



(51) International Patent Classification:

C12N 15/11 (2006.01)

(21) International Application Number:

PCT/US2018/032172

(22) International Filing Date:

10 May 2018 (10.05.2018)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/504,522 10 May 2017 (10.05.2017) US

62/504,527 10 May 2017 (10.05.2017) US

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

(72) Inventors: AUPEPIN DE LAMOTHE-DREUZY, Edouard; 123 Pinckney Street, Boston, Massachusetts 02114 (US). HEATH, Jack; 1 Wildwood Street, Winchester, Massachusetts 01890 (US). GORI, Jennifer Leah; 20 Spalding Street, Unit 1, Jamaica Plain, Massachusetts 02130 (US). BARRERA, Luis; 11 Morton Street #1, Somerville, Massachusetts 02145 (US).

(74) Agent: PROCHNOW, Courtney; Perkins Coie LLP, P.O. Box 1247, Seattle, Washington 98111-1247 (US).

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

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

Declarations under Rule 4.17:

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

Published:

- 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/RNA-GUIDED NUCLEASE SYSTEMS AND METHODS

(57) Abstract: Genome editing systems, guide RNAs, dead guide RNAs, and CRISPR-mediated methods are provided for altering portions of a target nucleic acid.



CRISPR/RNA-GUIDED NUCLEASE SYSTEMS AND METHODS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] The present application claims the benefit of United States Provisional Application No. 62/504,527, filed May 10, 2017 and United States Provisional Application No. 62/504,522, filed May 10, 2017; the contents of each of which is hereby incorporated by reference in its entirety.

SEQUENCE LISTING

[0002] This application contains a Sequence Listing, which was submitted in ASCII format via EFS-Web, and is hereby incorporated by reference in its entirety. The ASCII copy, created on May 10, 2018, is named 8014WO01_SequenceListing.txt and is 345 kilobytes in size.

FIELD

[0003] This disclosure relates to genome editing systems and methods for altering a target nucleic acid sequence, or modulating expression of a target nucleic acid sequence, and applications thereof in connection with enhancing editing of a target nucleic acid and applications thereof.

BACKGROUND

[0004] CRISPRs (Clustered Regularly Interspaced Short Palindromic Repeats) 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 complementary to the viral genome, mediates targeting of an RNA-guided nuclease to a target sequence in the viral genome. The RNA-guided nuclease, in turn, cleaves and thereby silences the viral target.

[0005] Recently, CRISPR-based genome editing systems have been adapted for genome editing in eukaryotic cells. The introduction of site-specific double strand breaks (DSBs) allows for target sequence alteration through endogenous DNA repair mechanisms, for example non-homologous end-joining (NHEJ) or homology-directed repair (HDR). Genome editing approaches made possible by CRISPR-based genome editing systems may revolutionize biology, and hold immense promise as tools for the treatment and prevention of disease. However, the activity of RNA-guided nucleases may be reduced or eliminated at certain genomic sites and/or

in certain cell types, limiting the circumstances in which genome editing strategies may be successfully implemented.

[0006] The differences in RNA-guided nuclease activity across loci and cell types may be explained, in part, by differences in chromatin structure between them. Without wishing to be bound by any theory, it is thought that genome editing system activity may be decreased at loci where chromatin structure limits the access of RNA-guided nucleases to target regions in the DNA relative to other loci having less restrictive chromatin structure. Genome editing systems and methods that permit the modification of restrictive chromatin structure may facilitate the editing of sites that have previously been difficult to edit, but such systems and methods have not previously been described in the art.

[0007] Additionally, hemoglobin (Hb) carries oxygen in erythrocytes or red blood cells (RBCs) from the lungs to tissues. During prenatal development and until shortly after birth, hemoglobin is present in the form of fetal hemoglobin (HbF), a tetrameric protein composed of two alpha (α)-globin chains and two gamma (γ)-globin chains. HbF is largely replaced by adult hemoglobin (HbA), a tetrameric protein in which the γ -globin chains of HbF are replaced with beta (β)-globin chains, through a process known as globin switching. The average adult makes less than 1% HbF out of total hemoglobin (Thein 2009). The α -hemoglobin gene is located on chromosome 16, while the β -hemoglobin gene (*HBB*), A gamma (γ A)-globin chain (*HBG1*, also known as gamma globin A), and G gamma (γ G)-globin chain (*HBG2*, also known as gamma globin G) are located on chromosome 11 within the globin gene cluster (also referred to as the globin locus).

[0008] Mutations in *HBB* can cause hemoglobin disorders (i.e., hemoglobinopathies) including sickle cell disease (SCD) and beta-thalassemia (β -Thal). Approximately 93,000 people in the United States are diagnosed with a hemoglobinopathy. Worldwide, 300,000 children are born with hemoglobinopathies every year (Angastiniotis 1998). Because these conditions are associated with *HBB* mutations, their symptoms typically do not manifest until after globin switching from HbF to HbA.

[0009] SCD is the most common inherited hematologic disease in the United States, affecting approximately 80,000 people (Brousseau 2010). SCD is most common in people of African ancestry, for whom the prevalence of SCD is 1 in 500. In Africa, the prevalence of SCD is 15 million (Aliyu 2008). SCD is also more common in people of Indian, Saudi Arabian and

Mediterranean descent. In those of Hispanic-American descent, the prevalence of sickle cell disease is 1 in 1,000 (Lewis 2014).

[0010] SCD is caused by a single homozygous mutation in the *HBB* gene, c.17A>T (HbS mutation). The sickle mutation is a point mutation (GAG>GTG) on *HBB* that results in substitution of valine for glutamic acid at amino acid position 6 in exon 1. The valine at position 6 of the β -hemoglobin chain is hydrophobic and causes a change in conformation of the β -globin protein when it is not bound to oxygen. This change of conformation causes HbS proteins to polymerize in the absence of oxygen, leading to deformation (i.e., sickling) of RBCs. SCD is inherited in an autosomal recessive manner, so that only patients with two HbS alleles have the disease. Heterozygous subjects have sickle cell trait, and may suffer from anemia and/or painful crises if they are severely dehydrated or oxygen deprived.

[0011] Sickle shaped RBCs cause multiple symptoms, including anemia, sickle cell crises, vaso-occlusive crises, aplastic crises, and acute chest syndrome. Sickle shaped RBCs are less elastic than wild-type RBCs and therefore cannot pass as easily through capillary beds and cause occlusion and ischemia (i.e., vaso-occlusion). Vaso-occlusive crisis occurs when sickle cells obstruct blood flow in the capillary bed of an organ leading to pain, ischemia, and necrosis. These episodes typically last 5-7 days. The spleen plays a role in clearing dysfunctional RBCs, and is therefore typically enlarged during early childhood and subject to frequent vaso-occlusive crises. By the end of childhood, the spleen in SCD patients is often infarcted, which leads to autosplenectomy. Hemolysis is a constant feature of SCD and causes anemia. Sickle cells survive for 10-20 days in circulation, while healthy RBCs survive for 90-120 days. SCD subjects are transfused as necessary to maintain adequate hemoglobin levels. Frequent transfusions place subjects at risk for infection with HIV, Hepatitis B, and Hepatitis C. Subjects may also suffer from acute chest crises and infarcts of extremities, end organs, and the central nervous system.

[0012] Subjects with SCD have decreased life expectancies. The prognosis for patients with SCD is steadily improving with careful, life-long management of crises and anemia. As of 2001, the average life expectancy of subjects with sickle cell disease was the mid-to-late 50's. Current treatments for SCD involve hydration and pain management during crises, and transfusions as needed to correct anemia.

[0013] Thalassemias (e.g., β -Thal, δ -Thal, and β/δ -Thal) cause chronic anemia. β -Thal is estimated to affect approximately 1 in 100,000 people worldwide. Its prevalence is higher in certain populations, including those of European descent, where its prevalence is approximately 1 in 10,000. β -Thal major, the more severe form of the disease, is life-threatening unless treated with lifelong blood transfusions and chelation therapy. In the United States, there are approximately 3,000 subjects with β -Thal major. β -Thal intermedia does not require blood transfusions, but it may cause growth delay and significant systemic abnormalities, and it frequently requires lifelong chelation therapy. Although HbA makes up the majority of hemoglobin in adult RBCs, approximately 3% of adult hemoglobin is in the form of HbA₂, an HbA variant in which the two γ -globin chains are replaced with two delta (Δ)-globin chains. δ -Thal is associated with mutations in the Δ hemoglobin gene (*HBD*) that cause a loss of *HBD* expression. Co-inheritance of the *HBD* mutation can mask a diagnosis of β -Thal (i.e., β/δ -Thal) by decreasing the level of HbA₂ to the normal range (Bouva 2006). β/δ -Thal is usually caused by deletion of the *HBB* and *HBD* sequences in both alleles. In homozygous ($\delta\delta/\delta\delta$ $\beta\beta/\beta\beta$) patients, *HBB* is expressed, leading to production of HbF alone.

[0014] Like SCD, β -Thal is caused by mutations in the *HBB* gene. The most common *HBB* mutations leading to β -Thal are: c.-136C>G, c.92+1G>A, c.92+6T>C, c.93-21G>A, c.118C>T, c.316-106C>G, c.25_26delAA, c.27_28insG, c.92+5G>C, c.118C>T, c.135delC, c.315+1G>A, c.-78A>G, c.52A>T, c.59A>G, c.92+5G>C, c.124_127delTTCT, c.316-197C>T, c.-78A>G, c.52A>T, c.124_127delTTCT, c.316-197C>T, c.-138C>T, c.-79A>G, c.92+5G>C, c.75T>A, c.316-2A>G, and c.316-2A>C. These and other mutations associated with β -Thal cause mutated or absent β -globin chains, which causes a disruption of the normal Hb α -hemoglobin to β -hemoglobin ratio. Excess α -globin chains precipitate in erythroid precursors in the bone marrow.

[0015] In β -Thal major, both alleles of *HBB* contain nonsense, frameshift, or splicing mutations that leads to complete absence of β -globin production (denoted β^0/β^0). β -Thal major results in severe reduction in β -globin chains, leading to significant precipitation of α -globin chains in RBCs and more severe anemia.

[0016] β -Thal intermedia results from mutations in the 5' or 3' untranslated region of *HBB*, mutations in the promoter region or polyadenylation signal of *HBB*, or splicing mutations within the *HBB* gene. Patient genotypes are denoted $\beta\delta/\beta^+$ or β^+/β^+ . $\beta\delta$ represents absent expression of a β -globin chain; β^+ represents a dysfunctional but present β -globin chain. Phenotypic

expression varies among patients. Since there is some production of β -globin, β -Thal intermedia results in less precipitation of α -globin chains in the erythroid precursors and less severe anemia than β -Thal major. However, there are more significant consequences of erythroid lineage expansion secondary to chronic anemia.

[0017] Subjects with β -Thal major present between the ages of 6 months and 2 years, and suffer from failure to thrive, fevers, hepatosplenomegaly, and diarrhea. Adequate treatment includes regular transfusions. Therapy for β -Thal major also includes splenectomy and treatment with hydroxyurea. If patients are regularly transfused, they will develop normally until the beginning of the second decade. At that time, they require chelation therapy (in addition to continued transfusions) to prevent complications of iron overload. Iron overload may manifest as growth delay or delay of sexual maturation. In adulthood, inadequate chelation therapy may lead to cardiomyopathy, cardiac arrhythmias, hepatic fibrosis and/or cirrhosis, diabetes, thyroid and parathyroid abnormalities, thrombosis, and osteoporosis. Frequent transfusions also put subjects at risk for infection with HIV, hepatitis B and hepatitis C.

[0018] β -Thal intermedia subjects generally present between the ages of 2-6 years. They do not generally require blood transfusions. However, bone abnormalities occur due to chronic hypertrophy of the erythroid lineage to compensate for chronic anemia. Subjects may have fractures of the long bones due to osteoporosis. Extramedullary erythropoiesis is common and leads to enlargement of the spleen, liver, and lymph nodes. It may also cause spinal cord compression and neurologic problems. Subjects also suffer from lower extremity ulcers and are at increased risk for thrombotic events, including stroke, pulmonary embolism, and deep vein thrombosis. Treatment of β -Thal intermedia includes splenectomy, folic acid supplementation, hydroxyurea therapy, and radiotherapy for extramedullary masses. Chelation therapy is used in subjects who develop iron overload.

[0019] Life expectancy is often diminished in β -Thal patients. Subjects with β -Thal major who do not receive transfusion therapy generally die in their second or third decade. Subjects with β -Thal major who receive regular transfusions and adequate chelation therapy can live into their fifth decade and beyond. Cardiac failure secondary to iron toxicity is the leading cause of death in β -Thal major subjects due to iron toxicity.

[0020] A variety of new treatments are currently in development for SCD and β -Thal. Delivery of an anti-sickling *HBB* gene via gene therapy is currently being investigated in clinical trials.

However, the long-term efficacy and safety of this approach is unknown. Transplantation with hematopoietic stem cells (HSCs) from an HLA-matched allogeneic stem cell donor has been demonstrated to cure SCD and β -Thal, but this procedure involves risks including those associated with ablation therapy, which is required to prepare the subject for transplant, increases risk of life-threatening opportunistic infections, and risk of graft vs. host disease after transplantation. In addition, matched allogeneic donors often cannot be identified. Thus, there is a need for improved methods of managing these and other hemoglobinopathies.

SUMMARY

[0021] The present disclosure provides genome editing systems and related methods which allow for the modulation of indel signatures and frequency of certain indels mediated by genome editing systems, and guide RNAs (gRNAs), and CRISPR-mediated methods. Strategies which allow for increasing the frequency of productive indels are also provided. Provided herein are genome editing systems, gRNAs, and CRISPR-mediated methods for altering one or more genes.

[0022] Provided herein are genome editing systems, gRNAs, and CRISPR-mediated methods for altering one or more γ -globin genes (e.g., *HBG1*, *HBG2*, or *HBG1* and *HBG2*), the erythroid specific enhancer of the *BCL11A* gene (*BCL11Ae*), or a combination thereof, and increasing expression of fetal hemoglobin (HbF). In certain embodiments, genome editing systems, gRNAs, and CRISPR-mediated methods may alter a 13 nucleotide (nt) target region that is 5' of the transcription site of the *HBG1*, *HBG2*, or *HBG1* and *HBG2* gene ("13 nt target region"). In certain embodiments, one or more gRNAs comprising a targeting domain set forth in SEQ ID NOs:251-901 or 940-942 may be used to introduce alterations in the 13 nt target region. In certain embodiments, genome editing systems, gRNAs, and CRISPR-mediated methods may alter a GATA1 binding motif in *BCL11Ae* that is in the +58 DNase I hypersensitive site (DHS) region of intron 2 of the *BCL11A* gene ("GATA1 binding motif in *BCL11Ae*"). In certain embodiments, one or more gRNAs comprising a targeting domain set forth in SEQ ID NOs:952-955 may be used to introduce alterations in the GATA1 binding motif in *BCL11Ae*. In certain embodiments, one or more gRNAs may be used to introduce alterations in the GATA1 binding motif in *BCL11Ae* and one or more gRNAs may be used to introduce alterations in the 13 nt target region of *HBG1* and/or *HBG2*. In certain embodiments, genome editing systems, gRNAs, and CRISPR-mediated methods may alter a region within 50, 100, 200, 300, 400, or 500 bp of a

proximal *HBG1/2* promoter sequence including the 13 nt target region ("proximal HBG1/2 promoter target sequence").

[0023] The inventors have also addressed a key unmet need in the field by identifying a strategy for increasing accessibility to the chromatin using an RNA-guided helicase and dead guide RNA to unwind the DNA within or proximal to the target region to be edited (*e.g.*, the 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*). This disclosure provides new and effective means of unwinding chromatin and thereby increasing accessibility of target regions to RNA-guided nucleases. Also provided herein are genome editing systems, guide RNAs, and CRISPR-mediated methods for unwinding and altering portions of a genome. Unwinding of the genome may be achieved using an RNA-guided helicase and/or a dead guide RNA configured to target an RNA-guided enzyme to a target region in DNA but not to support a cleavage event.

[0024] In one aspect, the disclosure relates to genome editing systems that may include an RNA-guided nuclease, a first guide RNA and a second guide RNA. In certain embodiments, the first and second guide RNAs may include first and second targeting domains complimentary to first and second sequences on opposite sides of positions of a 13 nt target region of a human *HBG1* or *HBG2* gene. One or both of the first and second sequences may overlap the 13 nt target region of the human *HBG1* or *HBG2* gene. The genome editing system may also include a nucleic acid template encoding a deletion of the 13 nt region of the human *HBG1* or *HBG2* gene. In certain embodiments, the RNA-guided nuclease may be an *S. pyogenes* Cas9 or a nickase, which optionally lacks RuvC activity. The first and second targeting domains may be complimentary to sequences immediately adjacent to a protospacer adjacent motif recognized by *S. pyogenes* Cas9. In certain embodiments, the first targeting domain may be complimentary to a sequence within positions c. -1,114 to -114 of a human *HBG1* or *HBG2* gene. In certain embodiments, at least one of the first and second targeting domains differ by no more than 3 nucleotides from a targeting domain listed in Table 7. The genome editing system may include first and second RNA-guided nucleases that, in some embodiments, are complexed with the first and second guide RNAs, respectively, forming first and second ribonucleoprotein complexes.

[0025] Continuing with this aspect of the disclosure, a genome editing system including any or all of the features described above may also include a third guide RNA, and optionally a fourth guide RNA. In certain embodiments, the third and fourth guide RNAs may include third and

fourth targeting domains complimentary to third and fourth sequences on opposite sides of positions of a GATA1 binding motif in *BCL11A* erythroid enhancer (*BCL11Ae*) of a human *BCL11A* gene. One or both of the third and fourth sequences may optionally overlap the GATA1 binding motif in *BCL11Ae* of the human *BCL11A* gene. The genome editing systems may also include a nucleic acid template encoding a deletion of the GATA1 binding motif in *BCL11Ae*. In certain embodiments, the RNA-guided nuclease may be an *S. pyogenes* Cas9. In certain embodiments, the third and fourth targeting domains may be complimentary to sequences immediately adjacent to a protospacer adjacent motif recognized by *S. pyogenes* Cas9. In certain embodiments, the RNA-guided nuclease may be a nickase, which optionally lacks RuvC activity. In certain embodiments, the third targeting domain may be complimentary to a sequence within 1000 nucleotides upstream of the GATA1 binding motif in *BCL11Ae*. In certain embodiments, the third targeting domain may be complimentary to a sequence within 100 nucleotides upstream of the GATA1 binding motif in *BCL11Ae*. In certain embodiments, one of the third and fourth targeting domains may be complimentary to a sequence within 100 nucleotides downstream of the GATA1 binding motif in *BCL11Ae*. In certain embodiments, the fourth targeting domain may be complimentary to a sequence within 50 nucleotides downstream of the GATA1 binding motif in *BCL11Ae*. In certain embodiments, at least one of the third and fourth targeting domains differ by no more than 3 nucleotides from a targeting domain listed in Table 9. In certain embodiments, the genome editing systems may further include first and second RNA-guided nucleases. In certain embodiments, the first and second RNA-guided nucleases may be complexed with the third and fourth guide RNAs, respectively, forming third and fourth ribonucleoprotein complexes.

[0026] Continuing with this aspect of the disclosure, a genome editing system including any or all of the features described above may also include an RNA-guided helicase. In certain embodiments, the RNA-guided helicase may unwind nucleic acid within or proximate to the 13 nt target region or GATA1 binding motif in *BCL11Ae* of the human *BCL11A* gene. In certain embodiments, the RNA-guided helicase may be a fifth RNA-guided nuclease configured to lack nuclease activity. In certain embodiments, the RNA-guided nuclease may be complexed to a dead guide RNA including a fifth targeting domain of 15 or fewer nucleotides in length. In certain embodiments, the RNA-guided nuclease and dead guide RNA are not configured to recruit an exogenous trans-acting factor to the target region. In certain embodiments, the fifth

targeting domain may be complimentary to a fifth sequence within or proximate to the 13 nt target region or GATA1 binding motif in *BCL11Ae* of the human *BCL11A* gene. In certain embodiments, the fifth targeting domain may include a nucleotide sequence that is identical to, or differs by no more than 1, 2, 3, 4, or 5 nucleotides from a nucleotide sequence set forth in Table 10. In certain embodiments, the fifth targeting domain may include a nucleotide sequence identical to the nucleotide sequence set forth in Table 10.

[0027] Another aspect of the disclosure relates to a method of altering a cell including contacting a cell with the genome editing systems described above and disclosed herein. In certain embodiments, the step of contacting the cell with the genome editing system may include contacting the cell with a solution including first and second ribonucleoprotein complexes. In certain embodiments, the step of contacting the cell with the solution may further include electroporating the cells, thereby introducing the first and second ribonucleoprotein complexes into the cell. In certain embodiments, the genome editing systems may further include contacting the cell with the genome editing system described above, in which the step of contacting the cell with the genome editing system may include contacting the cell with a solution including first, second, third, and optionally, fourth ribonucleoprotein complexes. In certain embodiments, the step of contacting the cell with the solution may further include electroporating the cells, thereby introducing the first, second, third, and optionally, fourth ribonucleoprotein complexes into the cell. In certain embodiments, the cell may be capable of differentiating into an erythroblast or a precursor of an erythroblast. In certain embodiments, the cell may be a CD34⁺ cell.

[0028] In one aspect, the disclosure relates to a CRISPR-mediated method of altering a cell including introducing a first DNA single strand break (SSB) or double strand break (DSB) within a genome of the cell between positions c. -614 to -102 of a human *HBG1* or *HBG2* gene and introducing a second SSB or DSB within the genome of the cell between positions c. -114 to -1 of the human *HBG1* or *HBG2* gene. In certain embodiments, the first and second SSBs or DSBs may be repaired by the cell in a manner that alters a 13 nt target region of the human *HBG1* or *HBG2* gene. In certain embodiments, the first and second SSBs or DSBs may be repaired by the cell in a manner that results in the deletion of all or part of a 13 nt target region of the human *HBG1* or *HBG2* gene. In certain embodiments, the first and second SSBs or DSBs may be repaired by the cell in a manner that results in the formation of at least one of an indel, a deletion,

or an insertion in the 13 nt target region of the human *HBG1* or *HBG2* gene. In certain embodiments, the first and second SSBs or DSBs may be repaired by the cell in an error prone manner. In certain embodiments, the CRISPR-mediated method may further include introducing a third DNA single strand break (SSB) or double strand break (DSB) within 500 nucleotides upstream of a GATA1 binding motif in *BCL11Ae* of a human *BCL11A* gene and introducing a fourth SSB or DSB within the genome of the cell within 100 nucleotides downstream of the GATA1 binding motif in *BCL11Ae* of the human *BCL11A* gene. In certain embodiments, the third and fourth SSBs or DSBs may be repaired by the cell in a manner that alters the GATA1 binding motif in *BCL11Ae* of the human *BCL11A* gene. In certain embodiments, the third and fourth SSBs or DSBs may be repaired by the cell in a manner that results in the deletion of all or part of the GATA1 binding motif in *BCL11Ae*. In certain embodiments, the third and fourth SSBs or DSBs may be repaired by the cell in a manner that results in the formation of at least one of an indel, a deletion, or an insertion in the GATA1 binding motif in *BCL11Ae*. In certain embodiments, the third and fourth SSBs or DSBs may be repaired by the cell in an error prone manner.

[0029] In one aspect, the disclosure relates to compositions including a plurality of cells generated by the method disclosed above, in which at least 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90% of the cells include an alteration of a sequence of a 13 nt target region of the human *HBG1* or *HBG2* gene or a plurality of cells generated by the method disclosed above, wherein at least 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90% of the cells include an alteration of a sequence of a 13 nt target region of the human *HBG1* or *HBG2* gene and at least 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90% of the cells include an alteration of a sequence of the GATA1 binding motif in *BCL11Ae*. In certain embodiments, at least a portion of the plurality of cells may be within an erythroid lineage. In certain embodiments, the plurality of cells may be characterized by an increased level of fetal hemoglobin expression relative to an unmodified plurality of cells. In certain embodiments, the level of fetal hemoglobin may be increased by at least 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90%. In certain embodiments, the compositions may further include a pharmaceutically acceptable carrier.

[0030] In one aspect, the disclosure relates to a method of altering a cell, that includes unwinding a chromatin segment within or proximal to a target region of a nucleic acid in a cell and generating a double stranded break (DSB) within the target region of the nucleic acid

whereby to alter the target region. In certain embodiments, the step of unwinding the chromatin segment may include contacting the chromatin segment with an RNA-guided helicase. In certain embodiments, the step of unwinding the chromatin does not include recruiting an exogenous trans-acting factor to the chromatin segment. The RNA-guided helicase may be an RNA-guided nuclease, and the RNA-guided nuclease may be complexed to a dead guide RNA (dgRNA) including a first targeting domain sequence of 15 or fewer nucleotides in length. In certain embodiments, the dgRNA may include modifications at the 5' or 3' end, including, but not limited to, an anti-reverse cap analog (ARCA) at the 5' end of the RNA, a polyA tail at the 3' end of the RNA, or both. In certain embodiments, the RNA-guided nuclease may be an enzymatically active RNA-guided nuclease or may be configured to lack nuclease activity. In certain embodiments, the targeting domain sequence of the dgRNA may be complementary to a sequence proximal to the target region. "Proximal to," in some embodiments herein, may mean within 10, 25, 50, 100, or 200 nucleotides of the target region. In certain embodiments, the step of unwinding the chromatin segment may not include forming a single or double-stranded break in the nucleic acid within the chromatin segment. In certain embodiments, the step of generating the DSB within the target region may include contacting the chromatin segment with an RNA-guided nuclease having nuclease activity. In certain embodiments, the RNA-guided nuclease having nuclease activity may be complexed to a gRNA including a targeting domain configured to overlap the target region.

[0031] Another aspect of the disclosure includes a method of inducing accessibility to a target region of a nucleic acid for editing in a cell including contacting the cell with an RNA-guided helicase and a dgRNA and unwinding DNA within or proximal to the target region with the RNA-guided helicase thereby inducing accessibility to the target region for editing. In various cases, the RNA-guided helicase and the dgRNA may be configured to associate within or proximal to the target region. In certain embodiments, the dgRNA may be configured such that it does not provide an RNA-guided nuclease cleavage event. In certain embodiments, the RNA-guided helicase and dgRNA may complex to form a dead ribonucleoprotein (RNP) that lacks cleavage activity. In certain embodiments, the dgRNA may include a targeting domain sequence of 15 or fewer nucleotides in length. In certain embodiments, the RNA-guided helicase may be an RNA-guided nuclease. In certain embodiments, the RNA-guided nuclease and dgRNA are not configured to recruit an exogenous trans-acting factor to the target region. In certain

embodiments, the RNA-guided nuclease may be a Cas9 or a Cas9-fusion protein. In certain embodiments, the Cas9 may be an enzymatically active Cas9 or an enzymatically dead Cas9. In certain embodiments, the step of unwinding the DNA does not comprise forming a single or double-stranded break in the DNA. In certain embodiments, the RNA-guided nuclease having nuclease activity may be complexed to a gRNA including a targeting domain configured to overlap the target region.

[0032] In another aspect, the disclosure relates to a method of increasing a rate of indel formation in a nucleic acid that includes unwinding double stranded DNA within or proximal to a target region of the nucleic acid using an RNA-guided helicase configured to associate within or proximal to the target region and generating a DSB within the target region. In certain embodiments, generating a DSB within the target region results in forming an indel at the target region. In certain embodiments, the DSB may be repaired in a manner forming an indel at the target region. In certain embodiments, the rate of indel formation in the gene achieved using the RNA-guided helicase is increased compared to a rate of indel formation in the gene achieved without using the RNA-guided helicase. In certain embodiments, the RNA-guided helicase may form an RNP complex with a dgRNA configured to associate within or proximal to the target region. In certain embodiments, the dgRNA may include a targeting domain sequence of 15 nucleotides or less in length. In certain embodiments, the RNA-guided helicase may be an RNA-guided nuclease. In certain embodiments, the RNA-guided nuclease may be a Cas9 or a Cas9-fusion protein. In certain embodiments, the Cas9 may be an enzymatically active Cas9 or an enzymatically dead Cas9. In certain embodiments, the RNA-guided nuclease and the dgRNA are not configured to recruit an exogenous trans-acting factor to the target region. In certain embodiments, the step of unwinding the double stranded DNA does not include forming a single or double-stranded break in the DNA.

[0033] In yet another aspect, this disclosure relates to a method of deleting a segment of a target nucleic acid in a cell that includes contacting the cell with an RNA-guided helicase and generating a DSB within the target region, whereby a segment of the target nucleic acid is deleted. In certain embodiments, the DSB may be repaired in a manner that deletes a segment of the target nucleic acid. In certain embodiments, the RNA-guided helicase may be configured to associate within or proximal to a target region of the target nucleic acid and unwind double stranded DNA (dsDNA) within or proximal to the target region. In certain embodiments, the

RNA-guided helicase may form a ribonucleoprotein complex with a dgRNA configured to associate within or proximal to the target region. In certain embodiments, the dgRNA may include a targeting domain sequence of 15 nucleotides or less in length. In certain embodiments, the RNA-guided helicase may be an RNA-guided nuclease. In certain embodiments, the RNA-guided nuclease may be a Cas9 or a Cas9-fusion protein. In certain embodiments, the Cas9 may be an enzymatically active Cas9 or an enzymatically dead Cas9. In certain embodiments, the RNA-guided nuclease and the dgRNA are not configured to recruit an exogenous trans-acting factor to the target region. In certain embodiments, the target nucleic acid may be a promoter region of a gene, a coding region of a gene, a non-coding region of a gene, an intron of a gene, or an exon of a gene. In certain embodiments, the segment of the target nucleic acid may be at least about 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, or 100 base pairs in length.

[0034] The disclosure also relates to a dead gRNA (dgRNA) molecule including a targeting domain comprising a truncation of a gRNA targeting domain. In certain embodiments, the gRNA targeting domain to be truncated may be a gRNA targeting domain set forth in Table 2 or Table 10. In certain embodiments, the gRNA targeting domain may be truncated from a 5' end of the gRNA targeting domain. In certain embodiments, the dgRNA may include a targeting domain sequence of 15 nucleotides or less in length. In certain embodiments, the first targeting domain may be the same as or may differ by no more than 3 nucleotides from a dgRNA targeting domain set forth in Table 10.

[0035] Another aspect of the disclosure relates to compositions including at least one polynucleotide encoding a plurality of gRNAs and an RNA-guided helicase, in which at least one gRNA may be a dgRNA configured such that it does not provide an RNA-guided nuclease cleavage event. In certain embodiments, the dgRNA may include a targeting domain sequence of 15 nucleotides or less in length. In certain embodiments, the RNA-guided helicase may be an RNA-guided nuclease. In certain embodiments, the RNA-guided nuclease may be a Cas9 or a Cas9-fusion protein. In certain embodiments, the Cas9 may be an enzymatically active Cas9 or an enzymatically dead Cas9. In certain embodiments, the RNA-guided nuclease and the dgRNA are not configured to recruit an exogenous trans-acting factor to the target region. In certain embodiments, the compositions further include a second RNA-guided nuclease configured to provide a cleavage event. In certain embodiments, the compositions further include a second gRNA configured to provide a cleavage event.

[0036] In another aspect, the disclosure relates to genome editing systems that include an RNA-guided nuclease and an RNA-guided helicase configured to associate with a target nucleic acid proximal to a target region of the target nucleic acid and induce a conformational change in the target region thereby promoting accessibility to the target region for the RNA-guided nuclease to form a break in the target region. The disclosure also relates to genome editing systems that include a dgRNA including a targeting domain sequence of 15 nucleotides or less in length, a first RNA-guided nuclease, and an RNA-guided helicase. In certain embodiments, the genome editing system further includes a gRNA. In certain embodiments, the gRNA and the first RNA-guided nuclease may associate with a target region in a target nucleic acid. In certain embodiments, the gRNA and the first RNA-guided nuclease may associate with a first PAM sequence in a target nucleic acid, wherein the first PAM sequence is facing outward. In certain embodiments, the RNA-guided helicase may be a second RNA-guided nuclease. In certain embodiments, the second RNA-guided nuclease and the dgRNA are not configured to recruit an exogenous trans-acting factor to the target region. In certain embodiments, the dgRNA and the second RNA-guided nuclease associate within or proximal to a target region in the target nucleic acid. In certain embodiments, the first RNA-guided nuclease and second RNA-guided nuclease may be complexed with the gRNA and dgRNAs, respectively, forming first and second ribonucleoprotein complexes.

[0037] In another aspect, the disclosure relates to a genome editing system that includes a dgRNA comprising a targeting domain sequence of 15 nucleotides or less in length, a first RNA-guided nuclease, and an RNA-guided helicase. In certain embodiments, the gRNA and the first RNA-guided nuclease may associate with a target region in a target nucleic acid. In certain embodiments, the gRNA and the first RNA-guided nuclease may associate with a first protospacer adjacent motif (PAM) sequence in a target nucleic acid. In certain embodiments, the first PAM sequence may be facing outward. In certain embodiments, the RNA-guided helicase may be a second RNA-guided nuclease. In certain embodiments, the second RNA-guided nuclease and the dgRNA are not configured to recruit an exogenous trans-acting factor to the target region. In certain embodiments, the dgRNA and the second RNA-guided nuclease may associate within or proximal to a target region in the target nucleic acid. In certain embodiments, the first RNA-guided nuclease and second RNA-guided nuclease may be complexed with the gRNA and dgRNAs, respectively, forming first and second ribonucleoprotein complexes. In

certain embodiments, the dgRNA and the second RNA-guided nuclease may associate with a second PAM sequence in a target nucleic acid, wherein the second PAM sequence may be facing outward.

[0038] In another aspect, the disclosure relates to a genome editing system that includes a dgRNA, a first gRNA comprising a second targeting domain sequence greater than 17 nucleotides in length, and an RNA-guided nuclease. In certain embodiments, the RNA-guided nuclease and the dgRNA may be configured to associate within a first target region in a target nucleic acid. In certain embodiments, the RNA-guided nuclease and the first gRNA may be configured to associate within a second target region and generate a double stranded break (DSB) in the target nucleic acid whereby to create an indel between the first target region and the second target region. In certain embodiments, the dgRNA may comprise a first targeting domain sequence of 15 nucleotides or less in length. In certain embodiments, the dgRNA has reduced or no RNA-guided nuclease cleavage activity. In certain embodiments, the dgRNA may be configured such that it does not provide an RNA-guided nuclease cleavage event. In certain embodiments, the dgRNA and the RNA-guided nuclease may associate with a first protospacer adjacent motif (PAM) sequence in the target nucleic acid. In certain embodiments, the first PAM sequence may be facing outward. In certain embodiments, the first gRNA and the RNA-guided nuclease may associate with a second PAM sequence in the target nucleic acid. In certain embodiments, the second PAM sequence may be facing outward.

[0039] In another aspect, the disclosure relates to a method of altering a cell, including contacting the cell with a dgRNA, a first gRNA comprising a second targeting domain sequence greater than 17 nucleotides in length, and an RNA-guided nuclease. In certain embodiments, the RNA-guided nuclease and the dgRNA may be configured to associate within a first target region in a target nucleic acid. In certain embodiments, the RNA-guided nuclease and the first gRNA may associate within a second target region and generate a double stranded break (DSB) in the target nucleic acid whereby to create an indel between the first target region and the second target region. In certain embodiments, the dgRNA may comprise a first targeting domain sequence of 15 nucleotides or less in length. In certain embodiments, the dgRNA has reduced or no RNA-guided nuclease cleavage activity. In certain embodiments, the dgRNA may be configured such that it does not provide an RNA-guided nuclease cleavage event. In certain embodiments, the dgRNA and the RNA-guided nuclease may associate with a first protospacer

adjacent motif (PAM) sequence in the target nucleic acid. In certain embodiments, the first PAM sequence may be facing outward. In certain embodiments, the first gRNA and the RNA-guided nuclease may associate with a second PAM sequence in the target nucleic acid. In certain embodiments, the second PAM sequence may be facing outward.

[0040] The disclosure herein also relates to methods of altering a cells, including contacting a cell with any of the genome editing systems disclosed herein. In certain embodiments, the step of contacting the cell may comprise contacting the cell with a solution comprising first and second ribonucleoprotein complexes. In certain embodiments, the step of contacting the cell with the solution further comprises electroporating the cells, thereby introducing the first and second ribonucleoprotein complexes into the cell.

[0041] In another aspect, the disclosure relates to cells that are altered using the methods disclosed herein. Cells that include a productive indel which results in HbF expression are also disclosed herein. In certain embodiments the indel may be produced by contacting the cell with a dgRNA, a first gRNA including a second targeting domain sequence greater than 17 nucleotides in length, and an RNA-guided nuclease. In certain embodiments, the RNA-guided nuclease and the dgRNA may be configured to associate within a first target region in a target nucleic acid. In certain embodiments, the RNA-guided nuclease and the first gRNA may associate with a second target region and generate a double stranded break (DSB) in the target nucleic acid whereby to create an indel between the first target region and the second target region. In certain embodiments, the cells disclosed herein, may be capable of differentiating into an erythroblast, erythrocyte, or a precursor of an erythrocyte or erythroblast. In certain embodiments, the cell may be a CD34⁺ cell.

[0042] A genome editing system or method including any of all of the features described above may include a target nucleic acid comprising a human *HBG1*, *HBG2* gene, or a combination thereof. In certain embodiments, the target region may be a CCAAT box target region of the human *HBG1*, *HBG2* gene, or a combination thereof. In certain embodiments, the first targeting domain sequence may be complementary to a first sequence on a side of a CCAAT box target region of the human *HBG1*, *HBG2* gene, or a combination thereof, in which the first sequence optionally overlaps the CCAAT box target region of the human *HBG1*, *HBG2* gene, or a combination thereof. In certain embodiments, the second targeting domain sequence may be complementary to a second sequence on a side of a CCAAT box target region of the human

HBG1, *HBG2* gene, or a combination thereof, in which the second sequence optionally overlaps the CCAAT box target region of the human *HBG1*, *HBG2* gene, or a combination thereof. In certain embodiments, the first targeting domain may comprise a truncation of a gRNA targeting domain. In certain embodiments, the gRNA targeting domain may include the gRNAs set forth in Table 2 or Table 10, and the gRNA targeting domain has been truncated from a 5' end of the gRNA targeting domain. In certain embodiments, the first targeting domain may be the same as or differs by no more than 3 nucleotides from a dgRNA targeting domain set forth in Table 10. In certain embodiments, the second targeting domain differs by no more than 3 nucleotides from a gRNA targeting domain set forth in Table 2 or Table 10. In certain embodiments, the indel may alter the CCAAT box target region indel. In certain embodiments, the indel may be a productive indel resulting in an increased level of fetal hemoglobin expression. In certain embodiments, the gRNA, dgRNA, or both may be *in vitro* synthesized or chemically synthesized.

[0043] In certain embodiments, a cell may include at least one modified allele of the *HBG* locus generated by any of the methods for altering a cell disclosed herein, in which the modified allele of the *HBG* locus comprises an alteration of the human *HBG1* gene, *HBG2*, gene, or a combination thereof.

[0044] In certain embodiments, an isolated population of cells may be modified by any of the methods for altering a cells disclosed herein, wherein the population of cells may include a distribution of indels that may be different from an isolated population of cells or their progenies of the same cell type that have not been modified by the method.

[0045] In certain embodiments, a plurality of cells may be generated by any of the methods for altering a cells disclosed herein, in which at least 20%, 30%, 40%, 50%, 60%, 70%, 80%, or 90% of the cells may include an alteration of a sequence in the CCAAT box target region of the human *HBG1* gene, *HBG2* gene or a combination thereof.

[0046] In certain embodiments, the cells disclosed herein may be used for a medicament. In certain embodiments, the cells may be for use in the treatment of β -hemoglobinopathy. In certain embodiments, β -hemoglobinopathy may be selected from the group consisting of sickle cell disease and beta-thalassemia.

[0047] In one aspect, the disclosure relates to compositions including a plurality of cells generated by a method including a dgRNA disclosed above, in which at least 20%, 30%, 40%,

50%, 60%, 70%, 80% or 90% of the cells include an alteration of a sequence of a CCAAT box target region of the human *HBG1* or *HBG2* gene or a plurality of cells generated by the method disclosed above, wherein at least 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90% of the cells include an alteration of a sequence of a CCAAT box target region of the human *HBG1* or *HBG2*. In certain embodiments, at least a portion of the plurality of cells may be within an erythroid lineage. In certain embodiments, the plurality of cells may be characterized by an increased level of fetal hemoglobin expression relative to an unmodified plurality of cells. In certain embodiments, the level of fetal hemoglobin may be increased by at least 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90%. In certain embodiments, the compositions may further include a pharmaceutically acceptable carrier.

[0048] In one aspect, the disclosure relates to a population of cells modified by a genome editing system including a dgRNA described above, wherein the population of cells comprise a higher percentage of a productive indel relative to a population of cells not modified by the genome editing system. The disclosure also relates to a population of cells modified by the genome editing system including a dgRNA described above, wherein a higher percentage of the population of cells are capable of differentiating into a population of cells of an erythroid lineage that express HbF relative to a population of cells not modified by the genome editing system. In certain embodiments, the higher percentage may be at least about 15%, at least about 20% , at least about 25%, at least about 30%, or at least about 40% higher. In certain embodiments, the cells may be hematopoietic stem cells. In certain embodiments, the cells may be capable of differentiating into an erythroblast, erythrocyte, or a precursor of an erythrocyte or erythroblast. In certain embodiments, the indel may be created by a repair mechanism other than microhomology-mediated end joining (MMEJ) repair.

[0049] The disclosure also relates to the use of any of the cells disclosed herein in the manufacture of a medicament for treating β -hemoglobinopathy in a subject.

[0050] In one aspect, the disclosure relates to a method of treating a β -hemoglobinopathy in a subject in need thereof, comprising administering to the subject the cells disclosed herein. In certain embodiments, a method of treating a β -hemoglobinopathy in a subject in need thereof, may include administering a population of modified hematopoietic cells to the subject, wherein one or more cells have been altered according to the methods of altering a cell disclosed herein.

[0051] This listing is intended to be exemplary and illustrative rather than comprehensive and limiting. Additional aspects and embodiments may be set out in, or apparent from, the remainder of this disclosure and the claims.

BRIEF DESCRIPTION OF THE DRAWINGS

[0052] The accompanying drawings are intended to provide illustrative, and schematic rather than comprehensive, examples of certain aspects and embodiments of the present disclosure. The drawings are not intended to be limiting or binding to any particular theory or model, and are not necessarily to scale. Without limiting the foregoing, nucleic acids and polypeptides may be depicted as linear sequences, or as schematic two- or three-dimensional structures; these depictions are intended to be illustrative rather than limiting or binding to any particular model or theory regarding their structure.

[0053] **Fig. 1** depicts, in schematic form, *HBG1* and *HBG2* gene(s) in the context of the γ -globin gene cluster on human chromosome 11. **Fig. 1.** Each gene in the γ -globin gene cluster is transcriptionally regulated by a proximal promoter. While not wishing to be bound by any particular theory, it is generally thought that A_γ and/or G_γ expression is activated by engagement between the proximal promoter with the distal strong erythroid-specific enhancer, the locus control region (LCR). Long-range transactivation by the LCR is thought to be mediated by alteration of chromatin configuration/confirmation. The LCR is marked by 4 erythroid specific DNase I hypersensitive sites (HS1-4) and 2 distal enhancer elements (5' HS and 3' HS1). γ -like gene globin gene expression is regulated in a developmental stage-specific manner, and expression of globin genes changes coincide with changes in the main site of blood production.

[0054] **Figs. 2A-2B** depict *HBG1* and *HBG2* genes, coding sequences (CDS) and small deletions and point mutations in and upstream of the *HBG1* and *HBG2* proximal promoters that have been identified in patients and associated with elevation of fetal hemoglobin (HbF). Core elements within the proximal promoters (CAAT box, 13 nt sequence) that have been deleted in some patients with hereditary persistence of fetal hemoglobin (HPFH). The 'target sequence' region of each locus, which has been screened for gRNA binding target sites, is also identified.

[0055] **Fig. 3** shows data from gRNA screening for incorporation of the 13 nt deletion in human K562 erythroleukemia cells. **Fig. 3A** Gene editing as determined by T7E1 endonuclease assay analysis (referred to interchangeably as a "T7E1 analysis") of *HBG1* and *HBG2* locus-specific PCR products amplified from genomic DNA extracted from K562 cells after electroporation

with DNA encoding *S. pyogenes*-specific gRNAs and plasmid DNA encoding *S. pyogenes* Cas9. **Fig. 3B** Gene editing as determined by DNA sequence analysis of PCR products amplified from the *HBG1* locus in genomic DNA extracted from K562 cells after electroporation with DNA encoding the indicated gRNA and Cas9 plasmid. **Fig. 3C** Gene editing as determined by DNA sequence analysis of PCR products amplified from the *HBG2* locus in genomic DNA extracted from K562 cells after electroporation with DNA encoding the indicated gRNA and Cas9 plasmid. For **Fig. 3B-C**, the types of editing events (insertions, deletions) and subtypes of deletions (13 nt target partially [12 nt HPFH] or fully [13-26 nt HPFH] deleted, other sequences deleted [other deletions]) are indicated by the differently shaded/patterned bars.

[0056] Fig. 4 depicts results of gene editing in human cord blood (CB) and human adult CD34⁺ cells after electroporation with RNPs complexed to in vitro transcribed *S. pyogenes* gRNAs that target a specific 13 nt sequence for deletion (*HBG* gRNAs Sp35 and Sp37). **Fig. 4A** depicts the percentage of indels detected by T7E1 analysis of *HBG1* and *HBG2* specific PCR products amplified from gDNA extracted from CB CD34⁺ cells treated with the indicated RNPs or donor matched untreated control cells (n=3 CB CD34⁺ cells, 3 separate experiments). Data shown represent the mean and error bars correspond to standard deviation across the 3 separate donors/experiments. **Fig. 4B** depicts the percentage of indels detected by T7E1 analysis of *HBG2* specific PCR product amplified from gDNA extracted from CB CD34⁺ cells or adult CD34⁺ cells treated with the indicated RNPs or donor matched untreated control cells (n=3 CB CD34⁺ cells, n=3 mobilized peripheral blood (mPB) CD34⁺ cells, 3 separate experiments). Data shown represent the mean and error bars correspond to standard deviation across the 3 separate donors/experiments. **Fig. 4C** (Top panel) depicts indels as detected by T7E1 analysis of *HBG2* PCR products amplified from gDNA extracted from human CB CD34⁺ cells electroporated with HBG Sp35 RNP or HBG Sp37 RNP +/- ssODN (unmodified or with PhTx modified 5' and 3' ends). The lower left panel shows the level of gene editing as determined by Sanger DNA sequence analysis of gDNA from cells edited with HBG Sp37 RNP and ssODN. The lower right panel shows the specific types of deletions detected within total deletions.

[0057] Fig. 5 depicts gene editing of *HBG* in adult human mobilized peripheral blood (mPB) CD34⁺ cells and induction of fetal hemoglobin in erythroid progeny of RNP treated cells after electroporation of mPB CD34⁺ cells with HBG Sp37 RNP +/- ssODN encoding the 13 nt deletion. **Fig. 5A** depicts the percentage of indels detected by T7E1 analysis of *HBG2* PCR

product amplified from gDNA extracted from mPB CD34⁺ cells treated with the RNP or donor matched untreated control cells. **Fig. 5B** depicts the fold change in *HBG* mRNA expression in day 7 erythroblasts that were differentiated from RNP treated and untreated donor matched control mPB CD34⁺ cells. mRNA levels are normalized to GAPDH and calibrated to the levels detected in untreated controls on the corresponding days of differentiation.

[0058] **Fig. 6** depicts the *ex vivo* differentiation potential of RNP treated and untreated mPB CD34⁺ cells from the same donor. **Fig. 6A** shows hematopoietic myeloid/erythroid colony forming cell (CFC) potential, where the number and subtype of colonies are indicated (GEMM: granulocyte-erythroid-monocyte-macrophage colony, E: erythroid colony, GM: granulocyte-macrophage colony, M: macrophage colony, G: granulocyte colony). **Fig. 6B** depicts the percentage of Glycophorin A expressed over the time course of erythroid differentiation as determined by flow cytometry analysis at the indicated time points and for the indicated samples.

[0059] **Fig. 7A** depicts indels detected by T7E1 analysis of *HBG* PCR product amplified from gDNA extracted from human mPB CD34⁺ cells treated with *HBG* RNPs (D10A paired nickases). For a subset of samples, cells also received ssODN encoding the 13 nt deletion plus silent SNPs to monitor for HDR (ssODN). **Fig. 7B** depicts DNA sequencing analysis for select subset of samples shown in **Fig. 7A**. The indels were subdivided according to the type of indel (insertion, 13 nt deletion, or other deletion).

[0060] **Fig. 8A** depicts the indels at the *HBG* target site after electroporation of mPB CD34⁺ cells with the indicated pairs of gRNAs complexed in D10A nickase and WT RNP pairs. **Fig. 8B** depicts the large deletion events (e.g., deletion of *HBG2*) after electroporation of mPB CD34⁺ cells with the indicated pairs of gRNAs complexed in D10A nickase and WT RNPs. **Fig. 8C** depicts DNA sequencing analysis and the subtypes of events (insertions, deletions) detected in gDNA from mPB CD34⁺ cells treated with paired D10A nickase pairs. **Fig. 8D** depicts DNA sequencing analysis and the subtypes of events (insertions, deletions) detected in gDNA from mPB CD34⁺ cells treated with paired WT RNP pairs.

[0061] **Fig. 9** depicts the summary of HbF protein and mRNA expression in the progeny of mPB CD34⁺ cells treated with paired RNPs targeting *HBG*, for the experiments shown in **Figs. 7** and **8**. HbF protein (by HPLC analysis) and HbF mRNA expression (ddPCR analysis) were evaluated in erythroid progeny of RNP treated human mPB CD34⁺ cells (background levels of

HbF detected in donor matched untreated controls were subtracted from the levels detected in progeny of RNP treated CD34⁺ cells).

[0062] Fig. 10 depicts the indel frequencies and *ex vivo* and *in vivo* short-term hematopoietic potential of CD34⁺ cells after treatment with different concentrations (0, 2.5, 3.75 μ M) of paired D10A nickase RNPs (SpA+Sp85). Indels were evaluated by T7E1 analysis (**Fig. 10A**) and by Illumina sequencing analysis (insertions and deletions, **Fig. 10B**). **Fig. 10C** depicts the % of HbF protein detected by HPLC analysis (%HbF = 100% x HbF/(HbF+HbA)). **Fig. 10D** depicts the hematopoietic activity of the RNP treated and donor matched untreated control CD34⁺ cells in colony forming cell (CFC) assays. CFCs shown are per thousand CD34⁺ cells plated. **Fig. 10E** depicts human blood CD45⁺ cell reconstitution of the peripheral blood in immunodeficient mice (NSG) 1 month after transplantation with donor matched human mPB CD34⁺ that were either untreated (0 μ M), or treated with one of two doses (2.5 and 3.75 μ M) of D10A RNP and paired gRNAs. **Fig. 10F** depicts human blood CD45⁺ cell reconstitution of the peripheral blood in immunodeficient mice (NSG) 2 months after transplantation. **Figs. 10G and 10H** depict the lineage distributions following human CD45⁺ blood cell reconstitution of NSG mice at 1 month (**Fig. 10G**) and 2 months (**Fig. 10H**).

[0063] Fig. 11a correlates HbF levels as assayed by HPLC and indel frequency as assessed by T7E1 analysis for two D10A nickase RNP pairs (SP37+SPB and SP37+SPA) delivered at the indicated concentrations to mPB CD34⁺ cells. HbF levels were analyzed in erythroid progeny (day 18) of edited CD34⁺ cells. HbF protein detected in donor-matched untreated controls were subtracted from edited samples. **Fig. 11b** depicts indel rates overlaid on hematopoietic colony forming cell (CFC) activity associated with CD34⁺ cells treated with the indicated D10A nickase pairs or untreated controls. **Fig. 11c** depicts human CD45⁺ blood cell reconstitution of immunodeficient NSG mice one month after transplantation of mPB CD34⁺ cells treated with indicated D10 RNP nickase pairs at the concentrations given or donor matched untreated controls. **Fig. 11d** depicts the human blood lineage distribution detected in the human CD45⁺ fraction in mouse peripheral blood one month post-transplant.

[0064] Fig. 12 depicts a target site for derepression of HbF, the GATA1 motif of the +58 DNase I hypersensitive site (DHS) erythroid specific enhancer of *BCL11A* (*BCL11Ae*) (genomic coordinates: chr2: 60,495,265 to 60,495,270).

[0065] **Fig. 13A** depicts the percentage of indels detected by T7E1 endonuclease analysis of *BCL11A* PCR products amplified from gDNA extracted from CB CD34⁺ cells treated with the indicated RNP +/- ssODN or donor matched untreated control cells. Data shown represent the mean of three 3 separate donors/experiments. **Fig. 13B** depicts indels detected by T7E1 endonuclease analysis of *BCL11A* PCR products amplified from gDNA extracted from CB CD34⁺ cells treated with the indicated WT RNP (single gRNA targeting the *BCL11A* erythroid enhancer complexed to WT *S. pyogenes* Cas9 having both RuvC and HNH activity) or paired nickase RNP (paired gRNAs targeting the *BCL11A* erythroid enhancer (*BCL11Ae*) complexed to *S. pyogenes* Cas9 nickases sharing the same HNH single stranded cutting activity (e.g., D10A), as well as the hematopoietic activity of cells in each condition.

[0066] **Fig. 14A** depicts the editing frequency of *BCL11Ae* (using single gRNA approach targeting the GATA1 motif) in adult human BM CD34⁺ cells. **Fig. 14B** depicts the monoallelic and biallelic editing detected in hematopoietic colonies (GEMMs, clonal progeny of *BCL11Ae* RNP treated CD34⁺ cells) as determined by DNA sequencing analysis. **Fig. 14C** depicts the kinetics of erythroblast maturation (enucleation as determined by DRAQ5⁺ cells detected by flow cytometry analysis). **Fig. 14D** depicts the acquisition of erythroid phenotype (Glycophorin A⁺ cells) in differentiated control and RNP-treated BM CD34⁺ cells, while **Fig. 14E** shows the fold increase in HbF⁺ cells as determined by flow cytometry analysis relative to HbF⁺ cells in untreated donor matched control samples.

[0067] **Fig. 15** depicts gene editing of *BCL11Ae* in adult human mPB CD34⁺ cells and induction of fetal hemoglobin in erythroid progeny of RNP and ssODN treated cells after electroporation of mPB CD34⁺ cells with *BCL11Ae* RNP + nonspecific ssODN (i.e., no homology to *BCL11Ae* target region). **Fig. 15A** depicts the percentage of indels detected by T7E1 analysis of *HBG2* PCR product amplified from gDNA extracted from mPB CD34⁺ cells treated with the *BCL11Ae* RNP and nonspecific ssODN or donor matched untreated control cells. **Fig. 15B** depicts the fold change in *HBG* mRNA expression in day 10 erythroblasts that were differentiated from *BCL11Ae* RNP treated and untreated donor matched control mPB CD34⁺ cells (mRNA levels are normalized to GAPDH and calibrated to the levels detected in untreated controls on the corresponding days of differentiation). **Fig. 15C** depicts the percentage of Glycophorin A expressed over the time course of erythroid differentiation of mPB CD34⁺ cells +/- treatment

with *BCL11Ae* RNP and nonspecific ssODN, as determined by flow cytometry analysis at the indicated time points and for the indicated samples.

[0068] Fig. 16 depicts a schematic of the -110 nt target region in the gamma hemoglobin gene (*HBG*) promoter (grey box) and the relative locations of homologous sequences to dead gRNAs (dgRNAs) and wild-type gRNAs. dgRNAs that have a truncated targeting domain sequence and do not promote Cas9 cutting are depicted (*i.e.*, Sp181 dgRNA and truncated (t)SpA dgRNA, **Table 10**) as white arrows. gRNAs that have a full-length targeting domain sequence, which promote Cas9 cutting are depicted as black arrows (*i.e.*, Sp35 and Sp37 gRNAs, **Table 10**).

[0069] Fig. 17 shows the percentage of edits determined by T7E1 endonuclease analysis of *HBG2* PCR product amplified from genomic DNA (gDNA) extracted from mobilized peripheral blood (mPB) CD34⁺ cells after codelivery of a dead ribonucleoprotein (dRNP) (*i.e.*, SpA dRNP) and a wild-type (WT) RNP (*i.e.*, Sp37 RNP). tSpA dRNP comprises WT Cas9 protein complexed to a truncated gRNA (tSpA dgRNA, **Table 10**) (*i.e.*, dead (d)RNA15-mer version of SpA) and Sp37 RNP comprises WT Cas9 protein complexed to gRNA Sp37 (**Table 10**).

[0070] Fig. 18 depicts the percentage of edits detected by T7E1 analysis of *HBG* PCR product amplified from gDNA extracted from mPB CD34⁺ cells after delivery of Sp35 RNP alone (*i.e.*, Sp35 gRNA complexed with WT Cas 9 protein) or codelivery of Sp35 RNP and dRNPs that target the same or opposite DNA strand as Sp35 RNP (*i.e.*, Sp181 dRNP (Sp181 dgRNA complexed with WT Cas9 protein) and tSpA dRNP (tSpA dgRNA complexed with WT Cas9 protein)) (see also **Fig. 16**). Black bars indicate the level of indels detected in the mPB CD34⁺ cells. White bars indicate the level of indels detected in the mPB CD34⁺ cells maintained in the day 7 erythroid progeny of edited cells.

[0071] Fig. 19 shows a schematic of the variety of insertions and deletions resulting from double strand breaks repaired through NHEJ. Each unique edit (*e.g.*, insertion or deletion) may serve as a unique identifier (or "barcode") for an individual cell or clone of cells descended therefrom.

[0072] Fig. 20 depicts a graphical rank ordering of the most abundant edited alleles in pre-infusion human HSCs and in lineages or tissue populations derived from long-term engrafting cells from two experimental replicates at 16-weeks post-infusion. Genomic DNA from cells electroporated with a ribonucleoprotein complex targeting the HBB locus was harvested and sequencing reads were aligned to an unedited or WT reference sequence. The frequency of individual edited alleles among the total number of reads from each sample was quantified and

ranked. White and grey bars represent to five most abundant unique alleles in each sample, with white bars representing the most abundant unique allele, and less frequent alleles being represented by progressively darker shades of grey. Black bars represent unique alleles outside of the top 5 in terms of frequency. These data indicate that the most frequent alleles in each sample represent a comparatively small fraction of the total reads, and that the distribution of reads varies across lineages or tissue populations derived from the same pre-infusion pool, indicating that diversity of edited alleles is preserved in long-term engrafting HSCs and their progeny.

[0073] Fig. 21 depicts a graphical rank-ordering of the abundance of edited alleles in pre-infusion human HSCs and in lineages or tissue populations derived from long-term engrafting cells in two experimental replicates at 16-weeks post-infusion. Editing and analysis were performed as described for **Fig. 20**, but the white bars correspond to the edited allele observed at the highest frequency in the pre-infusion edited HSC sample, and progressively darker bars correspond to less frequently observed alleles in the pre-infusion sample. Bars of the same color represent the same edited allele in each sample. Black bars represent unique alleles outside of the top 5 in any of the samples shown. The figure indicates that the frequency of individual alleles in tissue populations or lineages derived from long-term engrafting HSCs varies from the frequency of the same alleles in pre-infusion samples, consistent with the relatively low level of representation of long-term engrafting HSCs in the bulk CD34⁺ cell population.

[0074] Fig. 22 depicts, in schematic form, a portion of the *HBG* promoter sequence and the relative locations of homologous sequences to truncated SpA (tSpA) dgRNA (**Table 10**) and Sp37 gRNA (**Table 10**). The -13 nt target region (“-114 to -102 del” GTCAAGGCTATTG) is shown by bolded letters. The sequence of the distal CCAAT box (“ATTGG”) is represented by dashed lines. PAM sequences are underlined.

[0075] Fig. 23 depicts the percentage of indels detected by Next Generation Sequencing (NGS) analysis of *HBG* PCR product amplified from gDNA extracted from mPB CD34⁺ cells after delivery of Sp37 RNP alone (*in vitro* transcribed (IVT) Sp37 gRNA complexed with WT *S. Pyogenes* Cas 9 protein) or codelivery of Sp37 RNP (IVT Sp37 gRNA complexed with WT *S. Pyogenes* Cas 9 protein) and tSpA dRNP (IVT tSpA dgRNA complexed with WT *S. Pyogenes* Cas9 protein). Black bars indicate the level of indels detected in the mPB CD34⁺ cells. Grey circles indicate the percentage of HbF protein expression as determined by UPLC tetramer

analysis of hemoglobin expression in the erythroid progeny of the edited cells. The hemoglobin tetramers, $\alpha_2\beta_2$ or $\alpha_2\gamma_2$, were dissociated by using a gradually increasing acetonitrile gradient which periodically elutes them from the liquid chromatography column. The relative quantity of each subunit was then determined using the area under the curve (AUC), with HbF % representing $AUC (A\gamma + G\gamma) / (A\gamma + G\gamma + \beta)$ or $HbF = \text{Total Gamma} / \text{Total Beta-like}$.

[0076] Fig. 24 depicts the percentage of indels detected by NGS analysis of *HBG* PCR product amplified from gDNA extracted from mPB CD34⁺ cells after delivery of Sp37 RNP alone (chemically synthesized Sp37 gRNA complexed with WT *S. Pyogenes* Cas 9 protein) or codelivery of Sp37 RNP (chemically synthesized Sp37 gRNA complexed with WT *S. Pyogenes* Cas 9 protein) and tSpA dRNP (chemically synthesized tSpA dgRNA complexed with WT *S. Pyogenes* Cas9 protein). Black bars indicate the level of indels detected in the mPB CD34⁺ cells. Grey circles indicate the percentage of HbF protein expression as determined by UPLC tetramer analysis of hemoglobin expression in erythroid progeny of the edited cells. The hemoglobin tetramers, $\alpha_2\beta_2$ or $\alpha_2\gamma_2$, were dissociated by using a gradually increasing acetonitrile gradient which periodically elutes them from the liquid chromatography column. The relative quantity of each subunit was then determined using the area under the curve (AUC), with HbF % representing $AUC (A\gamma + G\gamma) / (A\gamma + G\gamma + \beta)$ or $HbF = \text{Total Gamma} / \text{Total Beta-like}$.

[0077] Fig. 25 depicts the percentage of indels detected by T7E1 analysis of *HBG2* PCR product amplified from gDNA extracted from mPB CD34⁺ cells after codelivery of varying concentrations of Sp37 RNP (IVT Sp37 gRNA complexed with *S. Pyogenes* WT Cas9 protein) and tSpA dRNP (IVT tSpA dgRNA complexed with *S. Pyogenes* WT Cas9 protein).

[0078] Figs. 26A-D depicts the deletion profiles observed with varying concentrations of active Sp37 RNP (IVT Sp37 gRNA complexed with *S. Pyogenes* WT Cas9 protein) codelivered with varying concentrations of inactive tSpA RNP (IVT tSpA gRNA complexed with *S. Pyogenes* WT Cas9 protein). **Fig. 26A** depicts the deletion profile observed with a 1:0 ratio of Sp37 RNP (8 μM):tSpA RNP (0 μM). **Fig. 26B** depicts the deletion profile observed with a 4:1 ratio of Sp37 RNP (6.4 μM):tSpA RNP (1.6 μM). **Fig. 26C** depicts the deletion profile observed with a 1:1 ratio of Sp37 RNP (4 μM):tSpA RNP (4 μM). **Fig. 26D** depicts the deletion profile observed with a 1:4 ratio of Sp37 RNP (1.6 μM):tSpA RNP (6.4 μM). For **Figs. 26A-D**, the X axis represents the position of the center of the deletions (0-based numbering) in relation to the transcription start site (TSS) of the *HBG* gene, and the Y axis represents the deletion length in

nucleotides. Each dot represents a unique deletion. The size of the dot represents the frequency of each deletion as detected by NGS.

[0079] Figs. 27A-B depict the most common deletions (representing >1% of all edited alleles) generated by Sp37 RNP (chemically synthesized Sp37 gRNA complexed with *S. Pyogenes* WT Cas9 protein) alone or Sp37 RNP (chemically synthesized Sp37 gRNA complexed with *S. Pyogenes* WT Cas9 protein) codelivered with tSpA RNP (chemically synthesized tSpA gRNA complexed with *S. Pyogenes* WT Cas9 protein). **Fig. 27A** depicts the most common deletions observed after editing by Sp37 RNP alone. The left graph indicates the size and the location of the deletions. Dark grey indicates deletions associated with microhomology domains (as determined by deletions whose sequence starts or ends with 2 nt or more of homology with the sequence flanking the other extremity of the deletion). Light grey indicates all other deletions. The right graph shows the contribution of each deletion to the overall indel profile (frequency of the deletion among the edited alleles as detected by NGS). **Fig. 27B** depicts the most common deletions observed after editing by Sp37 RNP codelivered with tSpA RNP. The left graph indicates the size of the deletions and the location of the deletions. Dark grey indicates deletions associated with microhomology domains (as determined by deletions whose sequence starts or ends with 2 nt or more of homology with the sequence flanking the other extremity of the deletion). Light grey indicates all other deletions. The right graph shows the contribution of each deletion to the overall indel profile (frequency of the deletion among the edited alleles as detected by NGS).

[0080] Figs. 28A-F depicts the characterization of productive, HbF-inducing deletions, at or near the distal CCAAT box. **Fig. 28A** depicts a schematic of the cell sorting analysis and genotype:phenotype analysis performed to determine which groups of deletions provide HbF induction. Briefly, mPB CD34+ cells were independently electroporated with single or paired RNP targeting the distal CCAAT box. After erythroid differentiation, the cells were sorted based on high HbF and low HbF protein expression. The indel frequency was determined by NGS analysis of the *HBG* PCR product amplified from the gDNA extracted from the high-HbF and low HbF cell fractions (right panel). **Fig. 28B** depicts the average fold enrichment of deletions disrupting only the proximal CCAAT box ("pCCAAT", pos. HBG -111: -115), deletions disrupting both the proximal CCAAT box and the distal CCAAT box ("pCCAAT & dCCAAT" respectively pos. HBG -84: -88 and pos. HBG -111: -115), deletions disrupting only the distal

CCAAT box (“dCCAAT” HBG -111: -115), deletions disrupting both the distal CCAAT box and the KLF binding area (“KLF & dCCAAT” respectively pos. HBG -111: -115 and pos. HBG -130: -144), deletions disrupting only the KLF1 binding area (“KLF”, as defined by the CACCC box and an additional 10nt at pos. HBG -130: -144), and deletions disrupting none of those genomic domains (“Others”). The values are depicted in a Tukey plot. Statistical analysis was performed by one ANOVA and post-hoc Tukey comparison test between all groups. Statistical significance is indicated when observed. **: $p < 0.01$. **Fig. 28C** depicts the average fold enrichment in the high HbF fraction of deletions disrupting the distal CCAAT box depending on the 5' boundary of the deletions (shown on the X axis). **Fig. 28D** depicts the average fold enrichment in the high HbF fraction of deletions disrupting the distal CCAAT box depending on the 3' boundary of the deletions (shown on the X axis). For **Figs. 28C and 28D**, “dCCAAT” represents the distal CCAAT box (pos. HBG -111: -115), “pCCAAT” represents the proximal CCAAT box (pos. HBG -84: -88), and KLF1 represents KLF1 binding site at the CACCC box (pos. HBG -140: -144). **Fig. 28E** depicts the average fold enrichment in the high HbF fraction of deletions disrupting 1, 2, 3, 4, or 5 nucleotide at the distal CCAAT box, regardless of the overall deletion length. A list of exemplary productive deletions that were shown to induce HbF expression (as defined by average enrichment in the high HbF fraction $> 2[\log_2]$) can be found within **Table 12**. **Fig. 28F** depicts the position and length of deletions frequently generated by Sp37 RNP or Sp37 RNP + tSpA dRNP (overlaid). Dark symbols indicate indels that were identified to be enriched in the high HbF fraction (as defined by average enrichment in the high HbF fraction $> 2[\log_2]$; regardless of the RNP used to generate the indels.). The number of nucleotides of the distal CCAAT box disrupted by the deletions is indicated by the shape of the symbol (circles : disrupts 0 nt of the distal CCAAT box, triangles : disrupts 1 to 2 nucleotides of the distal CCAAT box, squares : disrupts 3 to 5 nucleotides of the distal CCAAT box).

[0081] Figs. 29A-B depict indels generated by Sp37 RNP (chemically synthesized Sp37 gRNA complexed with *S. Pyogenes* WT Cas9 protein). **Fig. 29A** depicts the length and position of deletions generated in the region of the target distal CCAAT box (pos. HBG -111: -115) with the size of each dot representing the relative frequency of each deletion, i.e. larger dots represent a more common deletion. The X-axis represent the center of the deletions in the 0-based coordinate system relative to the *HBG* TSS (i.e., the deletion HBG:-112:-115 [1-based] is centered at -113 nt [0-based] from the *HBG* TSS. The Y axis represents the deletion length in

nucleotides. Each dot represents a unique deletion. The Sp37 gRNA cut site is indicated by an arrow at position X=103 nt [0-based] from the *HBG* TSS. The shaded grey triangular region depicts the area in which indels disrupt the distal CCAAT box, predicted to represent productive indels leading to HbF expression, based on results from Example 12. For example, small deletions originating from the Sp37 gRNA cut site and extending less than 8 nt downstream (i.e., deletion HBG -104:-108, which is 5 nts in length and centered at -105.5nt from the *HBG* TSS [0-based]), do not disrupt the dCCAAT box (at position HBG -111:-115) and is thus not covered by the shaded area. However, larger deletions, originating from the Sp37 gRNA cut site and extending further downstream of position HBG:-110 (i.e., HBG -104:-123, which is 20 nts in length and centered at -113 from the HBG TSS [0-based]), do disrupt the dCCAAT box and are thus predicted to promote HbF induction based on results from Example 12, as depicted by their location within the shaded region. Due to the repetitive nature of the sequence at this locus, some deletions may have ambiguous coordinates. For example, one 27 nt deletion removing one of the duplicated sequences, CTATTGGTCAAG, surrounding the proximal and distal CCAAT box (at position HBG-82:-93 and position HBG-109:121 ; depicted in **Fig. 31**), can be described as HBG:-82:-108, HBG:-94:-120, or any 27 nt long deletion in between. This deletion, as well as the deletion at position HBG:-78:-103, has the particularity to generate a single CCAAT box promoter that could be generated as the result of a distal CCAAT box deletion or proximal CCAAT box deletion. According to the observed HbF expression observed for both those deletions (as determined by enrichment of those indel in the high HbF fraction, as reported in **Table 12**) and the reported essential role of the proximal CCAAT box for *HBG* expression (Martyn 2017), those deletions were classified as distal CCAAT box deletions, and were thus included in the shaded area. **Fig. 29B** depicts the fraction of predicted productive deletions (based on results from Example 12) generated by Sp37 RNP among all indels. The fraction of other deletions and insertions is also detailed. Insertions= 28%, Deletions (Other)= 37%, dCCAAT Del. (1-3nt)= 5%, and dCCAAT Del. (4-5nt)= 30%.

[0082] Figs. 30A-B depict indels generated by Sp37 RNP (chemically synthesized Sp37 gRNA complexed with WT Cas9 protein) codelivered with tSpA RNP (chemically synthesized tSpA gRNA complexed with WT Cas9 protein), as determined by NGS analysis of the HBG PCR product from samples whose total editing and HbF results are described in **Fig. 24**. **Fig. 30A** depicts the length and position of deletions generated in the regions near and at the proximal

CCAAT box (pos. HBG 84: -88), distal CCAAT box (pos. HBG -111: -115), and CACCC box (KLF1 binding site) (pos. HBG -140: -144), with the size of each dot representing the relative frequency of each deletion, i.e. larger dots represent a more common deletion. The X-axis represents the center of the deletions in the 0-based coordinate system relative to the *HBG* TSS (i.e., the deletion HBG:-112:-115 [1-based] is centered at -113nt [0-based] from the *HBG* TSS). The Y axis represents the deletion length in nucleotides. Each dot represents a unique deletion. The Sp37 gRNA cut site is indicated by an arrow at position X=103 nt [0-based] from the *HBG* TSS. The shaded grey triangular region depicts the area in which indels are disrupting the distal CCAAT box, predicted to represent productive indels leading to HbF expression, based on results from Example 12. For example, small deletions originating from the Sp37 gRNA cut site and extending less than 8 nt downstream (i.e., deletion HBG -104:-108, of length 5 and centered at -105.5 nt from the HBG TSS [0-based]), do not disrupt the dCCAAT box (at position HBG -111:-115) and are thus not covered by the shaded area. However, larger deletions, originating from the Sp37 gRNA cut site and extending further downstream of position HBG:-110 (i.e., HBG -104:-123 of length 20 and centered at -113 from the HBG TSS [0-based]), do disrupt the dCCAAT box and are thus predicted to promote HbF induction based on results from Example 12, as depicted by their location within the shaded region. Due to the repetitive nature of the sequence at this locus, some deletions may have ambiguous coordinates. For example, one 27 nt deletion removing one of the duplicated sequence, CTATTGGTCAAG, surrounding the proximal and distal CCAAT box (at position HBG-82:-93 and position HBG-109:121; depicted in **Fig 31**), can be described as HBG:-82:-108, HBG:-94:-120, or any 27 nt long deletion in between. This deletion, as well as the deletion at position HBG:-78:-103 have the particularity to generate a single CCAAT box promoter that could be described as the result of a distal CCAAT box deletion or proximal CCAAT box deletion. According to the observed HbF expression observed for both those deletions (as determined by enrichment of those indel in the high HbF fraction, as reported in **Table 12**) and the reported essential role of the proximal CCAAT box for HBG expression (Martyn 2017), those deletions were classified as productive distal CCAAT box deletions, and thus included in the shaded area. **Fig. 30B** depicts the fraction of predicted productive deletions (based on results from Example 12) generated by Sp37 RNP among all indels. The fraction of other deletions and insertions is also detailed. Insertions= 23%, Deletions (Other)= 40%, dCCAAT Del. (1-3nt)= 7%, and dCCAAT Del. (4-5nt)= 30%.

[0083] Fig. 31 depicts a 27 nt deletion commonly generated by Sp37 RNP alone or paired with tSpA RNP. This deletion has the particularity of removing one of the duplicated sequence, CTATTGGTCAAG, surrounding the proximal and distal CCAAT box (at position HBG-82:-93 and position HBG-109:121). Due to the repetitive nature of the sequence at this locus, it can be described as HBG:-82:-108, HBG:-94:-120, or any 27 nt long deletion in between. As depicted in the figure showing four examples of coordinates describing the same 27 nt long deletion, any deletion matching the former criteria would lead to a unique resulting sequence after DNA repair. This deletion is also an example of deletion associated with micro-homologies. The 5' end of the deleted sequence (using coordinates HBG:-82:-108) contains a 12 nt long sequence perfectly homologous to the sequence immediately flanking the 3' end of the deletion. This deletion, whose high frequency after editing CD34+ cells with Sp37 RNP is unusual considering its length, is likely a favored repair outcome due to the presence of homologous domain facilitating DSB repair by micro-homology mediated end joining.

[0084] Figs. 32A-C depict editing detected in mPB CD34+ cells infused in nonirradiated NOD.B6.SCID Il2ry^{-/-} Kit(W41/W41) (NBSGW) mice. mPB CD34+ cells were electroporated with Sp37 RNP (comprising IVT Sp37 gRNA (SEQ ID NO:333, **Table 10**) complexed with *S. Pyogenes* wildtype Cas9) delivered alone (8 μ M) or co-delivered with tSpA dRNP (6.4 μ M + 1.6 μ M (Sp37 + tSpA)) (comprising IVT tSpA dgRNA (SEQ ID NO:970, **Table 10**) complexed with *S. Pyogenes* wildtype Cas9). **Fig. 32A** depicts the percentage of editing as determined by NGS for the pre-infusion modified mPB CD34+ cells ("Preinfusion") and for the human engrafted cells extracted from the bone marrow of the mice 16 weeks after infusion of the modified mPB CD34+ cells ("16week - M[1-10]", representing long term NBSGW-repopulating hematopoietic stem cells). Indels that disrupt the CCAAT box and are not associated with micro-homology domains ("CCAAT-disruption"), indels that disrupt the CCAAT box and are associated with micro-homology domains ("CCAAT-disruption MH associated"), indels that do not disrupt the CCAAT box and are associated with micro-homology domains ("MH-associated") and all other indels ("Other indels") were identified for each of the 10 mice tested (M1-10). **Fig. 32B.** depicts the deletion profile observed after treatment with Sp37 RNP in the pre-infusion mPB CD34+ cells and from the human cells collected from the bone marrow of the NBSGW mice 16 weeks after infusion of the modified mPB CD34+ cells (long term NBSGW-repopulating hematopoietic stem cells) as detected by NGS analysis. **Fig. 32C** depicts the deletion profile observed after

treatment with Sp37 RNP + tSpA dRNP in the pre-infusion mPB CD34+ cells and from the human cells collected from the bone marrow of the NBSGW mice 16 weeks after infusion of the modified mPB CD34+ cells (long term NBSGW-repopulating hematopoietic stem cells) as detected by NGS analysis. For **Figure 32B and 32C**, the center of the deletion relative to the transcription start site (TSS) of the *HBG* gene is shown on the X axis [0-based numbering] and the nucleotide length of the deletion is shown on the Y axis. Each symbol represents a unique deletion. The size of the symbol represents the frequency of each deletion or the average frequency of each deletion across all analyzed mice for 16-weeks *in vivo* samples) as detected by NGS, i.e., larger symbols represent a more common deletion. CCAAT box disrupting and non-CCAAT box disrupting deletions associated with micro-homology domains are shown, as well as those which are not associated with micro-homology domains.

DETAILED DESCRIPTION

Definitions and abbreviations

[0085] Unless otherwise specified, each of the following terms has the meaning associated with it in this section.

[0086] The indefinite articles "a" and "an" refer to at least one of the associated noun, and are used interchangeably with the terms "at least one" and "one or more." For example, "a module" means at least one module, or one or more modules.

[0087] The conjunctions "or" and "and/or" are used interchangeably as non-exclusive disjunctions.

[0088] "Domain" is used to describe a segment of a protein or nucleic acid. Unless otherwise indicated, a domain is not required to have any specific functional property.

[0089] The term "exogenous trans-acting factor" refers to any peptide or nucleotide component of a genome editing system that both (a) interacts with an RNA-guided nuclease or gRNA by means of a modification, such as a peptide or nucleotide insertion or fusion, to the RNA-guided nuclease or gRNA, and (b) interacts with a target DNA to alter a helical structure thereof. Peptide or nucleotide insertions or fusions may include, without limitation, direct covalent linkages between the RNA-guided nuclease or gRNA and the exogenous trans-acting factor, and/or non-covalent linkages mediated by the insertion or fusion of RNA/protein interaction domains such as MS2 loops and protein/protein interaction domains such as a PDZ, Lim or SH1,

2 or 3 domains. Other specific RNA and amino acid interaction motifs will be familiar to those of skill in the art. Trans-acting factors may include, generally, transcriptional activators.

[0090] An "indel" is an insertion and/or deletion in a nucleic acid sequence. An indel may be the product of the repair of a DNA double strand break, such as a double strand break formed by a genome editing system of the present disclosure. An indel is most commonly formed when a break is repaired by an "error prone" repair pathway such as the NHEJ pathway described below.

[0091] "Gene conversion" refers to the alteration of a DNA sequence by incorporation of an endogenous homologous sequence (e.g., a homologous sequence within a gene array). "Gene correction" refers to the alteration of a DNA sequence by incorporation of an exogenous homologous sequence, such as an exogenous single-or double stranded donor template DNA. Gene conversion and gene correction are products of the repair of DNA double-strand breaks by HDR pathways such as those described below.

[0092] Indels, gene conversion, gene correction, and other genome editing outcomes are typically assessed by sequencing (most commonly by "next-gen" or "sequencing-by-synthesis" methods, though Sanger sequencing may still be used) and are quantified by the relative frequency of numerical changes (e.g., ± 1 , ± 2 or more bases) at a site of interest among all sequencing reads. DNA samples for sequencing may be prepared by a variety of methods known in the art, and may involve the amplification of sites of interest by polymerase chain reaction (PCR), the capture of DNA ends generated by double strand breaks, as in the GUIDEseq process described in Tsai 2016 (incorporated by reference herein) or by other means well known in the art. Genome editing outcomes may also be assessed by *in situ* hybridization methods such as the FiberComb™ system commercialized by Genomic Vision (Bagneux, France), and by any other suitable methods known in the art.

[0093] "Target region" refers to a region of a nucleic acid that includes one or more nucleotides to be edited. In certain embodiments, the target region may be a target region of a genome.

[0094] "Productive indel" refers to an indel (deletion and/or insertion) that results in HbF expression. In certain embodiments, a productive indel may induce HbF expression. In certain embodiments, a productive indel may result in an increased level of HbF expression. Exemplary examples of productive indels are listed in **Table 12**.

[0095] "Alt-HDR," "alternative homology-directed repair," or "alternative HDR" are used interchangeably to refer to the process of repairing DNA damage using a homologous nucleic

acid (e.g., an endogenous homologous sequence, e.g., a sister chromatid, or an exogenous nucleic acid, e.g., a template nucleic acid). Alt-HDR is distinct from canonical HDR in that the process utilizes different pathways from canonical HDR, and can be inhibited by the canonical HDR mediators, RAD51 and BRCA2. Alt-HDR is also distinguished by the involvement of a single-stranded or nicked homologous nucleic acid template, whereas canonical HDR generally involves a double-stranded homologous template.

[0096] "Canonical HDR," "canonical homology-directed repair" or "cHDR" refer to the process of repairing DNA damage using a homologous nucleic acid (e.g., an endogenous homologous sequence, e.g., a sister chromatid, or an exogenous nucleic acid, e.g., a template nucleic acid). Canonical HDR typically acts when there has been significant resection at the double strand break, forming at least one single stranded portion of DNA. In a normal cell, cHDR 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.

[0097] Unless indicated otherwise, the term "HDR" as used herein encompasses both canonical HDR and alt-HDR.

[0098] "Non-homologous end joining" or "NHEJ" refers to ligation mediated repair and/or non-template mediated repair including canonical NHEJ (cNHEJ) and alternative NHEJ (altNHEJ), which in turn includes microhomology-mediated end joining (MMEJ), single-strand annealing (SSA), and synthesis-dependent microhomology-mediated end joining (SD-MMEJ).

[0099] "Replacement" or "replaced," when used with reference to a modification of a molecule (e.g., a nucleic acid or protein), does not require a process limitation but merely indicates that the replacement entity is present.

[0100] "Subject" means a human, mouse, or non-human primate. A human subject can be any age (e.g., an infant, child, young adult, or adult), and may suffer from a disease, or may be in need of alteration of a gene.

[0101] "Treat," "treating," and "treatment" mean the treatment of a disease in a subject (e.g., a human subject), including one or more of inhibiting the disease, i.e., arresting or preventing its development or progression; relieving the disease, i.e., causing regression of the disease state; relieving one or more symptoms of the disease; and curing the disease.

[0102] "Prevent," "preventing," and "prevention" refer to the prevention of a disease in a subject, including (a) avoiding or precluding the disease; (b) affecting the predisposition toward the disease; or (c) preventing or delaying the onset of at least one symptom of the disease.

[0103] A "kit" refers to any collection of two or more components that together constitute a functional unit that can be employed for a specific purpose. By way of illustration (and not limitation), one kit according to this disclosure can include a gRNA complexed or able to complex with an RNA-guided nuclease, and accompanied by (e.g., suspended in, or suspendable in) a pharmaceutically acceptable carrier. In certain embodiments, the kit can include a dead guide RNA. The kit can be used to introduce the complex into, for example, a cell or a subject, for the purpose of causing a desired genomic alteration in such cell or subject. The components of a kit can be packaged together, or they may be separately packaged. Kits according to this disclosure also optionally include directions for use (DFU) that describe the use of the kit e.g., according to a method of this disclosure. The DFU can be physically packaged with the kit, or it can be made available to a user of the kit, for instance by electronic means.

[0104] The terms "polynucleotide", "nucleotide sequence", "nucleic acid", "nucleic acid molecule", "nucleic acid sequence", and "oligonucleotide" refer to a series of nucleotide bases (also called "nucleotides") in DNA and RNA, and mean any chain of two or more nucleotides. The polynucleotides, nucleotide sequences, nucleic acids etc. can be chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. They can be modified at the base moiety, sugar moiety, or phosphate backbone, for example, to improve stability of the molecule, its hybridization parameters, etc. A nucleotide sequence typically carries genetic information, including, but not limited to, the information used by cellular machinery to make proteins and enzymes. These terms include double- or single-stranded genomic DNA, RNA, any synthetic and genetically manipulated polynucleotide, and both sense and antisense polynucleotides. These terms also include nucleic acids containing modified bases.

[0105] Conventional IUPAC notation is used in nucleotide sequences presented herein, as shown in **Table 1**, below (*see also* Cornish-Bowden 1985, incorporated by reference herein). It should be noted, however, that "T" denotes "Thymine or Uracil" in those instances where a sequence may be encoded by either DNA or RNA, for example in gRNA targeting domains.

Table 1: IUPAC nucleic acid notation

Character	Base
A	Adenine
T	Thymine
G	Guanine
C	Cytosine
U	Uracil
K	G or T/U
M	A or C
R	A or G
Y	C or T/U
S	C or G
W	A or T/U
B	C, G or T/U
V	A, C or G
H	A, C or T/U
D	A, G or T/U
N	A, C, G or T/U

[0106] The terms "protein," "peptide" and "polypeptide" are used interchangeably to refer to a sequential chain of amino acids linked together via peptide bonds. The terms include individual proteins, groups or complexes of proteins that associate together, as well as fragments or portions, variants, derivatives and analogs of such proteins. Peptide sequences are presented herein using conventional notation, beginning with the amino or N-terminus on the left, and proceeding to the carboxyl or C-terminus on the right. Standard one-letter or three-letter abbreviations can be used.

[0107] The term "variant" refers to an entity such as a polypeptide, polynucleotide or small molecule that shows significant structural identity with a reference entity but differs structurally from the reference entity in the presence or level of one or more chemical moieties as compared with the reference entity. In many embodiments, a variant also differs functionally from its reference entity. In general, whether a particular entity is properly considered to be a "variant" of a reference entity is based on its degree of structural identity with the reference entity.

[0108] The notation "CCAAT box target region" and the like refer to a sequence that is 5' of the transcription start site (TSS) of the *HBG1* and/or *HBG2* gene. CCAAT boxes are highly conserved motifs within the promoter region of α -like and β -like globin genes. The regions

within or near the CCAAT box play important roles in globin gene regulation. For example, the γ -globin distal CCAAT box is associated with hereditary persistence of fetal hemoglobin. A number of transcription factors have been reported to bind to the duplicated CCAAT box region of the γ -globin promoter, e.g., NF-Y, COUP-TFII (NF-E3), CDP, GATA1/NF-E1 and DRED (Martyn 2017). While not wishing to be bound by theory, it is believed that the binding sites of the transcriptional activator NF-Y overlaps with transcriptional repressors at the γ -globin promoter. HPFH mutations present within the distal γ -globin promoter region, e.g., within or near the CCAAT box, may alter the competitive binding of those factors and thus contribute to the increased γ -globin expression and elevated levels of HbF. Genomic locations provided herein for *HBG1* and *HBG2* are based on the coordinates provided in NCBI Reference Sequence NC_000011, “Homo sapiens chromosome 11, GRCh38.p12 Primary Assembly,” (Version NC_000011.10). The distal CCAAT box of *HBG1* and *HBG2* is positioned at *HBG1* and *HBG2* c.-111 to -115 (Genomic location is Hg38 Chr11:5,249,968 to Chr11:5,249,972 and Hg38 Chr11:5,254,892 to Chr11:5,254,896, respectively). The *HBG1* c.-111 to -115 region is exemplified in SEQ ID NO:902 (*HBG1*) at positions 2823-2827, and the *HBG2* c.-111 to -115 region is exemplified in SEQ ID NO:903 (*HBG2*) at positions 2747-2751. In certain embodiments, the “CCAAT box target region” denotes the region that is at or near the distal CCAAT box and includes the nucleotides of the distal CCAAT box and 25 nucleotides upstream (5') and 25 nucleotides downstream (3') of the distal CCAAT box (i.e., *HBG1/2* c.-86 to -140 (Genomic location is Hg38 Chr11:5249943 to Hg38 Chr11:5249997 and Hg38 Chr11:5254867 to Hg38 Chr11:5254921, respectively)). The *HBG1* c.-86 to -140 region is exemplified in SEQ ID NO:902 (*HBG1*) at positions 2798-2852, and the *HBG2* c.-86 to -140 region is exemplified in SEQ ID NO:903 (*HBG2*) at positions 2723-2776. In certain embodiments, the “CCAAT box target region” denotes the region that is at or near the distal CCAAT box and includes the nucleotides of the distal CCAAT box and 35 nucleotides upstream (5'), 30 nucleotides upstream (5'), 25 nucleotides upstream (5'), 20 nucleotides upstream (5'), 15 nucleotides upstream (5'), 10 nucleotides upstream (5'), or 5 nucleotides upstream (5') and 35 nucleotides downstream (3'), 30 nucleotides downstream (3'), 25 nucleotides downstream (3'), 20 nucleotides downstream (3'), 15 nucleotides downstream (3'), 10 nucleotides downstream (3'), or 5 nucleotides downstream (3') of the distal CCAAT box. In certain embodiments, the “CCAAT box target region” denotes the region that is at or near the distal CCAAT box and includes the nucleotides of the distal

CCAAT box and 5 nucleotides upstream (5') and 5 nucleotides downstream (3') of the distal CCAAT box (i.e., *HBG1*/2 c.-106 to -120 (Genomic location is Hg38 Chr11:5249963 to Hg38 Chr11:5249977 (*HGB1* and Hg38 Chr11:5254887 to Hg38 Chr11:5254901, respectively)). The *HBG1* c.-106 to -120 region is exemplified in SEQ ID NO:902 (*HBG1*) at positions 2818-2832, and the *HBG2* c.-106 to -120 region is exemplified in SEQ ID NO:903 (*HBG2*) at positions 2742-2756. The term "CCAAT box target site alteration" and the like refer to alterations (e.g., deletions, insertions, mutations) of one or more nucleotides of the CCAAT box target region. Examples of exemplary CCAAT box target region alterations include, without limitation, the 1 nt deletion, 4 nt deletion, 11 nt deletion, 13 nt deletion, and 18 nt deletion, and -117 G>A alteration. Additional exemplary CCAAT box target region alterations include the productive indels set forth in **Table 12**. As used herein, the terms "CCAAT box" and "CAAT box" can be used interchangeably.

[0109] The notations "c.-114 to -102 region," "c.-102 to -114 region," "-102:-114," "13 nt target region" and the like refer to a sequence that is 5' of the transcription start site (TSS) of the *HBG1* and/or *HBG2* gene at the genomic location Hg38 Chr11:5,249,959 to Hg38 Chr11:5,249,971 and Hg38 Chr11:5,254,883 to Hg38 Chr11:5,254,895, respectively. The *HBG1* c.-102 to -114 region is exemplified in SEQ ID NO:902 (*HBG1*) at positions 2824-2836 and the *HBG2* c.-102 to -114 region is exemplified in SEQ ID NO:903 (*HBG2*) at positions 2748-2760. The term "13 nt deletion" and the like refer to deletions of the 13 nt target region.

[0110] The notations "c.-121 to -104 region," "c.-104 to -121 region," "-104:-121," "18 nt target region," and the like refer to a sequence that is 5' of the transcription start site (TSS) of the *HBG1* and/or *HBG2* gene at the genomic location Hg38 Chr11:5,249,961 to Hg38 Chr11:5,249,978 and Hg38 Chr11:5,254,885 to Hg38 Chr11:5,254,902, respectively. The *HBG1* c.-104 to -121 region is exemplified in SEQ ID NO:902 (*HBG1*) at positions 2817-2834, and the *HBG2* c.-104 to -121 region is exemplified in SEQ ID NO:903 (*HBG2*) at positions 2741-2758. The term "18 nt deletion" and the like refer to deletions of the 18 nt target region.

[0111] The notations "c.-105 to -115 region," "c.-115 to -105 region," "-105:-115," "11 nt target region," and the like refer to a sequence that is 5' of the transcription start site (TSS) of the *HBG1* and/or *HBG2* gene at the genomic location Hg38 Chr11:5,249,962 to Hg38 Chr11:5,249,972 and Hg38 Chr11:5,254,886 to Hg38 Chr11:5,254,896, respectively. The *HBG1* c.-105 to -115 region is exemplified in SEQ ID NO:902 (*HBG1*) at positions 2823-2833, and the

HBG2 c.-105 to -115 region is exemplified in SEQ ID NO:903 (*HBG2*) at positions 2747-2757. The term "11 nt deletion" and the like refer to deletions of the 11 nt target region.

[0112] The notations "c.-115 to -112 region," "c.-112 to -115 region," "-112:-115," "4 nt target region," and the like refer to a sequence that is 5' of the transcription start site (TSS) of the *HBG1* and/or *HBG2* gene at the genomic location Hg38 Chr11:5,249,969 to Hg38 Chr11:5,249,972 and Hg38 Chr11:5,254,893 to Hg38 Chr11:5,254,896, respectively. The *HBG1* c.-112 to -115 region is exemplified in SEQ ID NO:902 at positions 2823-2826, and the *HBG2* c.-112 to -115 region is exemplified in SEQ ID NO:903 (*HBG2*) at positions 2747-2750. The term "4 nt deletion" and the like refer to deletions of the 4 nt target region.

[0113] The notations "c.-116 region," "HBG-116," "1 nt target region," and the like refer to a sequence that is 5' of the transcription start site (TSS) of the *HBG1* and/or *HBG2* gene at the genomic location Hg38 Chr11:5,249,973 and Hg38 Chr11:5,254,897, respectively. The *HBG1* c.-116 region is exemplified in SEQ ID NO:902 at position 2822, and the *HBG2* c.-116 region is exemplified in SEQ ID NO:903 (*HBG2*) at position 2746. The term "1 nt deletion" and the like refer to deletions of the 1 nt target region.

[0114] The notations "c.-117 G>A region," "HBG-117 G>A," "-117 G>A target region" and the like refer to a sequence that is 5' of the transcription start site (TSS) of the *HBG1* and/or *HBG2* gene at the genomic location Hg38 Chr11:5,249,974 to Hg38 Chr11:5,249,974 and Hg38 Chr11:5,254,898 to Hg38 Chr11:5,254,898, respectively. The *HBG1* c.-117 G>A region is exemplified by a substitution from guanine (G) to adenine (A) in SEQ ID NO:902 at position 2821, and the *HBG2* c.-117 G>A region is exemplified by a substitution from G to A in SEQ ID NO:903 (*HBG2*) at position 2745. The term "-117 G>A alteration" and the like refer to a substitution from G to A at the -117G>A target region.

[0115] The term "proximal HBG1/2 promoter target sequence" denotes the region within 50, 100, 200, 300, 400, or 500 bp of a proximal *HBG1/2* promoter sequence including the 13 nt target region. Alterations by genome editing systems according to this disclosure facilitate (e.g., cause, promote or tend to increase the likelihood of) upregulation of HbF production in erythroid progeny.

[0116] The term "GATA1 binding motif in *BCL11Ae*" refers to the sequence that is the GATA1 binding motif in the erythroid specific enhancer of *BCL11A* (*BCL11Ae*) that is in the +58 DNase I hypersensitive site (DHS) region of intron 2 of the *BCL11A* gene. The genomic coordinates for

the GATA1 binding motif in *BCL11Ae* are chr2: 60,495,265 to 60,495,270. The +58 DHS site comprises a 115 base pair (bp) sequence as set forth in SEQ ID NO:968. The +58 DHS site sequence, including ~500 bp upstream and ~200 bp downstream is set forth in SEQ ID NO:969.

Overview

[0117] The various embodiments of this disclosure generally relate to genome editing systems configured to introduce alterations (*e.g.*, a deletion or insertion, or other mutation) into chromosomal DNA that enhance transcription of the *HBG1* and/or *HBG2* genes, which encode the γ A and γ G subunits of hemoglobin, respectively. Exemplary mutations are made in or around the CCAAT box target region, 13 nt target region and/or into the GATA1 binding motif in *BCL11Ae* of *HBG1* and/or *HBG2*.

Targeted genome editing for fetal hemoglobin induction

[0118] Fetal hemoglobin (HbF) expression can be induced using various genome strategies. For example, HbF expression can be induced through targeted disruption of the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and or the erythroid cell specific expression of a transcriptional repressor, *BCL11A* (*BCL11Ae*) (also discussed in commonly-assigned International Patent Publication No. WO 2015/148860 by Friedland et al. (“Friedland”), published Oct. 1, 2015, which is incorporated by reference in its entirety herein), which encodes a repressor that silences *HBG1* and *HBG2* (Canvers 2015). In certain embodiments, the region of *BCL11Ae* targeted for disruption may be the GATA1 binding motif in *BCL11Ae*. In certain embodiments, genome editing systems disclosed herein may be used to introduce alterations into the GATA1 binding motif in *BCL11Ae*, the CCAAT box target region, and/or the 13 nt target region of *HBG1* and/or *HBG2*.

[0119] The genome editing systems of this disclosure can include an RNA-guided nuclease such as Cas9 or Cpf1 and one or more gRNAs having a targeting domain that is complementary to a sequence in or near the target region, and optionally one or more of a DNA donor template that encodes a specific mutation (such as a deletion or insertion) in or near the target region, and/or an agent that enhances the efficiency with which such mutations are generated including, without limitation, a random oligonucleotide, a small molecule agonist or antagonist of a gene product involved in DNA repair or a DNA damage response, or a peptide agent.

[0120] A variety of approaches to the introduction of mutations into the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1

binding motif in *BCL11Ae* may be employed in the embodiments of the present disclosure. In one approach, a single alteration, such as a double-strand break, is made within the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*, and is repaired in a way that disrupts the function of the region, for example by the formation of an indel or by the incorporation of a donor template sequence that encodes the deletion of the region. In a second approach, two or more alterations are made on either side of the region, resulting in the deletion of the intervening sequence, including the CCAAT box target region, 13 nt target region and/or the GATA1 binding motif in *BCL11Ae*.

[0121] The treatment of hemoglobinopathies by gene therapy and/or genome editing is complicated by the fact that the cells that are phenotypically affected by the disease, erythrocytes or RBCs, are enucleated, and do not contain genetic material encoding either the aberrant hemoglobin protein (Hb) subunits nor the γ A or γ G subunits targeted in the exemplary genome editing approaches described above. This complication is addressed, in certain embodiments of this disclosure, by the alteration of cells that are competent to differentiate into, or otherwise give rise to, erythrocytes. Cells within the erythroid lineage that are altered according to various embodiments of this disclosure include, without limitation, hematopoietic stem and progenitor cells (HSCs), erythroblasts (including basophilic, polychromatic and/or orthochromatic erythroblasts), proerythroblasts, polychromatic erythrocytes or reticulocytes, embryonic stem (ES) cells, and/or induced pluripotent stem (iPSC) cells. These cells may be altered *in situ* (e.g., within a tissue of a subject) or *ex vivo*. Implementations of genome editing systems for *in situ* and *ex vivo* alteration of cells is described under the heading "Implementation of genome editing systems: delivery, formulations, and routes of administration" below.

[0122] In certain embodiments, alterations that result in induction of γ A and/or γ G expression are obtained through the use of a genome editing system comprising an RNA-guided nuclease and at least one gRNA having a targeting domain complementary to a sequence within the 13 nt target region of *HBG1* and/or *HBG2* or proximate thereto (e.g., within 10, 20, 30, 40, or 50, 100, 200, 300, 400 or 500 bases of the 13 nt target region). As is discussed in greater detail below, the RNA-guided nuclease and gRNA form a complex that is capable of associating with and altering the 13 nt target region or a region proximate thereto. Examples of suitable targeting domains directed to the 13 nt target region of *HBG1* and/or *HBG2* or proximate thereto for use in

the embodiments disclosed herein include, without limitation, those set forth in SEQ ID NOs:251-901, 940-942.

[0123] In certain embodiments, alterations that result in induction of HbF expression are obtained through the use of a genome editing system comprising an RNA-guided nuclease and at least one gRNA having a targeting domain complementary to a sequence within the GATA1 binding motif in *BCL11Ae* or proximate thereto (e.g., within 10, 20, 30, 40, or 50, 100, 200, 300, 400 or 500 bases of the GATA1 binding motif in *BCL11Ae*). In certain embodiments, the RNA-guided nuclease and gRNA form a complex that is capable of associating with and altering the GATA1 binding motif in *BCL11Ae*. Examples of suitable targeting domains directed to the GATA1 binding motif in *BCL11Ae* for use in the embodiments disclosed herein include, without limitation, those set forth in SEQ ID NOs:952-955.

[0124] The genome editing system can be implemented in a variety of ways, as is discussed below in detail. As an example, a genome editing system of this disclosure can be implemented as a ribonucleoprotein complex or a plurality of complexes in which multiple gRNAs are used. This ribonucleoprotein complex can be introduced into a target cell using art-known methods, including electroporation, as described in commonly-assigned International Patent Publication No. WO 2016/182959 by Jennifer Gori ("Gori"), published Nov. 17, 2016, which is incorporated by reference in its entirety herein.

[0125] The ribonucleoprotein complexes within these compositions are introduced into target cells by art-known methods, including without limitation electroporation (e.g., using the Nucleofection™ technology commercialized by Lonza, Basel, Switzerland or similar technologies commercialized by, for example, Maxcyte Inc. Gaithersburg, Maryland) and lipofection (e.g., using Lipofectamine™ reagent commercialized by Thermo Fisher Scientific, Waltham Massachusetts). Alternatively, or additionally, ribonucleoprotein complexes are formed within the target cells themselves following introduction of nucleic acids encoding the RNA-guided nuclease and/or gRNA. These and other delivery modalities are described in general terms below and in Gori.

[0126] Cells that have been altered *ex vivo* according to this disclosure can be manipulated (e.g., expanded, passaged, frozen, differentiated, de-differentiated, transduced with a transgene, etc.) prior to their delivery to a subject. The cells are, variously, delivered to a subject from which

they are obtained (in an "autologous" transplant), or to a recipient who is immunologically distinct from a donor of the cells (in an "allogeneic" transplant).

[0127] In some cases, an autologous transplant includes the steps of obtaining, from the subject, a plurality of cells, either circulating in peripheral blood, or within the marrow or other tissue (e.g., spleen, skin, etc.), and manipulating those cells to enrich for cells in the erythroid lineage (e.g., by induction to generate iPSCs, purification of cells expressing certain cell surface markers such as CD34, CD90, CD49f and/or not expressing surface markers characteristic of non-erythroid lineages such as CD10, CD14, CD38, etc.). The cells are, optionally or additionally, expanded, transduced with a transgene, exposed to a cytokine or other peptide or small molecule agent, and/or frozen/thawed prior to transduction with a genome editing system targeting the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*. The genome editing system can be implemented or delivered to the cells in any suitable format, including as a ribonucleoprotein complex, as separated protein and nucleic acid components, and/or as nucleic acids encoding the components of the genome editing system.

[0128] However it is implemented, a genome editing system may include, or may be co-delivered with, one or more factors that improve the viability of the cells during and after editing, including without limitation an aryl hydrocarbon receptor antagonist such as StemRegenin-1 (SR1), UM171, LGC0006, alpha-naphthoflavone, and CH-223191, and/or an innate immune response antagonist such as cyclosporin A, dexamethasone, resveratrol, a MyD88 inhibitory peptide, an RNAi agent targeting Myd88, a B18R recombinant protein, a glucocorticoid, OxPAPC, a TLR antagonist, rapamycin, BX795, and a RLR shRNA. These and other factors that improve the viability of the cells during and after editing are described in Gori, under the heading "I. Optimization of Stem Cells" from page 36 through page 61, which is incorporated by reference herein.

[0129] The cells, following delivery of the genome editing system, are optionally manipulated e.g., to enrich for HSCs and/or cells in the erythroid lineage and/or for edited cells, to expand them, freeze/thaw, or otherwise prepare the cells for return to the subject. The edited cells are then returned to the subject, for instance in the circulatory system by means of intravenous delivery or delivery or into a solid tissue such as bone marrow.

[0130] Functionally, alteration of the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae* using the compositions, methods and genome editing systems of this disclosure results in significant induction, among hemoglobin-expressing cells, of γ A and/or γ G subunits (referred to interchangeably as HbF expression), e.g., at least 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50% or greater induction of γ A and/or γ G subunit expression relative to unmodified controls. This induction of protein expression is generally the result of alteration of the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae* (expressed, e.g., in terms of the percentage of total genomes comprising indel mutations within the plurality of cells) in some or all of the plurality of cells that are treated, e.g., at least 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50% of the plurality of cells comprise at least one allele comprising a sequence alteration, including, without limitation, an indel, insertion, or deletion in or near the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*.

[0131] The functional effects of alterations caused or facilitated by the genome editing systems and methods of the present disclosure can be assessed in any number of suitable ways. For example, the effects of alterations on expression of fetal hemoglobin can be assessed at the protein or mRNA level. Expression of HBG1 and HBG2 mRNA can be assessed by digital droplet PCR (ddPCR), which is performed on cDNA samples obtained by reverse transcription of mRNA harvested from treated or untreated samples. Primers for HBG1, HBG2, HBB, and/or HBA may be used individually or multiplexed using methods known in the art. For example, ddPCR analysis of samples may be conducted using the QX200™ ddPCR system commercialized by Bio Rad (Hercules, CA), and associated protocols published by BioRad. Fetal hemoglobin protein may be assessed by high pressure liquid chromatography (HPLC), for example, according to the methods discussed on pp. 143-44 of Chang 2017, incorporated by reference herein, or fast protein liquid chromatography (FPLC) using ion-exchange and/or reverse phase columns to resolve HbF, HbB and HbA and/or γ A and γ G globin chains as is known in the art.

[0132] It should be noted that the rate at which the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in

BCL11Ae is altered in the target cells can be modified by the use of optional genome editing system components such as oligonucleotide donor templates. Donor template design is described in general terms below under the heading "Donor template design." Donor templates for use in targeting the 13 nt target region may include, without limitation, donor templates encoding alterations (e.g., deletions) of *HBG1* c.-114 to -102 (corresponding to nucleotides 2824-2836 of SEQ ID NO: 902), *HBG1* c.-225 to -222 (corresponding to nucleotides 2716-2719 of SEQ ID NO:902)), and/or *HBG2* c.-114 to -102 (corresponding to nucleotides 2748-2760 of SEQ ID NO:903). Exemplary 5' and 3' homology arms, and exemplary full-length donor templates encoding deletions such as c. -114 to -102 are also presented below (SEQ ID NOS: 904-909). Donor templates used herein may be non-specific templates that are non-homologous to regions of DNA within or near the target sequence. In certain embodiments, donor templates for use in targeting the 13 nt target region may include, without limitation, non-target specific templates that are nonhomologous to regions of DNA within or near the 13 nt target region. For example, a non-specific donor template for use in targeting the 13 nt target region may be non-homologous to the regions of DNA within or near the 13 nt target region and may comprise a donor template encoding the deletion of *HBG1* c.-225 to -222 (corresponding to nucleotides 2716-2719 of SEQ ID NO:902). In certain embodiments, donor templates for use in targeting the GATA1 binding motif in *BCL11Ae* may include, without limitation, non-target specific templates that are nonhomologous to regions of DNA within or near GATA1 binding motif in *BCL11Ae* target sequence. Other donor templates for use in targeting *BCL11Ae* may include, without limitation, donor templates including alternations (e.g., deletions) of *BCL11Ae*, including, without limitation, the GATA1 motif in *BCL11Ae*.

RNA-Guided helicases and dead guide RNAs

[0133] Various embodiments of the present disclosure also generally relate to genome editing systems configured to alter the helical structure of a nucleic acid to enhance genome editing of a target region (e.g., the CCAAT box target region, 13 nt target region, proximal *HBG1/2* promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*) in the nucleic acid, and methods and compositions thereof. Many embodiments relate to the observation that positioning an event that alters the helical structure of DNA within or adjacent to target regions in nucleic acid may improve the activity of genome editing systems directed to such target regions. Without wishing to be bound by any theory, it is thought that alterations of helical structure (e.g.,

by unwinding) within or proximal to DNA target regions may induce or increase accessibility of a genome editing system to the target region, resulting in increased editing of the target regions by the genome editing system.

[0134] CRISPR nucleases evolved primarily to defend bacteria against viral pathogens, whose genomes are not naturally organized into chromatin. By contrast, when eukaryotic genomes are organized into nucleosomal units comprising genomic DNA segments coiled around histones. CRISPR nucleases from several bacterial families have been found to be inactive for editing eukaryotic DNA, suggesting the ability to edit nucleosome-bound DNA might differ across enzymes (Ran 2015). Biochemical evidence shows that *S. pyogenes* Cas9 can cleave DNA efficiently at nucleosome edges, but has reduced activity when the target site is positioned near the center of nucleosome dyad (Hinz 2016).

[0135] In many cell types, target sites of interest may be strongly bound by nucleosomes, or may only possess adjacent PAMs for enzymes that do not edit efficiently in the presence of nucleosomes. In this case, the problematic nucleosomes could be displaced first by using adjacent target sites that are closer to the nucleosome edge or are bound by an enzyme that is more effective at binding nucleosomal DNA. However, cleavage at these adjacent sites could be detrimental to the therapeutic strategy. Therefore, having a programmable enzyme that binds these adjacent sites but does not cleave can enable more efficient functional editing.

[0136] It will be evident to the skilled artisan that the simplified systems and methods described herein offer several advantages over competing approaches. For example, a related strategy using catalytically inactive (dead) enzymes targeting sites adjacent to the site where editing is desired has been described in the literature (Chen 2017). However, this strategy entails a potential safety issue: if the full-length gRNAs complexed with a catalytically inactive nuclease molecule dissociates and later reassociates with a catalytically active nuclease enzyme, the gRNAs could introduce undesirable off-target edits. In contrast, the systems and methods of the present disclosure eliminates this risk because it relies on the observation that a dead gRNA (dgRNA) (gRNAs with a targeting domain of 15 nucleotides or less) allow an RNA-guided nuclease to bind, but not cleave, its target site. Thus, the dgRNAs provided herein will not support nuclease activity irrespective of their association with any particular RNA-guided nuclease molecule. By using these dead gRNAs, adjacent target sites can be used to aid in

nucleosome displacement without the risk of guide RNA swapping between active and inactive enzyme.

[0137] Another related strategy utilizes recruitment of exogenous trans-acting factors to facilitate nucleosome displacement. However, the systems and methods of this disclosure are advantageous over this strategy because they do not require gRNA modifications beyond truncation of the targeting domain, do not require the recruitment of exogenous trans-acting factors, and do not require transcriptional activation to achieve increased rates of editing.

[0138] Additionally, the use of dead gRNAs in the genome editing systems of the present disclosure are advantageous because they are not expected to result in any new delivery/solubility or folding/manufacturing considerations relative to genome editing systems utilizing full-length gRNAs. However, a skilled artisan might expect to encounter such problems in genome editing systems that utilize an exogenous trans-acting factor, which may entail large fusion proteins and/or RNA insertions or fusions. Further, dead gRNA strategies are likely to be capable of implementation using existing manufacturing, delivery, and other commercial processes that have been designed for wild-type nuclease products with relatively few substantial changes.

[0139] A variety of approaches to the unwinding and alteration of nucleic acid are employed in the various embodiments of this disclosure. One approach comprises unwinding (or opening of) a chromatin segment within or proximal to a target region (*e.g.*, the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*) of a nucleic acid in a cell and generating a double stranded break (DSB) within the target region of the nucleic acid whereby the target region is altered. In certain embodiments, the DSB may be repaired in a manner that alters the target region. Unwinding the chromatin segment using the methods provided herein may facilitate increased access of catalytically active RNPs (*e.g.*, catalytically active RNA-guided nucleases and gRNAs) to the chromatin to allow for more efficient editing of the DNA. For example, these methods may be used to edit target regions in chromatin that are difficult for a ribonucleoprotein (*e.g.*, RNA-guided nuclease complexed to gRNA) to access because the chromatin is occupied by nucleosomes, such as closed chromatin. In certain embodiments, the unwinding of the chromatin segment occurs via RNA-guided helicase activity. In certain embodiments, the unwinding step does not require recruiting an exogenous trans-acting factor to the chromatin segment. In certain

embodiments, the step of unwinding the chromatin segment does not comprise forming a single or double-stranded break in the nucleic acid within the chromatin segment.

[0140] In certain embodiments of the approaches and methods described above, the alteration of DNA helical structure is achieved through the action of an "RNA-guided helicase," which term is generally used to refer to a molecule, typically a peptide, that (a) interacts (e.g., complexes) with a gRNA, and (b) together with the gRNA, associates with and unwinds, but does not cleave, a target site. RNA-guided helicases may, in certain embodiments, comprise RNA-guided nucleases configured to lack nuclease activity. However, the inventors have observed that even a cleavage-competent RNA-guided nuclease may be adapted for use as an RNA-guided helicase by complexing it to a dead gRNA having a truncated targeting domain of 15 or fewer nucleotides in length. Complexes of wild-type RNA-guided nucleases with dead gRNAs exhibit reduced or eliminated RNA-cleavage activity, but appear to retain helicase activity. RNA-guided helicases and dead gRNAs are described in greater detail below.

[0141] Regarding RNA-guided helicases, according to the present disclosure an RNA-guided helicase may comprise any of the RNA-guided nucleases disclosed herein and *infra* under the heading entitled "RNA-guided nucleases," including, without limitation, a Cas9 or Cpf1 RNA-guided nuclease. The helicase activity of these RNA-guided nucleases allow for unwinding of DNA, providing increased access of genome editing system components (e.g., without limitation, catalytically active RNA-guided nuclease and gRNAs) to the desired target region to be edited (e.g., the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*). In certain embodiments, the RNA-guided nuclease may be a catalytically active RNA-guided nuclease with nuclease activity. In certain embodiments, the RNA-guided helicase may be configured to lack nuclease activity. For example, in certain embodiments, the RNA-guided helicase may be a catalytically inactive RNA-guided nuclease that lacks nuclease activity, such as a catalytically dead Cas9 molecule, which still provides helicase activity. In certain embodiments, an RNA-guided helicase may form a complex with a dead gRNA, forming a dead RNP that cannot cleave nucleic acid. In other embodiments, the RNA-guided helicase may be a catalytically active RNA-guided nuclease complexed to a dead gRNA, forming a dead RNP that cannot cleave nucleic acid. In certain embodiments, the RNA-guided nuclease is not configured to recruit an exogenous trans-acting factor to the desired target region to be edited (e.g., the CCAAT box target region, 13 nt target

region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*).

[0142] Turning to dead gRNAs, these include any of the dead gRNAs discussed herein and *infra* under the heading entitled "Dead gRNA molecules." Dead gRNAs (also referred to herein as "dgRNAs") may be generated by truncating the 5' end of a gRNA targeting domain sequence, resulting in a targeting domain sequence of 15 nucleotides or fewer in length. In certain embodiments, a dgRNA may be generated by truncating the 5' end of any one of a gRNA targeting domain sequence disclosed herein in **Table 2** or **Table 10**. Dead guide RNA molecules according to the present disclosure include dead guide RNA molecules that have reduced, low, or undetectable cleavage activity. The targeting domain sequences of dead guide RNAs may be shorter in length by 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides compared to the targeting domain sequence of active guide RNAs. Dead gRNA molecules may comprise targeting domains complementary to regions proximal to or within a target region (*e.g.*, the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*) in a target nucleic acid. In certain embodiments, "proximal to" may denote the region within 10, 25, 50, 100, or 200 nucleotides of a target region (*e.g.*, the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*). In certain embodiments, dead gRNAs comprise targeting domains complementary to the transcription strand or non-transcription strand of DNA. In certain embodiments, the dead guide RNA is not configured to recruit an exogenous trans-acting factor to a target region (*e.g.*, the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*).

[0143] Also provided herein are methods of increasing a rate of indel formation in a target nucleic acid by unwinding DNA within or proximal to the target region (*e.g.*, the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*) using an RNA-guided helicase, generating a DSB within the target region, and forming an indel in the target region through repair of the DSB. The step of unwinding the DNA using an RNA-guided helicase provides for increased indel formation compared to a method of forming indels that does not use a helicase.

[0144] This disclosure further encompasses methods of deleting a segment of a target nucleic acid in a cell, comprising contacting the cell with an RNA-guided helicase and generating a

double strand break (DSB) within the target region (*e.g.*, the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*). In certain embodiments, the RNA-guided helicase is configured to associate within or proximal to a target region of the target nucleic acid and unwind double stranded DNA (dsDNA) within or proximal to the target region. In certain embodiments, the target nucleic acid is a promoter region of a gene, a coding region of a gene, a non-coding region of a gene, an intron of a gene, or an exon of a gene. In certain embodiments, the segment of the target nucleic acid to be deleted may be at least about 15, 20, 25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, or 100 base pairs in length. In certain embodiments, the DSB is repaired in a manner that deletes the segment of the target nucleic acid.

[0145] Genome editing systems configured to introduce alterations of helical structure may be implemented in a variety of ways, as is discussed below in detail. As an example, a genome editing system of this disclosure can be implemented as a ribonucleoprotein complex or a plurality of complexes in which multiple gRNAs are used. In certain embodiments, a ribonucleoprotein complex of the genome editing system may be an RNA-guided helicase complexed to a dead guide RNA. Ribonucleoprotein complexes can be introduced into a target cell using art-known methods, including electroporation, as described in Gori. Genome editing systems incorporating RNA-guided helicases may also be modified in any suitable manner, including without limitation by the inclusion of one or more of a DNA donor template that encodes a specific mutation (such as a deletion or insertion) in or near the target region, and/or an agent that enhances the efficiency with which such mutations are generated including, without limitation, a random oligonucleotide, a small molecule agonist or antagonist of a gene product involved in DNA repair or a DNA damage response, or a peptide agent. These modifications are described in greater detail below, under the heading "Genome Editing Strategies." For clarity, this disclosure includes compositions comprising one or more gRNAs, dead gRNAs, RNA-guided helicases, RNA-guided nucleases, or a combination thereof.

[0146] While several of the exemplary embodiments above have focused on DNA unwinding, it should be noted that other helical alterations are within the scope of the present disclosure. These include, without limitation, overwinding, underwinding, increase or decrease of torsional strain on DNA strands within or proximate to a target region (*e.g.*, through topoisomerase activity), denaturation or strand separation, and/or other suitable alterations resulting in

modifications of chromatin structure. Each of these alterations may be catalyzed by an RNA-guided activity, or by the recruitment of an endogenous factor to a target region.

[0147] Also provided herein are genome editing systems and methods of altering one or more indels (e.g., indel signature) generated by an active guide. As the inventors have discovered herein, pairing of a dead RNP (dRNP) (i.e., a dead guide RNA complexed with an RNA-guided nuclease) and an active RNP (i.e., an active guide RNA complexed with an RNA-guided nuclease) can result in a change of the directionality of the indels (e.g., indel signature) generated by the active RNP alone (without a dRNP). As shown in the examples below, the use of the dead guide RNA may result in an increased frequency of larger deletions extending from the active guide RNA cut site toward the dead guide RNA binding site. Thus, the dead guide RNA may be used to effectively “orient” deletion editing toward a desired target site. In certain embodiments, the use of the dead guide RNA with an active guide RNA may increase the frequency of deletions that are not associated with micro-homologies.

[0148] Although the examples disclosed in the Examples section below are directed to alterations of the CCAAT box target region, skilled artisans would contemplate that the genome editing systems, methods, cells and compositions described herein may be used to alter any other target region, for example, without limitation, to increase the frequency of deletions at the target region, increase the frequency of deletions at the target region that are not associated with micro-homologies (e.g., not repaired via MMEJ).

[0149] This overview has focused on a handful of exemplary embodiments that illustrate the principles of genome editing systems and CRISPR-mediated methods of altering cells. For clarity, however, this disclosure encompasses modifications and variations that have not been expressly addressed above, but will be evident to those of skill in the art. With that in mind, the following disclosure is intended to illustrate the operating principles of genome editing systems more generally. What follows should not be understood as limiting, but rather illustrative of certain principles of genome editing systems and CRISPR-mediated methods utilizing these systems, which, in combination with the instant disclosure, will inform those of skill in the art about additional implementations and modifications that are within its scope.

Genome editing systems

[0150] The term "genome editing system" refers to any system having RNA-guided DNA editing activity. Genome editing systems of the present disclosure include at least two

components adapted from naturally occurring CRISPR systems: a gRNA and an RNA-guided nuclease. These two components form a complex that is capable of associating with a specific nucleic acid sequence and editing the DNA in or around that nucleic acid sequence, for instance by making one or more of a single-strand break (an SSB or nick), a double-strand break (a DSB) and/or a point mutation.

[0151] In certain embodiments, the genome editing systems in this disclosure may include a helicase for unwinding DNA. In certain embodiments, the helicase may be an RNA-guided helicase. In certain embodiments, the RNA-guided helicase may be an RNA-guided nuclease as described herein, such as a Cas9 or Cpf1 molecule. In certain embodiments, the RNA-guided nuclease is not configured to recruit an exogenous trans-acting factor to a target region. In certain embodiments, the RNA-guided nuclease may be configured to lack nuclease activity. In certain embodiments, the RNA-guided helicase may be complexed with a dead guide RNA as disclosed herein. For example, the dead guide RNA (dgRNA) may comprise a targeting domain sequence less than 15 nucleotides in length. In certain embodiments, the dead guide RNA is not configured to recruit an exogenous trans-acting factor to a target region.

[0152] Naturally occurring CRISPR systems are organized evolutionarily into two classes and five types (Makarova 2011, incorporated by reference herein), and while genome editing systems of the present disclosure may adapt components of any type or class of naturally occurring CRISPR system, the embodiments presented herein are generally adapted from Class 2, and type II or V CRISPR systems. Class 2 systems, which encompass types II and V, are characterized by relatively large, multidomain RNA-guided nuclease proteins (e.g., Cas9 or Cpf1) and one or more guide RNAs (e.g., a crRNA and, optionally, a tracrRNA) that form ribonucleoprotein (RNP) complexes that associate with (i.e., target) and cleave specific loci complementary to a targeting (or spacer) sequence of the crRNA. Genome editing systems according to the present disclosure similarly target and edit cellular DNA sequences, but differ significantly from CRISPR systems occurring in nature. For example, the unimolecular guide RNAs described herein do not occur in nature, and both guide RNAs and RNA-guided nucleases according to this disclosure may incorporate any number of non-naturally occurring modifications.

[0153] Genome editing systems can be implemented (e.g., administered or delivered to a cell or a subject) in a variety of ways, and different implementations may be suitable for distinct applications. For instance, a genome editing system is implemented, in certain embodiments, as

a protein/RNA complex (a ribonucleoprotein, or RNP), which can be included in a pharmaceutical composition that optionally includes a pharmaceutically acceptable carrier and/or an encapsulating agent, such as, without limitation, a lipid or polymer micro- or nano-particle, micelle, or liposome. In certain embodiments, a genome editing system is implemented as one or more nucleic acids encoding the RNA-guided nuclease and guide RNA components described above (optionally with one or more additional components); in certain embodiments, the genome editing system is implemented as one or more vectors comprising such nucleic acids, for instance a viral vector such as an adeno-associated virus (see section below under the heading "Implementation of genome editing systems: delivery, formulations, and routes of administration"); and in certain embodiments, the genome editing system is implemented as a combination of any of the foregoing. Additional or modified implementations that operate according to the principles set forth herein will be apparent to the skilled artisan and are within the scope of this disclosure.

[0154] It should be noted that the genome editing systems of the present disclosure can be targeted to a single specific nucleotide sequence, or may be targeted to — and capable of editing in parallel — two or more specific nucleotide sequences through the use of two or more guide RNAs. The use of multiple gRNAs is referred to as "multiplexing" throughout this disclosure, and can be employed to target multiple, unrelated target sequences of interest, or to form multiple SSBs or DSBs within a single target domain and, in some cases, to generate specific edits within such target domain. For example, International Patent Publication No. WO 2015/138510 by Maeder et al. ("Maeder"), which is incorporated by reference herein, describes a genome editing system for correcting a point mutation (C.2991+1655A to G) in the human CEP290 gene that results in the creation of a cryptic splice site, which in turn reduces or eliminates the function of the gene. The genome editing system of Maeder utilizes two guide RNAs targeted to sequences on either side of (i.e., flanking) the point mutation, and forms DSBs that flank the mutation. This, in turn, promotes deletion of the intervening sequence, including the mutation, thereby eliminating the cryptic splice site and restoring normal gene function.

[0155] As another example, WO 2016/073990 by Cotta-Ramusino et al. ("Cotta-Ramusino"), which is incorporated by reference herein, describes a genome editing system that utilizes two gRNAs in combination with a Cas9 nickase (a Cas9 that makes a single strand nick such as *S. pyogenes* D10A), an arrangement termed a "dual-nickase system." The dual-nickase system of

Cotta-Ramusino is configured to make two nicks on opposite strands of a sequence of interest that are offset by one or more nucleotides, which nicks combine to create a double strand break having an overhang (5' in the case of Cotta-Ramusino, though 3' overhangs are also possible). The overhang, in turn, can facilitate homology directed repair events in some circumstances. And, as another example, International Patent Publication No. WO 2015/070083 by Palestrant et al. (incorporated by reference herein) describes a gRNA targeted to a nucleotide sequence encoding Cas9 (referred to as a "governing RNA"), which can be included in a genome editing system comprising one or more additional gRNAs to permit transient expression of a Cas9 that might otherwise be constitutively expressed, for example in some virally transduced cells. These multiplexing applications are intended to be exemplary, rather than limiting, and the skilled artisan will appreciate that other applications of multiplexing are generally compatible with the genome editing systems described here.

[0156] As disclosed herein, in certain embodiments, genome editing systems may comprise multiple gRNAs that may be used to introduce mutations into the GATA1 binding motif in *BCL11Ae*, the CCAAT box target region, or the 13 nt target region of *HBG1* and/or *HBG2*. In certain embodiments, genome editing systems disclosed herein may comprise multiple gRNAs used to introduce mutations into the GATA1 binding motif in *BCL11Ae* and the 13 nt target region of *HBG1* and/or *HBG2*.

[0157] Genome editing systems can, in some instances, form double strand breaks that are repaired by cellular DNA double-strand break mechanisms such as NHEJ or HDR. These mechanisms are described throughout the literature (see, e.g., Davis 2014 (describing Alt-HDR), Frit 2014 (describing Alt-NHEJ), and Iyama 2013 (describing canonical HDR and NHEJ pathways generally), all of which are incorporated by reference herein).

[0158] Where genome editing systems operate by forming DSBs, such systems optionally include one or more components that promote or facilitate a particular mode of double-strand break repair or a particular repair outcome. For instance, Cotta-Ramusino also describes genome editing systems in which a single stranded oligonucleotide "donor template" is added; the donor template is incorporated into a target region of cellular DNA that is cleaved by the genome editing system, and can result in a change in the target sequence.

[0159] In certain embodiments, genome editing systems modify a target sequence, or modify expression of a gene in or near the target sequence, without causing single- or double-strand

breaks. For example, a genome editing system may include an RNA-guided nuclease fused to a functional domain that acts on DNA, thereby modifying the target sequence or its expression. As one example, an RNA-guided nuclease can be connected to (e.g., fused to) a cytidine deaminase functional domain, and may operate by generating targeted C-to-A substitutions. Exemplary nuclease/deaminase fusions are described in Komor 2016, which is incorporated by reference herein. Alternatively, a genome editing system may utilize a cleavage-inactivated (i.e., a "dead") nuclease, such as a dead Cas9 (dCas9), and may operate by forming stable complexes on one or more targeted regions of cellular DNA, thereby interfering with functions involving the targeted region(s) including, without limitation, mRNA transcription, chromatin remodeling, etc. In certain embodiments, a genome editing system may include an RNA-guided helicase that unwinds DNA within or proximal to the target sequence, without causing single- or double-stranded breaks. For example a genome editing system may include an RNA-guided helicase configured to associate within or near the target sequence to unwind DNA and induce accessibility to the target sequence. In certain embodiments, the RNA-guided helicase may be complexed to a dead guide RNA that is configured to lack cleavage activity allowing for unwinding of the DNA without causing breaks in the DNA.

Guide RNA (gRNA) molecules

[0160] The terms "guide RNA" and "gRNA" refer to any nucleic acid that promotes the specific association (or "targeting") of an RNA-guided nuclease such as a Cas9 or a Cpf1 to a target sequence such as a genomic or episomal sequence in a cell. gRNAs can be unimolecular (comprising a single RNA molecule, and referred to alternatively as chimeric), or modular (comprising more than one, and typically two, separate RNA molecules, such as a crRNA and a tracrRNA, which are usually associated with one another, for instance by duplexing). gRNAs and their component parts are described throughout the literature, for instance in Briner 2014, which is incorporated by reference), and in Cotta-Ramusino. Examples of modular and unimolecular gRNAs that may be used according to the embodiments herein include, without limitation, the sequences set forth in SEQ ID NOs:29-31 and 38-51. Examples of gRNA proximal and tail domains that may be used according to the embodiments herein include, without limitation, the sequences set forth in SEQ ID NOs:32-37.

[0161] In bacteria and archaea, type II CRISPR systems generally comprise an RNA-guided nuclease protein such as Cas9, a CRISPR RNA (crRNA) that includes a 5' region that is

complementary to a foreign sequence, and a trans-activating crRNA (tracrRNA) that includes a 5' region that is complementary to, and forms a duplex with, a 3' region of the crRNA. While not intending to be bound by any theory, it is thought that this duplex facilitates the formation of — and is necessary for the activity of — the Cas9/gRNA complex. As type II CRISPR systems were adapted for use in gene editing, it was discovered that the crRNA and tracrRNA could be joined into a single unimolecular or chimeric guide RNA, in one non-limiting example, by means of a four nucleotide (e.g., GAAA) "tetraloop" or "linker" sequence bridging complementary regions of the crRNA (at its 3' end) and the tracrRNA (at its 5' end) (Mali 2013; Jiang 2013; Jinek 2012; all incorporated by reference herein).

[0162] Guide RNAs, whether unimolecular or modular, include a "targeting domain" that is fully or partially complementary to a target domain within a target sequence, such as a DNA sequence in the genome of a cell where editing is desired. Targeting domains are referred to by various names in the literature, including without limitation "guide sequences" (Hsu et al., Nat Biotechnol. 2013 Sep; 31(9): 827–832, ("Hsu"), incorporated by reference herein), "complementarity regions" (Cotta-Ramusino), "spacers" (Briner 2014) and generically as "crRNAs" (Jiang). Irrespective of the names they are given, targeting domains are typically 10-30 nucleotides in length, and in certain embodiments are 16-24 nucleotides in length (for instance, 16, 17, 18, 19, 20, 21, 22, 23 or 24 nucleotides in length), and are at or near the 5' terminus of in the case of a Cas9 gRNA, and at or near the 3' terminus in the case of a Cpf1 gRNA.

[0163] In addition to the targeting domains, gRNAs typically (but not necessarily, as discussed below) include a plurality of domains that may influence the formation or activity of gRNA/Cas9 complexes. For instance, as mentioned above, the duplexed structure formed by first and secondary complementarity domains of a gRNA (also referred to as a repeat:anti-repeat duplex) interacts with the recognition (REC) lobe of Cas9 and can mediate the formation of Cas9/gRNA complexes (Nishimasu et al., Cell 156, 935-949, February 27, 2014 ("Nishimasu 2014") and Nishimasu et al., Cell 162, 1113-1126, August 27, 2015 ("Nishimasu 2015"), both incorporated by reference herein. It should be noted that the first and/or second complementarity domains may contain one or more poly-A tracts, which can be recognized by RNA polymerases as a termination signal. The sequence of the first and second complementarity domains are, therefore, optionally modified to eliminate these tracts and promote the complete *in vitro*

transcription of gRNAs, for instance through the use of A-G swaps as described in Briner 2014, or A-U swaps. These and other similar modifications to the first and second complementarity domains are within the scope of the present disclosure.

[0164] Along with the first and second complementarity domains, Cas9 gRNAs typically include two or more additional duplexed regions that are involved in nuclease activity *in vivo* but not necessarily *in vitro*. (Nishimasu 2015). A first stem-loop one near the 3' portion of the second complementarity domain is referred to variously as the "proximal domain," (Cotta-Ramusino) "stem loop 1" (Nishimasu 2014 and 2015) and the "nexus" (Briner 2014). One or more additional stem loop structures are generally present near the 3' end of the gRNA, with the number varying by species: *S. pyogenes* gRNAs typically include two 3' stem loops (for a total of four stem loop structures including the repeat:anti-repeat duplex), while *S. aureus* and other species have only one (for a total of three stem loop structures). A description of conserved stem loop structures (and gRNA structures more generally) organized by species is provided in Briner 2014.

[0165] While the foregoing description has focused on gRNAs for use with Cas9, it should be appreciated that other RNA-guided nucleases exist which utilize gRNAs that differ in some ways from those described to this point. For instance, Cpf1 ("CRISPR from Prevotella and Francisella 1") is a recently discovered RNA-guided nuclease that does not require a tracrRNA to function (Zetsche 2015b, incorporated by reference herein). A gRNA for use in a Cpf1 genome editing system generally includes a targeting domain and a complementarity domain (alternately referred to as a "handle"). It should also be noted that, in gRNAs for use with Cpf1, the targeting domain is usually present at or near the 3' end, rather than the 5' end as described above in connection with Cas9 gRNAs (the handle is at or near the 5' end of a Cpf1 gRNA).

[0166] Those of skill in the art will appreciate, however, that although structural differences may exist between gRNAs from different prokaryotic species, or between Cpf1 and Cas9 gRNAs, the principles by which gRNAs operate are generally consistent. Because of this consistency of operation, gRNAs can be defined, in broad terms, by their targeting domain sequences, and skilled artisans will appreciate that a given targeting domain sequence can be incorporated in any suitable gRNA, including a unimolecular or chimeric gRNA, or a gRNA that includes one or more chemical modifications and/or sequential modifications (substitutions, additional

nucleotides, truncations, etc.). Thus, for economy of presentation in this disclosure, gRNAs may be described solely in terms of their targeting domain sequences.

[0167] More generally, skilled artisans will appreciate that some aspects of the present disclosure relate to systems, methods and compositions that can be implemented using multiple RNA-guided nucleases. For this reason, unless otherwise specified, the term gRNA should be understood to encompass any suitable gRNA that can be used with any RNA-guided nuclease, and not only those gRNAs that are compatible with a particular species of Cas9 or Cpf1. By way of illustration, the term gRNA can, in certain embodiments, include a gRNA for use with any RNA-guided nuclease occurring in a Class 2 CRISPR system, such as a type II or type V or CRISPR system, or an RNA-guided nuclease derived or adapted therefrom.

[0168] In certain embodiments, the gRNAs and dgRNAs described herein may be prepared using art-known methods, including, without limitation, chemical synthesis and *in vitro* transcription (IVT). IVT typically involves the transcription of RNA from a DNA template by means of a bacterial RNA polymerase such as T7 polymerase. Chemical synthesis methods are described throughout the literature, for instance by Beaucage and Carruthers, *Curr Protoc Nucleic Acid Chem.* 2001 May; Chapter 3: Unit 3.3 (Beaucage & Carruthers), which is incorporated by reference in its entirety herein. Other methods of synthesis that may be used to prepare the gRNAs and dgRNAs described herein are set forth in International Patent Application No. PCT/US17/69019 filed December 29, 2017, which is incorporated by reference in its entirety herein.

gRNA design

[0169] Methods for selection and validation of target sequences as well as off-target analyses have been described previously (see, e.g., Mali 2013; Hsu 2013; Fu 2014; Heigwer 2014; Bae 2014; Xiao 2014; all incorporated by reference herein). As a non-limiting example, gRNA design may involve the use of a software tool to optimize the choice of potential target sequences corresponding to a user's target sequence, e.g., to minimize total off-target activity across the genome. While off-target activity is not limited to cleavage, the cleavage efficiency at each off-target sequence can be predicted, e.g., using an experimentally-derived weighting scheme.

These and other guide selection methods are described in detail in Maeder and Cotta-Ramusino.

[0170] With respect to selection of gRNA targeting domain sequences directed to HBG1/2 target sites (e.g., the 13 nt target region), an in-silico gRNA target domain identification tool was

utilized, and the hits were stratified into four tiers. For *S. pyogenes*, tier 1 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1/2 13 nt target region), specifically within 400 bp of either end of the target site, (2) a high level of orthogonality, and (3) the presence of 5' G. Tier 2 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1/2 13 nt target region), specifically within 400 bp of either end of the target site, and (2) a high level of orthogonality. Tier 3 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1/2 13 nt target region), specifically within 400 bp of either end of the target site and (2) the presence of 5' G. Tier 4 targeting domains were selected based on distance upstream or downstream from either end of the target site (i.e., HBG1/2 13 nt target region), specifically within 400 bp of either end of the target site.

[0171] For *S. aureus*, tier 1 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1/2 13 nt target region), specifically within 400 bp of either end of the target site, (2) a high level of orthogonality, (3) the presence of 5' G, and (4) PAM having the sequence NNGRRT (SEQ ID NO:204). Tier 2 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1/2 13 nt target), specifically within 400 bp of either end of the target site, (2) a high level of orthogonality, and (3) PAM having the sequence NNGRRT (SEQ ID NO:204). Tier 3 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1/2 13 nt target region), specifically within 400 bp of either end of the target site, and (2) PAM having the sequence NNGRRT (SEQ ID NO:204). Tier 4 targeting domains were selected based on (1) distance upstream or downstream from either end of the target site (i.e., HBG1/2 13 nt target), specifically within 400 bp of either end of the target site, and (2) PAM having the sequence NNGRRV (SEQ ID NO:205).

[0172] Table 2, below, presents targeting domains for *S. pyogenes* and *S. aureus* gRNAs, broken out by (a) tier (1, 2, 3 or 4) and (b) HBG1 or HBG2.

Table 2: gRNA targeting domain sequences for HBG1/2 target sites

		HBG1	HBG2
<i>S. pyogenes</i>	Tier 1	251-256	760-764
	Tier 2	257-274	765-781
	Tier 3	275-300	275-281, 283-300
	Tier 4	301-366	301-311, 313-342, 344-348, 350-366, 782, 783
<i>S. aureus</i>	Tier 1	367-376	784-791
	Tier 2	343, 377-393	778, 792-803
	Tier 3	357, 365, 394-461	357, 365, 394-461
	Tier 4	252-254, 256, 268, 272-274, 292, 295, 347, 348, 353, 360-362, 366, 598-759	292, 295, 347, 348, 353, 360-362, 366, 462-468 476-481, 489-587, 601-607, 614-620, 640-666, 674-679, 687-693, 708-714, 733-753, 762-764, 775, 779-781, 804-901

[0173] gRNAs may be designed to target the erythroid specific enhancer of *BCL11A* (*BCL11Ae*) to disrupt expression of a transcriptional repressor, BCL11A (Friedland). gRNAs were designed to target the GATA1 binding motif that is in the erythroid specific enhancer of *BCL11A* that is in the +58 DHS region of intron 2 (i.e., the GATA1 binding motif in *BCL11Ae*), where the +58 DHS enhancer region comprises the sequence set forth in SEQ ID NO:968. Targeting domain sequences of gRNAs that were designed to target disruption of the GATA1 binding motif in *BCL11Ae*, include, but are not limited to, the sequences set forth in SEQ ID NOs:952-955. Targeting domain sequences plus PAM (NGG) of gRNAs that were designed to target disruption of the GATA1 binding motif in *BCL11Ae*, include, but are not limited to, the sequences set forth in SEQ ID NOs:960-963.

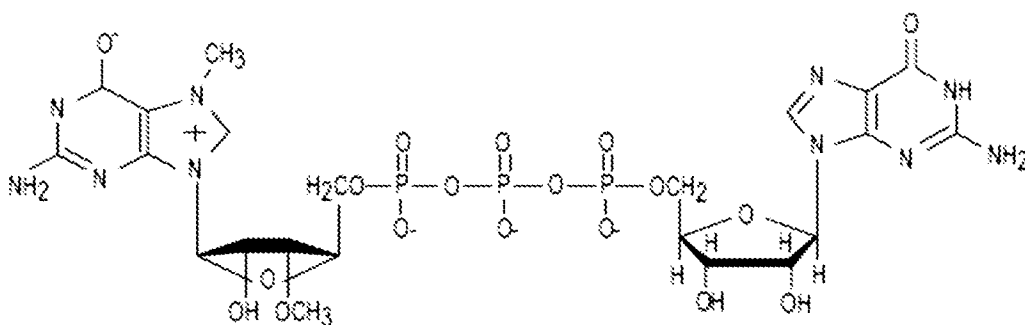
gRNA modifications

[0174] The activity, stability, or other characteristics of gRNAs can be altered through the incorporation of certain modifications. As one example, transiently expressed or delivered nucleic acids can be prone to degradation by, e.g., cellular nucleases. Accordingly, the 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 believed that certain modified gRNAs described herein can exhibit a reduced innate immune response when introduced into cells. Those of skill in the art will be aware of certain cellular responses commonly observed in cells, e.g., mammalian cells, in response to exogenous nucleic acids,

particularly those of viral or bacterial origin. Such responses, which can include induction of cytokine expression and release and cell death, may be reduced or eliminated altogether by the modifications presented herein.

[0175] Certain exemplary modifications discussed in this section can be included at any position within a gRNA sequence including, without limitation at or near the 5' end (e.g., within 1-10, 1-5, or 1-2 nucleotides of the 5' end) and/or at or near the 3' end (e.g., within 1-10, 1-5, or 1-2 nucleotides of the 3' end). In some cases, modifications are positioned within functional motifs, such as the repeat-anti-repeat duplex of a Cas9 gRNA, a stem loop structure of a Cas9 or Cpf1 gRNA, and/or a targeting domain of a gRNA.

[0176] As one example, the 5' end of a gRNA can include 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)), as shown below:



The cap or cap analog can be included during either chemical synthesis or *in vitro* transcription of the gRNA.

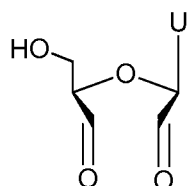
[0177] Along similar lines, the 5' end of the gRNA can lack a 5' triphosphate group. For instance, *in vitro* transcribed gRNAs can be phosphatase-treated (e.g., using calf intestinal alkaline phosphatase) to remove a 5' triphosphate group.

[0178] Another common modification involves the addition, at the 3' end of a gRNA, of a plurality (e.g., 1-10, 10-20, or 25-200) of adenine (A) residues referred to as a polyA tract. The polyA tract can be added to a gRNA during chemical synthesis, following *in vitro* transcription using a polyadenosine polymerase (e.g., *E. coli* Poly(A)Polymerase), or *in vivo* by means of a polyadenylation sequence, as described in Maeder.

[0179] It should be noted that the modifications described herein can be combined in any suitable manner, e.g., a gRNA, whether transcribed *in vivo* from a DNA vector, or *in vitro*

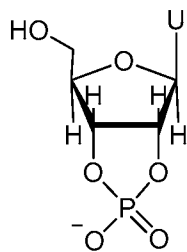
transcribed gRNA, can include either or both of a 5' cap structure or cap analog and a 3' polyA tract.

[0180] Guide RNAs 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.

[0181] The 3' terminal U ribose can be modified with a 2'3' cyclic phosphate as shown below:



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

[0182] Guide RNAs can contain 3' nucleotides which can be stabilized against degradation, e.g., by incorporating one or more of the modified nucleotides described herein. In certain embodiments, 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.

[0183] In certain embodiments, sugar-modified ribonucleotides can be incorporated into the gRNA, 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 certain embodiments, the phosphate backbone can be modified as described herein, e.g., with a phosphothioate (PhTx) group. In

certain 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.

[0184] Guide RNAs can also 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. Any suitable moiety can be used to provide such bridges, include without limitation 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).

[0185] In certain 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')).

[0186] Generally, gRNAs 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 certain embodiments, a gRNA comprises a 4'-S, 4'-Se or a 4'-C-aminomethyl-2'-O-Me modification.

[0187] In certain embodiments, deaza nucleotides, e.g., 7-deaza-adenosine, can be incorporated into the gRNA. In certain embodiments, O- and N-alkylated nucleotides, e.g., N6-methyl adenosine, can be incorporated into the gRNA. In certain embodiments, one or more or all of the nucleotides in a gRNA are deoxynucleotides.

Dead gRNA molecules

[0188] Dead guide RNA (dgRNA) molecules according to the present disclosure include dead guide RNA molecules that comprise reduced, low, or undetectable cleavage activity. The targeting domain sequences of dead guide RNAs are shorter in length by 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 nucleotides compared to the targeting domain sequence of active guide RNAs. In certain embodiments, dead guide RNA molecules may comprise a targeting domain comprising 15 nucleotides or fewer in length, 14 nucleotides or fewer in length, 13 nucleotides or fewer in length, 12 nucleotides or fewer in length, or 11 nucleotides or fewer in length. In some embodiments, dead guide RNAs are configured such that they do not provide an RNA guided-nuclease cleavage event. Dead guide RNAs may be generated by removing the 5' end of a gRNA targeting domain sequence, which results in a truncated targeting domain sequence. For example, if a gRNA sequence, configured to provide a cleavage event (i.e., 17 nucleotides or more in length), has a targeting domain sequence that is 20 nucleotides in length, a dead guide RNA may be created by removing 5 nucleotides from the 5' end of the gRNA sequence. For example, dgRNAs used herein may comprise a targeting domain set forth in Table 2 or Table 10 that has been truncated from the 5' end of the gRNA sequence and comprises 15 nucleotides or fewer in length. In certain embodiments, the dgRNA may be configured to bind (or associate with) a nucleic acid sequence within or proximal to a target region (e.g., the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*) to be edited. For example, any of the dgRNAs set forth in Table 10 may be employed to bind a nucleic acid sequence proximal to the 13 nt target region or CCAAT box target region. In certain embodiments, proximal to may denote the region within 10, 25, 50, 100, or 200 nucleotides of a target region (e.g., the CCAAT box target region, 13 nt target region, proximal HBG1/2 promoter target sequence, and/or the GATA1 binding motif in *BCL11Ae*). In certain embodiments, the dead guide RNA is not configured to recruit an exogenous trans-acting factor to a target region. In certain embodiments, the dgRNA is configured such that it does not provide a DNA cleavage event when complexed with an RNA-

guided nuclease. Skilled artisans will appreciate that dead guide RNA molecules may be designed to comprise targeting domains complementary to regions proximal to or within a target region in a target nucleic acid. In certain embodiments, dead guide RNAs comprise targeting domain sequences that are complementary to the transcription strand or non-transcription strand of double stranded DNA. The dgRNAs herein may include modifications at the 5' and 3' end of the dgRNA as described for guide RNAs in the section "gRNA modifications" herein. For example, in certain embodiments, dead guide RNAs may include an anti-reverse cap analog (ARCA) at the 5' end of the RNA. In certain embodiments, dgRNAs may include a polyA tail at the 3' end.

[0189] In certain embodiments, the use of a dead guide RNA with the genome editing systems and methods disclosed herein may increase the total editing level of an active guide RNA. In certain embodiments, the use of a dead guide RNA with the genome editing systems disclosed herein and methods thereof may increase the frequency of deletions. In certain embodiments, the deletions may extend from the cut site of the active guide RNA toward the dead guide RNA binding site. In this way the dead guide RNA can change the directionality of an active guide RNA and orient editing toward a desired target region.

RNA-guided nucleases

[0190] RNA-guided nucleases according to the present disclosure include, but are not limited to, naturally-occurring Class 2 CRISPR nucleases such as Cas9, and Cpf1, as well as other nucleases derived or obtained therefrom. In functional terms, RNA-guided nucleases are defined as those nucleases that: (a) interact with (e.g., complex with) a gRNA; and (b) together with the gRNA, associate with, and optionally cleave or modify, a target region of a DNA that includes (i) a sequence complementary to the targeting domain of the gRNA and, optionally, (ii) an additional sequence referred to as a "protospacer adjacent motif," or "PAM," which is described in greater detail below. As the following examples will illustrate, RNA-guided nucleases can be defined, in broad terms, by their PAM specificity and cleavage activity, even though variations may exist between individual RNA-guided nucleases that share the same PAM specificity or cleavage activity. Skilled artisans will appreciate that some aspects of the present disclosure relate to systems, methods and compositions that can be implemented using any suitable RNA-guided nuclease having a certain PAM specificity and/or cleavage activity. For this reason, unless otherwise specified, the term RNA-guided nuclease should be understood as a generic

term, and not limited to any particular type (e.g., Cas9 vs. Cpf1), species (e.g., *S. pyogenes* vs. *S. aureus*) or variation (e.g., full-length vs. truncated or split; naturally-occurring PAM specificity vs. engineered PAM specificity, etc.) of RNA-guided nuclease.

[0191] The PAM sequence takes its name from its sequential relationship to the "protospacer" sequence that is complementary to gRNA targeting domains (or "spacers"). Together with protospacer sequences, PAM sequences define target regions or sequences for specific RNA-guided nuclease / gRNA combinations.

[0192] Various RNA-guided nucleases may require different sequential relationships between PAMs and protospacers. In general, Cas9s recognize PAM sequences that are 3' of the protospacer as visualized relative to the top or complementary strand.

[0193] Cpf1, on the other hand, generally recognizes PAM sequences that are 5' of the protospacer.

In addition to recognizing specific sequential orientations of PAMs and protospacers, RNA-guided nucleases can also recognize specific PAM sequences. *S. aureus* Cas9, for instance, recognizes a PAM sequence of NNGRRT or NNGRRV, wherein the N residues are immediately 3' of the region recognized by the gRNA targeting domain. *S. pyogenes* Cas9 recognizes NGG PAM sequences. And *F. novicida* Cpf1 recognizes a TTN PAM sequence. PAM sequences have been identified for a variety of RNA-guided nucleases, and a strategy for identifying novel PAM sequences has been described by Shmakov 2015. It should also be noted that engineered RNA-guided nucleases can have PAM specificities that differ from the PAM specificities of reference molecules (for instance, in the case of an engineered RNA-guided nuclease, the reference molecule may be the naturally occurring variant from which the RNA-guided nuclease is derived, or the naturally occurring variant having the greatest amino acid sequence homology to the engineered RNA-guided nuclease). Examples of PAMs that may be used according to the embodiments herein include, without limitation, the sequences set forth in SEQ ID NOs:199-205. In certain embodiments, a PAM sequence in a target nucleic acid may be configured such that the PAM is facing outward (e.g., PAM-out). In certain embodiments, a PAM facing outward may be a PAM sequence that is located outside of the target region, with each protospacer (or genomic region with homology to the guide) falling within the region encompassed by the two respective PAMs.

[0194] In addition to their PAM specificity, RNA-guided nucleases can be characterized by their DNA cleavage activity: naturally-occurring RNA-guided nucleases typically form DSBs in target nucleic acids, but engineered variants have been produced that generate only SSBs (discussed above; see also Ran 2013, incorporated by reference herein), or that do not cut at all.

Cas9

[0195] Crystal structures have been determined for *S. pyogenes* Cas9 (Jinek 2014), and for *S. aureus* Cas9 in complex with a unimolecular guide RNA and a target DNA (Nishimasu 2014; Anders 2014; and Nishimasu 2015).

[0196] A naturally occurring Cas9 protein comprises two lobes: a recognition (REC) lobe and a nuclease (NUC) lobe; each of which comprise particular structural and/or functional domains. The REC lobe comprises an arginine-rich bridge helix (BH) domain, and at least one REC domain (e.g., a REC1 domain and, optionally, a REC2 domain). The REC lobe does not share structural similarity with other known proteins, indicating that it is a unique functional domain. While not wishing to be bound by any theory, mutational analyses suggest specific functional roles for the BH and REC domains: the BH domain appears to play a role in gRNA:DNA recognition, while the REC domain is thought to interact with the repeat:anti-repeat duplex of the gRNA and to mediate the formation of the Cas9/gRNA complex.

[0197] The NUC lobe comprises a RuvC domain, an HNH domain, and a PAM-interacting (PI) domain. The RuvC domain shares structural similarity to retroviral integrase superfamily members and cleaves the non-complementary (i.e., bottom) strand of the target nucleic acid. It may be formed from two or more split RuvC motifs (such as RuvC I, RuvCII, and RuvCIII in *S. pyogenes* and *S. aureus*). The HNH domain, meanwhile, is structurally similar to HNH endonuclease motifs, and cleaves the complementary (i.e., top) strand of the target nucleic acid. The PI domain, as its name suggests, contributes to PAM specificity. Examples of polypeptide sequences encoding Cas9 RuvC-like and Cas9 HNH-like domains that may be used according to the embodiments herein are set forth in SEQ ID NOs:15-23, 52-123 (RuvC-like domains) and SEQ ID NOs:24-28, 124-198 (HNH-like domains).

[0198] While certain functions of Cas9 are linked to (but not necessarily fully determined by) the specific domains set forth above, these and other functions may be mediated or influenced by other Cas9 domains, or by multiple domains on either lobe. For instance, in *S. pyogenes* Cas9, as described in Nishimasu 2014, the repeat:antirepeat duplex of the gRNA falls into a groove

between the REC and NUC lobes, and nucleotides in the duplex interact with amino acids in the BH, PI, and REC domains. Some nucleotides in the first stem loop structure also interact with amino acids in multiple domains (PI, BH and REC1), as do some nucleotides in the second and third stem loops (RuvC and PI domains). Examples of polypeptide sequences encoding Cas9 molecules that may be used according to the embodiments herein are set forth in SEQ ID NOs: 1-2, 4-6, 12, 14.

Cpf1

[0199] The crystal structure of *Acidaminococcus sp.* Cpf1 in complex with crRNA and a double-stranded (ds) DNA target including a TTTN PAM sequence has been solved (Yamano 2016, incorporated by reference herein). Cpf1, like Cas9, has two lobes: a REC (recognition) lobe, and a NUC (nuclease) lobe. The REC lobe includes REC1 and REC2 domains, which lack similarity to any known protein structures. The NUC lobe, meanwhile, includes three RuvC domains (RuvC-I, -II and -III) and a BH domain. However, in contrast to Cas9, the Cpf1 REC lobe lacks an HNH domain, and includes other domains that also lack similarity to known protein structures: a structurally unique PI domain, three Wedge (WED) domains (WED-I, -II and -III), and a nuclease (Nuc) domain.

[0200] While Cas9 and Cpf1 share similarities in structure and function, it should be appreciated that certain Cpf1 activities are mediated by structural domains that are not analogous to any Cas9 domains. For instance, cleavage of the complementary strand of the target DNA appears to be mediated by the Nuc domain, which differs sequentially and spatially from the HNH domain of Cas9. Additionally, the non-targeting portion of Cpf1 gRNA (the handle) adopts a pseudoknot structure, rather than a stem loop structure formed by the repeat:antirepeat duplex in Cas9 gRNAs.

Modifications of RNA-guided nucleases

[0201] The RNA-guided nucleases described above have activities and properties that can be useful in a variety of applications, but the skilled artisan will appreciate that RNA-guided nucleases can also be modified in certain instances, to alter cleavage activity, PAM specificity, or other structural or functional features.

[0202] Turning first to modifications that alter cleavage activity, mutations that reduce or eliminate the activity of domains within the NUC lobe have been described above. Exemplary mutations that may be made in the RuvC domains, in the Cas9 HNH domain, or in the Cpf1 Nuc

domain are described in Ran 2013 and Yamano 2016, as well as in Cotta-Ramusino. In general, mutations that reduce or eliminate activity in one of the two nuclease domains result in RNA-guided nucleases with nickase activity, but it should be noted that the type of nickase activity varies depending on which domain is inactivated. As one example, inactivation of a RuvC domain of a Cas9 will result in a nickase that cleaves the complementary or top strand.

[0203] On the other hand, inactivation of a Cas9 HNH domain results in a nickase that cleaves the bottom or non-complementary strand.

[0204] Modifications of PAM specificity relative to naturally occurring Cas9 reference molecules has been described for both *S. pyogenes* (Kleinstiver 2015a) and *S. aureus* (Kleinstiver 2015b). Modifications that improve the targeting fidelity of Cas9 have also been described (Kleinstiver 2016). Each of these references is incorporated by reference herein.

[0205] RNA-guided nucleases have been split into two or more parts (see, e.g., Zetsche 2015a; Fine 2015; both incorporated by reference).

[0206] RNA-guided nucleases can be, in certain embodiments, size-optimized or truncated, for instance via one or more deletions that reduce the size of the nuclease while still retaining gRNA association, target and PAM recognition, and cleavage activities. In certain embodiments, RNA guided nucleases are bound, covalently or non-covalently, to another polypeptide, nucleotide, or other structure, optionally by means of a linker. Exemplary bound nucleases and linkers are described by Guilinger 2014, which is incorporated by reference herein.

[0207] RNA-guided nucleases also optionally include a tag, such as, but not limited to, a nuclear localization signal to facilitate movement of RNA-guided nuclease protein into the nucleus. In certain embodiments, the RNA-guided nuclease can incorporate C- and/or N-terminal nuclear localization signals. Nuclear localization sequences are known in the art and are described in Maeder and elsewhere.

[0208] The foregoing list of modifications is intended to be exemplary in nature, and the skilled artisan will appreciate, in view of the instant disclosure, that other modifications may be possible or desirable in certain applications. For brevity, therefore, exemplary systems, methods and compositions of the present disclosure are presented with reference to particular RNA-guided nucleases, but it should be understood that the RNA-guided nucleases used may be modified in ways that do not alter their operating principles. Such modifications are within the scope of the present disclosure.

RNA-guided helicases

[0209] RNA-guided helicases according to the present disclosure include, but are not limited to, naturally-occurring RNA-guided helicases that are capable of unwinding nucleic acid. As discussed *supra*, catalytically active RNA-guided nucleases cleave or modify a target region of DNA. It has also been shown that certain RNA-guided nucleases, such as Cas9, also have helicase activity that enables them to unwind nucleic acid. In certain embodiments, the RNA-guided helicases according to the present disclosure may be any of the RNA-nucleases described herein and *supra* in the section entitled "RNA-guided nucleases." In certain embodiments, the RNA-guided nuclease is not configured to recruit an exogenous trans-acting factor to a target region. In certain embodiments, an RNA-guided helicase may be an RNA-guided nuclease configured to lack nuclease activity. For example, in certain embodiments, an RNA-guided helicase may be a catalytically inactive RNA-guided nuclease that lacks nuclease activity, but still retains its helicase activity. In certain embodiments, an RNA-guided nuclease may be mutated to abolish its nuclease activity (*e.g.*, dead Cas9), creating a catalytically inactive RNA-guided nuclease that is unable to cleave nucleic acid, but which can still unwind DNA. In certain embodiments, an RNA-guided helicase may be complexed with any of the dead guide RNAs as described herein. For example, a catalytically active RNA-guided helicase (*e.g.*, Cas9 or Cpf1) may form an RNP complex with a dead guide RNA, resulting in a catalytically inactive dead RNP (dRNP). In certain embodiments, a catalytically inactive RNA-guided helicase (*e.g.*, dead Cas9) and a dead guide RNA may form a dRNP. These dRNPs, although incapable of providing a cleavage event, still retain their helicase activity that is important for unwinding nucleic acid.

Nucleic acids encoding RNA-guided nucleases

[0210] Nucleic acids encoding RNA-guided nucleases, *e.g.*, Cas9, Cpf1 or functional fragments thereof, are provided herein. Examples of nucleic acid sequences encoding Cas9 molecules that may be used according to the embodiments herein are set forth in SEQ ID NOs:3, 7-11, 13. Exemplary nucleic acids encoding RNA-guided nucleases have been described previously (see, *e.g.*, Cong 2013; Wang 2013; Mali 2013; Jinek 2012).

[0211] In some cases, a nucleic acid encoding an RNA-guided nuclease can be a synthetic nucleic acid sequence. For example, the synthetic nucleic acid molecule can be chemically modified. In certain embodiments, an mRNA encoding an RNA-guided nuclease will have one

or more (e.g., all) of the following properties: it can be capped; polyadenylated; and substituted with 5-methylcytidine and/or pseudouridine.

[0212] Synthetic nucleic acid sequences can also 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. Examples of codon optimized Cas9 coding sequences are presented in Cotta-Ramusino.

[0213] In addition, or alternatively, a nucleic acid encoding an RNA-guided nuclease may comprise a nuclear localization sequence (NLS). Nuclear localization sequences are known in the art.

Functional analysis of candidate molecules

[0214] Candidate RNA-guided nucleases, gRNAs, and complexes thereof, can be evaluated by standard methods known in the art (see, e.g., Cotta-Ramusino). The stability of RNP complexes may be evaluated by differential scanning fluorimetry, as described below.

Differential Scanning Fluorimetry (DSF)

[0215] The thermostability of ribonucleoprotein (RNP) complexes comprising gRNAs and RNA-guided nucleases can be measured via DSF. The DSF 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.

[0216] A DSF assay can be performed according to any suitable protocol, and can be employed in any suitable setting, including without limitation (a) testing different conditions (e.g., different stoichiometric ratios of gRNA: RNA-guided nuclease protein, different buffer solutions, etc.) to identify optimal conditions for RNP formation; and (b) testing modifications (e.g., chemical modifications, alterations of sequence, etc.) of an RNA-guided nuclease and/or a gRNA to identify those modifications that improve RNP formation or stability. One readout of a DSF assay is a shift in melting temperature of the RNP complex; a relatively high shift suggests that the RNP complex is more stable (and may thus have greater activity or more favorable kinetics of formation, kinetics of degradation, or another functional characteristic) relative to a reference RNP complex characterized by a lower shift. When the DSF assay is deployed as a screening tool, a threshold melting temperature shift may be specified, so that the output is one or more RNPs having a melting temperature shift at or above the threshold. For instance, the threshold

can be 5-10°C (e.g., 5°, 6°, 7°, 8°, 9°, 10°) or more, and the output may be one or more RNPs characterized by a melting temperature shift greater than or equal to the threshold.

[0217] Two non-limiting examples of DSF assay conditions are set forth below:

[0218] To determine the best solution to form RNP complexes, a fixed concentration (e.g., 2 µM) of Cas9 in water+10x SYPRO Orange® (Life Technologies cat#S-6650) is 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' 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.

[0219] The second assay consists of mixing various concentrations of gRNA with fixed concentration (e.g., 2 µM) Cas9 in optimal buffer from assay 1 above and incubating (e.g., at RT for 10') 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.

Genome editing strategies

[0220] The genome editing systems described above are used, in various embodiments of the present disclosure, to generate edits in (i.e., to alter) targeted regions of DNA within or obtained from a cell. Various strategies are described herein to generate particular edits, and these strategies are generally described in terms of the desired repair outcome, the number and positioning of individual edits (e.g., SSBs or DSBs), and the target sites of such edits.

[0221] Genome editing strategies that involve the formation of SSBs or DSBs are characterized by repair outcomes including: (a) deletion of all or part of a targeted region; (b) insertion into or replacement of all or part of a targeted region; or (c) interruption of all or part of a targeted region. This grouping is not intended to be limiting, or to be binding to any particular theory or model, and is offered solely for economy of presentation. Skilled artisans will appreciate that the listed outcomes are not mutually exclusive and that some repairs may result in other outcomes. The description of a particular editing strategy or method should not be understood to require a particular repair outcome unless otherwise specified.

[0222] Replacement of a targeted region generally involves the replacement of all or part of the existing sequence within the targeted region with a homologous sequence, for instance through gene correction or gene conversion, two repair outcomes that are mediated by HDR pathways. HDR is promoted by the use of a donor template, which can be single-stranded or double stranded, as described in greater detail below. Single or double stranded templates can be exogenous, in which case they will promote gene correction, or they can be endogenous (e.g., a homologous sequence within the cellular genome), to promote gene conversion. Exogenous templates can have asymmetric overhangs (i.e., the portion of the template that is complementary to the site of the DSB may be offset in a 3' or 5' direction, rather than being centered within the donor template), for instance as described by Richardson 2016 (incorporated by reference herein). In instances where the template is single stranded, it can correspond to either the complementary (top) or non-complementary (bottom) strand of the targeted region.

[0223] Gene conversion and gene correction are facilitated, in some cases, by the formation of one or more nicks in or around the targeted region, as described in Ran and Cotta-Ramusino. In some cases, a dual-nickase strategy is used to form two offset SSBs that, in turn, form a single DSB having an overhang (e.g., a 5' overhang).

[0224] Interruption and/or deletion of all or part of a targeted sequence can be achieved by a variety of repair outcomes. As one example, a sequence can be deleted by simultaneously generating two or more DSBs that flank a targeted region, which is then excised when the DSBs are repaired, as is described in Maeder for the LCA10 mutation. As another example, a sequence can be interrupted by a deletion generated by formation of a double strand break with single-stranded overhangs, followed by exonucleolytic processing of the overhangs prior to repair.

[0225] One specific subset of target sequence interruptions is mediated by the formation of an indel within the targeted sequence, where the repair outcome is typically mediated by NHEJ pathways (including Alt-NHEJ). NHEJ is referred to as an "error prone" repair pathway because of its association with indel mutations. In some cases, however, a DSB is repaired by NHEJ without alteration of the sequence around it (a so-called "perfect" or "scarless" repair); this generally requires the two ends of the DSB to be perfectly ligated. Indels, meanwhile, are thought to arise from enzymatic processing of free DNA ends before they are ligated that adds and/or removes nucleotides from either or both strands of either or both free ends.

[0226] Because the enzymatic processing of free DSB ends may be stochastic in nature, indel mutations tend to be variable, occurring along a distribution, and can be influenced by a variety of factors, including the specific target site, the cell type used, the genome editing strategy used, etc. Even so, it is possible to draw limited generalizations about indel formation: deletions formed by repair of a single DSB are most commonly in the 1-50 bp range, but can reach greater than 100-200 bp. Insertions formed by repair of a single DSB 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.

[0227] Indel mutations – and genome editing systems configured to produce indels – are useful for interrupting target sequences, for example, when the generation of a specific final sequence is not required and/or where a frameshift mutation would be tolerated. They can also be useful in settings where particular sequences are preferred, insofar as the certain sequences desired tend to occur preferentially from the repair of an SSB or DSB at a given site. Indel mutations are also a useful tool for evaluating or screening the activity of particular genome editing systems and their components. In these and other settings, indels can be characterized by (a) their relative and absolute frequencies in the genomes of cells contacted with genome editing systems and (b) the distribution of numerical differences relative to the unedited sequence, e.g., ± 1 , ± 2 , ± 3 , etc. As one example, in a lead-finding setting, multiple gRNAs can be screened to identify those gRNAs that most efficiently drive cutting at a target site based on an indel readout under controlled conditions. Guides that produce indels at or above a threshold frequency, or that produce a particular distribution of indels, can be selected for further study and development. Indel frequency and distribution can also be useful as a readout for evaluating different genome editing system implementations or formulations and delivery methods, for instance by keeping the gRNA constant and varying certain other reaction conditions or delivery methods.

Multiplex strategies

[0228] Genome editing systems according to this disclosure may also be employed for multiplex gene editing to generate two or more DSBs, either in the same locus or in different loci. Any of the RNA-guided nucleases and gRNAs disclosed herein may be used in genome editing systems for multiplex gene editing. Strategies for editing that involve the formation of multiple DSBs, or SSBs, are described in, for instance, Cotta-Ramusino.

[0229] As disclosed herein, multiple gRNAs may be used in genome editing systems to introduce alterations (e.g., deletions, insertions) into the 13 nt target region of *HBG1* and/or *HBG2*. In certain embodiments, one or more gRNAs comprising a targeting domain set forth in SEQ ID NOs:251-901, 940-942 may be used to introduce alterations in the 13 nt target region of *HBG1* and/or *HBG2*. In other embodiments, multiple gRNAs may be used in genome editing systems to introduce alterations into the GATA1 binding motif in *BCL11Ae*. In certain embodiments, one or more gRNAs comprising a targeting domain set forth in SEQ ID NOs:952-955 may be used to introduce alterations in the GATA1 binding motif in *BCL11Ae*. Multiple gRNAs may also be used in genome editing systems to introduce alterations into the GATA1 binding motif in *BCL11Ae* and the 13 nt target region of *HBG1* and/or *HBG2*. In certain embodiments, one or more gRNAs comprising a targeting domain set forth in SEQ ID NOs:952-955 may be used to introduce alterations in the GATA1 binding motif in *BCL11Ae* and one or more gRNAs comprising a targeting domain set forth in SEQ ID NOs:251-901, 940-942 may be used to introduce alterations in the 13 nt target region of *HBG1* and/or *HBG2*.

Donor template design

[0230] Donor template design is described in detail in the literature, for instance in Cotta-Ramusino. DNA oligomer donor templates (oligodeoxynucleotides or ODNs), which can be single stranded (ssODNs) or double-stranded (dsODNs), can be used to facilitate HDR-based repair of DSBs or to boost overall editing rate, and are particularly useful for introducing alterations into a target DNA sequence, inserting a new sequence into the target sequence, or replacing the target sequence altogether.

[0231] Whether single-stranded or double stranded, donor templates generally include regions that are homologous to regions of DNA within or near (e.g., flanking or adjoining) a target sequence to be cleaved. These homologous regions are referred to here as "homology arms," and are illustrated schematically below:

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

[0232] The homology arms can have any suitable length (including 0 nucleotides if only one homology arm is used), and 3' and 5' homology arms can have the same length, or can differ in length. The selection of appropriate homology arm lengths can be influenced by a variety of factors, such as the desire to avoid homologies or microhomologies with certain sequences such as Alu repeats or other very common elements. For example, a 5' homology arm can be

shortened to avoid a sequence repeat element. In other embodiments, a 3' homology arm can be shortened to avoid a sequence repeat element. In some embodiments, both the 5' and the 3' homology arms can be shortened to avoid including certain sequence repeat elements. In addition, some homology arm designs can improve the efficiency of editing or increase the frequency of a desired repair outcome. For example, Richardson 2016, which is incorporated by reference herein, found that the relative asymmetry of 3' and 5' homology arms of single stranded donor templates influenced repair rates and/or outcomes.

[0233] Replacement sequences in donor templates have been described elsewhere, including in Cotta-Ramusino et al. A replacement sequence can be any suitable length (including zero nucleotides, where the desired repair outcome is a deletion), and typically includes one, two, three or more sequence modifications relative to the naturally-occurring sequence within a cell in which editing is desired. One common sequence modification involves the alteration of the naturally-occurring sequence to repair a mutation that is related to a disease or condition of which treatment is desired. Another common sequence modification involves the alteration of one or more sequences that are complementary to, or then, the PAM sequence of the RNA-guided nuclease or the targeting domain of the gRNA(s) being used to generate an SSB or DSB, to reduce or eliminate repeated cleavage of the target site after the replacement sequence has been incorporated into the target site.

[0234] Where a linear ssODN is used, it can be configured to (i) anneal to the nicked strand of the target nucleic acid, (ii) anneal to the intact strand of the target nucleic acid, (iii) anneal to the plus strand of the target nucleic acid, and/or (iv) anneal to the minus strand of the target nucleic acid. An ssODN may have any suitable length, e.g., about, at least, or no more than 150-200 nucleotides (e.g., 150, 160, 170, 180, 190, or 200 nucleotides).

[0235] It should be noted that a template nucleic acid can also be a nucleic acid vector, such as a viral genome or circular double stranded DNA, e.g., a plasmid. Nucleic acid vectors comprising donor templates can include other coding or non-coding elements. For example, a template nucleic acid can be delivered as part of a viral genome (e.g., in an AAV or lentiviral genome) that includes certain genomic backbone elements (e.g., inverted terminal repeats, in the case of an AAV genome) and optionally includes additional sequences coding for a gRNA and/or an RNA-guided nuclease. In certain embodiments, the donor template can be adjacent to, or flanked by, target sites recognized by one or more gRNAs, to facilitate the formation of free

DSBs on one or both ends of the donor template that can participate in repair of corresponding SSBs or DSBs formed in cellular DNA using the same gRNAs. Exemplary nucleic acid vectors suitable for use as donor templates are described in Cotta-Ramusino, which is incorporated by reference.

[0236] Whatever format is used, a template nucleic acid can be designed to avoid undesirable sequences. In certain embodiments, one or both homology arms can be shortened to avoid overlap with certain sequence repeat elements, e.g., Alu repeats, LINE elements, etc.

[0237] In certain embodiments, silent, non-pathogenic SNPs may be included in the ssODN donor template to allow for identification of a gene editing event.

[0238] In certain embodiments, a donor template may be a non-specific template that is non-homologous to regions of DNA within or near a target sequence to be cleaved. In certain embodiments, donor templates for use in targeting the GATA1 binding motif in *BCL11Ae* may include, without limitation, non-target specific templates that are nonhomologous to regions of DNA within or near the GATA1 binding motif in *BCL11Ae*. In certain embodiments, donor templates for use in targeting the 13 nt target region may include, without limitation, non-target specific templates that are nonhomologous to regions of DNA within or near the 13 nt target region.

Target cells

[0239] Genome editing systems according to this disclosure can be used to manipulate or alter a cell, e.g., to edit or alter a target nucleic acid. The manipulating can occur, in various embodiments, *in vivo* or *ex vivo*.

[0240] A variety of cell types can be manipulated or altered according to the embodiments of this disclosure, and in some cases, such as *in vivo* applications, a plurality of cell types are altered or manipulated, for example by delivering genome editing systems according to this disclosure to a plurality of cell types. In other cases, however, it may be desirable to limit manipulation or alteration to a particular cell type or types. In some embodiments, the cell is an animal or plant cell. In some embodiments, the cell is a mammalian, primate, or human cell. In some embodiments, the cell is a human cell, e.g., a cell form described herein. For instance, it can be desirable in some instances to edit a cell with limited differentiation potential or a terminally differentiated cell, such as a photoreceptor cell in the case of Maeder, in which modification of a genotype is expected to result in a change in cell phenotype. In other cases,

however, it may be desirable to edit a less differentiated, multipotent or pluripotent, stem or progenitor cell. By way of example, the cell may be an embryonic stem cell, induced pluripotent stem cell (iPSC), hematopoietic stem/progenitor cell (HSPC), or other stem or progenitor cell type that differentiates into a cell type of relevance to a given application or indication.

[0241] As a corollary, the cell being altered or manipulated is, variously, a dividing cell or a non-dividing cell, depending on the cell type(s) being targeted and/or the desired editing outcome.

[0242] When cells are manipulated or altered *ex vivo*, the cells can be used (e.g., administered to a subject) immediately, or they can be maintained or stored for later use. Those of skill in the art will appreciate that cells can be maintained in culture or stored (e.g., frozen in liquid nitrogen) using any suitable method known in the art.

Implementation of genome editing systems: delivery, formulations, and routes of administration

[0243] As discussed above, the genome editing systems of this disclosure can be implemented in any suitable manner, meaning that the components of such systems, including without limitation the RNA-guided nuclease, gRNA, and optional donor template nucleic acid, can be delivered, formulated, or administered in any suitable form or combination of forms that results in the transduction, expression or introduction of a genome editing system and/or causes a desired repair outcome in a cell, tissue or subject. **Tables 3** and **4** set forth several, non-limiting examples of genome editing system implementations. Those of skill in the art will appreciate, however, that these listings are not comprehensive, and that other implementations are possible. With reference to **Table 3** in particular, the table lists several exemplary implementations of a genome editing system comprising a single gRNA and an optional donor template. However, genome editing systems according to this disclosure can incorporate multiple gRNAs, multiple RNA-guided nucleases, and other components such as proteins, and a variety of implementations will be evident to the skilled artisan based on the principles illustrated in the table. In the table, [N/A] indicates that the genome editing system does not include the indicated component.

Table 3: Genome editing components

RNA-guided Nuclease	gRNA	Donor Template	Comments
Protein	RNA	[N/A]	An RNA-guided nuclease protein complexed with a gRNA molecule (an RNP complex)
Protein	RNA	DNA	An RNP complex as described above plus a single-stranded or double stranded donor template.
Protein	DNA	[N/A]	An RNA-guided nuclease protein plus gRNA transcribed from DNA.
Protein	DNA	DNA	An RNA-guided nuclease protein plus gRNA-encoding DNA and a separate DNA donor template.
Protein	DNA		An RNA-guided nuclease protein and a single DNA encoding both a gRNA and a donor template.
DNA			A DNA or DNA vector encoding an RNA-guided nuclease, a gRNA and a donor template.
DNA	DNA	[N/A]	Two separate DNAs, or two separate DNA vectors, encoding the RNA-guided nuclease and the gRNA, respectively.
DNA	DNA	DNA	Three separate DNAs, or three separate DNA vectors, encoding the RNA-guided nuclease, the gRNA and the donor template, respectively.
DNA		[N/A]	A DNA or DNA vector encoding an RNA-guided nuclease and a gRNA
DNA		DNA	A first DNA or DNA vector encoding an RNA-guided nuclease and a gRNA, and a second DNA or DNA vector encoding a donor template.
DNA	DNA		A first DNA or DNA vector encoding an RNA-guided nuclease and second DNA or DNA vector encoding a gRNA and a donor template.
DNA			A first DNA or DNA vector encoding an RNA-guided nuclease and a donor template, and a second
DNA			

		DNA or DNA vector encoding a gRNA
	DNA	A DNA or DNA vector encoding an RNA-guided nuclease and a donor template, and a gRNA
	RNA	
RNA	[N/A]	An RNA or RNA vector encoding an RNA-guided nuclease and comprising a gRNA
RNA	DNA	An RNA or RNA vector encoding an RNA-guided nuclease and comprising a gRNA, and a DNA or DNA vector encoding a donor template.

[0244] **Table 4** summarizes various delivery methods for the components of genome editing systems, as described herein. Again, the listing is intended to be exemplary rather than limiting.

Table 4

Delivery Vector/Mode		Delivery into Non-Dividing Cells	Duration of Expression	Genome Integration	Type of Molecule Delivered
Physical (e.g., electroporation, particle gun, Calcium Phosphate transfection, cell compression or squeezing)		YES	Transient	NO	Nucleic Acids and Proteins
<i>Viral</i>	Retrovirus	NO	Stable	YES	RNA
	Lentivirus	YES	Stable	YES/NO with modifications	RNA
	Adenovirus	YES	Transient	NO	DNA
	Adeno-Associated Virus (AAV)	YES	Stable	NO	DNA
	Vaccinia Virus	YES	Very Transient	NO	DNA
	Herpes Simplex Virus	YES	Stable	NO	DNA
<i>Non-Viral</i>	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
<i>Biological Non-Viral Delivery</i>	Attenuated Bacteria	YES	Transient	NO	Nucleic Acids

<i>Vehicles</i>	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

Nucleic acid-based delivery of genome editing systems

[0245] Nucleic acids encoding the various elements of a genome editing system according to the present disclosure can be administered to subjects or delivered into cells by art-known methods or as described herein. For example, RNA-guided nuclease-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.

[0246] Nucleic acids encoding genome editing systems or components thereof can be delivered directly to cells as naked DNA or RNA, for instance by means of transfection or electroporation, or can be conjugated to molecules (e.g., N-acetylgalactosamine) promoting uptake by the target cells (e.g., erythrocytes, HSCs). Nucleic acid vectors, such as the vectors summarized in **Table 4**, can also be used.

[0247] Nucleic acid vectors can comprise one or more sequences encoding genome editing system components, such as an RNA-guided nuclease, a gRNA and/or a donor template. A vector can also comprise a sequence encoding a signal peptide (e.g., for nuclear localization, nucleolar localization, or mitochondrial localization), associated with (e.g., inserted into or fused to) a sequence coding for a protein. As one example, a nucleic acid vectors can include a Cas9 coding sequence that includes one or more nuclear localization sequences (e.g., a nuclear localization sequence from SV40).

[0248] The nucleic acid vector can also include any suitable number of regulatory/control elements, e.g., promoters, enhancers, introns, polyadenylation signals, Kozak consensus

sequences, or internal ribosome entry sites (IRES). These elements are well known in the art, and are described in Cotta-Ramusino.

[0249] Nucleic acid vectors according to this disclosure include recombinant viral vectors.

Exemplary viral vectors are set forth in **Table 4**, and additional suitable viral vectors and their use and production are described in Cotta-Ramusino. Other viral vectors known in the art can also be used. In addition, viral particles can be used to deliver genome editing system components in nucleic acid and/or peptide form. For example, "empty" viral particles can be assembled to contain any suitable cargo. Viral vectors and viral particles can also be engineered to incorporate targeting ligands to alter target tissue specificity.

[0250] In addition to viral vectors, non-viral vectors can be used to deliver nucleic acids encoding genome editing systems according to the present disclosure. One important category of non-viral nucleic acid vectors are nanoparticles, which can be organic or inorganic.

Nanoparticles are well known in the art, and are summarized in Cotta-Ramusino. Any suitable nanoparticle design can be used to deliver genome editing system components or nucleic acids encoding such components. For instance, organic (e.g., lipid and/or polymer) nanoparticles can be suitable for use as delivery vehicles in certain embodiments of this disclosure. Exemplary lipids for use in nanoparticle formulations, and/or gene transfer are shown in **Table 5**, and **Table 6** lists exemplary polymers for use in gene transfer and/or nanoparticle formulations.

Table 5: 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
<i>N</i> -[1-(2,3-Dioleyloxy)propyl] <i>N,N,N</i> -trimethylammonium chloride	DOTMA	Cationic
1,2-Dioleyloxy-3-trimethylammonium-propane	DOTAP	Cationic
Diocadecylamidoglycylspermine	DOGS	Cationic
<i>N</i> -(3-Aminopropyl)- <i>N,N</i> -dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide	GAP-DLRIE	Cationic
Cetyltrimethylammonium bromide	CTAB	Cationic
6-Lauroxyhexyl ornithinate	LHON	Cationic
1-(2,3-Dioleyloxypropyl)-2,4,6-trimethylpyridinium	2Oc	Cationic
2,3-Dioleyloxy- <i>N</i> -[2(sperminecarboxamido-ethyl)- <i>N,N</i> -dimethyl-1-propanaminium trifluoroacetate	DOSPA	Cationic
1,2-Dioleoyl-3-trimethylammonium-propane	DOPA	Cationic
<i>N</i> -(2-Hydroxyethyl)- <i>N,N</i> -dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide	MDRIE	Cationic

Dimyristooxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3 β -[<i>N</i> -(<i>N</i> ', <i>N</i> '-Dimethylaminoethane)-carbamoyl]cholesterol	DC-Chol	Cationic
Bis-guanidium-tren-cholesterol	BGTC	Cationic
1,3-Diodeoxy-2-(6-carboxy-spermyl)-propylamide	DOSPER	Cationic
Dimethyloctadecylammonium bromide	DDAB	Cationic
Diocadecylamidoglycylspermidin	DSL	Cationic
rac-[(2,3-Dioctadecyloxypropyl)(2-hydroxyethyl)]-dimethylammonium chloride	CLIP-1	Cationic
rac-[2(2,3-Dihexadecyloxypropyl-oxy-methyloxy)ethyl]trimethylammonium bromide	CLIP-6	Cationic
Ethyl dimyristoyl phosphatidylcholine	EDMPC	Cationic
1,2-Distearoyl- <i>N,N</i> -dimethyl-3-aminopropane	DSDMA	Cationic
1,2-Dimyristoyl-trimethylammonium propane	DMTAP	Cationic
<i>O,O</i> '-Dimyristyl- <i>N</i> -lysyl aspartate	DMKE	Cationic
1,2-Distearoyl-sn-glycero-3-ethylphosphocholine	DSEPC	Cationic
<i>N</i> -Palmitoyl D-erythro-sphingosyl carbamoyl-spermine	CCS	Cationic
<i>N-t</i> -Butyl- <i>N</i> 0-tetradecyl-3-tetradecylaminopropionamidine	diC14-amidine	Cationic
Octadecenolyoxy[ethyl-2-heptadecenyl-3 hydroxyethyl]imidazolium chloride	DOTIM	Cationic
<i>N</i> 1-Cholesteryl oxycarbonyl-3,7-diazanonane-1,9-diamine	CDAN	Cationic
2-(3-[Bis(3-amino-propyl)-amino]propylamino)- <i>N</i> -ditetradecylcarbamoylme-ethyl-acetamide	RPR209120	Cationic
1,2-dilinoleyloxy-3- dimethylaminopropane	DLinDMA	Cationic
2,2-dilinoleyl-4-dimethylaminoethyl-[1,3]- dioxolane	DLin-KC2-DMA	Cationic
dilinoleyl- methyl-4-dimethylaminobutyrate	DLin-MC3-DMA	Cationic

Table 6: Polymers used for gene transfer

Polymer	Abbreviation
Poly(ethylene)glycol	PEG
Polyethylenimine	PEI
Dithiobis(succinimidylpropionate)	DSP
Dimethyl-3,3'-dithiobispropionimide	DTBP
Poly(ethylene imine) biscarbamate	PEIC
Poly(L-lysine)	PLL
Histidine modified PLL	
Poly(<i>N</i> -vinylpyrrolidone)	PVP
Poly(propylenimine)	PPI
Poly(amidoamine)	PAMAM
Poly(amido ethylenimine)	SS-PAEI
Triethylenetetramine	TETA
Poly(β -aminoester)	
Poly(4-hydroxy-L-proline ester)	PHP
Poly(allylamine)	

Poly(α -[4-aminobutyl]-L-glycolic acid)	PAGA
Poly(D,L-lactic-co-glycolic acid)	PLGA
Poly(<i>N</i> -ethyl-4-vinylpyridinium bromide)	
Poly(phosphazene)s	PPZ
Poly(phosphoester)s	PPE
Poly(phosphoramidate)s	PPA
Poly(<i>N</i> -2-hydroxypropylmethacrylamide)	pHPMA
Poly (2-(dimethylamino)ethyl methacrylate)	pDMAEMA
Poly(2-aminoethyl propylene phosphate)	PPE-EA
Chitosan	
Galactosylated chitosan	
<i>N</i> -Dodacylated chitosan	
Histone	
Collagen	
Dextran-spermine	D-SPM

[0251] Non-viral vectors optionally include targeting modifications to improve uptake and/or selectively target certain cell types. These targeting modifications can include e.g., cell specific antigens, monoclonal antibodies, single chain antibodies, aptamers, polymers, sugars (e.g., N-acetylgalactosamine (GalNAc)), and cell penetrating peptides. Such vectors also optionally use fusogenic and endosome-destabilizing peptides/polymers, undergo acid-triggered conformational changes (e.g., to accelerate endosomal escape of the cargo), and/or incorporate a stimuli-cleavable polymer, 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.

[0252] In certain embodiments, one or more nucleic acid molecules (e.g., DNA molecules) other than the components of a genome editing system, e.g., the RNA-guided nuclease component and/or the gRNA component described herein, are delivered. In certain embodiments, the nucleic acid molecule is delivered at the same time as one or more of the components of the Genome editing system. In certain embodiments, 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 Genome editing system are delivered. In certain embodiments, the nucleic acid molecule is delivered by a different means than one or more of the components of the genome editing system, e.g., the RNA-guided nuclease component and/or the gRNA 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 RNA-guided nuclease molecule component and/or the gRNA component can be delivered by electroporation, e.g., such that the toxicity caused by nucleic acids (e.g., DNAs) can be reduced. In certain embodiments, the nucleic acid molecule encodes a therapeutic protein, e.g., a protein described herein. In certain embodiments, the nucleic acid molecule encodes an RNA molecule, e.g., an RNA molecule described herein.

Delivery of RNPs and/or RNA encoding genome editing system components

[0253] RNPs (complexes of gRNAs and RNA-guided nucleases) and/or RNAs encoding RNA-guided nucleases and/or gRNAs, can be delivered into cells or administered to subjects by art-known methods, some of which are described in Cotta-Ramusino. *In vitro*, RNA-guided nuclease-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, GalNAc- or other conjugate-mediated delivery, and combinations thereof, can also be used for delivery *in vitro* and *in vivo*. A protective, interactive, non-condensing (PINC) system may be used for delivery.

[0254] *In vitro* delivery via electroporation comprises mixing the cells with the RNA encoding RNA-guided nucleases and/or gRNAs, 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. Systems and protocols for electroporation are known in the art, and any suitable electroporation tool and/or protocol can be used in connection with the various embodiments of this disclosure.

Route of administration

[0255] Genome editing systems, or cells altered or manipulated using such systems, can be administered to subjects by any suitable mode or route, whether local or systemic. Systemic modes of administration include oral and parenteral routes. Parenteral routes include, by way of example, intravenous, intramarrow, intrarterial, intramuscular, intradermal, subcutaneous, intranasal, and intraperitoneal routes. Components administered systemically can be modified or formulated to target, e.g., HSCs, hematopoietic stem/progenitor cells, or erythroid progenitors or precursor cells.

[0256] 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 portal vein. In certain embodiments, significantly smaller amounts of the components (compared with

systemic approaches) can 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 component are administered systemically.

[0257] Administration can be provided as a periodic bolus (for example, intravenously) or as continuous infusion from an internal reservoir or from an external reservoir (for example, from an intravenous bag or implantable pump). Components can be administered locally, for example, by continuous release from a sustained release drug delivery device.

[0258] In addition, components can 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 can be useful, however, the choice of the appropriate system will depend upon rate of release required by a particular application. Both non-degradable and degradable release systems can be used. Suitable release systems include polymers and polymeric matrices, non-polymeric matrices, or inorganic and organic excipients and diluents such as, but not limited to, calcium carbonate and sugar (for example, trehalose). Release systems may be natural or synthetic. However, synthetic release systems are preferred because generally they are more reliable, more reproducible and produce more defined release profiles. The release system material can be selected so that components having different molecular weights are released by diffusion through or degradation of the material.

[0259] Representative synthetic, biodegradable polymers include, for example: polyamides such as poly(amino acids) and poly(peptides); polyesters such as poly(lactic acid), poly(glycolic acid), poly(lactic-co-glycolic acid), and poly(caprolactone); poly(anhydrides); polyorthoesters; polycarbonates; and chemical derivatives thereof (substitutions, additions of chemical groups, for example, alkyl, alkylene, hydroxylations, oxidations, and other modifications routinely made by those skilled in the art), copolymers and mixtures thereof. Representative synthetic, non-degradable polymers include, for example: polyethers such as poly(ethylene oxide), poly(ethylene glycol), and poly(tetramethylene oxide); vinyl polymers-polyacrylates and polymethacrylates such as methyl, ethyl, other alkyl, hydroxyethyl methacrylate, acrylic and

methacrylic acids, and others such as poly(vinyl alcohol), poly(vinyl pyrrolidone), and poly(vinyl acetate); poly(urethanes); cellulose and its derivatives such as alkyl, hydroxyalkyl, ethers, esters, nitrocellulose, and various cellulose acetates; polysiloxanes; and any chemical derivatives thereof (substitutions, additions of chemical groups, for example, alkyl, alkylene, hydroxylations, oxidations, and other modifications routinely made by those skilled in the art), copolymers and mixtures thereof.

[0260] Poly(lactide-co-glycolide) microsphere can also be used. Typically the microspheres are composed of a polymer of lactic acid and glycolic acid, which are structured to form hollow spheres. The spheres can be approximately 15-30 microns in diameter and can be loaded with components described herein. In some embodiments, genome editing systems, system components and/or nucleic acids encoding system components, are delivered with a block copolymer such as a poloxamer or a poloxamine.

Multi-modal or differential delivery of components

[0261] Skilled artisans will appreciate, in view of the instant disclosure, that different components of genome editing systems disclosed herein can be delivered together or separately and simultaneously or nonsimultaneously. Separate and/or asynchronous delivery of genome editing system components can be particularly desirable to provide temporal or spatial control over the function of genome editing systems and to limit certain effects caused by their activity.

[0262] Different or differential modes as used herein refer to modes of delivery that confer different pharmacodynamic or pharmacokinetic properties on the subject component molecule, e.g., a RNA-guided nuclease molecule, gRNA, template nucleic acid, or payload. For example, the modes of delivery can result in different tissue distribution, different half-life, or different temporal distribution, e.g., in a selected compartment, tissue, or organ.

[0263] 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.

[0264] By way of example, the components of a genome editing system, e.g., a RNA-guided nuclease and a gRNA, can be delivered by modes that differ in terms of resulting half-life or persistent of the delivered component the body, or in a particular compartment, tissue or organ. In certain embodiments, a gRNA can be delivered by such modes. The RNA-guided nuclease

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.

[0265] More generally, in certain embodiments, 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.

[0266] 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.

[0267] 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.

[0268] 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.

[0269] In certain embodiments, 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.

[0270] In certain embodiments, the second mode of delivery comprises a relatively transient element, e.g., an RNA or protein.

[0271] In certain embodiments, the first component comprises gRNA, 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 RNA-guided nuclease molecule, is delivered in a transient

manner, for example as mRNA or as protein, ensuring that the full RNA-guided nuclease molecule/gRNA complex is only present and active for a short period of time.

[0272] 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.

[0273] 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.

[0274] 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 certain embodiments, a first component, e.g., a gRNA is delivered by a first delivery mode that results in a first spatial, e.g., tissue, distribution. A second component, e.g., a RNA-guided nuclease molecule is delivered by a second delivery mode that results in a second spatial, e.g., tissue, distribution. In certain embodiments, 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 certain embodiments, 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.

[0275] When the RNA-guided nuclease 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 and the RNA-guided nuclease 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.

EXAMPLES

[0276] The principles and embodiments described above are further illustrated by the non-limiting examples that follow:

Example 1: Screening of *S. pyogenes* gRNAs delivered to K562 cells as ribonucleoprotein complexes for use in causing 13 nt deletions in *HBG1* and *HBG2* regulatory regions

[0277] gRNAs targeting a 26 nt fragment spanning and including the 13 nucleotides at the 13 nt target region of *HBG1* and *HBG2* were designed by standard methods. After gRNAs were designed *in silico* and tiered, a subset of the gRNAs were selected and screened for activity and specificity in human K562 cells. The gRNAs selected for screening are set forth in **Table 7**. Briefly, gRNAs were *in vitro* transcribed and then complexed with *S. pyogenes* wildtype (Wt) Cas9 protein to form ribonucleoprotein complexes (RNPs). The gRNAs complexed to *S. pyogenes* Cas9 protein were modified sgRNAs ((e.g., 5' ARCA capped and 3' polyA (20A) tail; **Table 7**) and target the *HBG1* and *HBG2* regulatory regions. To allow for direct comparison of the activity of these RNPs in K562 cells and human CD34⁺ cells, RNPs were first delivered to K562 cells by electroporation (Amaza Nucleofector).

[0278] Three days after RNP electroporation, gDNA was extracted from K562 cells and then the *HBG1* and *HBG2* loci were PCR amplified from the gDNA. Gene editing was evaluated in the PCR products by T7E1 endonuclease assay analysis. Eight out of nine RNPs supported a high percentage of NHEJ. Sp37 RNP, the only gRNA shown to be active in human CD34⁺ cells (<10% editing in CD34⁺ cells) was highly active in K562 cells, with >60% indels detected at both *HBG1* and *HBG2* and eight cut in both the *HBG1* and *HBG2* targeted regions in the promoter sequences (**Fig. 3A**).

Table 7: Selected gRNAs for screening in K562 cells or CD34⁺ cells

gRNA ID	Targeting domain sequence (RNA)	Targeting domain sequence (DNA)	Targeting domain sequence plus PAM (NGG) (RNA)	Targeting domain sequence plus PAM (NGG) (DNA)	Sense
Sp9	GGCUAUUGGU CAAGGCA (SEQ ID NO:277)	GGCTATTGGTC AAGGCA (SEQ ID NO:910)	GGCUAUUGGUCA AGGCAAGG (SEQ ID NO:920)	GGCTATTGGTC AAGGCAAGG (SEQ ID NO:930)	Antisense
Sp36	CAAGGCUAUU GGUCAAGGCA (SEQ ID NO:338)	CAAGGCTATTG GTCAAGGCA (SEQ ID NO:911)	CAAGGCUAUUGG UCAAGGCAAGG (SEQ ID NO:921)	CAAGGCTATTG GTCAAGGCAAG G (SEQ ID NO:931)	Antisense
Sp40	UGCCUUGUCA AGGCUAU	TGCCTTGTC GGCTAT	UGCCUUGUCAAG GCUAUUGG	TGCCTTGTC GGCTATTGG	Antisense

	(SEQ ID NO:327)	(SEQ ID NO:912)	(SEQ ID NO:922)	(SEQ ID NO:932)	
Sp42	GUUUGCCUUG UCAAGGCUAU (SEQ ID NO:299)	GTTTGCCTTGTC AAGGCTAT (SEQ ID NO:913)	GUUUGCCUUGUC AAGGCUAUUGG (SEQ ID NO:923)	GTTTGCCTTGTC AAGGCTATTGG (SEQ ID NO:933)	Antisense
Sp38	GACCAAUAGC CUUGACA (SEQ ID NO:276)	GACCAATAGCC TTGACA (SEQ ID NO:914)	GACCAAUAGCCU UGACAAGG (SEQ ID NO:924)	GACCAATAGCC TTGACAAGG (SEQ ID NO:934)	Sense
Sp37	CUUGACCAAU AGCCUUGACA (SEQ ID NO:333)	CTTGACCAATA GCCTTGACA (SEQ ID NO:915)	CUUGACCAAUAG CCUUGACAAGG (SEQ ID NO:925)	CTTGACCAATA GCCTTGACAAG G (SEQ ID NO:935)	Sense
Sp43	GUCAAGGCUA UUGGUCA (SEQ ID NO:278)	GTCAAGGCTAT TGGTCA (SEQ ID NO:916)	GUCAAGGCUAUU GGUCAAGG (SEQ ID NO:926)	GTCAAGGCTAT TGGTCAAGG (SEQ ID NO:936)	Antisense
Sp35	CUUGUCAAGG CUAUUGGUCA (SEQ ID NO:339)	CTTGTCAAGGC TATTGGTCA (SEQ ID NO:917)	CUUGUCAAGGCU AUUGGUCAAGG (SEQ ID NO:927)	CTTGTCAAGGC TATTGGTCAAG G (SEQ ID NO:937)	Antisense
Sp41	UCAAGUUUGC CUUGUCA (SEQ ID NO:310)	TCAAGTTTGCCT TGTC (SEQ ID NO:918)	UCAAGUUUGCCU UGUCAAGG (SEQ ID NO:928)	TCAAGTTTGCCT TGTCAGG (SEQ ID NO:938)	Antisense
Sp34	UGGUCAAGUU UGCCUUGUCA (SEQ ID NO:340)	TGGTCAAGTTT GCCTTGTC (SEQ ID NO:919)	UGGUCAAGUUUG CCUUGUCAAGG (SEQ ID NO:929)	TGGTCAAGTTT GCCTTGTCAG G (SEQ ID NO:939)	Antisense
Sp85	AGUAUCCAGU GAGGCCA (SEQ ID NO:940)	AGTATCCAGTG AGGCCA (SEQ ID NO:943)	AGUAUCCAGUGA GGCCAGGG (SEQ ID NO:946)	AGTATCCAGTG AGGCCAGGG (SEQ ID NO:949)	Antisense
SpA	GGCAAGGCUG GCCAACCCA (SEQ ID NO:941)	GGCAAGGCTGG CCAACCCAT (SEQ ID NO:944)	GGCAAGGCUGGC CAACCCAUGGG (SEQ ID NO:947)	GGCAAGGCTGG CCAACCCATGG G (SEQ ID NO:950)	Sense
SpB	UAUUUGCAUU GAGAUAGUGU (SEQ ID NO:942)	TATTTGCATTGA GATAGTGT (SEQ ID NO:945)	UAUUUGCAUUGA GAUAGUGUGGG (SEQ ID NO:948)	TATTTGCATTGA GATAGTGTGGG (SEQ ID NO:951)	Sense

[0279] The *HBG1* and *HBG2* PCR products for the K562 cells that were targeted with the eight active sgRNAs were then analyzed by DNA sequencing analysis and scored for insertions and deletions detected. The deletions were subdivided into precise 13 nt deletions at the target site, 13 nt target site inclusive and proximal small deletions (18-26 nt), 12 nt deletions (i.e., partial deletion) of the 13 nt target site, >26 nt deletions that span a portion of the HPFH target site, and other deletions, e.g., deletions proximal to but outside the HPFH target site. Seven of the eight

sgRNAs targeted deletion of the 13 nt (HPFH mutation induction) (**Fig. 3B**) for *HBG1*. At least five of the eight sgRNAs also supported targeted deletion of the 13 nt in *HBG2* promoter region (**Fig. 3C**). Note that DNA sequence results for *HBG2* in cells treated with HBG Sp34 sgRNA were not available. These data indicate that Cas9 and sgRNA support precise induction of the 13 nt deletions. **Figs. 3B-3C** depict examples of the types of deletions observed in target sequences in *HBG1*.

Example 2: Cas9 RNP containing gRNA targeting the 13 nt deletion mutation supports gene editing in human hematopoietic stem/progenitor cells

[0280] Of the RNPs containing different gRNAs tested in human cord blood (CB) CD34⁺ cells, only Sp37 resulted in detectable editing at the target site in the *HBG1* and *HBG2* promoters as determined by T7E1 analysis of indels in *HBG1* and *HBG2* specific PCR products amplified from gDNA extracted from electroporated CB CD34⁺ cells from a three cord blood donors (**Fig. 4A**). The average level of editing detected in cells electroporated with Cas9 protein complexed to Sp37 was 5 ± 2 % indels at *HBG1* and 3 ± 1 % indels detected at *HBG2* (3 separate experiments, and CB donors).

[0281] Next, three *S. pyogenes* gRNAs whose target sites are within the HBG promoter (Sp35, Sp36, Sp37) were complexed to wild-type *S. pyogenes* Cas9 protein to form ribonucleoprotein complexes. These *HBG* targeted RNPS were electroporated into CB CD34⁺ cells (n=3 donors) and adult mobilized peripheral blood (mPB) CD34⁺ cell donors (n=3 donors). Then the level of insertions/deletions at the target site was analyzed by T7E1 endonuclease analysis of the *HBG2* PCR products amplified from genomic DNA extracted from the samples approximately 3 days after Cas9 RNP delivery. Each of these RNPs supported only low level gene editing in both the CB and adult CD34⁺ cells across 3 donors and 3 separate experiments (**Fig. 4B**).

[0282] To increase gene editing and the occurrence of the 13 nt deletion at the target site, single strand deoxynucleotide donor repair templates (ssODNs) that encoded 87 nt and 89 nt of homology on each side of the targeted deletion site was generated. The ssODNs, either unmodified at the ends (*i.e.*, ssODN1, SEQ ID NO:906, **Table 8**) or modified to contain phosphothioates (PhTx) at the 5' and 3' ends (*i.e.*, PhTx ssODN1, SEQ ID NO:909, **Table 8**). The ssODN was designed to 'encode' the 13 nt deletion with sequence homology arms engineered flanking this absent sequence to create a perfect deletion.

Table 8: Single strand deoxynucleotide donor repair templates (ssODN)

ssODN ID	SEQ ID NO	Sequence
ssODN1 5' homology arm	904	GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCG GCCCCTGGCCTCACTGGATACTCTAAGACTATTGGTCA AGTTTGCCTT
ssODN1 3' homology arm	905	<u>GTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCC AGGGACCGTTTCAGACAGATATTTGCATTGAGATAGTGTG GGGAAGGGG</u>
ssODN1	906	GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCG GCCCCTGGCCTCACTGGATACTCTAAGACTATTGGTCA AGTTTGCCTTGTCAAGGCAAGGCTGGCCAACCCATGGGTG GAGTTTAGCCAGGGACCGTTTCAGACAGATATTTGCATTGA GATAGTGTGGGGAAGGGG
PhTx ssODN1 5' homology arm	907	*GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCC GCCCCCTGGCCTCACTGGATACTCTAAGACTATTGGTC AAGTTTGCCTT
PhTx ssODN1 3' homology arm	908	<u>GTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCC AGGGACCGTTTCAGACAGATATTTGCATTGAGATAGTGTG GGGAAGGGG*</u>
PhTx ssODN1	909	*GGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCC GCCCCCTGGCCTCACTGGATACTCTAAGACTATTGGTC AAGTTTGCCTTGTCAAGGCAAGGCTGGCCAACCCATGGGT GGAGTTTAGCCAGGGACCGTTTCAGACAGATATTTGCATTG AGATAGTGTGGGGAAGGGG*
The homology arms flanking the deletion are indicated by bold [5' homology arm] and underline [3' homology arm]). Note the absence of the 13 bp sequence in ssODN1 and PhTx ssODN1. *Represents modification by phosphothioate.		

[0283] ssODN1 and PhTx ssODN1 were co-delivered with RNP targeting *HBG* containing the Sp37 gRNA (*HBG* Sp37 RNP) or *HBG* Sp35 (*HBG* Sp35 RNP) to CB CD34⁺ cells. Co-delivery of the ssODN donor encoding the 13 nt deletion with *HBG* Sp35 RNP or *HBG* Sp37 RNP led to a 6-fold and 5-fold increase in gene editing of the target site, respectively, as determined by T7E1 analysis of the *HBG2* PCR product (**Fig. 4C**). DNA sequencing analysis (Sanger sequencing) of the *HBG2* PCR product indicated that 20% gene editing in cells that were treated with *HBG* Sp37 RNP and the PhTx modified ssODN1, with 15% deletions and 5% insertions (**Fig. 4C**, lower left panel). Further analysis of the specific type and size of deletions at the target site revealed that 75% of the total deletions detected contained the 13 nt deletion (which

included deletion at c. -110 of the CAAT box in the proximal promoter), the absence of which is associated with elevation of HbF expression (**Fig. 4C**, lower right panel). The remaining ¼ of deletions were partial deletions that did not span the full 13 nt deletion. These data indicate that co-delivery of a homologous ssODN that is engineered to have a deletion supported precise gene editing (deletion) at *HBG* in human CD34⁺ cells.

Example 3: Cas9 RNP targeting the 13 nt deletion mutation supports gene editing in human adult mobilized peripheral blood hematopoietic stem/progenitor cells with increased *HBG* expression in erythroblast progeny.

[0284] To determine whether editing *HBG* with Cas9 RNP complexed to Sp37 gRNA or Sp35 gRNA (i.e., the gRNAs that target the 13 nt deletion that is associated with HPFH) in the promoter of *HBG* supports an increase in *HBG* expression in erythroid progeny of edited CD34⁺ cells, human adult CD34⁺ cells from mobilized peripheral blood (mPB) were electroporated with the RNPs. Briefly, mPB CD34⁺ cells were prestimulated for 2 days with human cytokines and PGE2 in StemSpan SFEM and then electroporated with Cas9 protein precomplexed to Sp35 and Sp37, respectively. T7E1 analysis of *HBG* PCR product indicated ~3% indels detected for mPB CD34⁺ cells treated with RNP complexed to Sp37 while no editing was detected for cells that were treated with RNP complexed to Sp35 (**Fig. 5A**).

[0285] In order to increase gene editing at the target site and to increase the occurrence of the 13 nt deletion at the target site, PhTx ssODN1 (SEQ ID NO:909) was co-delivered with the precomplexed RNP targeting *HBG* containing the Sp37 gRNA. Co-delivery of the ssODN donor encoding the 13 nt deletion led to a nearly 2-fold increase in gene editing of the target site (**Fig. 5A**). To determine whether editing *HBG* increases production of fetal hemoglobin in erythroid progeny of edited adult CD34⁺ cells, the cells were differentiated into erythroblasts by culture for up to 18 days in the presence of human cytokines (erythropoietin, SCF, IL3), human plasma (Octoplas), and other supplements (hydrocortisone, heparin, transferrin). Over the time course of differentiation, mRNA was collected to evaluate *HBG* gene expression in the erythroid progeny of RNP treated mPB CD34⁺ cells and donor matched negative (untreated) controls. By day 7 of differentiation, erythroblast progeny of human CD34⁺ cells that were treated with *HBG* Sp37 RNP and 13 nt deletion encoding ssODN (~5% indels detected in gDNA from the bulk cell population by T7E1 analysis) exhibited a 2-fold increase in *HBG* mRNA production (**Fig. 5B**). Importantly, CD34⁺ cells that were electroporated with *HBG* RNP maintained their *ex vivo*

hematopoietic activity (i.e., no difference in the quantity or diversity of erythroid and myeloid colonies compared to untreated donor matched CD34⁺ cell negative control), as determined in hematopoietic colony forming cell (CFC) assays (**Fig. 6A**). Furthermore, the erythroblasts differentiated from RNP treated CD34⁺ cells maintained the kinetics of differentiation observed for donor matched untreated control cells as determined by flow analysis for acquisition of erythroid phenotype (%Glycophorin A⁺ cells) (**Fig. 6B**). These data indicate that targeted disruption of *HBG1/HBG2* proximal promoter region supported an increase in *HBG* expression in erythroid progeny of RNP treated adult hematopoietic stem/progenitor cells without altering differentiation potential.

Example 4: Cas9 RNP targeting the HPFH mutation supports gene editing in human adult mobilized peripheral blood hematopoietic stem/progenitor cells with increased *HBG* expression in erythroblast progeny

[0286] To determine whether co-delivery of paired nickase RNPs targeting *HBG* would increase targeted disruption of the proximal *HBG* promoter, mPB CD34⁺ cells were cultured for 2 days with human cytokines and PGE2 in StemSpan SFEM and then electroporated with *S. pyogenes* D10A Cas9 protein precomplexed to two gRNAs that target sites flanking the site of the 13 nt deletion. The targeting domain sequences for gRNAs used in nickase pairs in this example (including, without limitation, SpA, Sp85 and SpB) are presented in **Table 7**. D10A nickase pairs were selected such that the PAMs for the targets were oriented outward and the distance between the cut sites were <100 nt. gRNAs were complexed with D10A Cas9 protein to form RNP complexes and then human CD34⁺ cells and paired nickase were subject to electroporation. To determine whether co-delivery of an ssODN that encoded the 13 nt deletion would increase editing and introduction of the mutation into the cells, in some experiments, ssODN1 was added to the cell RNP mixture prior to electroporation. Approximately 3 days after electroporation, gDNA was extracted from the RNP treated cells and analyzed by T7E1 endonuclease assay and/or Sanger DNA sequencing of *HBG2* PCR products amplified from the extracted gDNA. Of the three D10A nickase pairs tested, indels detected by T7E1 endonuclease analysis were increased for one nickase pair (gRNAs SpA+Sp85) samples for which ssODN1 was included (**Fig. 7A**). DNA sequencing analysis was performed on limited samples shown in **Fig. 7A**. DNA sequencing analysis showed up to ~27% indels at the target site, with insertions as the dominant indel detected, followed by deletions of the targeted region (area between the cut sites

of the paired nickases), and the 13 nt deletion mutation was also detected at a frequency of 2-3% when ssODN1 encoding the deletion was co-delivered (**Fig. 7B**). Silent, non-pathogenic SNPs were included in the ssODN1 donor template, and were detected in the sequences that contained the 13 nt deletion, indicating that creation of the HFPH mutation occurred through an HDR event.

Example 5: D10A paired RNPs electroporated into adult CD34⁺ cells supports induction of HbF protein in erythroid progeny.

[0287] To further optimize editing conditions in mPB CD34⁺ cells at the target site and to evaluate editing in additional human cell donors, human mPB CD34⁺ cells were electroporated with D10A Cas9 and WT Cas9 paired RNPs targeting *HBG*. The most efficient guide pair for both D10A Cas9 and WT Cas9 RNPs was Sp37+SpA, which supported >30% indels as determined by T7E1 endonuclease analysis of *HBG2* PCR products (**Fig. 8A**). Given that editing at both *HBG1* and *HBG2* could result in large deletions of *HBG2* and the intergenic region between *HBG2* and *HBG1*, indels were further characterized in order to capture local indels by T7E1 endonuclease assay and sequencing and large deletion by ddPCR analysis. Large deletions were detected in all samples at variable frequencies for both D10A Cas9 and WT Cas9 RNP nickase pairs (**Fig. 8B**). Illumina sequencing analysis of indels correlated with indels determined by T7E1 analysis (**Fig. 8C-8D**).

[0288] To determine whether CD34⁺ cells edited with dual nickases at the *HBG* promoter gave rise to erythroid progeny with elevated HbF expression, donor matched RNP treated and untreated controls were induced toward erythroid differentiation and then evaluated for maintenance of indels during differentiation and for expression of HbF mRNA and protein. The level of editing (as determined by T7E1 endonuclease assay) was evaluated over the first 2 weeks of erythroid differentiation in the progeny of RNP treated cells prior to enucleation. Indels were detected in the erythroid progeny at every time point assayed suggesting that the editing that occurred in the CD34⁺ cells was maintained during erythroid differentiation and that edited CD34⁺ cells maintain erythroid differentiation potential.

The levels of *HBG* mRNA (day 10 of differentiation) and HbF protein (day 20-23 of differentiation) were quantified by ddPCR and HPLC analysis (according to the HPLC method described in Chang 2017 at pp. 143-44, incorporated by reference herein), respectively (**Fig. 9**). A ~2-fold increase (+40% in in *HBG* transcripts vs. unedited donor matched control) was

observed for *HBG:HbA* ratio (data not shown) and the ratio of *HbF/HbF+HbA* (i.e., *HBG* mRNA / *HGB+HBB* mRNA) increased to 30% above the level detected in donor matched untreated control samples.

[0289] For the D10A Cas9 nickase pairs, upregulation of HbF mRNA and protein was detected in erythroid progeny (**Fig. 9**). With respect to HbF protein analysis, two pairs supported 20% HbF induction for two D10A nickase pairs. No HbF upregulation was detected in erythroid progeny of WT Cas9 RNP treated CD34⁺ cells (data not shown).

Example 6: Increasing the dose of RNP increases total editing efficiency in human adult CD34⁺ cells at the *HBG* locus.

[0290] The concentration of D10A Cas9 RNP for the nickase pair SpA+Sp85 was increased (2.5 μ M standard concentration and 3.7 μ M) and delivered to mPB CD34⁺ cells by electroporation. The increased RNP concentration supported an increase in indels at the *HBG* target site to >30% (**Fig. 10A**) as determined by T7E1 endonuclease analysis of the *HBG* PCR product amplified for gDNA extracted 3 days after electroporation of CD34⁺ cells. Sequencing analysis indicated that increasing the RNP concentration increased insertions (**Fig. 10B**). Erythroid progeny of RNP treated CD34⁺ cells also had an increase in HbF protein production (**Fig. 10C**). Importantly, the hematopoietic colony forming potential was maintained after editing (**Fig. 10D**). These cells were then transplanted into immunodeficient mice and their engraftment 1 month (**Fig. 10E**) and 2 months (**Fig. 10F**) after transplantation was evaluated by sampling the peripheral blood and measuring the percentage of human CD45⁺ cells. Early engraftment data showed no difference in engraftment between recipient cohorts of donor matched untreated controls (0 μ M RNP) and mice transplanted with RNP treated cells. Furthermore, there was no difference in human blood lineage distribution (myeloid, B cell, T cell) within the human CD45⁺ fraction among cohorts at indicated time points (**Fig. 10G-H**).

[0291] Two additional D10A nickase pairs were also tested in RNP dose response studies in adult mPB CD34⁺ cells (Sp37+SpA, Sp37+SpB). Here, mPB CD34⁺ cells were electroporated with D10A paired nickases delivered at 0, 2.5, and 3.75 μ M of total RNP. RNP treated cells were differentiated into erythroid progeny and the HbF protein levels (%HbF/HbF+HbA) were analyzed by HPLC analysis. The indel frequency detected in CD34⁺ cells was plotted with the HbF levels detected in erythroid progeny in order to correlate editing and HbF induction (**Fig. 11A**). RNP treated and untreated control mPB CD34⁺ cells were also differentiated into colonies

to evaluate *ex vivo* hematopoietic activity. Colony forming cell (CFC) activity was maintained for the progeny of RNP treated and donor matched untreated control CD34⁺ cells (**Fig. 11B**). There was no difference in the percentage of human CD45⁺ cells in the mouse peripheral blood 1 month after transplantation and no difference in blood lineage distribution (**Fig. 11C-D**) for cells exposed to different D10A RNP pairs at different doses compared to untreated donor matched control CD34⁺ cells.

Example 7: Co-delivery of RNP targeting the erythroid specific enhancer of *BCL11A* and a non-specific (N) single strand deoxynucleotide sequence or paired RNPs increases gene editing in human CD34⁺ cells and supports induction of fetal hemoglobin expression in erythroid progeny

[0292] Fetal hemoglobin expression can be induced through targeted disruption of the erythroid cell specific expression of a transcriptional repressor, *BCL11A* (Canvers 2015). One potential strategy to increase HbF expression through a gene editing strategy is to multiplex gene editing for introduction of 13 nt deletion associated in the *HBG* proximal promoter and also for targeted disruption of the GATA1 binding motif in the erythroid specific enhancer of *BCL11A* that is in the +58 DHS region of intron 2 of the *BCL11A* gene (**Fig. 12**). In order to accomplish this multiplex strategy to increase HbF expression through multiplex gene editing, the effect of disruption of *BCL11A* erythroid enhancer (*BCL11Ae*) must first be determined as a single editing event.

[0293] In this experiment, CB CD34⁺ cells were electroporated with *S. pyogenes* WT Cas9 complexed to *in vitro* transcribed sgRNA targeting the GATA1 motif in the +58 DHS region of intron 2 of *BCL11A* gene (gRNA SpK, **Table 9**) (**Fig. 13A**). To determine whether co-delivery of a non-target specific ssODN would increase editing of the target sequence, *BCL11Ae* RNP was co-delivered with ssODN (which is nonhomologous to the *BCL11Ae* target sequence, also called a non-specific ssODN) in CB CD34⁺ cells. T7E1 analysis of *BCL11A* erythroid enhancer PCR product from gDNA extracted from CB CD34⁺ cells treated with *BCL11Ae* RNP indicated that ~5% indels was achieved (**Fig. 13A**). Co-delivery of *BCL11Ae* RNP with a non-target specific ssODN increase in indels by 5- fold to 20% as detected by T7E1 endonuclease analysis. Illumina sequencing analysis indicated that >90% of edits had disruption of the GATA1 motif in the +DHS 58 region enhancer in intron 2 of the *BCL11A* gene (data not shown). To increase editing, human CB CD34⁺ cells were electroporated with WT Cas9 RNP (single gRNAs complexed to WT Cas9) or with WT Cas9 paired RNPs (paired gRNAs complexed to WT Cas9),

so that the cut sites in each pair flank the target site for excision of the GATA1 motif (gRNAs SpC, SpK, SpM, SpN) (**Table 9**). Two of the single gRNAs and two pairs had >50% indels as determined by T7E1 endonuclease analysis (**Fig. 13B**).

Table 9: Select gRNA sequences targeting *BCL11A* erythroid enhancer for screening in CD34⁺ cells

gRNA ID	Targeting domain sequence (RNA)	Targeting domain sequence (DNA)	Targeting domain sequence plus PAM (NGG) (RNA)	Targeting domain sequence plus PAM (NGG) (DNA)	Sense
SpK	CUAACAGUUG CUUUUAUCAC (SEQ ID NO:952)	CTAACAGTTG CTTTTATCAC (SEQ ID NO:956)	CUAACAGUUGC UUUUUAUCACAG G (SEQ ID NO:960)	CTAACAGTTGCT TTTATCACAGG (SEQ ID NO:964)	Antisense
SpM	GGGCGUGGGU GGGUAGAAG (SEQ ID NO:953)	GGGCGTGGGT GGGTAGAAG (SEQ ID NO:957)	GGGCGUGGGUG GGGUAGAAGAG G (SEQ ID NO:961)	GGGCGTGGGTGG GGTAGAAGAGG (SEQ ID NO:965)	Antisense
SpN	CUCUUAGACA UAACACACCA (SEQ ID NO:954)	CTCTTAGACA TAACACACCA (SEQ ID NO:958)	CUCUUAGACAU AACACACCAGG G (SEQ ID NO:962)	CTCTTAGACATA ACACACCAGGG (SEQ ID NO:966)	Antisense
SpC	AUCAGAGGCC AAACCCUUC (SEQ ID NO:955)	ATCAGAGGCC AAACCCCTTCC (SEQ ID NO:959)	AUCAGAGGCCA AAACCCUUCUGG (SEQ ID NO:963)	ATCAGAGGCCAA ACCCTTCCTGG (SEQ ID NO:967)	Sense

[0294] Next, human adult bone marrow CD34⁺ cells were electroporated with the *BCL11Ae* RNP. DNA sequencing analysis of the *BCL11A* PCR product amplified from gDNA extracted from marrow CD34⁺ cells indicated 15% gene editing comprised of insertions and deletions (**Fig. 14A**). Importantly, all deletions resulted in deletion of the GATA1 motif and all insertions disrupted GATA1 motif through addition of a small number of bp in the motif. CD34⁺ cells were plated into colony forming assays and the mixed hematopoietic colonies (GEMMs), which correspond to CD34⁺ cell clones, were picked. gDNA was isolated and analyzed by Illumina sequencing to quantify monoallelic and biallelic disruption of the target site. Most GEMMs differentiated from the CD34⁺ cell clones had monoallelic disruption and biallelic disruption was also detected, with the overall indel rate ~2/3 higher compared to what was detected in the bulk CD34⁺ cell population (**Fig. 14B**). This was likely a reflection of the percentage of common myeloid progenitors (CMPs) that give rise to GEMMs that make up a larger fraction of the heterogeneous CD34⁺ cells versus the other lineages present, but not captured/differentiated in the short-term CFC assays. The RNP treated marrow CD34⁺ cells also maintained similar kinetics of erythroid maturation (enucleation, **Fig. 14C**) and differentiation (phenotype

acquisition, **Fig. 14D**) compared to donor matched untreated control cells. Erythroid progeny of edited marrow CD34⁺ cells exhibited ~5-fold increase in HbF induction as determined by flow cytometry analysis (**Fig. 14E**).

[0295] Gene editing and induction of fetal hemoglobin was also evaluated in human adult mPB CD34⁺ cells. Co-delivery of *BCL11Ae* RNP and a non-specific ssODN supported ~20% indels at the target site (**Fig. 15A**). To evaluate early induction of fetal hemoglobin in erythroid progeny of edited cells, mPB CD34⁺ cells were differentiated into erythroblasts and induction of fetal hemoglobin transcription (*HBG* mRNA) was evaluated by qRT-PCR analysis. The erythroid progeny of *BCL11Ae* RNP treated CD34⁺ cells exhibited a 2-fold induction of *HBG* mRNA compared to untreated controls, suggesting induction of fetal hemoglobin expression (**Fig. 15B**). The RNP treated marrow CD34⁺ cells also maintained similar kinetics of differentiation (phenotype acquisition, **Fig. 15C**) compared to donor matched untreated control cells.

Example 8: Co-delivery of *S. pyogenes* Cas9 protein complexed to a truncated (15-mer) "dead" gRNA increases editing of the *HBG* promoter region in adult mobilized peripheral blood (mPB) CD34⁺ cells.

[0296] Delivery of a single wild-type (WT) ribonucleoprotein (RNP) (*e.g.*, WT Cas9 protein complexed to Sp37 guide RNA (gRNA), see **Table 10**) targeting the *HBG* promoter supports ~1.5% indels editing in human CD34⁺ cells (*see* International Patent Application No. PCT/US17/22377 by Gori et al., filed March 14, 2017, which is incorporated by reference herein). It was hypothesized that co-delivery of a dead RNP (dRNP), comprised of a catalytically active WT Cas9 protein and a truncated dead gRNA (dgRNA) that binds proximal to the target site in the *HBG* promoter (-110 nt) would increase the accessibility of a WT RNP (*e.g.*, catalytically active WT Cas9 complexed to a full-length gRNA (*e.g.*, Sp37 gRNA, see **Table 10**)) to the target site. Therefore, dead guide RNAs (dgRNAs) were designed that target the regions proximal to the -110 target site in the *HBG* promoter and have a truncated targeting domain (see **Fig. 16** and **Table 10**).

[0297] To increase editing at the target site in mobilized peripheral blood (mPB) CD34⁺ cells, WT Cas9 protein was complexed to a truncated gRNA (*i.e.*, dead (d)RNA15-mer version of wild-type SpA, which was truncated (t) at the 5' end of the gRNA sequence (tSpA dgRNA, see **Table 10**); tSpA dRNP). RNP comprised of dgRNA complexed to WT Cas9 is able to bind to sequence but does not cut genomic DNA homologous to the gRNA sequence. To determine the

optimal ratio of dead RNP:WT RNP for the assay, tSpA dRNP:Sp37 WT RNP mixed at different ratios were Nucleofected™ into mPB CD34⁺ cells, keeping the total RNP concentration constant at 3.75 μ M.

Table 10: List of selected guide RNAs and dead guide RNAs

Guide RNA (gRNA)* or Dead guide (dgRNA) ID	Targeting domain sequence (RNA)	Targeting domain sequence (DNA)	Sense
Sp35 gRNA	CUUGUCAAGGCUAUUGGUCA (SEQ ID NO: 339)	CTTGTC AAGGCTATTGGTC A (SEQ ID NO: 917)	Antisense
Sp37 gRNA	CUUGACCAAUAGCCUUGACA (SEQ ID NO: 333)	CTTGACCAATAGCCTTGAC A (SEQ ID NO: 915)	Sense
SpA gRNA	GGCAAGGCUGGCCAACCCAU (SEQ ID NO: 340)	GGCAAGGCTGGCCAACCCA T (SEQ ID NO: 919)	Sense
tSpA dgRNA	GGCUGGCCAACCCAU (SEQ ID NO:970)	GGCTGGCCAACCCAT (SEQ ID NO:971)	Sense
Sp180 dgRNA	GCCGGCGGCUGGCUA (SEQ ID NO:972)	GCCGGCGGCTGGCTA (SEQ ID NO:973)	
Sp181 dgRNA	AGUGAGGCCAGGGGC (SEQ ID NO:974)	AGTGAGGCCAGGGGC (SEQ ID NO:975)	
Sp182 dgRNA	UUAGAGUAUCCAGUG (SEQ ID NO:976)	TTAGAGTATCCAGTG (SEQ ID NO:977)	
* None of the guide RNAs or dead guide RNAs in Table 10 and used in the experiments of Example 8 are modified to recruit an exogenous trans-acting factor.			

[0298] tSpA dRNP co-delivered with Sp37 WT RNP at a ratio of 1:4 (dRNP : Total RNP ratio 1:5; 0.75 μ M dRNP : 3.75 μ M Total RNP) supported a ~4.3-fold increase in indels (as determined by T7E1 endonuclease analysis of *HBG2* PCR product amplified from gDNA extracted from CD34⁺ cells) compared to CD34⁺ cells treated with 3.75 μ M live Sp37 WT RNP alone (**Fig. 17**). These data show that dRNP paired with WT RNP can increase editing at a target in adult CD34⁺ cells.

[0299] To determine whether co-delivery of dead RNP would increase editing of *HBG* target site, Sp181 dRNP (comprising Sp181 dgRNA (**Table 10**)) and tSpA dRNP (comprising tSpA

dgRNA (**Table 10**) targeting the same strand of Sp35) were co-delivered with Sp35 by Maxcyte electroporation into mPB CD34⁺ cells.

[0300] Electroporation of Sp35 WT RNP (3.75 μ M) alone does not support detectable indels (by T7E1 endonuclease analysis) (**Fig. 18**). However, co-delivery of Sp35 WT RNP (3 μ M) with either dRNP (Sp181 dRNP and tSpA dRNP) (Sp181 dgRNA (**Table 10**) or tSpA dgRNA (**Table 10**), at 0.75 μ M) increased indels from 0% to 10% compared to CD34⁺ cells electroporated with Sp35 WT RNP alone (by T7E1 endonuclease analysis) (**Fig. 18**). Importantly, the level of indels detected in the mPB CD34⁺ cells was maintained in the day 7 erythroid progeny of edited cells (**Fig. 18**, white bars).

[0301] Additional dead/WT pairs of RNPs were tested to determine the effect of co-delivery of dead/WT RNPs on editing the target site in the *HBG* promoter (-110 nt) and resulting expression of HbF. Dead/WT pairs of RNPs (as shown in **Table 11**) were co-delivered by electroporation into mPB CD34⁺ cells. dRNP was codelivered with WT RNP (*i.e.*, Sp35 gRNA+ tSpA dgRNA, Sp35 gRNA +Sp181 dgRNA, and Sp37 gRNA+ tSpA dgRNA) at a ratio of 1:4 (dRNP : Total RNP ratio 1:5; 0.75 μ M dRNP : 3.75 μ M Total RNP).

Table 11: Percentage Editing and HbF Production by Co-Delivery of RNPs

gRNA Pairs	RNPs	Strategy	% Editing (T7E1)	%HbF
Sp37 gRNA +SpA gRNA	Sp37 D10A RNP +SpA D10A RNP	D10A*	22.44	11.2
Sp85 gRNA + SpA gRNA	Sp85 D10A RNP +SpA D10A RNP	D10A	10.5	5
Sp36 gRNA +Sp85 gRNA	Sp36 D10A RNP +Sp85 D10A RNP	D10A	5	5.3
Sp35 gRNA +tSpA dgRNA	Sp35 WT RNP +tSpA dRNP	WT Live/Dead	10	9.6
Sp35 gRNA +Sp181 dgRNA	Sp35 WT RNP +Sp181 dRNP	WT Live/Dead	8.27	12.71
Sp37 gRNA +tSpA dgRNA	Sp37 WT RNP +tSpA dRNP	WT Live/Dead	33.66	13.48
*= D10A is a Cas9 nickase that makes a single strand nick.				

[0302] tSpA dRNP co-delivered with Sp35 WT RNP, Sp181 dRNP co-delivered with Sp35 WT RNP, and tSpA dRNP co-delivered with Sp37 WT RNP supported editing of the *HBG* promoter (as determined by T7E1 endonuclease analysis of *HBG2* PCR product amplified from gDNA extracted from CD34⁺ cells) and resulted in induction of HbF protein (as determined by HPLC

analysis of hemoglobin expression in erythroid progeny according to the HPLC method described in Chang 2017 at pp. 143-44 and/or UPLC analysis, incorporated by reference herein) (**Table 3**). These data show that dRNP paired with WT RNP can support editing at a target region in adult CD34⁺ cells, resulting in HbF protein expression in erythroid progeny of the edited adult CD34⁺ cells.

Example 9: Tracking edited HSC contribution to hematopoiesis based on tracking edited alleles in their progeny *in vivo*.

[0303] DNA lesions created by paired Cas9 WT and nickases (such as the D10A and N863A mutants) can lead to a variety of repair outcomes, including a wide spectrum of insertions and deletions in the region proximal to the nicks (Bothmer 2017). However, in contrast to wild-type Cas9, the repair outcomes induced by paired nickases are more diverse and have a more uniform distribution of frequencies of specific indels (Bothmer 2017).

[0304] The diverse repair outcomes obtained after repair of double strand breaks made by WT CRISPR nucleases or after paired nicking can then be used to estimate the diversity of edited HSCs that are contributing to blood production. This is important because for life-long hematopoiesis from an edited cell pool, multiple edited HSCs must retain their ability to produce blood and self-renew for a human life span. For a specific target site, editing using a CRISPR with a gRNA that has specificity for one target site in the genome suggests that only one site will be modified, presenting the challenge of distinguishing edited alleles among the many HSCs, and thus hard to determine whether multiple HSCs are contributing to hematopoiesis. However, this presents a unique advantage to tracking the edited cells based on subtle differences in DNA repair outcomes that can occur within each allele and in each cell. The unique alleles are distinguished from each other based on indel characteristics including the type and size of the edit (insertion, deletion, insertion/deletion, and number of nucleotides deleted or inserted) and on their relative distance to the cut site and within the amplicon. For example, each deletion or insertion observed when sequencing the cell population can be characterized by its position in the genome, its length, and in the case of insertions, its sequence. The combination of these features can be used as an indel barcode to track the persistence of HSCs and their differentiation into mature blood cells as a measure of diversity after editing (**Fig. 19**). Importantly, unlike in other CRISPR indel barcoding approaches or in contrast to gene therapy approaches, the indel barcode is a potentially functional edit at the target locus, requiring no

further modification of the genome for purposes of tracking. Although it is possible for different cells to be independently edited in a way that creates the same edit, tracking by indel barcodes can establish a lower bound on the diversity of a population. Because each allele in a diploid cell can receive a different indel, that possibility must also be accounted for in diversity estimates.

[0305] It is hypothesized that if multiple unique indels in hematopoietic stem cells (HSCs) and progeny are detected then edited HSC diversity is maintained after editing. A method of tracking unique edited alleles to determine whether HSC diversity is maintained is disclosed herein. First, an RNA-guided nuclease and guide RNA complexed to form a ribonucleoprotein (RNP) complex for editing is electroporated into cells and several CD34⁺ cells repair the DNA slightly differently to create unique alleles. Over time after transplantation into an animal, edited HSCs repopulate the blood system and can be collected and sorted based on the different tissues and lineages to evaluate specific unique indels in the long-term engrafted HSCs and in differentiated progeny.

[0306] An *HBB* locus was used as a model to illustrate this method and determine whether HSC diversity is maintained. CD34⁺ cells were electroporated with D10A nickase RNPs targeting the *HBB* locus as described using the methods for electroporation provided in Example 1. Before transplantation, genomic DNA was harvested from an aliquot of the bulk pre-infusion CD34⁺ cell product, sequenced, and reads aligned to a reference sequence encompassing the target site at the *HBB* locus. The remainder (majority bulk) of the CD34⁺ cells were transplanted into mice. Four months after transplantation, human cells were purified from the hematopoietic organs of the mice (peripheral blood [PB], spleen, and bone marrow [BM]) and the human cell lineages (myeloid, erythroid, lymphoid, CD34⁺ or HSCs) were further purified. The genomic DNA was isolated from all of these human cells derived from the engrafted edited HSCs and sequenced (sequencing reads were aligned to the reference locus). The percentage of each unique edited allele over the total sum of all edited alleles detected was plotted to determine their relative contribution (**Fig. 20**). The black bars represent a group of all unique alleles occurring at low frequencies of total edited alleles. White and grey bars correspond to the top five most abundant unique alleles ranked (**Fig. 20**). In the bulk CD34⁺ cell preinfusion product, the top five most abundant clones together make up less than 10% of total edited alleles, consistent with the diversity and heterogeneity of cell types within the bulk CD34⁺ cell population. There are many unique alleles at less than 1% each grouped into the black bar. An analysis of the top five

most abundant alleles in mouse 1 (that is, *in vivo* after transplantation of and long-term engraftment of edited HSCs derived from the heterogeneous preinfusion product) indicates that there are shared unique HSC alleles across tissues and in different lineages (**Fig. 21**). An analysis of the top 5 most abundant clones in mouse 2 indicates that, for the most part, different unique alleles are present *in vivo* compared to mouse 1, but there are also shared alleles across tissues and lineages (**Fig. 21**). These data show that multiple edited HSCs are making blood *in vivo* and that a repertoire of unique edits are detected across multiple tissues. This demonstrates that there is diversity among the alleles and that no one edited allele is dominant over others. This method provides a means to survey for any unintended effects of alleles at the target site providing a readout on safety of editing.

Example 10: Co-delivery of active RNP with inactive RNP increases total editing at the distal CCAAT box target site.

[0307] To determine whether co-delivery of inactive RNP (i.e., dead RNP (“dRNP”), dead guide RNA (“dgRNA”) complexed with active Cas9) would increase editing of wild-type active RNP (i.e., active gRNA complexed with active Cas9) at the distal CCAAT box target site of the *HBG* promoter, 8 μ M Sp37 RNP (comprising Sp37 gRNA (SEQ ID NO:333, **Table 10**) complexed with *S. Pyogenes* wildtype Cas9) was delivered alone or co-delivered with 8 μ M tSpA dRNP (comprising tSpA dgRNA (SEQ ID NO: 970, **Table 10**) complexed with *S. Pyogenes* wildtype Cas9) by electroporation into mPB CD34+ cells. In this experiment, Sp37 gRNA and tSpA dgRNA were generated via *in vitro* transcription (IVT). **Fig. 22** shows the distal CCAAT box target site at the *HBG* promoter and the relative locations of homologous sequences to Sp37 gRNA, which binds the CCAAT box target site, and tSpA dgRNA, which binds proximal to the CCAAT box target site. Co-delivery of inactive tSpA dRNP increased total editing of active Sp37 RNP by more than two times the level of editing shown with delivery of active Sp37 RNP alone (**Fig. 23**). Although the addition of tSpA dRNP resulted in an increase in editing, the increase in editing did not correspond to an increase in HbF protein expression *ex vivo* (**Fig. 23**).

[0308] To determine whether gRNAs that have been chemically synthesized influence editing compared with gRNA generated via IVT, 8 μ M Sp37 RNP (comprising chemically synthesized Sp37 gRNA (SEQ ID NO:333, **Table 10**) complexed with *S. Pyogenes* wildtype Cas9) was delivered or co-delivered with 8 μ M tSpA dRNP (comprising chemically synthesized tSpA dgRNA (SEQ ID NO:970, **Table 10**) complexed with *S. Pyogenes* wildtype Cas9) by

electroporation into mPB CD34⁺ cells. Co-delivery of inactive tSpA dRNP comprising chemically synthesized dgRNA with Sp37 RNP comprising chemically synthesized gRNA increased total editing by active Sp37 RNP (as determined by NGS (Illumina sequencing analysis) of HBG1 PCR product amplified from gDNA extracted from CD34⁺ cells) (**Fig. 24**). A 6-fold increase in editing with RNP comprising synthetic gRNA was seen versus RNP comprising gRNA generated via IVT (compare **Fig. 23** with **Fig. 24**). Similar to the RNP comprising IVT gRNA, RNP comprising chemically synthesized gRNA increased editing, but did not result in an increase in HbF protein expression (as determined by UPLC tetramer analysis of hemoglobin expression in erythroid progeny of the edited HSPCs). These data indicate that pairing with a dgRNA can increase the total editing level of active gRNA.

Example 11: Co-delivery of active RNP with an inactive RNP binding proximally alters the indel profile.

[0309] Different concentrations of inactive dRNP and active RNP were tested to evaluate the effect of co-delivering a proximally binding dRNP with an active RNP on the indel profile. Briefly, active Sp37 RNP (comprising IVT Sp37 gRNA (SEQ ID NO:333, **Table 10**) complexed with *S. Pyogenes* wildtype Cas9) at decreasing concentrations (6.4 μ M, 4.0 μ M, 1.6 μ M, 0 μ M) was co-delivered with increasing concentrations (1.6 μ M, 4.0 μ M, 6.4 μ M, 8.0 μ M) of inactive tSpA dRNP (comprising IVT tSpA dgRNA (SEQ ID NO:970, **Table 10**) complexed with *S. Pyogenes* wildtype Cas9) by electroporation into mPB CD34⁺ cells. T7E1 endonuclease analysis indicated that the editing activity of Sp37 RNP at *HBG2* can be increased by co-delivering an inactive RNP (**Fig. 25**). NGS sequencing analysis was performed to quantify the type and frequency of indels generated when delivering Sp37 RNP only or Sp37 RNP paired with tSpA dRNP (**Figs. 26A-D**). Co-delivering tSpA dRNP increased the frequency of deletions extending from the active gRNA cut site toward the dgRNA binding site. Increasing the relative dose of tSpA RNP accentuated that effect (**Figs. 26A-D**). These data confirm that as the dgRNA concentration increases, the directionality of the deletions generated by the active gRNA changes in favor of deletions extending toward the dgRNA binding site.

[0310] The most common deletions (contributing to >1% of all indels) generated with tSpA dRNP co-delivered with Sp37 RNP extended from the Sp37 target site to the tSpA target site, but they did not extend past the tSpA target site (**Fig. 27B**). This indicates that the maximum length

of indels extending from the wildtype cut site to the inactive binding site does not frequently extend beyond the inactive binding site.

[0311] Next, the deletions were analyzed to determine whether they are likely to have been generated by microhomology-mediated end joining (MMEJ). While NHEJ repair occurs during all phases of the cell cycle, it is uncertain whether MMEJ repair can occur at high frequency in the G0/G1 phase (Truong 2013), where most of the long-term repopulating hematopoietic stem cells reside. MMEJ repair of DSBs rely on the annealing of DNA ends by base pairing of short microhomology sequences. The sequences of the deletions were systematically analyzed and compared with the sequence flanking the deletion to identify deletions associated with microhomologies, which may have been generated via the MMEJ repair pathway (as determined by deletions whose sequence starts or ends with 2 nt or more of homology with the sequence flanking the other extremity of the deletion (**Fig. 31**). Sequencing analysis demonstrated that the most common deletions generated using Sp37 RNP alone were associated with microhomologies, and were thus likely generated by micro-homology mediated end-joining (MMEJ) repair. In particular, the largest deletions frequently observed (>5 nt long and contributing to $\geq 1\%$ of all indels) using Sp37 RNP alone were all associated with microhomologies whereas co-delivery of tSpA dRNP with Sp37 RNP increased the frequency of large deletions extending toward the inactive dgRNA binding site without relying on micro-homology domains for repair (**Figs. 27A-B**).

[0312] These data show that pairing with a dgRNA can change the editing outcome (directionality and use of micro-homologies for repair) of an active gRNA. Thus, dgRNA can be used to “orientate” deletions toward a desired target site to increase the frequency of target site disrupting indels. In addition, whereas deletions of length >5 nt are only frequently generated by CRISPRCas9 in the presence micro-homology domains (van Overbeek 2016), the dgRNA can be used to increase the frequency of large deletions, independently of the presence of microhomology domains, or in cells lacking the machinery to use the MMEJ repair pathway.

Example 12: Distal CCAAT box disrupting indels promote HbF expression.

[0313] DNA lesions created by single or paired Cas9 RNP (using WT or nickase Cas9 protein, such as the D10A and N863A mutants) can lead to a variety of repair outcomes, including a wide spectrum of insertions and deletions. The phenotypic outcome in edited cells may vary according to the type of generated indels. Characterizing a target domain whose disruption is

required to achieve a desired phenotypic outcome can allow the design of dedicated editing approaches aiming at maximizing the frequency of cells in a population of cells having the desired genotype and thus the desired phenotype. In particular, targeting the *HBG* distal CCAAT box is likely to disrupt the binding of regulatory proteins repressing the expression of the *HBG* genes. The length and position of indels may affect their capacity to de-repress *HBG* expression, for example, based on their capacity to prevent the binding of repressive factors or whether they disrupt nearby essential regulatory elements.

[0314] To identify indels resulting in productive HbF induction at the proximity of the distal CCAAT box, HSPCs were electroporated by single or paired Cas9 RNP targeting the distal CCAAT box, differentiated into erythroid cells and sorted based on HbF expression into a low HbF expressing fraction and a high HbF expressing fraction. The frequency of each indel was quantified in both fraction, allowing to identify indels enriched in the high-HbF fraction (**Fig 28A**).

[0315] Briefly, 8 different sets of single or paired Cas9 RNP:

- (1) 3 μ M Sp37 RNP (comprising IVT Sp37 gRNA (SEQ ID NO:333, **Table 10**) complexed with *S. Pyogenes* D10A Cas9) and 3 μ M SpA dRNP (comprising IVT SpA gRNA (SEQ ID NO:340, **Table 10**) complexed with *S. Pyogenes* D10A Cas9);
- (2) 4 μ M Sp37 RNP (comprising IVT Sp37 gRNA (SEQ ID NO:333, **Table 10**) complexed with *S. Pyogenes* D10A Cas9) and 4 μ M tSpA dRNP (comprising IVT SpA gRNA (SEQ ID NO:340, **Table 10**) complexed with *S. Pyogenes* WT Cas9);
- (3) 8 μ M Sp37 RNP (comprising IVT Sp37 gRNA (SEQ ID NO:333, **Table 10**) complexed with *S. Pyogenes* WT Cas9);
- (4) 6.4 μ M Sp35 RNP (comprising IVT Sp35 gRNA (SEQ ID NO:339, **Table 10**) complexed with *S. Pyogenes* WT Cas9) and 1.6 μ M tSpD RNP (comprising IVT tSpD (targeting domain sequence set forth in SEQ ID NO:978 (RNA) and SEQ ID NO:979 (DNA)) complexed with *S. Pyogenes* wildtype Cas9);
- (5) 8 μ M Sp35 RNP (comprising chemically synthesized Sp35 gRNA (SEQ ID NO:339, **Table 10**) complexed with *S. Pyogenes* WT Cas9);
- (6) 8 μ M Sp35 RNP (comprising chemically synthesized Sp35 gRNA (SEQ ID NO:339, **Table 10**) complexed with *S. Pyogenes* WT Cas9);

(7) 8 μ M Sp35 RNP (comprising chemically synthesized Sp35 gRNA (SEQ ID NO:339, **Table 10**) complexed with *S. Pyogenes* WT Cas9) and 2 μ M tSpA dRNP (comprising chemically synthesized tSpA dgRNA (SEQ ID NO:970, **Table 10**) complexed with *S. Pyogenes* WT Cas9); and

(8) 8 μ M Sp37 RNP (comprising chemically synthesized Sp37 gRNA (SEQ ID NO:333, **Table 10**) complexed with *S. Pyogenes* WT Cas9) and 2 μ M tSpA dRNP (comprising IVT tSpA dgRNA (SEQ ID NO:970, **Table 10**) complexed with *S. Pyogenes* WT Cas9) were delivered by electroporation into mPB CD34+ cells and sequenced. Erythroid cells derived from RNP treated HSPCs were sorted based on HbF expression into a low HbF expressing fraction and a high HbF expressing fraction (as determined by the level of staining using an anti-gamma chain antibody) by fluorescence-activated cell sorting. The frequency of each indel was quantified in both fraction by amplification of the *HBG* PCR amplicon from the extracted gDNA and NGS analysis.

[0316] It was observed that most of the indels enriched in the high HbF fraction (i.e., associated with HbF expression) were indels disrupting the distal CCAAT box (**Fig. 28B**, see “dCCAAT”). Deletions that only disrupted the proximal CCAAT or the KLF binding area (**Fig. 28B**, see “pCCAAT” and “KLF” respectively) had a detrimental effect or no effect on HbF induction. Deletions that disrupted the distal CCAAT box and that co-disrupted the proximal CCAAT box or the KLF1 binding area appeared to diminish HbF production when compared to deletions only disrupting the distal CCAAT box (**Fig. 28B**, compare “dCCAAT & pCCAAT” and “dCCAAT & KLF” with “dCCAAT”). While the distal CCAAT box disruption appeared to be necessary for HbF induction, the effect of the deletion length both 5’ and 3’ of the distal CCAAT box was analyzed. It was determined that deletions of the distal CCAAT box extending up to 22 nt upstream (starting 3’ of pos. HBG -88) and up to 15 nt downstream (ending 5’ of pos. HBG -131) of the dCCAAT box (positions flanked by the proximal CCAAT box (“pCCAAT”) and KLF1 binding area (“KLF”)) were able to support HbF induction (**Figs. 28C** and **Fig. 28D**). In addition, it was determined that deletions of ≥ 3 nt of the dCCAAT were more frequently associated with high HbF expression (**Fig. 28E**). A list of example indels that were shown to induce HbF expression (i.e., productive indels) is found in **Table 12**.

Table 12: Exemplary deletions that induce HbF expression*

Position Relative to <i>HBG</i> Transcription Start Site	Genomic Location 1 (<i>HBG1</i>)	Genomic Location 2 (<i>HBG2</i>)
c.-114 to -102	Hg38 Chr11:5249959 to Hg38 Chr11:5249971	Hg38 Chr11:5254883 to Hg38 Chr11:5254895
-119:-104	Hg38 Chr11:5249961 to Hg38 Chr11:5249976	Hg38 Chr11:5254885 to Hg38 Chr11:5254900
-121:-104	Hg38 Chr11:5249961 to Hg38 Chr11:5249978	Hg38 Chr11:5254885 to Hg38 Chr11:5254902
-104:-78	Hg38 Chr11:5249935 to Hg38 Chr11:5249961	Hg38 Chr11:5254859 to Hg38 Chr11:5254885
-108:-82	Hg38 Chr11:5249939 to Hg38 Chr11:5249965	Hg38 Chr11:5254863 to Hg38 Chr11:5254889
-115:-112	Hg38 Chr11:5249942 to Hg38 Chr11:5249945	Hg38 Chr11:5254893 to Hg38 Chr11:5254896
-113:-107	Hg38 Chr11:5249964 to Hg38 Chr11:5249970	Hg38 Chr11:5254888 to Hg38 Chr11:5254894
-114:-107	Hg38 Chr11:5249964 to Hg38 Chr11:5249971	Hg38 Chr11:5254888 to Hg38 Chr11:5254895
-123:-113	Hg38 Chr11:5249970 to Hg38 Chr11:5249980	Hg38 Chr11:5254894 to Hg38 Chr11:5254904
-119:-104	Hg38 Chr11: 5249961 to Hg38 Chr11: 5249976	Hg38 Chr11:5254885 to Hg38 Chr11:5254900
-119:-102	Hg38 Chr11: 5249959 to Hg38 Chr11: 5249976	Hg38 Chr11:5254883 to Hg38 Chr11:5254900
-127:-103	Hg38 Chr11: 5249960 to Hg38 Chr11: 5249984	Hg38 Chr11:5254884 to Hg38 Chr11:5254908
* >2 [log2] enrichment in high-HbF fraction not inclusive		

Example 13: Co-delivery of active RNP with an inactive RNP can be used to change the type of indels disrupting the distal CCAAT-box.

[0317] Sequencing analysis resulting from the electroporation of the Sp37 RNP alone showed that 35% of the generated indels disrupted the distal CCAAT box (**Figs. 29A-B** and **30A-B**). Briefly, 8 μ M Sp37 RNP (comprising chemically synthesized Sp37 gRNA (SEQ ID NO:333, **Table 10**) complexed with *S. Pyogenes* wildtype Cas9) was delivered alone or co-delivered with 8 μ M tSpA dRNP (comprising chemically synthesized tSpA dgRNA (SEQ ID NO:970, **Table 10**) complexed with *S. Pyogenes* wildtype Cas9) by electroporation into mPB CD34+ cells. Small deletions close to the Sp37 cut site did not disrupt the distal CCAAT box (see **Fig. 22** identifying the Sp37 cut site) and **Figs. 29A-B**). Most of the distal CCAAT box disruptions generated by Sp37 RNP alone were formed by 27 nt long deletions, which are associated with microhomology domains and thus likely to occur via micro-homology mediated end joining (**Figures 29A** and **Fig. 31**). Codelivery of tSpA dRNP with Sp37 RNP did not increase the frequency of CCAAT box disrupting indels (compare **Figs. 29B** and **30B**). However, the addition of the tSpA dRNP increased the frequency of distal CCAAT box deletions extending from the active Sp37 gRNA cut site toward the tSpA dgRNA binding site, including HbF-inducing distal CCAAT box deletions (compare **Fig. 29A** with **Fig. 30A** and **Fig. 28F**).

Example 14: Pairing with an inactive dead guide RNA changes the editing outcome in hematopoietic stem cells.

[0318] To determine whether co-delivery of dead RNP increases editing of the distal CCAAT-box target site in long term repopulating hematopoietic stem cells in vivo, mPB CD34+ cells electroporated with dRNP and active RNP were infused into nonirradiated NOD.B6.SCID Il2ry-/- Kit(W41/W41) ("NBSGW") mice. Briefly, 8 μ M Sp37 RNP (comprising *in vitro* transcribed Sp37 gRNA (SEQ ID NO:333, **Table 10**) complexed with *S. Pyogenes* wildtype Cas9) was delivered alone or co-delivered with 8 μ M tSpA dRNP (comprising *in vitro* transcribed tSpA dgRNA (SEQ ID NO:970, **Table 10**) complexed with *S. Pyogenes* wildtype Cas9) by Maxcyte HBGM34-2 electroporation into mPB CD34+ cells. Cells were then infused into NBSGW mice (n= 5) and the percentage of editing in bone marrow engrafted human cells was evaluated after 16 weeks *in vivo*. Co-delivery with dRNP resulted in increased total editing by Sp37 RNP compared with Sp37 RNP alone (**Fig. 32A**). In particular, co-delivery with dRNP resulted in increased distal CCAAT box disrupting deletions (**Fig. 32B**).

[0319] Deletions associated with micro-homology domains were observed in cells pre-infusion, but were not frequently observed *in vivo* 16 weeks after transplantation (**Figs. 32A-C**). These data indicate that indels associated with micro-homologies occur at lower frequencies in HSCs and therefore do not frequently persist long term after transplantation of edited mPB CD34+ cells. MMEJ repair is likely favored in the fraction of CD34 cells consisting of actively cycling progenitors. Those progenitors that do not self-renew and thus have a short life span *in vivo*. The drop in the frequency of micro-homology associated indels observed between pre-infusion and after 16 weeks post infusion may be explained by the progressive disappearance *in vivo* of those short lived progenitors.

[0320] Surprisingly, the co-delivery of tSpA dRNP with Sp37 RNP increased the frequency of distal CCAAT box deletions not associated with micro-homologies pre-infusion (**Fig. 32C**). While most distal CCAAT box deletions generated by Sp37 RNP alone are associated with micro-homologies and were not frequently observed *in vivo* 16 weeks after transplantation, distal CCAAT box deletions generated by co-delivery of the inactive tSpA dRNP and not associated with micro-homologies were better maintained *in vivo* 16 weeks after transplantation (**Figs. 32A-C**).

[0321] These results demonstrate that co-delivery of inactive dRNP can increase the frequency of indels in cells where the MMEJ repair pathway is reduced, such as quiescent hematopoietic stem cells. In addition, whereas deletions >5 nt are only frequently generated by CRISPRCas9 single RNP in the presence of micro-homologies, using a dRNP promoted the frequency of >5 nt deletions independently of micro-homologies. Finally, using a dRNP that promotes deletions disrupting the target *HBG* distal CCAAT box can be used to increase the frequency of predicted productive (HbF inducing), distal CCAAT box deletions in hematopoietic stem cells.

SEQUENCES

[0322] Genome editing system components according to the present disclosure (including without limitation, RNA-guided nucleases, guide RNAs, donor template nucleic acids, nucleic acids encoding nucleases or guide RNAs, and portions or fragments of any of the foregoing), are exemplified by the nucleotide and amino acid sequences presented in the Sequence Listing. The sequences presented in the Sequence Listing are not intended to be limiting, but rather illustrative of certain principles of genome editing systems and their component parts, which, in combination with the instant disclosure, will inform those of skill in the art about additional

implementations and modifications that are within the scope of this disclosure. A list of the sequences presented is provided in the following **Table 13**.

Table 13: Sequences presented in the Sequence Listing:

SEQ ID NOS:	Description
1-2, 4-6, 12, 14	Cas9 polypeptides
3, 7-11, 13	Cas9 coding sequences
15-23, 52-123	Cas9 RuvC-like domains
24-28, 124-198	Cas9 HNH-like domains
29-31, 38-51	Full-length modular and unimolecular gRNAs
32-37	gRNA proximal and tail domains
199-205	PAM sequences
251-901, 940-942, 952-955	gRNA targeting domains (RNA)– see Tables 2, 7, 9, 10
910-919, 943-945, 956-959	gRNA targeting domains (DNA)– see Tables 7, 9, 10
920-929, 946-948, 960-963	gRNA targeting domains plus PAM (NGG) (RNA) – see Tables 7, 9
930-939, 949-951, 964-967	gRNA targeting domains plus PAM (NGG) (DNA) – see Tables 7, 9
970, 972, 974, 976	dgRNA targeting domains (RNA) – see Table 10
971, 973, 975, 977	dgRNA targeting domains (DNA) – see Table 10
902, 903	Human HBG1, 2 promoter sequences including HPFH deletion site
904-909	Oligonucleotide donor sequences and homology arms – see Table 8
968-969	<i>BCL11A</i> e sequences

INCORPORATION BY REFERENCE

[0323] 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

[0324] Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments described herein. Such equivalents are intended to be encompassed by the following claims.

REFERENCES

- Aliyu et al. Am J Hematol 83:63-70 (2008)
- Angastiniotis & Modell Ann N Y Acad Sci 850:251-269 (1998)
- Anders et al. Nature 513(7519):569-573 (2014)
- Bae et al. Bioinformatics 30(10):1473-1475 (2014)
- Bothmer et al. Nat Commun 8:13905 (2017)
- Bouva Hematologica 91(1):129-132 (2006)
- Briner et al. Mol Cell 56(2):333-339 (2014)
- Brousseau Am J Hematol 85(1):77-78 (2010)
- Canvers et al. Nature 527(12):192-197 (2015)
- Chang et al. Mol Ther Methods Clin Dev 4:137-148 (2017)
- Chen et al. Nat Commun 8:14958 (2017)
- Cong et al. Science 399(6121):819-823 (2013)
- Cornish-Bowden Nucleic Acids Res 13(9):3021-3030 (1985)
- Davis & Maizels Proc Natl Acad Sci USA 111(10):E924-E932 (2014)
- Fine et al. Sci Rep 5:10777 (2015)
- Frit et al. DNA Repair (Amst.) 17:81-97 (2014)
- Fu et al. Nat Biotechnol 32(3):279-284 (2014)
- Guilinger et al. Nat Biotechnol 32(6):577-582 (2014)
- Heigwer et al. Nat Methods 11(2):122-123 (2014)
- Hinz et al. J Biol Chem 291(48):24851-24856 (2016)
- Hsu et al. Nat Biotechnol 31(9):827-832 (2013)
- Iyama & Wilson DNA Repair (Amst.) 12(8):620-636 (2013)
- Jiang et al. Nat Biotechnol 31(3):233-239 (2013)
- Jinek et al. Science 337(6096):816-821 (2012)
- Jinek et al. Science 343(6176):1247997 (2014)
- Kleinstiver et al. Nature 523(7561):481-485 (2015a)

- Kleinstiver et al. *Nat Biotechnol* 33(12):1293-1298 (2015b)
- Kleinstiver et al. *Nature* 529(7587):490-495 (2016)
- Komor et al. *Nature* 533(7603):420-424 (2016)
- Lee et al. *Nano Lett* 12(12):6322-6327 (2012)
- Lewis "Medical-Surgical Nursing: Assessment and Management of Clinical Problems"
(2014)
- Makarova et al. *Nat Rev Microbiol* 9(6):467-477 (2011)
- Mali et al. *Science* 339(6121):823-826 (2013)
- Martyn et al. *Biochim. Biophys Acta* 1860 (5):525-536 (2017)
- Nishimasu et al. *Cell* 156(5):935-949 (2014)
- Nishimasu et al. *Cell* 162(5):1113-1126 (2015)
- Ran et al. *Cell* 154(6):1380-1389 (2013)
- Ran et al. *Nature* 520(7546):186-191 (2015)
- Richardson et al. *Nat Biotechnol* 34(3):339-344 (2016)
- Shmakov et al. *Mol Cell* 60:385–397 (2015)
- Thein *Hum Mol Genet* 18(R2):R216-223 (2009)
- Truong et al. *PNAS* 110(19):7720-5 (2013)
- Tsai et al. *Nat Biotechnol* 34(5):483 (2016)
- van Overbeek et al. *Cell* 163(4):633-646 (2016)
- Wang et al. *Cell* 153(4):910-918 (2013)
- Xiao et al. *Bioinformatics* 30(8):1180-1182 (2014)
- Yamano et al. *Cell* 165(4):949-962 (2016)
- Zetsche et al. *Nat Biotechnol* 33(2):139-142 (2015a)
- Zetsche et al. *Cell* 163(3):759–771 (2015b)

CLAIMS

1. A genome editing system comprising:
 - (a) a dead guide RNA (dgRNA) comprising a first targeting domain sequence;
 - (b) a first guide RNA (first gRNA) comprising a second targeting domain sequence greater than 17 nucleotides in length; and
 - (c) an RNA-guided nuclease,wherein the RNA-guided nuclease and the dgRNA are configured to associate within a first target region in a target nucleic acid; and
wherein the RNA-guided nuclease and the first gRNA are configured to associate within a second target region in the target nucleic acid and generate a double stranded break (DSB) in the target nucleic acid whereby to create an indel between the first target region and the second target region.
2. The genome editing system of claim 1, wherein the dgRNA comprises a first targeting domain sequence of 15 nucleotides or less in length.
3. The genome editing system of claim 1 or 2, wherein the dgRNA comprises reduced or no RNA-guided nuclease cleavage activity.
4. The genome editing system of any one of claims 1-3, wherein the dgRNA and the RNA-guided nuclease associate with a first protospacer adjacent motif (PAM) sequence in the target nucleic acid, wherein the first PAM sequence is facing outward.
5. The genome editing system of any one of claims 1-4, wherein the first gRNA and the RNA-guided nuclease associate with a second PAM sequence in the target nucleic acid, wherein the second PAM sequence is facing outward.
6. The genome editing system of any one of claims 1-5, wherein the target nucleic acid is a human *HBG1*, *HBG2* gene, or a combination thereof.
7. The genome editing system of any one of claims 1-6, wherein the first targeting domain sequence is complementary to a first sequence on a side of a CCAAT box target region of the human *HBG1*, *HBG2* gene, or a combination thereof,
wherein the first sequence optionally overlaps the CCAAT box target region of the human *HBG1*, *HBG2* gene, or a combination thereof,

8. The genome editing system of any one of claims 1-7, wherein the second targeting domain sequence is complementary to a second sequence on the first side of a CCAAT box target region of the human *HBG1*, *HBG2* gene, or a combination thereof, and

wherein the second sequence optionally overlaps the CCAAT box target region of the human *HBG1*, *HBG2* gene, or a combination thereof.

9. The genome editing system of any one of claims 1-8, wherein the first targeting domain comprises a truncation of a gRNA targeting domain set forth in Table 2 or Table 10, wherein the gRNA targeting domain has been truncated from a 5' end of the gRNA targeting domain.

10. The genome editing system of any one of claims 1-9, wherein the first targeting domain is the same as or differs by no more than 3 nucleotides from a dgRNA targeting domain set forth in Table 10.

11. The genome editing system of any one of claims 1-10, wherein the second targeting domain is the same as or differs by no more than 3 nucleotides from a gRNA targeting domain set forth in Table 2 or Table 10.

12. The genome editing system of any one of claims 1-11, wherein the indel alters the CCAAT box target region of the human *HBG1*, *HBG2* gene, or a combination thereof.

13. The genome editing system of any one of claims 1-12, wherein the indel is a productive indel resulting in an increased level of fetal hemoglobin expression.

14. The genome editing system of any one of claims 1-13, wherein the gRNA, dgRNA, or both is chemically synthesized.

15. A method of altering a cell, comprising: contacting the cell with
(a) a dead guide RNA (dgRNA) comprising a first targeting domain sequence;
(b) a first guide RNA (first gRNA) comprising a second targeting domain sequence greater than 17 nucleotides in length; and

(c) an RNA-guided nuclease;
wherein the RNA-guided nuclease and the dgRNA are configured to associate within a first target region in a target nucleic acid;

wherein the RNA-guided nuclease and the first gRNA associate within a second target region and generate a double stranded break (DSB) in the target nucleic acid whereby to create an indel between the first target region and the second target region.

16. The method of claim 15, wherein the dgRNA comprises a first targeting domain sequence of 15 nucleotides or less in length.

17. The method of claim 15 or 16, wherein the dgRNA is configured such that it does not provide an RNA-guided nuclease cleavage event.

18. The method of any one of claims 15-17, wherein the dgRNA and the RNA-guided nuclease associate with a first protospacer adjacent motif (PAM) sequence in the target nucleic acid, wherein the first PAM sequence is facing outward.

19. The method of any one of claims 15-18, wherein the first gRNA and the RNA-guided nuclease associate with a second PAM sequence in the target nucleic acid, wherein the second PAM sequence is facing outward.

20. The method of any one of claims 15-19, wherein the target nucleic acid is a human *HBG1*, *HBG2* gene, or a combination thereof.

21. The method of any one of claims 15-20, wherein the first targeting domain sequence is complementary to a first sequence on a side of a CCAAT box target of the human *HBG1*, *HBG2* gene, or a combination thereof,

wherein the first sequence optionally overlaps the CCAAT box target region of the human *HBG1*, *HBG2* gene, or a combination thereof,

22. The method of any one of claims 15-21, wherein the second targeting domain sequence is complementary to a second sequence on the first side of a CCAAT box target region of the human *HBG1*, *HBG2* gene, or a combination thereof, and

wherein the second sequence optionally overlaps the CCAAT box target region of the human *HBG1*, *HBG2* gene, or a combination thereof.

23. The method of any one of claims 15-22, wherein the first targeting domain comprises a truncation of a gRNA targeting domain set forth in Table 2 or Table 10, wherein the gRNA targeting domain has been truncated from a 5' end of the gRNA targeting domain.

24. The method of any one of claims 15-23, wherein the first targeting domain is the same as or differs by no more than 3 nucleotides from a dgRNA targeting domain set forth in Table 10.

25. The method of any one of claims 15-24, wherein the second targeting domain is the same as or differs by no more than 3 nucleotides from a gRNA targeting domain set forth in Table 2 or Table 10.

26. The method of any one of claims 15-25, wherein the indel alters the CCAAT box target region.

27. The method of any one of claims 15-26, wherein the indel is a productive indel resulting in an increased level of fetal hemoglobin expression.

28. The method of any one of claims 15-27, wherein the gRNA, dgRNA, or both is chemically synthesized.

29. A method of altering a cell, comprising contacting a cell with the genome editing system of any of claims 1-14.

30. The method of claim 29, wherein the cell is capable of differentiating into an erythroblast, erythrocyte, or a precursor of an erythrocyte or erythroblast.

31. The method of claim 29 or 30, wherein the cell is a CD34+ cell.

32. A modified cell comprising a productive indel which results in an increase in HbF expression, wherein the indel is produced by contacting the cell with

(a) a dead guide RNA (dgRNA) comprising a first targeting domain sequence;

(b) a first guide RNA (gRNA) comprising a second targeting domain sequence greater than 17 nucleotides in length; and

(c) an RNA-guided nuclease;

wherein the RNA-guided nuclease and the dgRNA are configured to associate within a first target region in a target nucleic acid; and

wherein the RNA-guided nuclease and the first gRNA associate with a second target region and generate a double stranded break (DSB) in the target nucleic acid whereby to create an indel between the first target region and the second target region.

33. A cell, altered by the method of any of claims 15-31.

34. The cell of claim 32 or 33, wherein the cell is capable of differentiating into an erythroblast, erythrocyte, or a precursor of an erythrocyte or erythroblast.

35. The cell of any of claims 32-34, wherein the cell is a CD34+ cell.

36. A dead gRNA (dgRNA) molecule comprising a first targeting domain sequence comprising a truncation of a gRNA targeting domain set forth in Table 2 or Table 10, wherein the gRNA targeting domain has been truncated from a 5' end of the gRNA targeting domain.

37. The dgRNA of claim 36, wherein the first targeting domain is the same as or differs by no more than 3 nucleotides from a dgRNA targeting domain set forth in Table 10.

38. A kit comprising, (a) gRNA molecule, or nucleic acid that encodes said gRNA, and one or more of the following: (b) an RNA-guided nuclease; (c) a dgRNA molecule of claim 36 or 37; and (d) a nucleic acid that encodes one or both of (b) or (c).

39. The kit of claim 38, comprising a nucleic acid that encodes one or more of (a), (b) and (c).

40. A method of treating a β -hemoglobinopathy in a subject in need thereof, comprising administering to the subject the cells of any of claims 32-35.

41. A method of treating a β -hemoglobinopathy in a subject in need thereof, comprising administering a population of modified hematopoietic cells, wherein one or more cells have been altered according to the methods of claims 15-31.

42. A population of cells modified by the genome editing system of claims 1-14, wherein the population of cells comprise a higher percentage of a productive indel relative to a population of cells not modified by the genome editing system of claims 1-14.

43. A population of cells modified by the genome editing system of claims 1-14, wherein a higher percentage of the population of cells are capable of differentiating into a population of cells of an erythroid lineage that express HbF relative to a population of cells not modified by the genome editing system of claims 1-14.

44. The population of cells of claim 42 or 43, wherein the higher percentage is at least about 15%, at least about 20% , at least about 25%, at least about 30%, or at least about 40% higher.

45. The population of cells of any one of claims 42-44, wherein the cells are hematopoietic stem cells.

46. The population of cells of any one of claims 42-45, wherein the cells are capable of differentiating into an erythroblast, erythrocyte, or a precursor of an erythrocyte or erythroblast.

47. The population of cells of any one of claims 42-46, wherein the indel is created by a repair mechanism other than microhomology-mediated end joining (MMEJ) repair.

FIG 1

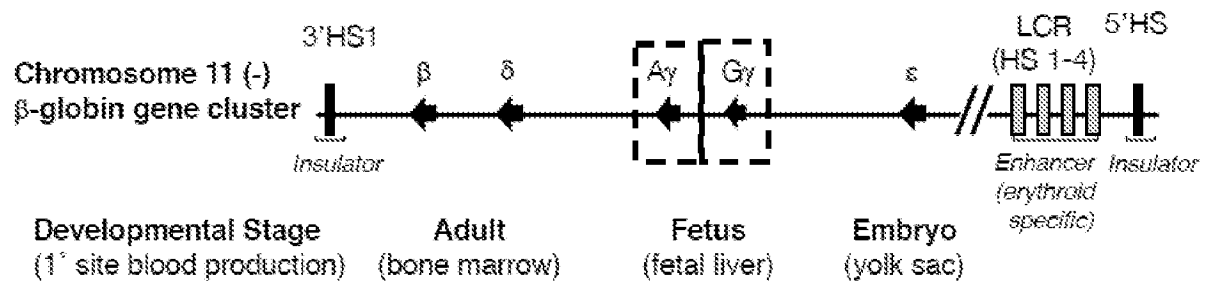


FIG 2A

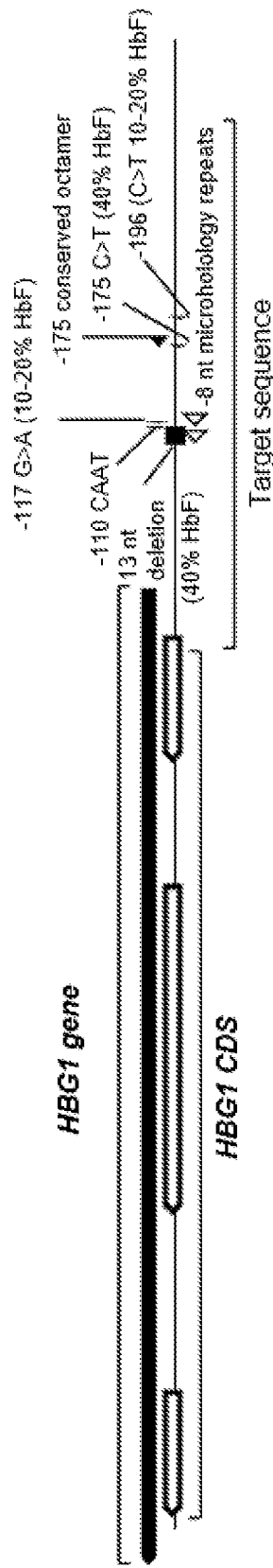


FIG 2B

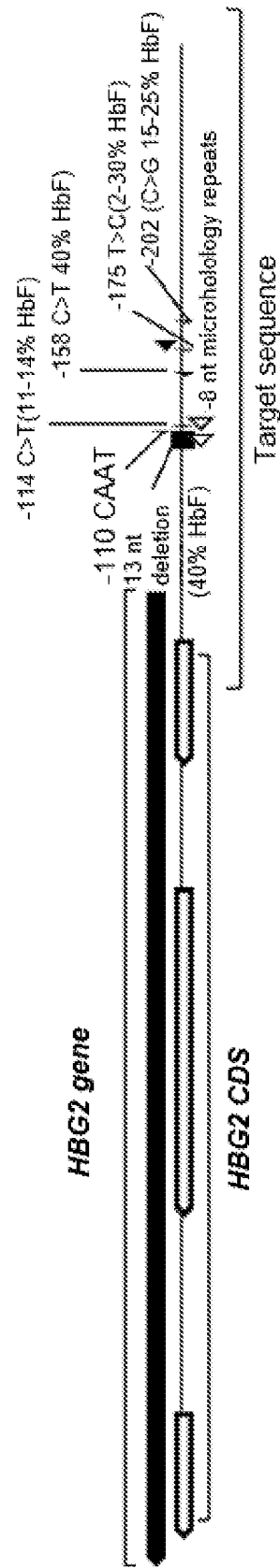


FIG 3A

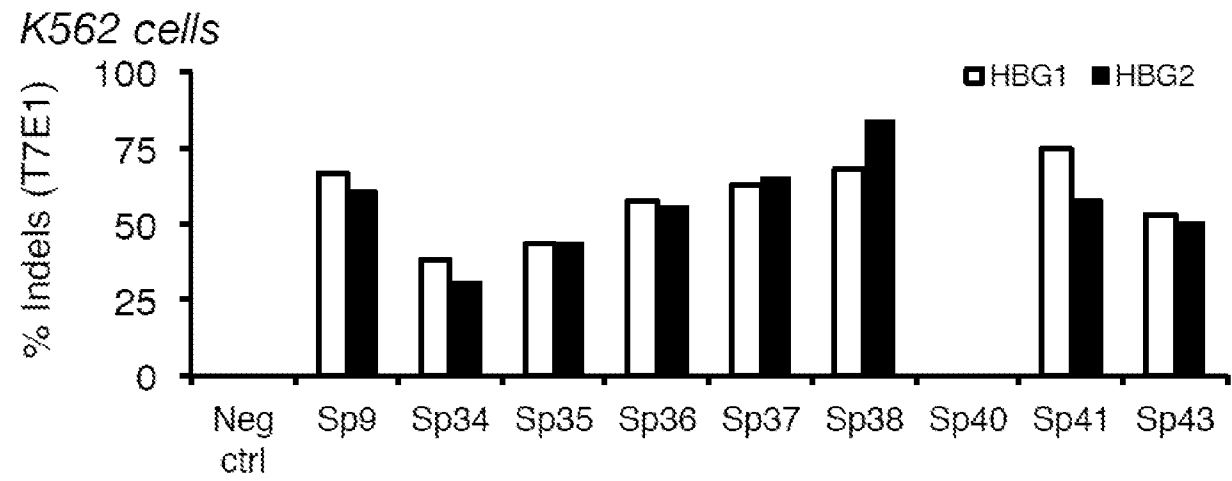


FIG 3B

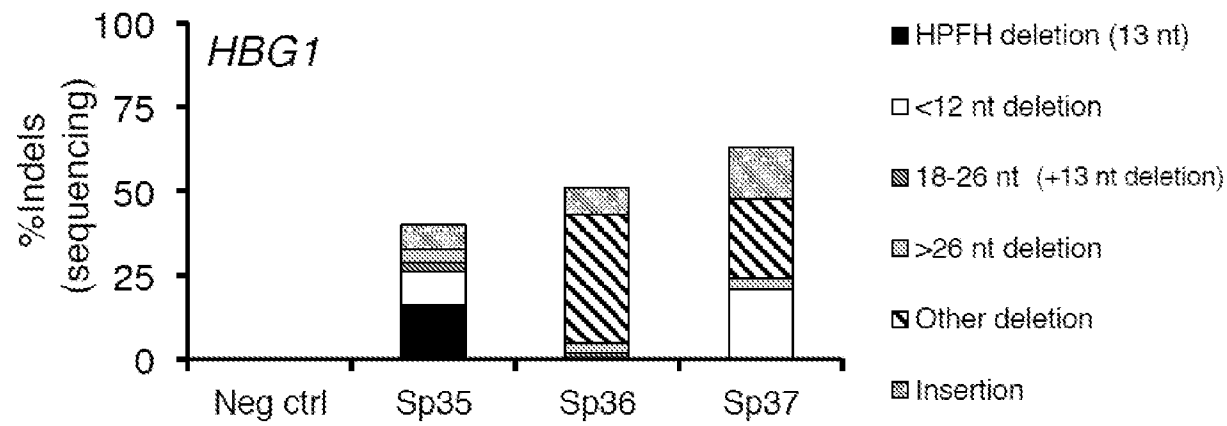


FIG 3C

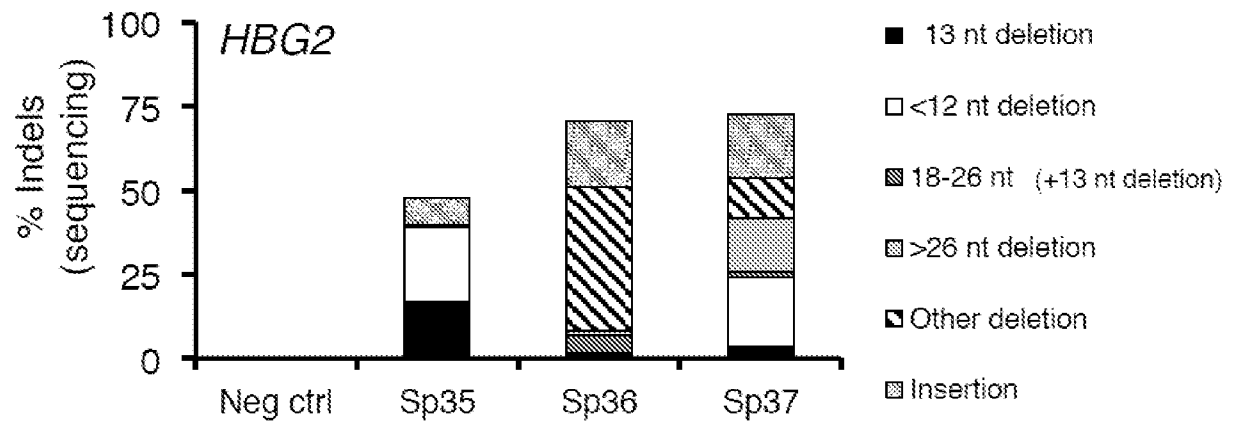


FIG 4A

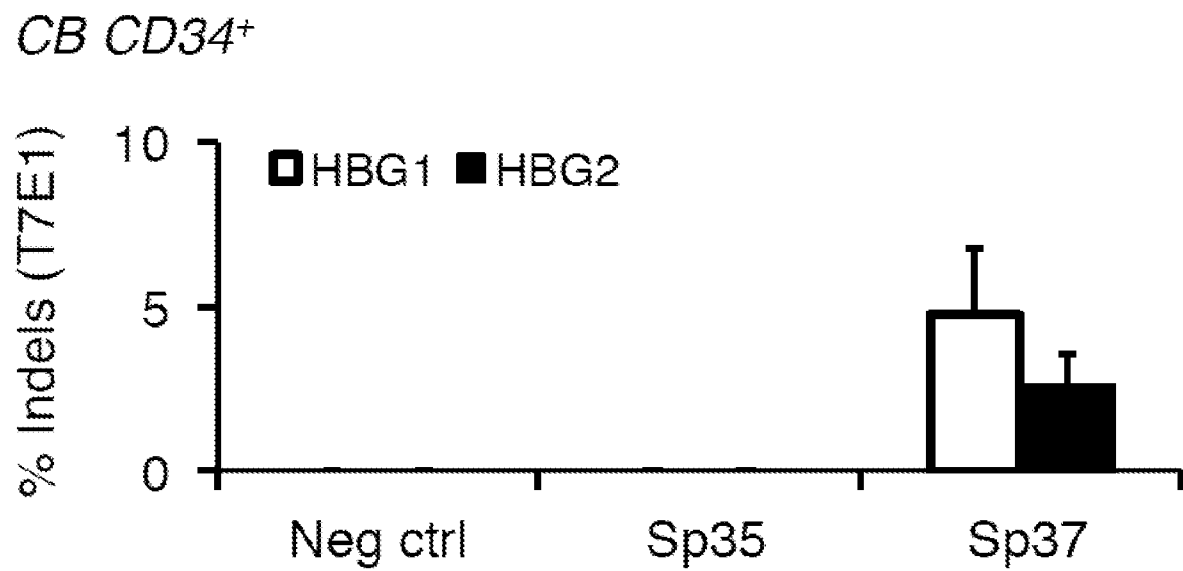


FIG 4B

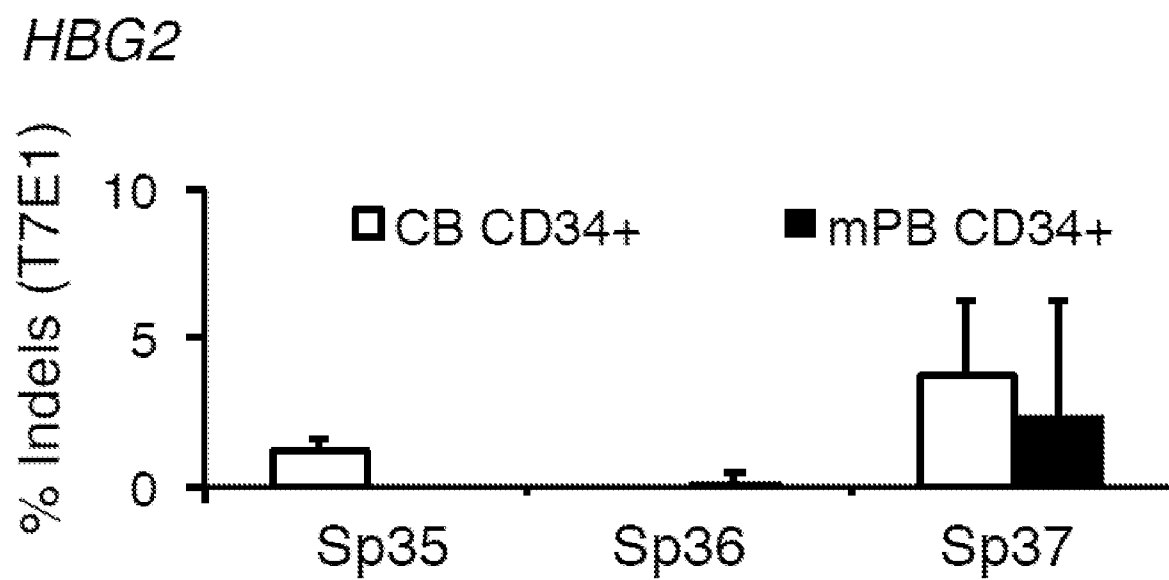


FIG 4C

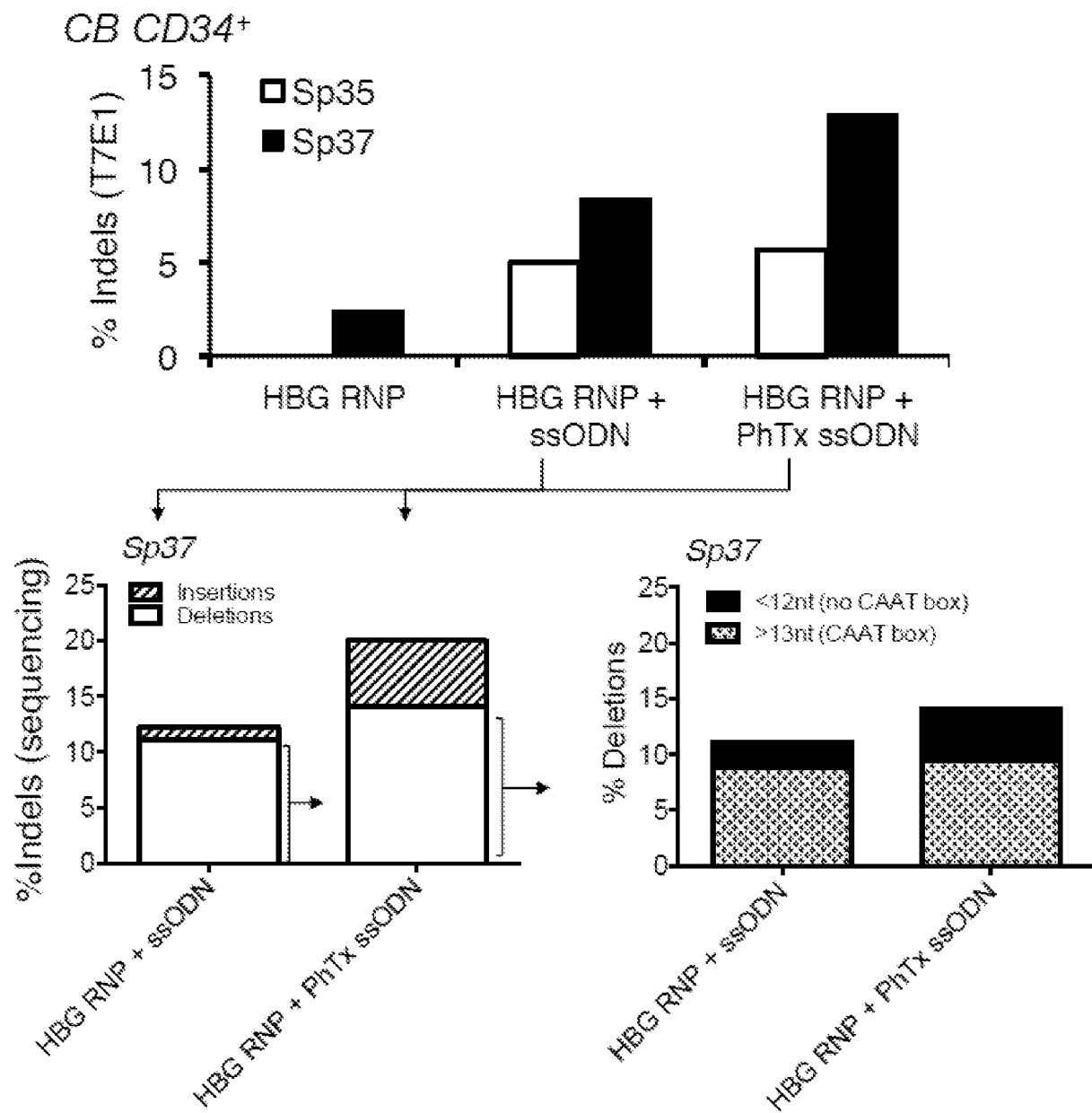


FIG 5A

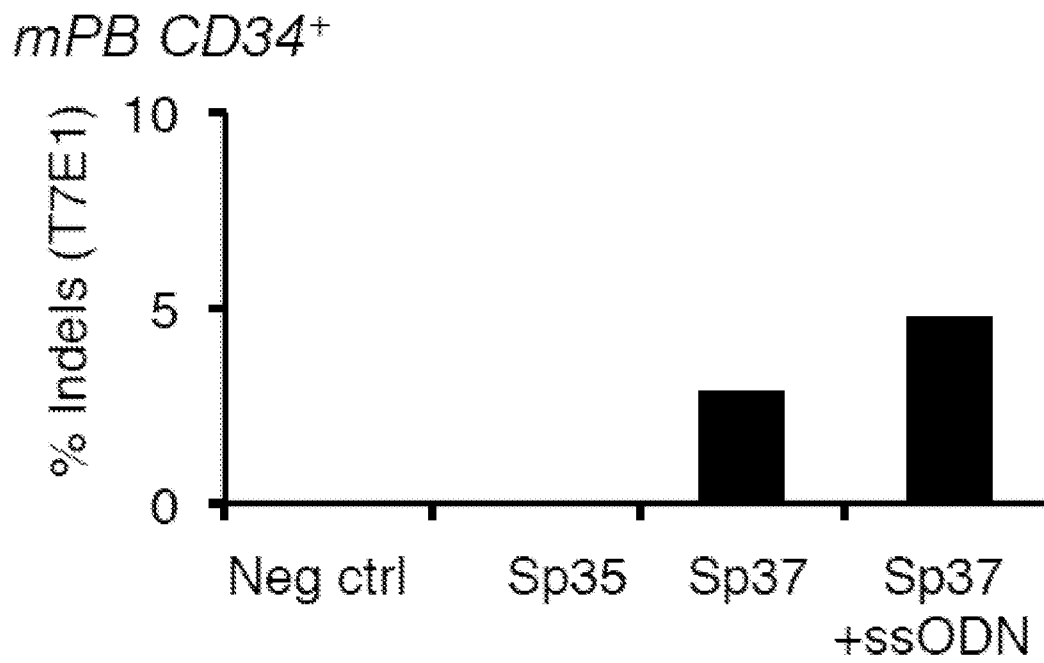


FIG 5B

mPB CD34⁺ erythroblasts

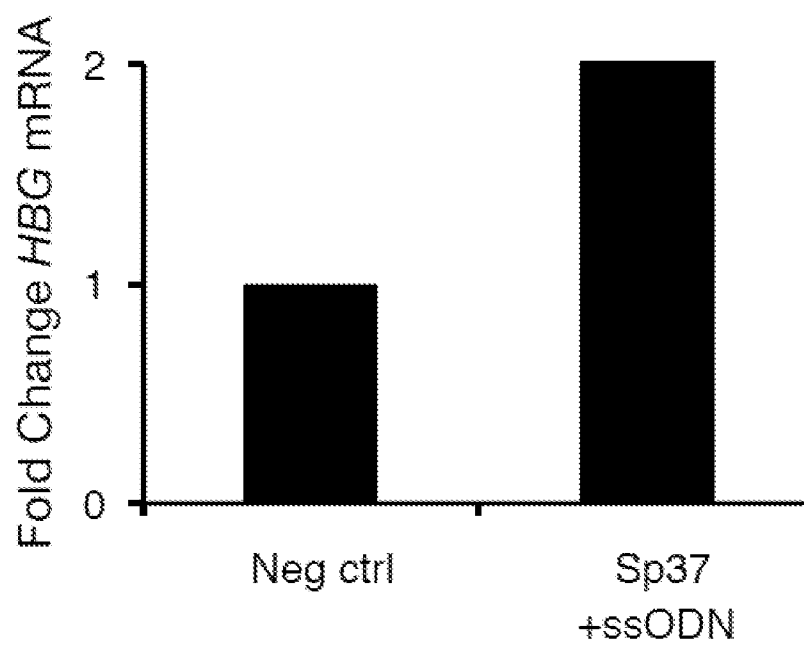


FIG 6A

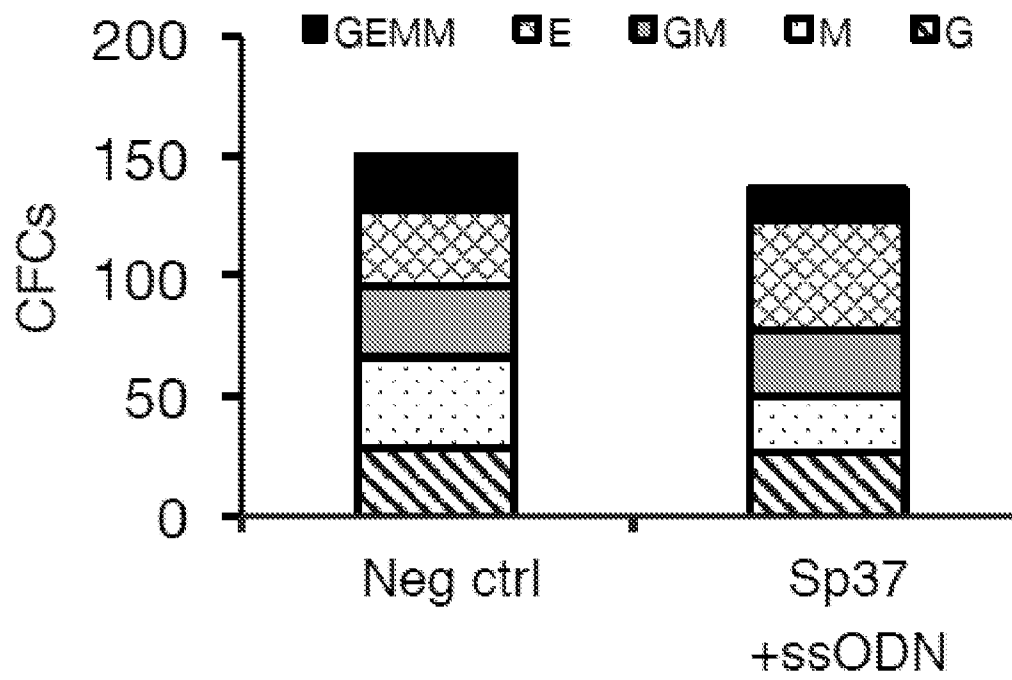


FIG 6B

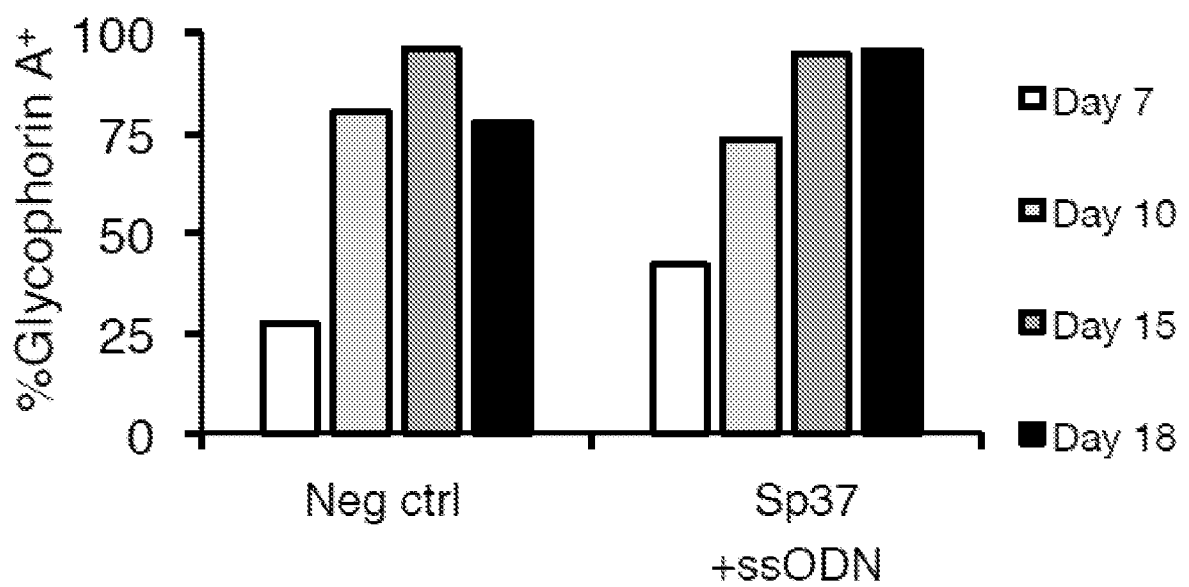


FIG 7A

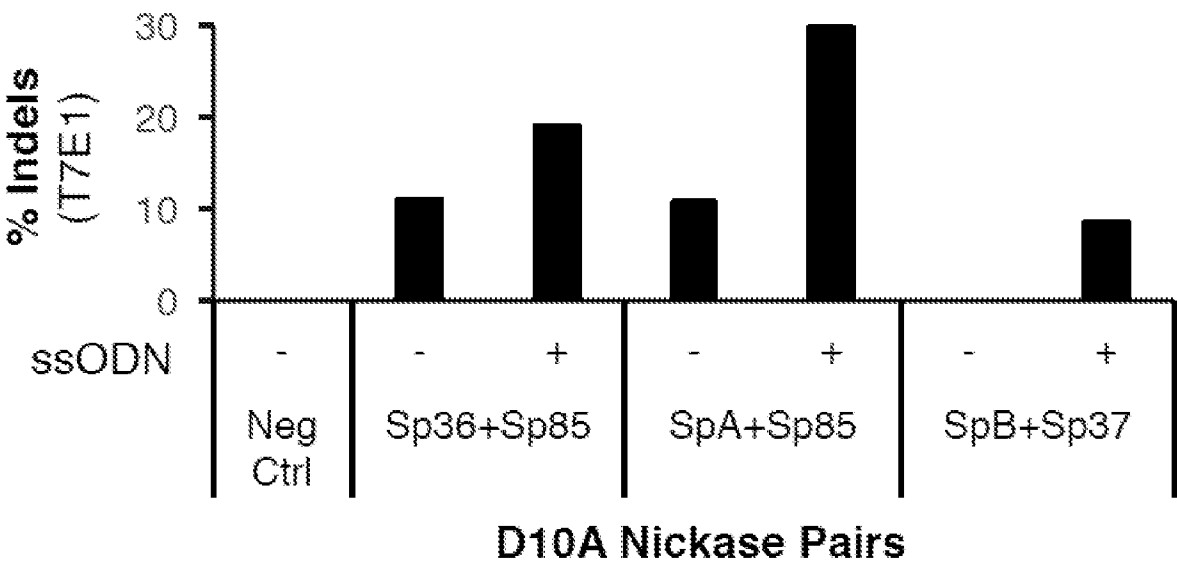


FIG 7B

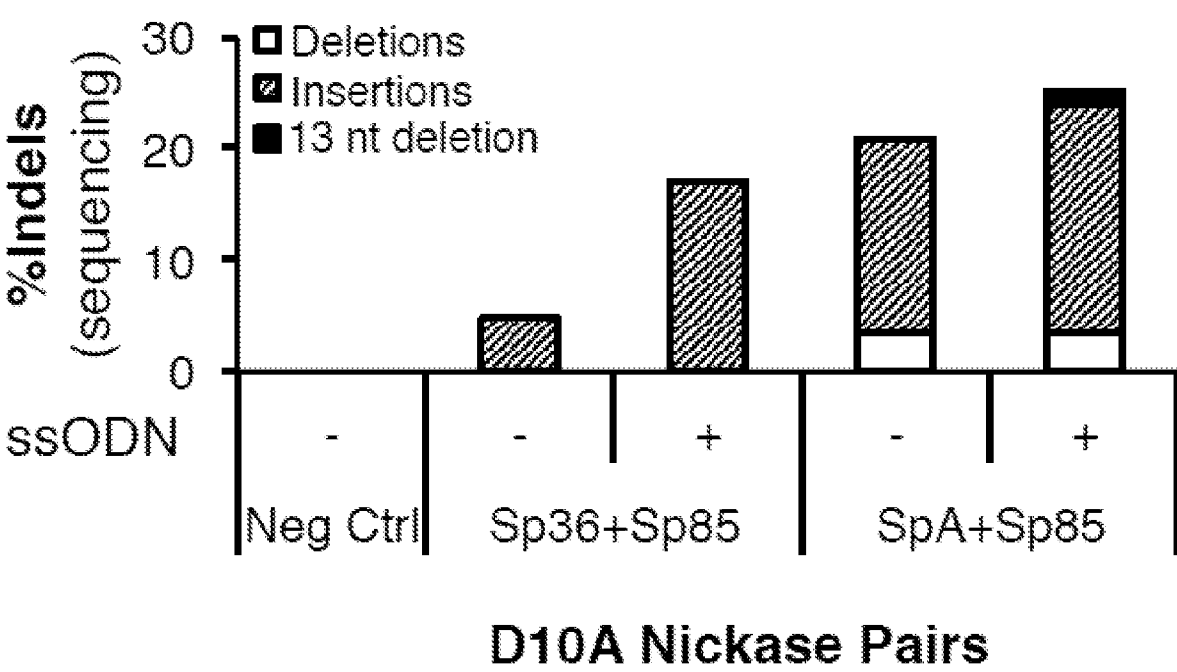


FIG 8A

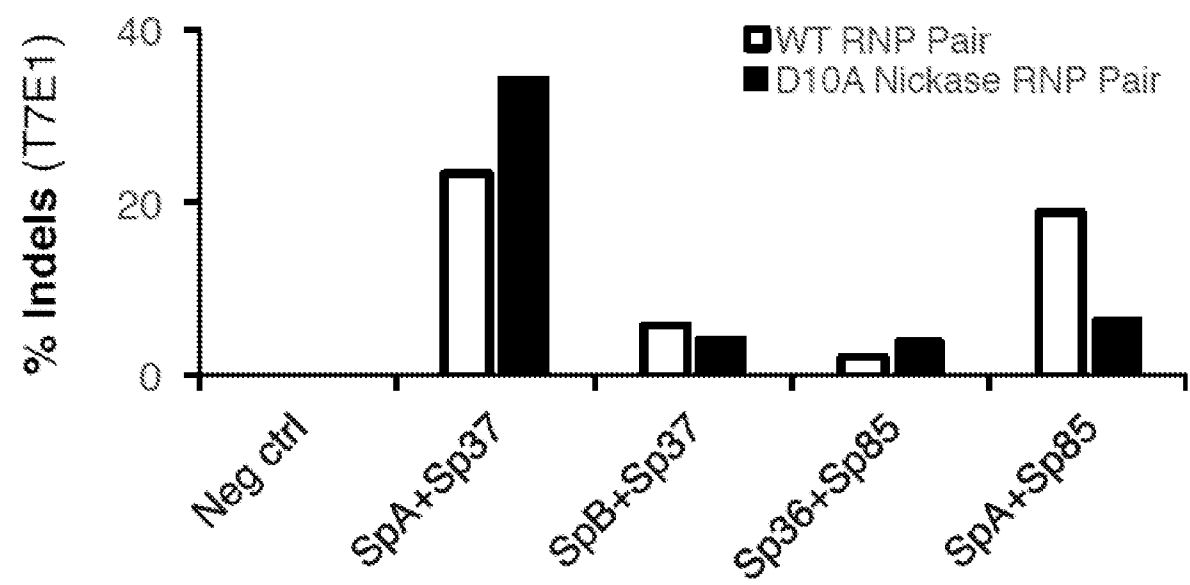


FIG 8B

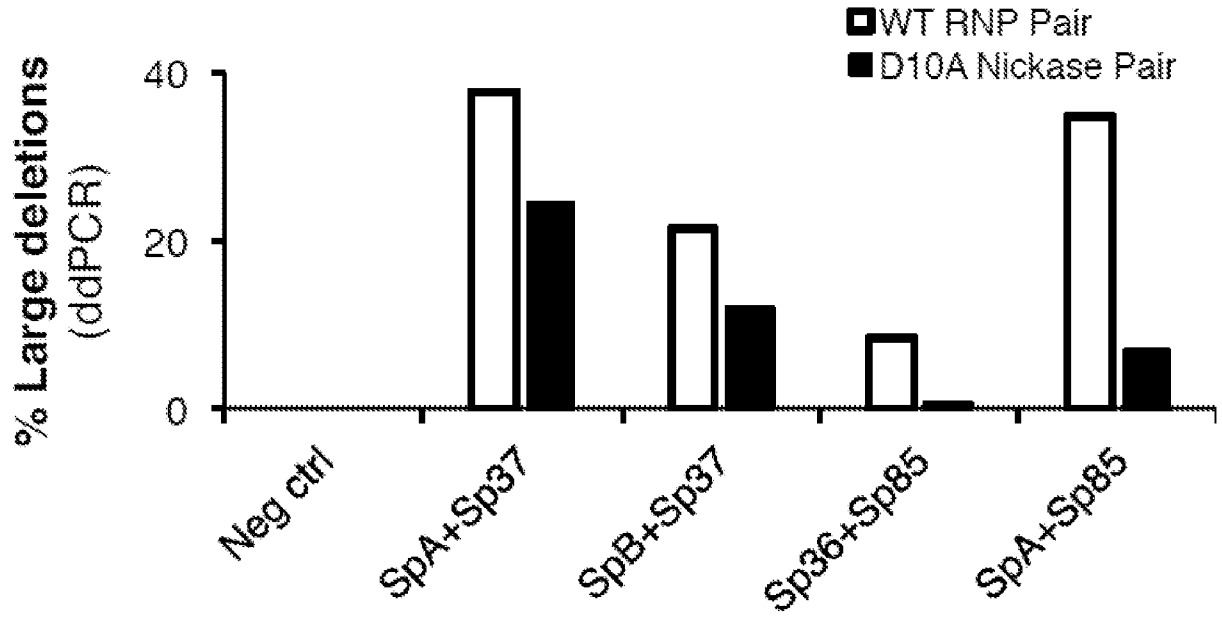


FIG 8C

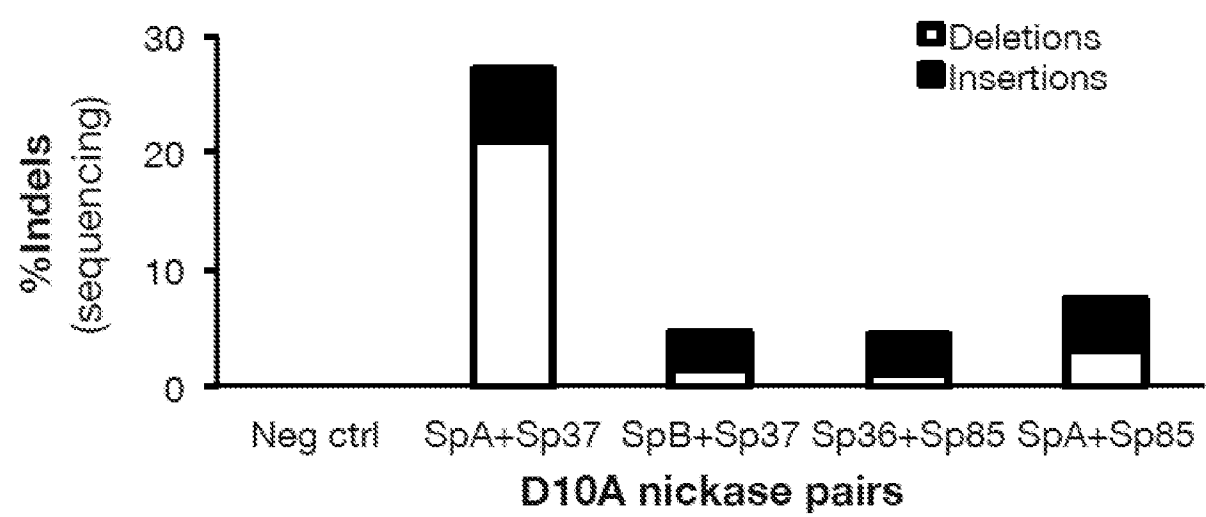


FIG 8D

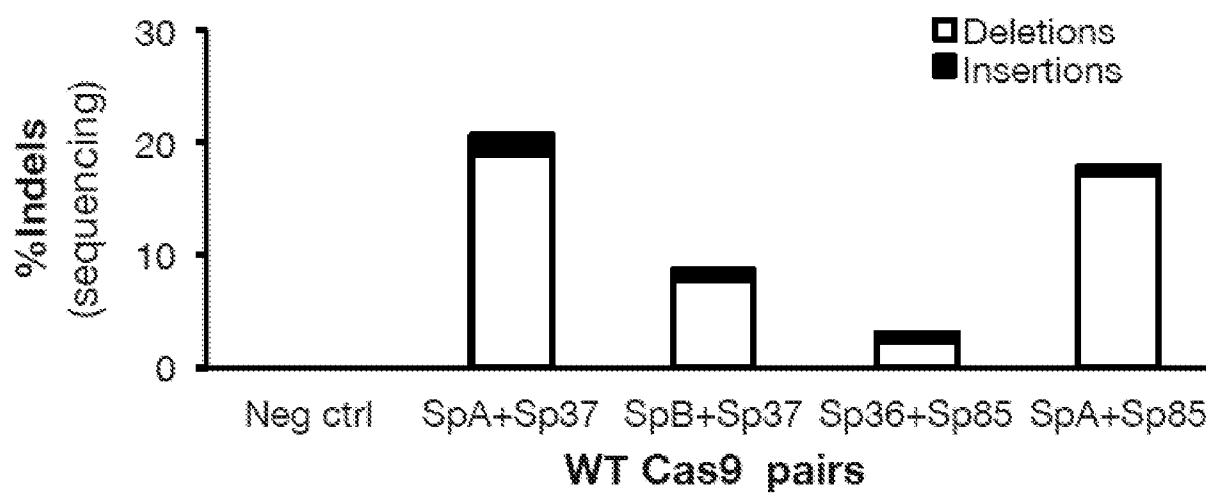


FIG 9

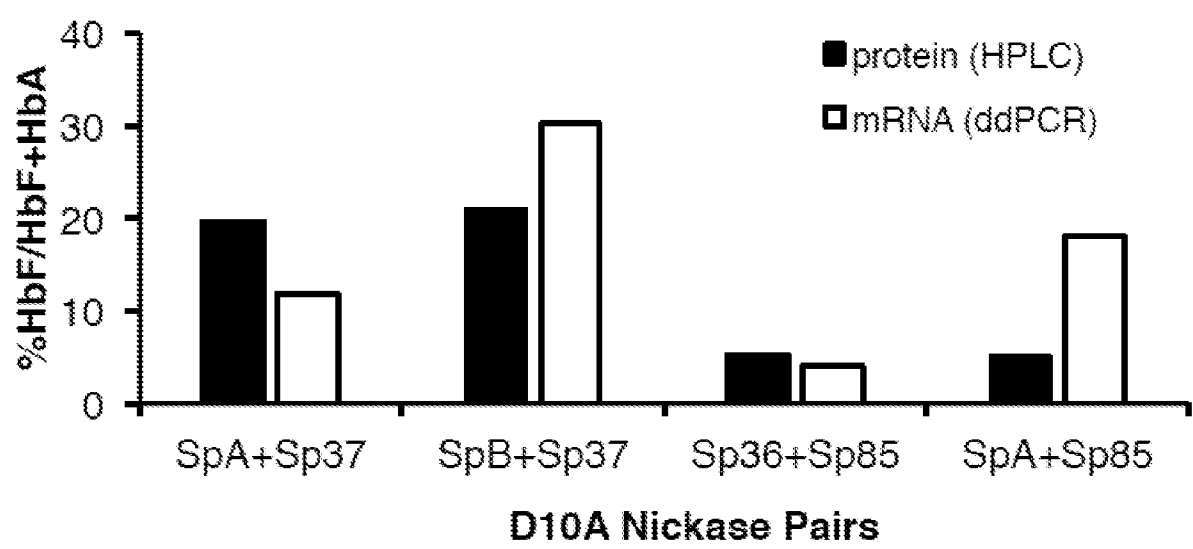


FIG 10A

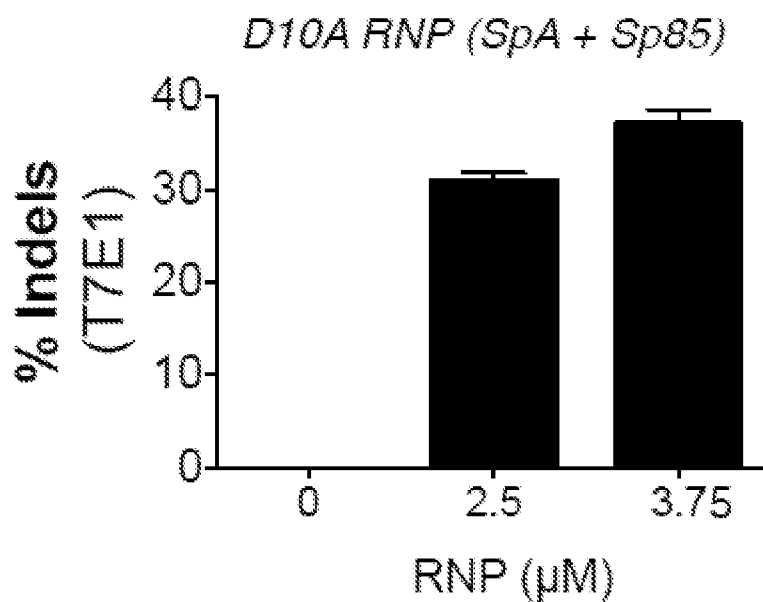


FIG 10B

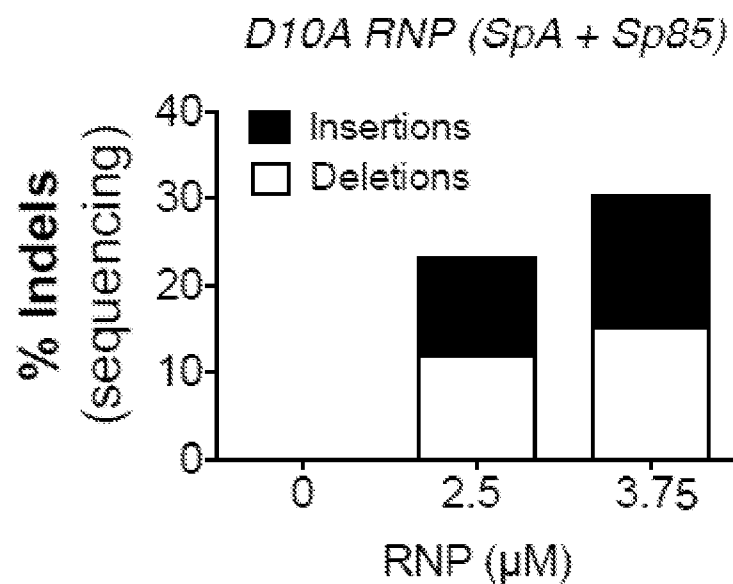


FIG 10C

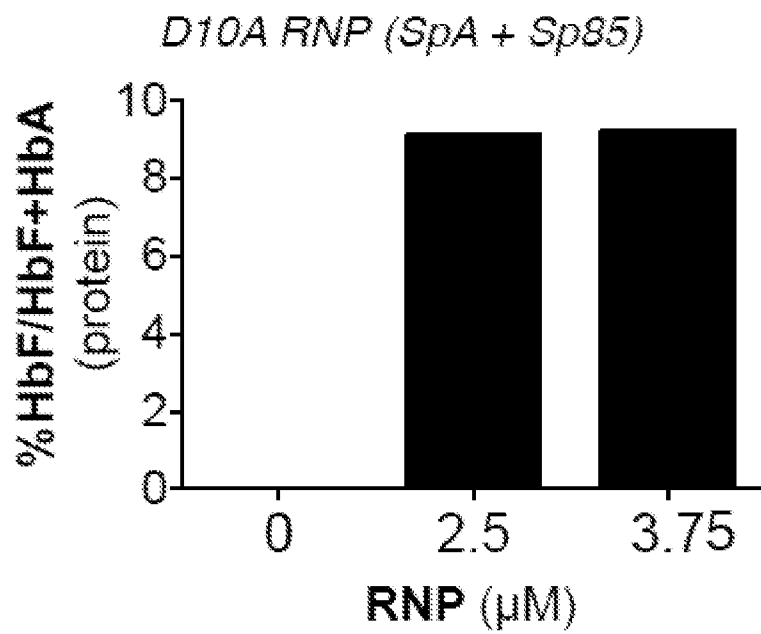


FIG 10D

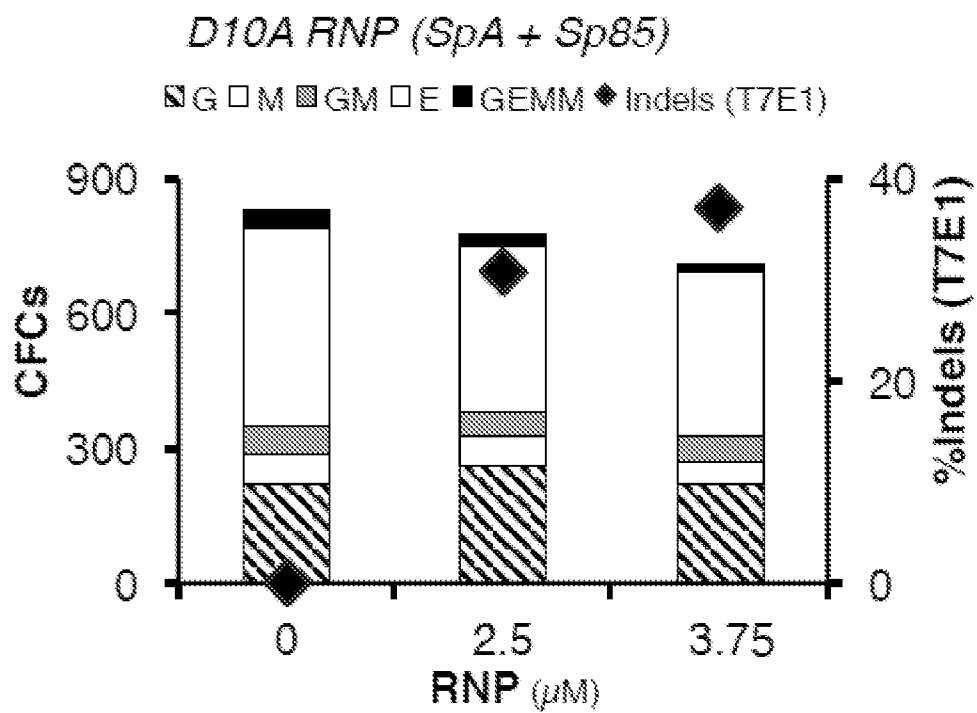


FIG 10E

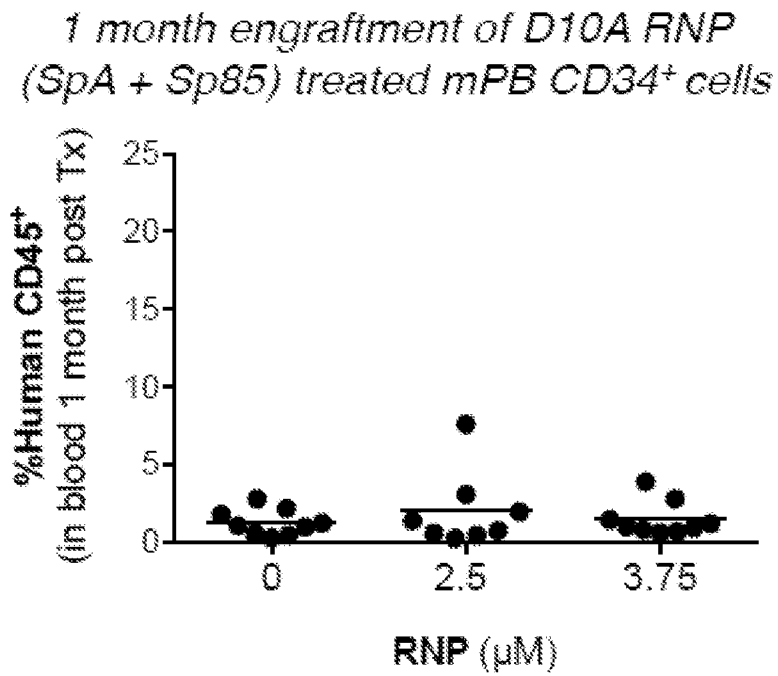


FIG 10F

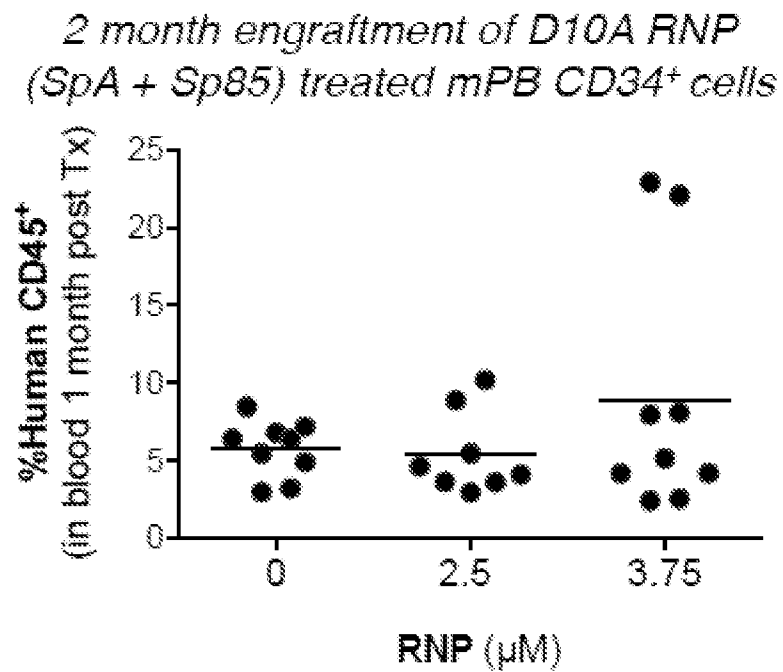


FIG 11A

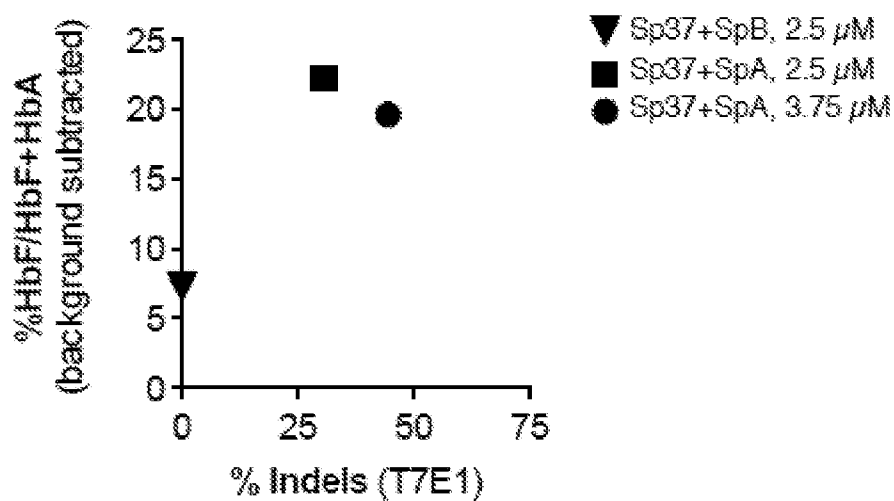


FIG 11B

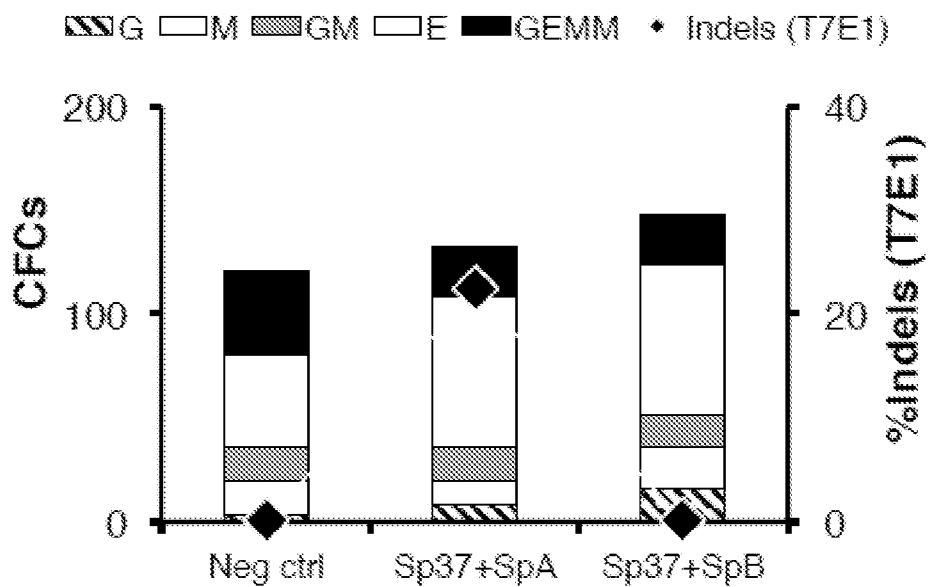


FIG 11C

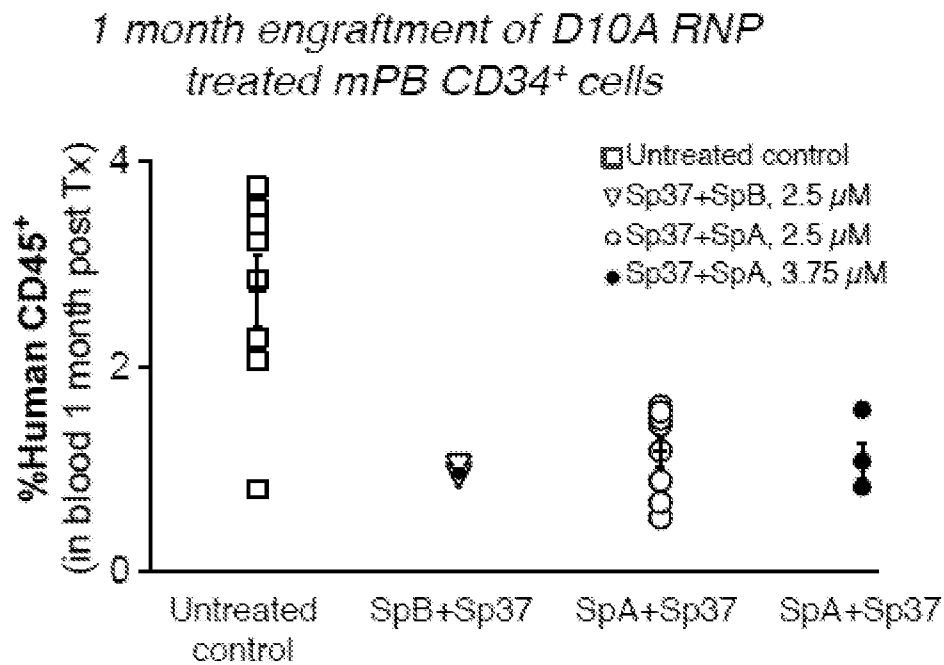


FIG 11D

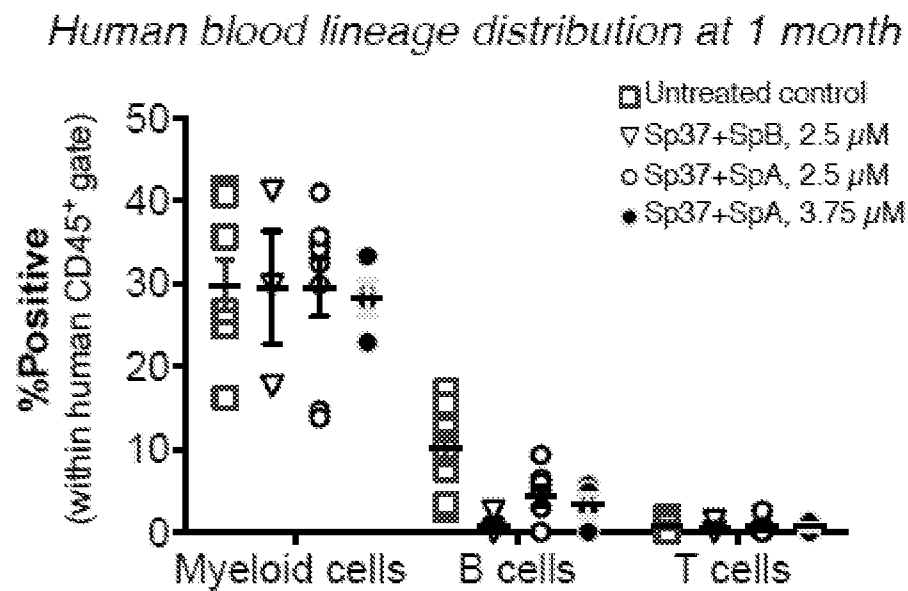


FIG 12

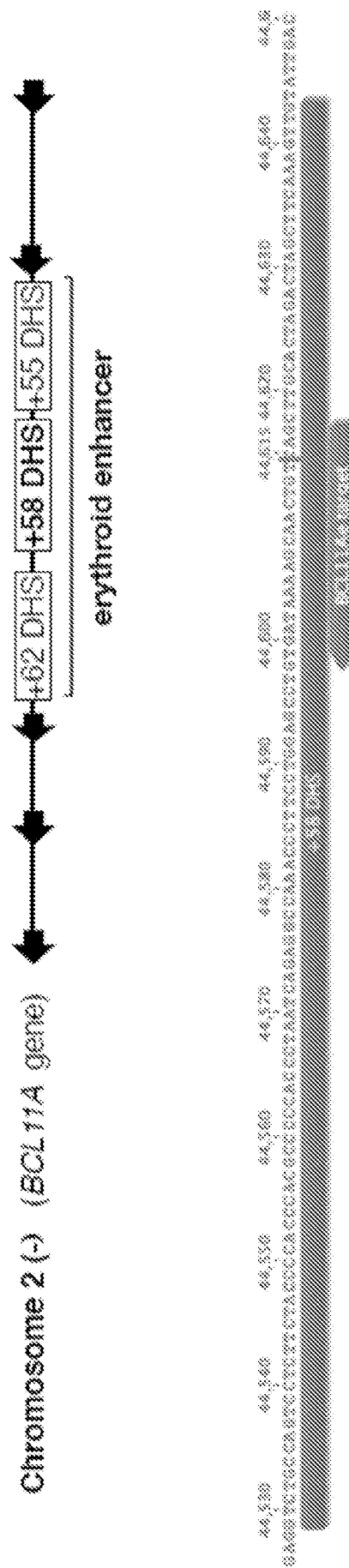


FIG 13A

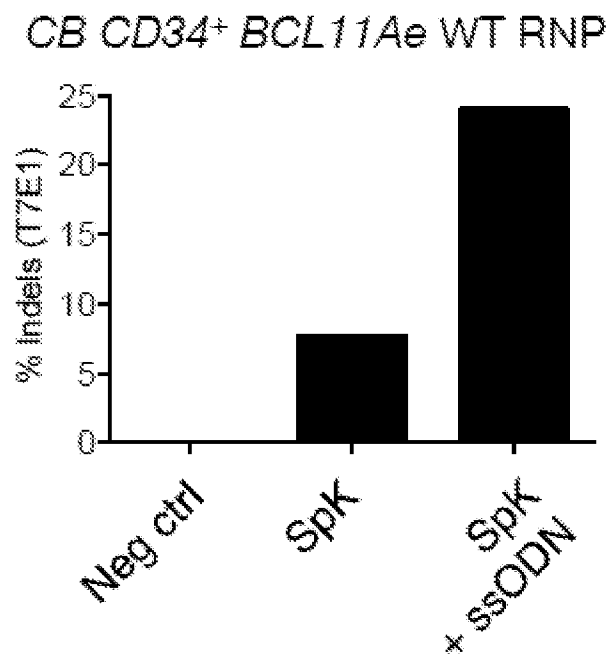


FIG 13B

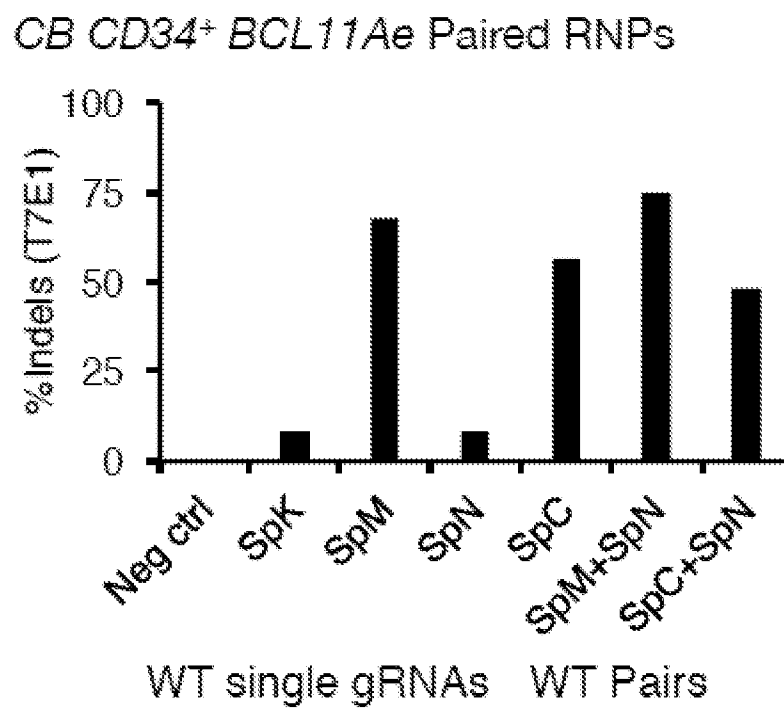


FIG 14A

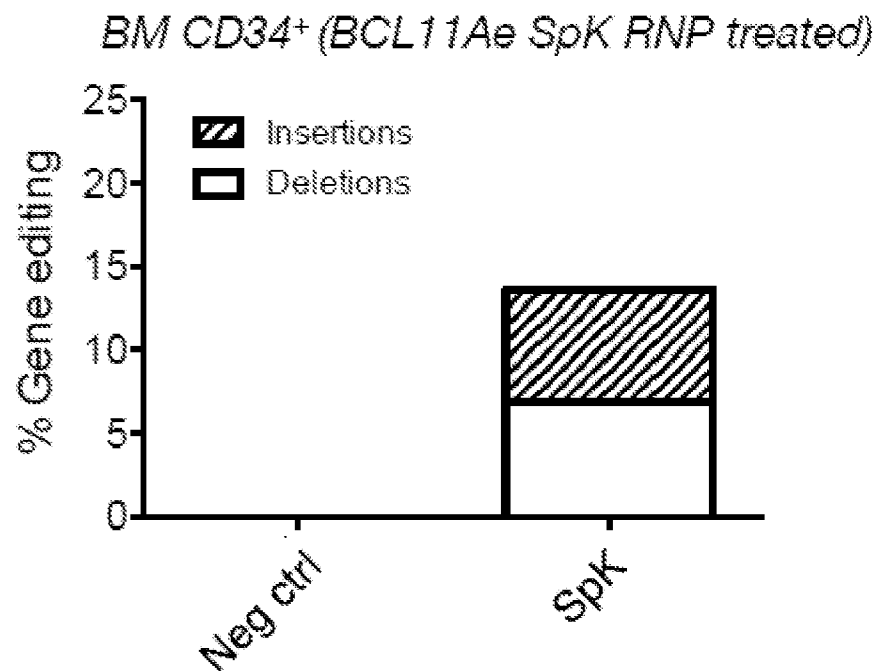


FIG 14B

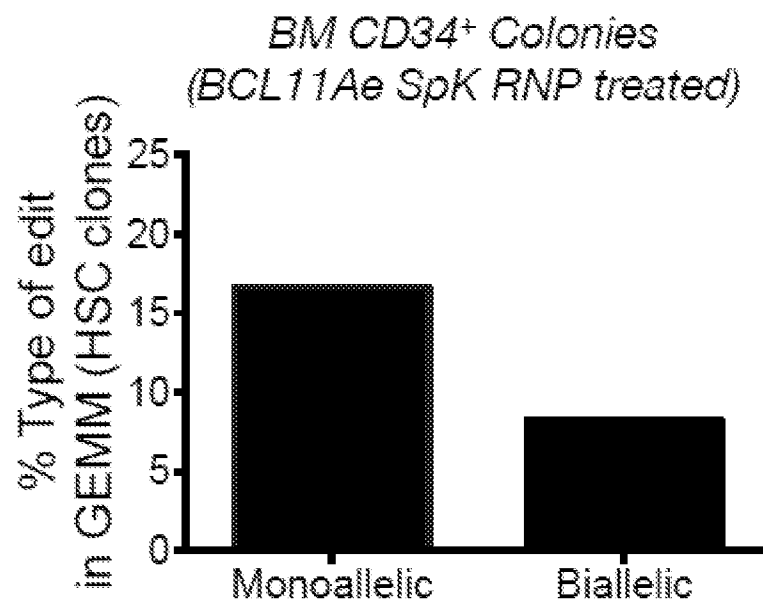


FIG 14C

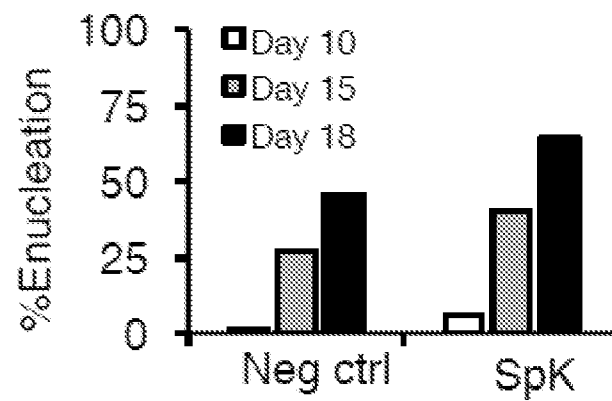


FIG 14D

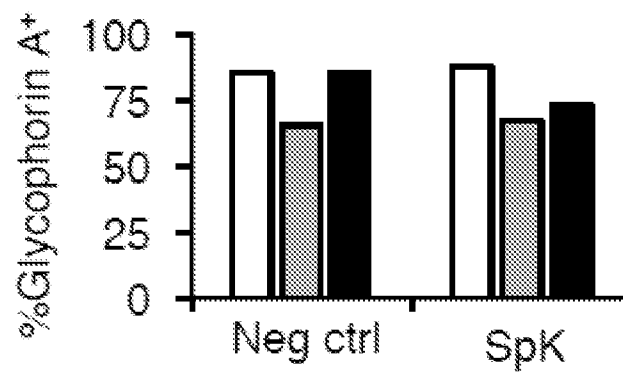


FIG 14E

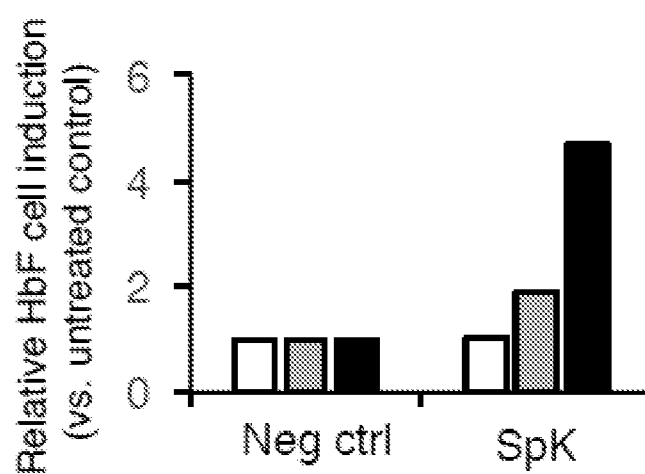


FIG 15A

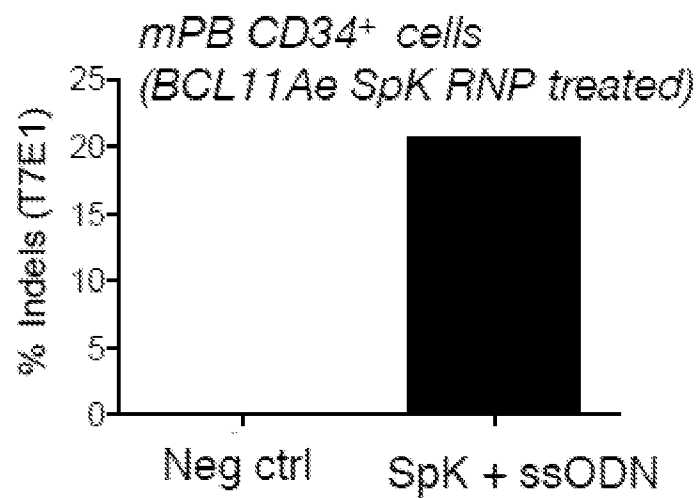


FIG 15B

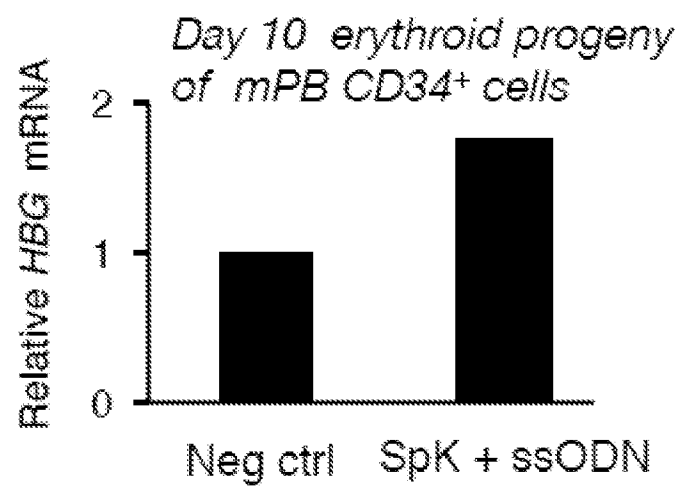


FIG 15C

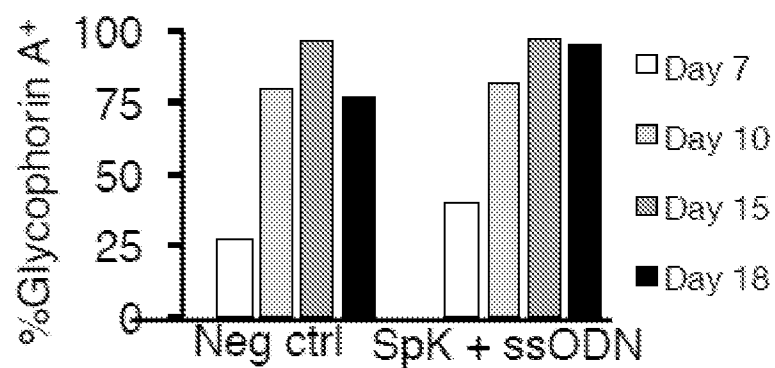


FIG 16

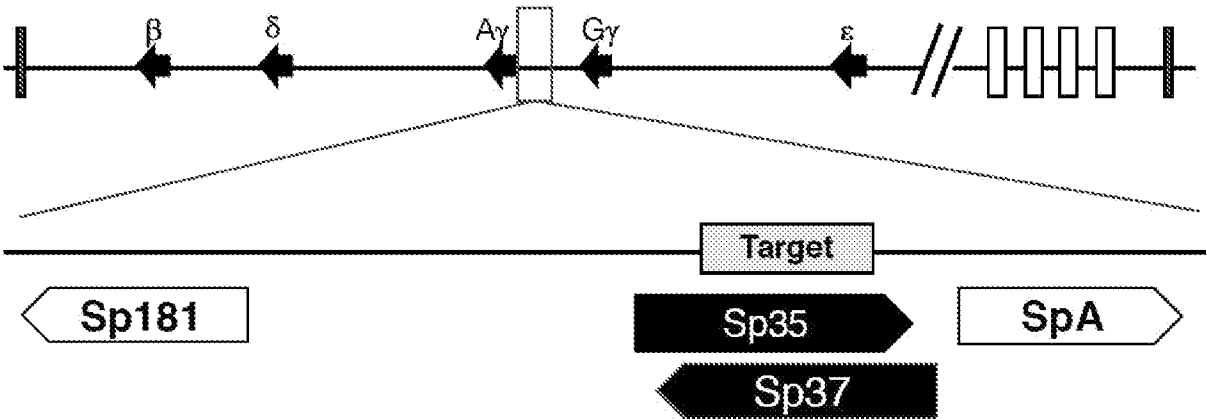


FIG 17

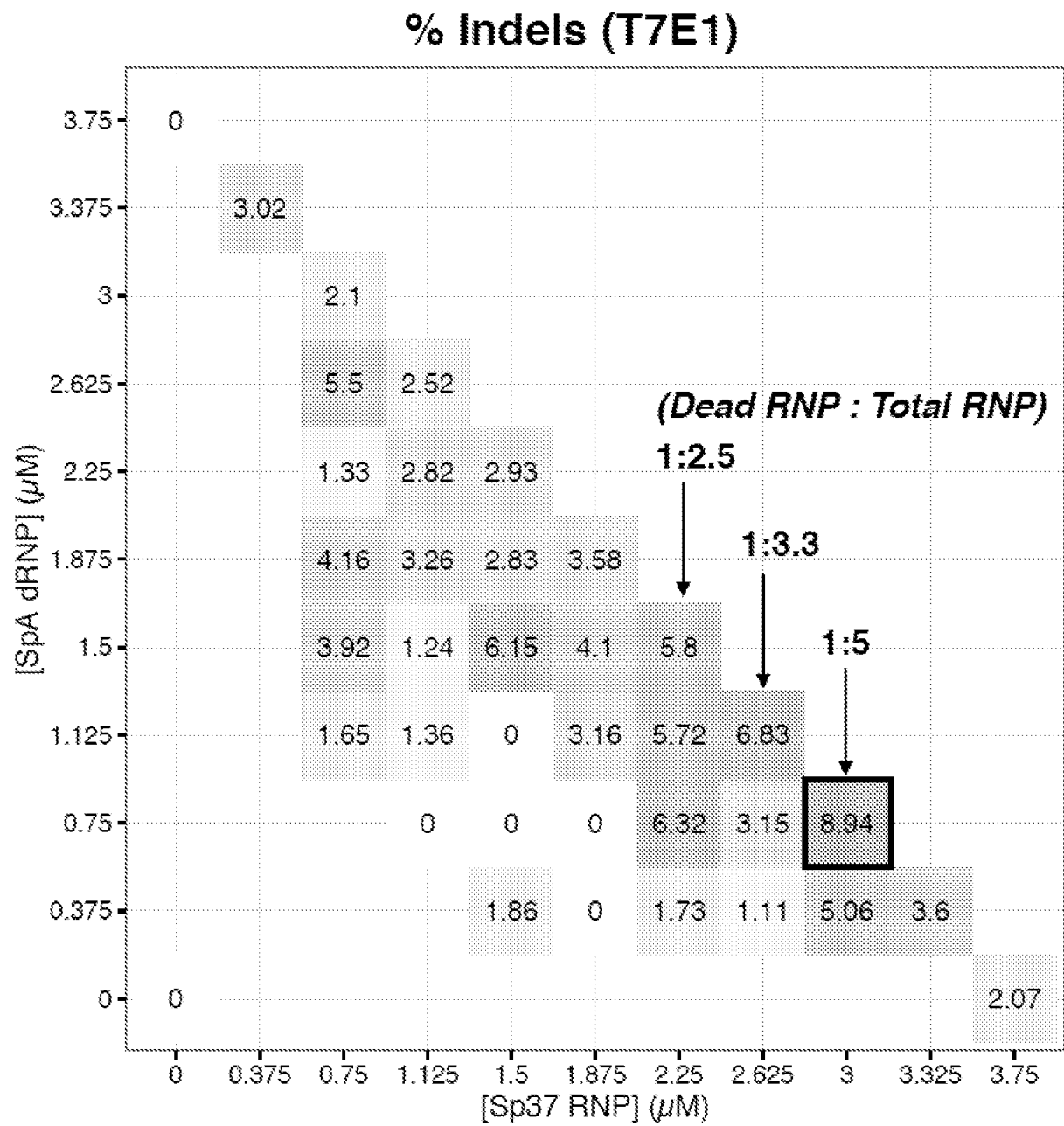


FIG 18

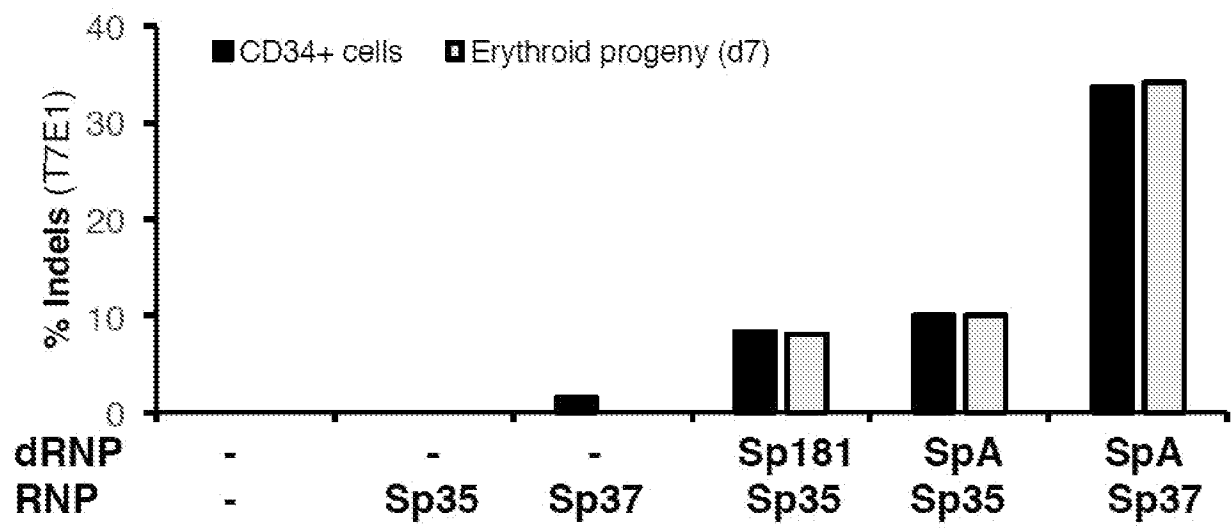


FIG 19

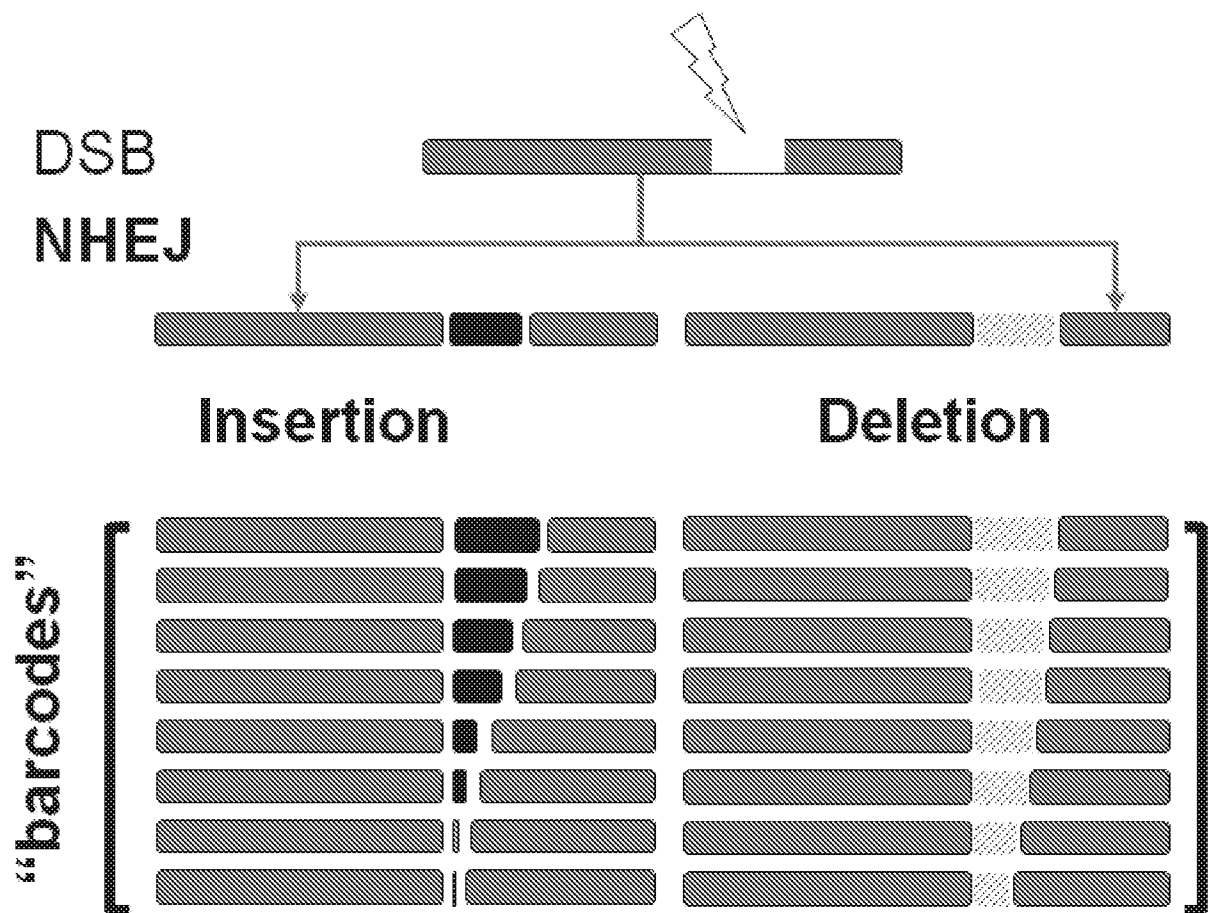


FIG 20

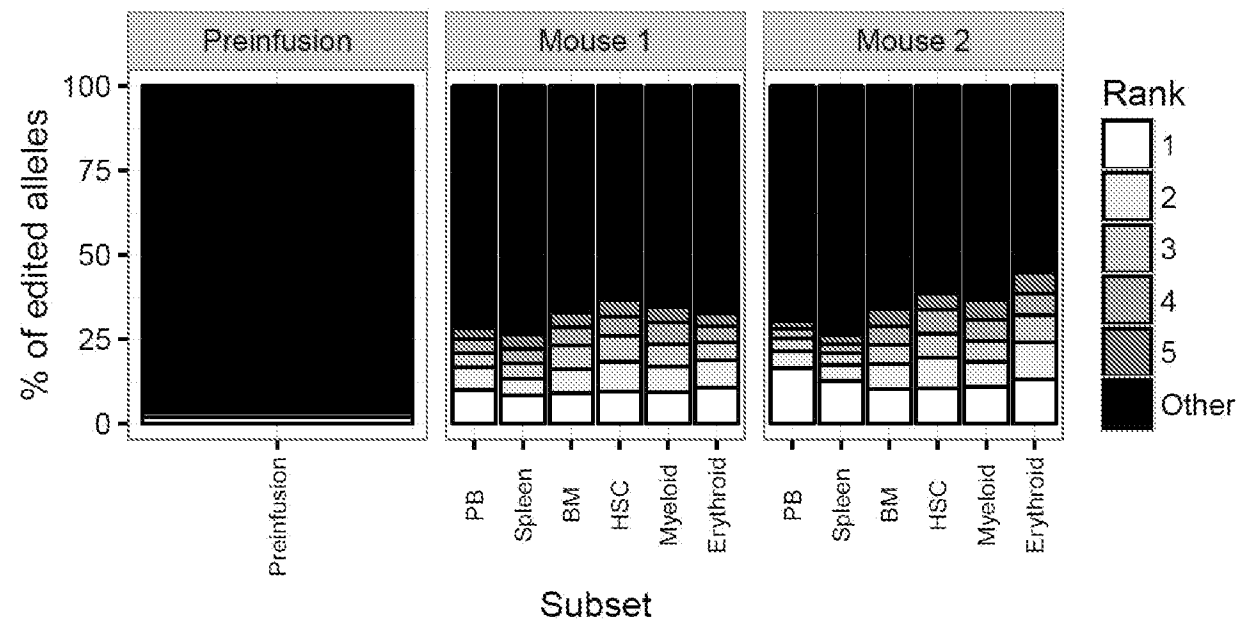


FIG 21

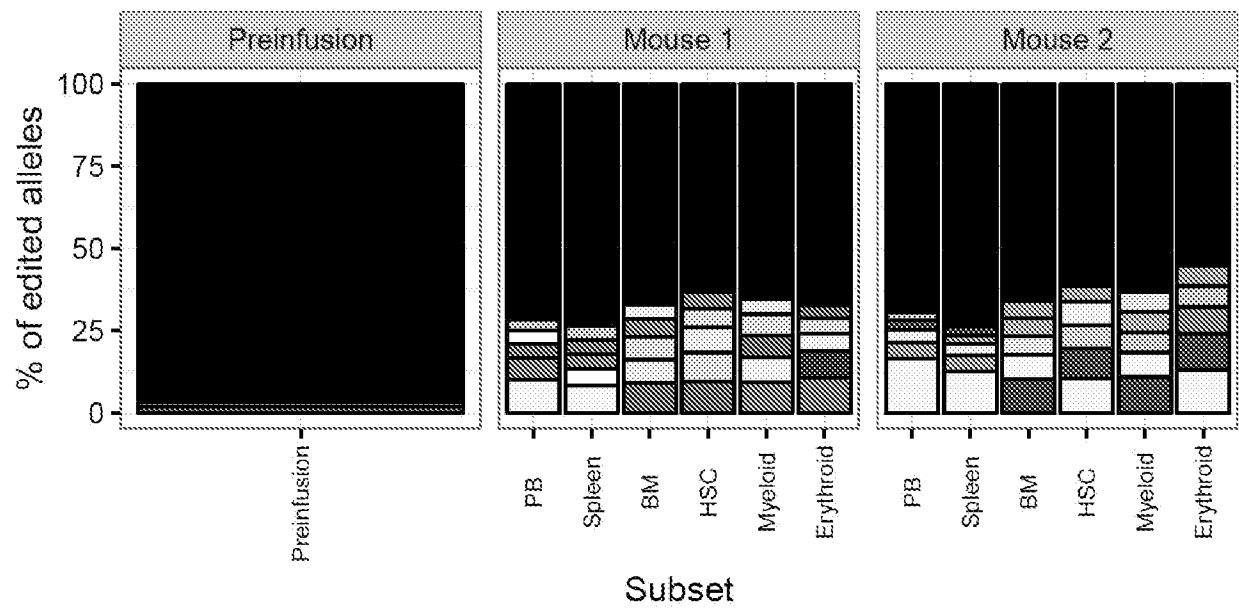


FIG 22

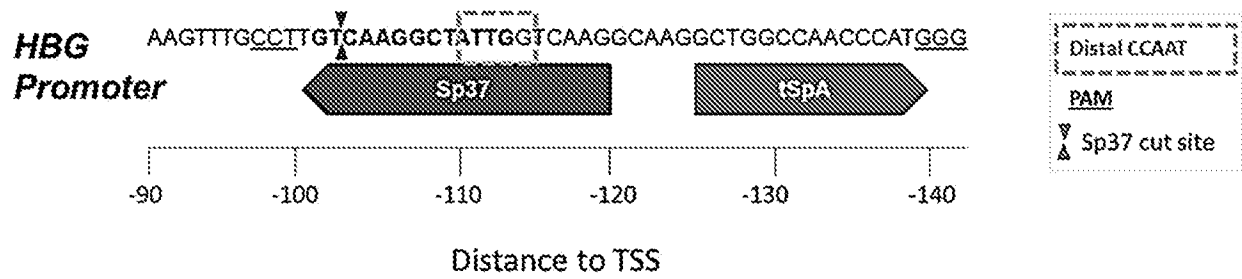


FIG 23

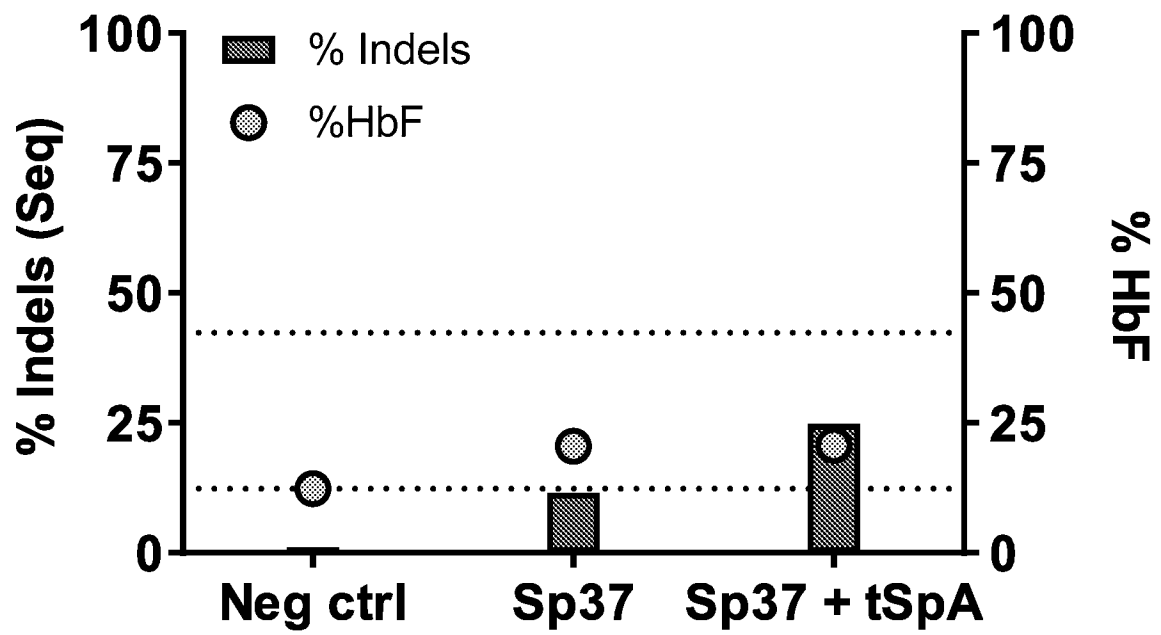


FIG 24

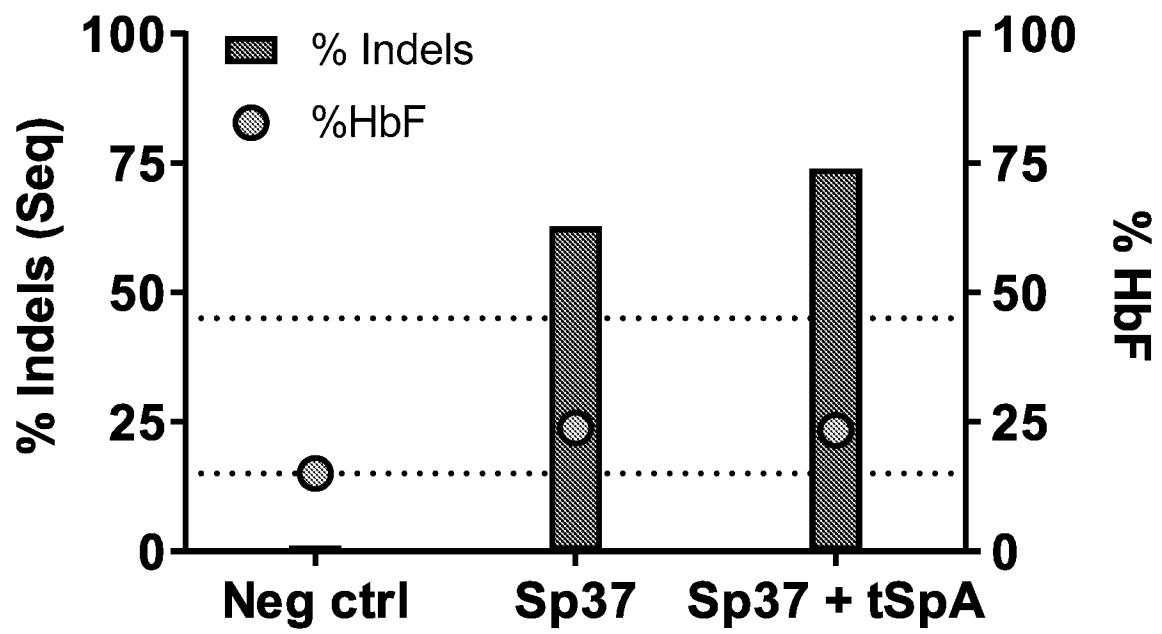


FIG 25

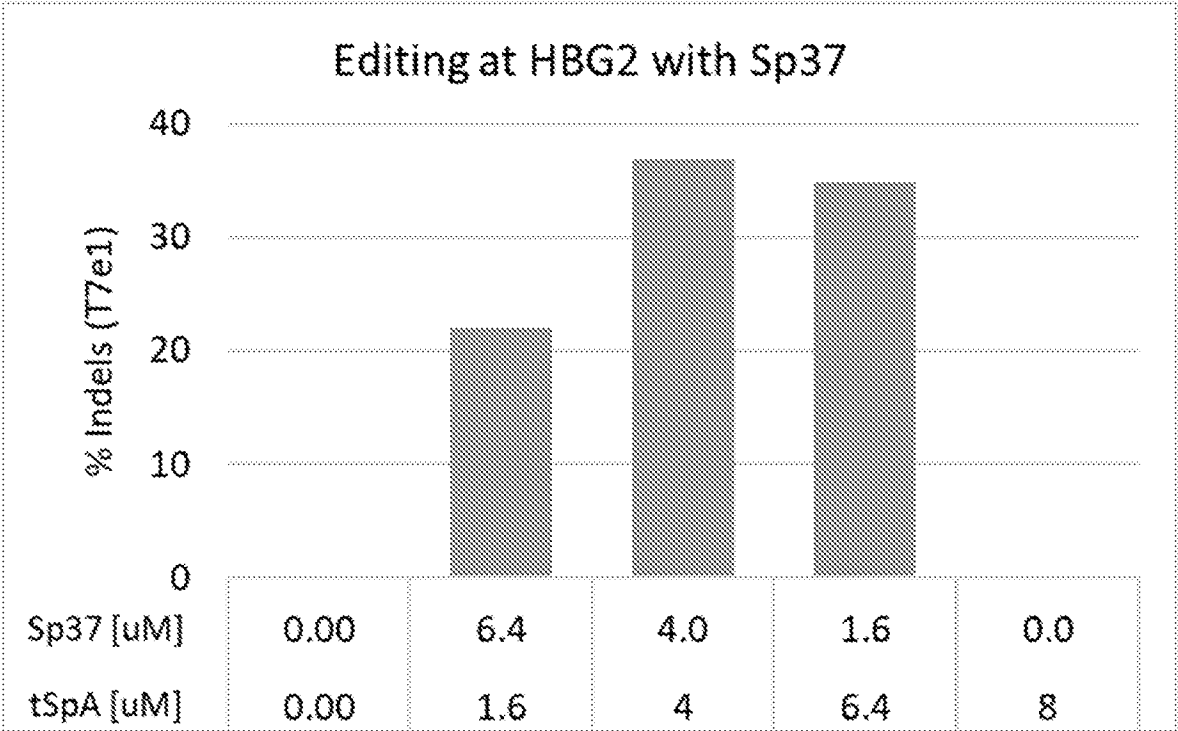


FIG 26A

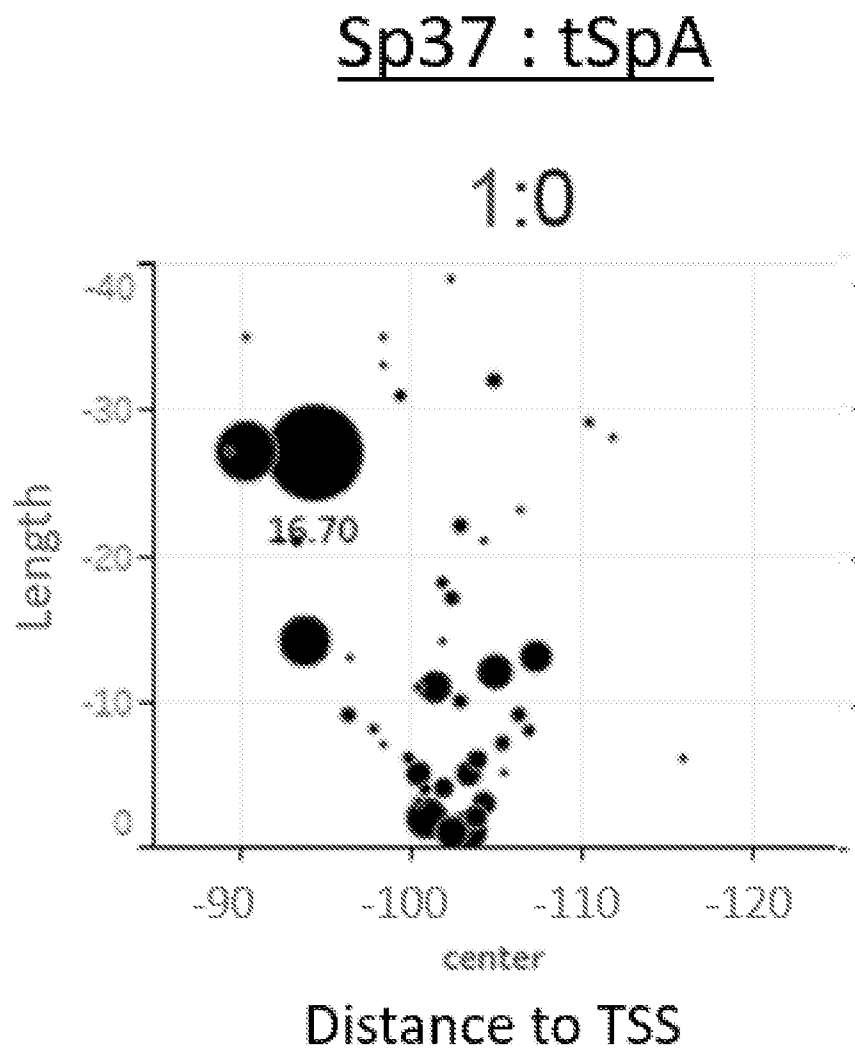


FIG 26B

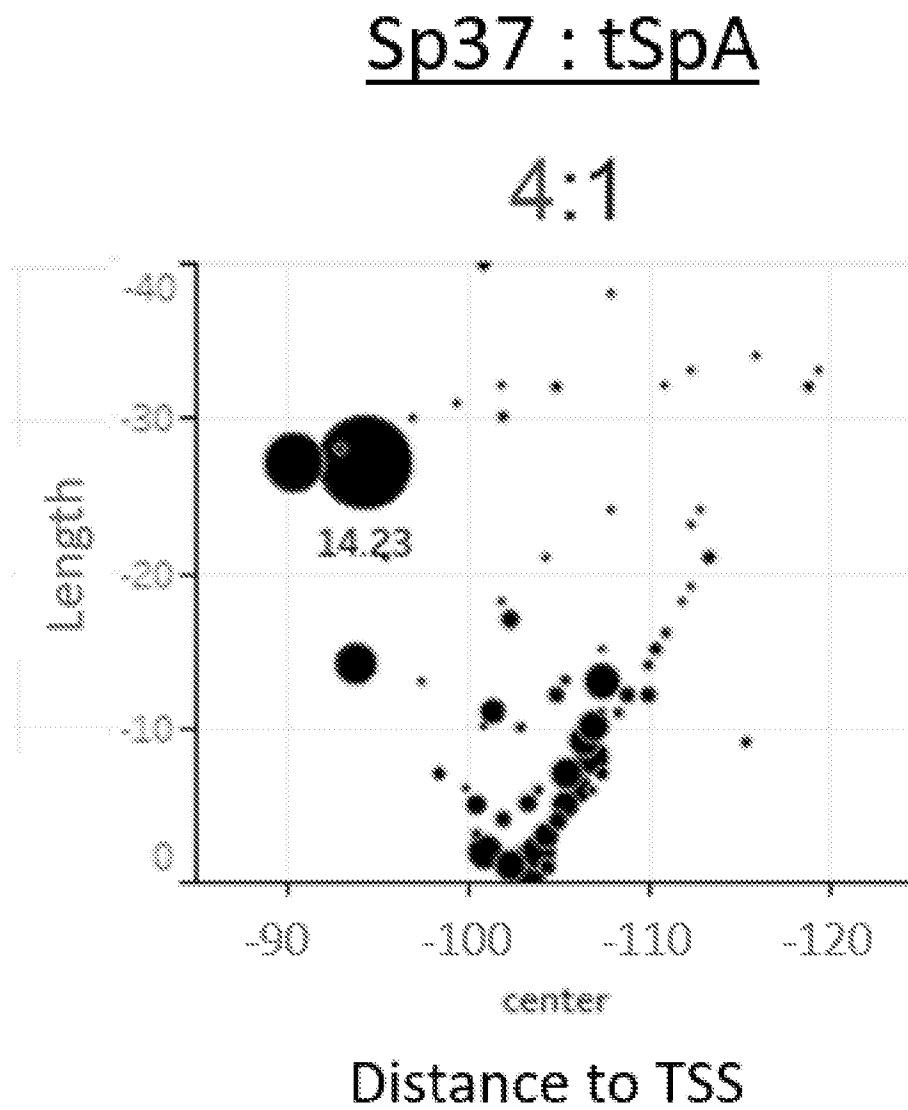


FIG 26C

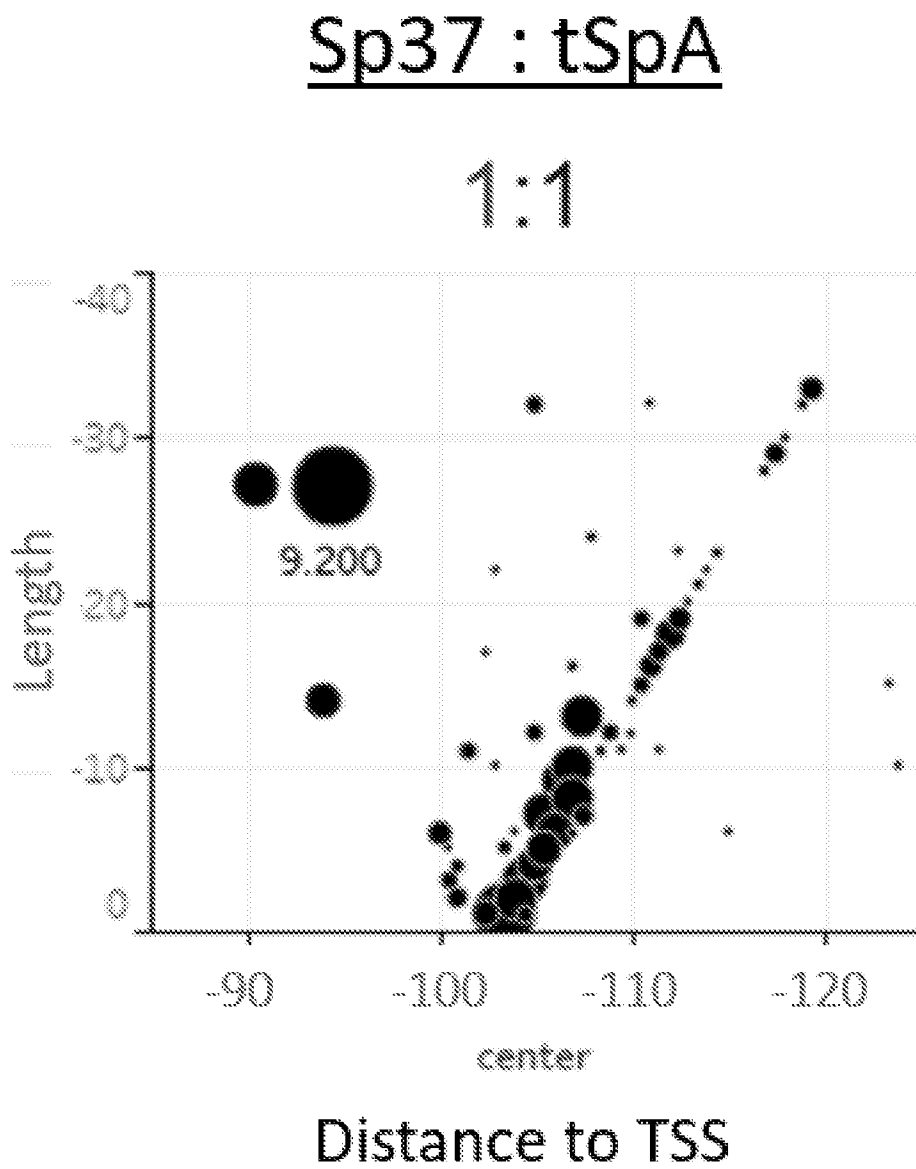


FIG 26D

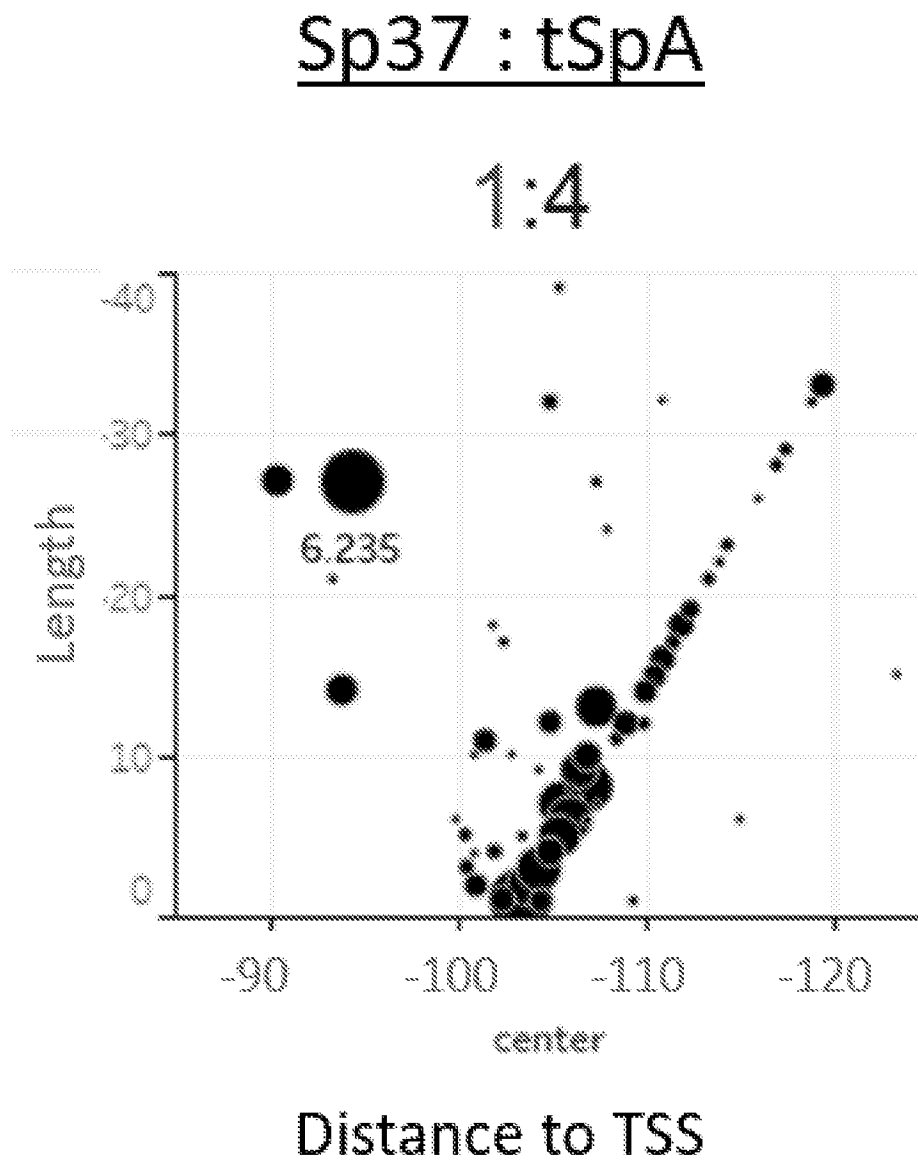


FIG 27A

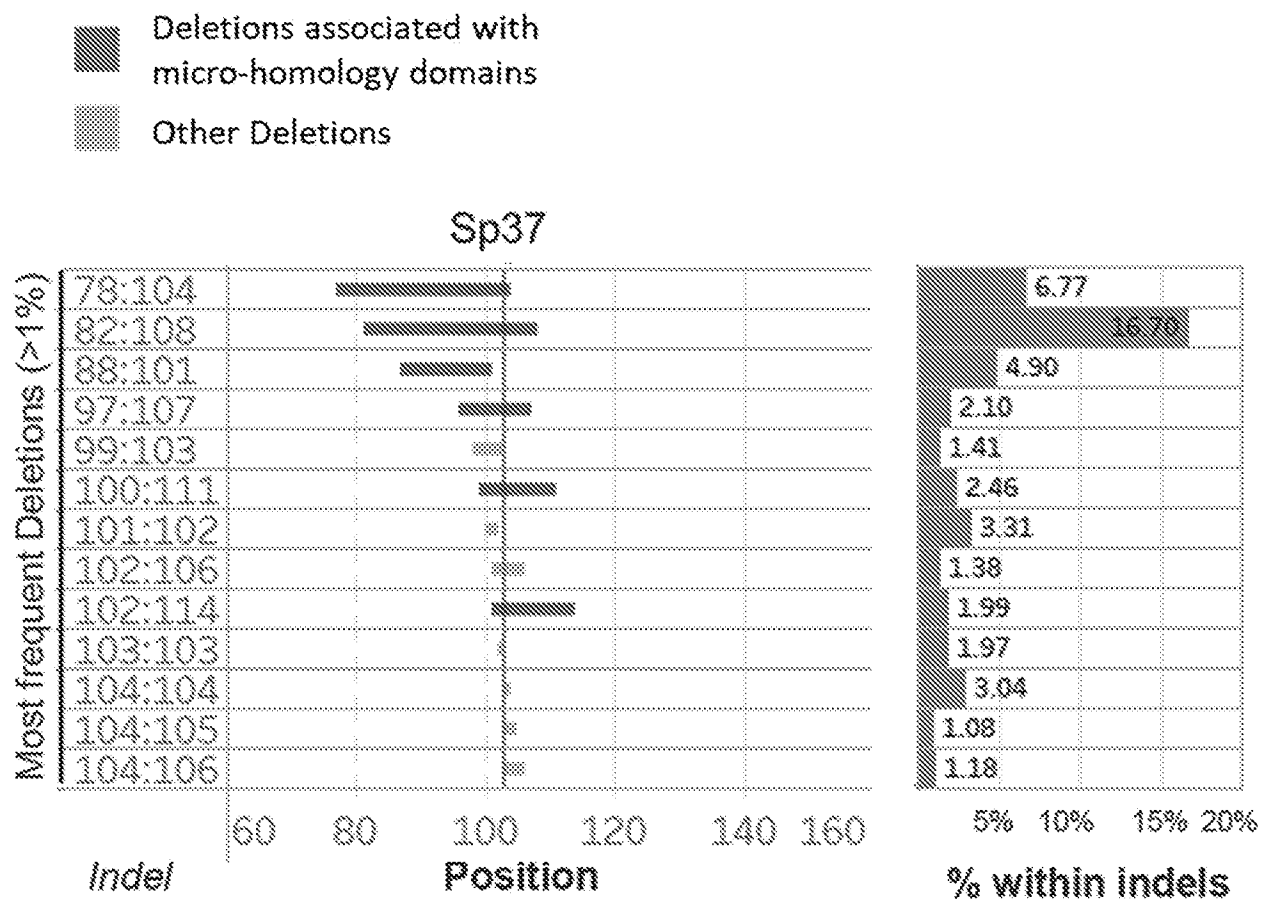


FIG 27B

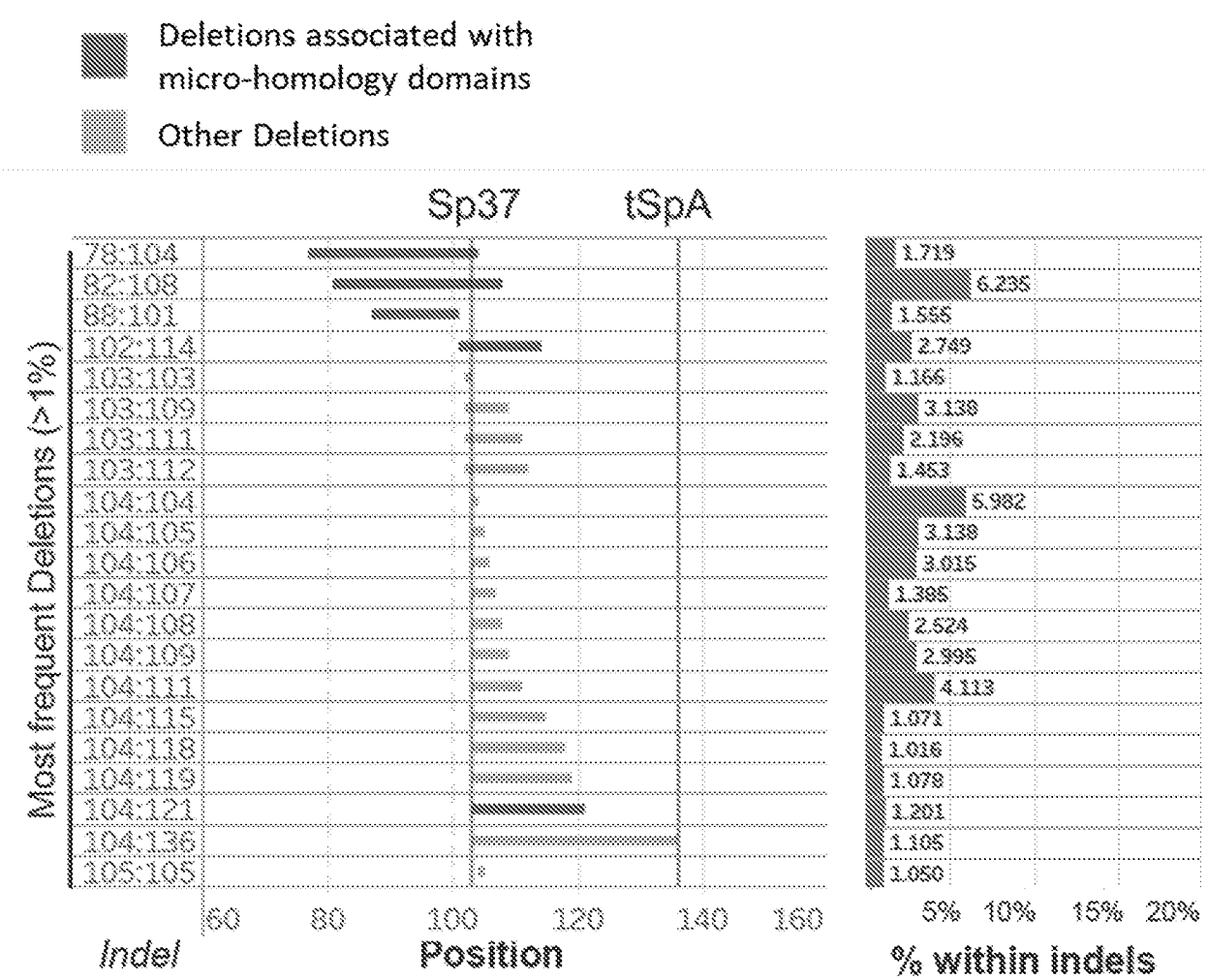


FIG. 28A

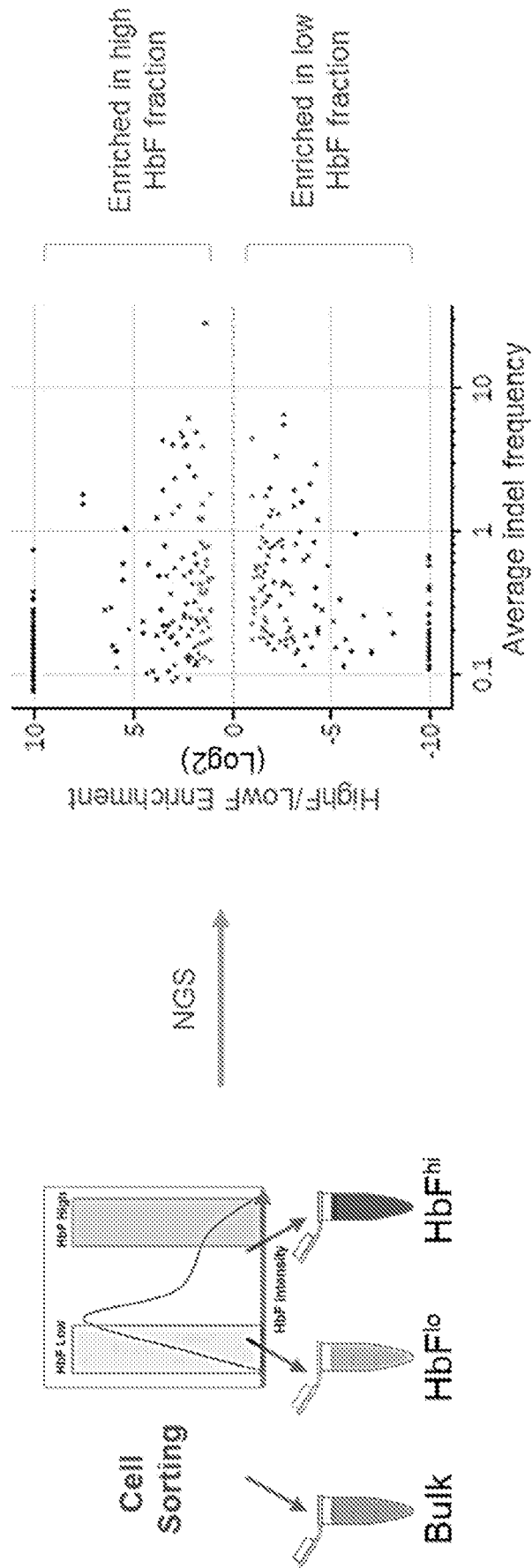


FIG 28B

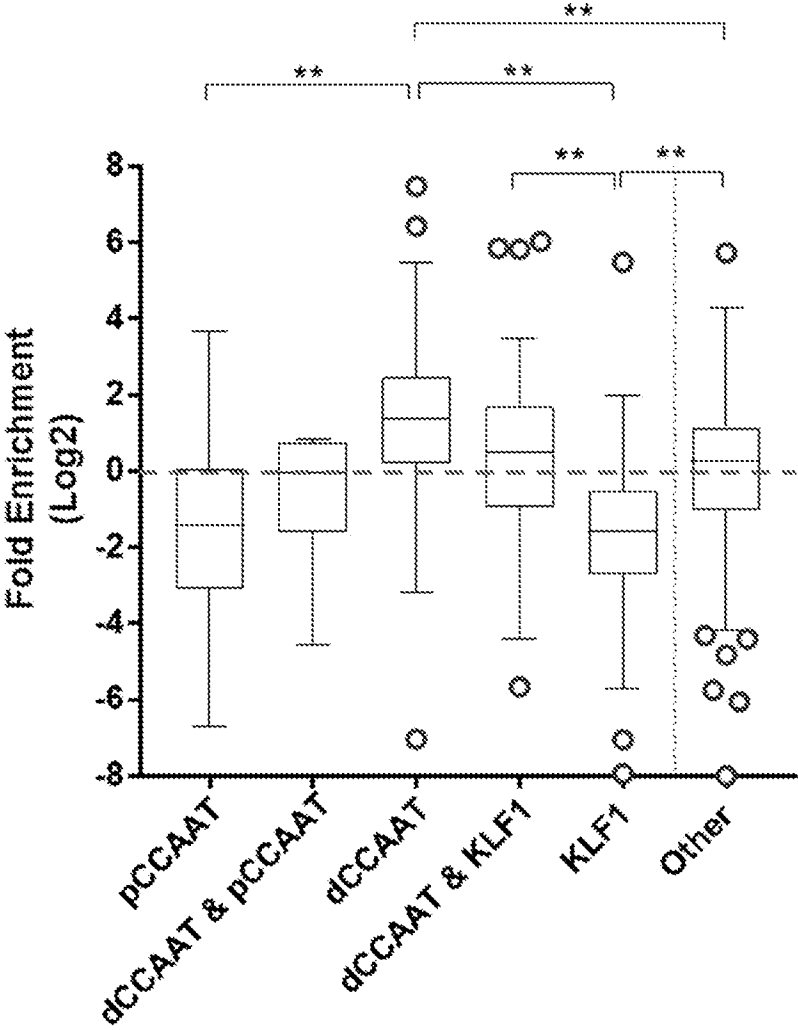


FIG 28C

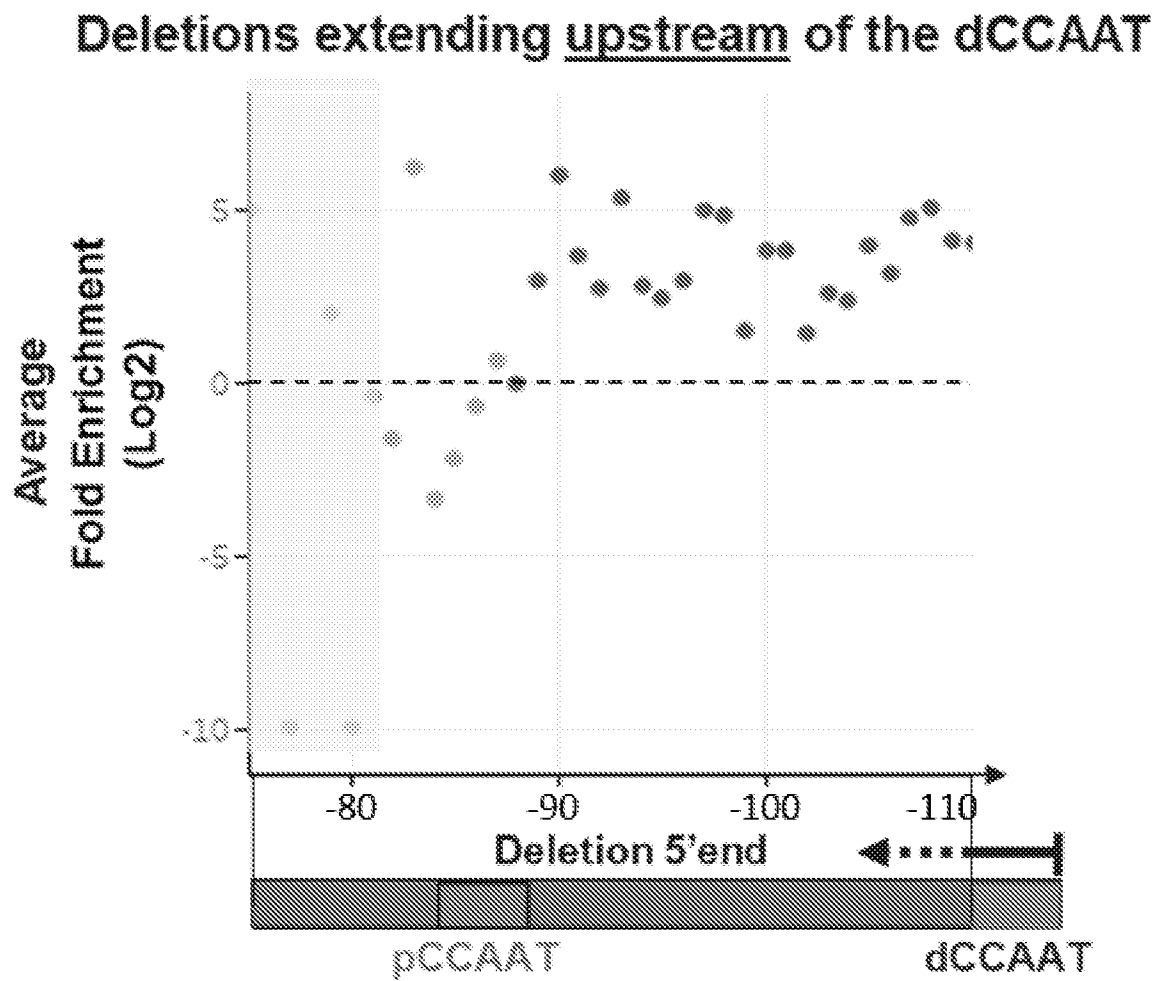


FIG 28D

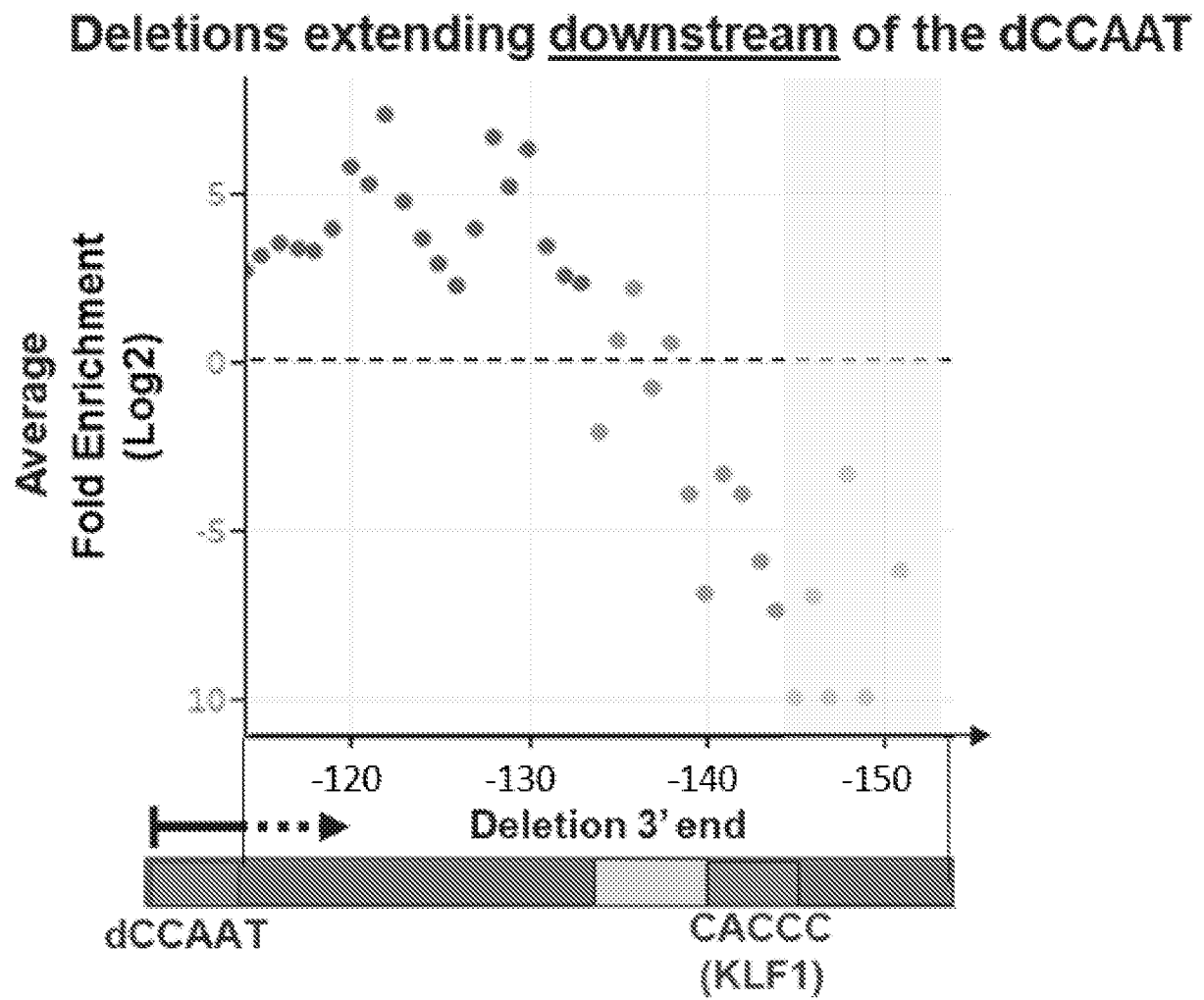


FIG 28E

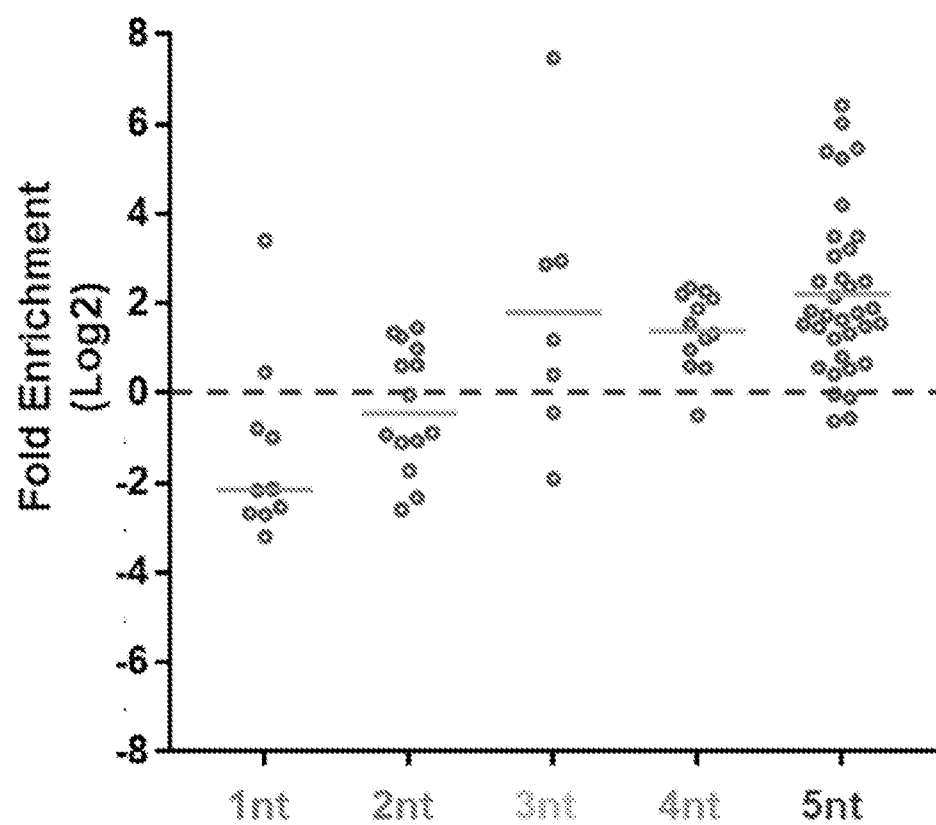
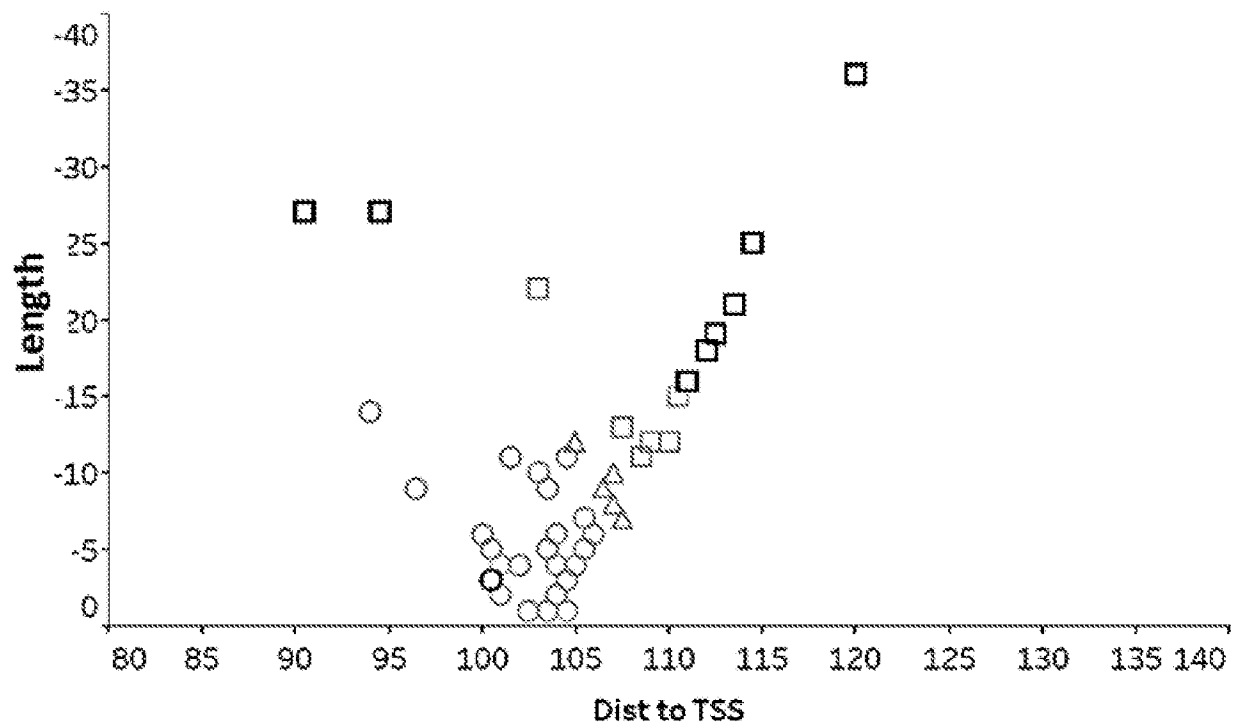


FIG 28F



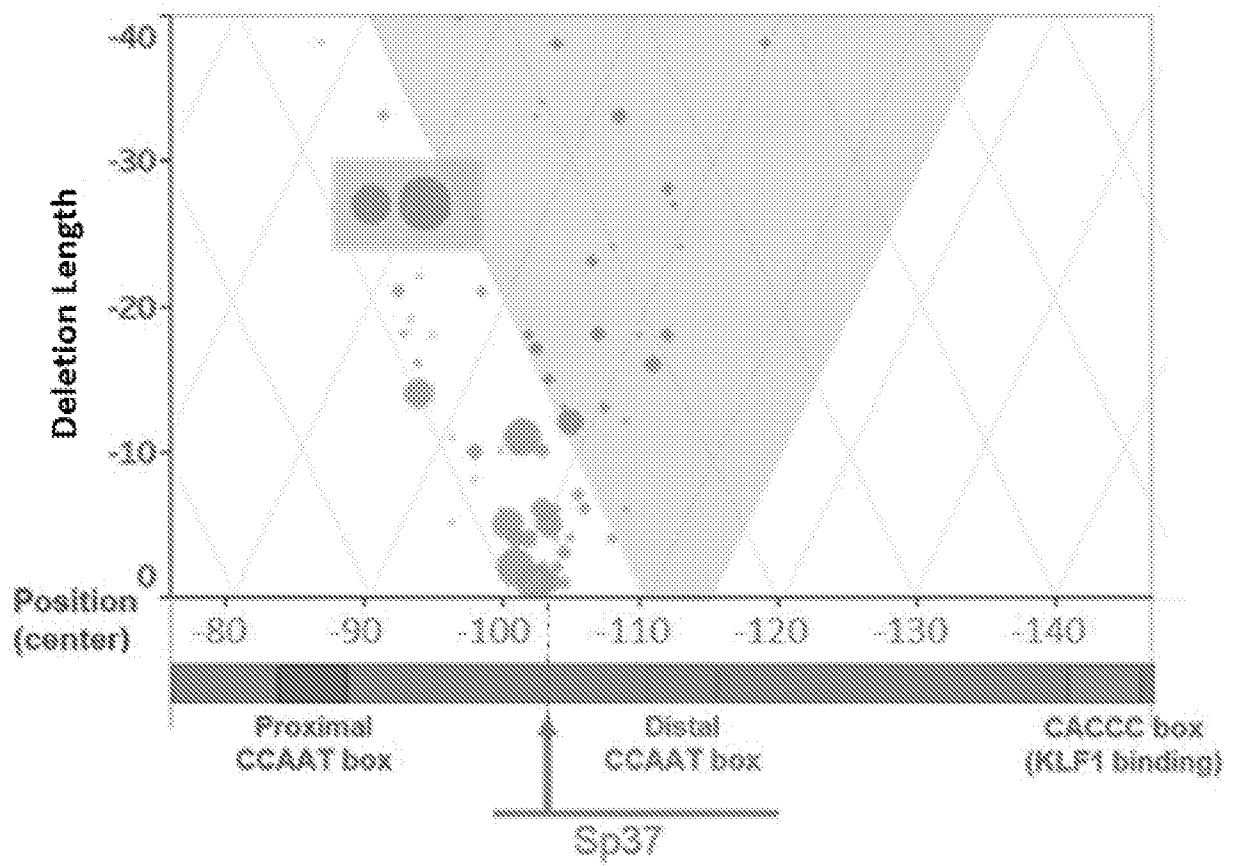
Enriched in high-HbF fraction (Enrichment ≥ 2 [log2])

- Not disrupting the distal CCAAT-box
- △ Disrupting 1-2nt of the distal CCAAT-box
- Disrupting 3-5nt of the distal CCAAT-box

**No/Low enrichment in high-HbF fraction
(Enrichment < 2 [log2])**

- Not disrupting the distal CCAAT-box
- △ Disrupting 1-2nt of the distal CCAAT-box
- Disrupting 3-5nt of the distal CCAAT-box

FIG 29A



SUBSTITUTE SHEET (RULE 26)

FIG 29B

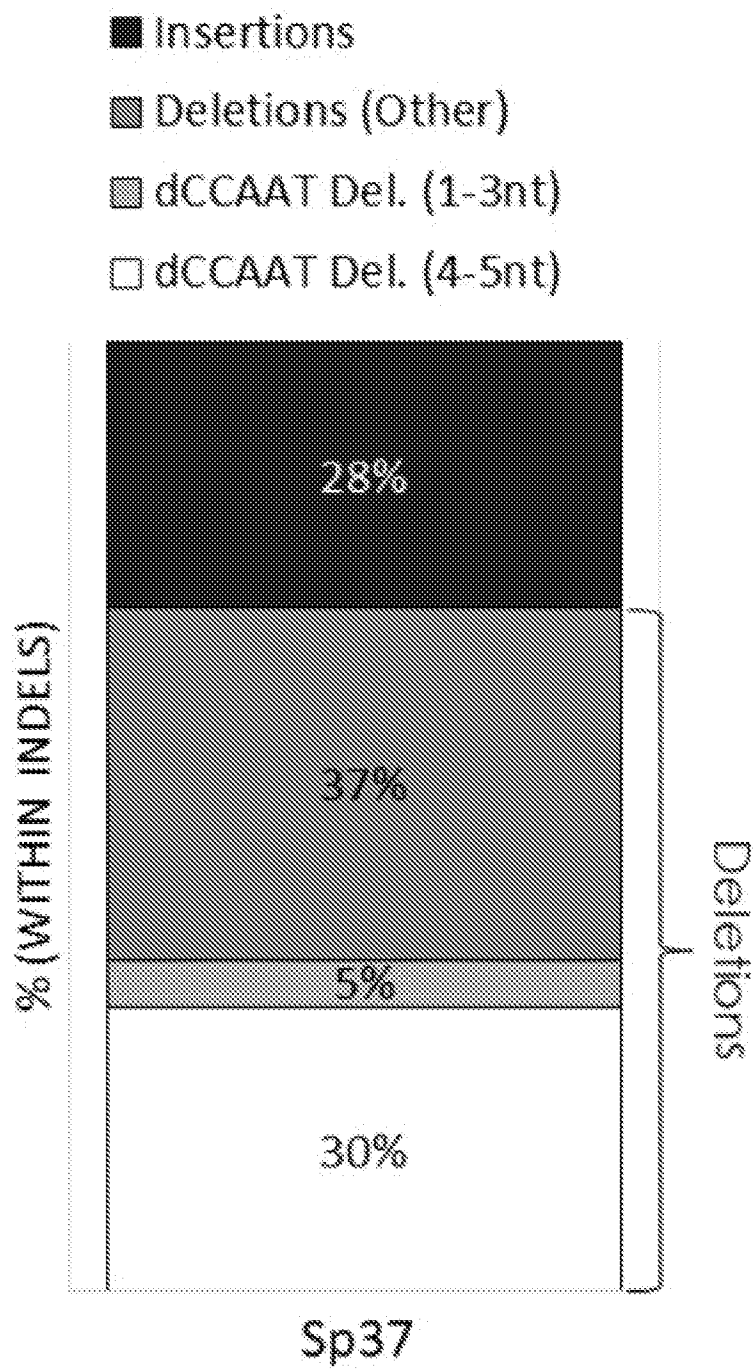


FIG 30A

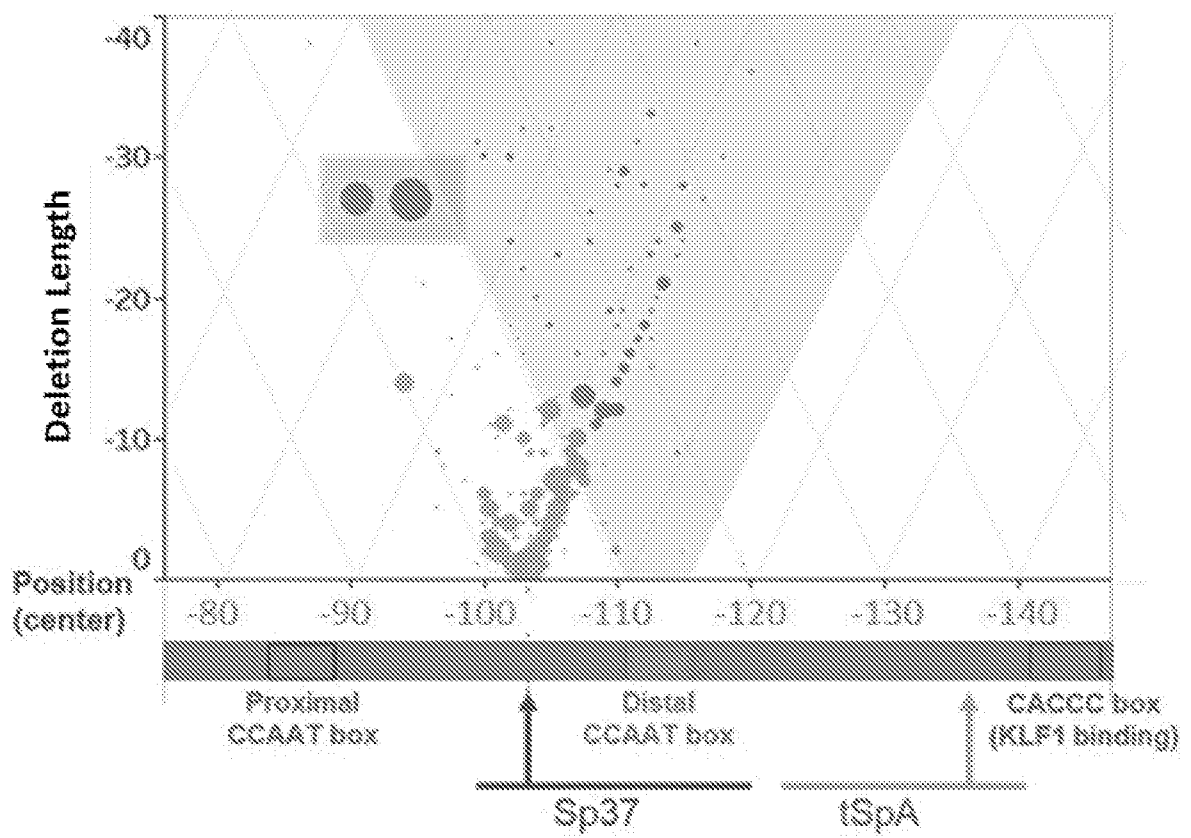


FIG 30B

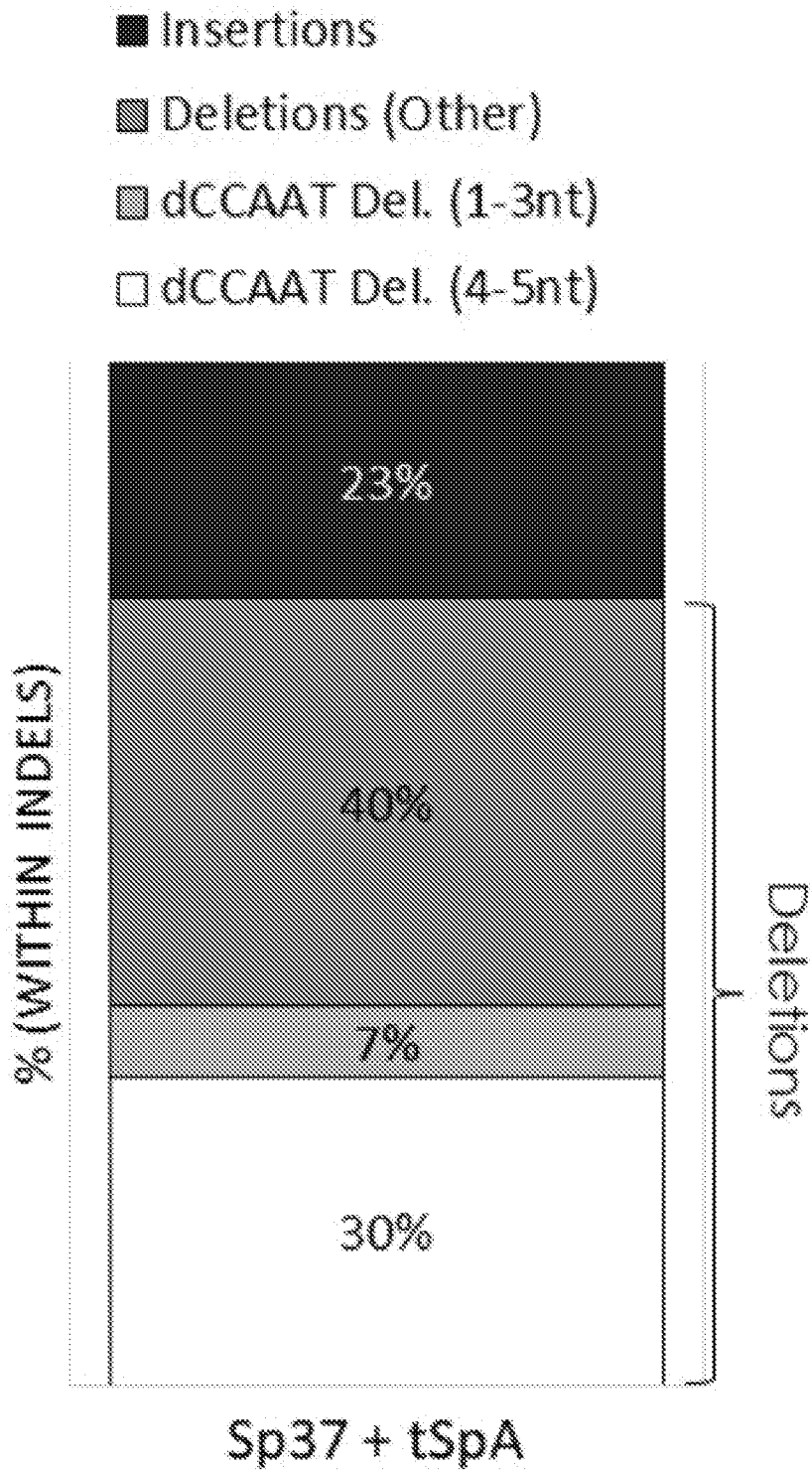


FIG 31

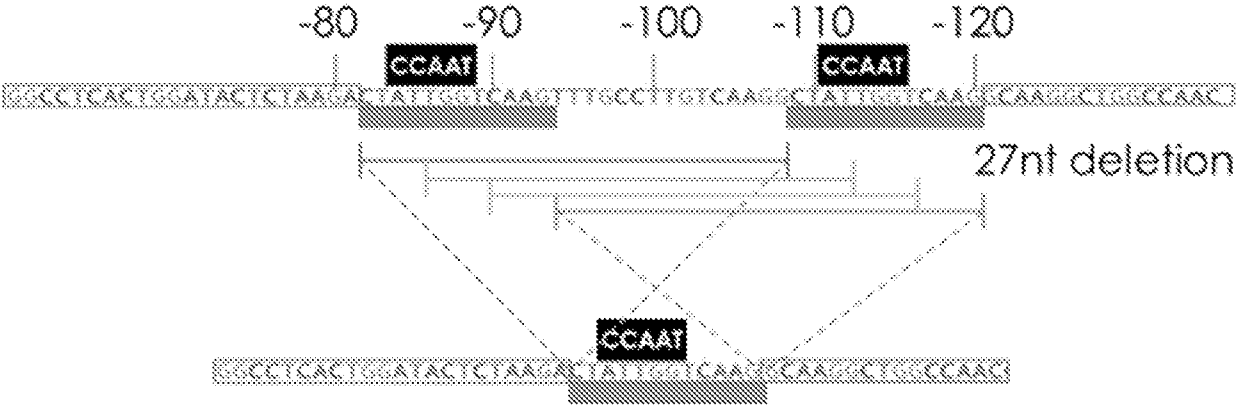


FIG 32A

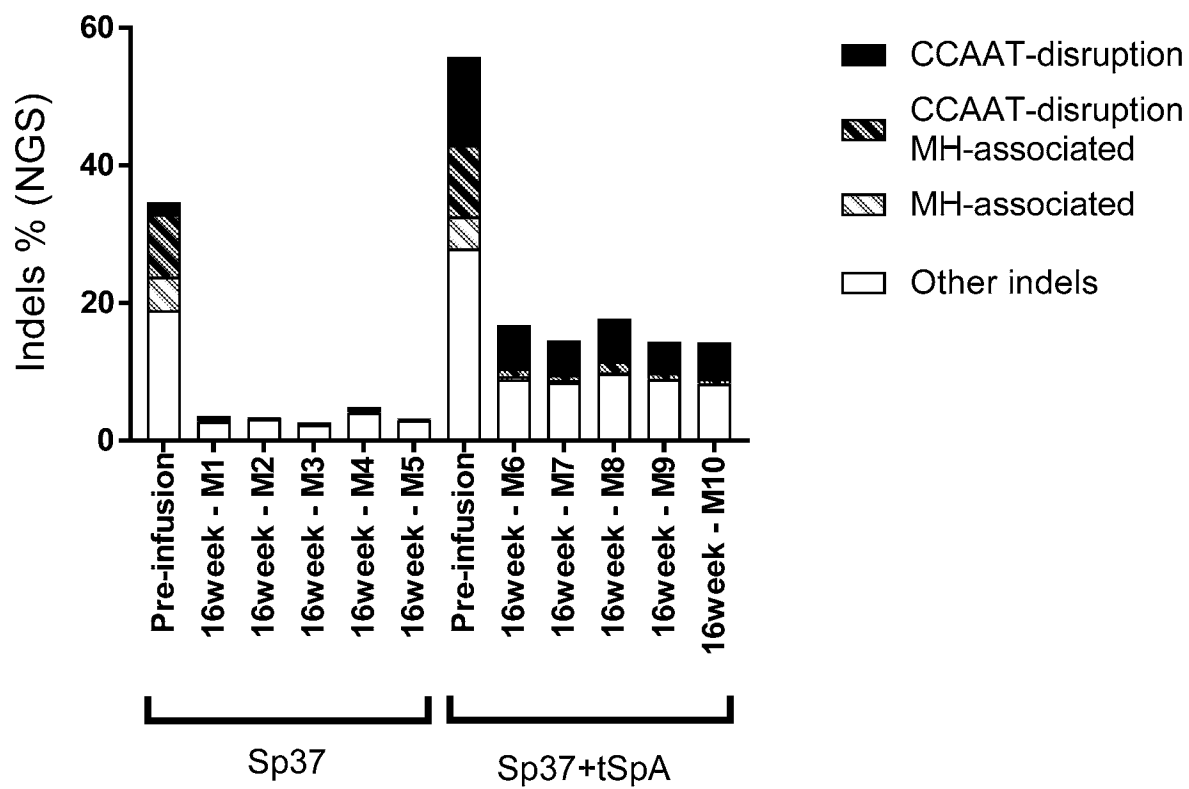


FIG. 32B

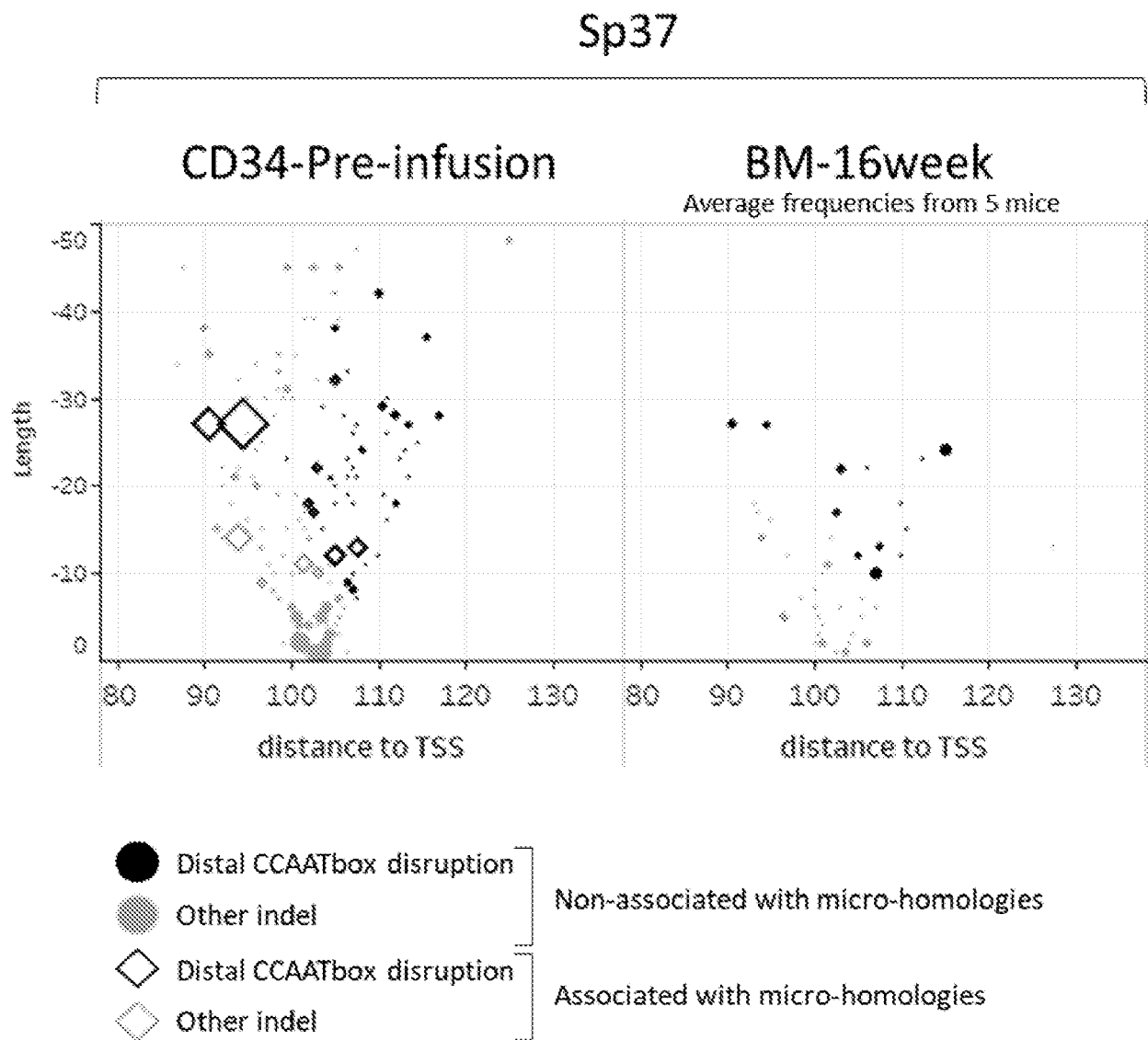


FIG. 32C

