid cells are treated *ex vivo*. In an embodiment, an *ex vivo* treated allogenic or autologous bone marrow or erythroid cells are administered to the subject. In an embodiment, an erythroid cell, *e.g.*, an autologous erythroid cell, is treated *ex vivo* and returned to the subject. In an embodiment, an autologous stem cell, is treated *ex vivo* and returned to the subject. In an embodiment, the modified HSCs are administered to the patient following no myeloablative pre-conditioning. In an embodiment, the modified HSCs are administered to the patient following mild myeloablative pre-conditioning such that following engraftment, some of the hematopoietic

30

cells are derived from the modified HSCs. In other aspects, the HSCs are administered after full myeloablation such that following engraftment, 100% of the hematopoietic cells are derived from the modified HSCs.

In an embodiment, the method comprises delivery of a gRNA molecule and Cas9 molecule by intravenous injection, intramuscular injection, subcutaneous injection, or intra-bone marrow (IBM) injection.

In an embodiment, the method comprises delivery of a gRNA molecule and/or a Cas9 molecule by an AAV. In an embodiment, the method comprises delivery of a gRNA molecule and/or a Cas9 molecule by a lentivirus. In an embodiment, the method comprises delivery of a gRNA molecule and/or a Cas9 molecule by a nanoparticle. In an embodiment, the method comprises delivery of a gRNA molecule by a parvovirus, *e.g.*, a modified parvovirus specifically designed to target bone marrow cells and/or CD4+ cells. In an embodiment, two or more gRNA molecules (*e.g.*, a second, third or fourth gRNA molecules) are delivered.

Other Methods to Treat Sickle Cell Disease

In healthy individuals, two beta-globin molecules pair with two alpha-globin molecules to form normal hemoglobin (Hb). In SCD, mutations in the *beta-globin (HBB)* gene, *e.g.*, a point mutation (GAG → GTG) that results in the substitution of valine for glutamic acid at amino acid position 6 of the beta-globin molecule, cause production of sickle hemoglobin (HbS). HbS is more likely to polymerize and leads to the characteristic sickle shaped red blood cells (RBCs). Sickle shaped RBCs give rise to multiple manifestations of disease, such as, anemia, sickle cell crises, vaso-occlusive crises, aplastic crises and acute chest syndrome. Alpha-globin can also pair with fetal hemoglobin (HbF), which significantly moderates the severe anemia and other symptoms of SCD. However, the expression of HbF is negatively regulated by the *BCL11A* gene product.

Methods and compositions disclosed herein provide a number of approaches for treating SCD. As is discussed *infra*, methods described herein provide for treating SCD by correcting a target position in the *HBB* gene to provide corrected, or functional, *e.g.*, wild type, beta-globin. Methods and compositions discussed herein can be used in combination with targeting a second gene, *e.g.*, a gene other than the *HBB* gene, *e.g.*, the *BCL11A* gene (also known as B-cell CLL/lymphoma 11A, *BCL11A-L*, *BCL11A-S*, *BCL11A-XL*, *CTIP1*, *HBFGTL5* and *ZNF*), *e.g.*, by a CRISPR/Cas related method. *BCL11A* encodes a zinc-finger protein that is involved in the regulation of globin gene expression. By altering the *BCL11A*

gene (*e.g.*, one or both alleles of the *BCL11A* gene), the levels of gamma globin can be increased. Gamma globin can replace beta globin in the hemoglobin complex and effectively carry oxygen to tissues, thereby ameliorating SCD disease phenotypes. Altering the *BCL11A* gene herein refers to reducing or eliminating (1) *BCL11A* gene expression, (2) *BCL11A* protein function, or (3) the level of *BCL11A* protein. In an embodiment, an SCD target knockdown position is targeted. In another embodiment, an SCD target knockdown position is targeted.

“SCD target knockout position”, as used herein, refers to a position in the *BCL11A* gene, which if altered, *e.g.*, disrupted by insertion or deletion of one or more nucleotides, *e.g.*, by NHEJ-mediated alteration, results in reduction or elimination of expression of functional *BCL11A* gene product. In an embodiment, the position is in the *BCL11A* coding region, *e.g.*, an early coding region. In an embodiment, the position is in the *BCL11A* non-coding region, *e.g.*, an enhancer region.

“SCD target knockdown position”, as used herein, refers to a position, *e.g.*, in the *BCL11A* gene, which if targeted by an eiCas9 or an eiCas9 fusion described herein, results in reduction or elimination of expression of functional *BCL11A* gene product. In an embodiment, transcription is reduced or eliminated. In an embodiment, the position is in the *BCL11A* promoter sequence. In an embodiment, a position in the promoter sequence of the *BCL11A* gene is targeted by an enzymatically inactive Cas9 (eiCas9) or an eiCas9-fusion protein, as described herein.

I. 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
 5 from 5' to 3':

a targeting domain complementary to a target domain in a *HBB* gene, *e.g.*, a
 targeting domain from any of SEQ ID NOs: 387-485, 6803-6871, or
 16010-16256;

a first complementarity domain;

10 a linking domain;

a second complementarity domain (which is complementary to the first
 complementarity domain);

a proximal domain; and

optionally, a tail domain.

15 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
HBB gene), *e.g.*, a targeting domain from any one of

SEQ ID NOs: 387-485, 6803-6871, or 16010-16256; and

20 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

25 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
 30 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 *HBB* target gene. The nucleic acid sequence in a
 target gene, *e.g.*, *HBB*, 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, *e.g.*, *HBB*. 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). Examples of suitable targeting domains for use in the methods, compositions, and kits described herein include those set forth in SEQ ID NOs: 387-485, 6803-6871, or 16010-16256.

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
 5 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
 10 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
 15 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
 20 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
 25 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
 30 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 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 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 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-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 an embodiment, the gRNA molecule, *e.g.*, a gRNA molecule comprising a targeting domain, which is complementary with the *HBB* gene, is a modular gRNA molecule. In another embodiment, the gRNA molecule is a unimolecular or chimeric gRNA molecule.

In an 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 by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence described herein, *e.g.*, a targeting domain sequence from any one of **Tables 1-3**. In an embodiment, the nucleic acid encodes a gRNA molecule comprising a targeting domain described herein, *e.g.*, selected from those in **Tables 1-3**.

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 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.

In certain embodiments, the targeting domain which is complementary with the *HBB* 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 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.

In an 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 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 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.

5

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*.

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 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 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 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 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 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).

5 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
10 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
15 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
20 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,
25 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.

30 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 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 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.

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 complementary. 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

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.

In certain embodiments, the 5' extension domain nucleotides do not comprise
 5 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
 modified with a phosphorothioate, or other modification(s) as set forth below. In certain
 10 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,
 5, 6, 7, or 8 modifications. In certain embodiments, the 5' extension domain comprises as
 15 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
 consecutive nucleotides, *e.g.*, two consecutive nucleotides that are within 5 nucleotides of the
 20 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
 domain, or within a region that is more than 5 nucleotides away from one or both ends of the
 25 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
 30 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.

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 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 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 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,

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 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.

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 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 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 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.

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, 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.

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:

5 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;

10 the linking domain is 1 to 5 nucleotides in length;

 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 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.

20 In certain embodiments, a unimolecular gRNA as disclosed herein comprises, 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;

25 a linking domain;

 a second complementarity domain;

 a proximal domain; and

 a tail domain,

wherein,

30 (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

(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.

5 In certain embodiments, the sequence from (a), (b), and/or (c) has at least 50%, 60%, 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.

10 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 are complementary to the corresponding nucleotides of the first complementarity domain.

15 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 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
20 is complementary to the target domain over the entire length of the targeting domain, the 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
25 amino acid sequence set forth in SEQ ID NO:45, wherein the targeting domain is listed as 20 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.

30 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 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.

5 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;

10 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

15 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
20 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.

25 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.

30 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
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.

5 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
10 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
15 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
20 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
25 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
30 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 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
5 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.*,
10 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
15 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
20 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
25 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
30 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.

5 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.

10 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.

15 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
20 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
25 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
30 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.

gRNA delivery

In certain embodiments of the methods provided herein, the methods comprise delivery of one or more (*e.g.*, two, three, or four) gRNA molecules as described herein. In certain of these embodiments, the gRNA molecules are delivered by intravenous injection, intramuscular injection, subcutaneous injection, or inhalation.

II. 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 *HBB* gene was obtained from the UCSC Genome browser and sequences were 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.

Targeting Domains for Use in Correcting the E6V Mutation in the *HBB* Gene

Targeting domains to target the E6V mutation in the *HBB* gene in conjunction with the methods disclosed herein were identified and ranked into tiers for *S. pyogenes*, *S. aureus* and *N. meningitidis*.

A. First Targeting Domain Selection Strategy

For *S. pyogenes* Cas9, targeting domains were identified and ranked into three tiers. Tier 1 targeting domains were selected based on (1) a reasonable distance upstream or downstream from either end of the target site, and (2) a high level of orthogonality. Tier 2 targeting domains were selected based on (1) a reasonable distance upstream or downstream from either end of the target site, and (2) presence of a 5'G. Tier 3 targeting domains were selected based on (1) a reasonable upstream or downstream from either end of the target site.

For *S. aureus* Cas9, targeting domains were identified manually by scanning genomic DNA sequence for the presence of PAM sequences. These gRNAs were not separated into tiers.

Note that tiers are non-inclusive (each targeting domain is listed only once for the strategy). The identified targeting domains are summarized below in Table 1.

Table 1. Nucleic Acid Sequences of *S. pyogenes* and *S. aureus* targeting domains

Tier	<i>S. pyogenes</i>	<i>S. aureus</i>*
Tier 1	SEQ ID NOs: 387-390	SEQ ID NOs: 442-485
Tier 2	SEQ ID NOs: 391-404	
Tier 3	SEQ ID NOs: 405-441	

*Not separated into tiers.

In certain embodiments, a point mutation in the *HBB* gene, *e.g.*, at E6, *e.g.*, E6V, is targeted, *e.g.*, for correction by gene conversion. In certain embodiments, the targeting domain of a gRNA molecule comprises a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from any one of **Table 1**. In certain embodiments, the targeting domain is selected from those in **Tables 1**. For example, in certain embodiments, the targeting domain is:

AAGGUGAACGUGGAUGAAGU (SEQ ID NO: 387);
GUAACGGCAGACUUCUCCUC (SEQ ID NO: 388);
GUGAACGUGGAUGAAGU (SEQ ID NO: 389);
ACGGCAGACUUCUCCUC (SEQ ID NO: 390); or
GUAACGGCAGACUUCUCCAC (SEQ ID NO: 16318).

In certain embodiments, when the SCD target point position is E6, *e.g.*, E6V, and two gRNAs are used to position two breaks, *e.g.*, two single stranded breaks, in the target nucleic acid sequence, the targeting domain of each gRNA molecule is selected from one of **Table 1**.

B. Second Targeting Domain Selection Strategy

For *S. pyogenes* Cas9, targeting domains were identified and ranked into four tiers. Tier 1 targeting domains were selected based on (1) a short distance upstream or downstream from either end of the target site, *e.g.*, within 100 bp upstream and 100 bp downstream of the E6V mutation, (2) a high level of orthogonality, and (3) the presence of a 5' G. Tier 2 targeting domains were selected based on (1) a short distance upstream or downstream from either end of the target site, *e.g.*, within 100 bp upstream and 100 bp downstream of the E6V mutation, and (2) a high level of orthogonality. Tier 3 targeting domains were selected based on (1) a short distance upstream or downstream from either end of the target site, *e.g.*, within 100 bp upstream and 100 bp downstream of the E6V mutation, (2) the presence of a 5' G. Tier 4 targeting domains were selected based on (1) a short distance upstream or downstream from either end of the target site, *e.g.*, within 100 bp upstream and 100 bp downstream of the E6V mutation.

For *S. aureus* Cas9, targeting domains were identified and ranked into 5 tiers. Tier 1 targeting domains were selected based on (1) a short distance upstream or downstream from

either end of the target site, *e.g.*, within 100 bp upstream and 100 bp downstream of the E6V mutation, (2) a high level of orthogonality, and (3) the presence of a 5' G. Tier 2 targeting domains were selected based on (1) a short distance upstream or downstream from either end of the target site, *e.g.*, within 100 bp upstream and 100 bp downstream of the E6V mutation, and (2) a high level of orthogonality. Tier 3 targeting domains were selected based on (1) a short distance upstream or downstream from either end of the target site, *e.g.*, within 100 bp upstream and 100 bp downstream of the E6V mutation, and (2) the presence of a 5' G. Tier 4 targeting domains were selected based on (1) a short distance upstream or downstream from either end of the target site, *e.g.*, within 100 bp upstream and 100 bp downstream of the E6V mutation. Tier 5 targeting domains were selected based on (1) a short distance upstream or downstream from either end of the target site, *e.g.*, within 100 bp upstream and 100 bp downstream of the E6V mutation, and (2) the PAM is NNGRRV.

Note that tiers are non-inclusive (each targeting domain is listed only once for the strategy). In certain instances, no targeting domain was identified based on the criteria of the particular tier. The identified targeting domains are summarized below in Table 2.

Table 2. Nucleic Acid Sequences of *S. pyogenes* and *S. aureus* targeting domains

Tier	<i>S. pyogenes</i>	<i>S. aureus</i>
Tier 1	SEQ ID NOs: 6803-6804	SEQ ID NO: 6855
Tier 2	SEQ ID NOs: 6805-6810	SEQ ID NO: 6856
Tier 3	SEQ ID NOs: 6811-6822	N/A
Tier 4	SEQ ID NOs: 6823-6854	N/A
Tier 5	N/A	SEQ ID NOs: 6857-6871

Exemplary gRNA pairs can be generated using the following targeting domain pairs: HBB-9 (SEQ ID NO: 6811) and HBB-11 (SEQ ID NO: 6813); HBB-9 (SEQ ID NO: 6811) and HBB-39 (SEQ ID NO: 6841); HBB-20 (SEQ ID NO: 6822) and HBB-11 (SEQ ID NO: 6813); and HBB-20 (SEQ ID NO: 6822) and HBB-39 (SEQ ID NO: 6841).

In certain embodiments, a point mutation in the *HBB* gene, *e.g.*, at E6, *e.g.*, E6V, is targeted, *e.g.*, for correction. In certain embodiments, the targeting domain of a gRNA molecule comprises a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from any one of **Table 2**. In certain embodiments, the targeting domain is selected from those in **Tables 2**. For example, in certain embodiments, the targeting domain is:

GGUGCACCUGACUCCUG (SEQ ID NO: 6803); or
GUAACGGCAGACUUCUCCAC (SEQ ID NO: 6804).

In certain embodiments, when the SCD target point position is E6, *e.g.*, E6V, and two gRNAs are used to position two breaks, *e.g.*, two single stranded breaks, in the target nucleic acid sequence, the targeting domain of each gRNA molecule is selected from one of **Table 2**.

C. Third Targeting Domain Selection Strategy

For *S. pyogenes* Cas9, targeting domains were identified and ranked into 3 tiers. Tier 1 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site, *e.g.*, within 200 bp upstream and 200 bp downstream of the E6V mutation and (2) a high level of orthogonality. Tier 2 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site, *e.g.*, within 200 bp upstream and 200 bp downstream of the E6V mutation, and (2) the presence of a 5' G. Tier 3 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site, *e.g.*, within 200 bp upstream and 200 bp downstream of the E6V mutation. For *S. aureus* Cas9, targeting domains were identified and ranked into 4 tiers. Tier 1 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site, *e.g.*, within 200 bp upstream and 200 bp downstream of the E6V mutation, (2) a high level of orthogonality, and (3) the PAM is NNGRRT. Tier 2 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site, *e.g.*, within 200 bp upstream and 200 bp downstream of the E6V mutation, (2) the presence of a 5' G, and (3) the PAM is NNGRRT. Tier 3 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site, *e.g.*, within 200 bp upstream and 200 bp downstream of the E6V mutation and (2) the PAM is NNGRRT. Tier 4 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site, *e.g.*, within 200 bp upstream and 200 bp downstream of the E6V mutation and (2) the PAM is NNGRRT.

For *N. meningitidis* Cas9, targeting domains were identified and ranked into 3 tiers. Tier 1 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site, *e.g.*, within 200 bp upstream and 200 bp downstream of the E6V mutation and (2) a high level of orthogonality. Tier 2 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site, *e.g.*, within 200 bp upstream and 200 bp downstream of the E6V mutation, and (2) the presence of a 5' G. Tier 3 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site, *e.g.*, within 200 bp upstream and 200 bp downstream of the E6V mutation.

Note that tiers are non-inclusive (each targeting domain is listed only once for the strategy). In certain instances, no targeting domain was identified based on the criteria of the particular tier. The identified targeting domains are summarized below in Table 3.

5 **Table 3. Nucleic Acid Sequences of *S. pyogenes* and *S. aureus* targeting domains**

Tier	<i>S. pyogenes</i>	<i>S. aureus</i>	<i>N. meningitidis</i>
Tier 1	SEQ ID NOs: 16010-16055	SEQ ID NOs: 16085-16098	SEQ ID NOs: 16253-16256
Tier 2	SEQ ID NOs: 16056-16063	N/A	
Tier 3	SEQ ID NOs: 16064-16084	N/A	
Tier 4	N/A	SEQ ID NOs: 16099-16252	

In certain embodiments, a point mutation in the *HBB* gene, *e.g.*, at E6, *e.g.*, E6V, is targeted, *e.g.*, for correction. In certain embodiments, the targeting domain of a gRNA molecule comprises a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence from any one of **Table 3**. In certain embodiments, the targeting domain is selected from those in **Table 3**.

In certain embodiments, when the SCD target point position is E6, *e.g.*, E6V, and two gRNA molecules are used to position two breaks, *e.g.*, two single stranded breaks, in the target nucleic acid sequence, the target domain of each gRNA is selected from one of **Table 3**.

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), *e.g.*, two gRNAs comprising the targeting domains of HBB-8 (SEQ ID NO: 388) and HBB-15 (SEQ ID NO: 387) are used. In certain embodiments, two 17-mer gRNAs are used to target two Cas9 nucleases or two Cas9 nickases, *e.g.*, two gRNAs comprising the targeting domains of HBB-35 (SEQ ID NO: 389) and HBB-53 (SEQ ID NO: 390) are used. Any of the targeting domains in the tables 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).

In certain embodiments, the targeting domain comprises a sequence that is the same as, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence described herein, *e.g.*, a targeting domain sequence selected from any one of **Tables 1-3**. In certain embodiments, the targeting domain is a targeting domain sequence described herein, *e.g.*, a targeting domain sequence selected from those in **Tables 1-3**.

HBB 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 an embodiment, the proximal domain and tail domain are taken together as a single domain.

In an 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.

III. Cas9 Molecules

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*, *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*, *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 trichosporium*, *Mobiluncus mulieris*, *Neisseria bacilliformis*, *Neisseria cinerea*, *Neisseria flavescens*, *Neisseria lactamica*, *Neisseria meningitidis*, *Neisseria* sp., *Neisseria wadsworthii*, *Nitrosomonas* sp., *Parvibaculum lavamentivorans*, *Pasteurella multocida*,
 5 *Phascolarctobacterium succinatutens*, *Ralstonia syzygii*, *Rhodopseudomonas palustris*, *Rhodovulum* sp., *Simonsiella muelleri*, *Sphingomonas* sp., *Sporolactobacillus vineae*, *Staphylococcus lugdunensis*, *Streptococcus* sp., *Subdoligranulum* sp., *Tistrella mobilis*, *Treponema* sp., or *Verminephrobacter eiseniae*. The amino acid sequences of exemplary Cas9 orthologs are set forth in SEQ ID NOs: 304-386.

10

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).

15

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*.

20

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.*

25

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
 30 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 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:

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);

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 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:

5 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

10 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.

15 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

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

20 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.

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
25 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, 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.

30

Additional RuvC-like domains

In addition to the N-terminal RuvC-like domain, the Cas9 molecule or Cas9 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.

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

5 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.

10 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:

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

X₁ is H or L;

X₂ is R or V; and

X₃ is E or V.

20 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).

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:

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

wherein

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);

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

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

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 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.

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

10 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;

15 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;

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);

20 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;

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);

25 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;

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);

30 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;

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.

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

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

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

$X_1-X_2-X_3-H-X_4-X_5-P-X_6-S-X_8-X_9-X_{10}-D-D-S-X_{14}-X_{15}-N-K-V-L-X_{19}-X_{20}-X_{21}-X_{22}-X_{23}-N$ (SEQ ID NO: 18),

wherein

- 10 X_1 is selected from D and E;
 X_2 is selected from L, I, R, Q, V, M and K;
 X_3 is selected from D and E;
 X_4 is selected from I, V, T, A and L (*e.g.*, A, I and V);
 X_5 is selected from V, Y, I, L, F and W (*e.g.*, V, I and L);
15 X_6 is selected from Q, H, R, K, Y, I, L, F and W;
 X_8 is selected from F, L, V, K, Y, M, I, R, A, E, D and Q (*e.g.*, F);
 X_9 is selected from L, R, T, I, V, S, C, Y, K, F and G;
 X_{10} is selected from K, Q, Y, T, F, L, W, M, A, E, G, and S;
 X_{14} is selected from I, L, F, S, R, Y, Q, W, D, K and H (*e.g.*, I, L and F);
20 X_{15} is selected from D, S, I, N, E, A, H, F, L, Q, M, G, Y and V;
 X_{19} is selected from T, V, C, E, S and A (*e.g.*, T and V);
 X_{20} is selected from R, F, T, W, E, L, N, C, K, V, S, Q, I, Y, H and A;
 X_{21} is selected from S, P, R, K, N, A, H, Q, G and L;
 X_{22} is selected from D, G, T, N, S, K, A, I, E, L, Q, R and Y; and
25 X_{23} 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: 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:

30 $X_1-V-X_3-H-I-V-P-X_6-S-X_8-X_9-X_{10}-D-D-S-X_{14}-X_{15}-N-K-V-L-T-X_{20}-X_{21}-X_{22}-X_{23}-N$ (SEQ ID NO:19),

wherein

X_1 is selected from D and E;

X_3 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);

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;

5 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;

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

10 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:

15 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),

wherein

X₂ is selected from I and V;

X₅ is selected from I and V;

20 X₇ is selected from A and S;

X₉ is selected from I and L;

X₁₀ is selected from K and T;

X₁₂ is selected from D and N;

X₁₆ is selected from R, K and L;

25 X₁₉ is selected from T and V;

X₂₀ is selected from S and R;

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).

30 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.

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

X₁' is selected from K and R;

5 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;

X₆' is selected from D, N and H;

10 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

15 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.

20 In certain embodiments, the HNH-like domain differs from a sequence of an HNH-like domain disclosed herein, *e.g.*, in SEQ ID Nos: 178-252, 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 SEQ ID Nos: 178-252 are present.

In certain embodiments, the HNH-like domain differs from a sequence of an HNH-like domain disclosed herein, *e.g.*, in SEQ ID Nos: 253-302, as many as 1 but not more than 25 2, 3, 4, or 5 residues. In certain embodiments, 1, 2, all 3 of the highly conserved residues of SEQ ID Nos: 253-302 are present.

Cas9 Activities

30 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 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.

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

- 5 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 acid and create a double stranded break, which in an embodiment is the presence of two nickase activities;
- 10 an endonuclease activity;
- an exonuclease activity; and
- a helicase activity, *i.e.*, the ability to unwind the helical structure of a double stranded 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
 15 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 hybridizes with. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with an HNH domain. In an embodiment, an eaCas9 molecule or
 20 eaCas9 polypeptide comprises cleavage activity associated with a RuvC domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises cleavage activity associated with an HNH domain and cleavage activity associated with a RuvC domain. In an embodiment, an eaCas9 molecule or eaCas9 polypeptide comprises an active, or cleavage competent, HNH domain and an inactive, or cleavage incompetent, RuvC domain. In an
 25 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.
 30 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 an 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 an 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 an 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 an 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 an 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 an 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 an 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 an 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. dysgalactiae* (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 an 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 an 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 an 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 an 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*,

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

In an 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.*

10 *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.*

15 *pyogenes*, *S. thermophilus*, *S. mutans*, or *L. innocua*.

In an 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*,

20 *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.*

25 *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.*

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

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

15 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

25 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

30 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 an 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 an 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 an 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 an 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.

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

In an 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 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*);

5 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 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.

10 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 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.

15 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 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

20 position 10 of SEQ ID NO:2, *e.g.*, can be substituted with an alanine. In an embodiment, the eaCas9 molecule or eaCas9 polypeptide differs from wild-type in the N-terminal RuvC-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

25 by 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 an embodiment, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology.

30 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. pyogenes* Cas9 sequence (SEQ ID NO:2) *e.g.*, can be substituted with an alanine. In an 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 molecule, *e.g.*, a naturally occurring Cas9 molecule such as a Cas9 molecule of *S. pyogenes*, *S. aureus*, or *S. thermophilus*. In an 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 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 an embodiment, a mutation(s) is present in a RuvC domain. In an embodiment, a mutation(s) is present in an HNH domain. In an 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 an 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.

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 an 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 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 an embodiment, a Cas9 molecule 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 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 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, *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: 16302 (D10A) or SEQ ID NO: 16303 (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; cleavage activity associated with an HNH domain; cleavage activity associated with an HNH domain and cleavage activity associated with a RuvC domain.

In an 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 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 an embodiment, the reference Cas9 molecule is the naturally occurring Cas9 molecule having the closest sequence identity or homology. In an 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.

5 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 an 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

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

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.

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 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 acid and create a double stranded break, which in an 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 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 an 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 an 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, 16300 and 16301.

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)	<i>y1727</i>
<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 ^{§§}
<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>csml</i>	NA	COG1353	MTH326, Rv2823c ^{§§}

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
		and <i>csx11</i>			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-B	<i>cmr3</i>	NA	COG1769 (RAMP)	PF1128
<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

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>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

IV. 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).

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.

5 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
10 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
15 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.
20 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.

25 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
30 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 µL. Cas9 protein molecule is programmed with equimolar amounts of pre-annealed gRNA molecule and titrated from 100 pM to 1 µM. 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 µM 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 µM 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.

V. Genome Editing Approaches

Mutations in a target gene may be corrected using one of the approaches discussed herein. In an embodiment, a mutation in a target gene is corrected by homology directed repair (HDR) using an exogenously provided template nucleic acid (see Section V.1), 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 (see Section V.1), referred to herein as gene correction.

V.1 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 at E6 of the *HBB* gene). While not wishing to be bound by theory, it is believed that HDR-mediated alteration of a target sequence within a target gene (*e.g.*, a *HBB* 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 other embodiments, 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 target sequence in a genome without the use of an exogenously provided donor template or template nucleic acid. While not wishing to be bound by theory, it is believed that alteration of the target sequence occurs by HDR with endogenous genomic donor sequence, in a process referred to herein as gene conversion. For example, the endogenous genomic donor sequence provides for alteration of the target sequence. It is contemplated that in an embodiment the endogenous genomic donor sequence is located on the same chromosome as the target sequence. It is further contemplated that in another embodiment the endogenous genomic donor sequence is located on a different chromosome from the target sequence. In certain embodiments, the endogenous genomic donor sequence comprises one or more nucleotides derived from the *HBD* gene. In other embodiments, the endogenous genomic donor sequence comprises one or more nucleotides derived from the *SMN2* gene. In certain embodiments, the endogenous genomic donor sequence comprises one or more nucleotides derived from the *p47-PHOX* pseudogene. Alteration of a target sequence by endogenous genomic donor sequence depends on cleavage by a Cas9 molecule. Cleavage by Cas9 can comprise a double-strand break or two single-strand breaks.

In an 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 an 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 an 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 an 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 an 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.

5 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
10 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.

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

15 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 overhanging DNA regions.

20 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 Holliday junction model (or double-strand break repair, DSB, pathway) or the synthesis-
25 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 the invading strand to fill the gap resulting from the resection. The end of the newly
30 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 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 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.

Mutations in the *HBB* gene that can be altered (*e.g.*, corrected) by HDR with a template nucleic acid or with endogenous genomic donor sequence include, *e.g.*, point mutation at E6, *e.g.*, E6V.

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.

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.

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 an 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 an 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 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 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;

5 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, 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
10 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.

15 In an embodiment, there is no overlap between the target sequences that are complementary to the targeting domains of the two gRNAs. In an embodiment, the gRNAs do not overlap and are separated by as much as 50, 100, or 200 nucleotides. In an 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
20 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 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

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,
30 100, 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 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 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, 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 an 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 an 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 an 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

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.

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, 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, *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.

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.

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 SSB machinery). In certain embodiments, the nicks are far enough apart that they 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 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' 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).

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 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
 5 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 (*e.g.*, no more than 90, 80, 70, 60, 50, 40, 30, 20, 10, or 5 bp away from each other). In
 10 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
 15 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
 20 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 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
 30 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 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 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 cleave the complementary strand, resulting in a double stranded break.

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

5 In an 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 believed that, in an embodiment, the sequence exposed by a cleavage event (*e.g.*, a single-strand cleavage event) mediated by a
 10 gRNA molecule and a Cas9 molecule (*e.g.*, a Cas9 nickase, *e.g.*, a Cas9 molecule containing D10A or N863A mutation) may affect (*e.g.*, increase or decrease) gene conversion 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 an embodiment, the
 15 exposed sequence comprises or consists of a 5' overhang. In another embodiment, the exposed sequence comprises or consists of a 3' overhang. In an 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.

20 In an 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,
 25 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 molecule/gRNA molecule mediated cleavage event can constitute a substrate used for
 30 homology search in gene conversion.

By way of example, when a pair of gRNA molecules targets a Cas9 nickase to mediate a cleavage event at a *HBB* locus, the sequence exposed by said cleavage event can probe any genomic locus for possible homology. While not to be bound by theory, it is believed that, in an embodiment, higher sequence identity of the exposed *HBB* region with a

genomic region can promote gene conversion more effectively. For example, the *HBD* gene has approximately 92% sequence homology with the *HBB* gene, therefore the *HBD* gene may be used as a donor template for effective gene conversion. If the exposed sequence contains a mutation with respect to the *HBD* sequence, the mutation may decrease gene conversion efficiency. Therefore, different gRNA molecules may give rise to exposed sequences having different homologies to a donor template.

In an embodiment, the exposed sequence differs by 1, 2, 3, 4, 5, 10, 25, 50, 100 or fewer, nucleotides with an endogenous homologous sequence. In an 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 an 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 an 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 an 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 an 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 an 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 an 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 an 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 an 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 an 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

5 nucleotides in length.

In an 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
10 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
15 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 an 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,

20 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,
25 300, 400, 500, 750, 1000, 2000, 3000, 4000, 5000, 6000, 7000, 8000, 9000, 10000, or 15000 nucleotides in length.

In an 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
30 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 an 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 an embodiment, the single-strand breaks are present on different strands. In another embodiment, the single-strand breaks are present on the same strand. In an embodiment, the cleavage event further comprises one or more (*e.g.*, two) of single-strand break, double-strand break, or both.

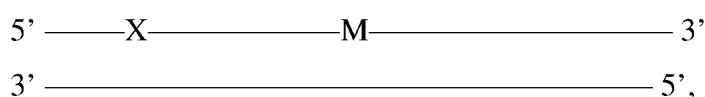
In an embodiment, the Cas9 molecule comprises HNH-like domain cleavage activity but has no, or no significant, N-terminal RuvC-like domain cleavage activity. In an embodiment, the eaCas9 molecule is an HNH-like domain nickase, *e.g.*, the Cas9 molecule comprises a mutation at D10, *e.g.*, D10A. In another embodiment, the eaCas9 molecule comprises N-terminal RuvC-like domain cleavage activity but has no, or no significant, HNH-like domain cleavage activity. In an 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 an embodiment, the first gRNA molecule positions a cleavage event on a strand that does not bind to the first gRNA molecule.

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

In an 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' overhang on each strand.

In an embodiment, the first gRNA molecule positions a cleavage event 5' to the target position on the first strand. In an embodiment, the first gRNA molecule positions a cleavage event 5' to the target position on the first strand, as shown in the diagram below:



wherein X is the cleavage event and M is the target position.

In an 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 an embodiment, the second gRNA molecule positions a cleavage event 3' to the target position
 5 (relative to the target position on the first strand) on the second strand, as shown in the diagram below:

$$\begin{array}{l} 5' \text{ ————— M ————— } 3' \\ 3' \text{ ————— X ————— } 5', \end{array}$$

wherein X is the cleavage event and M is the target position.

10 In an embodiment, the second gRNA molecule positions a cleavage event 5' to the target position on the second strand. In an embodiment, the second gRNA molecule positions a cleavage event 5' to the target position on the second strand, as shown in the diagram below:

$$\begin{array}{l} 5' \text{ ————— M ————— } 3' \\ 15 \quad 3' \text{ ————— M ————— X ————— } 5', \end{array}$$

wherein X is the cleavage event and M is the target position.

In an 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 an 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, as shown in the diagram below:

$$\begin{array}{l} 5' \text{ — X ————— M ————— } 3' \\ 25 \quad 3' \text{ ————— X ————— } 5', \end{array}$$

wherein X is the cleavage event and M is the target position.

In an embodiment, the first gRNA molecule positions a cleavage event 3' to the target position on the first strand. In an embodiment, the first gRNA molecule positions a cleavage event 3' to the target position on the first strand, as shown in the diagram below:

$$\begin{array}{l} 30 \quad 5' \text{ ————— M ————— X ————— } 3' \\ 3' \text{ ————— } 5', \end{array}$$

wherein X is the cleavage event and M is the target position.

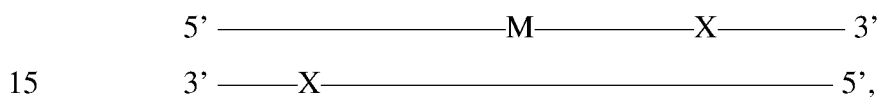
In an embodiment, the second gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand. In an

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, as shown in the diagram below:



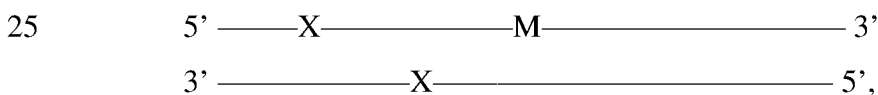
wherein X is the cleavage event and M is the target position.

In an 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 an 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, as shown in the diagram below:



wherein X is the cleavage event and M is the target position.

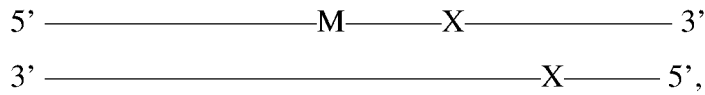
In an embodiment, the first gRNA molecule positions a cleavage event 5' to the target position on the first strand, and wherein the second gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 5' overhang. In an 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, as shown in the diagram below:



wherein X is the cleavage event and M is the target position.

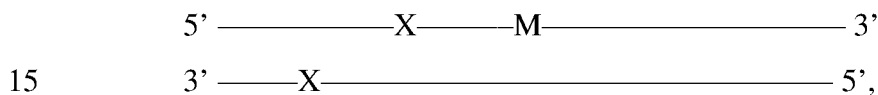
In an embodiment, the first gRNA molecule positions a cleavage event 3' to the target position on the first strand, and wherein the second gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand, *e.g.*, to produce a 5' overhang. In an 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 5' overhang, as shown in the diagram below:



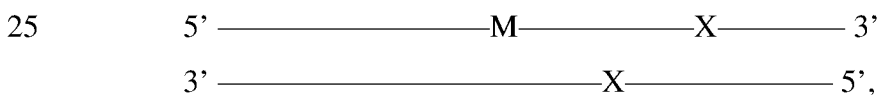
5 wherein X is the cleavage event and M is the target position.

In an 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' overhang. In an 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' overhang, as shown in the diagram below:



wherein X is the cleavage event and M is the target position.

In an 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 an 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, as shown in the diagram below:



wherein X is the cleavage event and M is the target position.

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

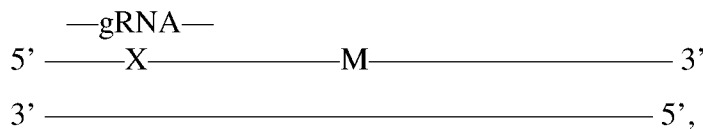
30 In an embodiment, the first gRNA molecule positions a cleavage event on a strand that binds to the gRNA molecule.

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

In an 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 second gRNA molecule bind to different strands, *e.g.*, resulting in a 5' overhang on each strand.

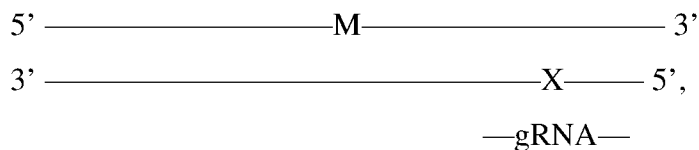
In an embodiment, the gRNA molecule, together with the Cas9 molecule (*e.g.*, a nickase), positions a cleavage event on a strand (*e.g.*, a first strand or a second strand),

In an 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 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). For example, this embodiment can be illustrated as shown in the diagram below:



wherein X is the cleavage event, M is the target position.

In an 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, 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). For example, this embodiment can be illustrated as shown in the diagram below:



wherein X is the cleavage event, M is the target position.

In an 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.*, 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). For example, this embodiment can be illustrated as shown in the diagram below:

5

5' ————— M ————— X ————— 3'

3' ————— 5',

—gRNA—

wherein X is the cleavage event, M is the target position.

In an 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). For example, this embodiment can be illustrated as shown in the diagram below:

5' ————— M ————— 3'
3' ————— X ————— 5',
—gRNA—

wherein X is the cleavage event, M is the target position.

20 In an 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). For example, this

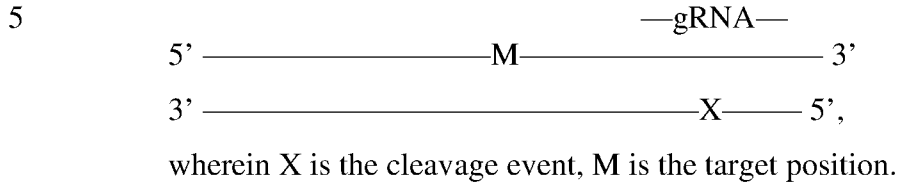
25 embodiment can be illustrated as shown in the diagram below:

5' —X— M— 3'
 3' ————— 5',
 —gRNA—

30 wherein X is the cleavage event, M is the target position.

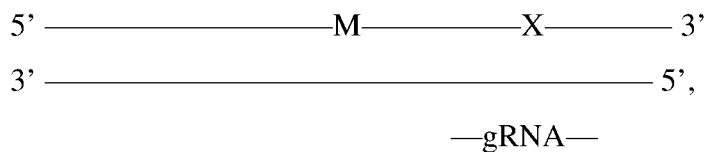
In an embodiment, the gRNA molecule positions a cleavage event 3' to the target position (relative to the target position on the first strand) on the second strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with 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). For example, this embodiment can be illustrated as shown in the diagram below:



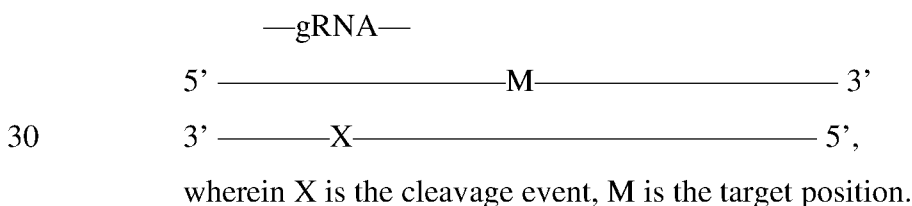
In an embodiment, the gRNA molecule positions a cleavage event 3' to the target position on the first strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with an N863A mutation), *e.g.*, to place a single-strand cleavage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

15 embodiment can be illustrated as shown in the diagram below:



wherein X is the cleavage event, M is the target position.

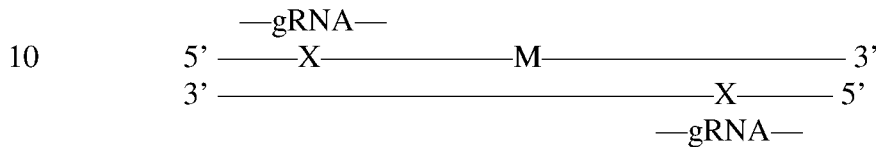
In an embodiment, the gRNA molecule positions a cleavage event 5' to the target position (relative to the target position on the first strand) on the second strand. This embodiment allows the use of a single Cas9 molecule, *e.g.*, a single Cas9 molecule that is a nickase (*e.g.*, a Cas9 molecule with 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). For example, this embodiment can be illustrated as shown in the diagram below:



In an 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.

For example, this embodiment can be illustrated as shown in the diagram below:



wherein X is the cleavage event, M is the target position.

In an 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 an 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,

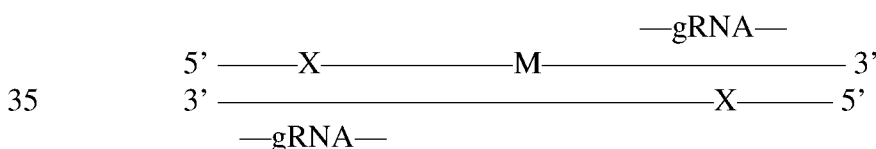
wherein:

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.

For example, this embodiment can be illustrated as shown in the diagram below:



wherein X is the cleavage event and M is the target position.

In an 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 an 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,

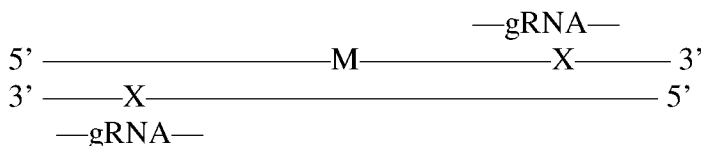
wherein:

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

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.

For example, this embodiment can be illustrated as shown in the diagram below:



wherein X is the cleavage event and M is the target position.

In an 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 an 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

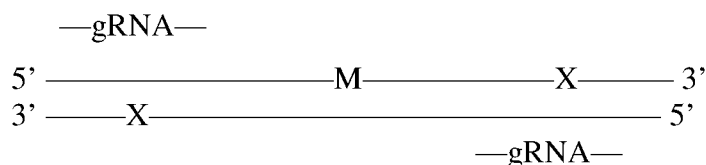
the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand other than the strand that binds to the second gRNA molecule,

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 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 N863A mutation), *e.g.*, to place single-strand cleavage events on each
side of the target position.

For example, this embodiment can be illustrated as shown in the diagram below:



wherein X is the cleavage event and M is the target position.

In an 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 an 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

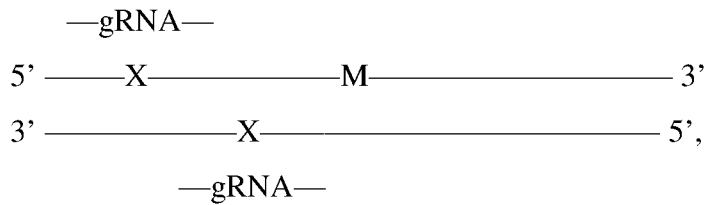
the second gRNA molecule, together with the Cas9 molecule, positions a cleavage
event on the strand that binds to the second gRNA molecule,

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 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.

For example, this embodiment can be illustrated as shown in the diagram below:



5 wherein X is the cleavage event and M is the target position.

In an 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.

10 In an 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,

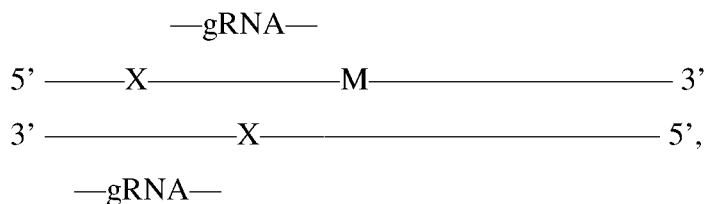
15 wherein:

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

20 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.

25 For example, this embodiment can be illustrated as shown in the diagram below:



30 wherein X is the cleavage event and M is the target position.

In an 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 an 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

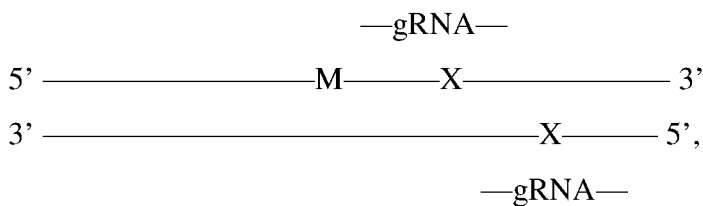
the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand that binds to the second gRNA molecule,
 wherein:

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.

For example, this embodiment can be illustrated as shown in the diagram below:



wherein X is the cleavage event and M is the target position.

In an 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 an 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,

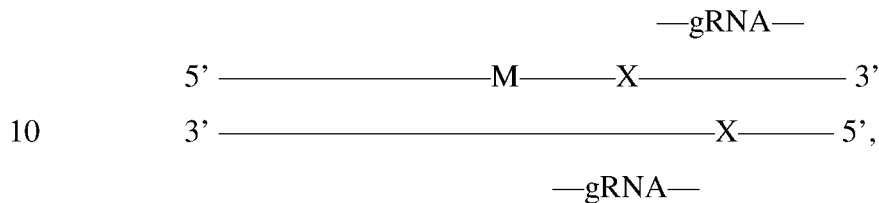
wherein:

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
 5 single-strand cleavage events on each side of the target position, *e.g.*, to produce a 5' overhang.

For example, this embodiment can be illustrated as shown in the diagram below:



wherein X is the cleavage event and M is the target position.

In an 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
 15 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 an embodiment:

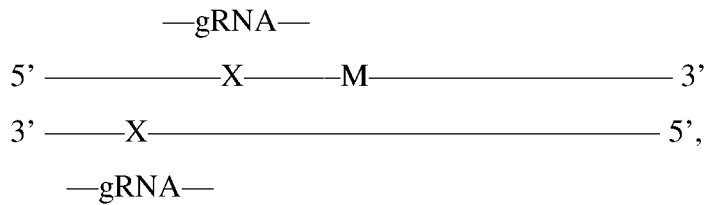
the gRNA molecule, together with the Cas9 molecule (*e.g.*, a nickase), positions a cleavage event on the strand that binds to the gRNA molecule; and
 20 the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand that binds to the second gRNA molecule,

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
 25 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 a D10A mutation), *e.g.*, to place single-
 30 strand cleavage events on one side of the target position, *e.g.*, to produce a 3' overhang.

For example, this embodiment can be illustrated as shown in the diagram below:



5 wherein X is the cleavage event and M is the target position.

In an 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.

10 In an 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,

15 wherein:

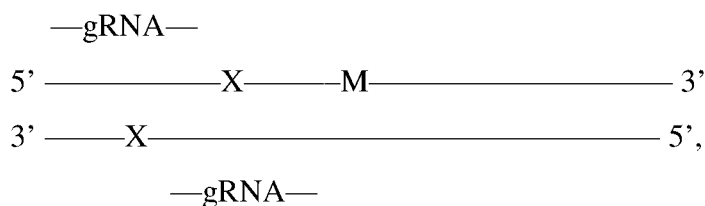
the gRNA molecule and the second gRNA molecule bind to different strands,

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

20 (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.

25 For example, this embodiment can be illustrated as shown in the diagram below:



30 wherein X is the cleavage event and M is the target position.

In an 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 an 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

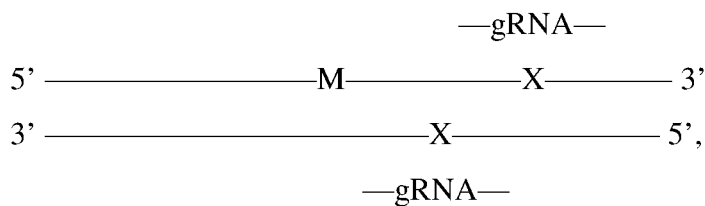
the second gRNA molecule, together with the Cas9 molecule, positions a cleavage event on the strand that binds to the second gRNA molecule, wherein:

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.

For example, this embodiment can be illustrated as shown in the diagram below:



wherein X is the cleavage event and M is the target position.

In an 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 an 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,

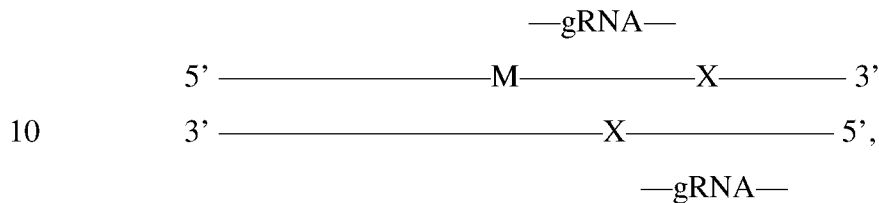
wherein:

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.

For example, this embodiment can be illustrated as shown in the diagram below:



wherein X is the cleavage event and M is the target position.

In an 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
 15 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.

Homology arms of the donor template

A homology arm should extend at least as far as the region in which end resection
 20 may occur, *e.g.*, in order to allow the resected single stranded overhang to find a complementary region within the donor template. The overall length could be limited by parameters such as plasmid size or viral packaging limits. In an 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,
 25 3000, 4000, or 5000 nucleotides. In some embodiments, the homology arm length is 50-100, 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
 30 target position can be a modified Cas9 molecule cleavage of the target nucleic acid and template nucleic acid directed modification, *e.g.*, correction, of the target position. In an 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 an embodiment, the target position is within a target sequence (*e.g.*, the sequence to which the gRNA binds). In an embodiment, a target position is upstream or downstream of a target sequence (*e.g.*, the sequence to which the gRNA binds).

5 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 an embodiment, the template nucleic acid is single stranded. In certain
10 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 an 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
15 template nucleic acid comprises endogenous genomic sequence.

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
20 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
25 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 an 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
30 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 *HBB* 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 *HBB* 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 an 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.

- 5 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 certain embodiments alt-HDR proceeds more efficiently when the template nucleic acid has extended homology 5' to the nick (*i.e.*, in
10 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
15 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
20 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'
25 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
30 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 conversion, the template nucleic acid is an endogenous homologous region. 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.

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. 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 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 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 bp 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
5 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
10 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 template nucleic acid is an adenovirus vector, *e.g.*, an AAV
15 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
20 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,
25 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.

30 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 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 an 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, 30, 40, or 50 silent mutations relative to the corresponding sequence in the genome of the cell to be altered. In certain embodiments, the template nucleic acid comprises at most 2, 3, 4, 5, 10, 20, 30, 40, or 50 silent mutations relative to the corresponding sequence in the genome of the cell to be altered. In an 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, 30, 40, or 50 silent mutations relative to the corresponding sequence in the genome of the cell to be altered. In certain embodiments, the template nucleic acid comprises at most 2, 3, 4, 5, 10, 20, 30, 40, or 50 silent mutations relative to the corresponding sequence in the genome of the cell to be altered.

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.

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 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 allele into a wild type allele, transforming a wild type allele into a mutant allele, and/or
 5 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-translated or non-transcribed region. Such alterations include an alteration in a control
 10 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 can be used to alter an unwanted structure, *e.g.*, an unwanted or mutant nucleotide.

15 Exemplary template nucleic acids (also referred to herein as donor constructs) to correction a mutation, *e.g.*, at E6, *e.g.*, E6V, in the *HBB* gene, are provided.

Suitable sequence for a 5' homology arm can be selected from (*e.g.*, includes a portion of) or include the nucleic acid sequence of SEQ ID NO: 16257 (5'H arm).

Suitable sequence for the 3' homology arm can be selected from (*e.g.*, includes a
 20 portion of) or include the nucleic acid sequence of SEQ ID NO: 16258 (3'H arm).

In some embodiments, the replacement sequence comprises or consists of an adenine (A) residue to correct the amino acid sequence to a glutamic acid (E) residue.

In some embodiments, to correct a mutation, *e.g.*, at E6, *e.g.*, E6V, in the *HBB* gene, the homology arms, *e.g.*, the 5' and 3' homology arms, may each comprise about 1000 base
 25 pairs (bp) of sequence flanking the most distal gRNAs (*e.g.*, 1100 bp of sequence on either side of the mutation). The 5' homology arm is shown as bold sequence, codon 6 is shown as underlined sequence, the inserted base to correct the mutation at E6, *e.g.*, E6V, is shown as boxed sequence, and the 3' homology arm is shown as no emphasis sequence:

30 **ATAGGAACTTGAATCAAGGAAATGATTTTAAAACGCAGTATTCTTAGTGGA**
CTAGAG
GAAAAAATAATCTGAGCCAAGTAGAAGACCTTTTCCCCTCCTACCCCTACTTTCTA
AGTCACAGAGGCTTTTTGTTCCCCCAGACACTCTTGCAGATTAGTCCAGGCAGAAA
CAGTTAGATGTCCCCAGTTAACCTCCTATTTGACACCACTGATTACCCCATGATAG
TCACACTTTGGGTTGTAAGTGACTTTTTATTTATTTGTATTTTGTACTGCATTAAGAG
GTCTCTAGTTTTTTATCTCTTGTTCCTCCAAACCTAATAAGTAAGTAATGCACAGAG
 35 **CACATTGATTTGTATTTATTCTATTTTATGACATAATTTATTAGCATGCATGAGCAA**
TTAAGAAAAACAACAACAAATGAATGCATATATATGTATATGTATGTGTATATAT
ACACACATATATATATATATTTTTTCTTTTCTTACCAGAAGGTTTAAATCCAAATAAG

GAGAAGATATGCTTAGAACCGAGGTAGAGTTTTTCATCCATTCTGTCCTGTAAGTATT
 TTGCATATTCTGGAGACGCAGGAAGAGATCCATCTACATATCCCAAAGCTGAATTAT
 GGTAGACAAAACCTCTTCCACTTTTAGTGCATCAACTTCTTATTTGTGTAATAAGAAA
 ATTGGGAAAACGATCTTCAATATGCTTACCAAGCTGTGATTCCAAATATTACGTAAA
 5 TACACTTGCAAAGGAGGATGTTTTTAGTAGCAATTTGTACTGATGGTATGGGGCCA
 AGAGATATATCTTAGAGGGAGGGCTGAGGGTTTGAAGTCCAACCTCCTAAGCCAGTG
 CCAGAAGAGCCAAGGACAGGTACGGCTGTCATCACTTAGACCTCACCTGTGGAGC
 CACACCCTAGGGTTGGCCAATCTACTCCCAGGAGCAGGGAGGGCAGGAGCCAGGG
 CTGGGCATAAAAAGTCAGGGCAGAGCCATCTATTGCTTACATTTGCTTCTGACACAAC
 10 TGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTCCTGAGGAGAAG
 TCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAGTTGGTGGTGAGGCCCTG
 GGCAGGTTGGTATCAAGGTTACAAGACAGGTTTAAGGAGACCAATAGAACTGGGCATG
 TGGAGACAGAGAAGACTCTTGGGTTTCTGATAGGCACTGACTCTCTCTGCCTATTGGTCT
 ATTTTCCCACCCTTAGGGCTGCTGGTGGTCTACCCTTGGACCCAGAGGTTCTTTGAGTCCTT
 15 TGGGGATCTGTCCACTCCTGATGCTGTTATGGGCAACCCTAAGGTGAAGGCTCATGGCAA
 GAAAGTGCTCGGTGCCTTTAGTGATGGCCTGGCTCACCTGGACAACCTCAAGGGCACCTT
 TGCCACACTGAGTGAGCTGCACTGTGACAAGCTGCACGTGGATCCTGAGAACTTCAGGG
 TGAGTCTATGGGACGCTTGATGTTTTCTTTCCCCTTCTTTTCTATGGTTAAGTTCATGTCAT
 AGGAAGGGGATAAGTAACAGGGTACAGTTTAGAATGGGAAACAGACGAATGATTGCAT
 20 CAGTGTGGAAGTCTCAGGATCGTTTTAGTTTCTTTTATTTGCTGTTTATAACAATTGTTTT
 CTTTTGTTTAATTCTTGCTTTCTTTTTTTTTCTTCTCCGCAATTTTTACTATTATACTTAATG
 CCTAACATTGTGTATAACAAAAGGAAATATCTCTGAGATACATTAAGTAACTTAAAAAA
 AAACCTTTACACAGTCTGCCTAGTACATTACTATTTGGAATATATGTGTGCTTATTTGCATA
 TTCATAATCTCCCTACTTTATTTTCTTTTATTTTAAATTGATACATAATCATTATACATATT
 25 TATGGGTAAAGTGTAATGTTTAAATATGTGTACACATATTGACCAAATCAGGGTAATTT
 TGCATTTGTAATTTTAAAAAATGCTTTCTTCTTTTAAATATACTTTTTGTTTATCTTATTTCT
 AATACTTTCCCTAATCTCTTTCTTTTCAAGGCAATAATGATACAATGTATCATGCCTCTTTG
 CACCATTCTAAAGAATAACAGTGATAATTTCTGGGTAAAGGCAATAGCAATATCTCTGCA
 TATAAATATTTCTGCATATAAATTGTAAGT (Template Construct 1; SEQ ID NO: 16259).

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
 35 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
 40 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
 45 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.

In one embodiment, an ssODN may be used to correct a mutation, *e.g.*, E6V in the *HBB* gene. For example, the ssODN may include 50 bp 5' and 3' homology arms as shown below. The 5' homology arm is shown as bold sequence, codon 6 is shown as underlined sequence, the inserted base to correct the E6V mutation is shown as boxed sequence, and the 3' homology arm is shown as no emphasis sequence.

ACTGTGTTCACTAGCAACCTCAAACAGACACCATGGTGCATCTGACTC
CTG**G**GGAGAAGTCTGCCGTTACTGCCCTGTGGGGCAAGGTGAACGTGGATGAAG
 T (Template Construct 2; SEQ ID NO: 16260)

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 HBB gene, the mutant form of which can cause Sickle Cell Disease. If gRNA HBB-6 with the 20-base target sequence CGUUACUGCCCUGUGGGGCA is used to insert a donor sequence including

CTCCTGAGGAGAAGTCTGC**CGTTACTGCCCTGTGGGGCA****AGgGTGAACGTG GATGAAGT** (SEQ ID NO: 16297), where the *italic* A is the base being corrected and the

bracketed bases are those that match the guide RNA, the donor sequence may be changed to **CTCCTGAGGAGAAGTCTGC****CGTTACTGCCCTGTGGGGCA****AGaTGAACGTGG ATGAAGT** (SEQ ID NO: 16298), where the lowercase a has been changed from a G (lower case g) at that position so that codon 15 still codes for the amino acid arginine but the PAM sequence AGG has been modified to AGA to reduce or eliminate Cas9 cleavage at that locus.

Upregulators of Gene Conversion

In certain embodiments, without being bound by theory, the methods provided herein involve up-regulating a gene conversion pathway(s). For instance, the methods may involve modulating (*e.g.*, stimulating or overexpressing) a component (*e.g.*, exactly one component, or one or more components, *e.g.*, two or three components) of a gene conversion pathway, *e.g.*, a BRCA1, BRCA2 and/or RAD51. In some embodiments, the up-regulator of gene conversion is a BRCA1. In other embodiments, the up-regulator of gene conversion is a BRCA2. In yet other embodiments, the up-regulator of gene conversion is a RAD51.

In some embodiments, the up-regulator of gene conversion is selected from the group consisting of a polypeptide of **Table 5**, or a polypeptide that comprises at least 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, or differs by no more than 50, 40, 30, 20, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, or 1, amino acid residues from a naturally occurring polypeptide of **Table 5**.

Table 5. Polypeptides that Promote Gene Conversion

Factor	Sequence	SEQ ID NOs
BRCA2	MPIGSKERPTFFEIFKTRCNKADLGPI SLNWFEE LSSEAPPYNSEPAEESEHKNNNYEPNLFKTPQRKPS YNQLASTPIIFKEQGLTLPLYQSPVKELDKFKLDLGRNVPSNRHKSRLTVKTKMDQADDVSCPLLNSCLS ESPVVLQCTHVTPQRDKSVVCGSLFHTPKFVKGRQTPKHISESLGAEVDPDMSWSSSLATPPTLSSTVLI VRNEEASETVFPHD TTANVKS YFSNHDESLKKNDRFIASVTDSENTNQREAA SHGFGKTS GNSFKVNSCK DHIGKMPNVLEDEVYETVVD TSEEDS FSLCF SKCRTKNLQKVRTSKTRKKIFHEANADECEKSKNQVKE KYSFVSEVEPNDDPLDSNVANQKPFESGSDKISKEVVPSLACEWSQLTSLGLNGAQMEKIPLLHISSCD QNISEKDLDDTENKRKKDFLTSENSLPRISSLPKSEKPLNEETVVNKRDEEQHLESHTDCILAVKQAISG	16304

	TSPVASSFQGIKKISIFRIESPKEITFNASFSGHMTDPNFKKETEASESGLEIHTVCSQKEDSLCPNLIDN GSWPATTTQNSVALKNAGLISTLKKKTNKFIYAIHDETSYKGGKIPKDQKSELINCSAQFEANAFEAPLT FANADSGLLHSSVKRSCSQNDSEPTLSLTSFSGTILRKCSRNETCSNNTVISQDLKYEAKCNKEKLQL FITPEADSLCLQEGQCENDPKSKKVSDIKEEVLAAACHPVQHSKVEYSDTDFQSQKSLLYDHENASTLI LTPTSKDVLNLVMSIRGKESYKMSDKLKGNNYESDVELTKNIPMEKNQDVCALNENYKNVELLPPEKYM RVASPSRKVQFNQNTNLRLVIQKNQEETTSISKITVNPDSSEELFSDNENNFFVQVANERNNLALGNTKELH ETDLTVCNPIFKNSTMVLYGDTGDKQATQVSIKKDLVYVLAENKNSVKQHIKMTLGQDLKSDISLNDI KIPEKNNDYMNKWAGLLGPI SNHSFGGSFRITASNKEIKLSEHNKSKMFFKDIEEQYPTSLACVEIVNT LALDNQKKLSKPQSINTVSAHLQSSVVVSDCKNSHITPQMLFSKQDFNSNHNLTSPQKAEITELSTILEE SGSQFEFTQFRKPSYILQKSTFEVPENQMTILKTTSEECRDADLHVIMNAPSIGQVDSKQFEGTVEIKR KFAGLLKNDCKNSASGYLTDENEVGRGFYSAHGTKLVNSTEALQKAVKLFSDIENISEETSAEVHPISL SSSKCHDSVVMFKIENHNDKTVSEKNKNCQLILQNNIEMTTGTFVEEITENYKRNTENEDNKYTAASRN SHNLEFDGSDSSKNDTVCIHKDETDLLFTDQHNICLKLGGQFMKEGNTQIKEDLSDLTFLEVAKAQEACH GNTSNKEQLTATKTEQNIKDFETSDTFFQTASGKNISVAKESFNKIVNFFDQKPEELHNFSLNSELHSDI RKNKMDILSYEETDVKHKLKESVPGTGNQLVTFQGGQPERDEKIKEPTLLGFHTASGKKVKIAKESLD KVKNLDFSEKQGTSEITFSHQWAKTLKYREACKDLELACETIEITAPKCKEMQNSLNNDKLVSIETV VPPKLLSDNLCRQTEENLKTSSIFLKVKVHENVEKETAKSPATCYTNQSPYSVIENSALAFYTSCSRKTS VSQTSLLLEAKRWLREGIFDGGQPERINTADYVGNLYENNSNSTIAENDKNHLSEKQDITYLSNSSMSNSYS YHSDEVYNDSGYLSKNKLDGIEPVLKNVEDQKNTSFSKVISNVKDNAYPQTVNEDICVEELVTSSSPC KNKNAIDLSISNSNNFEVGPAPAFIAGKIVCVSHETIKKVKDIFDTSFSKVIKENNENKSKICQTKIM AGCYEALDDSEDILHNSLDNDECSTHSHKVFADIQSEELQHNQNMGLEKVSISPDCVSLSTSDICKC SIGKLHKSVSANTCGIFSTASGKSVQVSDASLQNAQVQFSEIEDSTKQVFSKVLFKSNEHSDQLTREEN TAIRTEPHLISQKGSYNVNVSSAFSGFSTASGKQVSILESSLHKVKGVLEEFDLIRTEHSLHYSPTSRO NVSKILPRVDKRNPEHCNVSEMEKTSCEKFLSNLNVGEGSSENNHSIKVSPYLSQFQQDKQLVLGKTK VSLVENIHVLGKEQASPKNVKMEIGKTETFSVDPVKTNIEVCSTYSKDSSENYFETEAVEIAKAFMEDDEL TDSKLP SHATHSLFTCPENEEMVLSNSRIGKRRGEPLILVGEPSIKRNLNNEFDRIENQEKSLKASKST PDGTIKDRRLFMHVSLEPITCVPFRTTAKERQEIQNPNTAPGQEFLSKSHLYEHLTLEKSSSNLAVSGH PFYQVSATRNKEMRHLITTTGRPTKVFPFPFKTSKSHFRVEQCVRNINLEENRQKQNDIDHGSDSSKNKIN DNEIHQFNKNNSNQAVAVTFTKCEEEPLDLITSLQNARDIQDMRIKKKQQRQVFPQPGSLYLAKTSTLPR ISLKAAGVGQVPSACSHKQLYTYGVSKHCIKINSKNAESFQFHTEDYFGKESLWTGKGIQLADGGWLIPS NDGKAGKEEFYRALCDTPGVDPKLSIRIWWYNHYRWIWKLAAMECAFPKEFANRCLSPERVLLQLKYRY DTEIDRSRRSAIKKIMERDDTAAKTLVLCVSDIISLSANISSETSSNKTSSADTQKVAIIELTGWAYVKA QLDPPLLAVLKNGLTVGQKIILHGAELVGSPPACTPLEAPESLMLKISANSTRPARWYTKLGFPPDPRP FPLPLSSLSFDGGNVGCVDVIIQRAYPIQWMEKTSGLYIFRNEREEKEAAYVEAQKRLALFTKIQ EEFEEHEENTTKPYLPSRALTRQQVRALQDGAELYEAVKNAADPAYLEGYFSEEQLRALNNHRQMLNDKK QAQIQLEIRKAMESAEQKEQGLSRDVTTVWKLRIVSYSKKEKDSVILSIWRPSSDLYSLTTEGKRYRIYH LATSKSKSKSERANIQLAATKTKYQQLPVSDIILFQIYQPREPLHFSKFLDPDFQPCSEVDLIGFVVS VVKKTGLAPFYVLSDECYNLLAIKFWIDLNEDIKPHMLIAASNQWRPESKSGLLTLFAGDFSVFSASP KEGHFQETFNKMKNTVENIDILCNEAENKLMHILHANDPKWSTPTKDCSTSGPYTAQIIPGTGNKLLMSSP NCEIYYQSPSLCLMAKRRKSVSTPVSAQMTSKCKGEKEIDDQKNCKRRALDFLSRLPLPPVSPICTFV SPAAQKAFQPPRSCGTYETPIKKKELNSPQMTFFKKFNEISLLESNSIADEELALINTQALLSGSTGEK QFISVSESTRTAPTSEDYLRKRRCTTSLIKEQESSQASTECEKNQDITITTKKYI (breast cancer 2, early onset, isoform CRA_c [Homo sapiens] CCDS 9344.1)	
BRCA1	MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITK RSLQESTRFSQLVEELLKIIICAFQLDTGLEANSYNFAKKENNSPEHLKDEVSI IQSMGYRNRARLLQS EPENPSLQETSLSVQLSNLGTVRTLRTKQRIQPQKTSVYIELGSDSSEDTVNKATYCSVGDQELLQITPQ GTRDEISLDSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERHPEKYQGSSVSNLHVPECGTNTNTHA SSLQHENSSLLLTCDRMNVEKAIEFCNKSQKQGLARSQHNWRWAGSKETCNDRRTPSTEKKVDLADPLCER KEWNNQKQLPCSENPRDTEVPWITLNSSIQKVNEWFSRSDDELGSDSDSHDGESESNKAVADVLDVLNEVD EYSGSEKIDLASDPHEALICKSERVHSKSVESNIEDKIFGKTYRKASLPNLSHVTENLIIGAFVTEP QIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGD SIQENKPNPILSLEKESAFKTKAEPISSSI NMELNLNIHNSKAPKKNRLRRKSSTRHIALELVSRN LSPPNCTELQIDSCSSSEI KKKKYNQMPVRHSRNLQLMGKPEPATGAKKSNKPNEQTSKRHSDTFPEL KLTNAPGSFTKCSNTSELKEFVNPSLPREEKEEKLETVKVSNNAEDPKDMLLSGERVLQTERSVESSSIS LVPGTDTGQTESISLLEVSTLKGAKTEPNKCVSQCAAFENPKGLIHGCSKDNRNDTEGFKYPLGHEVNSH RETSIEMEESGLDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKKQSPKVTFECEQKEENQ GKNESNIKPVQTVNIITAGFPVVGQKQKDPVDNAKCSIKGGSRFCLSSQFRGNETGLITPNKHGLLQNPYRI PPLFPKISFVKTKCKKNLLEENFEHSMSPEREMGNENIPSTVSTISRNNIRENVFKEASSNINEVGSS TNEVGSSINEIGSSDENIQAE LGRNRGPKLNAMLRLGVLQPEVYKQSLPGSNCKHPEIKKQYEEVVQTV NTDFSPYILSDNLEQPMGSSSHASQVCSETPDLLDDGEIKEDTSAENDIKESSAVFSKSVQKGLSRSP SPFTHTHLAQGYRRGAKKLESSEENLSEDEELPCFQHLFLGKVNIPSPQSTRHSTVATECLSKNTEENL LSLKNSLNDSCNQVILAKASQEHHLSEETKCSASLFSQQCELEDLTANTNTQDPFLIGSSQMRHQSES QGVGLSDKELVSDDEERGTGLEENNQEEQSMDSNLGEAASGCESETSVSEDCSGLSSQSDILTTQQRDTM QHNLIKQLQEMAELEAVLEQHGSSQPSNSYPSIISDSSALEDLRNPEQSTSEKAVLTSQKSSEYPISQNP GLSADKFEVSDSSSKNKEPGVRESKCPSLDDRWMHSCSGSLQNRNYPSEELIKVVDVEEQGLE ESGPHDLTETSYLPRQDLEGTPLYESGISLFSDDPESDPEDRAPESARVGNIPSSTSALKVPQLKVAES AQSPAAHTTDTAGYNAMEEVSREKPELTASTERNKRMMSVVSGLTPEEFMLVYKFARKHHITLNLNI TEETHVVMKTADEFVCERTILKYFLGIAGGKWVVSYFWVTQSIKERKMLNEHDFEVRGDDVNGRNNHQGPK RARESQRKIFRGLEICCYGPFNTMPDQLEWMVQLCGASVKELSSFTLGTGVHPIVVVQDPWATEDNG FHAIGQMCEAPVVTREWVLDVALYQCQELDTYLIPQIPHSHY (breast cancer type 1 susceptibility protein isoform 1 [Homo sapiens] CCDS 11453.1) MLKLLNQKKGPSQCPLCKNDITKRSQESTRFSQLVEELLKIIICAFQLDTGLEANSYNFAKKENNSPEH LKDEVSI IQSMGYRNRARLLQSEPENPSLQETSLSVQLSNLGTVRTLRTKQRIQPQKTSVYIELGSDS EDTVNKATYCSVGDQELLQITPQGTREISLDSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERH PEKYQGSSVSNLHVPECGTNTHASSLQHENSLLLTCDRMNVEKAIEFCNKSQKQGLARSQHNWRWAGSKET	16305- 16309

	CNDRRTPESTEKKVDLNDPLCERKEWNQKQLPCSENPRDTEVPWITLNSSIQKVNEWFSRSDDELLGSDSDHGESESNKAVADVLDVLENEVDEYSGSSEKIDLLASDPHEALICKSERVHKSVESENIEDKIFGKTYRKASLPNLSHVTENLIIGAFVTEPQIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELELNHNKAPKKNRLRRKSSSTRHIALELVVSRNLSPNCTELQIDSCSSSEIKKKKYNQMPVRHSRNLQLMGKEPATGAKKSNKPNEQTSKRHSDSTFPPELKLTNAPGSFTKCSNTSELKEFVNPSLPREKEEKLETVKVSNNAEDEPKDMLSGSERVLQTERSVESSSISLVPGTQYGTQESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNHRSRETSIEMEESLDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKKQSPKVTFECEQKEENQKKNESNIKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPKISFVTKCKKNLLEENFEHSMSPEREMGNENIPSTVSTISRNNIRENVFKEASSSINEVGSSSTNEVGSSINEIGSSDENIAELGRNRGPKLNAMRLGLVQPEVYKQSLPGSNCKHPEIKKQYEEVVQTVNTDFSPYILSDNLEQPMGSSHASQVCSETPDDLDDGEIKEDTSFAENDIKESSAVFSKSVQKELSRSPSPFTHTHLAQGYRRGAKKLESSEENLSEDEELPCFQHLFLGKVNINIPQSSTRHSTVATECLSKNTEENLLSLKNSLNDCSNQVILAKASQEHHLSEETKCSASLFSQCSELEDLTANTNTQDPFLIGSSQKMRHQSSESQGVGLSDKELVSDDEERGTLGLENNQEEQSMDSNLGEAASGCESETSVSEDCSGLSSQSDILTTQQRDTMQHNLIKLQQEMAELEAVLEQHGSGQPSNSYPSIISDSSALEDLRNPEQSTSEKVDHISVLTSQKSSEYPISQNPEGLSADKFEVSADSSTSKNKEPGVERSSPSKCPSLDDRWMHSCSGSLQNRNYPSEELIKVVDVEEQLEESGPHDLTETSYLPRQDLEGTPLYESGISLFSDDPESDPSEDRAPE SARVGNIPSSSALKVPQLKVAESAQSPAAHTTDTAGYNAMEESVSREKPELTASTERNKRMMSVSGLTPEEFMLVYKFARKHHITLTNLI TEETHVVMKTDAEFVCERTLKYFLGIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDDVVNGRNHQQPKRARESQRKIFRGLICCYGPFTNMPTDQLEWMVQLCGASVVKELSSFTLGTGVHPIVVQPDWATEDNGFHAIGQMCEAPVVTREWVLDVALYQCQELDTYLIPQIPHSHY (breast cancer type 1 susceptibility protein isoform 2 [Homo sapiens] CCDS 11459.2) MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITK RSLQESTRFSQLVEELLKIIICAFQLDTGLEANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRKRLQSEPENPSLQETSLSVQLSNLGTVRTLRTRKQRIQPKTSVYIELGSDSSEDTVKNATYCSVGDQELLQITPQ GTRDEISLDSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERHPEKYQGSSVSNLHVEPCGTNTHA SSLQHENSLLLTkdRMNVEKAEEFCNKSQKQGLARSQHNRWAGSKETCNDRRTPESTEKKVDLNDPLCER KEWNQKQLPCSENPRDTEVPWITLNSSIQKVNEWFSRSDDELLGSDSDHGESESNKAVADVLDVLENEVD EYSGSEKIDLLASDPHEALICKSERVHKSVESENIEDKIFGKTYRKASLPNLSHVTENLIIGAFVTEP QIIQERPLTNKLRKRRTSGLHPEDFIKKADLAVQKTPEMINQGTNQTEQNGQVMNITNSGHENKTKGDSIQNEKNPNPIESLEKESAFKTKAEPISSSISNMELELNHNKAPKKNRLRRKSSSTRHIALELVVSRNLSPNCTELQIDSCSSSEIKKKKYNQMPVRHSRNLQLMGKEPATGAKKSNKPNEQTSKRHSDSTFPPELKL TNAPGSFTKCSNTSELKEFVNPSLPREKEEKLETVKVSNNAEDEPKDMLSGSERVLQTERSVESSSISLVPGTQYGTQESISLLEVSTLGKAKTEPNKCVSQCAAFENPKGLIHGCSKDNNDTEGFKYPLGHEVNHRSRETSIEMEESLDAQYLQNTFKVSKRQSFAPFSNPGNAEEECATFSAHSGSLKKQSPKVTFECEQKEENQKKNESNIKPVQTVNITAGFPVVGQKDKPVDNAKCSIKGGSRFCLSSQFRGNETGLITPNKHGLLQNPYRIPPLFPKISFVTKCKKNLLEENFEHSMSPEREMGNENIPSTVSTISRNNIRENVFKEASSSINEVGSSSTNEVGSSINEIGSSDENIAELGRNRGPKLNAMRLGLVQPEVYKQSLPGSNCKHPEIKKQYEEVVQTVNTDFSPYILSDNLEQPMGSSHASQVCSETPDDLDDGEIKEDTSFAENDIKESSAVFSKSVQKELSRSPSPFTHTHLAQGYRRGAKKLESSEENLSEDEELPCFQHLFLGKVNINIPQSSTRHSTVATECLSKNTEENLLSLKNSLNDCSNQVILAKASQEHHLSEETKCSASLFSQCSELEDLTANTNTQDPFLIGSSQKMRHQSSESQGVGLSDKELVSDDEERGTLGLENNQEEQSMDSNLGEAASGCESETSVSEDCSGLSSQSDILTTQQRDTMQHNLIKLQQEMAELEAVLEQHGSGQPSNSYPSIISDSSALEDLRNPEQSTSEKVDHISVLTSQKSSEYPISQNPEGLSADKFEVSADSSTSKNKEPGVERSSPSKCPSLDDRWMHSCSGSLQNRNYPSEELIKVVDVEEQLEESGPHDLTETSYLPRQDLEGTPLYESGISLFSDDPESDPSEDRAPE SARVGNIPSSSALKVPQLKVAESAQSPAAHTTDTAGYNAMEESVSREKPELTASTERNKRMMSVSGLTPEEFMLVYKFARKHHITLTNLI TEETHVVMKTDAEFVCERTLKYFLGIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDDVVNGRNHQQPKRARESQRKIFRGLICCYGPFTNMPTDQLEWMVQLCGASVVKELSSFTLGTGVHPIVVQPDWATEDNGFHAIGQMCEAPVVTREWVLDVALYQCQELDTYLIPQIPHSHY (breast cancer type 1 susceptibility protein isoform 2 [Homo sapiens], CCDS 11456.2) MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITK RSLQESTRFSQLVEELLKIIICAFQLDTGLEANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRKRLQSEPENPSLQETSLSVQLSNLGTVRTLRTRKQRIQPKTSVYIELGSDSSEDTVKNATYCSVGDQELLQITPQ GTRDEISLDSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERHPEKYQGEAASGCESETSVSEDCS GLSSQSDILTTQQRDTMQHNLIKLQQEMAELEAVLEQHGSGQPSNSYPSIISDSSALEDLRNPEQSTSEKVLTSQKSSEYPISQNPEGLSADKFEVSADSSTSKNKEPGVERSSPSKCPSLDDRWMHSCSGSLQNRNYPSEELIKVVDVEEQLEESGPHDLTETSYLPRQDLEGTPLYESGISLFSDDPESDPSEDRAPE SARVGNIPSSSALKVPQLKVAESAQSPAAHTTDTAGYNAMEESVSREKPELTASTERNKRMMSVSGLTPEEFMLVYKFARKHHITLTNLI TEETHVVMKTDAEFVCERTLKYFLGIAGGKVVVSYFWVTQSIKERKMLNEHDFEVRGDDVVNGRNHQQPKRARESQRKIFRGLICCYGPFTNMPTDQLEWMVQLCGASVVKELSSFTLGTGVHPIVVQPDWATEDNGFHAIGQMCEAPVVTREWVLDVALYQCQELDTYLIPQIPHSHY (breast cancer type 1 susceptibility protein isoform 2 [Homo sapiens] CCDS 11454.2) MDLSALRVEEVQNVINAMQKILECPICLELIKEPVSTKCDHIFCKFCMLKLLNQKKGPSQCPLCKNDITK RSLQESTRFSQLVEELLKIIICAFQLDTGLEANSYNFAKKENNSPEHLKDEVSIIQSMGYRNRKRLQSEPENPSLQETSLSVQLSNLGTVRTLRTRKQRIQPKTSVYIELGSDSSEDTVKNATYCSVGDQELLQITPQ GTRDEISLDSAKKAACEFSETDVTNTEHHQPSNNDLNTTEKRAAERHPEKYQGEAASGCESETSVSEDCS GLSSQSDILTTQQRDTMQHNLIKLQQEMAELEAVLEQHGSGQPSNSYPSIISDSSALEDLRNPEQSTSEKVLTSQKSSEYPISQNPEGLSADKFEVSADSSTSKNKEPGVERSSPSKCPSLDDRWMHSCSGSLQNRNYPSEELIKVVDVEEQLEESGPHDLTETSYLPRQDLEGTPLYESGISLFSDDPESDPSEDRAPE SARVGNIP	
--	---	--

	SSTSALKVPQLKVAESAQSPAAHTTDTAGYNAMEESVSREKPELTASTERVNKRMSMVVSGLTPEEFML VYKFARKHHITLNLITEETHVVMKTDAEFVCERTLKYFLGIAGGKWVVSFYFWVTQSIKERKMLNEHDF EVRGDEVNNGRNHQGPKRARESQRKIFRGLEICCYGPFNTMPTGCPPNCGCAARCLDRGQWLPCNWADV (breast cancer type 1 susceptibility protein isoform 2 [Homo sapiens] CCDS 11455.2)	
RAD51	MAMQMQLLEANADTSVEEESFGPQPISRLEQCGINANDVKKLEEAGFHTVEAVAYAPKKELINIKGISEAK ADKILAEAAKLVPMTTATEFHQRRSEIIQITTGSKELDKLLQGGIETGSITEMFGEFRTGKTQICHTL AVTCQLPIDRGGGEGKAMYIDTEGTFRPERLLAVAERYGLSGSDVLDNVAYARAFNTDHQTQLLYQASAM MVESRYALLIVDSATALYRTDYSGRGELSARQMHLARFLRMLLRLADEFVAVVITNQVVAQVDGAAMFA ADPKKPIGGNIIAHASTTRLYLRKGRGETRICKIYDSPCLPEAEAMFAINADGVGDAKD (RAD51 [Homo sapiens], CCDS 10062.1) MAMQMQLLEANADTSVEEESFGPQPISRLEQCGINANDVKKLEEAGFHTVEAVAYAPKKELINIKGISEAK ADKILTESRSVARLECNVILVYCTLRSLSGSSDSPASASRVVGTGGIETGSITEMFGEFRTGKTQICHT LAVTCQLPIDRGGGEGKAMYIDTEGTFRPERLLAVAERYGLSGSDVLDNVAYARAFNTDHQTQLLYQASAM MMVESRYALLIVDSATALYRTDYSGRGELSARQMHLARFLRMLLRLADEFVAVVITNQVVAQVDGAAMFA AADPKKPIGGNIIAHASTTRLYLRKGRGETRICKIYDSPCLPEAEAMFAINADGVGDAKD (RAD51 [Homo sapiens], CCDS 53931.1) MAMQMQLLEANADTSVEEESFGPQPISRLEQCGINANDVKKLEEAGFHTVEAVAYAPKKELINIKGISEAK ADKILAEAAKLVPMTTATEFHQRRSEIIQITTGSKELDKLLQGGIETGSITEMFGEFRTGKTQICHTL AVTCQLPIDRGGGEGKAMYIDTEGTFRPERLLAVAERYGLSGSDVLDNVAYARAFNTDHQTQLLYQASAM MVESRYALLIVDSATALYRTDYSGRGELSARQMHLARFLRMLLRLADEFVAVVITNQVVAQVDGAAMFA AADPKKPIGGNIIAHASTTRLYLRKGRGETRICKIYDSPCLPEAEAMFAINADGVGDAKD (RAD51 [Homo sapiens], CCDS 53932.1)	16310– 16312

In certain embodiments of the methods provided herein, the frequency of preferred repair outcomes generated using a Cas9 molecule described herein may be increased by modulating (*e.g.*, stimulating or overexpressing) a component (*e.g.*, exactly one component, or one or more components, *e.g.*, two or three components) of a gene conversion pathway, *e.g.*, a BRCA1, BRCA2 and/or RAD51. In some embodiments, the frequency of gene conversion resulting from a Cas9 nuclease (*e.g.*, wild type Cas9 nuclease) induced-lesion in a target position of a target cell overexpressing a gene conversion 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 conversion resulting from a Cas9 nickase (*e.g.*, a Cas9 D10A nickase) induced-lesion in a target position in the absence of overexpression of a gene conversion pathway.

In some embodiments, the frequency of gene conversion resulting from a Cas9 nuclease (*e.g.*, wild type Cas9 nuclease) induced-lesion in a target position of a target cell overexpressing a gene conversion 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, as compared to the frequency of gene conversion

resulting from a Cas9 nickase (*e.g.*, a Cas9 D10A nickase) induced-lesion in a target position in the absence of overexpression of a gene conversion pathway.

In certain embodiments of the methods provided herein, the frequency of preferred repair outcomes generated using a Cas9 molecule described herein may be increased by modulating (*e.g.*, stimulating or overexpressing) a component (*e.g.*, exactly one component, or one or more components, *e.g.*, two or three components) of a gene conversion pathway, *e.g.*, a BRCA1, BRCA2 and/or RAD51. In some embodiments, the frequency of gene conversion resulting from a Cas9 nickase (*e.g.*, a Cas9 D10A nickase) induced-lesion in a target position of a target cell overexpressing a gene conversion 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 conversion resulting from a Cas9 nickase (*e.g.*, a Cas9 D10A nickase) induced-lesion in a target position in the absence of overexpression of a gene conversion pathway.

In some embodiments, the frequency of gene conversion resulting from a Cas9 nickase (*e.g.*, a Cas9 D10A nickase) induced-lesion in a target position of a target cell overexpressing a gene conversion 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, as compared to the frequency of gene conversion resulting from a Cas9 nickase (*e.g.*, a Cas9 D10A nickase) induced-lesion in a target position in the absence of overexpression of a gene conversion pathway.

V.2 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 *eaCas9* molecules and single strand, or nickase, *eaCas9* 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 Cas9 molecules and single strand, or nickase, Cas9 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.*,
 5 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
 10 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).

15 V.3 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
 20 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
 25 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
 30 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 an 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.

V.4 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.

V. 5 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
 5 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
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60
 65
 70
 75
 80
 85
 90
 95
 100
 105
 110
 115
 120
 125
 130
 135
 140
 145
 150
 155
 160
 165
 170
 175
 180
 185
 190
 195
 200
 205
 210
 215
 220
 225
 230
 235
 240
 245
 250
 255
 260
 265
 270
 275
 280
 285
 290
 295
 300
 305
 310
 315
 320
 325
 330
 335
 340
 345
 350
 355
 360
 365
 370
 375
 380
 385
 390
 395
 400
 405
 410
 415
 420
 425
 430
 435
 440
 445
 450
 455
 460
 465
 470
 475
 480
 485
 490
 495
 500
 505
 510
 515
 520
 525
 530
 535
 540
 545
 550
 555
 560
 565
 570
 575
 580
 585
 590
 595
 600
 605
 610
 615
 620
 625
 630
 635
 640
 645
 650
 655
 660
 665
 670
 675
 680
 685
 690
 695
 700
 705
 710
 715
 720
 725
 730
 735
 740
 745
 750
 755
 760
 765
 770
 775
 780
 785
 790
 795
 800
 805
 810
 815
 820
 825
 830
 835
 840
 845
 850
 855
 860
 865
 870
 875
 880
 885
 890
 895
 900
 905
 910
 915
 920
 925
 930
 935
 940
 945
 950
 955
 960
 965
 970
 975
 980
 985
 990
 995
 1000
 1005
 1010
 1015
 1020
 1025
 1030
 1035
 1040
 1045
 1050
 1055
 1060
 1065
 1070
 1075
 1080
 1085
 1090
 1095
 1100
 1105
 1110
 1115
 1120
 1125
 1130
 1135
 1140
 1145
 1150
 1155
 1160
 1165
 1170
 1175
 1180
 1185
 1190
 1195
 1200
 1205
 1210
 1215
 1220
 1225
 1230
 1235
 1240
 1245
 1250
 1255
 1260
 1265
 1270
 1275
 1280
 1285
 1290
 1295
 1300
 1305
 1310
 1315
 1320
 1325
 1330
 1335
 1340
 1345
 1350
 1355
 1360
 1365
 1370
 1375
 1380
 1385
 1390
 1395
 1400
 1405
 1410
 1415
 1420
 1425
 1430
 1435
 1440
 1445
 1450
 1455
 1460
 1465
 1470
 1475
 1480
 1485
 1490
 1495
 1500
 1505
 1510
 1515
 1520
 1525
 1530
 1535
 1540
 1545
 1550
 1555
 1560
 1565
 1570
 1575
 1580
 1585
 1590
 1595
 1600
 1605
 1610
 1615
 1620
 1625
 1630
 1635
 1640
 1645
 1650
 1655
 1660
 1665
 1670
 1675
 1680
 1685
 1690
 1695
 1700
 1705
 1710
 1715
 1720
 1725
 1730
 1735
 1740
 1745
 1750
 1755
 1760
 1765
 1770
 1775
 1780
 1785
 1790
 1795
 1800
 1805
 1810
 1815
 1820
 1825
 1830
 1835
 1840
 1845
 1850
 1855
 1860
 1865
 1870
 1875
 1880
 1885
 1890
 1895
 1900
 1905
 1910
 1915
 1920
 1925
 1930
 1935
 1940
 1945
 1950
 1955
 1960
 1965
 1970
 1975
 1980
 1985
 1990
 1995
 2000
 2005
 2010
 2015
 2020
 2025
 2030
 2035
 2040
 2045
 2050
 2055
 2060
 2065
 2070
 2075
 2080
 2085
 2090
 2095
 2100
 2105
 2110
 2115
 2120
 2125
 2130
 2135
 2140
 2145
 2150
 2155
 2160
 2165
 2170
 2175
 2180
 2185
 2190
 2195
 2200
 2205
 2210
 2215
 2220
 2225
 2230
 2235
 2240
 2245
 2250
 2255
 2260
 2265
 2270
 2275
 2280
 2285
 2290
 2295
 2300
 2305
 2310
 2315
 2320
 2325
 2330
 2335
 2340
 2345
 2350
 2355
 2360
 2365
 2370
 2375
 2380
 2385
 2390
 2395
 2400
 2405
 2410
 2415
 2420
 2425
 2430
 2435
 2440
 2445
 2450
 2455
 2460
 2465
 2470
 2475
 2480
 2485
 2490
 2495
 2500
 2505
 2510
 2515
 2520
 2525
 2530
 2535
 2540
 2545
 2550
 2555
 2560
 2565
 2570
 2575
 2580
 2585
 2590
 2595
 2600
 2605
 2610
 2615
 2620
 2625
 2630
 2635
 2640
 2645
 2650
 2655
 2660
 2665
 2670
 2675
 2680
 2685
 2690
 2695
 2700
 2705
 2710
 2715
 2720
 2725
 2730
 2735
 2740
 2745
 2750
 2755
 2760
 2765
 2770
 2775
 2780
 2785
 2790
 2795
 2800
 2805
 2810
 2815
 2820
 2825
 2830
 2835
 2840
 2845
 2850
 2855
 2860
 2865
 2870
 2875
 2880
 2885
 2890
 2895
 2900
 2905
 2910
 2915
 2920
 2925
 2930
 2935
 2940
 2945
 2950
 2955
 2960
 2965
 2970
 2975
 2980
 2985
 2990
 2995
 3000
 3005
 3010
 3015
 3020
 3025
 3030
 3035
 3040
 3045
 3050
 3055
 3060
 3065
 3070
 3075
 3080
 3085
 3090
 3095
 3100
 3105
 3110
 3115
 3120
 3125
 3130
 3135
 3140
 3145
 3150
 3155
 3160
 3165
 3170
 3175
 3180
 3185
 3190
 3195
 3200
 3205
 3210
 3215
 3220
 3225
 3230
 3235
 3240
 3245
 3250
 3255
 3260
 3265
 3270
 3275
 3280
 3285
 3290
 3295
 3300
 3305
 3310
 3315
 3320
 3325
 3330
 3335
 3340
 3345
 3350
 3355
 3360
 3365
 3370
 3375
 3380
 3385
 3390
 3395
 3400
 3405
 3410
 3415
 3420
 3425
 3430
 3435
 3440
 3445
 3450
 3455
 3460
 3465
 3470
 3475
 3480
 3485
 3490
 3495
 3500
 3505
 3510
 3515
 3520
 3525
 3530
 3535
 3540
 3545
 3550
 3555
 3560
 3565
 3570
 3575
 3580
 3585
 3590
 3595
 3600
 3605
 3610
 3615
 3620
 3625
 3630
 3635
 3640
 3645
 3650
 3655
 3660
 3665
 3670
 3675
 3680
 3685
 3690
 3695
 3700
 3705
 3710
 3715
 3720
 3725
 3730
 3735
 3740
 3745
 3750
 3755
 3760
 3765
 3770
 3775
 3780
 3785
 3790
 3795
 3800
 3805
 3810
 3815
 3820
 3825
 3830
 3835
 3840
 3845
 3850
 3855
 3860
 3865
 3870
 3875
 3880
 3885
 3890
 3895
 3900
 3905
 3910
 3915
 3920
 3925
 3930
 3935
 3940
 3945
 3950
 3955
 3960
 3965
 3970
 3975
 3980
 3985
 3990
 3995
 4000
 4005
 4010
 4015
 4020
 4025
 4030
 4035
 4040
 4045
 4050
 4055
 4060
 4065
 4070
 4075
 4080
 4085
 4090
 4095
 4100
 4105
 4110
 4115
 4120
 4125
 4130
 4135
 4140
 4145
 4150
 4155
 4160
 4165
 4170
 4175
 4180
 4185
 4190
 4195
 4200
 4205
 4210
 4215
 4220
 4225
 4230
 4235
 4240
 4245
 4250
 4255
 4260
 4265
 4270
 4275
 4280
 4285
 4290
 4295
 4300
 4305
 4310
 4315
 4320
 4325
 4330
 4335
 4340
 4345
 4350
 4355
 4360
 4365
 4370
 4375
 4380
 4385
 4390
 4395
 4400
 4405
 4410
 4415
 4420
 4425
 4430
 4435
 4440
 4445
 4450
 4455
 4460
 4465
 4470
 4475
 4480
 4485
 4490
 4495
 4500
 4505
 4510
 4515
 4520
 4525
 4530
 4535
 4540
 4545
 4550
 4555
 4560
 4565
 4570
 4575
 4580
 4585
 4590
 4595
 4600
 4605
 4610
 4615
 4620
 4625
 4630
 4635
 4640
 4645
 4650
 4655
 4660
 4665
 4670
 4675
 4680
 4685
 4690
 4695
 4700
 4705
 4710
 4715
 4720
 4725
 4730
 4735
 4740
 4745
 4750
 4755
 4760
 4765
 4770
 4775
 4780
 4785
 4790
 4795
 4800
 4805
 4810
 4815
 4820
 4825
 4830
 4835
 4840
 4845
 4850
 4855
 4860
 4865
 4870
 4875
 4880
 4885
 4890
 4895
 4900
 4905
 4910
 4915
 4920
 4925
 4930
 4935
 4940
 4945
 4950
 4955
 4960
 4965
 4970
 4975
 4980
 4985
 4990
 4995
 5000
 5005
 5010
 5015
 5020
 5025
 5030
 5035
 5040
 5045
 5050
 5055
 5060
 5065
 5070
 5075
 5080
 5085
 5090
 5095
 5100
 5105
 5110
 5115
 5120
 5125
 5130
 5135
 5140
 5145
 5150
 5155
 5160
 5165
 5170
 5175
 5180
 5185
 5190
 5195
 5200
 5205
 5210
 5215
 5220
 5225
 5230
 5235
 5240
 5245
 5250
 5255
 5260
 5265
 5270
 5275
 5280
 5285
 5290
 5295
 5300
 5305
 5310
 5315
 5320
 5325
 5330
 5335
 5340
 5345
 5350
 5355
 5360
 5365
 5370
 5375
 5380
 5385
 5390
 5395
 5400
 5405
 5410
 5415
 5420
 5425
 5430
 5435
 5440
 5445
 5450
 5455
 5460
 5465
 5470
 5475
 5480
 5485
 5490
 5495
 5500
 5505
 5510
 5515
 5520
 5525
 5530
 5535
 5540
 5545
 5550
 5555
 5560
 5565
 5570
 5575
 5580
 5585
 5590
 5595
 5600
 5605
 5610
 5615
 5620
 5625
 5630
 5635
 5640
 5645
 5650
 5655
 5660
 5665
 5670
 5675
 5680
 5685
 5690
 5695
 5700
 5705
 5710
 5715
 5720
 5725
 5730
 5735
 5740
 5745
 5750
 5755
 5760
 5765
 5770
 5775
 5780
 5785
 5790
 5795
 5800
 5805
 5810
 5815
 5820
 5825
 5830
 5835
 5840
 5845
 5850
 5855
 5860
 5865
 5870
 5875
 5880
 5885
 5890
 5895
 5900
 5905
 5910
 5915
 5920
 5925
 5930
 5935
 5940
 5945
 5950
 5955
 5960
 5965
 5970
 5975
 5980
 5985
 5990
 5995
 6000
 6005
 6010
 6015
 6020
 6025
 6030
 6035
 6040
 6045
 6050
 6055
 6060
 6065
 6070
 6075
 6080
 6085
 6090
 6095
 6100
 6105
 6110
 6115
 6120
 6125
 6130
 6135
 6140
 6145
 6150
 6155
 6160
 6165
 6170
 6175
 6180
 6185
 6190
 6195
 6200
 6205
 6210
 6215
 6220
 6225
 6230
 6235
 6240
 6245
 6250
 6255
 6260
 6265
 6270
 6275
 6280
 6285
 6290
 6295
 6300
 6305
 6310
 6315
 6320
 6325
 6330
 6335
 6340
 6345
 6350
 6355
 6360
 6365
 6370
 6375
 6380
 6385
 6390
 6395
 6400
 6405
 6410
 6415
 6420
 6425
 6430
 6435
 6440
 6445
 6450
 6455
 6460
 6465
 6470
 6475
 6480
 6485
 6490
 6495
 6500
 6505
 6510
 6515
 6520
 6525
 6530
 6535
 6540
 6545
 6550
 6555
 6560
 6565
 6570
 6575
 6580
 6585
 6590
 6595
 6600
 6605
 6610
 6615
 6620
 6625
 6630
 6635
 6640
 6645
 6650
 6655
 6660
 6665
 6670
 6675
 6680
 6685
 6690
 6695
 6700
 6705
 6710
 6715
 6720
 6725
 6730
 6735
 6740
 6745
 6750
 6755
 6760
 6765
 6770
 6775
 6780
 6785
 6790
 6795
 6800
 6805
 6810

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 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 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 having the RuvC activity inactivated, *e.g.*, a Cas9 molecule having a mutation at D10, *e.g.*, the D10A mutation.

In an embodiment, the gRNA is used with a Cas9 nickase molecule having RuvC activity, *e.g.*, a Cas9 molecule having the HNH activity inactivated, *e.g.*, a Cas9 molecule having a mutation at 840, *e.g.*, the H840A.

In an embodiment, the gRNAs are used with a Cas9 nickase molecule having RuvC activity, *e.g.*, a Cas9 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 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

(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 having the RuvC activity inactivated, *e.g.*, a Cas9 molecule having a mutation at D10, *e.g.*, the D10A mutation.

In certain embodiments, the gRNAs are used with a Cas9 nickase molecule having RuvC activity, *e.g.*, a Cas9 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 having the HNH activity inactivated, *e.g.*, a Cas9 molecule having a mutation at N863, *e.g.*, the N863A mutation.

VI. Target Cells

Cas9 molecules and gRNA molecules, *e.g.*, a Cas9 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 (*e.g.*, a *HBB* 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 *HBB* target gene, *e.g.*, as described herein. In one embodiment, the expression of the *HBB* target gene is modulated, *e.g.*, *in vivo*. In another embodiment, the expression of the *HBB* target gene is modulated, *e.g.*, *ex vivo*.

The Cas9 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 *HBB* gene and/or modulating the expression of the *HBB* 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 myeloablative pre-conditioning. In other embodiments, the modified HSCs are administered after mild myeloblastic conditioning such that, following engraftment, some of the hematopoietic cells are derived from the modified HSCs. In still other embodiments, the modified HSCs are administered after full myeloblation 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 an 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.

- 5 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.

10

VII. Delivery, Formulations and Routes of Administration

- The components, *e.g.*, a Cas9 molecule and gRNA molecule (*e.g.*, a Cas9 molecule/gRNA molecule complex), and a donor template nucleic acid, or all three, can be delivered, formulated, or administered, in a variety of forms, see, *e.g.*, **Tables 6-7**. In certain
- 15 embodiments, one Cas9 molecule and two or more (*e.g.*, 2, 3, 4, or more) different gRNA molecules are delivered, *e.g.*, by an AAV vector. In certain embodiments, the sequence encoding the Cas9 molecule and 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 Cas9 or gRNA component is delivered encoded in DNA, the DNA will
- 20 typically include a control region, *e.g.*, comprising a promoter, to effect expression. Useful promoters for Cas9 molecule sequences include CMV, SFFV, EFS, EF-1a, PGK, CAG, and CBH promoters, or a blood cell specific promoter. In an 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.
- 25 Promoters with similar or dissimilar strengths can be selected to tune the expression of components. Sequences encoding a Cas9 molecule can comprise a nuclear localization signal (NLS), *e.g.*, an SV40 NLS. In an embodiment, the sequence encoding a Cas9 molecule comprises at least two nuclear localization signals. In an embodiment a promoter for a Cas9 molecule or a gRNA molecule can be, independently, inducible, tissue specific, or cell
- 30 specific.

Table 6 provides examples of how the components can be formulated, delivered, or administered.

Table 6

Elements			
Cas9 Molecule(s)	gRNA Molecule(s)	Optional Donor Template Nucleic Acid	Comments
DNA	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA molecule are transcribed from DNA. In this embodiment, they are encoded on separate molecules. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	DNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA molecule are transcribed from DNA. In this embodiment, they are encoded on separate molecules. In this embodiment, the donor template is provided on the same DNA molecule that encodes the gRNA molecule.
DNA		DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA molecule are transcribed from DNA, here from a single molecule. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, and a gRNA molecule are transcribed from DNA. In this embodiment, they are encoded on separate molecules. In this embodiment, the donor template is provided on the same DNA molecule that encodes the Cas9 molecule.
DNA	RNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is transcribed from DNA, and a gRNA molecule is provided as <i>in vitro</i> transcribed or synthesized RNA. In this embodiment, the donor template is provided as a separate DNA molecule.
DNA	RNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is transcribed from DNA, and a gRNA molecule is provided as <i>in vitro</i> transcribed or synthesized RNA. In this embodiment, the donor template is provided on the same DNA molecule that encodes the Cas9 molecule.
mRNA	RNA	DNA	In this embodiment, a Cas9 molecule,

			typically an eaCas9 molecule, is translated from <i>in vitro</i> transcribed mRNA, and a gRNA molecule is provided as <i>in vitro</i> transcribed or synthesized RNA. In this embodiment, the donor template is provided as a DNA molecule.
mRNA	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is translated from <i>in vitro</i> transcribed mRNA, and a gRNA molecule is transcribed from DNA. In this embodiment, the donor template is provided as a separate DNA molecule.
mRNA	DNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is translated from <i>in vitro</i> transcribed mRNA, and a gRNA molecule is transcribed from DNA. In this embodiment, the donor template is provided on the same DNA molecule that encodes the gRNA molecule.
Protein	DNA	DNA	In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is provided as a protein, and a gRNA molecule is transcribed from DNA. In this embodiment, the donor template is provided as a separate DNA molecule.
Protein	DNA		In this embodiment, a Cas9 molecule, typically an eaCas9 molecule, is provided as a protein, and a gRNA is transcribed from DNA. In this embodiment, the donor template is provided on the same DNA molecule that encodes the gRNA molecule.
Protein	RNA	DNA	In this embodiment, an eaCas9 molecule is provided as a protein, and a gRNA molecule is provided as transcribed or synthesized RNA. In this embodiment, the donor template is provided as a DNA molecule.

Table 7 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 7

Delivery Vector/Mode	Delivery into Non-Dividing	Duration of Expression	Genome Integration	Type of Molecule Delivered
----------------------	----------------------------	------------------------	--------------------	----------------------------

		Cells			
	Physical (e.g., electroporation, particle gun, calcium phosphate transfection, cell compression or squeezing)	YES	Transient	NO	Nucleic Acids and Proteins
Viral	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 a Cas9 molecule and/or one or more gRNA molecule and/or a donor template

5 Nucleic acids encoding Cas9 molecules (e.g., eaCas9 molecules), gRNA molecules, donor template nucleic acid, or any combination (e.g., two or all) thereof, can be administered to subjects or delivered into cells by art-known methods or as described herein. For example, Cas9-encoding and/or gRNA-encoding DNA, as well as donor template nucleic

acids, 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 Cas9 molecules (*e.g.*, eaCas9 molecules) and/or gRNA molecules can be conjugated to molecules (*e.g.*, N-acetylgalactosamine) promoting uptake by the target cells (*e.g.*, erythrocytes, HSCs). Donor template molecules can likewise be conjugated to molecules (*e.g.*, N-acetylgalactosamine) promoting uptake by the target cells (*e.g.*, erythrocytes, HSCs).

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a vector (*e.g.*, viral vector/virus or plasmid).

Vectors can comprise a sequence that encodes a Cas9 molecule and/or a gRNA molecule and/or a donor template with high homology to the region (*e.g.*, target sequence) being targeted. In certain embodiments, the donor template comprises all or part of a target sequence. Exemplary donor templates are a repair template, *e.g.*, a gene correction template, or a gene mutation template, *e.g.*, point mutation (*e.g.*, single nucleotide (nt) substitution) template. A vector can also comprise a sequence encoding a signal peptide (*e.g.*, for nuclear localization, nucleolar localization, or mitochondrial localization), fused, *e.g.*, to a Cas9 molecule sequence. For example, the vectors can comprise a nuclear localization sequence (*e.g.*, from SV40) fused to the sequence encoding the Cas9 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 Cas9 molecule and/or 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 Cas9 molecule and/or 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 an 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 Cas9- and/or 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.

In an embodiment, the Cas9- and/or gRNA-encoding nucleic acid sequence is delivered by a recombinant lentivirus. In an embodiment, the donor template nucleic acid 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 Cas9- and/or gRNA-encoding nucleic acid sequence is delivered by a recombinant adenovirus. In an embodiment, the donor template nucleic acid is

delivered by a recombinant adenovirus. In some embodiments, the adenovirus is engineered to have reduced immunity in human.

In some embodiments, the Cas9- and/or gRNA-encoding nucleic acid sequence is delivered by a recombinant AAV. In an embodiment, the donor template nucleic acid 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 which anneal together to form double stranded DNA.

In an 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.

In an embodiment, the Cas9- and/or 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.

In an embodiment, the Cas9- and/or gRNA-encoding DNA is delivered by a chimeric AAV capsid. In an embodiment, the donor template nucleic acid 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 an 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.

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a hybrid virus, *e.g.*, a hybrid of one or more of the viruses described herein. In an 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), with other viral sequences being replaced

by an expression cassette encoding the protein to be expressed, *e.g.* Cas9. 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 (Cas9 and gRNA) to only the target cell. The specificity of the vector can also be mediated by microRNA-dependent control of transgene expression. In an 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 an 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.

In some embodiments, the Cas9- and/or 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 combination thereof.

In an embodiment, delivery via electroporation comprises mixing the cells with the Cas9-and/or gRNA-encoding DNA in a cartridge, chamber or cuvette and applying one or more electrical impulses of defined duration and amplitude. In an embodiment, delivery via electroporation is performed using a system in which cells are mixed with the Cas9-and/or gRNA-encoding DNA 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.

In some embodiments, the Cas9- and/or gRNA-encoding DNA is delivered by a combination of a vector and a non-vector based method. In an embodiment, the donor template nucleic acid is delivered by a combination of a vector and a non-vector based method. For example, virosomes combine liposomes with an 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) a sequence encoding a gRNA molecule comprising a targeting domain that is complementary with a target domain in the *HBB* gene, and (b) a sequence encoding a Cas9 molecule. In an embodiment, (a) and (b) are 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 an embodiment, the nucleic acid molecule is an AAV vector. Exemplary AAV vectors 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 another embodiment, (a) is 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 (b) is present on a second nucleic acid molecule, *e.g.*, a second vector, *e.g.*, a second vector, *e.g.*, a second AAV vector. The first and second nucleic acid molecules may be AAV vectors. In yet another embodiment, the nucleic acid may further comprise (c) a sequence that encodes a second, third and/or fourth gRNA molecule as described herein. In an embodiment, the

nucleic acid comprises (a), (b) and (c). Each of (a) and (c) 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 an embodiment, the nucleic acid molecule is an AAV vector.

5 In another embodiment, (a) and (c) are on different vectors. For example, (a) 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 (c) 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 an embodiment, the first and second nucleic acid molecules are AAV vectors. In yet another embodiment, each of (a), (b), and (c)
10 are present on the same nucleic acid molecule, *e.g.*, the same vector, *e.g.*, the same viral vector, *e.g.*, an AAV vector. In an embodiment, the nucleic acid molecule is an AAV vector. In an alternate embodiment, one of (a), (b), and (c) is encoded on a first nucleic acid molecule, *e.g.*, a first vector, *e.g.*, a first viral vector, *e.g.*, a first AAV vector; and a second and third of (a), (b), and (c) is encoded on a second nucleic acid molecule, *e.g.*, a second
15 vector, *e.g.*, a second vector, *e.g.*, a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

 In an embodiment, (a) is present on a first nucleic acid molecule, *e.g.*, a first vector, *e.g.*, a first viral vector, a first AAV vector; and (b) and (c) are present on a second nucleic acid molecule, *e.g.*, a second vector, *e.g.*, a second vector, *e.g.*, a second AAV vector. The
20 first and second nucleic acid molecule may be AAV vectors. In another embodiment, (b) is 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 (a) and (c) are present on a second nucleic acid molecule, *e.g.*, a second vector, *e.g.*, a second vector, *e.g.*, a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

25 In another embodiment, (c) is 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 (b) and (a) are present on a second nucleic acid molecule, *e.g.*, a second vector, *e.g.*, a second vector, *e.g.*, a second AAV vector. The first and second nucleic acid molecule may be AAV vectors.

 In another embodiment, each of (a), (b) and (c) are present on different nucleic acid
30 molecules, *e.g.*, different vectors, *e.g.*, different viral vectors, *e.g.*, different AAV vector. For example, (a) may be on a first nucleic acid molecule, (b) on a second nucleic acid molecule, and (c) on a third nucleic acid molecule. The first, second and third nucleic acid molecule may be AAV vectors.

In another embodiment, when a third and/or fourth gRNA molecule are present, each of (a), (b), (c)(i), (c)(ii) and (c)(iii) 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 an embodiment, the nucleic acid molecule is an AAV vector. In an alternate embodiment, each of (a), (b), (c)(i), (c)(ii) and (c)(iii) 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 (a), (b), (c)(i), (c)(ii) and (c)(iii) may be present on more than one nucleic acid molecule, but fewer than five nucleic acid molecules, *e.g.*, AAV vectors.

In another embodiment, when (d) a template nucleic acid is present, each of (a), (b), and (d) 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 an embodiment, the nucleic acid molecule is an AAV vector. In an alternate embodiment, each of (a), (b), and (d) 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 (a), (b), and (d) may be present on more than one nucleic acid molecule, but fewer than three nucleic acid molecules, *e.g.*, AAV vectors.

In another embodiment, when (d) a template nucleic acid is present, each of (a), (b), (c)(i) and (d) 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 an embodiment, the nucleic acid molecule is an AAV vector. In an alternate embodiment, each of (a), (b), (c)(i) and (d) 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 (a), (b), (c)(i) and (d) may be present on more than one nucleic acid molecule, but fewer than four nucleic acid molecules, *e.g.*, AAV vectors.

In another embodiment, when (d) a template nucleic acid is present, each of (a), (b), (c)(i), (c)(ii) and (d) 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 an embodiment, the nucleic acid molecule is an AAV vector. In an alternate embodiment, each of (a), (b), (c)(i), (c)(ii) and (d) 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 (a), (b), (c)(i), (c)(ii) and (d) may be present on more than one nucleic acid molecule, but fewer than five nucleic acid molecules, *e.g.*, AAV vectors.

In another embodiment, when (d) a template nucleic acid is present, each of (a), (b), (c)(i), (c)(ii), (c)(iii) and (d) 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 an embodiment, the nucleic acid molecule is an AAV vector. In an alternate embodiment, each of (a), (b), (c)(i), (c)(ii), (c)(iii) and (d) 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 (a), (b), (c)(i), (c)(ii), (c)(iii) and (d) may be present on more than one nucleic acid molecule, but fewer than six nucleic acid molecules, *e.g.*, AAV vectors.

The nucleic acids described herein may comprise a promoter operably linked to the sequence that encodes the gRNA molecule of (a), *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 of (c), *e.g.*, a promoter described herein. The promoter and second promoter differ from one another. In an embodiment, the promoter and second promoter are the same.

The nucleic acids described herein may further comprise a promoter operably linked to the sequence that encodes the Cas9 molecule of (b), *e.g.*, a promoter described herein.

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 an 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 8**.

Table 8. 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-	GAP-DLRIE	Cationic

propanaminium bromide		
Cetyltrimethylammonium bromide	CTAB	Cationic
6-Lauroxyhexyl ornithinate	LHON	Cationic
1-(2,3-Dioleoyloxypropyl)-2,4,6-trimethylpyridinium	2Oc	Cationic
2,3-Dioleoyloxy-N-[2(sperminecarboxamido-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
3 β -[N-(N',N'-Dimethylaminoethane)-carbamoyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Dioleoyloxy-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic
Dimethyloctadecylammonium bromide	DDAB	Cationic
Diocetadecylamidoglicylspermidin	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
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-ditetradecylcarbamoylme-ethyl-acetamide	RPR209120	Cationic
1,2-dilinoleoyloxy-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 9**.

Table 9. Polymers Used for Gene Transfer

Polymer	Abbreviation
Poly(ethylene)glycol	PEG
Polyethylenimine	PEI
Dithiobis(succinimidylpropionate)	DSP
Dimethyl-3,3'-dithiobispropionimidate	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 an 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 an embodiment, the vehicle uses fusogenic and endosome-destabilizing peptides/polymers. In an embodiment, the vehicle undergoes acid-triggered conformational changes (*e.g.*, to accelerate endosomal escape of the cargo). In an 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 an embodiment, the delivery vehicle is a biological non-viral delivery vehicle. In an 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 an 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 an 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 an 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 an embodiment, one or more nucleic acid molecules (*e.g.*, DNA molecules) other than the components of a Cas system, *e.g.*, the Cas9 molecule component and/or the gRNA molecule component described herein, are delivered. In an embodiment, the nucleic acid molecule is delivered at the same time as one or more of the components of the Cas system are delivered. In an 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 components of the Cas system are delivered. In an embodiment, the nucleic acid molecule is delivered by a different means than one or more of the components of the Cas system, *e.g.*, the Cas9 molecule component and/or the gRNA molecule component, 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 Cas9 molecule component and/or 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 an embodiment, the nucleic acid molecule encodes a therapeutic protein, *e.g.*, a protein described herein. In an embodiment, the nucleic acid molecule encodes an RNA molecule, *e.g.*, an RNA molecule described herein.

Delivery of RNA encoding a Cas9 molecule

RNA encoding Cas9 molecules and/or gRNA molecules, can be delivered into cells, *e.g.*, target cells described herein, by art-known methods or as described herein. For

example, Cas9-encoding and/or 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. Cas9-encoding and/or gRNA-encoding RNA can be conjugated to molecules) promoting uptake by the target cells (*e.g.*, target cells described herein).

In an embodiment, delivery via electroporation comprises mixing the cells with the RNA encoding Cas9 molecules and/or gRNA molecules, with or without donor template nucleic acid molecules, in a cartridge, chamber or cuvette and applying one or more electrical impulses of defined duration and amplitude. In an embodiment, delivery via electroporation is performed using a system in which cells are mixed with the RNA encoding Cas9 molecules and/or gRNA molecules, with or without donor template nucleic acid 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. Cas9-encoding and/or gRNA-encoding RNA can be conjugated to molecules to promote uptake by the target cells (*e.g.*, target cells described herein).

Delivery of Cas9

Cas9 molecules can be delivered into cells by art-known methods or as described herein. For example, Cas9 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 protein can be conjugated to molecules promoting uptake by the target cells (*e.g.*, target cells described herein).

In an embodiment, delivery via electroporation comprises mixing the cells with the Cas9 molecules and/or gRNA molecules, with or without donor nucleic acid, in a cartridge, chamber or cuvette and applying one or more electrical impulses of defined duration and amplitude. In an embodiment, delivery via electroporation is performed using a system in which cells are mixed with the Cas9 molecules and/or gRNA molecules, with or without donor nucleic acid 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. Cas9-encoding and/or 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 include, by way of example, intravenous, intramarrow, intrarterial, intramuscular,

5 intradermal, subcutaneous, intranasal 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 the trabecular bone or intrafemoral injection into the marrow space, and infusion into the
10 portal vein. In an 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 potentially toxic side effects that may occur when therapeutically effective amounts of a
15 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 example, by continuous release from a sustained release drug delivery device.

20 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
25 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
30 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.

5 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
10 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
15 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

20 Separate delivery of the components of a Cas system, *e.g.*, the Cas9 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 an embodiment, the Cas9 molecule and the gRNA molecule are delivered by
25 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 molecule, gRNA molecule, template nucleic acid, or payload. For example, the modes of delivery can result in different tissue distribution, different half-life, or different temporal
30 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 molecule and a gRNA molecule, can be delivered by modes that differ in terms of resulting half-life or persistence of the delivered component in the body, or in a particular compartment, tissue or organ. In an embodiment, a gRNA molecule can be delivered by such modes. The Cas9 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 an 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.

In an 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.

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 molecule, is delivered in a transient manner, for example as mRNA or 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.

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 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 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 an 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 second component, *e.g.*, a Cas9 molecule is delivered by a second delivery mode that results in a second spatial, *e.g.*, tissue, distribution. In an 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 an 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 molecule is delivered in a virus delivery vector, a liposome, or polymeric nanoparticle, 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 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 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 *HBB* gene, *e.g.*, a gRNA molecule as described herein;

(b) a Cas9 molecule, *e.g.*, a Cas9 molecule as described herein; and optionally, (c) a second, third and/or fourth gRNA that targets *HBB* gene, *e.g.*, a gRNA molecule; and optionally, (d) a template nucleic acid, as described herein. In an embodiment, the method comprises contacting said cell with (a) and (b). In an embodiment, the method comprises contacting said cell with (a), (b), and (c). In an embodiment, the method comprises contacting said cell with (a), (b), (c) and (d). In an embodiment, the gRNA targets the *HBB* gene and no exogenous template nucleic acid is contacted with the cell. The targeting domain of the gRNA molecule of (a) and optionally (c) may be selected from a targeting domain sequence described herein, *e.g.*, a targeting domain sequence selected from any of **Tables 1-3**, or a targeting domain sequence that differs by no more than 1, 2, 3, 4, or 5 nucleotides from, a targeting domain sequence described herein, *e.g.*, a targeting domain sequence from any of **Tables 1-3**.

In an embodiment, the method comprises contacting a cell from a subject suffering from or likely to develop SCD. The cell may be from a subject having a mutation at an SCD target position in the *HBB* gene. In an 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 contacting step may be performed *in vivo*. In an embodiment, the method of altering a cell as described herein comprises acquiring knowledge of the sequence at an SCD target position in said cell, prior to the contacting step. Acquiring knowledge of the sequence at an SCD target position in the cell may be by sequencing the *HBB* gene, or a portion of the *HBB* gene. In an 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 at least one of (a), (b), and (c). In an 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 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 an 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 at least one of (a), (b), (c) and (d). In an 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 molecule of (b), a nucleic acid which encodes a gRNA molecule of (a) and a template nucleic acid of (d), 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 an 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 vector, a modified AAV3 vector, an AAV6 vector, a modified AAV6 vector, an AAV8 vector or an AAV9 vector.

In an embodiment, contacting comprises delivering to the cell a Cas9 molecule of (b), as a protein or an mRNA, and a nucleic acid which encodes (a) and optionally a second, third and/or fourth gRNA molecule of (c).

In an embodiment, contacting comprises delivering to the cell a Cas9 molecule of (b), as a protein or an mRNA, said gRNA molecule of (a), as an RNA, and optionally said second, third and/or fourth gRNA molecule of (c), as an RNA.

In an embodiment, contacting comprises delivering to the cell a gRNA of (a) as an RNA, optionally said second, third and/or fourth gRNA molecule of (c) as an RNA, and a nucleic acid that encodes the Cas9 molecule of (b).

Disclosed herein are also methods of treating or preventing a subject suffering from or likely to develop SCD, *e.g.*, altering the structure, *e.g.*, sequence, of a target nucleic acid of the subject, comprising contacting the subject (or a cell from the subject) with:

- (a) a gRNA that targets the *HBB* gene, *e.g.*, a gRNA disclosed herein;
- (b) a Cas9 molecule, *e.g.*, a Cas9 molecule disclosed herein; and

optionally, (c)(i) a second gRNA that targets the *HBB* gene, *e.g.*, a second gRNA disclosed herein, and further optionally, (c)(ii) a third gRNA molecule, and still further optionally, (c)(iii) a fourth gRNA molecule that targets the *HBB* gene, *e.g.*, a third and fourth gRNA disclosed herein. The method of treating a subject may further comprise contacting the subject (or a cell from the subject) with (d) a template nucleic acid (in an embodiment where an exogenous template is used), *e.g.*, a template nucleic acid disclosed herein. In an embodiment, a template nucleic acid is used when the method of treating a subject uses HDR to alter the sequence of the target nucleic acid of the subject. In an embodiment, the gRNA molecule targets the *HBB* gene and no exogenous template nucleic acid is contacted with the subject (or a cell from the subject). In an embodiment, contacting comprises contacting with (a) and (b). In an embodiment, contacting comprises contacting with (a), (b), and (c)(i). In an embodiment, contacting comprises contacting with (a), (b), (c)(i) and (c)(ii). In an embodiment, contacting comprises contacting with (a), (b), (c)(i), (c)(ii) and (c)(iii). In an

embodiment, contacting comprises contacting with (a), (b), (c)(i) and (d). In an embodiment, contacting comprises contacting with (a), (b), (c)(i), (c)(ii) and (d). In an embodiment, contacting comprises contacting with (a), (b), (c)(i), (c)(ii), (c)(iii) and (d).

The targeting domain of the gRNA molecule of (a) or (c) (*e.g.*, (c)(i), (c)(ii), or (c)(iii)) may be selected from a targeting domain sequence described herein, *e.g.*, a targeting domain sequence selected from any of **Tables 1-3**, or a targeting domain sequence that differs by no more than 1, 2, 3, 4, or 5 nucleotides from a targeting domain sequence described herein, *e.g.*, a targeting domain sequence from any of **Tables 1-3**.

In an embodiment, the method comprises acquiring knowledge of the sequence (*e.g.*, a mutation) of an SCD target position in said subject. In an embodiment, the method comprises acquiring knowledge of the sequence (*e.g.*, a mutation) of an SCD target position in said subject by sequencing the *HBB* gene or a portion of the *HBB* gene. In an embodiment, the method comprises correcting a mutation at an SCD target position in the *HBB* gene. In an embodiment, the method comprises correcting a mutation at an SCD target position in the *HBB* gene by HDR.

When the method comprises correcting the mutation at an SCD target position by HDR, a Cas9 of (b), at least one gRNA molecule, *e.g.*, a gRNA molecule of (a) and a template nucleic acid of (d) are included in the contacting step. In an embodiment, a cell of the subject is contacted *ex vivo* with (a), (b), (d) 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 an embodiment, a cell of the subject is contacted *in vivo* with (a), (b), (d) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In an embodiment, the cell of the subject is contacted *in vivo* by intravenous delivery of (a), (b), (d) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In an embodiment, the cell of the subject is contacted *in vivo* by intramuscular delivery of (a), (b), (d) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In an embodiment, the cell of the subject is contacted *in vivo* by subcutaneous delivery of (a), (b), (d) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In an embodiment, the cell of the subject is contacted *in vivo* by intra-bone marrow (IBM) delivery of (a), (b), (d) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

In an 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), (d) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

In an embodiment, contacting comprises delivering to said subject said Cas9 molecule of (b), as a protein or mRNA, and a nucleic acid which encodes (a), a nucleic acid of (d) and
5 optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

In an embodiment, contacting comprises delivering to the subject the Cas9 molecule of (b), as a protein or mRNA, the gRNA molecule of (a), as an RNA, a nucleic acid of (d) and optionally the second, third and/or fourth gRNA molecule of (c), as an RNA.

In an embodiment, contacting comprises delivering to the subject the gRNA molecule
10 of (a), as an RNA, optionally said second, third and/or fourth gRNA molecule of (c), as an RNA, a nucleic acid that encodes the Cas9 molecule of (b), and a nucleic acid of (d).

In an 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 an embodiment, said cell is returned to the subject's body.

In an embodiment, a populations of cells from a subject is contacted *ex vivo* with (a),
15 (b) and optionally (c) to correct the E6V mutation in the *HBB* gene and a second population of cells from the subject is contacted *ex vivo* with (a), (b) and optionally (c) to introduce a mutation in a second gene, *e.g.*, the *BCL11A* gene, or to knockout or knock down a second gene, *e.g.*, the *BCL11A* gene. A mixture of the two cell populations may be returned to the
20 subject's body to treat or prevent SCD.

In an 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 an embodiment, the cell of the subject is contacted *in vivo* by intravenous delivery of (a), (b) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii). In an
25 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 an 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 an
30 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 an 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 (c)(iii).

In an embodiment, contacting comprises delivering to said subject said Cas9 molecule of (b), as a protein or mRNA, and a nucleic acid which encodes (a) and optionally (c)(i), further optionally (c)(ii), and still further optionally (c)(iii).

5 In an embodiment, contacting comprises delivering to the subject the Cas9 molecule of (b), as a protein or mRNA, the gRNA molecule of (a), as an RNA, and optionally the second, third and/or fourth gRNA molecule of (c), as an RNA.

In an embodiment, contacting comprises delivering to the subject the gRNA molecule of (a), as an RNA, optionally said second, third and/or fourth gRNA molecule of (c), as an RNA, and a nucleic acid that encodes the Cas9 molecule of (b).

10 In one embodiment, disclosed herein are kits comprising compositions of the invention and instructions for use.

Ex vivo delivery

15 In some embodiments, components described in **Table 6** 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 7**.

VIII. Modified Nucleosides, Nucleotides, and Nucleic Acids

20 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.

25 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;

30 (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 an embodiment, every base of a gRNA is modified, *e.g.*, all bases have a modified phosphate group, *e.g.*, all are phosphorothioate groups. In an embodiment, all, or substantially all, of the phosphate groups of a unimolecular (or chimeric) or modular gRNA molecule are replaced with phosphorothioate groups.

In an 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 an embodiment, the modified nucleic acids comprise one, two, three or more modified nucleotides. In an 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 an 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 an 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 an 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.

Definitions of Chemical Groups

5 As used herein, “alkyl” is meant to refer to a saturated hydrocarbon group which is straight-chained or branched. Example alkyl groups include methyl (Me), ethyl (Et), propyl (*e.g.*, n-propyl and isopropyl), butyl (*e.g.*, n-butyl, isobutyl, t-butyl), pentyl (*e.g.*, n-pentyl, isopentyl, neopentyl), and the like. An alkyl group can contain from 1 to about 20, from 2 to about 20, from 1 to about 12, from 1 to about 8, from 1 to about 6, from 1 to about 4, or from 10 1 to about 3 carbon atoms.

As used herein, “aryl” refers to monocyclic or polycyclic (*e.g.*, having 2, 3 or 4 fused rings) aromatic hydrocarbons such as, for example, phenyl, naphthyl, anthracenyl, phenanthrenyl, indanyl, indenyl, and the like. In an embodiment, aryl groups have from 6 to about 20 carbon atoms.

15 As used herein, “alkenyl” refers to an aliphatic group containing at least one double bond.

As used herein, “alkynyl” refers to a straight or branched hydrocarbon chain containing 2-12 carbon atoms and characterized in having one or more triple bonds. Examples of alkynyl groups include, but are not limited to, ethynyl, propargyl, and 3- 20 hexynyl.

As used herein, “arylalkyl” or “aralkyl” refers to an alkyl moiety in which an alkyl hydrogen atom is replaced by an aryl group. Aralkyl includes groups in which more than one hydrogen atom has been replaced by an aryl group. Examples of “arylalkyl” or “aralkyl” include benzyl, 2-phenylethyl, 3-phenylpropyl, 9-fluorenyl, benzhydryl, and trityl groups.

25 As used herein, “cycloalkyl” refers to a cyclic, bicyclic, tricyclic, or polycyclic non-aromatic hydrocarbon groups having 3 to 12 carbons. Examples of cycloalkyl moieties include, but are not limited to, cyclopropyl, cyclopentyl, and cyclohexyl.

As used herein, “heterocyclyl” refers to a monovalent radical of a heterocyclic ring system. Representative heterocyclyls include, without limitation, tetrahydrofuranyl, 30 tetrahydrothienyl, pyrrolidinyl, pyrrolidonyl, piperidinyl, pyrrolinyl, piperazinyl, dioxanyl, dioxolanyl, diazepinyl, oxazepinyl, thiazepinyl, and morpholinyl.

As used herein, “heteroaryl” refers to a monovalent radical of a heteroaromatic ring system. Examples of heteroaryl moieties include, but are not limited to, imidazolyl, oxazolyl,

thiazolyl, triazolyl, pyrrolyl, furanyl, indolyl, thiophenyl pyrazolyl, pyridinyl, pyrazinyl, pyridazinyl, pyrimidinyl, indoliziny, purinyl, naphthyridinyl, quinolyl, and pteridinyl.

Phosphate Backbone Modifications

Phosphate Group

In an embodiment, the phosphate group of a modified nucleotide can be modified by replacing one or more of the oxygens with a different substituent. Further, the modified nucleotide, *e.g.*, modified nucleotide present in a modified nucleic acid, can include the wholesale replacement of an unmodified phosphate moiety with a modified phosphate as described herein. In an embodiment, the modification of the phosphate backbone can include alterations that result in either an uncharged linker or a charged linker with unsymmetrical charge distribution.

Examples of modified phosphate groups include, phosphorothioate, phosphoroselenates, borano phosphates, borano phosphate esters, hydrogen phosphonates, phosphoroamidates, alkyl or aryl phosphonates and phosphotriesters. In an embodiment, one of the non-bridging phosphate oxygen atoms in the phosphate backbone moiety can be replaced by any of the following groups: sulfur (S), selenium (Se), BR₃ (wherein R can be, *e.g.*, hydrogen, alkyl, or aryl), C (*e.g.*, an alkyl group, an aryl group, and the like), H, NR₂ (wherein R can be, *e.g.*, hydrogen, alkyl, or aryl), or OR (wherein R can be, *e.g.*, alkyl or aryl). The phosphorous atom in an unmodified phosphate group is achiral. However, replacement of one of the non-bridging oxygens with one of the above atoms or groups of atoms can render the phosphorous atom chiral; that is to say that a phosphorous atom in a phosphate group modified in this way is a stereogenic center. The stereogenic phosphorous atom can possess either the “R” configuration (herein Rp) or the “S” configuration (herein Sp).

Phosphorodithioates have both non-bridging oxygens replaced by sulfur. The phosphorus center in the phosphorodithioates is achiral which precludes the formation of oligoribonucleotide diastereomers. In an embodiment, modifications to one or both non-bridging oxygens can also include the replacement of the non-bridging oxygens with a group independently selected from S, Se, B, C, H, N, and OR (R can be, *e.g.*, alkyl or aryl).

The phosphate linker can also be modified by replacement of a bridging oxygen, (*i.e.*, the oxygen that links the phosphate to the nucleoside), with nitrogen (bridged phosphoroamidates), sulfur (bridged phosphorothioates) and carbon (bridged

methylenephosphonates). The replacement can occur at either linking oxygen or at both of the linking oxygens.

Replacement of the Phosphate Group

The phosphate group can be replaced by non-phosphorus containing connectors. In an embodiment, the charge phosphate group can be replaced by a neutral moiety.

Examples of moieties which can replace the phosphate group can include, without limitation, *e.g.*, methyl phosphonate, hydroxylamino, siloxane, carbonate, carboxymethyl, carbamate, amide, thioether, ethylene oxide linker, sulfonate, sulfonamide, thioformacetal, formacetal, oxime, methyleneimino, methylenemethylimino, methylenehydrazo, methylenedimethylhydrazo and methyleneoxymethylimino.

Replacement of the Ribophosphate Backbone

Scaffolds that can mimic nucleic acids can also be constructed wherein the phosphate linker and ribose sugar are replaced by nuclease resistant nucleoside or nucleotide surrogates. In an embodiment, the nucleobases can be tethered by a surrogate backbone. Examples can include, without limitation, the morpholino, cyclobutyl, pyrrolidine and peptide nucleic acid (PNA) nucleoside surrogates.

Sugar Modifications

The modified nucleosides and modified nucleotides can include one or more modifications to the sugar group. For example, the 2' hydroxyl group (OH) can be modified or replaced with a number of different "oxy" or "deoxy" substituents. In an embodiment, modifications to the 2' hydroxyl group can enhance the stability of the nucleic acid since the hydroxyl can no longer be deprotonated to form a 2'-alkoxide ion. The 2'-alkoxide can catalyze degradation by intramolecular nucleophilic attack on the linker phosphorus atom.

Examples of "oxy"-2' hydroxyl group modifications can include alkoxy or aryloxy (OR, wherein "R" can be, *e.g.*, alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or a sugar); polyethyleneglycols (PEG), $O(CH_2CH_2O)_nCH_2CH_2OR$ wherein R can be, *e.g.*, H or optionally substituted alkyl, and n can be an integer from 0 to 20 (*e.g.*, from 0 to 4, from 0 to 8, from 0 to 10, from 0 to 16, from 1 to 4, from 1 to 8, from 1 to 10, from 1 to 16, from 1 to 20, from 2 to 4, from 2 to 8, from 2 to 10, from 2 to 16, from 2 to 20, from 4 to 8, from 4 to 10, from 4 to 16, and from 4 to 20). In an embodiment, the "oxy"-2' hydroxyl group modification can include "locked" nucleic acids (LNA) in which the 2' hydroxyl can be connected, *e.g.*, by a C_{1-6} alkylene or C_{1-6} heteroalkylene bridge, to the 4' carbon of the same ribose sugar, where exemplary bridges can include methylene, propylene, ether, or amino

bridges; O-amino (wherein amino can be, *e.g.*, NH₂; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino) and aminoalkoxy, O(CH₂)_n-amino, (wherein amino can be, *e.g.*, NH₂; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino). In an embodiment, the “oxy”-2’ hydroxyl group modification can include the methoxyethyl group (MOE), (OCH₂CH₂OCH₃, *e.g.*, a PEG derivative).

“Deoxy” modifications can include hydrogen (i.e. deoxyribose sugars, *e.g.*, at the overhang portions of partially ds RNA); halo (*e.g.*, bromo, chloro, fluoro, or iodo); amino (wherein amino can be, *e.g.*, NH₂; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, diheteroarylamino, or amino acid); NH(CH₂CH₂NH)_nCH₂CH₂-amino (wherein amino can be, *e.g.*, as described herein), -NHC(O)R (wherein R can be, *e.g.*, alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or sugar), cyano; mercapto; alkyl-thio-alkyl; thioalkoxy; and alkyl, cycloalkyl, aryl, alkenyl and alkynyl, which may be optionally substituted with *e.g.*, an amino as described herein.

The sugar group can also contain one or more carbons that possess the opposite stereochemical configuration than that of the corresponding carbon in ribose. Thus, a modified nucleic acid can include nucleotides containing *e.g.*, arabinose, as the sugar. The nucleotide “monomer” can have an alpha linkage at the 1’ position on the sugar, *e.g.*, alpha-nucleosides. The modified nucleic acids can also include “abasic” sugars, which lack a nucleobase at C-1’. These abasic sugars can also be further modified at one or more of the constituent sugar atoms. The modified nucleic acids can also include one or more sugars that are in the L form, *e.g.*, L-nucleosides.

Generally, RNA includes the sugar group ribose, which is a 5-membered ring having an oxygen. Exemplary modified nucleosides and modified nucleotides can include, without limitation, replacement of the oxygen in ribose (*e.g.*, with sulfur (S), selenium (Se), or alkylene, such as, *e.g.*, methylene or ethylene); addition of a double bond (*e.g.*, to replace ribose with cyclopentenyl or cyclohexenyl); ring contraction of ribose (*e.g.*, to form a 4-membered ring of cyclobutane or oxetane); ring expansion of ribose (*e.g.*, to form a 6- or 7-membered ring having an additional carbon or heteroatom, such as for example, anhydrohexitol, altritol, mannitol, cyclohexanyl, cyclohexenyl, and morpholino that also has a phosphoramidate backbone). In an embodiment, the modified nucleotides can include multicyclic forms (*e.g.*, tricyclo; and “unlocked” forms, such as glycol nucleic acid (GNA) (*e.g.*, R-GNA or S-GNA, where ribose is replaced by glycol units attached to phosphodiester

bonds), threose nucleic acid (TNA, where ribose is replaced with α -L-threofuranosyl- (3'→2')).

Modifications on the Nucleobase

5 The modified nucleosides and modified nucleotides described herein, which can be incorporated into a modified nucleic acid, can include a modified nucleobase. Examples of nucleobases include, but are not limited to, adenine (A), guanine (G), cytosine (C), and uracil (U). These nucleobases can be modified or wholly replaced to provide modified nucleosides and modified nucleotides that can be incorporated into modified nucleic acids. The
10 nucleobase of the nucleotide can be independently selected from a purine, a pyrimidine, a purine or pyrimidine analog. In an embodiment, the nucleobase can include, for example, naturally-occurring and synthetic derivatives of a base.

Uracil

15 In an embodiment, the modified nucleobase is a modified uracil. Exemplary nucleobases and nucleosides having a modified uracil include without limitation pseudouridine (ψ), pyridin-4-one ribonucleoside, 5-aza-uridine, 6-aza-uridine, 2-thio-5-aza-uridine, 2-thio-uridine (s2U), 4-thio-uridine (s4U), 4-thio-pseudouridine, 2-thio-
20 pseudouridine, 5-hydroxy-uridine (ho⁵U), 5-aminoallyl-uridine, 5-halo-uridine (*e.g.*, 5-iodo-uridine or 5-bromo-uridine), 3-methyl-uridine (m³U), 5-methoxy-uridine (mo⁵U), uridine 5-oxyacetic acid (cmo⁵U), uridine 5-oxyacetic acid methyl ester (mcmo⁵U), 5-carboxymethyl-uridine (cm⁵U), 1-carboxymethyl-pseudouridine, 5-carboxyhydroxymethyl-uridine (chm⁵U),
25 5-carboxyhydroxymethyl-uridine methyl ester (mchm⁵U), 5-methoxycarbonylmethyl-uridine (mcm⁵U), 5-methoxycarbonylmethyl-2-thio-uridine (mcm⁵s2U), 5-aminomethyl-2-thio-uridine (nm⁵s2U), 5-methylaminomethyl-uridine (mnm⁵U), 5-methylaminomethyl-2-thio-
30 uridine (mnm⁵s2U), 5-methylaminomethyl-2-seleno-uridine (mnm⁵se²U), 5-carbamoylmethyl-uridine (ncm⁵U), 5-carboxymethylaminomethyl-uridine (cmnm⁵U), 5-carboxymethylaminomethyl-2-thio-uridine (cmnm⁵s2U), 5-propynyl-uridine, 1-propynyl-pseudouridine, 5-taurinomethyl-uridine (τ cm⁵U), 1-taurinomethyl-pseudouridine, 5-taurinomethyl-2-thio-uridine (τ m⁵s2U), 1-taurinomethyl-4-thio-pseudouridine, 5-methyl-
uridine (m⁵U, *i.e.*, having the nucleobase deoxythymine), 1-methyl-pseudouridine (m¹ ψ), 5-methyl-2-thio-uridine (m⁵s2U), 1-methyl-4-thio-pseudouridine (m¹s⁴ ψ), 4-thio-1-methyl-pseudouridine, 3-methyl-pseudouridine (m³ ψ), 2-thio-1-methyl-pseudouridine, 1-methyl-1-deaza-pseudouridine, 2-thio-1-methyl-1-deaza-pseudouridine, dihydrouridine (D),

dihydropseudouridine, 5,6-dihydrouridine, 5-methyl-dihydrouridine (m^5D), 2-thio-dihydrouridine, 2-thio-dihydropseudouridine, 2-methoxy-uridine, 2-methoxy-4-thio-uridine, 4-methoxy-pseudouridine, 4-methoxy-2-thio-pseudouridine, N1-methyl-pseudouridine, 3-(3-amino-3-carboxypropyl)uridine (acp^3U), 1-methyl-3-(3-amino-3-carboxypropyl)pseudouridine ($acp^3\psi$), 5-(isopentenylaminomethyl)uridine (inm^5U), 5-(isopentenylaminomethyl)-2-thio-uridine (inm^5s2U), α -thio-uridine, 2'-O-methyl-uridine (Um), 5,2'-O-dimethyl-uridine (m^5Um), 2'-O-methyl-pseudouridine (ψm), 2-thio-2'-O-methyl-uridine ($s2Um$), 5-methoxycarbonylmethyl-2'-O-methyl-uridine (mcm^5Um), 5-carbamoylmethyl-2'-O-methyl-uridine (ncm^5Um), 5-carboxymethylaminomethyl-2'-O-methyl-uridine ($cmnm^5Um$), 3,2'-O-dimethyl-uridine (m^3Um), 5-(isopentenylaminomethyl)-2'-O-methyl-uridine (inm^5Um), 1-thio-uridine, deoxythymidine, 2'-F-ara-uridine, 2'-F-uridine, 2'-OH-ara-uridine, 5-(2-carbomethoxyvinyl) uridine, 5-[3-(1-E-propenylamino)uridine, pyrazolo[3,4-d]pyrimidines, xanthine, and hypoxanthine.

Cytosine

In an embodiment, the modified nucleobase is a modified cytosine. Exemplary nucleobases and nucleosides having a modified cytosine include without limitation 5-azacytidine, 6-aza-cytidine, pseudoisocytidine, 3-methyl-cytidine (m^3C), N4-acetyl-cytidine (act), 5-formyl-cytidine (f^5C), N4-methyl-cytidine (m^4C), 5-methyl-cytidine (m^5C), 5-halo-cytidine (*e.g.*, 5-iodo-cytidine), 5-hydroxymethyl-cytidine (hm^5C), 1-methyl-pseudoisocytidine, pyrrolo-cytidine, pyrrolo-pseudoisocytidine, 2-thio-cytidine ($s2C$), 2-thio-5-methyl-cytidine, 4-thio-pseudoisocytidine, 4-thio-1-methyl-pseudoisocytidine, 4-thio-1-methyl-1-deaza-pseudoisocytidine, 1-methyl-1-deaza-pseudoisocytidine, zebularine, 5-azazebularine, 5-methyl-zebularine, 5-aza-2-thio-zebularine, 2-thio-zebularine, 2-methoxycytidine, 2-methoxy-5-methyl-cytidine, 4-methoxy-pseudoisocytidine, 4-methoxy-1-methyl-pseudoisocytidine, lysidine (k^2C), α -thio-cytidine, 2'-O-methyl-cytidine (Cm), 5,2'-O-dimethyl-cytidine (m^5Cm), N4-acetyl-2'-O-methyl-cytidine (ac^4Cm), N4,2'-O-dimethyl-cytidine (m^4Cm), 5-formyl-2'-O-methyl-cytidine (f^5Cm), N4,N4,2'-O-trimethyl-cytidine (m^4_2Cm), 1-thio-cytidine, 2'-F-ara-cytidine, 2'-F-cytidine, and 2'-OH-ara-cytidine.

Adenine

In an embodiment, the modified nucleobase is a modified adenine. Exemplary nucleobases and nucleosides having a modified adenine include without limitation 2-amino-purine, 2,6-diaminopurine, 2-amino-6-halo-purine (*e.g.*, 2-amino-6-chloro-purine), 6-halo-

purine (*e.g.*, 6-chloro-purine), 2-amino-6-methyl-purine, 8-azido-adenosine, 7-deaza-adenosine, 7-deaza-8-aza-adenosine, 7-deaza-2-amino-purine, 7-deaza-8-aza-2-amino-purine, 7-deaza-2,6-diaminopurine, 7-deaza-8-aza-2,6-diaminopurine, 1-methyl-adenosine (m^1A), 2-methyl-adenosine (m^2A), N6-methyl-adenosine (m^6A), 2-methylthio-N6-methyl-adenosine (ms^2m^6A), N6-isopentenyl-adenosine (i^6A), 2-methylthio-N6-isopentenyl-adenosine (ms^2i^6A), N6-(cis-hydroxyisopentenyl)adenosine (io^6A), 2-methylthio-N6-(cis-hydroxyisopentenyl)adenosine (ms^2io^6A), N6-glycinylicarbamoyl-adenosine (g^6A), N6-threonylicarbamoyl-adenosine (t^6A), N6-methyl-N6-threonylicarbamoyl-adenosine (m^6t^6A), 2-methylthio-N6-threonylicarbamoyl-adenosine (ms^2g^6A), N6,N6-dimethyl-adenosine (m^6_2A), N6-hydroxynorvalylcarbamoyl-adenosine (hn^6A), 2-methylthio-N6-hydroxynorvalylcarbamoyl-adenosine (ms^2hn^6A), N6-acetyl-adenosine (ac^6A), 7-methyl-adenosine, 2-methylthio-adenosine, 2-methoxy-adenosine, α -thio-adenosine, 2'-O-methyl-adenosine (Am), N^6 ,2'-O-dimethyl-adenosine (m^6Am), N^6 -Methyl-2'-deoxyadenosine, N6,N6,2'-O-trimethyl-adenosine (m^6_2Am), 1,2'-O-dimethyl-adenosine (m^1Am), 2'-O-ribosyladenosine (phosphate) ($Ar(p)$), 2-amino-N6-methyl-purine, 1-thio-adenosine, 8-azido-adenosine, 2'-F-ara-adenosine, 2'-F-adenosine, 2'-OH-ara-adenosine, and N6-(19-amino-pentaoxanonadecyl)-adenosine.

Guanine

In an embodiment, the modified nucleobase is a modified guanine. Exemplary nucleobases and nucleosides having a modified guanine include without limitation inosine (I), 1-methyl-inosine (m^1I), wyosine (imG), methylwyosine (mimG), 4-demethyl-wyosine (imG-14), isowyosine (imG2), wybutosine (yW), peroxywybutosine (o_2yW), hydroxywybutosine (OHyW), undermodified hydroxywybutosine (OHyW*), 7-deaza-guanosine, queuosine (Q), epoxyqueuosine (oQ), galactosyl-queuosine (galQ), mannosyl-queuosine (manQ), 7-cyano-7-deaza-guanosine ($preQ_0$), 7-aminomethyl-7-deaza-guanosine ($preQ_1$), archaeosine (G^+), 7-deaza-8-aza-guanosine, 6-thio-guanosine, 6-thio-7-deaza-guanosine, 6-thio-7-deaza-8-aza-guanosine, 7-methyl-guanosine (m^7G), 6-thio-7-methyl-guanosine, 7-methyl-inosine, 6-methoxy-guanosine, 1-methyl-guanosine (m^1G), N2-methyl-guanosine (m^2G), N2,N2-dimethyl-guanosine (m^2_2G), N2,7-dimethyl-guanosine (m^2_7G), N2, N2,7-dimethyl-guanosine ($m^2_2,7G$), 8-oxo-guanosine, 7-methyl-8-oxo-guanosine, 1-methyl-6-thio-guanosine, N2-methyl-6-thio-guanosine, N2,N2-dimethyl-6-thio-guanosine, α -thio-guanosine, 2'-O-methyl-guanosine (Gm), N2-methyl-2'-O-methyl-guanosine (m^2Gm), N2,N2-dimethyl-2'-O-methyl-guanosine (m^2_2Gm), 1-methyl-2'-O-methyl-guanosine (m^1Gm),

N²,7-dimethyl-2'-O-methyl-guanosine (m²,7Gm), 2'-O-methyl-inosine (Im), 1,2'-O-dimethyl-inosine (m'Im), O⁶-phenyl-2'-deoxyinosine, 2'-O-ribosylguanosine (phosphate) (Gr(p)), 1-thio-guanosine, O⁶-methyl-guanosine, O⁶-Methyl-2'-deoxyguanosine, 2'-F-ara-guanosine, and 2'-F-guanosine.

5

Exemplary Modified gRNAs

In some embodiments, the modified nucleic acids can be modified gRNAs. It is to be understood that any of the gRNAs described herein can be modified in accordance with this section, including any gRNAs that comprises a targeting domain from SEQ ID NOs: 387-
10 485, 6803-6871, and 16010-16256.

As discussed above, transiently expressed or delivered nucleic acids can be prone to degradation by, *e.g.*, cellular nucleases. Accordingly, in one aspect the modified gRNAs described herein can contain one or more modified nucleosides or nucleotides which introduce stability toward nucleases. While not wishing to be bound by theory it is also
15 believed that certain modified gRNAs described herein can exhibit a reduced innate immune response when introduced into a population of cells, particularly the cells of the present disclosure. As noted above, 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
20 the interferons, and cell death.

While some of the exemplary modification discussed in this section may be included at any position within the gRNA sequence, in some embodiments, a gRNA molecule comprises a modification at or near its 5' end (*e.g.*, within 1-10, 1-5, or 1-2 nucleotides of its 5' end). In some embodiments, a gRNA comprises a modification at or near its 3' end (*e.g.*,
25 within 1-10, 1-5, or 1-2 nucleotides of its 3' end). In some embodiments, a gRNA molecule comprises both a modification at or near its 5' end and a modification at or near its 3' end.

In an embodiment, the 5' end of a gRNA is modified by the inclusion of a eukaryotic mRNA cap structure or cap analog (*e.g.*, a G(5')ppp(5')G cap analog, a m⁷G(5')ppp(5')G cap analog, or a 3'-O-Me-m⁷G(5')ppp(5')G anti reverse cap analog (ARCA)). The cap or
30 cap analog can be included during either chemical synthesis or *in vitro* transcription of the gRNA.

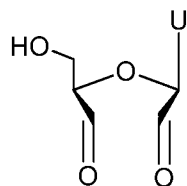
In an embodiment, an *in vitro* transcribed gRNA is modified by treatment with a phosphatase (*e.g.*, calf intestinal alkaline phosphatase) to remove the 5' triphosphate group.

In an embodiment, the 3' end of a gRNA is modified by the addition of one or more (e.g., 25-200) adenine (A) residues. The polyA tract can be contained in the nucleic acid (e.g., plasmid, PCR product, viral genome) encoding the gRNA, or can be added to the gRNA during chemical synthesis, or following *in vitro* transcription using a polyadenosine polymerase (e.g., *E. coli* Poly(A)Polymerase).

In another aspect, methods and compositions discussed herein provide methods and compositions for gene editing by using a gRNA molecule which comprises a polyA tail. In one embodiment, a polyA tail of undefined length ranging from 1 to 1000 nucleotide is added enzymatically using a polymerase such as *E. coli* polyA polymerase (E-PAP). In one embodiment, the polyA tail of a specified length (e.g., 1, 5, 10, 20, 30, 40, 50, 60, 100, or 150 nucleotides) is encoded on a DNA template and transcribed with the gRNA via an RNA polymerase (e.g., T7 RNA polymerase). In one embodiment, a polyA tail of defined length (e.g., 1, 5, 10, 20, 30, 40, 50, 60, 100, or 150 nucleotides) is synthesized as a synthetic oligonucleotide and ligated on the 3' end of the gRNA with either an RNA ligase or a DNA ligase with or without a splinted DNA oligonucleotide complementary to the guide RNA and the polyA oligonucleotide. In one embodiment, the entire gRNA including a defined length of polyA tail is made synthetically, in one or several pieces, and ligated together by either an RNA ligase or a DNA ligase with or without a splinted DNA oligonucleotide.

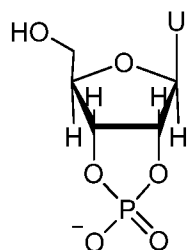
In an embodiment, *in vitro* transcribed gRNA molecule contains both a 5' cap structure or cap analog and a 3' polyA tract. In an embodiment, an *in vitro* transcribed gRNA is modified by treatment with a phosphatase (e.g., calf intestinal alkaline phosphatase) to remove the 5' triphosphate group and comprises a 3' polyA tract.

In some embodiments, gRNAs can be modified at a 3' terminal U ribose. For example, the two terminal hydroxyl groups of the U ribose can be oxidized to aldehyde groups and a concomitant opening of the ribose ring to afford a modified nucleoside as shown below:



wherein "U" can be an unmodified or modified uridine.

In another embodiment, the 3' terminal U can be modified with a 2'3' cyclic phosphate as shown below:



wherein “U” can be an unmodified or modified uridine.

In some embodiments, the gRNA molecules may contain 3' nucleotides which can be stabilized against degradation, *e.g.*, by incorporating one or more of the modified nucleotides described herein. In this embodiment, *e.g.*, uridines can be replaced with modified uridines, *e.g.*, 5-(2-amino)propyl uridine, and 5-bromo uridine, or with any of the modified uridines described herein; adenosines and guanosines can be replaced with modified adenosines and guanosines, *e.g.*, with modifications at the 8-position, *e.g.*, 8-bromo guanosine, or with any of the modified adenosines or guanosines described herein.

In some embodiments, sugar-modified ribonucleotides can be incorporated into the gRNA molecule, *e.g.*, wherein the 2' OH-group is replaced by a group selected from H, -OR, -R (wherein R can be, *e.g.*, alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or sugar), halo, -SH, -SR (wherein R can be, *e.g.*, alkyl, cycloalkyl, aryl, aralkyl, heteroaryl or sugar), amino (wherein amino can be, *e.g.*, NH₂; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, diheteroarylamino, or amino acid); or cyano (-CN). In some embodiments, the phosphate backbone can be modified as described herein, *e.g.*, with a phosphothioate group. In some embodiments, one or more of the nucleotides of the gRNA can each independently be a modified or unmodified nucleotide including, but not limited to 2'-sugar modified, such as, 2'-O-methyl, 2'-O-methoxyethyl, or 2'-Fluoro modified including, *e.g.*, 2'-F or 2'-O-methyl, adenosine (A), 2'-F or 2'-O-methyl, cytidine (C), 2'-F or 2'-O-methyl, uridine (U), 2'-F or 2'-O-methyl, thymidine (T), 2'-F or 2'-O-methyl, guanosine (G), 2'-O-methoxyethyl-5-methyluridine (Teo), 2'-O-methoxyethyladenosine (Aeo), 2'-O-methoxyethyl-5-methylcytidine (m5Ceo), and any combinations thereof.

In some embodiments, a gRNA can include “locked” nucleic acids (LNA) in which the 2' OH-group can be connected, *e.g.*, by a C1-6 alkylene or C1-6 heteroalkylene bridge, to the 4' carbon of the same ribose sugar, where exemplary bridges can include methylene, propylene, ether, or amino bridges; O-amino (wherein amino can be, *e.g.*, NH₂; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino) and aminoalkoxy or O(CH₂)_n-amino (wherein amino can be,

e.g., NH₂; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino).

In some embodiments, a gRNA can include a modified nucleotide which is multicyclic (*e.g.*, tricyclo; and “unlocked” forms, such as glycol nucleic acid (GNA) (*e.g.*, R-GNA or S-GNA, where ribose is replaced by glycol units attached to phosphodiester bonds), or threose nucleic acid (TNA, where ribose is replaced with α -L-threofuranosyl-(3'→2')).

Generally, gRNA molecules include the sugar group ribose, which is a 5-membered ring having an oxygen. Exemplary modified gRNAs can include, without limitation, replacement of the oxygen in ribose (*e.g.*, with sulfur (S), selenium (Se), or alkylene, such as, *e.g.*, methylene or ethylene); addition of a double bond (*e.g.*, to replace ribose with cyclopentenyl or cyclohexenyl); ring contraction of ribose (*e.g.*, to form a 4-membered ring of cyclobutane or oxetane); ring expansion of ribose (*e.g.*, to form a 6- or 7-membered ring having an additional carbon or heteroatom, such as for example, anhydrohexitol, altritol, mannitol, cyclohexanyl, cyclohexenyl, and morpholino that also has a phosphoramidate backbone). Although the majority of sugar analog alterations are localized to the 2' position, other sites are amenable to modification, including the 4' position. In an embodiment, a gRNA comprises a 4'-S, 4'-Se or a 4'-C-aminomethyl-2'-O-Me modification.

In some embodiments, deaza nucleotides, *e.g.*, 7-deaza-adenosine, can be incorporated into the gRNA molecule. In some embodiments, O- and N-alkylated nucleotides, *e.g.*, N6-methyl adenosine, can be incorporated into the gRNA molecule. In some embodiments, one or more or all of the nucleotides in a gRNA molecule are deoxynucleotides.

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: Cloning and Initial Screening of gRNA Molecules

The suitability of candidate gRNA molecules can be evaluated as described in this example. Although described for a chimeric gRNA molecule, the approach can also be used to evaluate modular gRNA molecules.

10

Cloning gRNAs into Vectors

For each gRNA molecule, a pair of overlapping oligonucleotides is designed and obtained. Oligonucleotides are annealed and ligated into a digested vector backbone containing an upstream U6 promoter and the remaining sequence of a long chimeric gRNA molecule. Plasmid is sequence-verified and prepped to generate sufficient amounts of transfection-quality DNA. Alternate promoters may be used to drive *in vivo* transcription (e.g., H1 promoter) or for *in vitro* transcription (e.g., a T7 promoter).

15

Cloning gRNAs in linear dsDNA molecule (STITCHR)

For each gRNA molecule, a single oligonucleotide is designed and obtained. The U6 promoter and the gRNA scaffold (e.g., including everything except the targeting domain, e.g., including sequences derived from the crRNA and tracrRNA, e.g., including a first complementarity domain; a linking domain; a second complementarity domain; a proximal domain; and a tail domain) are separately PCR amplified and purified as dsDNA molecules. The gRNA-specific oligonucleotide is used in a PCR reaction to stitch together the U6 and the gRNA scaffold, linked by the targeting domain specified in the oligonucleotide.

20

Resulting dsDNA molecule (STITCHR product) is purified for transfection. Alternate promoters may be used to drive *in vivo* transcription (e.g., H1 promoter) or for *in vitro* transcription (e.g., T7 promoter). Any gRNA scaffold may be used to create gRNAs compatible with Cas9s from any bacterial species.

25

Initial gRNA Screen

Each gRNA to be tested is transfected, along with a plasmid expressing Cas9 and a small amount of a GFP-expressing plasmid into human cells. In preliminary experiments, these cells can be immortalized human cell lines such as 293T, K562 or U2OS.

30

Alternatively, primary human cells may be used. In this case, cells may be relevant to the eventual therapeutic cell target (for example, an erythroid cell). The use of primary cells

similar to the potential therapeutic target cell population may provide important information on gene targeting rates in the context of endogenous chromatin and gene expression.

Transfection may be performed using lipid transfection (such as Lipofectamine or Fugene) or by electroporation (such as Lonza Nucleofection). Following transfection, GFP
5 expression can be determined either by fluorescence microscopy or by flow cytometry to confirm consistent and high levels of transfection. These preliminary transfections can comprise different gRNAs and different targeting approaches (17-mers, 20-mers, nuclease, dual-nickase, etc.) to determine which gRNAs/combinations of gRNAs give the greatest activity.

10 Efficiency of cleavage with each gRNA may be assessed by measuring NHEJ-induced indel formation at the target locus by a T7E1-type assay or by sequencing. Alternatively, other mismatch-sensitive enzymes, such as CeiI/Surveyor nuclease, may also be used.

For the T7E1 assay, PCR amplicons are approximately 500-700bp with the intended
15 cut site placed asymmetrically in the amplicon. Following amplification, purification and size-verification of PCR products, DNA is denatured and re-hybridized by heating to 95°C and then slowly cooling. Hybridized PCR products are then digested with T7 Endonuclease I (or other mismatch-sensitive enzyme) which recognizes and cleaves non-perfectly matched DNA. If indels are present in the original template DNA, when the amplicons are denatured
20 and re-annealed, this results in the hybridization of DNA strands harboring different indels and therefore lead to double-stranded DNA that is not perfectly matched. Digestion products may be visualized by gel electrophoresis or by capillary electrophoresis. The fraction of DNA that is cleaved (density of cleavage products divided by the density of cleaved and uncleaved) may be used to estimate a percent NHEJ using the following equation: %NHEJ =
25 $(1 - (1 - \text{fraction cleaved})^{1/2})$. The T7E1 assay is sensitive down to about 2-5% NHEJ.

Sequencing may be used instead of, or in addition to, the T7E1 assay. For Sanger sequencing, purified PCR amplicons are cloned into a plasmid backbone, transformed, minipreped and sequenced with a single primer. Sanger sequencing may be used for determining the exact nature of indels after determining the NHEJ rate by T7E1.

30 Sequencing may also be performed using next generation sequencing techniques. When using next generation sequencing, amplicons may be 300-500bp with the intended cut site placed asymmetrically. Following PCR, next generation sequencing adapters and barcodes (for example Illumina multiplex adapters and indexes) may be added to the ends of

the amplicon, *e.g.*, for use in high throughput sequencing (for example on an Illumina MiSeq). This method allows for detection of very low NHEJ rates.

Example 2: Assessment of Gene Targeting by NHEJ

5 The gRNAs that induce the greatest levels of NHEJ in initial tests can be selected for further evaluation of gene targeting efficiency. In this case, cells are derived from disease subjects and, therefore, harbor the relevant mutation.

Following transfection (usually 2-3 days post-transfection,) genomic DNA may be isolated from a bulk population of transfected cells and PCR may be used to amplify the
10 target region. Following PCR, gene targeting efficiency to generate the desired mutations (either knockout of a target gene or removal of a target sequence motif) may be determined by sequencing. For Sanger sequencing, PCR amplicons may be 500-700 bp long. For next generation sequencing, PCR amplicons may be 300-500 bp long. If the goal is to knockout gene function, sequencing may be used to assess what percent of alleles have undergone
15 NHEJ-induced indels that result in a frameshift or large deletion or insertion that would be expected to destroy gene function. If the goal is to remove a specific sequence motif, sequencing may be used to assess what percent of alleles have undergone NHEJ-induced deletions that span this sequence.

Example 3: Assessment of Gene Targeting by HDR

20 The gRNAs that induce the greatest levels of NHEJ in initial tests can be selected for further evaluation of gene targeting efficiency. In this case, cells are derived from disease subjects and, therefore, harbor the relevant mutation.

Following transfection (usually 2-3 days post-transfection,) genomic DNA may be
25 isolated from a bulk population of transfected cells and PCR may be used to amplify the target region. Following PCR, gene targeting efficiency can be determined by several methods.

Determination of gene targeting frequency involves measuring the percentage of alleles that have undergone homologous directed repair (HDR) with the exogenously
30 provided donor template or endogenous genomic donor sequence and which therefore have incorporated the desired correction. If the desired HDR event creates or destroys a restriction enzyme site, the frequency of gene targeting may be determined by a RFLP assay. If no restriction site is created or destroyed, sequencing may be used to determine gene targeting frequency. If a RFLP assay is used, sequencing may still be used to verify the desired HDR

event and ensure that no other mutations are present. If an exogenously provided donor template is employed, at least one of the primers is placed in the endogenous gene sequence outside of the region included in the homology arms, which prevents amplification of donor template still present in the cells. Therefore, the length of the homology arms present in the donor template may affect the length of the PCR amplicon. PCR amplicons can either span the entire donor region (both primers placed outside the homology arms) or they can span only part of the donor region and a single junction between donor and endogenous DNA (one internal and one external primer). If the amplicons span less than the entire donor region, two different PCRs should be used to amplify and sequence both the 5' and the 3' junction.

If the PCR amplicon is short (less than 600bp) it is possible to use next generation sequencing. Following PCR, next generation sequencing adapters and barcodes (for example Illumina multiplex adapters and indexes) may be added to the ends of the amplicon, *e.g.*, for use in high throughput sequencing (for example on an Illumina MiSeq). This method allows for detection of very low gene targeting rates.

If the PCR amplicon is too long for next generation sequencing, Sanger sequencing can be performed. For Sanger sequencing, purified PCR amplicons are cloned into a plasmid backbone (for example, TOPO cloned using the LifeTech Zero Blunt® TOPO® cloning kit), transformed, miniprep and sequenced.

The same or similar assays described above can be used to measure the percentage of alleles that have undergone HDR with endogenous genomic donor sequence and which therefore have incorporated the desired correction.

Example 4: Gene Targeting of the *HBB* Locus by CRISPR/Cas9 to Investigate Repair Pathway Choice in Response to Different Types of DNA Lesions

The CRISPR/Cas9 system was used to target the human *HBB* gene in the region of the sickle cell anemia-causing mutation.

To examine how the nature of the targeted break affects the frequency of different DNA repair outcomes, blunt double-strand breaks, single-strand nicks, and dual-nicks in which the nicks are placed on opposite strands and leave either 3' or 5' overhangs of varying lengths, were introduced by utilizing the wild type Cas9 nuclease, as well as two different Cas9 nickases. Several different DNA repair outcomes including indel mutations resulting from non-homologous end-joining, homology-dependent repair (HDR) using an exogenous donor as a template (*i.e.*, gene correction), and HDR using the closely related *HBD* gene as an endogenous template (*i.e.*, gene conversion), were characterized using either single-strand

oligonucleotide (ssODN) or plasmid DNA donors. The frequency of these various repair outcomes under different conditions offer insight into the mechanisms of DNA repair and how it is impacted by the nature of the DNA break. The data also indicates a therapeutic approach in which correction of the sickle-cell mutation is efficiently mediated through HDR

5 with either a donor template or with the *HBD* gene.

In this study different gRNA molecules for the *HBB* region that surrounds the nucleotides encoding the amino acid most commonly mutated in sickle cell disease had been tested in 293T cells with wild type Cas9 molecule. The gRNAs that induced similar high rates of NHEJ and had PAMs facing in opposite orientations (*e.g.*, each PAM is oriented 5' relative to a mutation on the first and second strands, respectively) were selected to test as pairs with Cas9 D10A and Cas9 N863A nickases.

As shown in **Fig. 1**, the gRNA pair 8/15 ("8gRNA"/ "15gRNA" pair) was selected as one of the best pairs of gRNA molecules. "8gRNA" has the targeting domain sequence of GUAACGGCAGACUUCUCCUC (SEQ ID NO: 388) and "15gRNA" has the targeting domain sequence of AAGGUGAACGUGGAUGAAGU (SEQ ID NO: 387). This pair of gRNAs in combination with the mutant Cas9 D10A would generate a 5' overhang of 47 bp, and in combination with the mutant N863A would generate a 3' overhang of 47 bp. Different overhang structures are believed to have different properties for engaging different repair mechanisms. To address repair efficacy and repair outcomes in a systematic manner, U2OS cells with respective Cas9 variants and gRNAs were used.

U2OS cells were electroporated with 200 ng of each gRNA and 750 ng of plasmid that encodes wild type Cas9 or mutant Cas9. Cells were collected 6 days after electroporation and genomic DNA was extracted. PCR amplification of the *HBB* locus was performed and subcloned into a Topo Blunt Vector. For each condition in each experiment 96 colonies were sequenced with Sanger sequencing. In the experiments assessing HDR efficacy, cells were electroporated with 2.5 ug of single stranded oligo or double stranded oligo in addition to the gRNA and the Cas9-encoding plasmid.

As shown in **Fig. 2**, the similar rates of overall modification at the *HBB* locus were generated using wild type Cas9 or Cas9 nickases (D10A, N863A), as detected by Sanger sequencing, suggesting a similar rate in error prone repair independent of the Cas9-induced DNA lesion.

Figs. 3A-3B show that a majority of the total gene editing events (about 3/4 of the total) were small deletions (<10 bp). This is consistent with the notion that wild type Cas9 generates a blunt end which are preferentially repaired by canonical NHEJ. In contrast,

deletions represented only about a quarter of the total events using either nickase (D10A or N863A). Moreover, larger deletions of ~50 bp that can be mapped to the region between the two nickase sites were observed (**Fig. 3A** and **3C**). The remaining gene-editing events were substantially different between the two nickases.

As shown in **Fig. 4A**, in the case of Cas9 D10A nickase, which leaves a 5' protruding end, the lesion is mostly repaired through a mechanism defined as gene conversion. In gene conversion, the *HBD* locus will serve as a template to repair the *HBB* gene. *HBD* is a highly similar gene (92% identity with *HBB*) that does not carry the sickle-cell mutation (**Fig. 4B**). **Fig. 5** shows that the majority of the *HBD* sequence that were incorporated in the *HBB* locus were in the region between the nickase cuts. In contrast, a low frequency of gene conversion was observed when the Cas9 N863 nickase was used (**Fig. 4A**). In the case of Cas9 N863A nickase, a majority of the gene editing events were insertions in which the inserted part was a duplication of the overhangs (**Figs. 6A-6B**).

Figure 7A shows compiled data of the gene editing events resulting from DNA lesions generated by the three different Cas9 variants. While similar rates of overall modification at the *HBB* locus were observed, the distribution of the resulting type of repair outcome dramatically differed. 12.4% of WT Cas9-induced modifications harbored the footprint of the highly homologous *HBD* gene (a phenomenon known as “gene conversion”). In gene conversion, the *HBD* locus will serve as a template to repair the *HBB* gene. *HBD* is a highly similar gene (92% identity with *HBB*) that does not carry the sickle-cell mutation (**Fig. 6B**). The *HBD* gene lies approximately 7.6 kb upstream of the *HBB* gene on chromosome 11. The rate of gene conversion from the *HBD* gene was significantly enhanced when lesions were induced using the Cas9 D10A nickase. 32.8% of D10A Cas9-induced modifications were resolved by gene conversion. In contrast, 3.5% of N863A Cas9-induced modifications were resolved by gene conversion. Overall, this data suggested that DNA cleavage induced by the Cas9 D10A nickase are particularly amenable for gene conversion from the endogenous *HBD* gene. These data also suggest that the different DNA ends generated with different Cas9 variants activate different DNA repair pathway responses.

Figure 7B shows the distribution of deletion size resulting from DNA lesions generated by the three different Cas9 variants. Specifically, WT Cas9-induced lesions were predominantly resolved as deletions. Deletions were also frequently observed when using the gRNA pair 8/15 with either the Cas9 D10A nickase or Cas9 N863A nickase. WT Cas9 induced deletions were predominantly small in size (median deletion size = 3 nucleotides (nts)). Cas9 D10A nickase-induced deletions and Cas9 N863A nickase-induced deletions

were significantly larger in size (median deletion size = 28 nts and 36 nts, for Cas9 N863A and Cas9 D10A, respectively).

To determine whether the DNA repair events observed were representative of events occurring in single cells, and not the result of PCR bias during amplification of the DNA from a population of cells, repair outcomes were analyzed from a single U2Os cell clones (~80 single clones per condition) after nucleofection with each of the three Cas9 variants. As shown in **Figure 8A**, the repair event distribution was very similar in individual clones to the distribution observed in parental population of mixed cells. Similar repair event distribution was observed using the myeloid leukemia cell line K562 (**Fig. 8B**). These data suggest that the DNA ends generated with the different Cas9 variants activate different DNA repair pathway responses.

To better understand the molecular pathways responsible for the gene conversion pathway, Cas9 D10A nickase-induced gene conversion tracts were analyzed in detail. The region located between the 8gRNA/ 15gRNA pair was converted to the *HBD* sequence in approximately 80% of all gene conversion cases, as detected by analyzing single nucleotide polymorphisms (SNPs) that perfectly matched the *HBD* sequence (**Fig. 9**). Directionality in the gene conversion tracts was observed, suggesting that perhaps the homology arm to the right of 8gRNA more frequently engages in homology pairing with the *HBD* locus due to the presence of longer stretches of uninterrupted homology, resulting in effective gene conversion (**Fig. 9**). In contrast, the arm to the left of 15gRNA harbors more homology-disrupting SNPs, which we speculated could result in a higher frequency of heteroduplex rejection at the *HBD* site.

Gene conversion is a highly precise mechanism that repairs DSBs during the S/G2 phases of the cell cycle through the homologous recombination (HR) pathway. The genetic requirements of HR have been characterized. Briefly, initially 3-5' resection leads to the exposure of a single stranded 3' overhang. Subsequently, BRCA2-dependent RAD51 loading onto the single stranded DNA initiates homology search and strand invasion. To determine whether gene conversion proceeds through the HR pathway, BRCA2 and RAD51 were knocked-down using siRNA in U2OS cells, and gene conversion analyzed following Cas9 D10A nickase-induced lesions (**Fig. 10A**). Gene conversion outcomes were significantly decreased when either BRCA2 or RAD51 were knocked down, indicating that both genes are required for efficient gene conversion and suggesting that gene conversion from the *HBD* locus proceeds through the HR pathway (**Fig. 10B**).

To investigate the fate of the predicted Cas9 N863A nickase-induced 3' overhang

structure, the repair outcome of these lesions was analyzed in detail. While the predominant repair outcome for these lesions was insertions (29.9% of all modifications, as compared to 9.0% for modifications induced by the Cas9 D10A nickase), the majority of the insertions mapped to the overhang (98.5% of insertions), while only 71.4% of insertions induced by the Cas9 D10 nickase mapped to the overhang (**Fig. 11A**). Cas9 N863A nickase-induced insertions are significantly longer than the insertions induced by Cas9 D10A nickase (**Fig. 11B**). Cas9 N863A nickase-induced insertions also predominantly stem from the central part of the overhang, which is different from the Cas9 D10A nickase-induced insertions, which mostly derive from the first 20 nts of the overhang structure (**Fig. 12A**). Moreover, Cas9 N863A nickase-induced insertions more frequently contained several (up to seven) repetitions of the entire 47 nts overhang structure generated upon cleavage, and in all cases of full overhang repetition, microhomology usage was evident (**Fig. 11A and 12B**). These observations suggest a mechanistic difference between the Cas9 N863A nickase- and Cas9 D10A nickase-induced insertions. To test the effect that different lesions had on the engagement of HDR, a donor template was provided as a single-strand oligo or as ds DNA donor. In both cases the length of the donor is approximately 170 bp with 60 bp of homology outside the nicks and with 8 mismatches (**Fig. 13A**). As shown in **Fig. 13B**, the Cas9 D10A nickase that resulted in a 5' overhang resulted in a significantly higher rate of HDR, especially when using the upper stand as a single-strand oligo donor. **Fig. 13C** shows different forms of donors (dsDNA, upper stand, and lower strand) and their contribution to HDR.

To further investigate the different repair outcomes resulting from the use of the different Cas9 variants, gene correction efficiency using a 179 nt ssODN that targets the *HBB* locus in the area of the sickle mutation was tested. Similar to the outcome observed during gene conversion, Cas9 D10A nickase, which creates a 5' overhang, yielded the highest rate of gene correction (23.8% for Cas9 D10A nickase, as compared to 7.7% for WT Cas9 ($p=0.0024$) and 7.5% for Cas9 N863A nickase ($p=0.0016$)) (**Fig. 14**). To further dissect the genetic requirements of the ssODN gene correction approach, HR components BRCA2 and RAD51 were knocked down using siRNA. No change in gene correction frequency was observed, indicating that gene correction using a ssODN proceeds through a pathway that is independent of BRCA2 and RAD51 (**Fig. 10C**).

Given the differences in repair outcome and repair pathway engagement observed using the different Cas9 variants using a paired nickase strategy, repair outcome and repair pathway engagement resulting from a single nicking strategy was investigated.

First, the repair outcome in cells treated with the Cas9 D10A nickase with either a single gRNA or two gRNA molecules was analyzed. As shown in **Fig. 15**, gene conversion was observed with Cas9 D10A nickase with a single gRNA molecule (*i.e.*, 8gRNA or 15gRNA), as compared to two gRNA molecules (8gRNA/ 15gRNA pair). With a single gRNA, a reduction in the overall frequency of gene editing events was observed, as compared to the use of a pair of gRNA molecules. However, a similar distribution of the different types of editing events was maintained.

In additional experiments, the repair outcome and repair pathway engagement was examined in cells treated with the 8gRNA in combination with either the Cas9 D10A nickase, which results in a single DNA nick in the target strand, with the Cas9 N863A nickase, which results in a single DNA nick in the non-target strand, or with the control WT Cas9, which results in a blunt double stranded break. The overall modification frequency observed for WT Cas9-induced lesions was significantly higher than that observed for the Cas9 D10A nickase- and Cas9 863A nickase-generated single nicks (66.3% for WT Cas9, as compared to 4.6% for Cas9 D10A nickase and 1.3% for Cas9 N863A nickase) (**Fig. 16A**). This difference suggests the presence of a highly efficient single nick repair mechanism. To determine whether the difference in detectable repair events between Cas9 D10A nickase-induced single nick lesions and Cas9 N863A nickase-induced single nick lesions (approximately 4 fold; 4.6% for Cas9 D10A nickase vs. 1.3% for Cas9 N863A nickase, $p=3.1 \times 10^{-6}$) was attributable to differences in cutting efficiency between the two nickase variants, an *in vitro* DNA cutting assay was performed in which a gel-shifted DNA band indicates successful DNA cleavage. No significant differences in cutting efficiency between the Cas9 D10A nickase or the Cas9 N863A nickase were observed (**Fig. 16B**), suggesting that either the DNA repair of Cas9 N863A nickase-induced single nick of the non-target strand proceeds more efficiently than the repair of the cleaved target DNA strand, or that Cas9 D10A nickase-induced nick results in a more mutagenic intermediate.

To determine whether gene correction and gene conversion events that result from single nicking using a Cas9 nickase variant proceed through similar pathways as described above for the paired nickase approach (*i.e.*, utilizing a gRNA pair and Cas9 nickase), BRCA2 and RAD51 were knocked-down using siRNA in U2OS cells, and repair outcomes analyzed. Gene conversion was dependent on RAD51 for WT Cas9-, Cas9 D10A nickase-, and Cas9 N863A nickase-induced single nicks (**Fig. 17A and 18A**). An approximately 6-fold increase in gene correction was observed upon RAD51 knockdown for Cas9 D10A nickase-induced single nicks (from 1.0% in FF control to 6.24% in siRAD51-treated cells). However, an

increase in gene correction was not observed upon RAD51 knockdown for Cas9 N863A nickase-induced single nicks (**Fig. 17A and 18A**). Similar results were obtained with BRCA2 depletion (**Fig. 17A**). An equivalent simultaneous increase in overall modifications for Cas9 D10A nickase-induced nicks was also observed upon RAD51 knockdown (from 6.1% with the FF control to 48.4% in siRAD51-treated cells), without a relative increase in gene correction efficiency (**Fig. 17A and 18A**). A significant increase in the size of deletions resulting from Cas9 D10A nickase-induced single nicks was also observed upon RAD51 knockdown (**Fig. 18B**). Overall, these observations suggest that Cas9 D10A nickase-induced nicks are converted to double stranded breaks during S-phase and are then repaired by more error prone pathways than HR. It is possible that Cas9 D10A nickase-induced lesions might not be sensed until S-phase, when they are converted to DSBs (see, *e.g.*, Mayle 2015; Saleh-Gohari 2005; Neelsen and Lopes 2015). This may be due to the inability of RAD51- and BRCA2-deficient cells to undergo replication fork remodeling to prevent replication fork run-off or undesirable processing at the Cas9 D10A nickase-induced nick (**Fig. 18C**, left panel) (see, *e.g.*, Neelsen and Lopes 2015; Schlacher 2011; Zellweger 2015). Alternatively, the high degree of locus disruption could be due to the inability of RAD51/BRCA2 deficient cells to complete HR if run-off has occurred (**Fig. 18C**, right panel). The observed increase in the size of the deletions in the absence of RAD51 could suggest the involvement of A-NHEJ pathways in the repair of Cas9 D10A nickase-induced nicks during S-phase. These data cumulatively provide evidence that both RAD51 and BRCA2 are critical in promoting high fidelity repair of nicks that escape the SSBR pathway.

In summary, Cas9 nickases (D10A and N863A) showed comparable levels of efficacy compared to wild type Cas9. Different DNA ends engage different repair pathways. The use of a wild type Cas9 generates a blunt end, which are preferentially repaired by canonical NHEJ. Use of a Cas9 nickase with two gRNAs generates either 3' or 5' overhangs, which are not suitable substrates to be repaired by canonical NHEJ but can be repaired by alternative pathways.

The 5' protruding end was mostly repaired through a mechanism called gene conversion in which the *HBB* gene is repaired by using the *HBD* locus as a template. Use of nickase is advantageous to promote HDR. In the experiments in which a donor was provided, a significantly higher rate of HDR was observed using the Cas9 D10A nickase as compared to the wild type Cas9. Also, a significantly higher rate of gene correction was observed using the Cas9 D10A nickase as compared to the wild type Cas9 or the Cas9 N863A nickase. The nature of the donor template influences the outcome as HDR was

preferentially observed when a single stranded (SS) oligo was used.

Methods

gRNA selection and production - A list of gRNAs directing *S. pyogenes* Cas9 were designed to target the human *hemoglobin beta (HBB)* locus requiring a NGG PAM (protospacer adjacent motif). Potential gRNAs were aligned against the human genome sequence to identify potential off-target sites. Variations in PAM sequence (NGG or NAG) were allowed for gRNAs with spacer lengths of 20 nucleotides. Primers containing gRNA sequences were synthesized (Integrated DNA Technologies (IDT)). gRNAs 8 and 15 were generated by cloning annealed gRNA sequence oligos containing the target sequence into pMLM3636 (Addgene plasmid # 43860) which contains a *S. pyogenes* Cas9-TRACR and a customizable U6-driven gRNA scaffold. Plasmid DNA was purified using MaxiPrep Kit (ThermoFisher) and sequences were confirmed by Sanger sequencing.

Cell lines and cell culture - HEK293 (ATCC #CRL-1573), K562 (ATCC #CCL-243) and U2-OS (ATCC #HTB-96) cells were maintained in Dulbecco's Modified Eagle Medium (DMEM; Life Technologies) supplemented with 10 % fetal bovine serum (FBS) and 5 % penicillin/streptomycin, and 2mM Glutamax. Cells were kept at 37 °C in a 5% CO₂ incubator.

Transfections/Electroporations - 2.5×10^5 cells were transfected using the Lonza 4D Nucleofector (AAF-1002B 4D-Nucleofector™ Core unit, AAF-1002X 4D-Nucleofector™ X unit, SE Cell Line 4D-Nucleofector X Kit S V4XC-1032) with 200 ng of gRNA plasmid and 750 ng of Cas9-variant encoding plasmid (pJDS246-WT; Cas9/pAF001-N863A; Cas9/pJDS271-D10ACas9) in the presence or absence of 50 pmol single-stranded deoxynucleotide donor (ssODN) using the DN-100 program. After nucleofection, cells were incubated at room temperature for 10 min. and then resuspended in complete DMEM before transfer to 6-well plates containing pre-warmed media. Media was changed 72 hours after nucleofection, and cells were harvested 5 days post nucleofection.

siRNA Knockdowns - To knock down BRCA2 and RAD51, immediately before nucleofection with Cas9-variant encoding plasmid and gRNA(s), cells were treated with 30 pmols of siBRCA2 (siGENOME Human BRCA2 (675) siRNA – SMARTpool; M-003462-

01-0005), 15 pmols of both siRAD51 (siGENOME Human RAD51 (5888) siRNA; D-003530-05-0005) , or 15 pmols of control firefly luciferase siRNA ("FF") (FF siRNA; CTM-127256). Samples were collected for gDNA extraction and further processing 5 days after nucleofection and knockdown efficiency assessed by Western Blot.

5

Genomic DNA Analysis - Genomic DNA was extracted from cells (Beckman Coulter Agencourt DNAdvance #A48706) according to the manufacturer's directions. The targeted genomic region of *HBB* was amplified by PCR in a 50 µl reaction mixture (10 µL of 5X Phusion HF buffer; 0.5 µM forward primer (F1_HBB, 5'-AGGCCATCACTAAAGGCACC (SEQ ID NO: 16313)); 0.5 µM reverse primer (R_HBB, 5'-TAAGCCAGTGCCAGAAGAGC (SEQ ID NO:16314)); 200 µM dNTP; 1.5 µL DMSO; 0.5 µl of Phusion polymerase; 100 ng of gDNA template) using the following PCR conditions: 30 s at 98°C for initial denaturation, followed by 30 cycles of 10 seconds (s) at 98°C for denaturation, 15s at 64°C for annealing, 30s at 72°C for extension, and 5 minutes at 72°C for the final extension. PCR products were purified using (1.8×) Agencourt AMPure XP beads (Beckman Coulter Agencourt AMPure XP - PCR Purification #A63882) as directed by the manufacturer. Amplified locus fragments were cloned into pCR4-TOPO vectors using the ZeroBlunt TOPO Cloning Kit (Life Technologies Cat. No. #K287540) and transformed in One Shot Top10 chemically competent *E. coli* cells. Cells were plated on carbenicillin LB agar plates and incubated overnight at 37°C. Plasmid DNA from 96 colonies per sample was sequenced at MacroGen Corp. and Genewiz, Inc. using an M13 forward or reverse primer.

Single cell cloning - 2.5×10^5 U2OS cells per condition were nucleofected using Lonza 4D Nucleofector (AAF-1002B 4D-Nucleofector™ Core unit, AAF-1002X 4D-Nucleofector™ X unit). The cells were plated in 6 well dishes. For each of the conditions, 0.3 cells/well were plated in 3X 96 well plates. After 3 weeks, the 96 well plates were visually scored using microscopy for single colony formations. More than 80 clones were picked for each condition and analyzed by TOPO cloning the PCR products of the amplified locus. Competent cells were plated onto divided bioassay plates with LB agar containing ampicillin (Molecular Devices Vented QTray, Fischer Scientific Cat. No. # NC0183882) and 6 colonies were picked for each subclone. Genomic DNA was extracted from cells using DNAdvance kits (Beckman Coulter Agencourt DNAdvance - Nucleic Acid Isolation from Mammalian Tissue #A48706). The *HBB* locus of the genomic DNA from individual clones was analyzed as described above. 6 colonies were analyzed for each subclone.

Illumina Next Generation Sequencing - Two rounds of PCR were performed on gDNA. The first round of amplification was used to amplify the *HBB* locus. Amplification was performed in a 50 µl reaction volume, (10 µL of 5X HercII buffer; 0.5 µM forward primer (HBB Miseq 1F: 5' CCATCTCATCCCTGCGTGTCTCCGACCACCAGCAGCCTAAGG (SEQ ID NO:16315)); 0.5 µM reverse primer (HBB Miseq 1R: 5' CCTCTCTATGGGCAGTCGGTGATGGCCATCTATTGCTTACATTTGCT (SEQ ID NO:16316)); 50 µM dNTP; 50mM MgCl₂; 0.5 µl of HercII polymerase; 250 ng of gDNA template) using the following conditions: 2 min. at 95°C for initial denaturation, followed by 30 cycles of 20 s at 95°C for denaturation, 20 s at 60°C for annealing, 20 s at 72°C for extension, and 3 min. at 72°C for the final extension. A second round of PCR amplification was performed to incorporate the P7 and P5 Illumina adapters and a unique 8-mer barcode sequence in a 50 µl reaction mixture (10 µL of 5X HercII buffer; 0.5 µM forward primer; 0.5 µM reverse primer; 50uM dNTP; 0.5ul of HercII polymerase; 3ul of round 1 PCR product) using the amplification conditions as the first round of PC R amplification. PCR products were purified using (1.8x) Agencourt AMPure XP beads, as directed by the manufacturer, eluted in 40 µl of H₂O, and size verified by capillary electrophoresis using the Qiagen QIAxcel Advanced System. PCR products were quantified using Picogreen, normalized into one library pool and sequenced on the Illumina MiSeq according to the manufacturer's protocols.

Quantification of editing events in NGS data - The CRISPResso software (v0.9.1)(see, e.g., Pinello *et al.*) was used to identify editing events in high-throughput sequencing data and determine the frequencies of insertion, deletion, gene conversion and gene correction events. To identify insertions and deletions events, CRISPResso default settings were used to obtain the number of reads with either insertions or deletions. To identify gene conversion and gene correction events, CRISPResso was used separately for each sample and potential HDR template (*i.e.*, the orthologous HBD sequence and the ssODN) with a required sequence similarity of 95% (using the `hdr_perfect_alignment_threshold` parameter).

Determination of insertion sizes - To identify insertions from Sanger sequencing data, the Needleman-Wunsch algorithm implementation in the Biopython (`pairwise2.globalms`)

package was used with the following parameters: base match score = 2, base mismatch penalty = -1, gap open penalty = -50, gap extend penalty = -1. Insertion sequences were identified by taking the contiguous bases aligned to gap characters in the reference sequence.

5 *Overhang analysis* - Overhang sequence was identified as located between the two predicted cleavage sites for guide RNA pairs, under the assumption that cleavage occurs three nucleotides 5' of the start of the NGG PAM. For the gRNA 8/15 pair, the predicted overhang sequence is:

TCATCCACGTTACCTTGCCCCACAGGGCAGTAACGGCAGACTTCTC (SEQ ID
10 NO:16317).

Insertion sequences were classified according to the absence of matching overhang sequence, the presence of a partial matching sequence (at least one 9-mer in common), or the presence of a full matching sequence (complete overhang sequence contained within the insertion sequence). Fisher's exact test was used to determine statistical significance of differences in
15 the proportion of sequences with partial or full overhang matches generated with Cas9 D10A nickase or Cas9 N863A nickase. Segments of the overhang sequence overrepresented in insertions were identified by extracting all 4-mer sequences present in each insertion sequence and identifying the positions in the overhang where the 4-mer sequence exactly
20 matched. 4-mers found multiple times within the overhang sequence were excluded from downstream analyses. Statistical significance of the distribution of 4-mer positions resulting from Cas9 D10A nickase- or Cas9 N863A nickase-induced lesions was tested using a permutation test (permTS function from the "perm" R package with default settings).

Example 5: Assessment of Gene Targeting in Hematopoietic Stem Cells

25 Transplantation of autologous CD34⁺ hematopoietic stem cells (HSCs, also known as hematopoietic stem/progenitor cells or HSPCs) genetically modified to correct the Sickle Cell Disease (SCD) mutation in the human β -hemoglobin gene (*HBB*) would prevent deformability (sickling) after deoxygenation in the erythrocyte progeny of corrected HSCs which could ameliorate symptoms associated with SCD. Genome editing with the
30 CRISPR/Cas9 platform precisely alters endogenous gene targets by creating an indel at the targeted cut site that can lead to knock down of gene expression at the edited locus. In this Example, genome editing in the human K562 bone marrow erythroleukemia cell line, which serve as a proxy for HSCs and which can be predictive of genome editing in HSCs, were

electroporated with Cas9 mRNA and gRNA HBB-8 and gRNA HBB-15 to induce gene editing at the human *HBB* locus.

K562 cells were grown in RPMI media (Life Technologies) containing 10% fetal bovine serum (FBS). *S. pyogenes* Cas9 mRNA and gRNA HBB-15 and gRNA HBB-8 were prepared by *in vitro* transcription using linearized plasmid DNA as templates and the Ambion mMessage mMachine® T7 Ultra Transcription kit (Life Technologies) according to the manufacturer's instructions. In this embodiment, both the Cas9 and gRNA were *in vitro* transcribed using a T7 polymerase. For example, a 5' ARCA cap was added to both RNA species simultaneous to transcription while a polyA tail was added after transcription to the 3' end of the RNA species by an *E. coli* polyA polymerase. Capped and tailed gRNA HBB-8 and gRNA HBB-15 were complexed at room temperature with *S. pyogenes* H-NLS-Cas9 protein at a molar ratio of ~25:1 (gRNA : Cas9 protein) in a total of 30 µg RNP. Briefly, three million K562 cells were suspended in 100 µL electroporation buffer and transferred to the RNP solution (13µL). In addition, K562 cells were electroporated with *S. pyogenes* Cas9 mRNA and each of the gRNA HBB-8 and gRNA HBB-15. For the mRNA/gRNA electroporation, 10 µg of gRNA HBB-8 (or 10 µg of *HBB* gRNA HBB-15) were mixed with 10 µg of Cas9 mRNA. Four million K562 cells were suspended in 100 µL electroporation buffer and then transferred to the mRNA/gRNA solution (13 µL). K562 cells mixed with either RNP or RNA were electroporated. At 48 hours after electroporation, K562 cells were enumerated by trypan blue exclusion and were determined to have >88% viability in the electroporated cell populations. Genomic DNA was extracted from K562 cells 48 hours after electroporation and *HBB* locus-specific PCR reactions were performed.

In order to detect indels at the *HBB* locus, T7E1 assays were performed on *HBB* locus-specific PCR products that were amplified from genomic DNA samples from electroporated K562 cells and the percentage of indels detected at the *HBB* locus was calculated (**Fig. 19**).

Co-delivery of 10 µg RNP which contains wild-type *S. pyogenes* Cas9 protein with *HBB* gRNA 8 or HBB gRNA 15 resulted in 26.8% and 16.1% indels, respectively, at the *HBB* locus in gDNA from K562 cells (molar ratio protein: gRNA 24:1). Co-delivery of Cas9 mRNA with gRNA HBB-8 or HBB-15 led to 66.9% and 29.5% indels at the *HBB* locus in gDNA from K562 cells (10 µg of each RNA/4 million cells). This example shows that delivery of Cas9 mRNA/gRNA and Cas9 RNPs leads to editing of the *HBB* locus in a relevant bone marrow derived hematopoietic cell line (K562 cells). Clinically, transplantation of autologous HSCs in which the *HBB* locus has been edited to correct the

genetic mutation that causes red blood cell sickling could be used to ameliorate symptoms of SCD.

Example 6: Contribution of Different Pairs of gRNA Molecules for Gene Correction

The CRISPR/Cas9 system was used to target the human *HBB* gene in the region of the sickle cell anemia-causing mutation.

Three different gRNA pairs were evaluated to examine how different pairs of gRNA molecules allocated at different distances can affect the frequency of different DNA repair outcomes in the presence of dual-nicks, in which the nicks are placed on opposite strands and leave either 3' or 5' overhangs of varying lengths. DNA repair outcomes including, *e.g.*, indel mutations resulting from non-homologous end-joining (NHEJ), homology-dependent repair (HDR) using a donor as a template, and HDR using the closely related *HBD* gene as an endogenous template, were characterized. The frequencies of these various repair outcomes under different conditions offer insights into the mechanism of DNA repair and how it is impacted by the nature of the DNA break, the length of the overhangs, and/or the nature of the sequence exposed.

In this study different gRNA molecules for the *HBB* region that surround the nucleotides encoding the amino acid that is most commonly mutated in sickle cell disease were tested in 293T cells with wild type Cas9. The gRNA molecules that induced similar high rates of NHEJ and had PAMs facing in opposite orientations were selected to test as pairs with Cas9 D10A and Cas9 N863A nickases.

As shown in **Fig. 20**, the gRNA pairs 8/15 ("8gRNA"/ "15gRNA" pair), 8/19 ("8gRNA"/ "19gRNA" pair) and 8/21 ("8gRNA"/ "21gRNA" pair) were selected. ("8gRNA"/ "15gRNA" pair). The targeting domain sequences for "8gRNA" and "15gRNA" are described above. "19gRNA" has the targeting domain sequence of CCUGUGGGGCAAGGUGAACG (SEQ ID NO: 419) and "21gRNA" has the targeting domain sequence of UGAAGUUGGUGGUGAGGCC (SEQ ID NO: 431). These pairs of gRNA molecules used in combination with the mutant Cas9 D10A generated a 5' overhang of 47 bp, 37 bp and 61 bp, respectively; in combination with the mutant Cas9 N863A generated a 3' overhang of 47 bp, 37 bp and 61 bp, respectively.

In this Example, U2OS cells were electroporated with 200 ng of each gRNA and 750 ng of plasmid that encodes mutant Cas9 (D10A or N863A). Cells were collected 6 days after electroporation and genomic DNA was extracted. PCR amplification of the *HBB* locus was performed and the product was subcloned into a Topo Blunt Vector. For each condition in

each experiment, 96 colonies were sequenced with Sanger sequencing. In the experiments assessing HDR efficacy, cells were electroporated with 2.5 µg of single stranded oligo or double stranded oligo in addition to the gRNA and the Cas9-encoding plasmid.

As shown in **Fig. 21**, the total percentages of all editing events detected by Sanger sequencing of the *HBB* locus were similar using the different pairs of gRNA molecules (Cas9 N863A was less active compared to Cas9 D10A). **Fig. 21** indicates that the overall editing frequencies for both Cas9 D10A and Cas9 N863A nickases were independent of overhang length and composition.

Fig. 22 shows that for the Cas9 N863A nickase, a majority of the gene editing events were insertions with all three pairs of gRNA molecules (Cas9 N863A had a higher level of insertions compared to Cas9 D10A). **Fig. 22** indicates that the Cas9 N863 nickase yielded a higher frequency of insertions, which was independent of overhang length and composition, while the Cas9 D10A insertions were overall less frequent, they appeared to be sensitive to overhang length and composition.

As shown in **Fig. 23**, the Cas9 N863A-induced insertions are mostly duplications of the overhang regions, as opposed to the overhangs coming from the Cas9 D10A, *i.e.*, with specific gRNA pair combinations, the Cas9 N863 nickase had a higher percentage of insertions derived from the overhang sequence than the Cas9 D10A nickase. Moreover, the lengths of the insertions between Cas9 D10A and Cas9 N863A were substantially different with longer insertions being observed using the Cas9 N863A nickase (**Fig. 24**).

As shown in **Fig. 25**, in the case of the Cas9 D10A nickase, which leaves a 5' protruding end, the lesion was mostly repaired through a mechanism of gene conversion with a significantly higher rate compared to the Cas9 N863A nickase. Not all gRNA pairs tested in this example behaved similarly. The 8/15 gRNA pair was mostly repaired using a gene conversion mechanism, while the 8/19 and 8/21 pairs yielded a significant reduction of gene conversion frequency. Specifically, the 8/21 pair displayed a 3 fold reduction in gene conversion efficiency. These data suggest that either the length of the overhang or the nature of the sequence exposed, or both, could affect the DNA repair machinery. These data indicate that gene conversion was depending on the type of Cas nickase and on the sequence and length of overhangs.

To test the effect that different pairs had on the engagement of HDR, a donor template was provided as a single-strand oligo or as a double stranded (ds) DNA donor. In both cases the length of the donor is approximately 170 bp with 60 bp of homology outside the nicks and with 8 mismatches (**Fig. 26**). As shown in **Fig. 26**, all three pairs resulted in similar

HDR frequencies, *i.e.*, the frequency of gene correction from exogenous donor was comparable despite of the length of overhangs. Neither Cas9 D10A nor Cas9 N863A induced gene correction from exogenous donor was dependent on overhang length and composition. **Fig. 26** indicates that gene correction from exogenous donor was more efficient with the Cas9 D10A nuclease compared to the Cas9 N863A nuclease.

Example 7: Contact between *S. pyogenes* Cas9 ribonucleoprotein complexed to gRNAs targeting the *HBB* genetic locus supports gene editing in adult human hematopoietic stem cells

Transplantation of autologous CD34⁺ hematopoietic stem cells (HSCs) collected from patients affected with hemoglobinopathies (*e.g.*, sickle cell disease [SCD], β -thalassemia), that have been genetically modified with a lentivirus vector that expresses non-sickling β -hemoglobin gene (*HBB*) has been shown to restore expression of functional adult hemoglobin (HbA) thus preventing the formation of sickle hemoglobin (HbSS), in erythroid cells derived from transduced CD34⁺ cells and ameliorating clinical symptoms in affected patients (Press Release from *Bluebird Bio*, June 13, 2015, “bluebird bio Reports New Beta-thalassemia major and Severe Sickle Cell Disease Data from HGB-205 study at EHA”). However, delivery of a transgene encoding a non-sickling β -hemoglobin does not correct the causative mutation or prevent expression of the mutant (*e.g.*, sickling) form of *HBB*. Furthermore, lentivirus vector transduction of CD34⁺ cells can lead to the occurrence of multiple transgene integration sites per cell, and the long-term effects of multiple transgene integration sites is currently undetermined.

In contrast, genome editing with the CRISPR/Cas9 platform precisely alters endogenous gene targets by creating an insertion or deletion (indel) at the cut site that can lead to gene disruption at the edited locus. Co-delivery of two gRNAs each targeting regions proximal to the single nucleotide polymorphism (SNP) that encodes HbSS (*e.g.*, GAG→GTG, which results in a change in the amino acid residue from glutamic acid to valine) co-delivered with a Cas9 D10A nickase supports a low level of homology directed repair (HDR) in human cell lines (*e.g.*, gene conversion using a region of homology in the *HBD* locus as DNA repair template).

In this Example, genome editing in adult human mobilized peripheral blood CD34⁺ HSCs after co-delivery of Cas9 D10A nickase with two gRNAs targeting the *HBB* locus was evaluated. The edited CD34⁺ cells were then differentiated into myeloid and erythroid cells to

determine the hematopoietic activity of the HSCs. Gene editing at the *HBB* locus was evaluated by T7E1 analysis and DNA sequencing. Expression of *HBB* protein was also analyzed in erythroid progeny.

Human CD34⁺ HSCs cells from mobilized peripheral blood (AllCells®) were thawed into StemSpan Serum-Free Expansion Medium (SFEM™, StemCell Technologies) containing 300 ng/mL each of human stem cell factor (SCF) and flt-3 ligand (FL), 100 ng/mL thrombopoietin (TPO), and 60 ng/mL of IL-6, and 10 μM PGE2 (Cayman Biochemicals; all other supplements were from PeproTech® unless otherwise indicated). Cells were grown for 3 days in a humidified incubator and 5% CO₂ 20% O₂. On day 3 (morning), media was replaced with fresh Stemspan-SFEM™ supplemented with human SCF, TPO, FL and PGE2. In the afternoon of day 3, 2.5 million CD34⁺ cells per sample were suspended in electroporation buffer.

The gRNA was generated by *in vitro* transcription using a T7 polymerase. A 5' ARCA cap was added to the RNA simultaneous to transcription while a polyA tail was added after transcription to the 3' end of the RNA species by an *E. coli* polyA polymerase.

After the gRNAs were *in vitro* transcribed and tailed, the quality and quantity of gRNAs were evaluated with the Bioanalyzer (Nanochip®) to determine RNA concentration and by Differential Scanning Fluorimetry (DSF) assay, a thermal shift assay that quantifies the change in the thermal denaturation temperature of Cas9 protein with and without complexing to gRNA. In DSF assays, the Cas9 protein was mixed with gRNA and allowed to form complexes for 10 minutes. Cas9 protein : gRNA were mixed at a molar ration of 1:1, and the DSF assay performed as a measure of Cas9 stability and as an indirect measure of gRNA quality, since a 1:1 ratio of gRNA:Cas9 should support a thermal shift if the gRNA is of good quality (**Figure 27A**).

For half of the samples, *in vitro* transcribed capped and tailed guide (g)RNAs *HBB*-8 and *HBB*-15 were added at a 2:1 molar ratio to 12.5μg D10A Cas9 ribonucleoprotein (RNP) (5μg RNP per million cells) to 2.5 million cells. “HBB-8” has the targeting domain sequence of GUAACGGCAGACUUCUCCUC (SEQ ID NO:388), and “HBB-15” has the targeting domain sequence of AAGGUGAACGUGGAUGAAGU (SEQ ID NO:387). D10A protein and gRNAs RNP complexes were transferred to 2.5 million adult CD34⁺ cells in electroporation buffer. The RNP/cell mixture was transferred to an electroporation cartridge, and the cells then electroporated (“Program 2” and “Program 3”).

For the second portion of CD34⁺ cell samples, equal amounts (5 μg or 10 μg [2XgRNA] each) of *in vitro* transcribed capped and tailed guide (g)RNAs *HBB*-8 and *HBB*-

15 were added to 10 µg of *in vitro* transcribed Cas9 D10A mRNA. The mRNA : gRNA: cell mixture was transferred to an electroporation cartridge and electroporated using Program 2 (P2).

For all samples, the cells were collected from the cartridge and placed at 37°C for 20
 5 minutes (recovery period). Then, the cells were either transferred to pre-warmed cytokine supplemented Stemspan-SFEM™ media and placed at 30°C for 2 hours (cold shock samples) or placed directly into 37°C. For the cold shocked samples, the cells were transferred to the 37°C incubator after the 2-hour incubation period at 30°C. At 24, 48, and 72 hours after electroporation, the CD34⁺ cells were counted by trypan blue exclusion (cell survival) and
 10 divided into 3 portions for the following analyses: a) flow cytometry analysis for assessment of viability by co-staining with 7-Aminoactinomycin-D (7-AAD) and allophycocyanin (APC)-conjugated Annexin-V antibody (eBioscience); b) flow cytometry analysis for maintenance of HSC phenotype (after co-staining with phycoerythrin (PE)-conjugated anti-human CD34 antibody (BD Biosciences) and APC-conjugated CD133 (Miltenyi Biotech; c)
 15 hematopoietic colony forming cell (CFC) analysis by plating 800 cells in semi-solid methylcellulose based Methocult medium (StemCell Technologies H4435) that supports differentiation of erythroid and myeloid blood cell colonies from HSCs and serves as a surrogate assay to evaluate HSC multipotency and differentiation potential *ex vivo*; d)
 genomic DNA analysis for detection of editing at the *HBB* locus. Genomic DNA was
 20 extracted from the HSCs at 48 and 72 hours after electroporation and *HBB* locus-specific PCR reactions were performed. The purified PCR products were analyzed for insertions/deletions (indels) in T7E1 assays and by DNA sequencing of individual clones (PCR product was transformed and sub-cloned into TOPO-vector, individual colonies picked, and plasmid DNA containing individual PCR products were sequenced).

25 Western blot analysis of cell lysates extracted from the Cas9/gRNA electroporated CD34⁺ cells indicated the presence of Cas9 protein at 72 hours after electroporation of CD34⁺ cells that received Cas9 RNP and gRNA pair. Very low levels of Cas9 protein were detected in the lysates of cells that were electroporated with Cas9 mRNA (**Figure 27B**).

Electroporated CD34⁺ cells maintained a stem cell phenotype (*e.g.*, co-expression of
 30 CD34 and CD133) and viability (*e.g.*, 75% AnnexinV⁺7AAD⁻) as determined by flow cytometry analysis (**Figure 28A**). The absolute number of viable CD34⁺ cells was maintained across most samples over a 72-hour culture period after electroporation (**Figure 28B**). In addition, gene edited CD34⁺ cells maintained *ex vivo* hematopoietic activity and multipotency

as indicated by their ability to give rise to erythroid (*e.g.*, CFU-E or CFU-GEMM) and myeloid (*e.g.*, CFU-G, -M or -GM) (**Figure 28C**).

Gene editing at the *HBB* locus was then evaluated at 72 hours after electroporation of D10A mRNA or RNP co-delivered with two gRNAs (HBB-8 and HBB-15). Briefly, genomic (g)DNA was isolated from electroporated CD34⁺ cells at 72 hours after electroporation, and PCR amplification of a ~607bp fragment of the *HBB* locus (which captured both of the individual genomic locations that were targeted by the two gRNAs HBB-8 and HBB-15) was performed. After cleanup of the *HBB* PCR product with AMPURE beads, insertions/deletions (indels) at the targeted genomic location were evaluated by T7E1 assay and by DNA sequencing. For the CD34⁺ cells electroporated with D10A mRNA and HBB targeting gRNA pair, no indels were detected (**Table 10**). In contrast to the negative results obtained after delivery of D10A mRNA, ~30-60% indels were detected by T7E1 and sequencing analysis of CD34⁺ cells that were electroporated with D10A RNP with the *HBB* targeting gRNA pair (**Table 10, Figure 29A-29C**). The cells that were cultured for 2 hours at 30°C (after a 20-minute recovery period at 37°C) exhibited 57% editing as determined by DNA sequencing. In addition, gene conversion (*e.g.*, *HBD* genomic sequence used as a template copy for DNA repair of the disrupted *HBB* locus) was detected in the gDNA from CD34⁺ cells that were 'cold shocked' (30°C incubation) at a frequency of 3% relative to the total gene editing events (**Figure 29C**).

Table 10. Summary of gene editing results in adult CD34⁺ cells 72 hours after co-delivery of D10A Cas9 and HBB specific gRNA pair. *Alternate electroporation program (P3).

D10A source	μg D10A	μg HBB-8 gRNA	μg HBB-15 gRNA	Temperature of 2-hour recovery	%indels (T7E1)	%indels (sequencing)
mRNA	10	5	5	37°C	0	ND
mRNA	10	5	5	30°C	0	ND
mRNA	10	10	10	37°C	0	ND
RNP	12.5	3.3	3.3	37°C	45	30
RNP	12.5	3.3	3.3	30°C	39	57
RNP*	12.5	3.3	3.3	37°C	39	39

To determine whether targeted disruption of the *HBB* locus induced changes in expression of b-hemoglobin protein, CFU-E colonies were picked, dissociated, fixed, permeabilized and stained with PE-conjugated mouse anti-human β -hemoglobin antibody (Santa Cruz Biotechnology®). The erythroid progeny of *HBB* gene edited cells exhibited a 7- to-10 fold reduction in β -hemoglobin expression compared to the CFU-Es differentiated from untreated control CD34⁺ cells (**Figure 30**). These data show that the progeny of gene edited cells retain erythroid differentiation potential and that gene editing events detected in the parental CD34⁺ cells result in reduced protein expression in erythroid progeny.

Example 8: Contact between *S. pyogenes* D10A Cas9 nickase ribonucleoprotein complexed to gRNAs targeting the *HBB* genetic locus supports gene editing in fresh umbilical cord blood derived human CD34⁺ hematopoietic stem cells

In this Example, genome editing in freshly collected umbilical cord blood (CB) CD34⁺ HSCs after co-delivery of D10A Cas9 nickase with two gRNAs targeting the *HBB* locus was evaluated. The edited CB CD34⁺ cells were then differentiated into myeloid and erythroid cells to determine the hematopoietic activity of the HSCs. Targeted disruption of the *HBB* locus was evaluated by T7E1 analysis and DNA sequencing.

Human CD34⁺ HSCs cells were isolated from freshly obtained human umbilical cord blood by ficoll gradient density centrifugation followed by MACS® (antibody conjugated immunomagnetic bead sorting) with mouse anti-human CD34⁺ immunomagnetic beads using the human CD34 cell enrichment kit and LS magnetic columns from Miltenyi Biotech. The cells were plated into StemSpan Serum-Free Expansion Medium (SFEM™, StemCell Technologies) containing 100 ng/mL each of human stem cell factor (SCF) and flt-3 ligand (FL), 20 ng/mL each of thrombopoietin (TPO) and IL-6, and 10 μ M PGE2 (Cayman Biochemicals; all other supplements were from Peprotech unless otherwise indicated). Cells were grown for 3 days in a humidified incubator and 5% CO₂ 20% O₂. On day 3 (morning), media was replaced with fresh Stemspan-SFEM™ supplemented with human SCF, TPO, FL and PGE2. In the afternoon of day 3, 2.5 million CD34⁺ cells per sample were suspended in electroporation buffer.

The gRNAs were generated by *in vitro* transcription using a T7 polymerase. A 5' ARCA cap was added to the RNA simultaneous to transcription while a polyA tail was added after transcription to the 3' end of the RNA species by an *E. coli* polyA polymerase. After the gRNAs were *in vitro* transcribed and tailed, the quality and quantity of gRNAs were

evaluated with the Bioanalyzer (Nanochip) to determine RNA concentration and by DSF assay performed as a measure of D10A Cas9 nickase RNP stability and as an indirect measure of gRNA quality.

In vitro transcribed capped and tailed guide gRNAs *HBB-8* and *HBB-15* were added at a 2:1 molar ratio (total gRNA : Cas9 protein) to 12.5µg D10A nickase ribonucleoprotein (RNP) (5µg RNP per million cells) to each of two samples each containing 2.5 million CB CD34⁺ cells. A third CB CD34⁺ cell aliquot was mixed with 25µg D10A nickase RNP and HBB gRNAs (total gRNA: D10A ratio at 2:1). For each experimental sample, the D10A RNP/cell mixture was transferred to an electroporation cartridge, and the cells then electroporated using Program 2.

For all samples, the cells were collected from the cartridge and placed at 37°C for 20 minutes (recovery period). For the CB CD34⁺ HSCs that were contacted with 12.5µg D10A nickase RNP, one sample was transferred to pre-warmed cytokine supplemented Stemspan-SFEM™ media and placed at 30°C for 2 hours (cold shock samples) or placed directly into the same media at 37°C. For the cold shocked samples, the cells were transferred to the 37°C incubator after the 2-hour incubation period at 30°C. At 24, 48, and 72 hours after electroporation, the CB CD34⁺ HSCs were counted by trypan blue exclusion (cell survival) and divided into 3 portions for the following analyses: a) flow cytometry analysis for assessment of viability by co-staining with 7-Aminoactinomycin-D (7-AAD) and allophycocyanin (APC)-conjugated Annexin-V antibody (eBioscience); b) flow cytometry analysis for maintenance of HSC phenotype (after co-staining with phycoerythrin (PE)-conjugated anti-human CD34 antibody (BD Biosciences) and APC-conjugated CD133 (Miltenyi Biotech; c) hematopoietic colony forming cell (CFC) analysis by plating 800 CB CD34⁺ HSCs in semi-solid methylcellulose based Methocult medium (StemCell Technologies H4435) that supports differentiation of erythroid and myeloid blood cell colonies from HSCs and serves as a surrogate assay to evaluate HSC multipotency and differentiation potential *ex vivo*; d) genomic DNA analysis for detection of editing at the *HBB* locus. Genomic DNA was extracted from the HSCs at 48 and 72 hours after electroporation and *HBB* locus-specific PCR reactions were performed. The purified PCR products were analyzed for insertions/deletions (indels) in T7E1 assays and by DNA sequencing of individual clones (PCR product was transformed and subcloned into TOPO-vector, individual colonies picked, and plasmid DNA containing individual PCR products were sequenced).

Electroporated CB CD34⁺ cells maintained a stem cell phenotype (*e.g.*, co-expression of CD34 and CD133, ~90% CD34⁺CD133⁺) and viability (*e.g.*, 91% AnnexinV⁻7AAD⁻) as

determined by flow cytometry analysis (**Figure 31A**). The absolute number of viable CD34⁺ cells was maintained across most samples over a 72-hour culture period after electroporation. In addition, gene edited CD34⁺ cells maintained *ex vivo* hematopoietic activity and multipotency as indicated by their ability to give rise to erythroid (*e.g.*, CFU-E or CFU-GEMM) and myeloid (*e.g.*, CFU-G, -M, or -GM) (**Figure 31B**).

Gene editing at the *HBB* locus was then evaluated at 72 hours after electroporation of D10A nickase RNP co-delivered with 2 gRNAs (HBB-8 and HBB-15). Briefly, gDNA was isolated from electroporated CD34⁺ cells at 72 hours after electroporation, and PCR amplification of a ~607bp fragment of the *HBB* locus (which captures both of the individual genomic locations that were targeted by gRNAs HBB-8 and HBB-15) was performed. After cleanup of the *HBB* PCR product with AMPURE beads, insertions/deletions (indels) at the targeted genomic location were evaluated by T7E1 assay and by DNA sequencing. For the CD34⁺ cells electroporated with 5 µg per million cells of D10A nickase RNP and HBB-specific gRNA pair, ~20% indels were detected by T7E1 analysis (**Table 11**). In contrast to adult CD34⁺ cells, a 2-hour incubation at 30°C did not alter the level of gene editing as determined by either T7E1 analysis or DNA sequence analysis. In addition, doubling the D10A nickase RNP/gRNA concentration to 10 µg RNP per million cells nearly doubled the gene editing at the *HBB* locus to 57%, as determined by DNA sequencing analysis (**Table 11, Figures 32A-32C**). Stratification of DNA repair events through DNA sequencing analysis revealed that ~50-70% of the sequence reads contained small insertions, ~20-40% contained large deletions, and ~8% showed evidence of *HBB/HBD* gene conversion events in the targeted *HBB* genomic location (**Figure 32C**).

Table 11. Summary of gene editing results in CB CD34⁺ cells 72 hours after co-delivery of D10A nickase RNP and HBB-specific gRNA pair.

Total µg D10A RNP	µg D10A RNP/1E6 cells	µg HBB-8 gRNA	µg HBB-15 gRNA	Temperature of 2-hour recovery	Electroporation Program	%indels (T7E1)	%indels (sequencing)
12.5	5	3.3	3.3	37°C	2	20	23
12.5	5	3.3	3.3	30°C	2	27	20
25	10	6.6	6.6	37°C	2	36	51

In contrast to adult CD34⁺ cells, CB CD34⁺ cells are fetal in origin and therefore the progeny of CB CD34⁺ cells express fetal hemoglobin (HbF) which contains γ -hemoglobin instead of β -hemoglobin. Given the lack of β -hemoglobin by CB erythroblasts, disruption of the *HBB* locus in this model system will not impact expression of hemoglobin protein. CB CD34⁺ cells are used as a model system for evaluation of gene editing in HSCs, since these umbilical cord blood derived CD34⁺ cells are more readily available for research use and reconstitute immune-deficient mouse xenografts more efficiently compared to adult CD34⁺ cells.

To determine whether gene edited cells retained their erythroid differentiation potential, the edited CD34⁺ cells were induced to differentiate into erythroblasts. Briefly, CD34⁺ cells were co-cultured with human plasma, holotransferrin, insulin, hydrocortisone, and cytokines (erythropoietin, SCF, IL3), for 20 days in which the latter 4 growth factors were added at different stages of differentiation to direct erythroid specification program. The cells were then evaluated by flow cytometry for the acquisition of erythroid phenotypic characteristics including: co-expression of the transferrin receptor (CD71) and Glycophorin A (CD235); expression of HbF, and enucleation (as indicated by the absence dsDNA detected by the dsDNA dye DRAQ5) and loss of CD45 expression. By day 18 of differentiation, the erythroblast progeny of edited CD34⁺ cells possessed this red blood cell phenotype (**Figures 33A-33C**). These data, along with the CFC data shown in **Figures 31A-31B** show that the gene editing does not negatively impact the differentiation potential of CD34⁺ cells.

In summary, the data in this Example indicate: 1) electroporation of fresh CB CD34⁺ HSCs with D10A nickase and paired capped/tailed gRNAs does not impact cell viability, proliferation potential, or multipotency; and 2) contact between CB CD34⁺ HSCs and 10 μ g D10A RNP (per million cells) supports >50% gene editing with HDR events (8% gene conversion events of total).

Example 9: Contact between *S. pyogenes* wild-type Cas9 RNP or D10A nickase RNP complexed to gRNAs targeting the *HBB* genetic locus supports up to 60% gene editing in human cord blood hematopoietic stem cells

In this Example, umbilical cord blood (CB) CD34⁺ HSCs were contacted with *S. pyogenes* wild-type Cas9 RNP, N863A nickase RNP, or D10A nickase RNP complexed with 2 gRNAs targeting the *HBB* locus. The percentage of gene editing and type of editing event (e.g., HDR (e.g., gene conversion) or NHEJ) were evaluated to determine the optimal Cas9

activity (*e.g.*, type of cut, *e.g.*, double-strand break from wild-type Cas9 or off-set nicks on opposite DNA strands by nickases) for gene editing in HSCs that would favor HDR (*e.g.*, gene conversion) over NHEJ (*e.g.*, ends of the DNA left exposed after the cut, *e.g.*, blunt ends left by wild-type Cas9 cut, 5' overhang left by D10A nickase cut or 3' overhang left by N863A nickase cut). Other optimizations included: 1) removal of endotoxin from Cas9 protein preparation to reduce toxicity of Cas9 protein; 2) use of 10 µg RNP per million CD34⁺ cells to increase gene editing (shown to double gene editing in fresh CB CD34⁺ cells compared to 5 µg RNP, as indicated in **Example 8**); 3) testing of human CD34⁺ cells that were isolated from cord blood (CB), cryopreserved, and confirm that gene editing was not impacted by cryopreservation (compared to freshly isolated HSCs, described in **Example 8**); and 4) evaluate Cas9 RNP levels in CD34⁺ cells over time to understand Cas9 RNP stability in HSCs.

Human CD34⁺ HSCs cells were isolated from freshly obtained human umbilical cord blood by ficoll gradient density centrifugation followed by MACS sorting with mouse anti-human CD34⁺ immunomagnetic beads. CD34⁺ cells were cryopreserved, thawed at a later date, and plated into StemSpan Serum-Free Expansion Medium (SFEM™, StemCell Technologies) containing 100 ng/mL each of human stem cell factor (SCF) and flt-3 ligand (FL), 20 ng/mL each of thrombopoietin (TPO) and IL-6, and 10 µM PGE2 (Cayman Biochemicals; all other supplements were from PeproTech® unless otherwise indicated).

Cells were grown for 3 days in a humidified incubator and 5% CO₂ 20% O₂. On day 3 (morning), media was replaced with fresh Stemspan-SFEM™ supplemented with human SCF, TPO, FL and PGE2. In the afternoon of day 3, 2.2 million CD34⁺ cells per sample were suspended in electroporation buffer.

The gRNAs were generated by *in vitro* transcription using a T7 polymerase. A 5' ARCA cap was added to the RNA simultaneous to transcription while a polyA tail was added after transcription to the 3' end of the RNA species by an *E. coli* polyA polymerase. After the gRNAs were *in vitro* transcribed and tailed, the quality and quantity of gRNAs were evaluated with the Bioanalyzer (Nanochip) to determine RNA concentration and by DSF assay performed as a measure of D10A Cas9 nickase RNP stability and as an indirect measure of gRNA quality.

In vitro transcribed capped and tailed guide gRNAs *HBB-8* and *HBB-15* were added at a 2:1 molar ratio (total gRNA : Cas9 protein) to 10µg RNP per million cells. The RNPs tested include the following: wild-type (WT) Cas9, endotoxin-free WT Cas9, N863A nickase,

D10A nickase. For each experimental sample, the D10A RNP/cell mixture was transferred to the electroporation cartridge, and the cells then electroporated using Program 2 (P2).

For all samples, the cells were collected from the cartridge and placed at 37°C for 20 minutes (recovery period). For the CB CD34⁺ cells that were contacted with 10µg D10A nickase RNP (per million cells), one sample was transferred to pre-warmed cytokine supplemented Stemspan-SFEM™ media and placed at 30°C for 2 hours (cold shock recovery) or placed directly into the same media at 37°C. For the cold shocked samples, the cells were transferred to the 37°C incubator after the 2-hour incubation period at 30°C. At 24, 48, and 72 hours after electroporation, the CB CD34⁺ cells were counted by trypan blue exclusion (cell survival) and divided into 3 portions for the following analyses: a) flow cytometry analysis for assessment of viability by co-staining with 7-Aminoactinomycin-D (7-AAD) and allophycocyanin (APC)-conjugated Annexin-V antibody (ebioscience); b) flow cytometry analysis for maintenance of HSC phenotype (after co-staining with phycoerythrin (PE)-conjugated anti-human CD34 antibody (BD Biosciences) and APC-conjugated CD133 (Miltenyi Biotech; c) hematopoietic colony forming cell (CFC) analysis by plating 800 cells in semi-solid methylcellulose based Methocult medium (StemCell Technologies H4435) that supports differentiation of erythroid and myeloid blood cell colonies from HSCs and serves as a surrogate assay to evaluate HSC multipotency and differentiation potential *ex vivo*; d) genomic DNA analysis for detection of editing at the *HBB* locus; and e) Western blot analysis of protein to evaluate the stability of Cas9 RNP in CD34⁺ HSCs. gDNA was extracted from the HSCs at 48 and 72 hours after electroporation and *HBB* locus-specific PCR reactions were performed. The purified PCR products were analyzed for insertions/deletions (indels) in T7E1 assays and by DNA sequencing of individual clones (PCR product was transformed and subcloned into TOPO-vector, individual colonies picked, and plasmid DNA containing individual PCR products were sequenced).

Electroporated CB CD34⁺ cells maintained a stem cell phenotype (*e.g.*, co-expression of CD34 and CD133, >90% CD34⁺CD133⁺) and as determined by flow cytometry analysis. The absolute number of viable CD34⁺ cells was maintained across most samples over a 72-hour culture period after electroporation (**Figure 34A**). Gene edited CD34⁺ cells maintained *ex vivo* hematopoietic activity and multipotency as indicated by their ability to give rise to erythroid (*e.g.*, CFU-E, or CFU-GEMM) and myeloid (*e.g.*, CFU-G, -M, or -GM) (**Figure 34B**).

Gene editing at the *HBB* locus was then evaluated at 72 hours after electroporation of WT Cas9, N863A, or D10A nickases co-delivered with 2 gRNAs (HBB-8 and HBB-15).

Briefly, gDNA was isolated from electroporated CD34⁺ cells at 72 hours after electroporation, and PCR amplification of a ~607bp fragment of the *HBB* locus (which captured both of the individual genomic locations that were targeted by gRNAs HBB-8 and HBB-15) was performed. After cleanup of the *HBB* PCR product with AMPURE beads, insertions/deletions (indels) at the targeted genomic location were evaluated by T7E1 assay and by DNA sequencing. For the CD34⁺ cells electroporated with WT Cas9 and endotoxin-free WT Cas9 the percentages of indels detected by T7E1 analysis at 72 hours was 59% and 51%, respectively (**Figure 35A**), which correlated with the indels detected by DNA sequencing (**Table 12**). CD34⁺ cells electroporation with N863A nickase and HBB-specific gRNA pair, had only 1% indels detected by T7E1 analysis. CD34⁺ HSCs electroporated with D10A nickase with and without cold shock supported gene editing at percentages of 39% and 48% indels detected by T7E1 analysis, respectively. In order to confirm that this low level of editing observed in CD34⁺ HSCs contacted with N863A nickase, was not due to the lack of N863A RNP contacting the cells, western blot analysis was performed (**Figure 35B**).

Cas9 protein was present in all electroporated samples (*e.g.*, cells that received WT Cas9, D10A nickase, and N863A nickase). For these samples, Cas9 protein was detected at 24 and 48 hours after electroporation, suggesting that the lack of N863A activity in the CD34⁺ HSCs was not due to the lack of protein.

Gene editing in CD34⁺ HSCs that contacted WT Cas9, endotoxin-free Cas9, and D10A nickase was 54-60%, based on DNA sequencing analysis (**Table 12, Figure 36A**). Stratification of DNA repair events through DNA sequencing analysis revealed that >90% of the gene editing events were deletions in CD34⁺ HSCs that contacted WT Cas9 and endotoxin-free WT Cas9 (insertions or combination of insertion and deletion comprised the remaining 3-6% of editing events) (**Figure 36A**). In contrast, gDNA from CD34⁺ HSCs that contacted D10A nickases had 3% HDR (*e.g.*, gene conversion), up to 75% insertions, and up to 22% deletions (**Figure 36B**).

In summary, the data in this Example indicate: 1) endotoxin removal does not negatively impact Cas9 functionality or CD34⁺ HSC cell viability, proliferation potential, or multipotency; 2) use of 10 µg D10A RNP supports 60% gene editing in CD34⁺ HSCs with HDR events (*e.g.*, gene conversion); and 3) after contacting CD34⁺ HSCs, WT Cas9 and nickase RNPs are detected for up to 48 hours, but is not detectable thereafter.

Table 12. Summary of gene editing results in CB CD34⁺ cells 72 hours after co-delivery of wild-type Cas9, N863A nickase, D10A nickase RNP and HBB-specific gRNA pair.

Cas9	Total μ g RNP/1E6 cells	Temperature of 2-hour recovery	Electroporation Program	%indels (T7E1)	%indels (sequencing)
WT	10	37°C	2	59	56
Endo-Free WT	10	37°C	2	51	60
N863A	10	37°C	2	1	ND
D10A	10	37°C	2	39	60
D10A	10	30°C	2	48	54

Example 10: Gene Conversion Modulation

The CRISPR/Cas9 system was used to target the human *HBB* gene in the region of the sickle cell anemia-causing mutation.

The nature of the targeted break affects the frequency of different DNA repair outcomes. Blunt double-strand breaks, single-strand nicks, and dual-nicks in which the nicks are placed on opposite strands and leave either 3' or 5' overhangs of varying lengths, were introduced by utilizing the wild type Cas9 nuclease, as well as two different Cas9 nickases. Several different DNA repair outcomes including, *e.g.*, indel mutations resulting from non-homologous end-joining (NHEJ) and homology-dependent repair (HDR) using the closely related *HBD* gene as an endogenous template, were characterized. The frequency of these various repair outcomes under different conditions offers insight into the mechanisms of DNA repair and how it is impacted by the nature of the DNA break. Here, we examined how different repair pathways affect the frequency of gene conversion (HDR from the closely related HBD gene as an endogenous template). Specifically, we evaluated the effect of down-regulation of the alternative NHEJ (alt-NHEJ) and HDR pathways on the gene conversion rate in the presence of dual-nicks.

In this study, different gRNAs targeting the *HBB* region that surround the nucleotides encoding the amino acid most commonly mutated in sickle cell disease were tested in 293T cells with wild type Cas9. The gRNAs that induced similar high rates of NHEJ and had PAMs facing in opposite orientations were selected to test as pairs with Cas9 D10A (which would generate a 5' overhang) and Cas9 N863A nickases (which would generate a 3' overhang).

As a first example of enzymes that modulate gene conversion frequency, we selected the polymerase Pol theta, which is a central player in the alt-NHEJ pathway. In this example, U2OS cells were electroporated with 200 ng of each gRNA (8 and 15), 750 ng of plasmid that encodes wild type Cas9 or mutant Cas9 (D10A or N863A) and 30 pmol of scrambled siRNA as a negative control or 30 pmol of siRNA against Pol Theta. Cells were collected 6 days after electroporation and genomic DNA was extracted. PCR amplification of the *HBB* locus was performed and subcloned into a Topo Blunt Vector. For each condition in each experiment 96 colonies were sequenced with Sanger sequencing.

As shown in **Fig. 37**, down regulation of Pol theta in the context of the D10A Cas9 does not affect the gene editing profile. In contrast, Pol theta down-regulation in the context of the N863A Cas9 nuclease leads to a strong reduction of the insertion frequency and an increase in the gene conversion rate. These data suggest that the 3' protruding ends generated by the N863A Cas9 nuclease are substrates for processing by Pol Theta, resulting in a high accumulation of insertions. Upon down-regulation of Pol Theta, the 3' protruding DNA ends are available for engaging the gene conversion pathway (**Fig. 37**).

Next, we dissected the genetic requirements of the gene conversion pathway in the context of different Cas9 nucleases (**Fig. 38**), using BRCA2 and RAD51 – central players of canonical HR – as an example. U2OS cell were electroporated with 200 ng of each gRNA (8 and 15), 750 ng of plasmid that encodes wild type Cas9 or mutant Cas9 (D10A or N863A) and 30 pmol of scrambled siRNA as a negative control or 30 pmol of siRNA against BRCA2 or RAD51. As shown in **Fig. 38**, the majority of the gene conversion events are mediated by BRCA2 and RAD51. These data are consistent with the notion that gene conversion using the HBD gene as a donor in cis is an event mediated by canonical HR.

In summary, different DNA ends engage different repair pathways. **Fig. 39** depicts a model of the genetic requirements of the gene conversion pathway in the context of different Cas9 nucleases (wild type, D10A or N863A). Use of a Cas9 nickase with two gRNAs generates either 3' or 5' overhangs, which are most likely not suitable substrates to be repaired by canonical NHEJ, but can be repaired by alternative pathways. The 3' end is predominantly a substrate for repair by the Alt NHEJ. The key player in the pathway is Pol Theta, which refills the ends, thereby generating insertions deriving from the overhangs.

The 5' protruding end is mostly repaired through gene conversion, in which the *HBB* gene is repaired using the homologous *HBD* locus as a template. This pathway is dependent

on BRCA2 and RAD51, which suggests the action of the canonical HR pathway.

In conclusion, these data support a therapeutic approach in which correction of the sickle-cell mutation is efficiently mediated through HDR from the endogenous *HBD* gene, and it elucidates strategies of how to promote or suppress this event.

5

Example 11: CRISPR/Cas9 RNP supports gene editing of T lymphocytes derived from a Sickle Cell Disease Patient *Ex Vivo*

To determine whether Cas9 RNP mediated gene editing could correct the sickle cell disease (SCD) mutation, peripheral blood mononuclear cells (MNCs) from a SCD patient were obtained (Conversant Bio). To expand the T lymphocyte fraction from bulk MNCs, MNCs were thawed into T cell culture media (X-VIVO 15 supplemented with human AB serum, N-acetylcysteine, Glutamax (L-glutamine), and human cytokines (IL-2, IL-7, IL-15)). The T lymphocyte fraction was activated with CD3/CD28 immunomagnetic beads. By day 3 of MNC culture in T cell media, the cell population was 72% CD3⁺. After 7 days, a population comprising >98% CD3⁺ T lymphocytes was obtained, most of which were also CD4⁺ (**Fig. 40A**). Genomic DNA (gDNA) was extracted from the CD3⁺ T lymphocytes and sequenced to confirm the presence of the SCD mutation.

After confirmation of the sickle mutation in the CD3⁺ T lymphocytes, gRNA was designed to target the SCD mutation (HBB-8-sickle gRNA targeting sequence: GUAACGGCAGACUUCUCCAC (SEQ ID NO:16318), underline indicates SCD mutation). The gRNA was *in vitro* transcribed from a DNA PCR product template. The HBB-8-sickle was modified to contain a 5' ARCA cap and a 3' polyA tail. Another gRNA HBB-15 (SEQ ID NO:387) was also *in vitro* transcribed and was modified to contain a 5' ARCA cap and a 3' polyA tail. The gRNAs were complexed to *S. pyogenes* Cas9 D10A nickase protein for RNP production.

To evaluate Cas RNP editing in SCD patient cells, 2 x 10⁶ cells were electroporated with Cas9 D10A nickase RNP, in which the nickase was complexed to 2 different gRNA pairs to target the sequence specific to the *HBB* gDNA of SCD patients (HBB-8-sickle and HBB-15). Dual D10A nickase RNP (10μg per 1 x 10⁶ cells) was co-electroporated with a single strand oligonucleotide donor (SSODN, 250 pmoles per 1 x 10⁶ cells). After electroporation, cells were plated onto T cell media with or without CD3/CD28 immunomagnetic beads to determine whether restimulation of T cells after electroporation would improve survival and viability of the gene-edited cells. Re-plating the edited cells into culture with CD3/CD28 beads for cell reactivation improved the total viability of the

electroporated cells in comparison to cells plated in T cell media without CD3/CD28 beads (% viability determined by flow cytometry analysis after staining with apoptosis stains 7-AAD and Annexin V; **Fig. 40B**). Gene editing was evaluated by DNA sequencing. The gRNA combination HBB-8-sickle and HBB-15 (each complexed to Cas9 D10A nickase) supported 48% total editing, as detected by T7E1 endonuclease assay analysis of the *HBB* PCR product (**Fig. 40C**). Ten percent higher editing was detected in gDNA from the T lymphocytes that were reactivated with CD3/CD28 beads after electroporation, compared to cells cultured in T cell medial alone. These data show that Cas9 RNP targeting the SCD mutation supports gene editing in SCD patient cells.

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)
- 5 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)
- 10 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)
- 15 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)
- 20 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)
- Marteijn *et al. Nat. Rev. Mol. Cell. Biol.* 15(7):465-481 (2014)
- Mayle *et al. Science* 349: 742-47 (2015)
- 25 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)
- Saleh-Gohari *et al. Mol. Cell. Biol.* 25: 7158-69 (2005)
- Schlacher *et al. Cancer Cell* 22: 106-16 (2012)
- 30 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)

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. A method of modifying an endogenous target gene in a cell, the method comprising:
contacting the cell with a first gRNA molecule, a first enzymatically active Cas9 (eaCas9) molecule, a second gRNA molecule, and a second eaCas9 molecule;
wherein the first gRNA molecule and the first eaCas9 molecule associate with the target gene and generate a first single strand cleavage event on a first strand of the target gene;
wherein the second gRNA molecule and the second eaCas9 molecule associate with the target gene and generate a second single strand cleavage event on a second strand of the target gene, 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 gene are repaired using an endogenous homologous region, thereby modifying the endogenous target gene in the cell.
2. The method of claim 1, wherein, after repair of the first overhang and the second overhang, the target gene comprises the sequence of the endogenous homologous region.
3. The method of any one of claims 1-2, wherein the cell is not contacted with an exogenous nucleic acid homologous to the target gene.
4. The method of any one of claims 1-3, wherein the first overhang and the second overhang are repaired by homology directed repair (HDR).
5. The method of claim 4, wherein the HDR is gene conversion.
6. The method of any one of claims 1-5, wherein the first overhang is a 5' overhang, and the second overhang is a 5' overhang.
7. The method of claim 6, wherein the first overhang and the second overhang undergo processing, exposing a first 3' overhang and a second 3' overhang.

8. The method of any one of claims 1-7, wherein the method is used to correct a mutation in the endogenous target gene, and wherein the mutation is located between the first single strand break and the second single strand break.
9. The method of any one of claims 1-8, wherein the method is used to correct a mutation in the endogenous target gene, and wherein the mutation is located within fewer than 50 nucleotides of the first single strand break, or within fewer than 50 nucleotides of the second single strand break.
10. The method of any one of claims 1-9, wherein the endogenous target gene has at least 80% sequence homology with the endogenous homologous region.
11. The method any of claims 1-10, wherein the method is used to correct a mutation in the endogenous target gene, and wherein the endogenous homologous region comprises a domain which is homologous to a domain in the endogenous target gene but which does not comprise the mutation.
12. The method of any one of claims 1-11, wherein the endogenous homologous region comprises at least one of: a control region which is homologous to a control region of the target gene; a coding region which is homologous to a coding region of the target gene; a non-coding region which is homologous to a non-coding region of the target gene; an intron which is homologous to an intron of the target gene; or an exon which is homologous to an exon of the target gene.
13. The method of any one of claims 1-12, wherein the first eaCas9 molecule is a first nickase molecule and the second eaCas9 molecule is a second nickase molecule.
14. The method of claim 13, wherein each nickase molecule forms a single strand break in the target gene.
15. The method of any one of claims 1-14, wherein the first eaCas9 molecule and the second eaCas9 molecule are the same species of eaCas9 molecule.

16. The method of claim 15, wherein the eaCas9 molecule comprises HNH-like domain cleavage activity but has no N-terminal RuvC-like domain cleavage activity.
17. The method of claim 15, wherein the eaCas9 molecule is an HNH-like domain nickase.
18. The method of any one of claims 15-17, wherein the eaCas9 molecule comprises a mutation at an amino acid position corresponding to amino acid position D10 of *Streptococcus pyogenes* Cas9.
19. The method of any one of claims 1-18, wherein the target gene is an *HBB* gene.
20. The method of claim 19, wherein the first gRNA molecule is a gRNA molecule comprising SEQ ID NO:387, and wherein the second gRNA molecule is a gRNA molecule comprising SEQ ID NO:16318.
21. The method of claim 19, wherein the first gRNA molecule is a gRNA molecule comprising any one of SEQ ID NOs: 387-485, 6803-6871, or 16010-16256, and wherein the second gRNA molecule is a gRNA molecule comprising any one of SEQ ID NOs: 387-485, 6803-6871, or 16010-16256.
22. The method of claim 19, wherein the cell is in, or is taken from, a patient having sickle cell disease or beta thalassemia.
23. The method of claim 19, wherein the endogenous homologous region comprises a region of an endogenous *HDB* gene.
24. The method of any one of claims 1-18, wherein the target gene is an *SMN1* gene.
25. The method of claim 24, wherein the cell is in, or is taken from, a patient having spinal muscular atrophy.
26. The method of claim 24, wherein the endogenous homologous region includes a region of an endogenous *SMN2* gene.

27. The method of any one of claims 1-18, wherein the target gene is an *NCF1* (p47-PHOX) gene.
28. The method of claim 27, wherein the cell is in, or is taken from, a patient having chronic granulomatous disease.
29. The method of claim 27, wherein the endogenous homologous region includes a region of a p47-PHOX pseudogene.
30. The method of any one of claims 1-29, wherein the cell is a population of cells, and wherein the first overhang and the second overhang in the target gene are repaired by gene conversion in about 12% to about 45% of the cells in the population of cells.
31. The method of any one of claims 1-29, wherein the cell is a population of cells, and wherein the first overhang and the second overhang in the target gene are repaired by non-homologous end joining (NHEJ) in less than 40% of the cells in the population of cells.
32. The method of claim 31, wherein the NHEJ results in a deletion in the target gene.
33. The method of claim 31, wherein the NHEJ results in an insertion of the target gene.
34. The method of any one of claims 1-33, wherein the cell is selected from the group consisting of a blood cell, a neuronal cell, an immune cell, a muscle cell, a stem cell, a progenitor cell, and a diseased cell.
35. The method of any one of claims 1-34, wherein the cell is a mammalian cell.
36. The method of any one of claims 1-35, wherein the cell is a human cell.
37. A cell altered by the method of any one of claims 1-36.
38. A pharmaceutical composition comprising the cell of claim 37.

39. A composition, comprising.
- a first non-naturally occurring gRNA molecule;
 - a first non-naturally occurring enzymatically active Cas9 (eaCas9) molecule
 - a second non-naturally occurring gRNA molecule; and
 - a second non-naturally occurring eaCas9 molecule;
- wherein the first gRNA molecule and the first eaCas9 molecule are designed to associate with a target gene and generate a first single strand cleavage event on a first strand of the target gene;
- wherein the second gRNA molecule and the second eaCas9 molecule are designed to associate with the target gene and generate a second single strand cleavage event on a second strand of the target gene, thereby forming a double strand break in the target gene having a first overhang and a second overhang; and
- wherein the first gRNA molecule, the first eaCas9 molecule, the second gRNA molecule and the second eaCas9 molecule are designed such that the first overhang and the second overhang in the target gene are repaired by gene conversion using an endogenous homologous region.
40. The composition of claim 39, wherein the first non-naturally occurring eaCas9 molecule is an HNH-like domain nickase, and wherein the second non-naturally occurring eaCas9 molecule is an HNH-like domain nickase.
41. The composition of claim 39 or claim 40, for use as a medicament.
42. The composition of claim 39 or claim 40, for use in the treatment of sickle cell disease or beta thalassemia.
43. The composition of claim 39 or claim 40, for use in the treatment of spinal muscular atrophy.
44. The composition of claim 39 or claim 40, for use in the treatment of chronic granulomatous disease.
45. A method of modifying a target region of a target gene in a mammalian cell, the method comprising:

generating, within the cell, a first single strand break on a first strand of the target gene and second single strand break on a second strand of the target gene, thereby forming a double strand break in the target gene having a first 5' overhang and a second 5' overhang;

wherein the target region of the target gene is located

- (a) between the first single strand break and the second single strand break,
- (b) within fewer than 50 nucleotides of the first single strand break, or
- (c) within fewer than 50 nucleotides of the second single strand break;

wherein the double strand break is repaired by gene conversion,

thereby modifying the target region of the target gene in the mammalian cell.

46. The method of claim 45, wherein the step of generating the first single strand break and the second single strand break comprises contacting the cell with at least one eaCas9 molecule, a first gRNA molecule, and a second gRNA molecule.

47. The method of claim 45, wherein the first gRNA molecule and the at least one eaCas9 molecule associate with the target gene and generate the first single strand break, and wherein the second gRNA molecule and the at least one eaCas9 molecule associate with the target gene and generate the second single strand break.

48. The method of any one of claims 45-47, wherein the double strand break is repaired by gene conversion using an endogenous homologous region.

49. The method of claim 48, wherein the endogenous homologous region is located within a gene cluster comprising the target gene.

50. The method of claim 48 or claim 49, wherein the target gene is an *HBB* gene, and wherein the endogenous homologous region is a region of an *HBD* gene.

51. The method of claim 50, wherein the first gRNA molecule comprises a targeting sequence comprising SEQ ID NO:387 and the second gRNA molecule comprises a targeting sequence comprising SEQ ID NO:16318.

52. The method of claim 47, wherein the at least one eaCas9 molecule is at least one eaCas9 nickase molecule.

53. The method of claim 52, wherein the at least one eaCas9 nickase molecule is at least one HNH-type nickase molecule.

54. A method of increasing the percentage of cells in a population of cells that modify a target region of a target gene by gene conversion using an endogenous homologous region, the method comprising

contacting the population of cells with a first gRNA molecule, a first enzymatically active Cas9 (eaCas9) molecule, a second gRNA molecule, and a second eaCas9 molecule;

wherein the first gRNA molecule and the first eaCas9 molecule associate with the target gene and generate a first single strand cleavage event on a first strand of the target gene;

wherein the second gRNA molecule and the second eaCas9 molecule associate with the target gene and generate a second single strand cleavage event on a second strand of the target gene, thereby forming a double strand break in the target gene having a first 5' overhang and a second 5' overhang; and

wherein the first 5' overhang and the second 5' overhang in the target gene are repaired by gene conversion using the endogenous homologous region,

thereby increasing the percentage of cells in the population of cells that modify the target region of the target gene by gene conversion using the endogenous homologous region.

55. The method of claim 54, wherein the first 5' overhang and the second 5' overhang in the target gene are repaired by gene conversion in about 12% to about 45% of the cells in the population of cells.

56. The method of claim 54, wherein the percentage of cells in the population of cells that modify the target region of the target gene by gene conversion using the endogenous homologous region is increased, as compared to a percentage of cells in a population of cells which would modify the target region of the target gene by gene conversion using the endogenous homologous region in the absence of the contacting.

57. The method of claim 54, wherein the percentage of cells in the population of cells that modify the target region of the target gene by gene conversion using the endogenous homologous region is increased, as compared to a percentage of cells in a population of cells which would modify the target region of the target gene by gene conversion using the endogenous homologous region after contacting the population of cells with a first gRNA molecule, a first enzymatically active Cas9 (eaCas9) molecule, a second gRNA molecule, and a second eaCas9 molecule which would generate a first 3' overhang and a second 3' overhang in the target gene.

58. The method of claim 54, wherein the percentage of cells in the population of cells that modify the target region of the target gene by gene conversion using the endogenous homologous region is increased, as compared to a percentage of cells in a population of cells which would modify the target region of the target gene by gene conversion using the endogenous homologous region after contacting the population of cells with a first gRNA molecule, a first enzymatically active Cas9 (eaCas9) molecule, a second gRNA molecule, and a second eaCas9 molecule which would generate a double stranded blunt end break in the target gene.

1/49

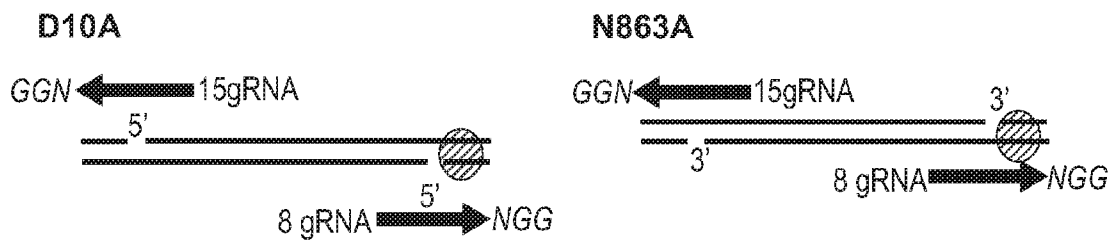


FIG. 1

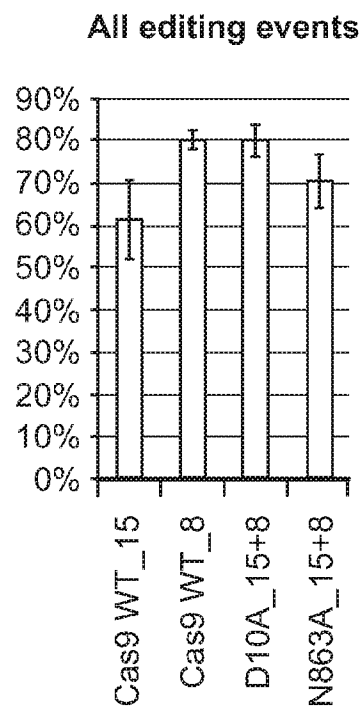


FIG. 2

2/49

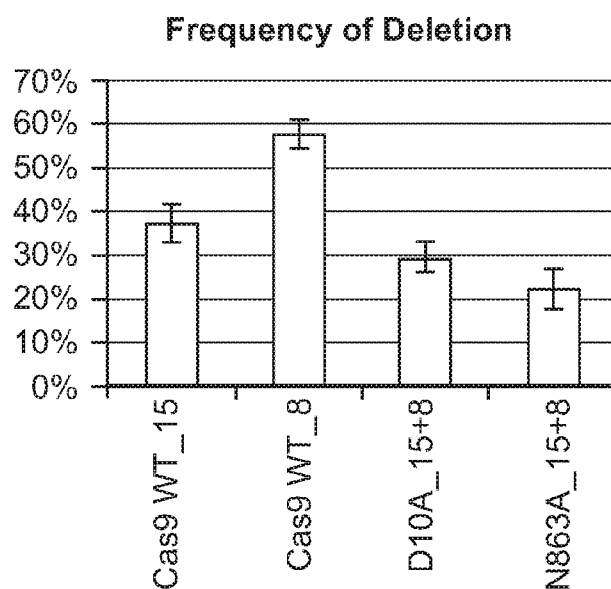


FIG. 3A

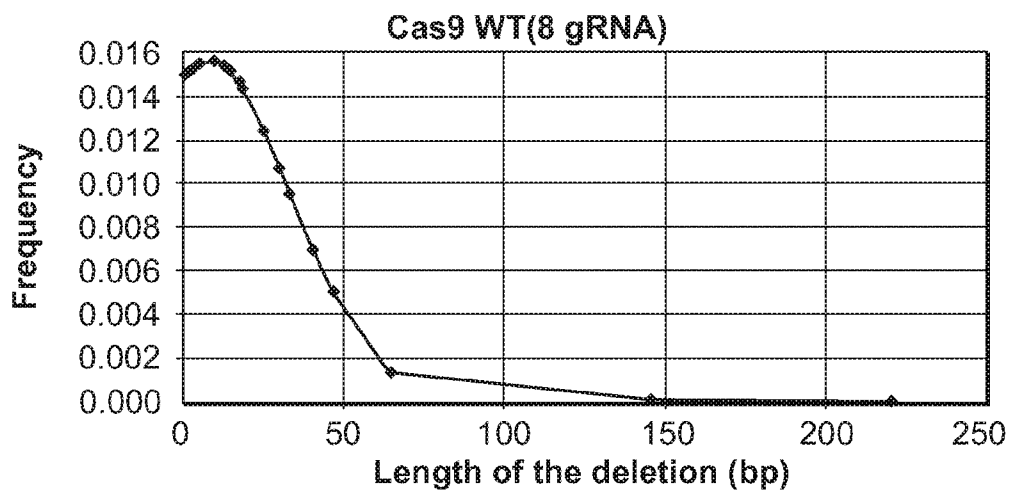


FIG. 3B

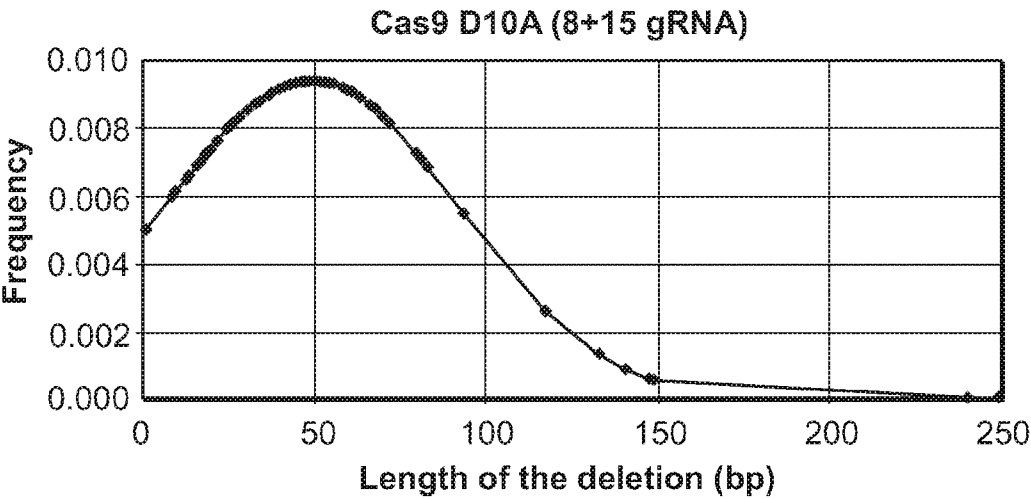


FIG. 3C

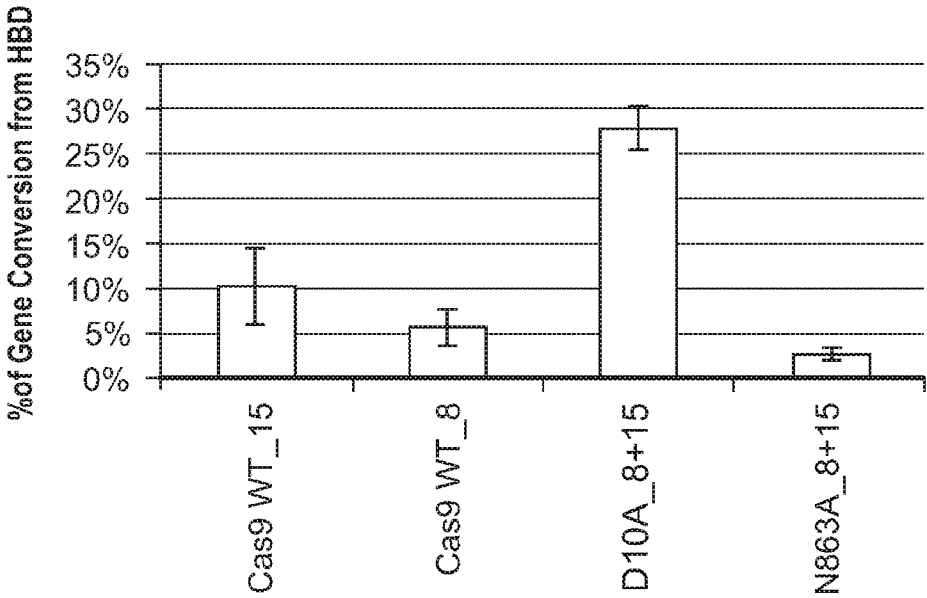


FIG. 4A

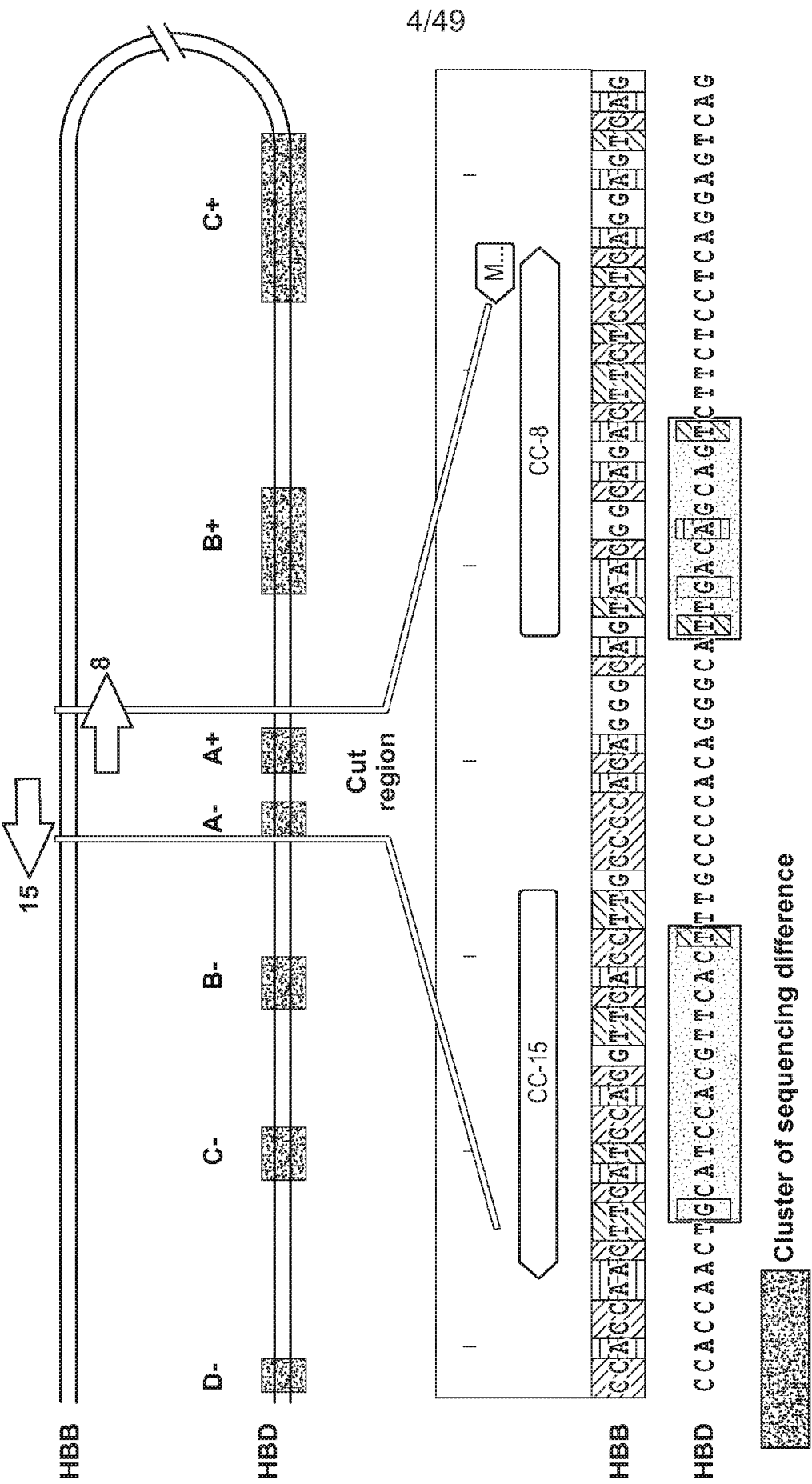
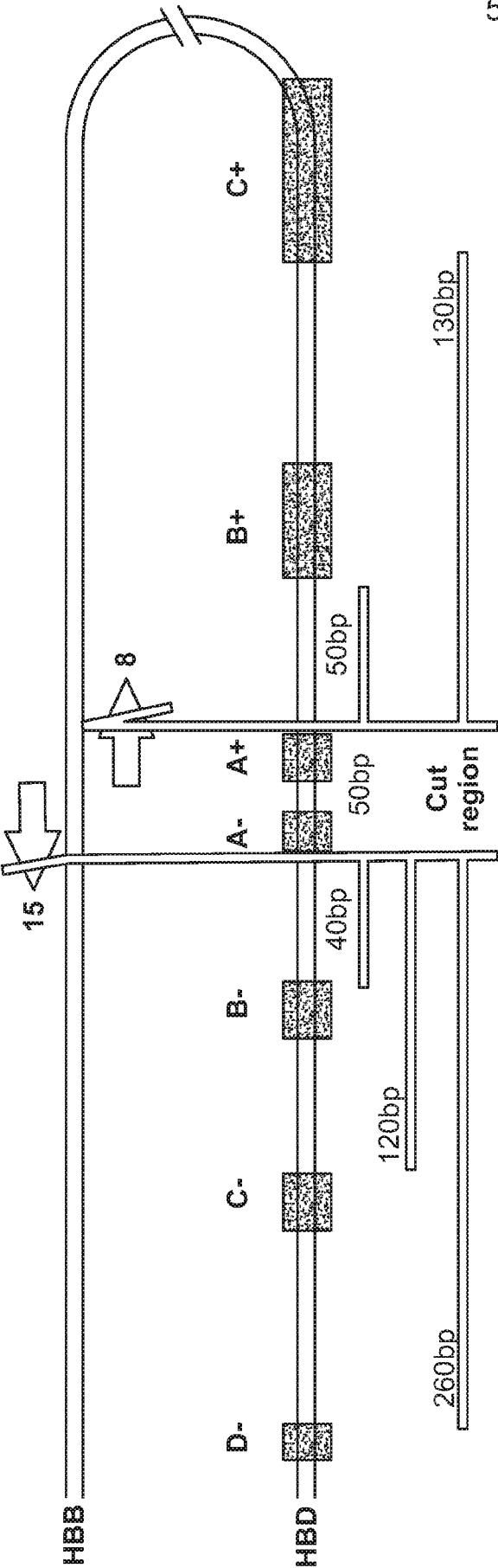


FIG. 4B



Length of the Gene Conversion

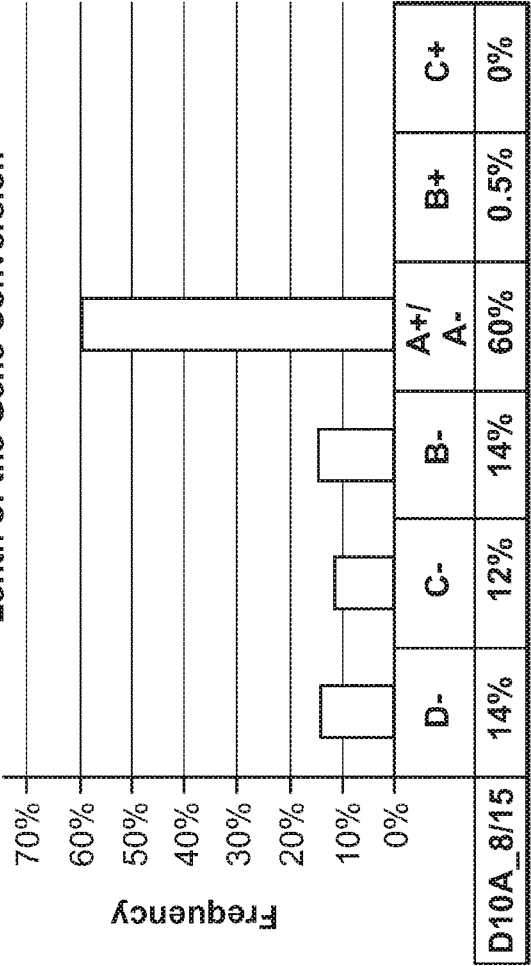
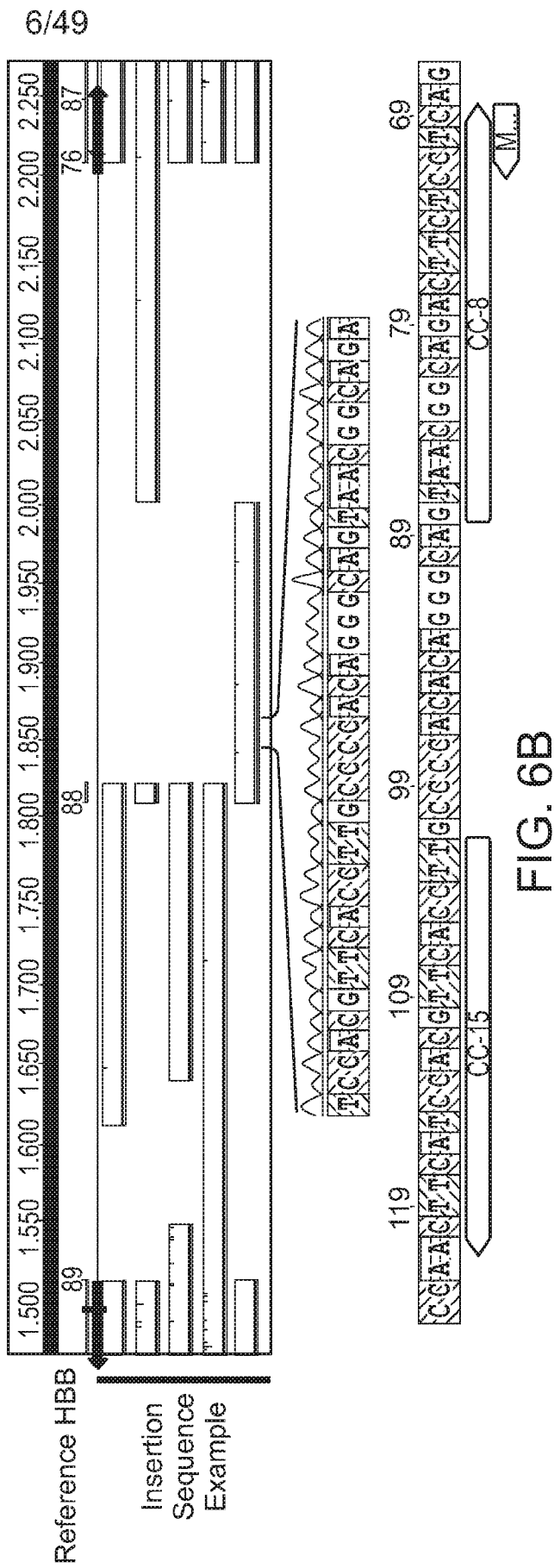
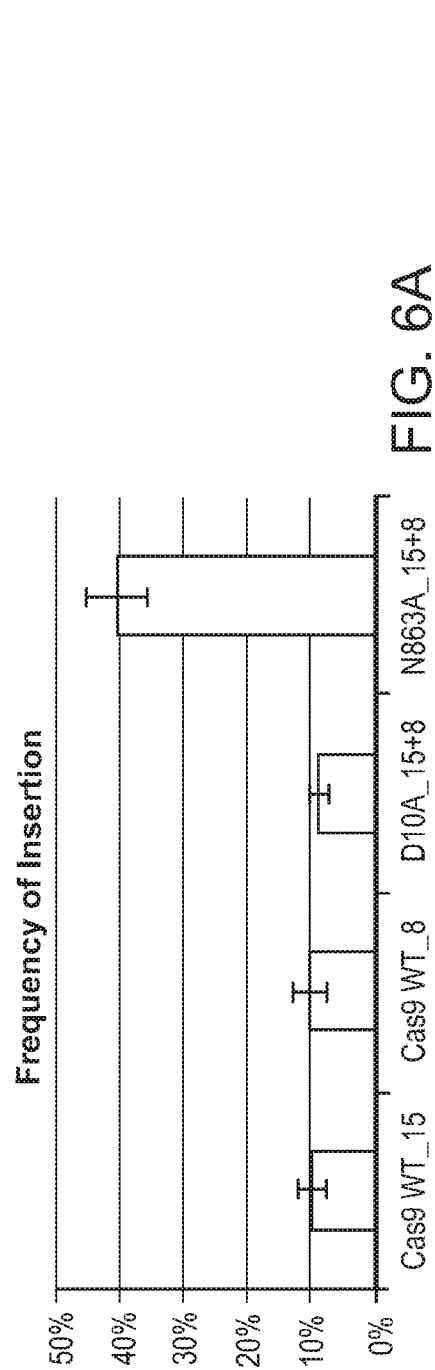


FIG. 5



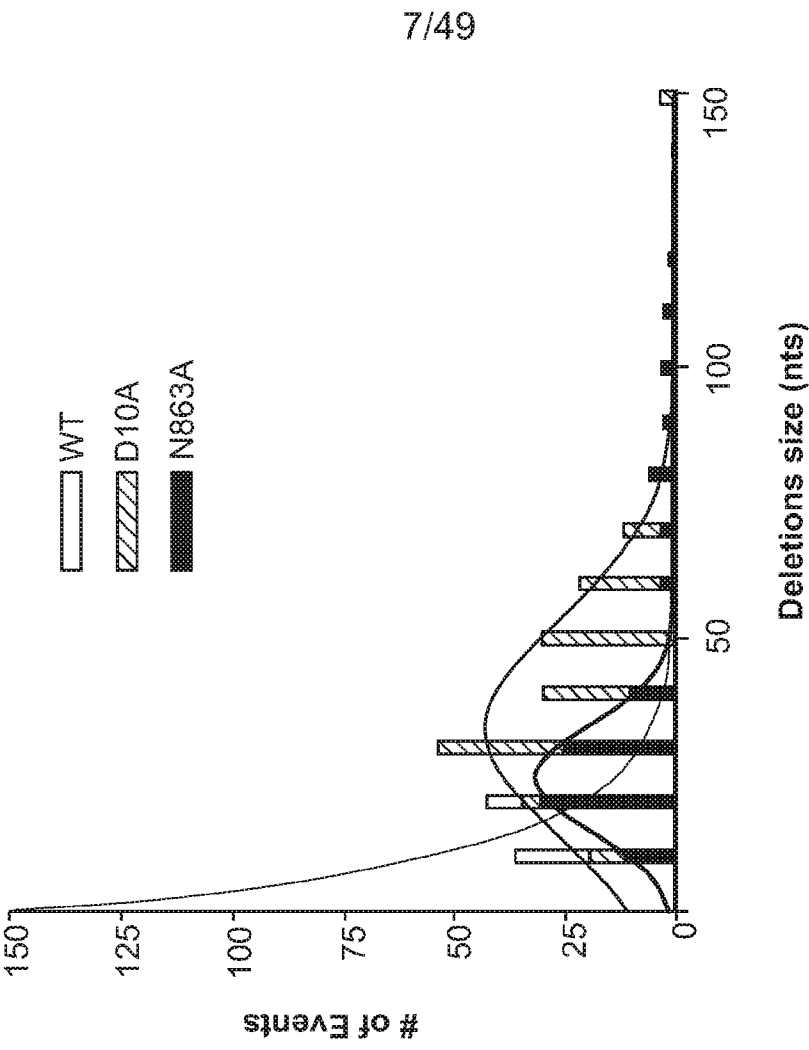


FIG. 7B

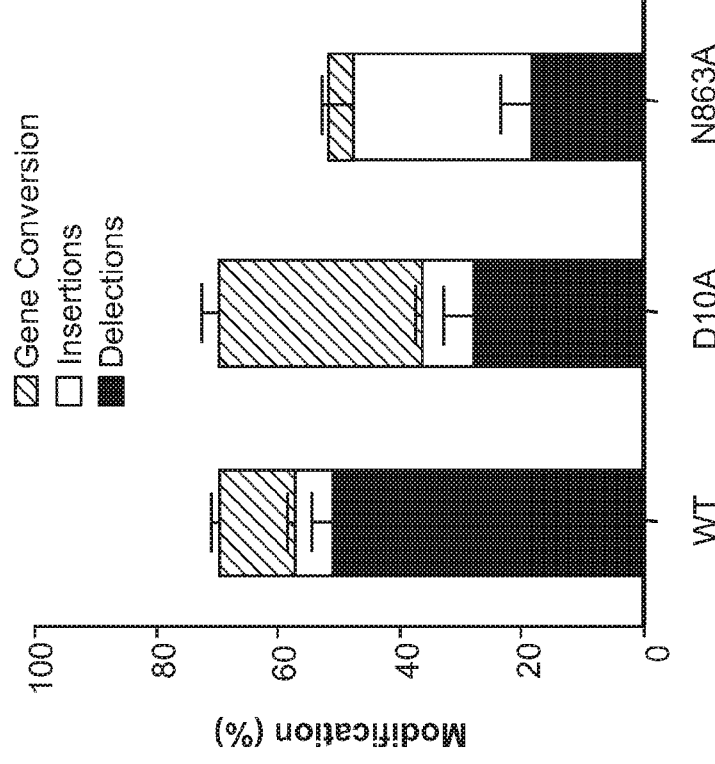
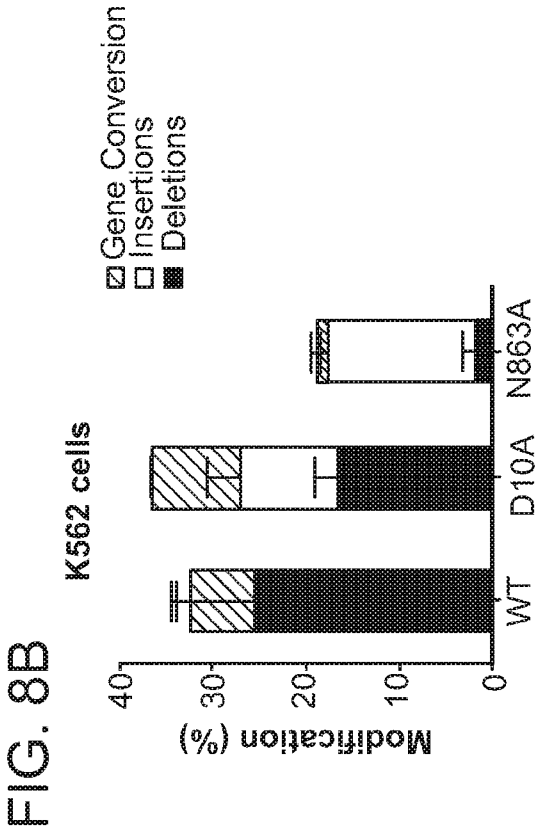
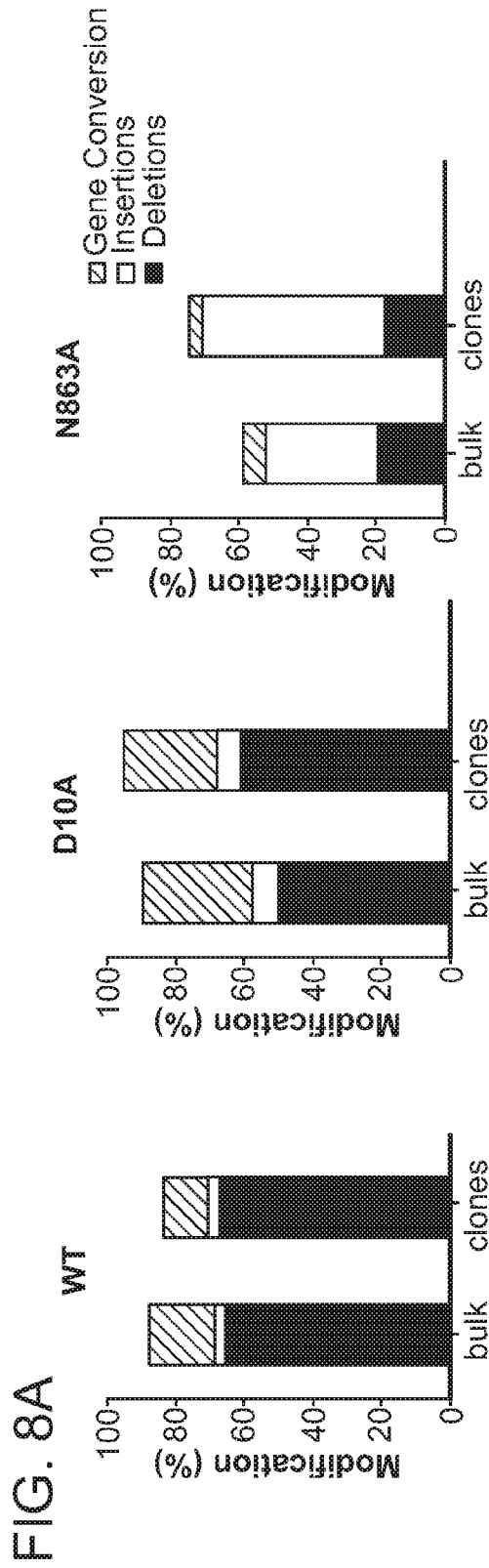


FIG. 7A



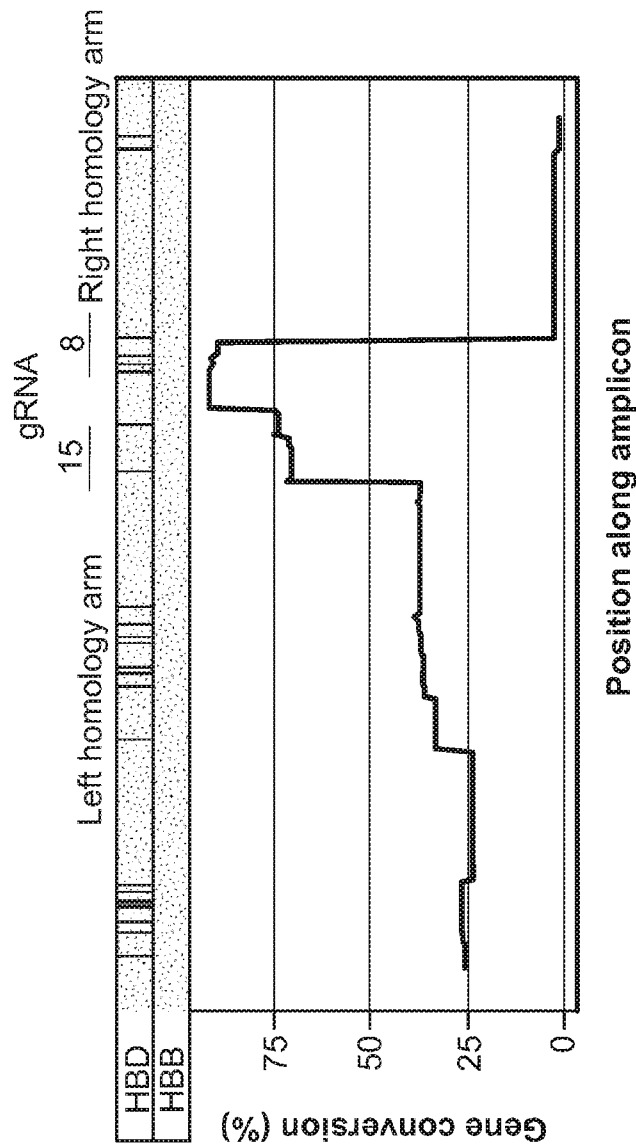


FIG. 9

10/49

FIG 10A

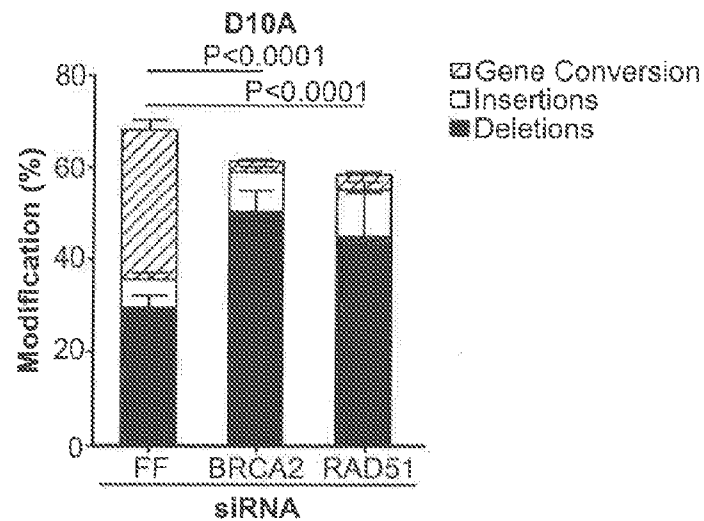
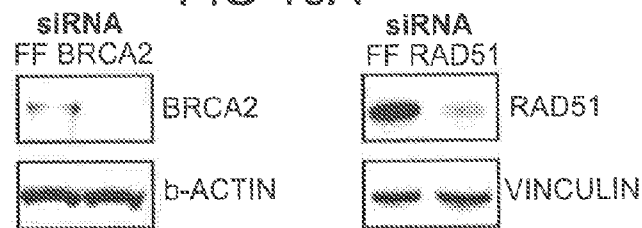


FIG. 10B

FIG. 10C

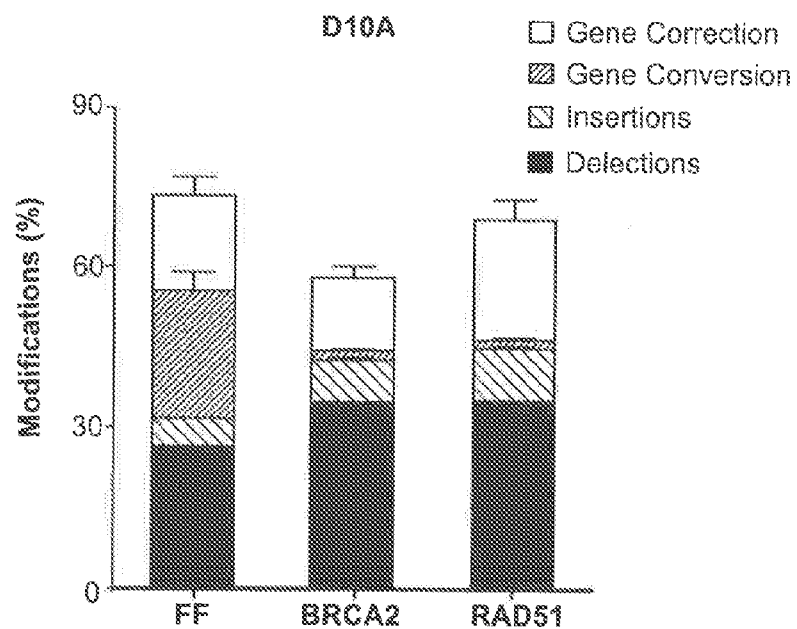
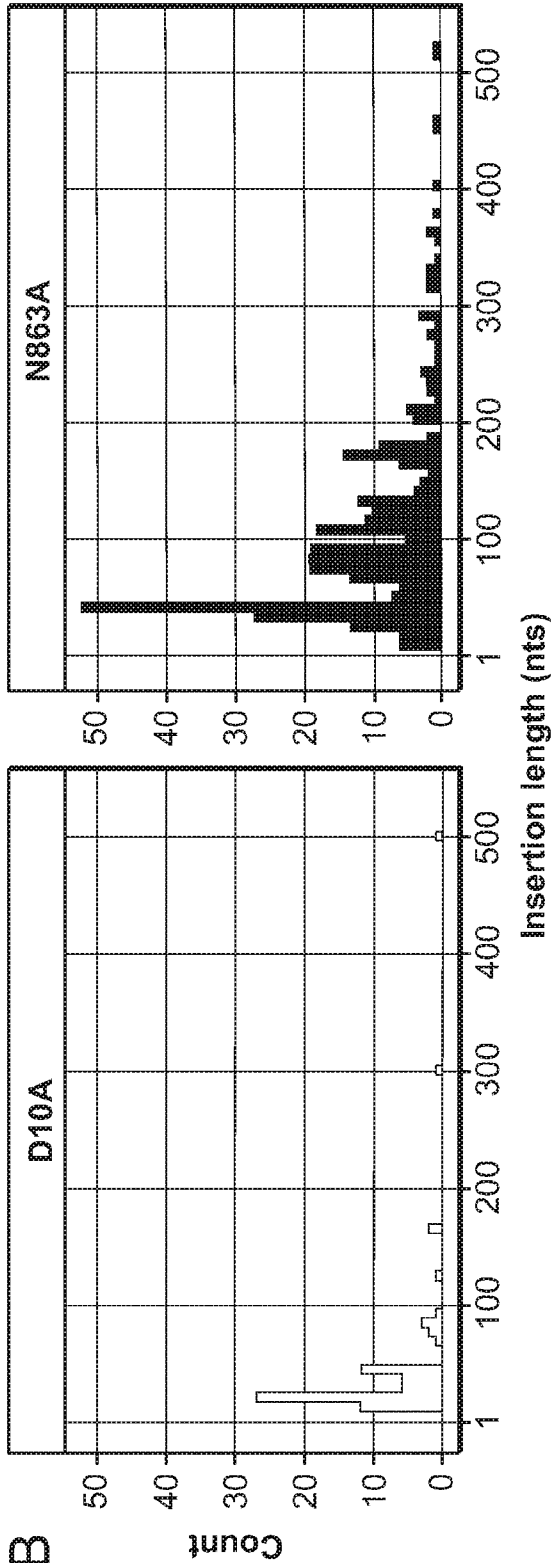


FIG. 11A

	D10A	N863A	p-value (D10A v. N863A)
# Insertions deriving from overhang (% of all)	75 (71.45%)	325 (98.5%)	1.022x10 ⁻¹⁵
# Insertions with full overhang repetition (% of all)	6 (5.7%)	66 (20%)	0.0004264
# Insertions analyzed	105	330	

p value: Fisher's exact test

FIG. 11B



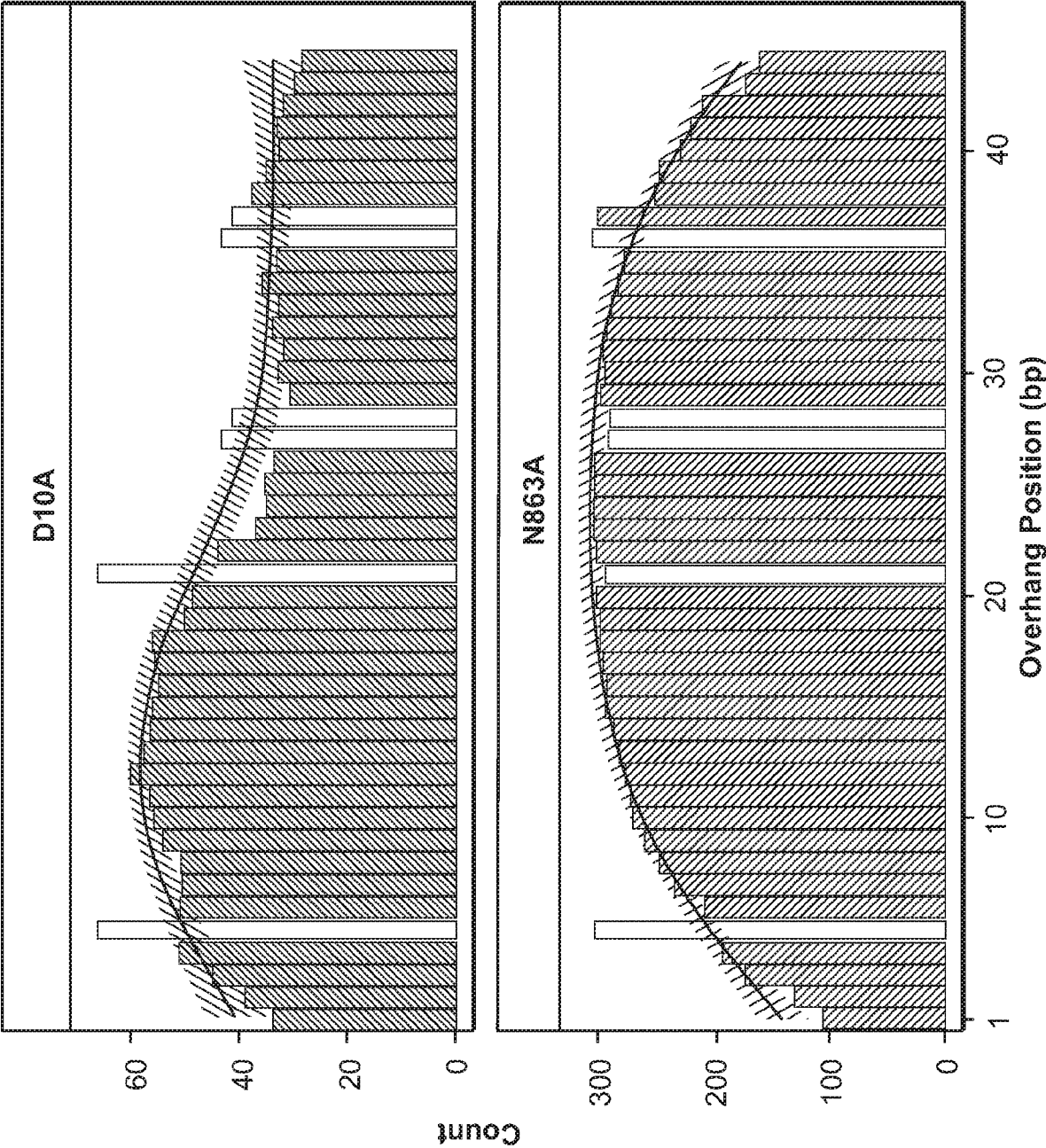
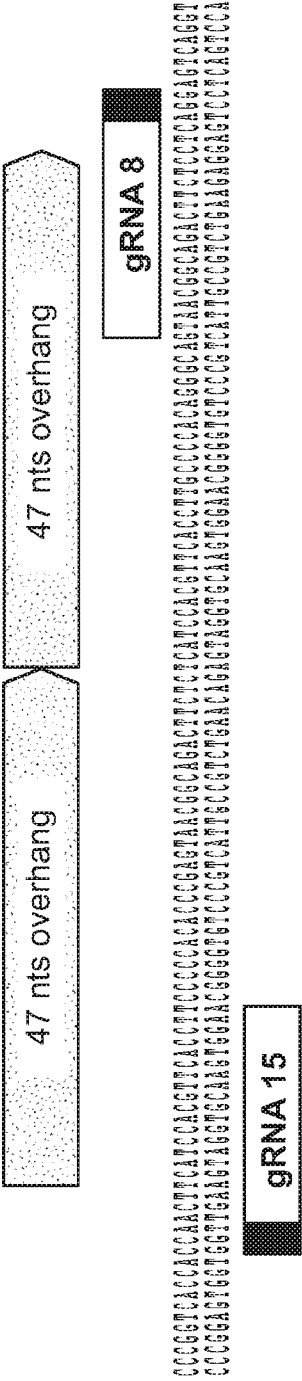


FIG. 12A

D10A Overhang Example



N863A Overhang Example

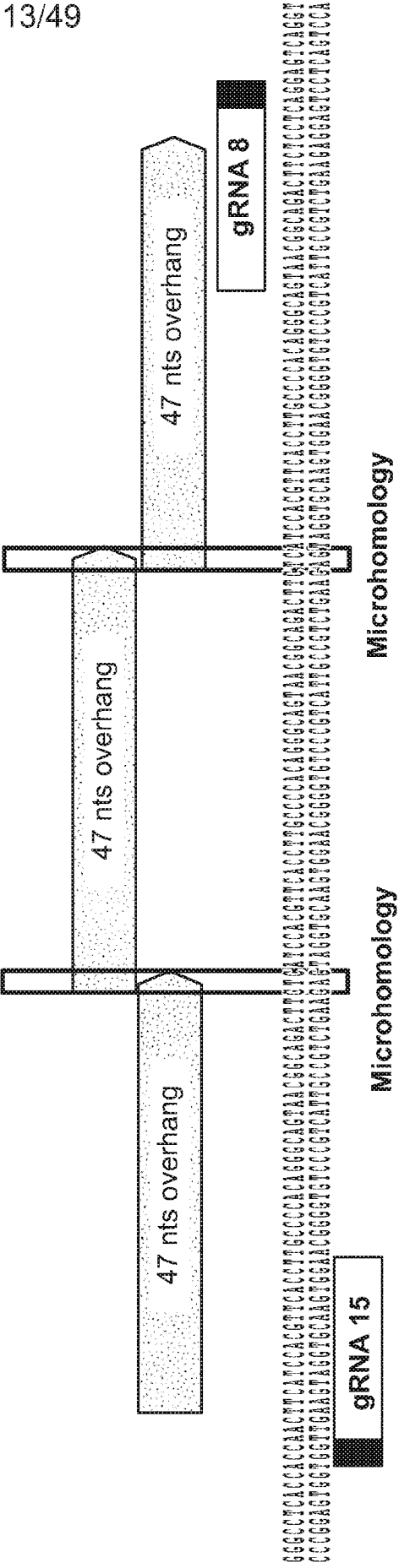


FIG. 12B

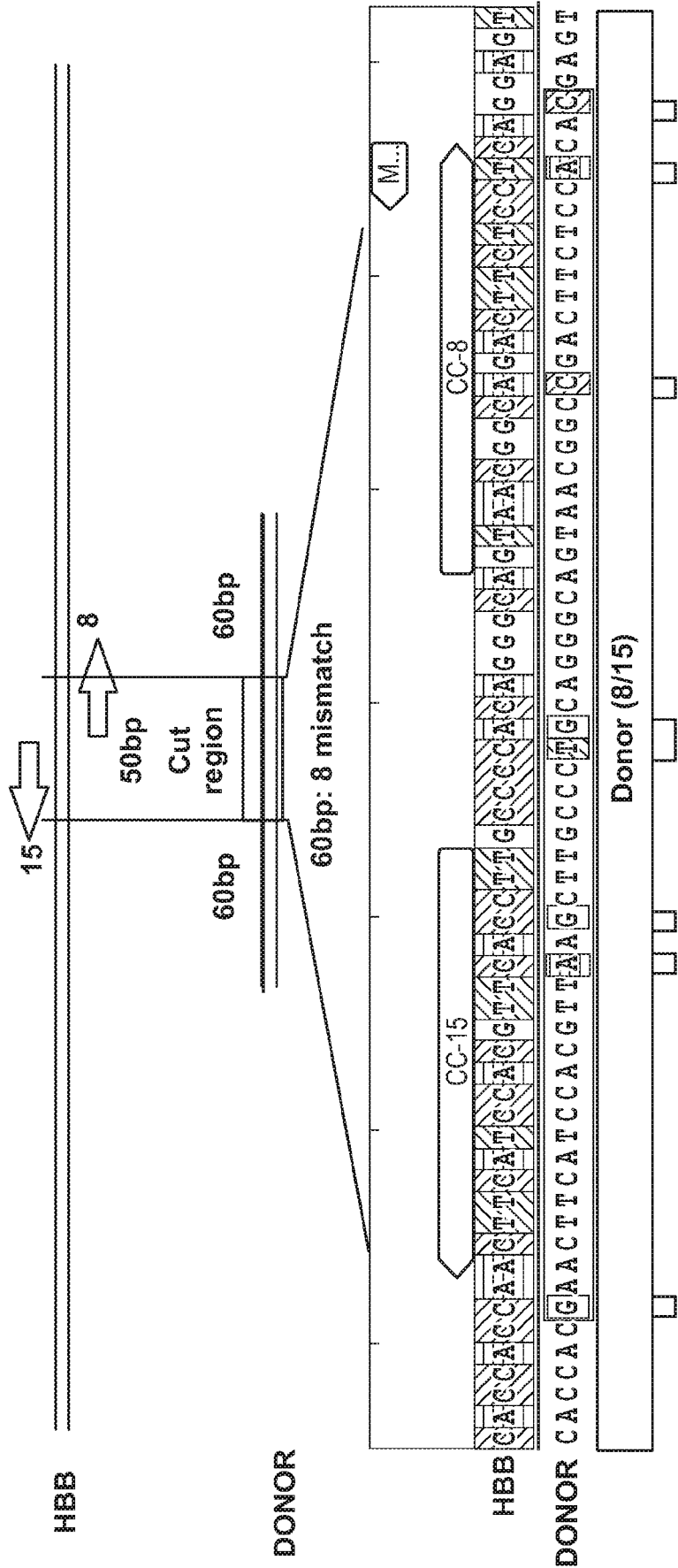


FIG. 13A

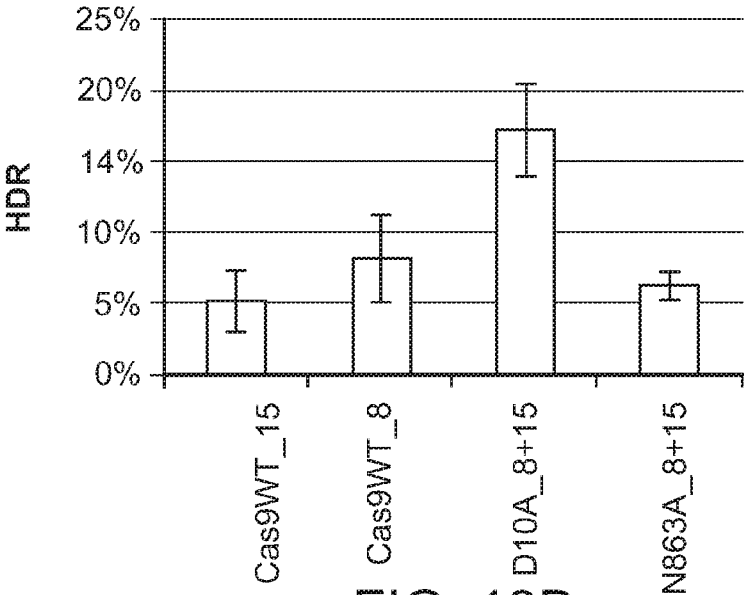


FIG. 13B

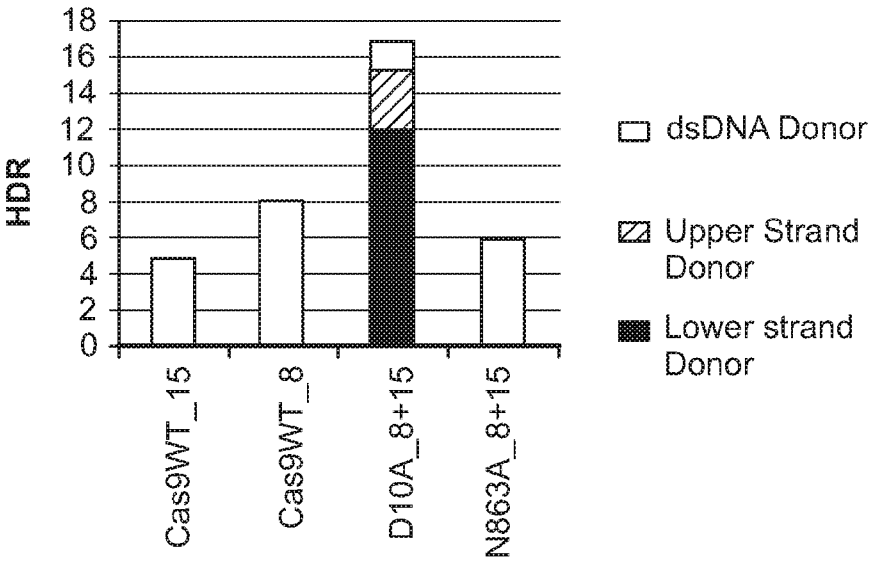


FIG. 13C

16/49

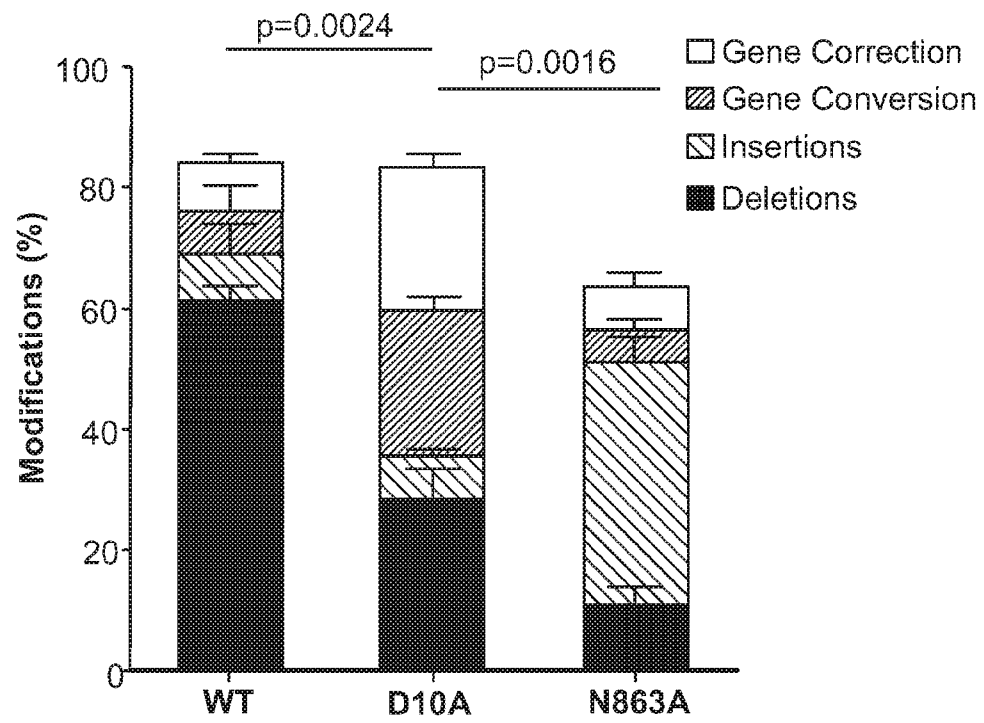


FIG. 14

17/49

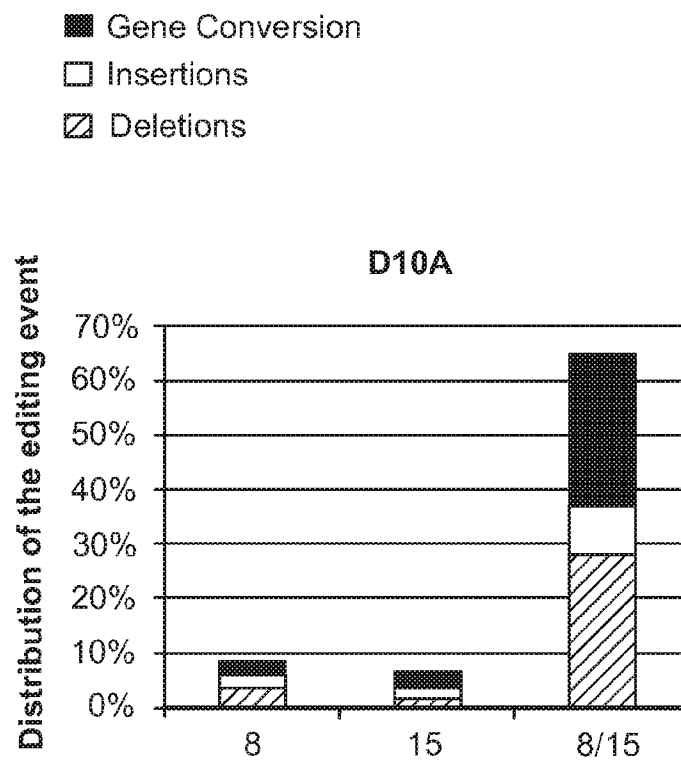


FIG. 15

FIG. 16A

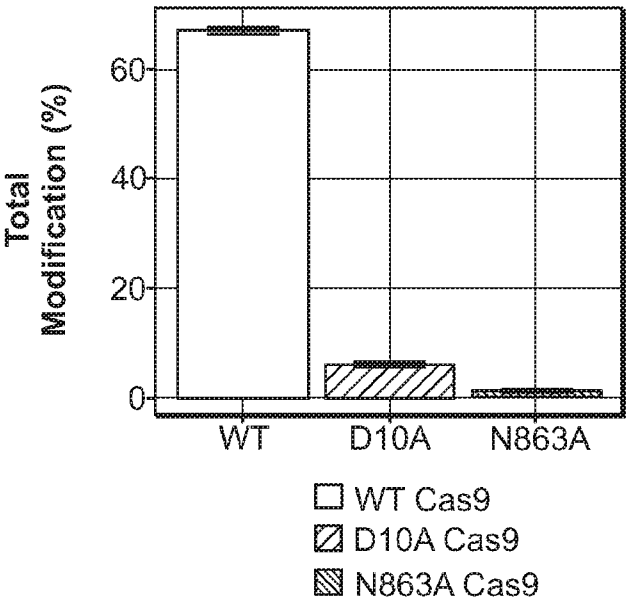
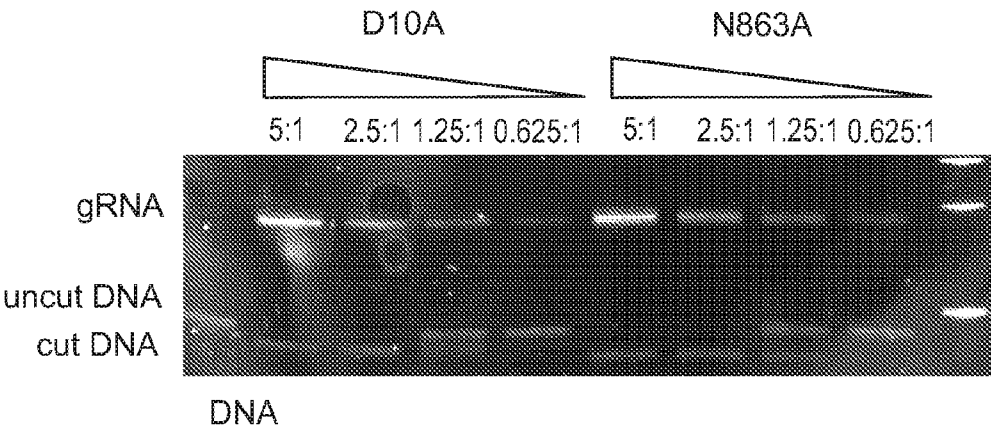


FIG. 16B



19/49

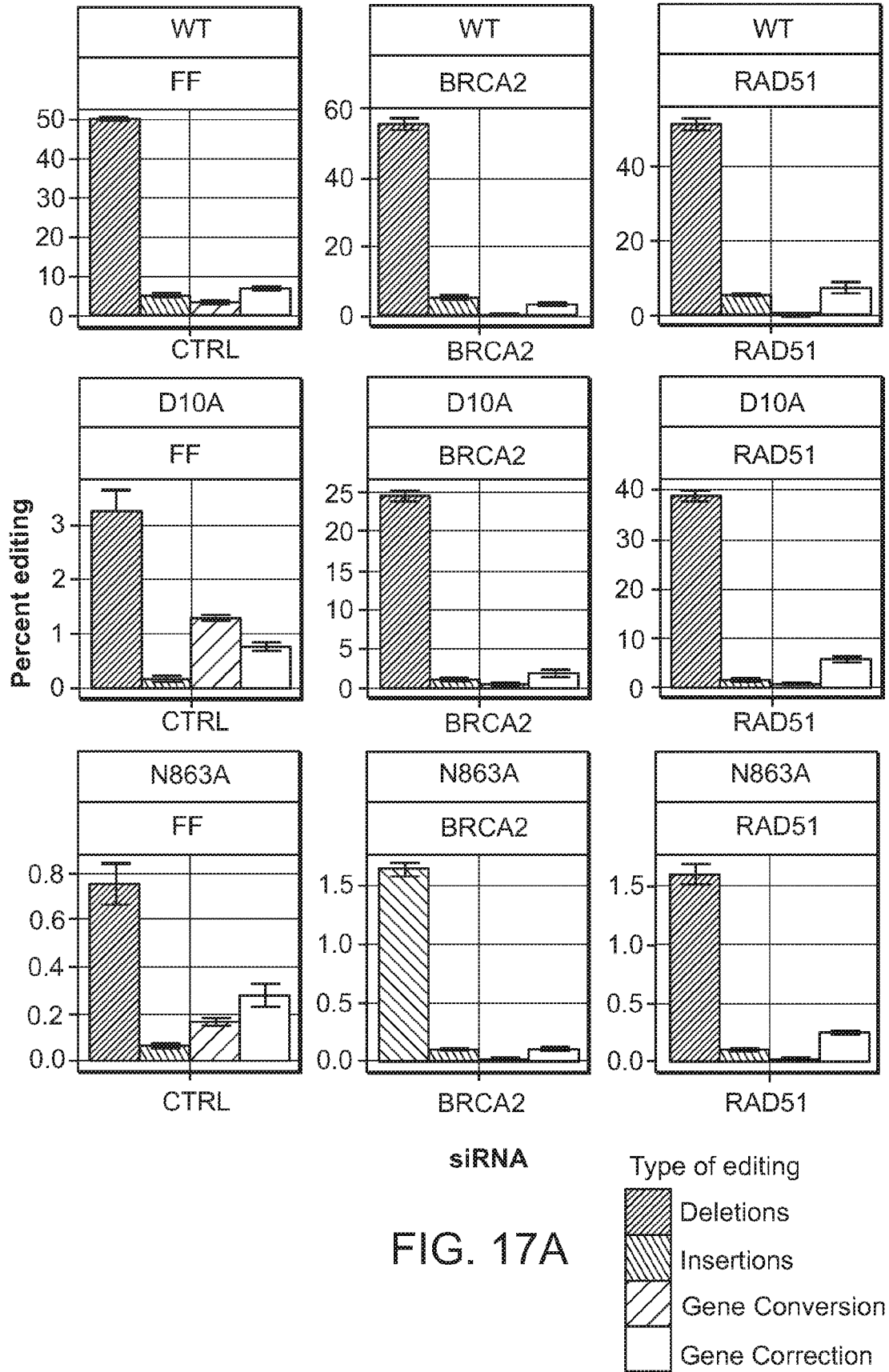
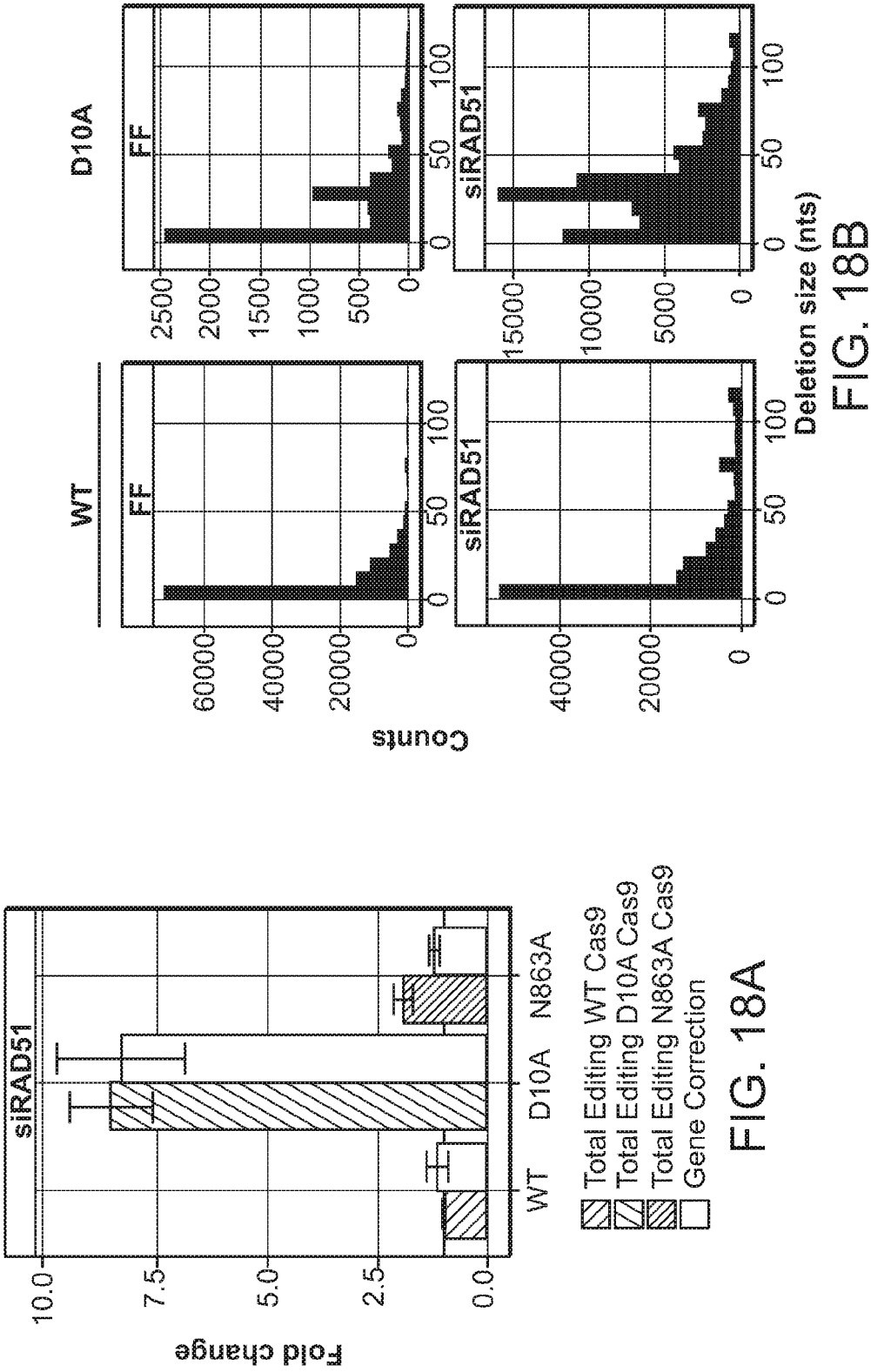


FIG. 17A



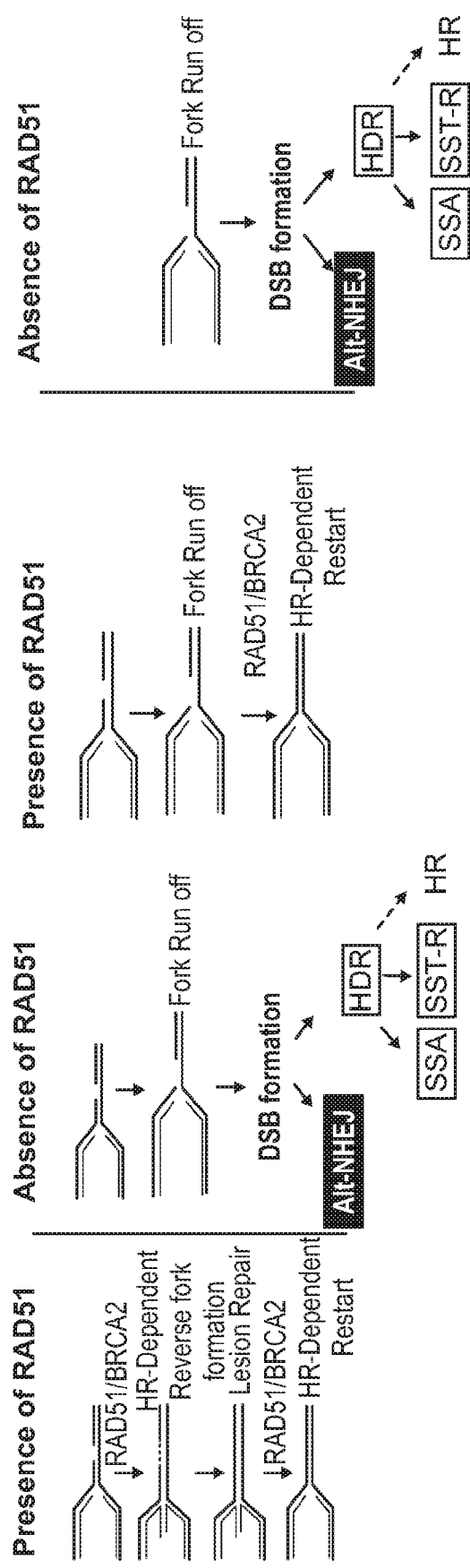


FIG. 18C

22/49

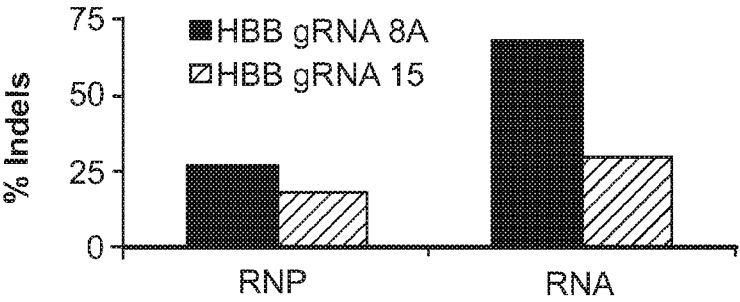


FIG. 19

23/49

FIG. 20

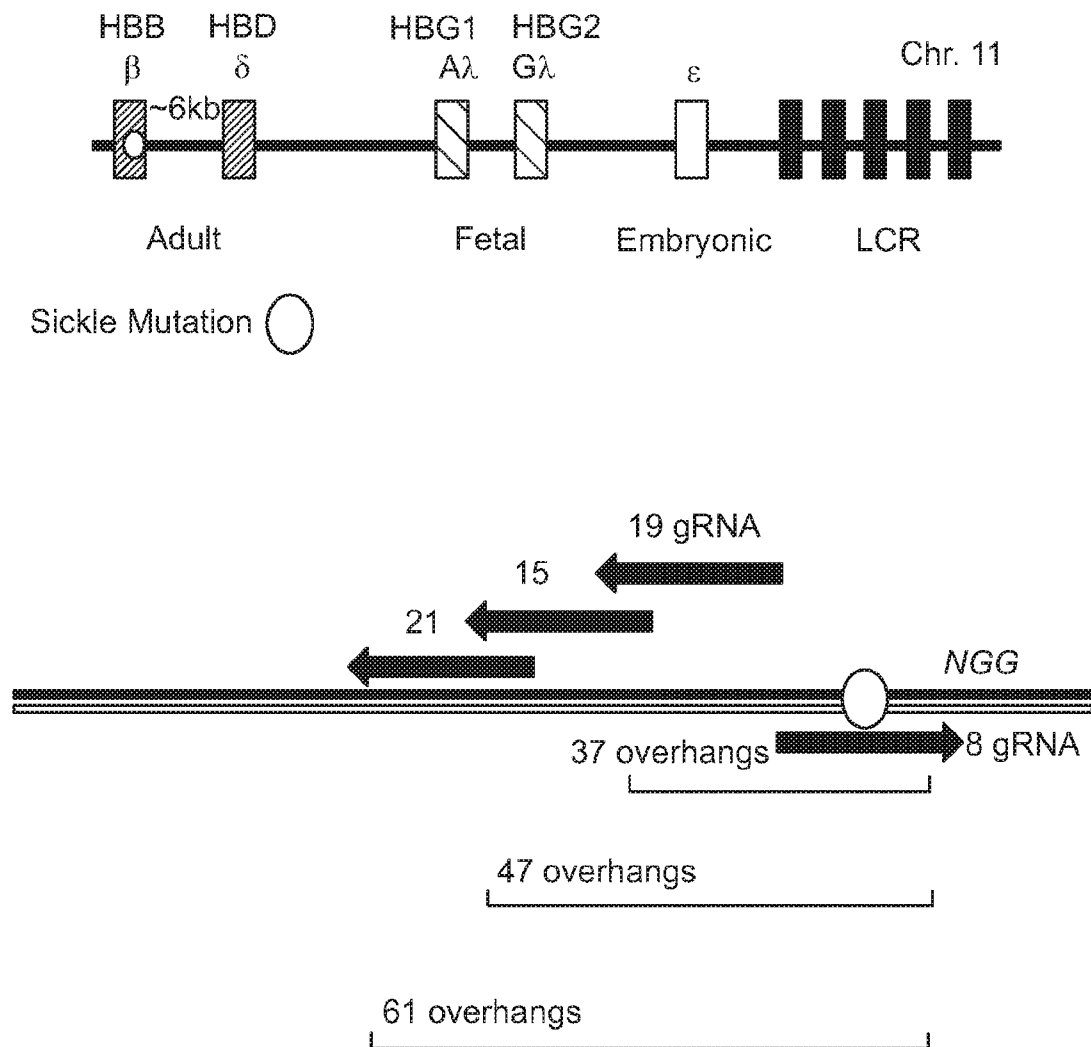
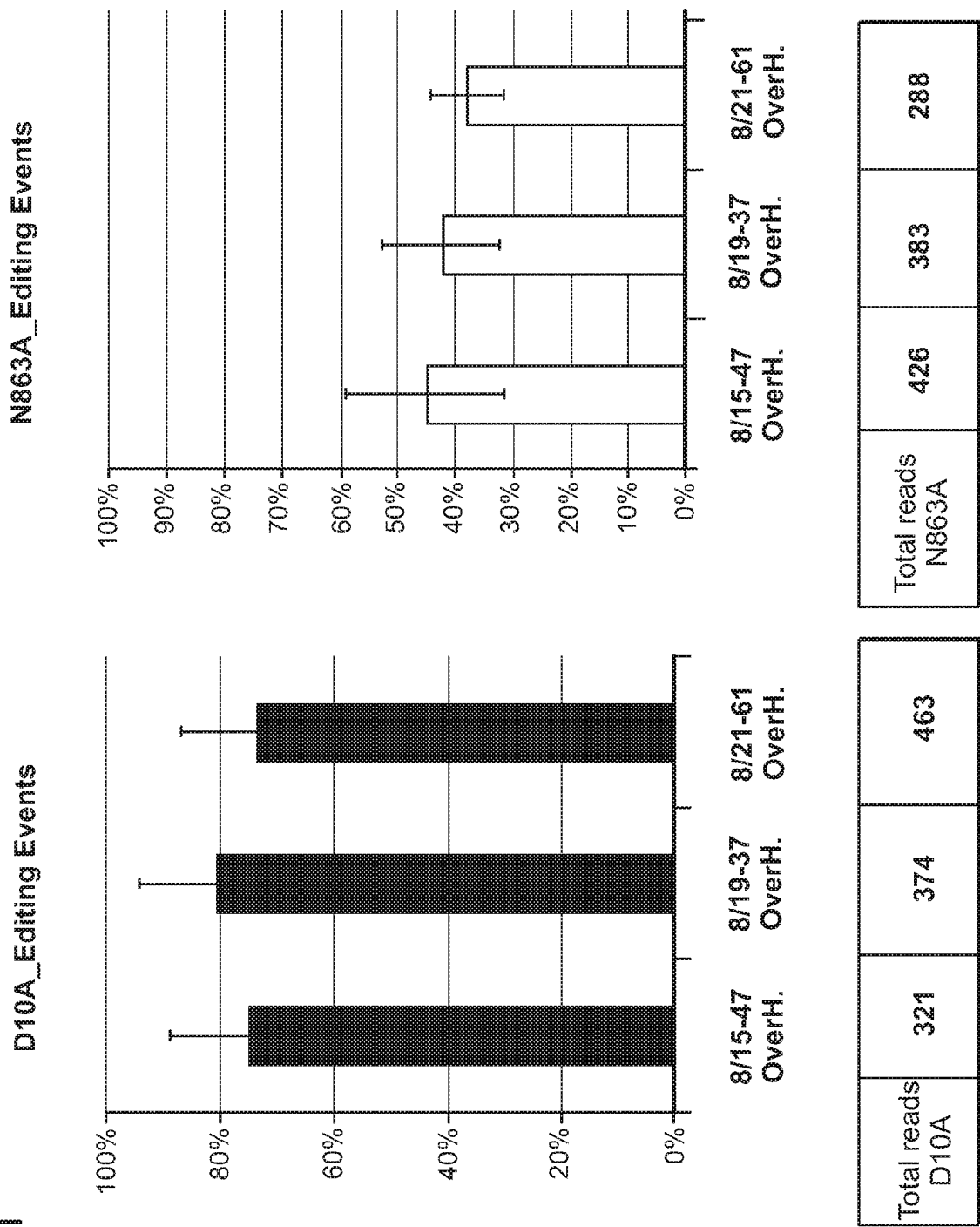


FIG. 21



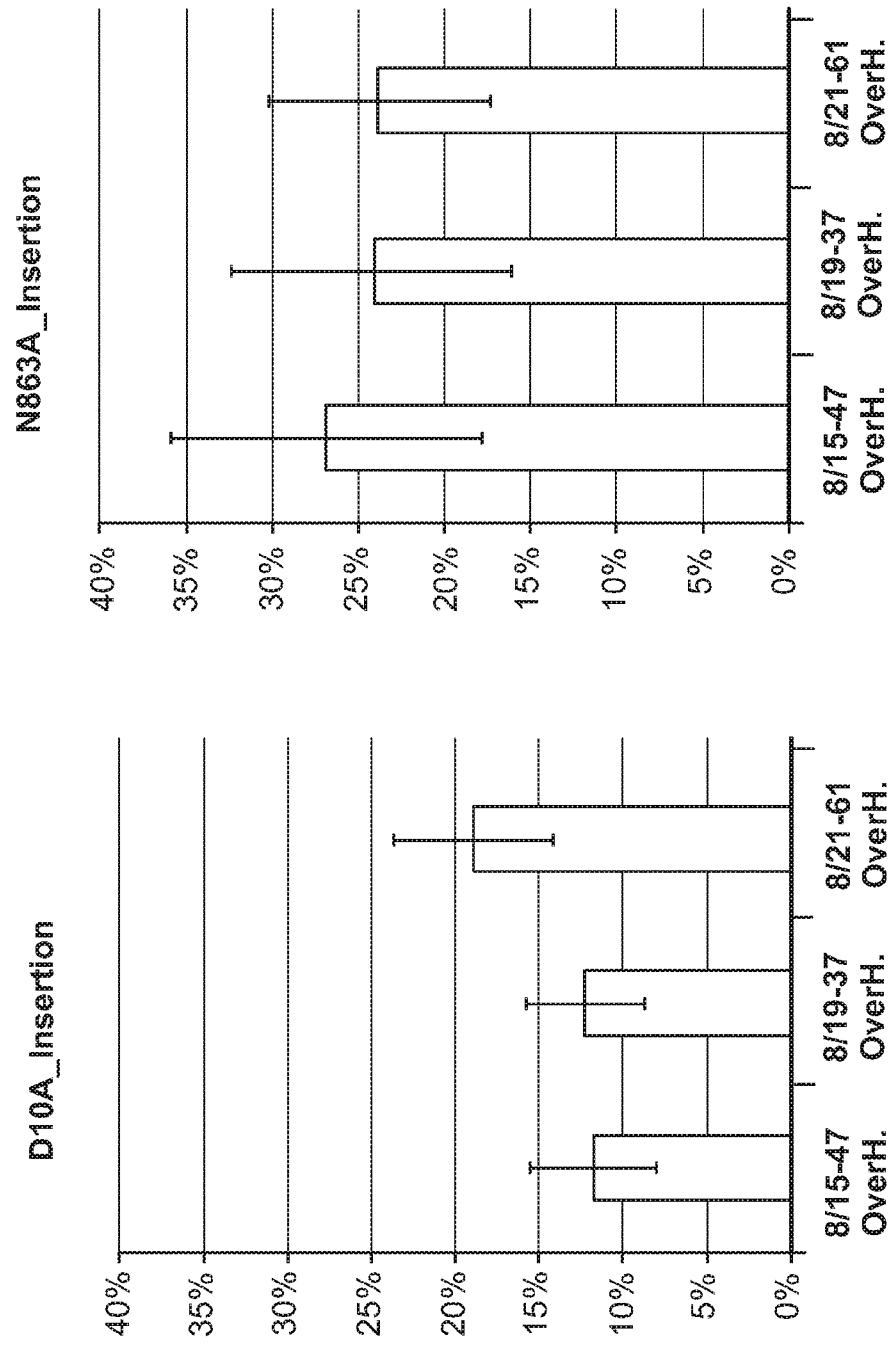


FIG. 22

26/49

% of Insertions coming from
the overhangs

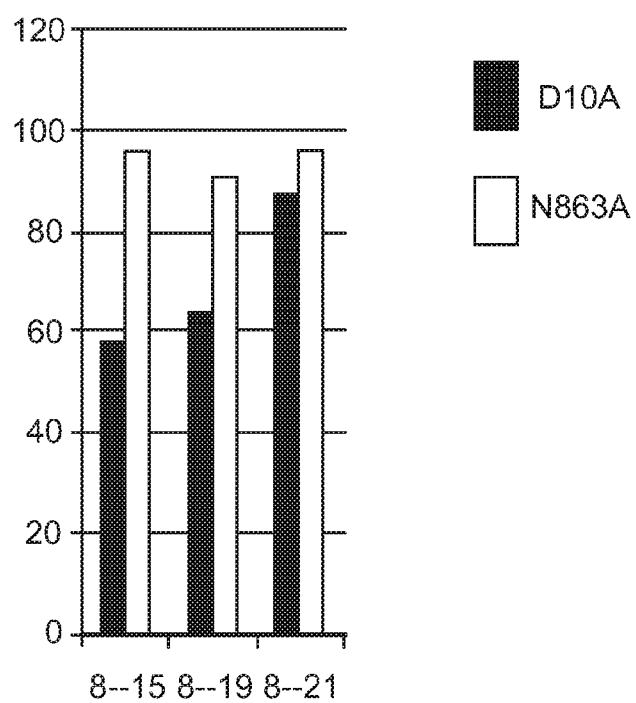


FIG. 23

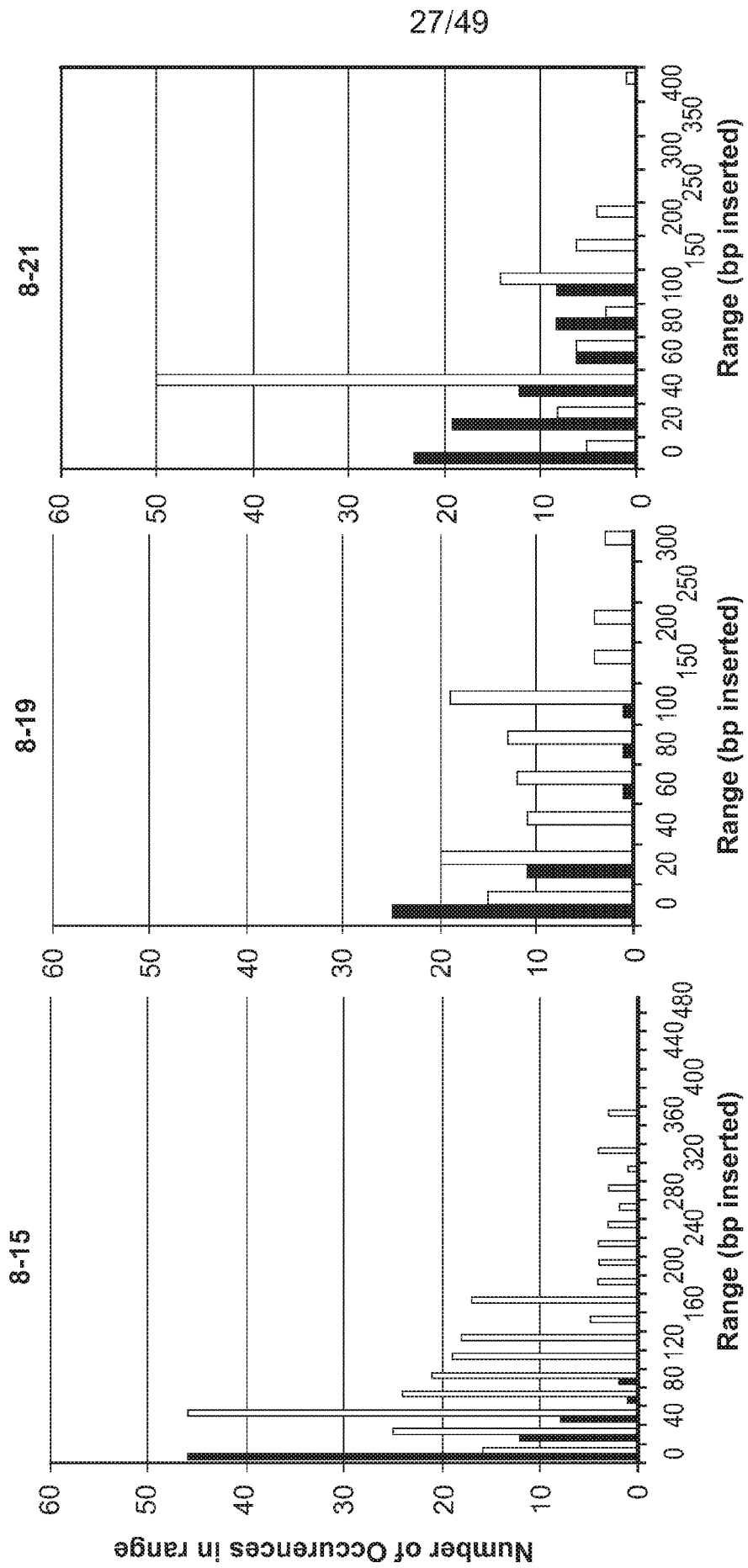


FIG. 24

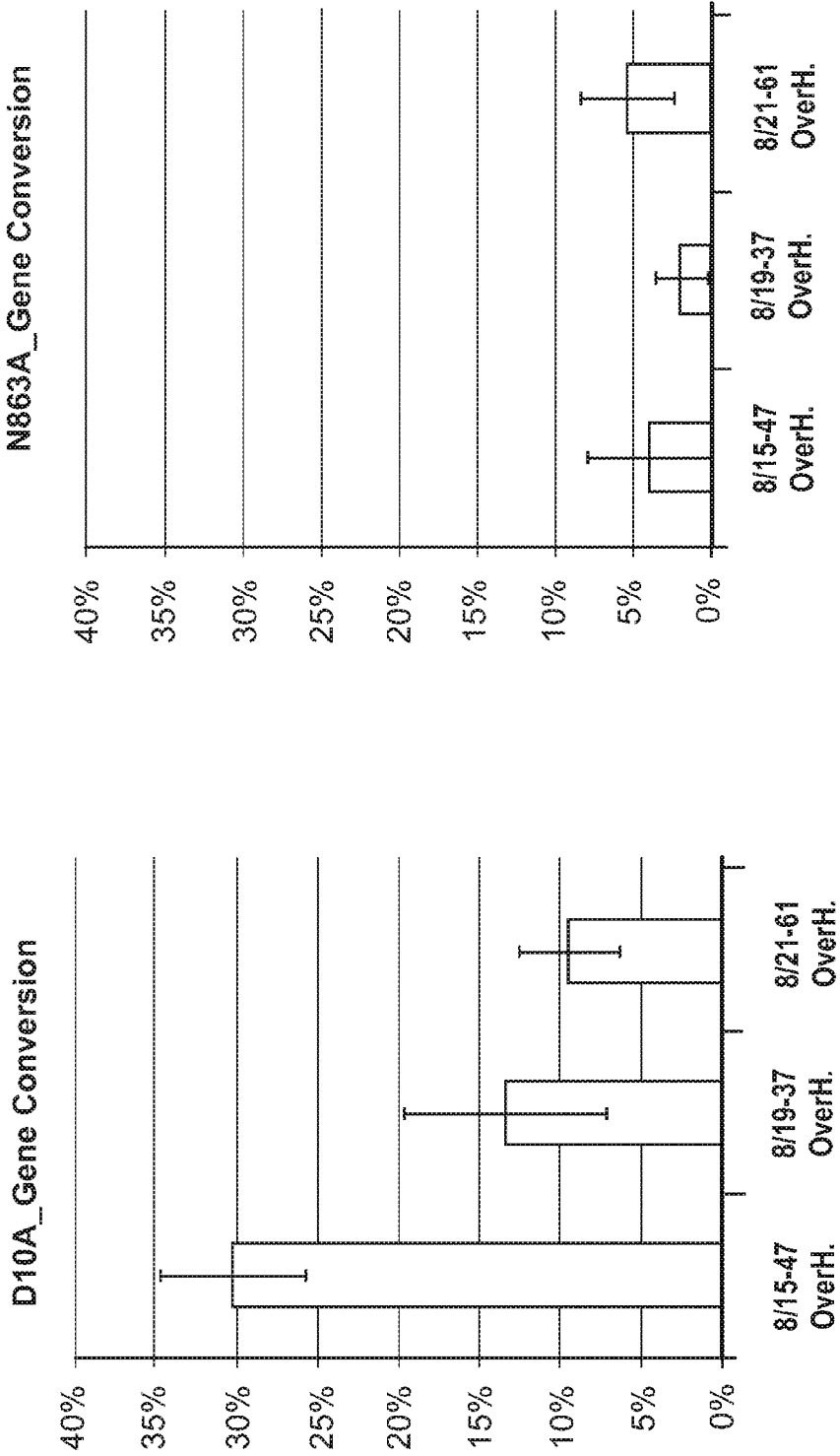


FIG. 25

FIG. 26

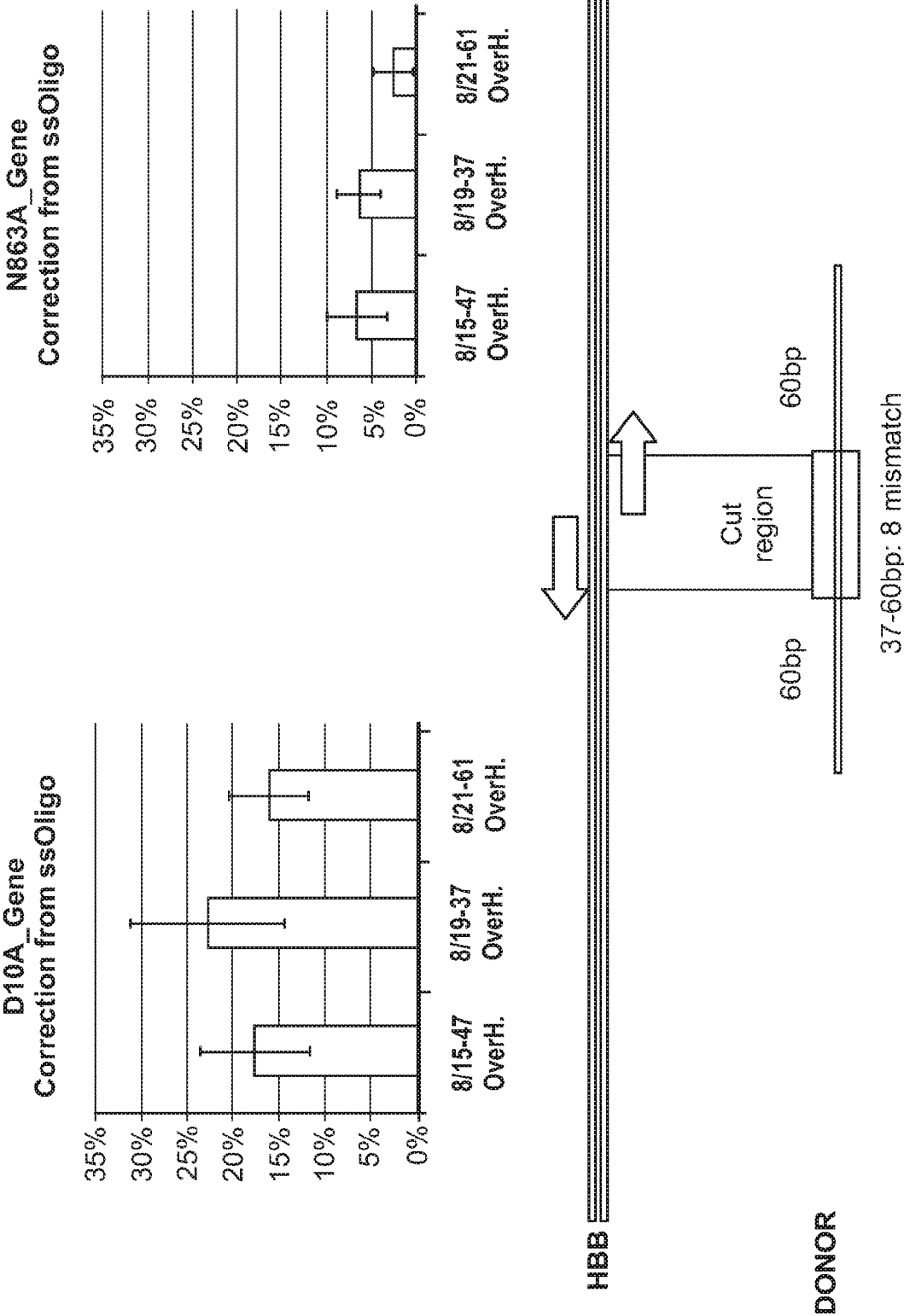
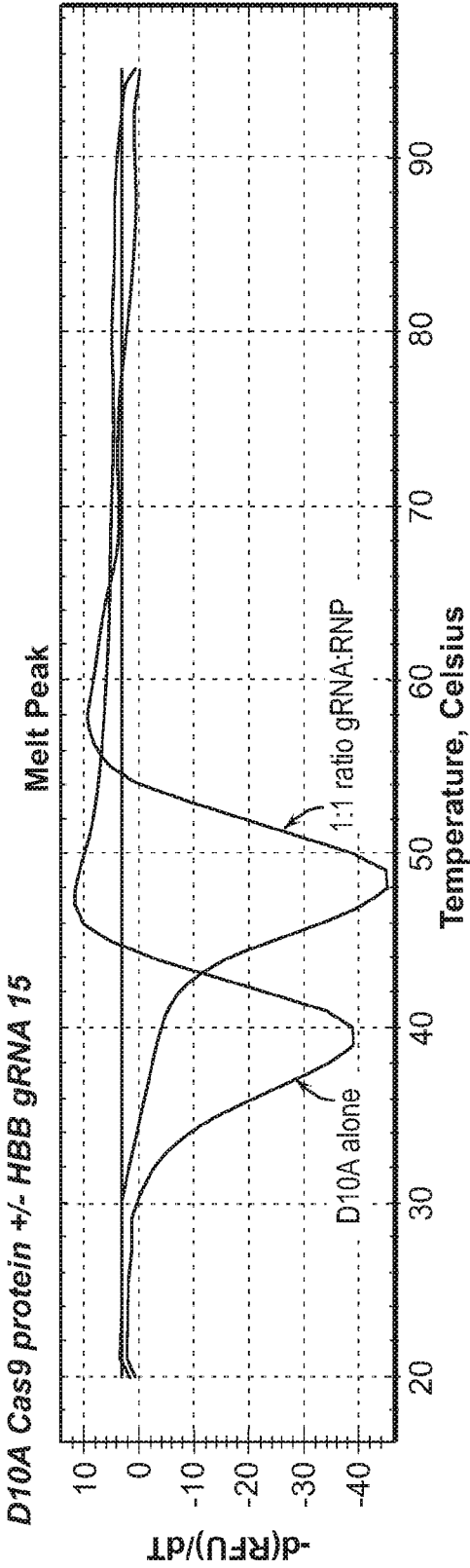
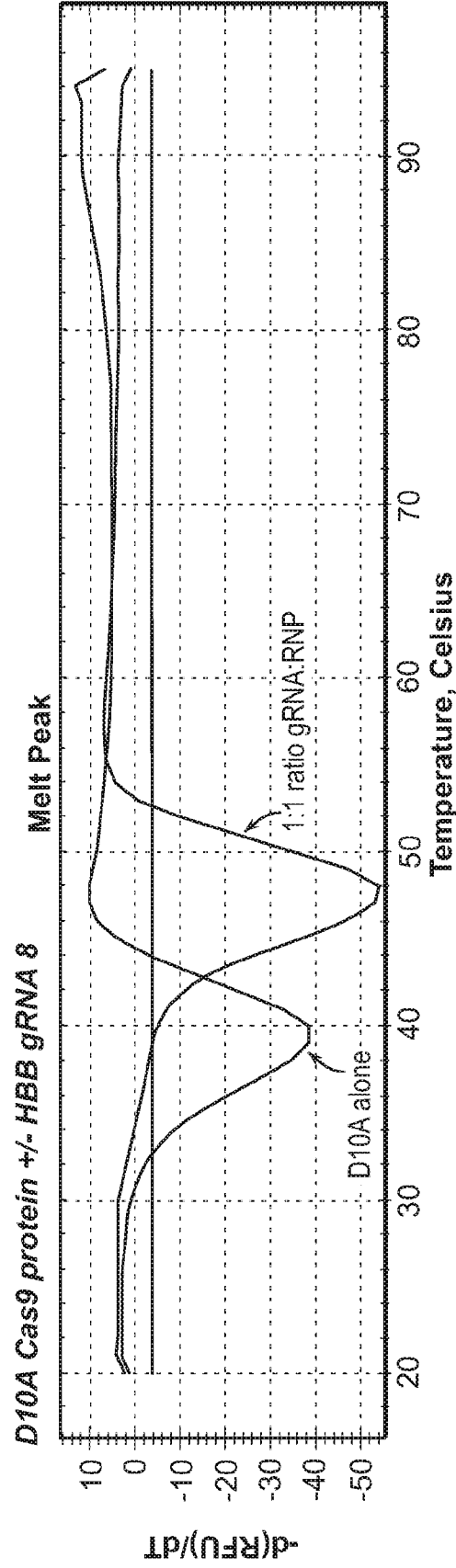


FIG. 27A



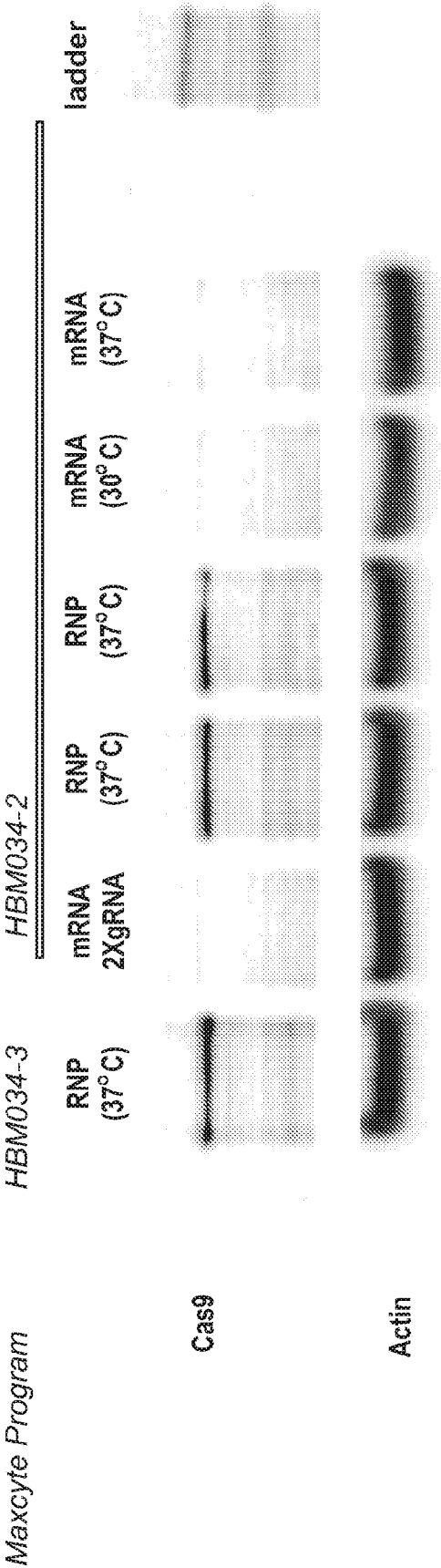


FIG. 27B

32/49

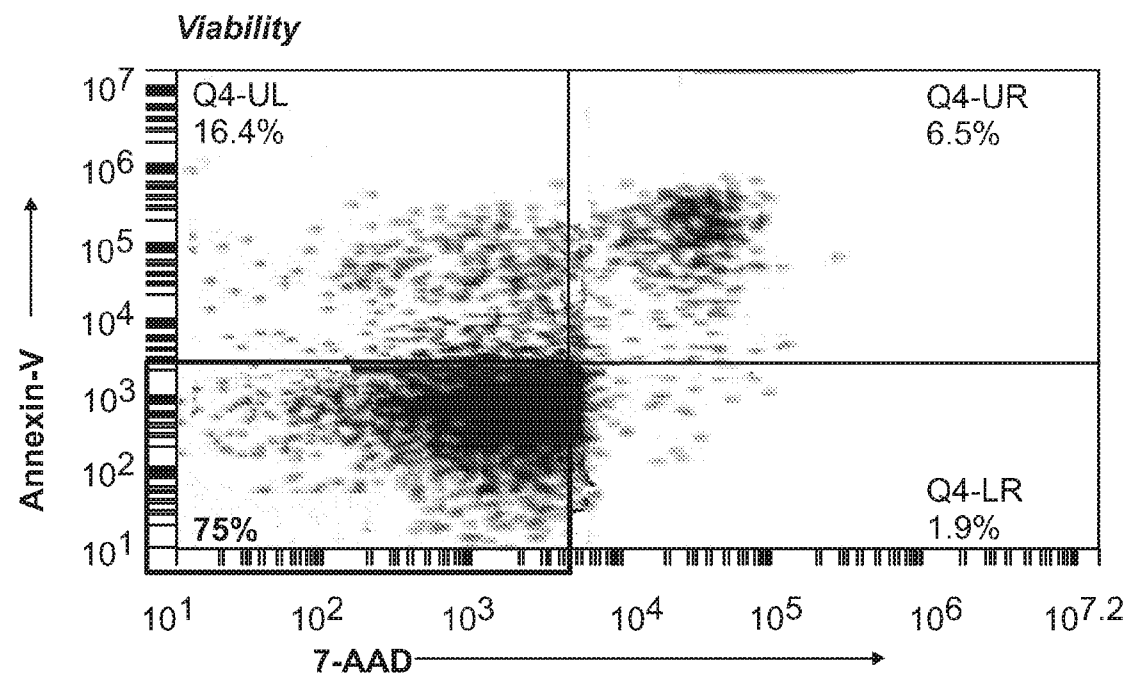
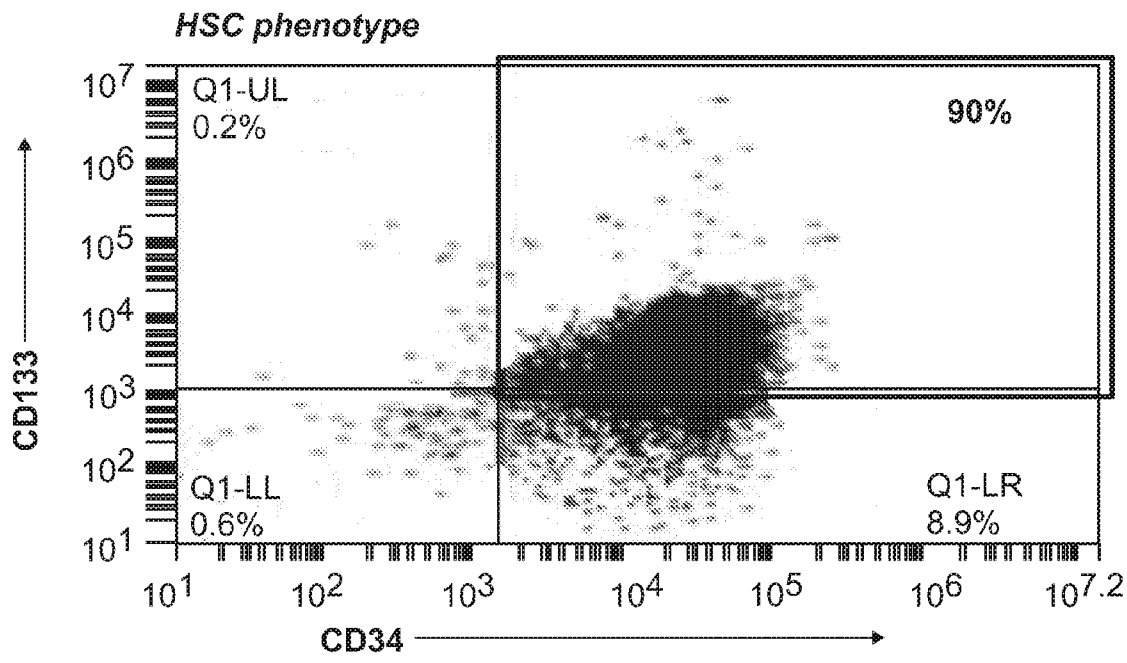


FIG. 28A

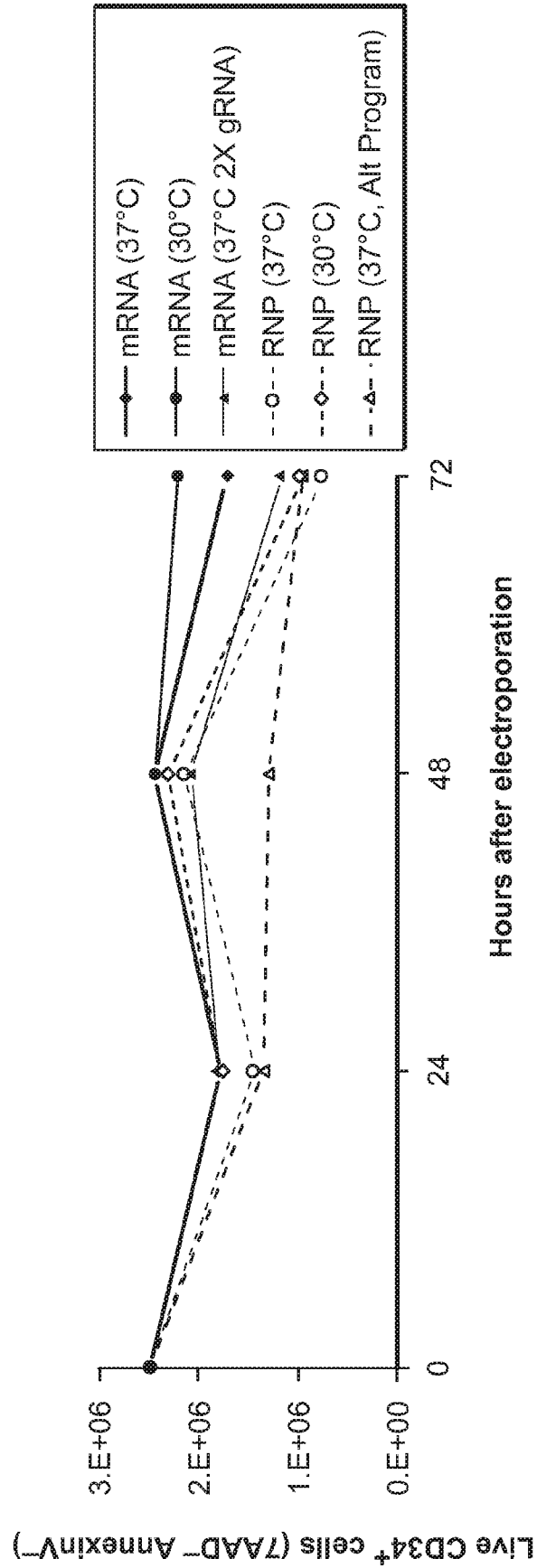


FIG. 28B

34/49

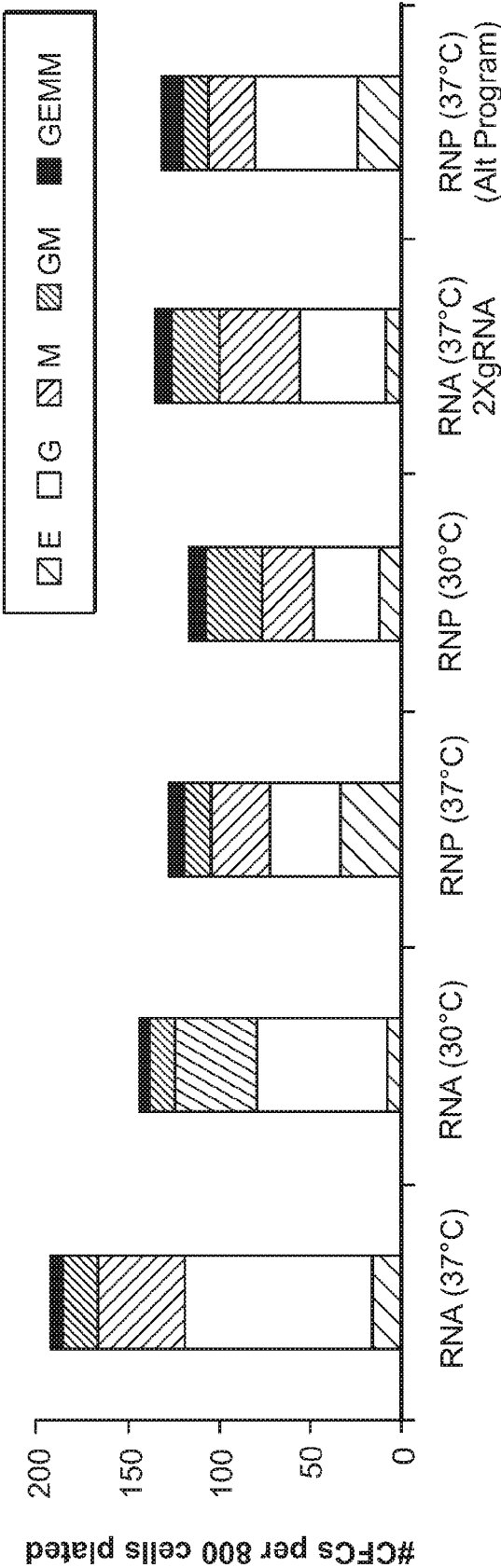
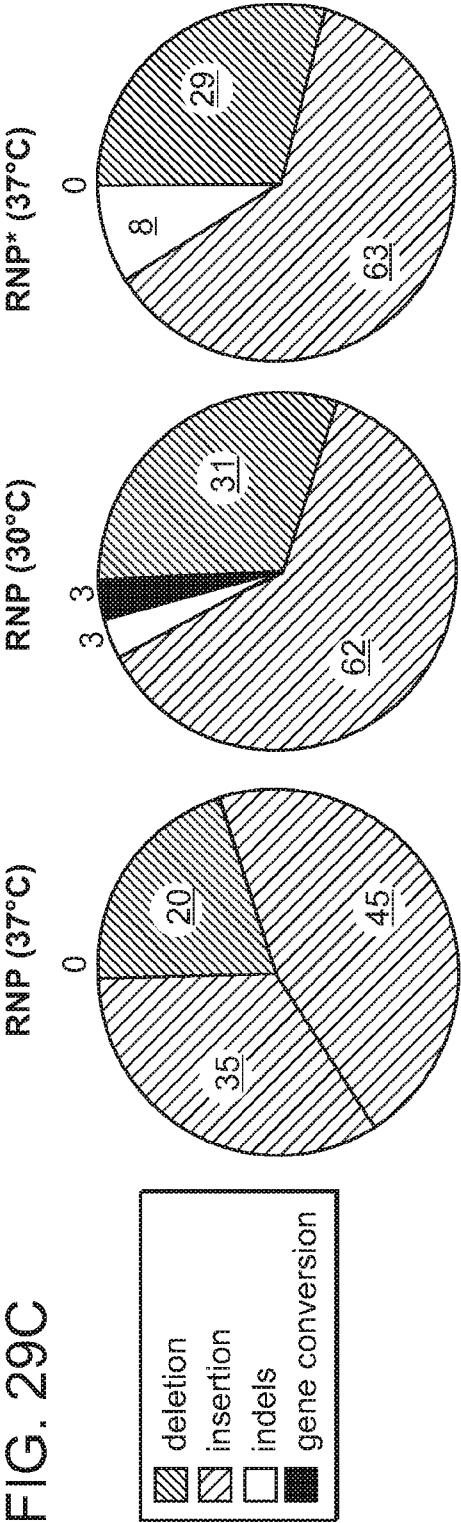
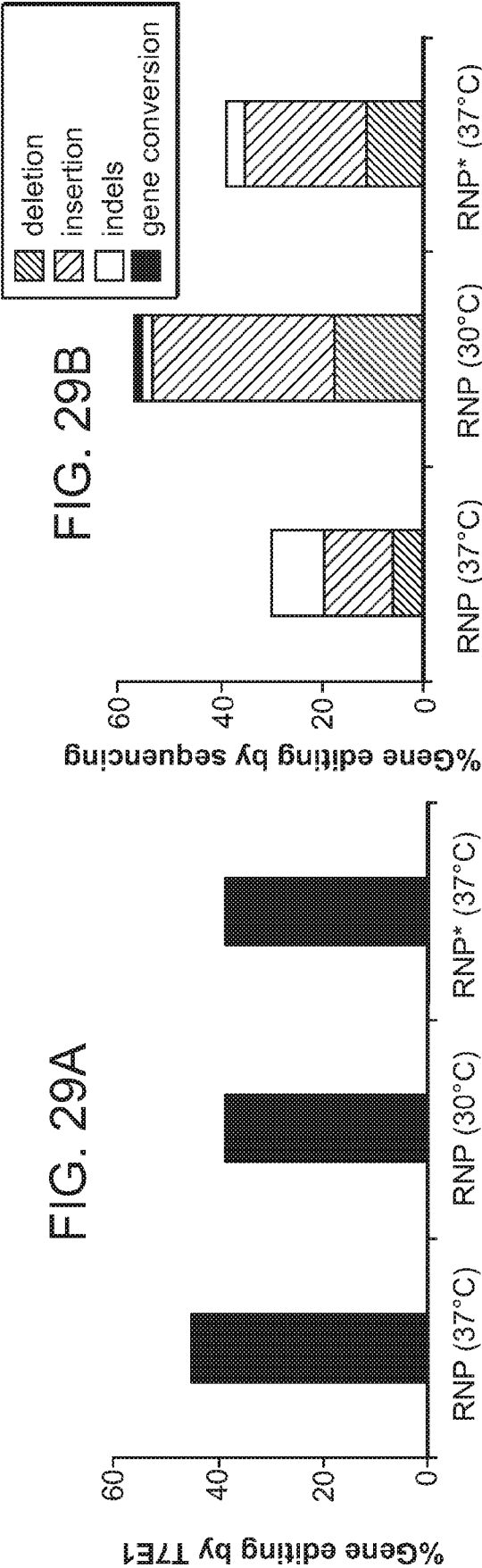


FIG. 28C



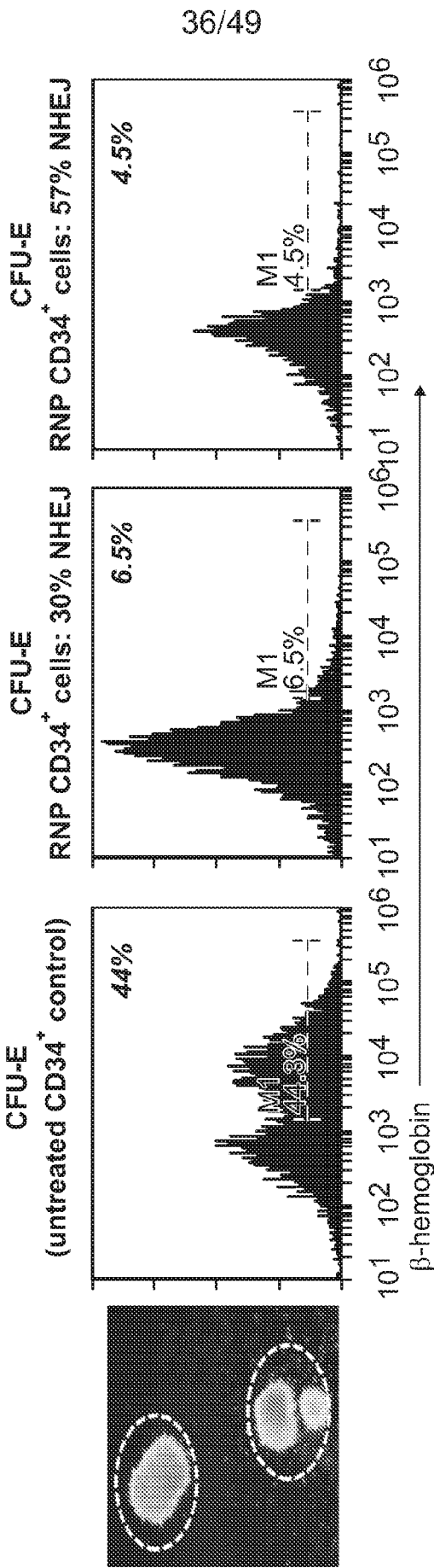


FIG. 30

37/49

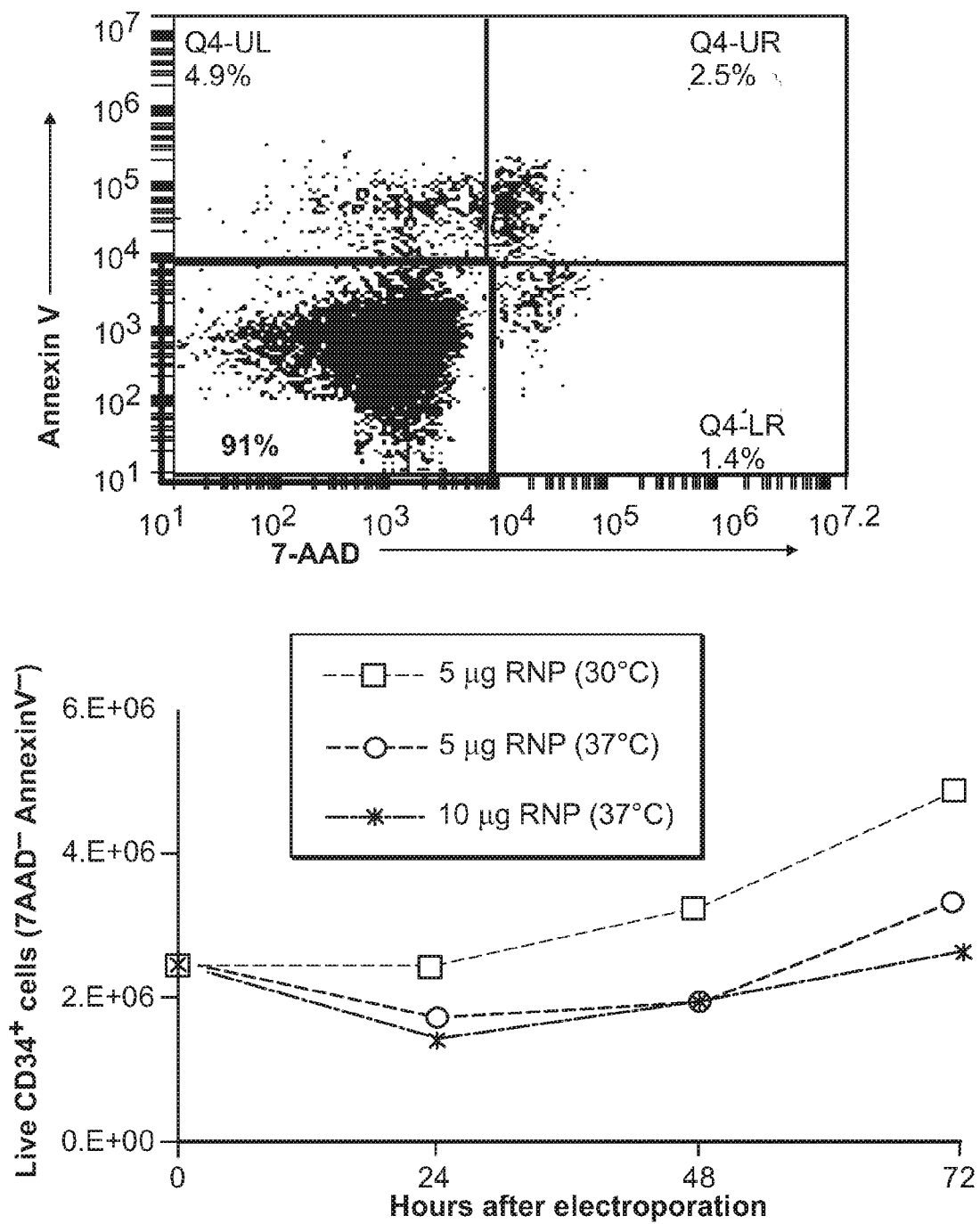


FIG. 31A

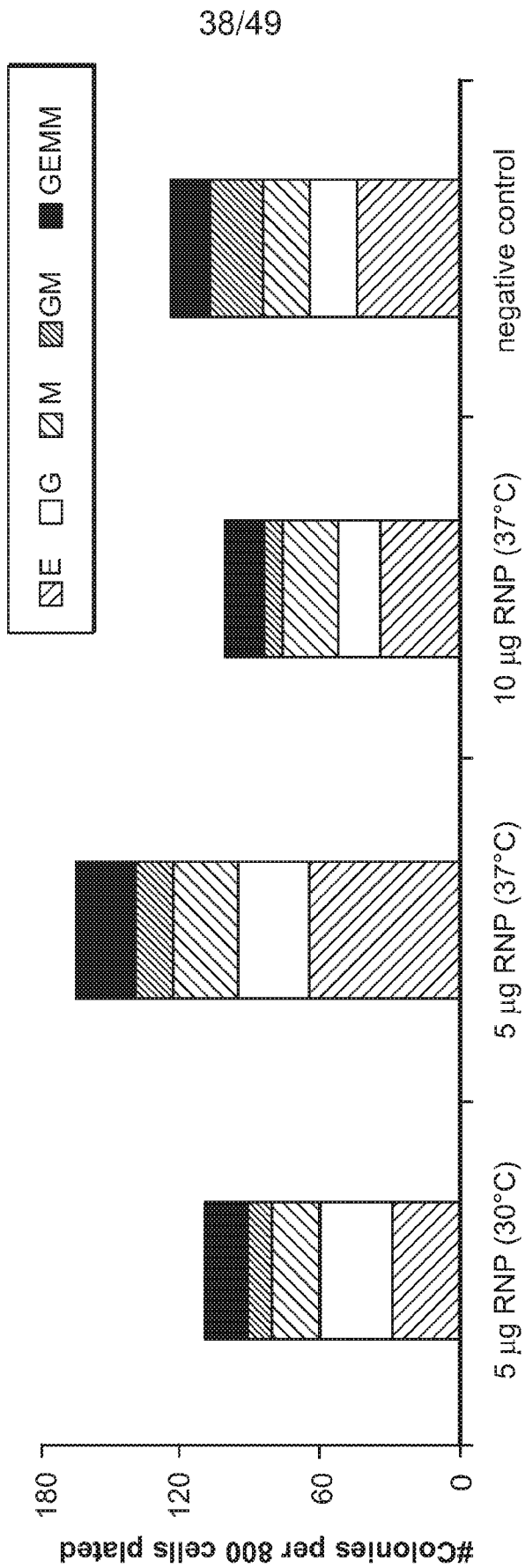


FIG. 31B

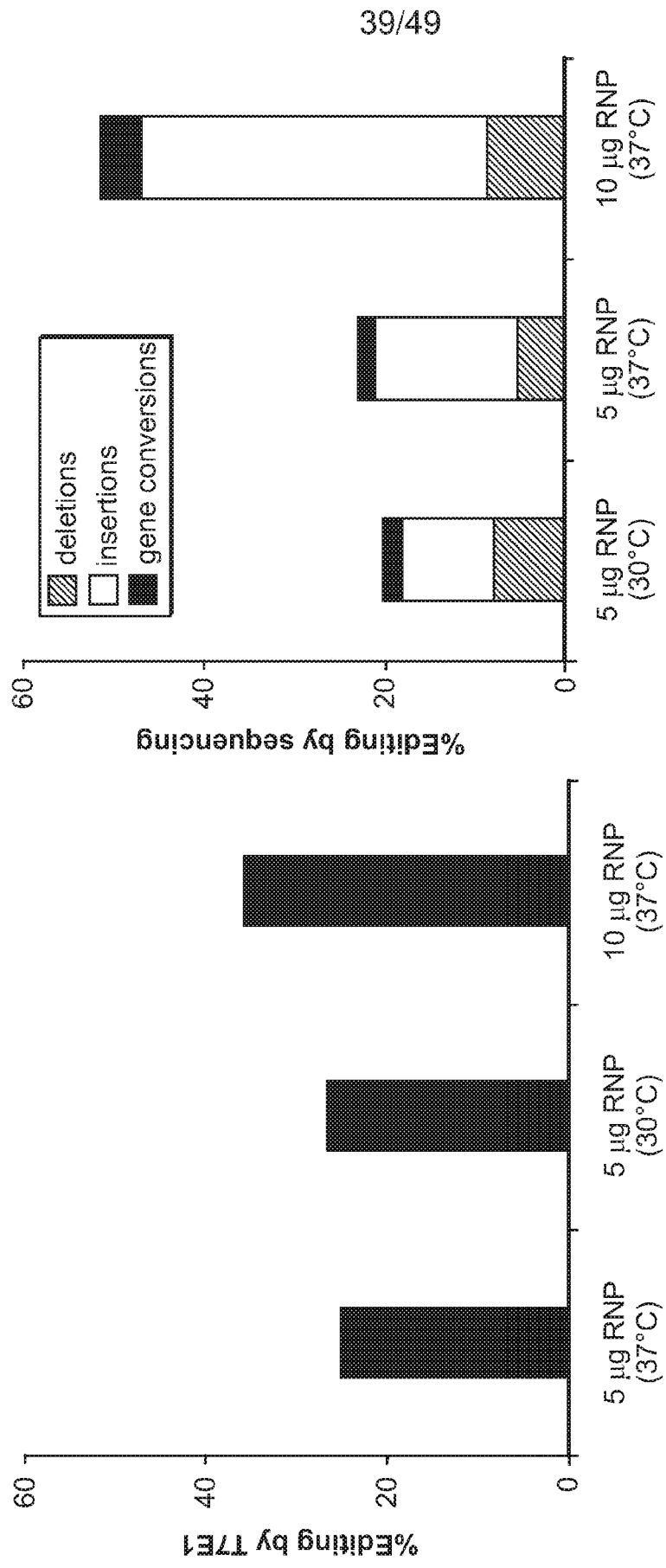


FIG. 32B

FIG. 32A

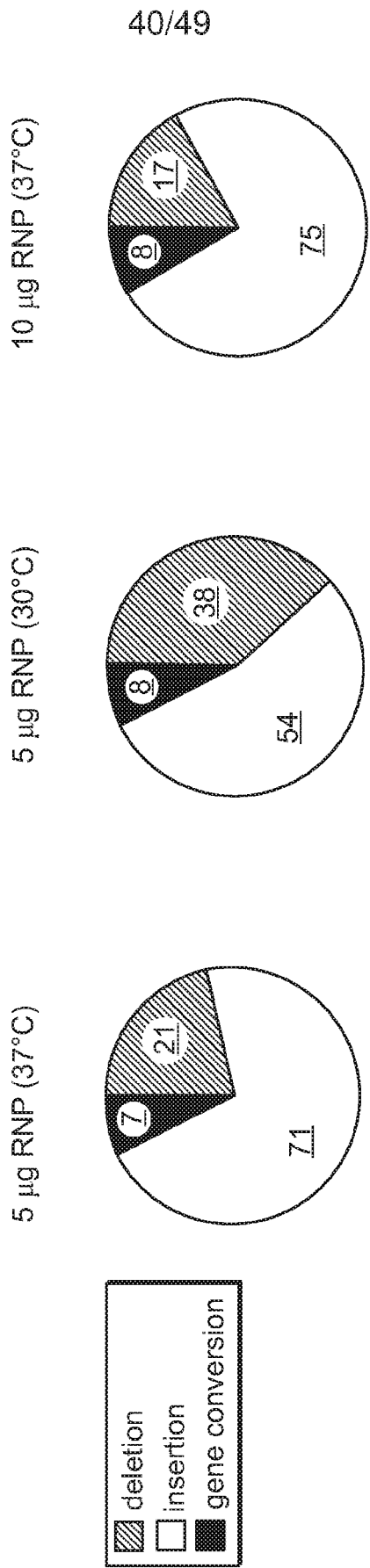


FIG. 32C

FIG. 33A

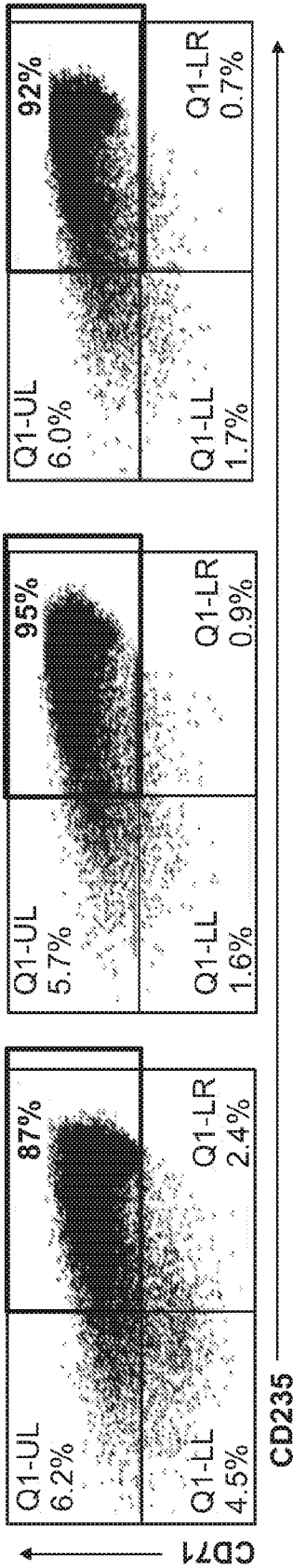


FIG. 33B

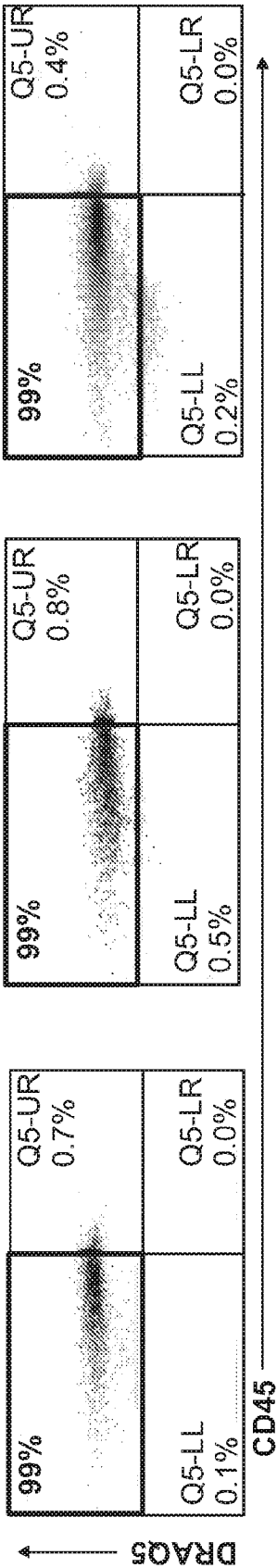
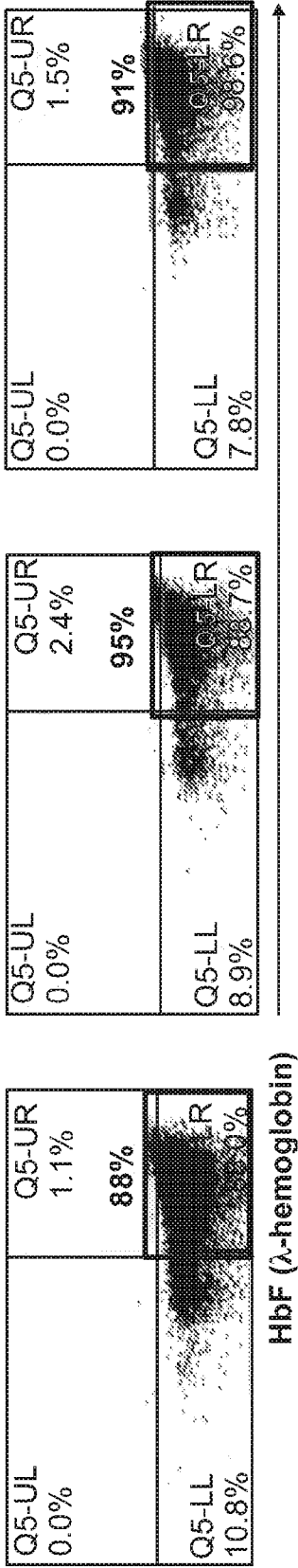


FIG. 33C



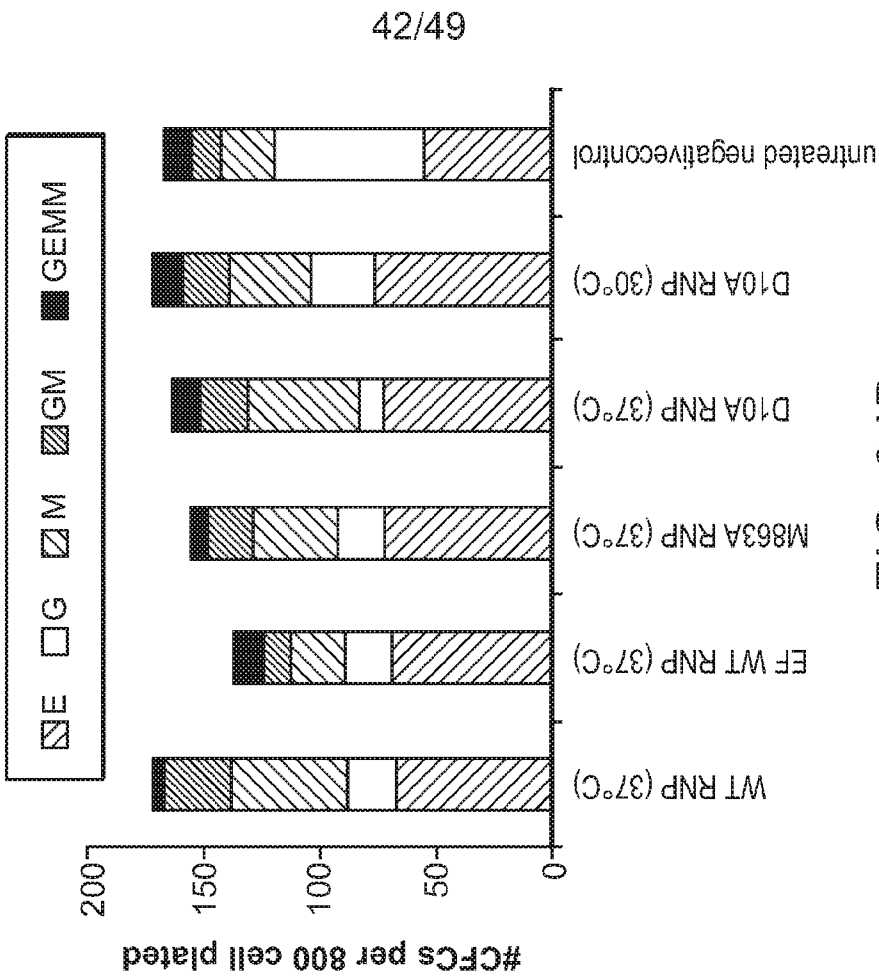
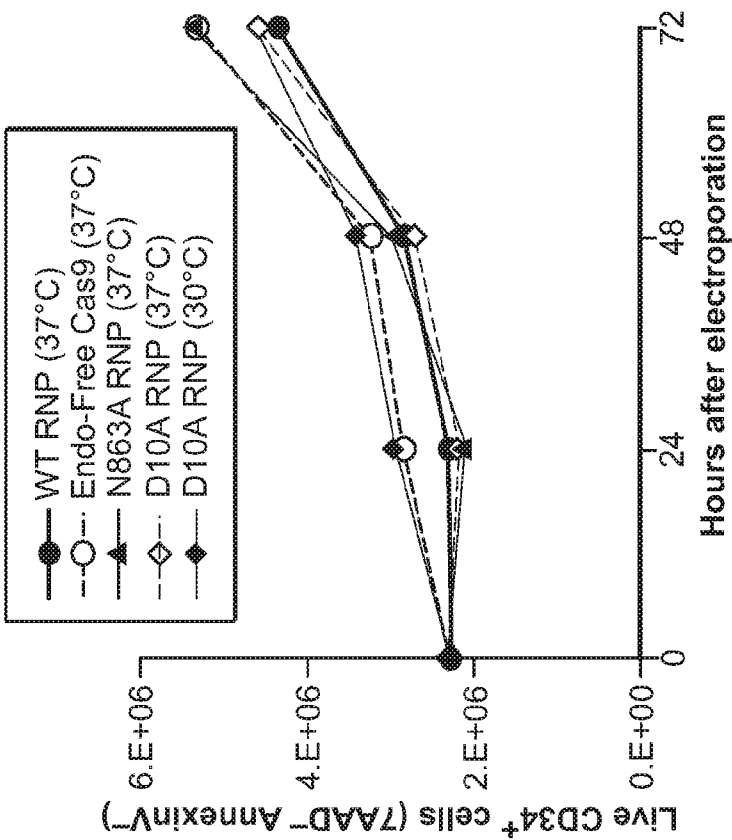


FIG. 35A

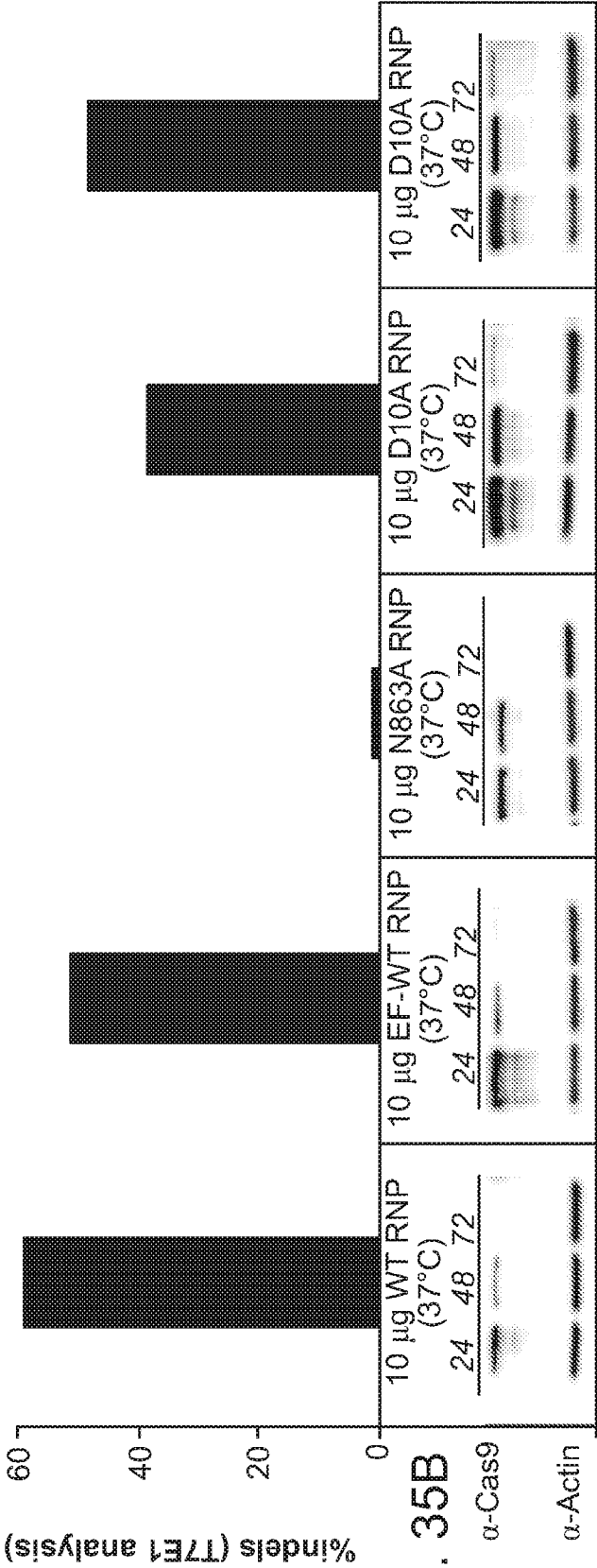


FIG. 36A

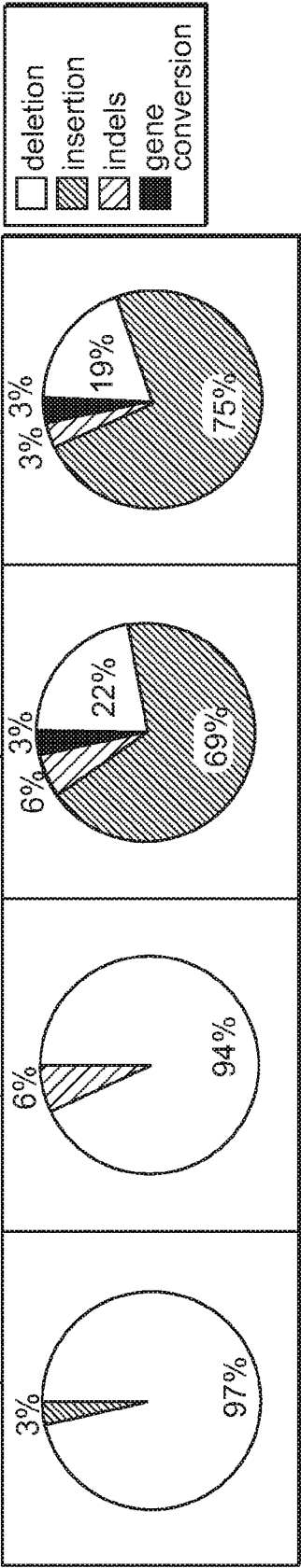
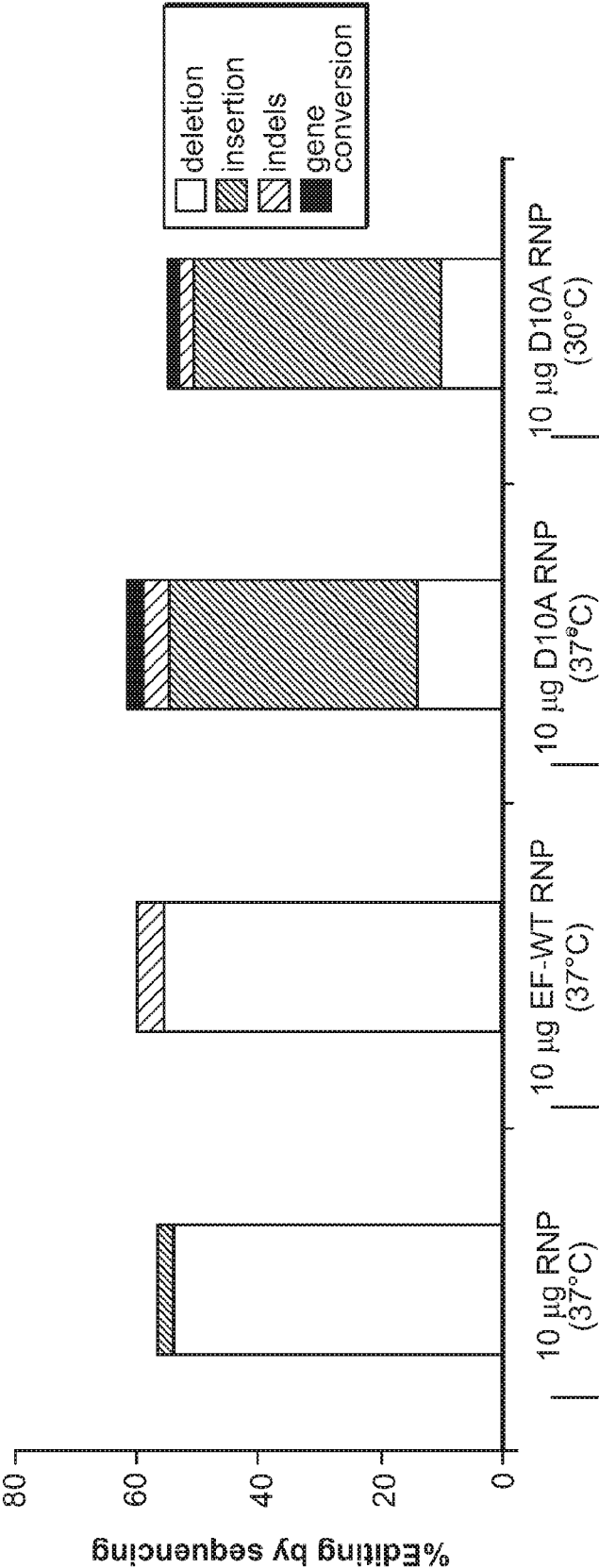
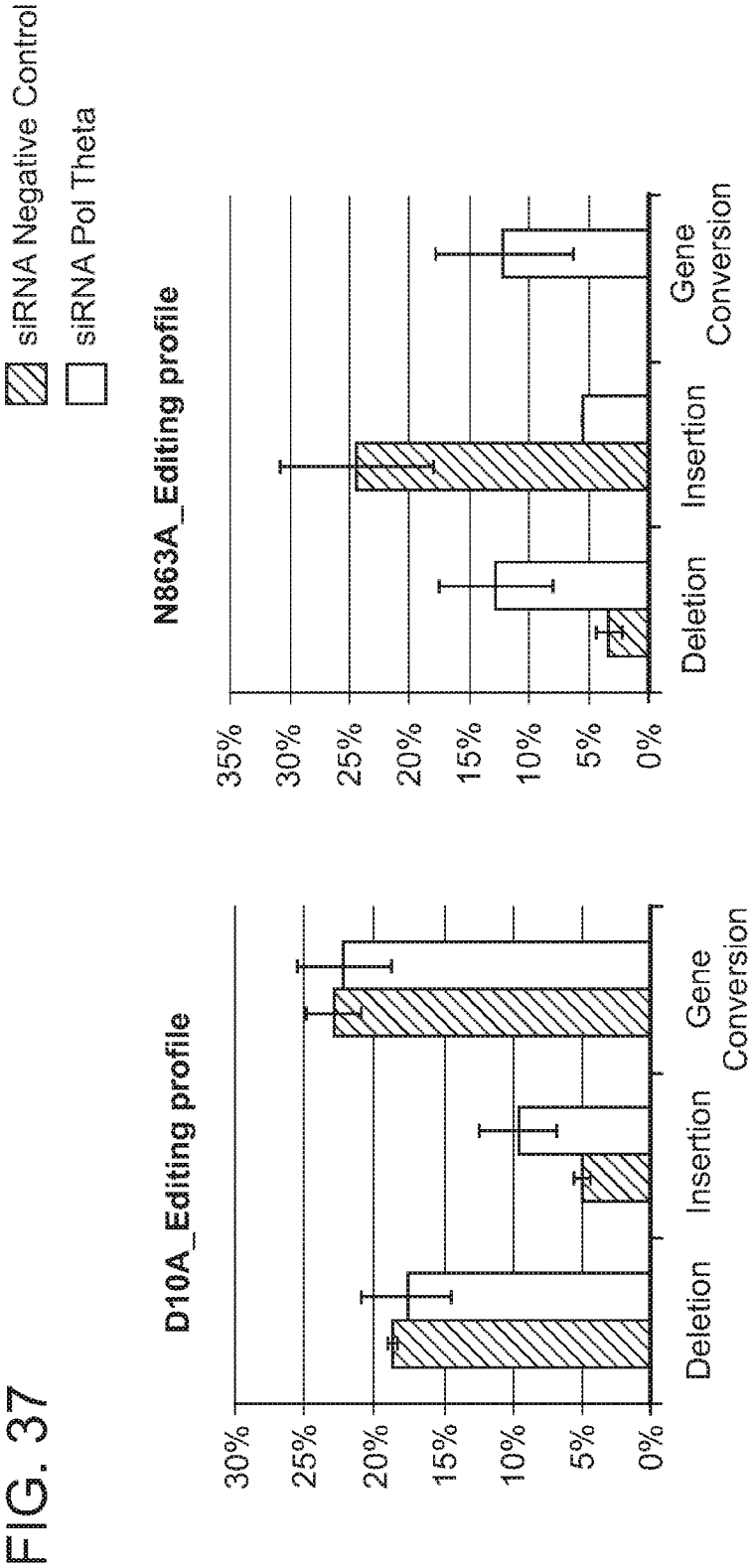


FIG. 36B



46/49

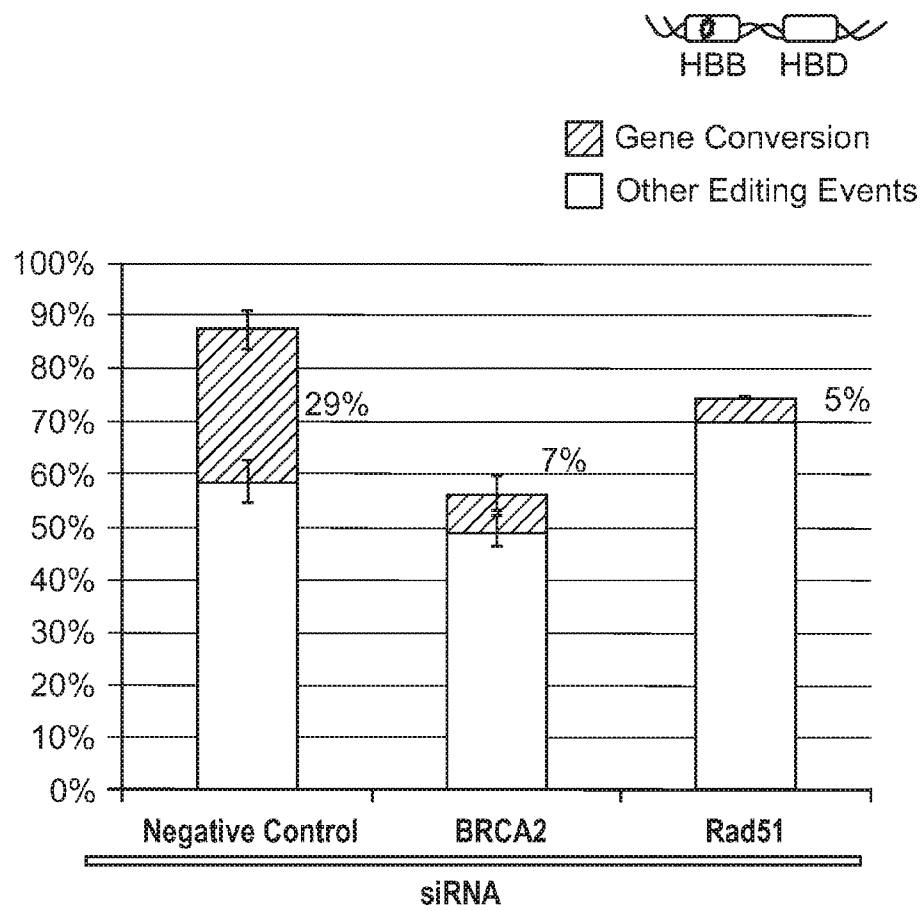


FIG. 38

FIG. 39

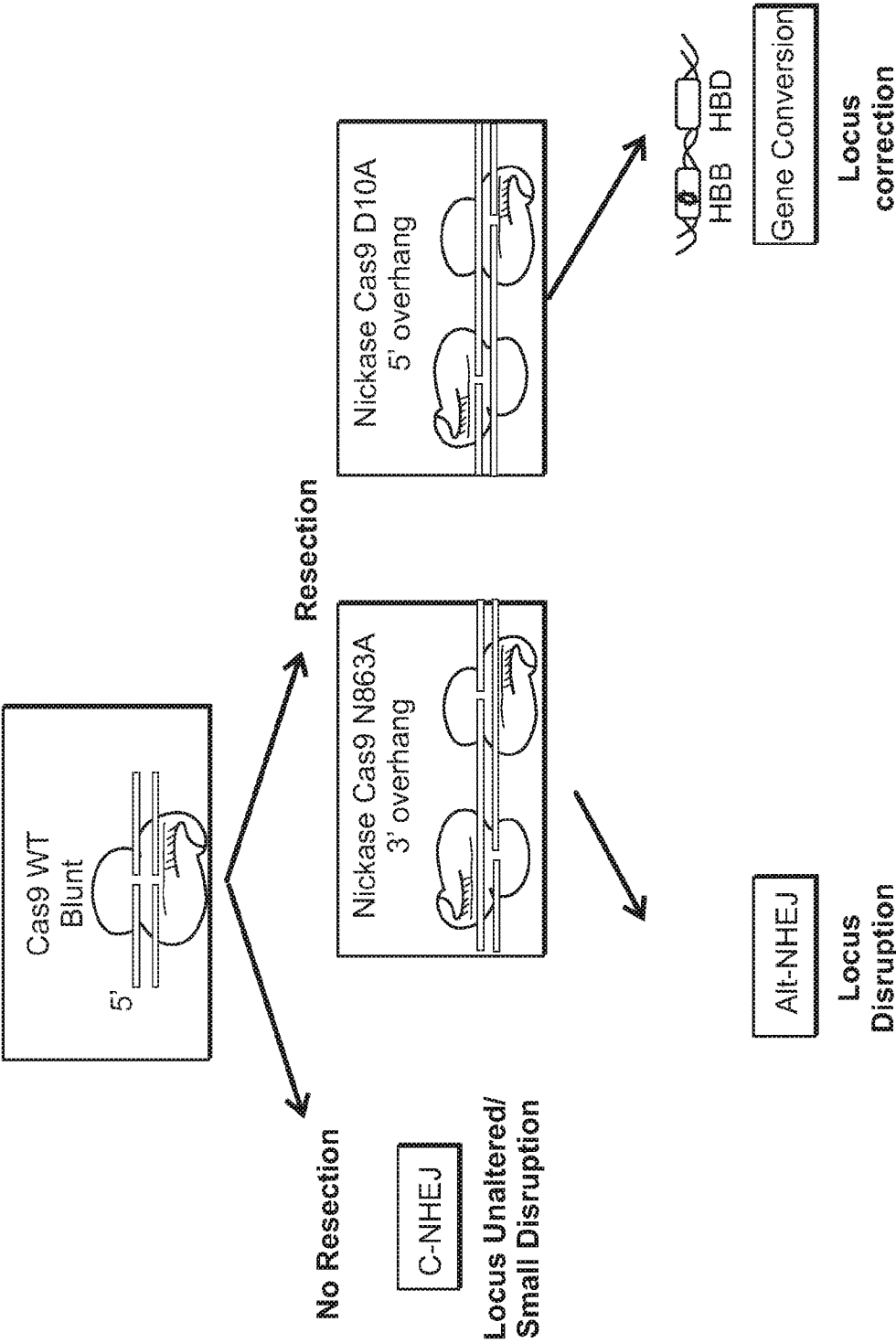
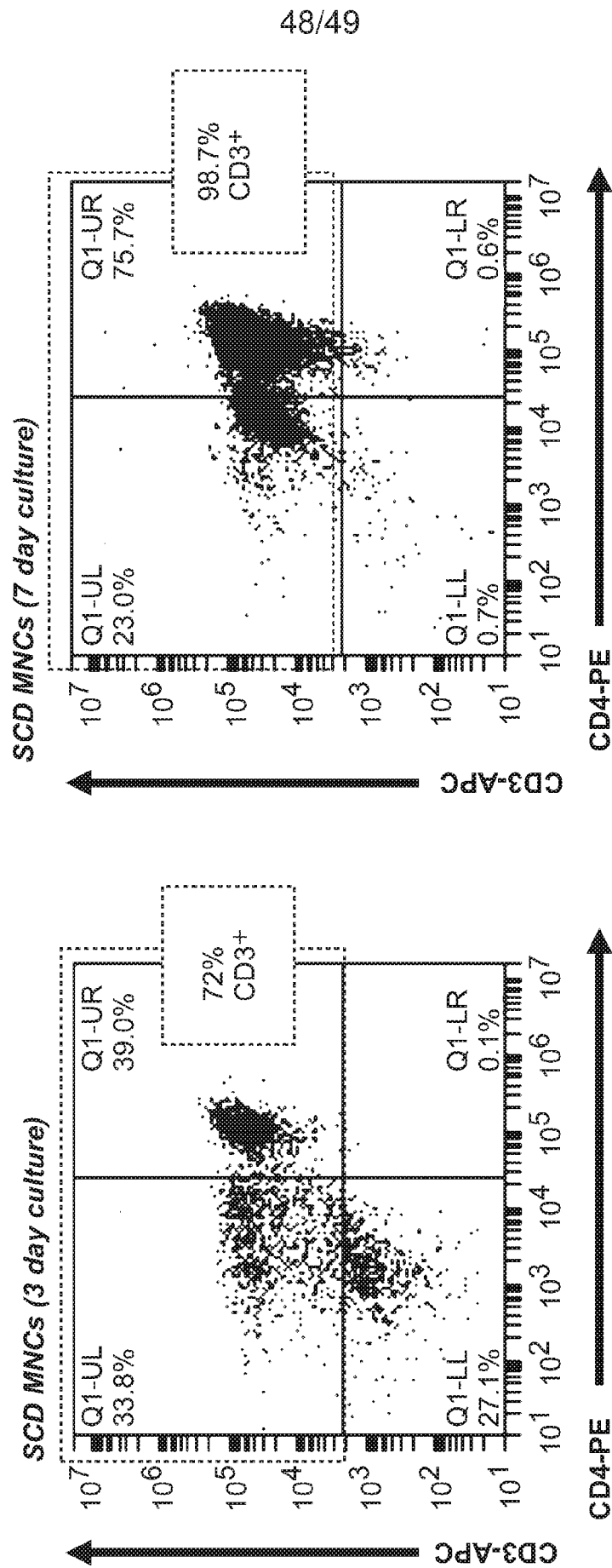


FIG. 40A



49/49

FIG. 40B



FIG. 40C

