MODIFIED CELLS AND METHODS FOR THE TREATMENT OF HEMOGLOBINOPATHIES

PRIORITY CLAIM

[0001] This application claims the benefit of U.S. Provisional Patent Application No. 62/945,190, filed December 8, 2019, and U.S. Provisional Patent Application No. 63/115,518, filed November 18, 2020, both of which are incorporated herein by reference in their entirety, including drawings.

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 December 8, 2020, is named SequenceListing.txt and is 699 KB 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 the alteration of genes encoding hemoglobin subunits and/or treatment of hemoglobinopathies.

BACKGROUND

[0004] 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 (Aγ)-globin chain (*HBG1*, also known as gamma globin A), and G gamma (Gγ)-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).

[0005] 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.

[0006] 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).

[0007] 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.

[0008] 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.

[0009] 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.

[0010] 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 (δo/δo βo/βo) patients, *HBG* is expressed, leading to production of HbF alone.

[0011] 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.

[0012] 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.

[0013] 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 βo/β+ or β+/β+. βo 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.

[0014] 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.

[0015] 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.

[0016] 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.

[0017] 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

[0018] Provided herein in certain embodiments are first populations of modified cells comprising a plurality of modified CD34+ or hematopoietic stem cells with one or more indels in an *HBG* gene promoter. In certain of these embodiments, the plurality of modified cells include an indel in a CCAAT box target region.

[0019] In certain embodiments of the first populations of modified cells provided herein, one or more of the cells in the plurality of modified cells include a HBG1/2 c.-104 to -121 deletion in a *HBG1* promoter, an *HBG2* promoter, or both. In certain of these embodiments, HBG1/2 c.-104 to -121 deletions make up 1% or more, 1.5% or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, 5.5% or more, 6% or more, 6.5% or more, 7% or more, 7.5% or more, 8% or more, 8.5% or more, 9% or more, 9.5% or more, 10% or more, 10.5% or more, 11% or more, 11.5% or more, 12% or more, 12.5% or more, 13% or more, 13.5% or more, 14% or more, 14.5% or more, 15% or more, or 15.5% or more of the indels in the plurality of modified cells as a whole. In certain of these embodiments, HBG1/2 c.-104 to -121 deletions make up less than 25% of the indels in the plurality of modified cells as a whole.

[0020] In certain embodiments of the first populations of modified cells provided herein, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, or 95% or more of the indels in the plurality of modified cells as a whole are deletions of at least 4 base pairs.

[0021] In certain embodiments of the first populations of modified cells provided herein, 25% or more, 30% or more, 35% or more, 40% or more, 45% or more, 50% or more, 55% or more, 60% or more, or 65% or more of the indels in the plurality of modified cells as a whole are deletions of at least 4 base pairs introduced by a repair mechanism other than microhomology-mediated end joining (MMEJ) repair.

[0022] In certain embodiments of the first populations of modified cells provided herein, 25% or more, 30% or more, 35% or more, 40% or more, 45% or more, 50% or more, 55% or more, 60% or more, or 65% or more of the indels in the plurality of modified cells as a whole are deletions of at least 4 base pairs introduced by non-homologous end joining (NHEJ) repair, e.g., canonical NHEJ repair.

[0023] In certain embodiments of the first populations of modified cells provided herein, 50% or more, 55% or more, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, or 92% or more of the indels in the plurality of modified cells as a whole are deletions of 1 to 25 base pairs.

[0024] In certain embodiments of the first populations of modified cells provided herein, 50% or more, 55% or more, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, or 85% or more of the indels in the plurality of modified cells as a whole are deletions of 3 to 25 base pairs.

[0025] In certain embodiments of the first populations of modified cells provided herein, 50% or more, 55% or more, 60% or more, 65% or more, 70% or more, 75% or more, or 80% or more of the indels in the plurality of modified cells as a whole are deletions of 4 to 25 base pairs.

[0026] In certain embodiments of the first populations of modified cells provided herein, 50% or more, 55% or more, 60% or more, 65% or more, 70% or more, or 72% or more of the indels in the plurality of modified cells as a whole are deletions of 5 to 25 base pairs.

[0027] In certain embodiments of the first populations of modified cells provided herein, the modified cells are produced by delivering a first RNP complex including a first gRNA comprising a first gRNA targeting domain and a Cpf1 RNA-guided nuclease or a modified Cpf1 RNA-guided nuclease to a first population of unmodified cells comprising a plurality of unmodified CD34+ or hematopoietic stem cells to generate indels. In certain of these embodiments, the first RNP complex is delivered to the first population of unmodified cells by electroporation. In certain embodiments, the first population of unmodified cells is from a subject having sickle cell disease. In certain embodiments, the first gRNA includes a 5' end and a 3' end, with a DNA extension at the 5' end and a 2'-O-methyl-3'-phosphorothioate modification at the 3' end. In certain of these embodiments, the DNA extension at the 5' end comprises a sequence set forth in any of SEQ ID NOs:1235-1250. In certain embodiments, the first gRNA targeting domain comprises the sequence set forth in SEQ ID NO:1254. In certain embodiments, the first gRNA comprises the sequence set forth in SEQ ID NO:1051. In certain embodiments, the modified Cpf1 RNA-guided nuclease comprises the sequence set forth in SEQ ID NO:1097. In certain embodiments, the first population of modified cells has higher fetal hemoglobin (HbF) levels than the first population of unmodified cells.

[0028] In certain embodiments of the first populations of modified cells provided herein, one or more of the cells in the plurality of modified cells include (a) a HBG1/2 c.-104 to -121 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (b) a HBG1/2 c.-110 to -115 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (c) a HBG1/2 c.-112 to -115 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (d) a HBG1/2 c.-113 to -115 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (e) a HBG1/2 c.-111 to -115 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (f) a HBG1/2 c.-111 to -117 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (g) a HBG1/2 c.-102 to -114 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (h) a HBG1/2 c.-114 to -118 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (i) a HBG1/2 c.-112 to -116 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; or (j) a HBG1/2 c.-113 to -117 deletion in a *HBG1* promoter, an *HBG2* promoter, or both.

[0029] In certain embodiments of the first populations of modified cells provided herein, the plurality of modified cells as a whole includes (a) a HBG1/2 c.-104 to -121 deletion in a *HBG1* promoter, an *HBG2* promoter, or both and (b) a HBG1/2 c.-110 to -115 deletion in a *HBG1* promoter, an *HBG2* promoter, or both. In certain embodiments, HBG1/2 c.-104 to -121 deletions make up 1% or more, 1.5% or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, 5.5% or more, 6% or more, 6.5% or more, 7% or more, 7.5% or more, 8% or more, 8.5% or more, 9% or more, 9.5% or more, 10% or more, 10.5% or more, 11% or more, 11.5% or more, 12% or more, 12.5% or more, 13% or more, 13.5% or more, 14% or more, 14.5% or more, or 15% or more of the indels in the plurality of modified cells as a whole. In certain embodiments, HBG1/2 c.-104 to -121 deletions make up 1% to 15.5% of the indels in the plurality of modified cells as a whole. In

certain embodiments, HBG1/2 c.-110 to -115 deletions make up 0.5% or more, 1% or more, 1.5% or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, or 5.5% or more of the indels in the plurality of modified cells as a whole. In certain embodiments, HBG1/2 c.-110 to -115 deletions make up 0.5% to 6% of the indels in the plurality of modified cells as a whole.

[0030] In certain embodiments of the first populations of modified cells provided herein, the plurality of modified cells as a whole include (a) a HBG1/2 c.-104 to -121 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (b) a HBG1/2 c.-110 to -115 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (c) a HBG1/2 c.-112 to -115 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (d) a HBG1/2 c.-113 to -115 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (e) a HBG1/2 c.-111 to -115 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (f) a HBG1/2 c.-111 to -117 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (g) a HBG1/2 c.-102 to -114 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (h) a HBG1/2 c.-114 to -118 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; (i) a HBG1/2 c.-112 to -116 deletion in a *HBG1* promoter, an *HBG2* promoter, or both; and (j) a HBG1/2 c.-113 to -117 deletion in a *HBG1* promoter, an *HBG2* promoter, or both. In certain embodiments, the HBG1/2 c.-104 to -121 deletions make up 1% or more, 1.5% or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, 5.5% or more, 6% or more, 6.5% or more, 7% or more, 7.5% or more, 8% or more, 8.5% or more, 9% or more, 9.5% or more, 10% or more, 10.5% or more, 11% or more, 11.5% or more, 12% or more, 12.5% or more, 13% or more, 13.5% or more, 14% or more, 14.5% or more, or 15% or more of the indels in the plurality of modified cells as a whole. In certain embodiments, the HBG1/2 c.-104 to -121 deletions make up 1% to 15.5% of the indels in the plurality of modified cells as a whole. In certain embodiments, the HBG1/2 c.-110 to -115 deletions make up 0.5% or more, 1% or more, 1.5% or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, or 5.5% or more of the indels in the plurality of modified cells as a whole. In certain embodiments, HBG1/2 c.-110 to -115 deletions make up 0.5% to 6% of the indels in the plurality of modified cells as a whole. In certain embodiments, the HBG1/2 c.-112 to -115 deletions make up 0.5% or more, 1% or more, 1.5% or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, or 5.5% or more of the indels in the plurality of modified cells as a whole. In certain embodiments, HBG1/2 c.-112 to -115 deletions make up 0.5% to 6% of the indels in the plurality of modified cells as a whole. In certain embodiments, the HBG1/2 c.-113 to -115 deletions make up 0.5% or more, 1% or more, 1.5% or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, or 5.5% or more of the indels in the plurality of modified cells as a whole. In certain embodiments, HBG1/2 c.-113 to -115 deletions make up 0.5% to 6% of the indels in the plurality of modified cells as a whole. In certain embodiments, the HBG1/2 c.-111 to -115 deletions make up 0.5% or more, 1% or more, 1.5%

or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, or 5.5% or more of the indels in the plurality of modified cells as a whole. In certain embodiments, HBG1/2 c.-111 to -115 deletions make up 0.5% to 6% of the indels in the plurality of modified cells as a whole. In certain embodiments, the HBG1/2 c.-111 to -117 deletions make up 0.5% or more, 1% or more, 1.5% or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, or 5.5% or more of the indels in the plurality of modified cells as a whole. In certain embodiments, HBG1/2 c.-111 to -117 deletions make up 0.5% to 6% of the indels in the plurality of modified cells as a whole. In certain embodiments, the HBG1/2 c.-102 to -114 deletions make up 0.5% or more, 1% or more, 1.5% or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, or 5.5% or more of the indels in the plurality of modified cells as a whole. In certain embodiments, HBG1/2 c.-102 to -114 deletions make up 0.5% to 6% of the indels in the plurality of modified cells as a whole. In certain embodiments, the HBG1/2 c.-114 to -118 deletions make up 0.5% or more, 1% or more, 1.5% or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, or 5.5% or more of the indels in the plurality of modified cells as a whole. In certain embodiments, HBG1/2 c.-114 to -118 deletions make up 0.5% to 6% of the indels in the plurality of modified cells as a whole. In certain embodiments, the HBG1/2 c.-112 to -116 deletions make up 0.5% or more, 1% or more, 1.5% or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, or 5.5% or more of the indels in the plurality of modified cells as a whole. In certain embodiments, HBG1/2 c.-112 to -116 deletions make up 0.5% to 6% of the indels in the plurality of modified cells as a whole. In certain embodiments, the HBG1/2 c.-113 to -117 deletions make up 0.5% or more, 1% or more, 1.5% or more, 2% or more, 2.5% or more. 3% or more, 3.5% or more, 4% or more, 4.5% or more, 5% or more, or 5.5% or more of the indels in the plurality of modified cells as a whole. In certain embodiments, HBG1/2 c.-113 to -117 deletions make up 0.5% to 6% of the indels in the plurality of modified cells as a whole.

[0031] In certain embodiments of the first populations of modified cells provided herein, the plurality of modified cells as a whole includes the 108 deletions present in all 14 samples in Table 25.

[0032] In certain embodiments of the first populations of modified cells provided herein, the plurality of modified cells as a whole includes the indels identified as having an "Ave % in Indel" of 0.1% or more in Table 25. In certain of these embodiments, the plurality of modified cells as a whole includes all of the indels in Table 25.

[0033] In certain embodiments of the first populations of modified cells provided herein, the plurality of modified cells as a whole include at least 10% more deletions of at least 4 base pairs than a second population of modified cells comprising a plurality of modified CD34+ or hematopoietic stem cells comprising a plurality of modified CD34+ or hematopoietic stem cells with one or more indels in an *HBG* gene promoter, where the indels of the second population of modified cells are generated by

delivering a second RNP complex including a second gRNA having a gRNA targeting domain comprising SEQ ID NO:339 and a Cas9 RNA-guided nuclease to a second population of unmodified cells comprising a plurality of unmodified CD34+ or hematopoietic stem cells. In certain of these embodiments, the second RNP complex is delivered to the second population of unmodified cells by electroporation. In certain embodiments, the second population of unmodified cells is from a subject having sickle cell disease. In certain embodiments, the first population of modified cells has higher HbF levels than the second population of modified cells. In certain of these embodiments, the plurality of modified cells in the second population include an indel in a CCAAT box target region.

[0034] Provided herein in certain embodiments are methods of inducing expression of HbF in a first population of modified cells comprising a plurality of modified CD34+ or hematopoietic stem cells with one or more indels in an *HBG* gene promoter, the method comprising delivering a first RNP complex including a first gRNA comprising a first gRNA targeting domain and a Cpf1 RNA-guided nuclease or a modified Cpf1 RNA-guided nuclease to a first population of unmodified cells comprising a plurality of unmodified CD34+ or hematopoietic stem cells to generate indels. In certain embodiments, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, or 95% or more of the resultant indels in the plurality of modified cells as a whole are deletions of at least 4 base pairs. In certain embodiments, the first population of modified cells exhibits increased HbF levels versus the first population of unmodified cells. In certain embodiments, the first RNP complex is delivered to the first population of unmodified cells by electroporation.

[0035] Provided herein in certain embodiments are methods of decreasing sickling in a first population of red blood cells (RBCs) cultured from a first population of modified cells comprising a plurality of modified CD34+ cells with one or more indels in an *HBG* gene promoter, the method comprising delivering a first RNP complex including a first gRNA comprising a first gRNA targeting domain and a Cpf1 RNA-guided nuclease or a modified Cpf1 RNA-guided nuclease to a first population of unmodified cells comprising a plurality of unmodified CD34+ cells to generate indels, then culturing the first population of RBCs from the first population of modified cells. In certain embodiments, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, or 95% or more of the resultant indels in the plurality of modified cells as a whole are deletions of at least 4 base pairs. In certain embodiments, the first population of RBCs exhibits significantly decreased sickling upon deoxygenation versus a second population of RBCs cultured from the first population of unmodified cells. In certain embodiments, the first population of RBCs sickle at a significantly lower oxygen tension, for example as measured by relative oxygen pressure, than the second population of RBCs. In certain embodiments, the first population of RBCs has a significantly higher minimum elongation index upon deoxygenation than the second population of RBCs. In certain embodiments, the first population of RBCs has a significantly higher velocity upon

deoxygenation than the second population of RBCs. In certain embodiments, the first population of RBCs has higher HbF levels than the second population of RBCs.

[0036] Provided herein in certain embodiments are methods of alleviating one or more symptoms of sickle cell disease in a subject in need thereof comprising delivering a first RNP complex including a first gRNA comprising a first targeting domain and a Cpf1 RNA-guided nuclease or a modified Cpf1 RNA-guided nuclease to a first population of unmodified cells comprising a plurality of unmodified CD34+ or hematopoietic stem cells to generate a first population of modified cells comprising a plurality of modified CD34+ or hematopoietic stem cells comprising one or more indels in an HBG gene promoter, and then administering the resultant first population of modified cells to the subject to alleviate one or more symptoms of sickle cell disease. In certain embodiments, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, or 95% or more of the resultant indels in the resultant plurality of modified CD34+ or hematopoietic stem cells as a whole are deletions of at least 4 base pairs. In certain embodiments, the methods further comprise detecting a population of modified erythroid progeny cells comprising a plurality of modified erythroid progeny cells cultured from the first population of modified cells at about 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, 8 weeks, 10 weeks, 12 weeks. 16 weeks, 20 weeks, 1 month, 2 months, 3 months, 4 months, 5 months, 6 months, 8 months, 12 months, 1 year, 2 years, 3 years, 4 years, 5 years, or more than 5 years after administration. In certain embodiments, the cultured cells may include bone marrow (BM)-engrafted CD34+ hematopoietic stem cells or blood cells derived therefrom, e.g., myeloid progenitor or differentiated myeloid cells, e.g., erythrocytes, mast cells, myoblasts; or lymphoid progenitors or differentiated lymphoid cells, e.g., T- or B- lymphocytes or NK cells. In certain embodiments, the method results in long-term engraftment of a plurality of HSC clones in bone marrow, e.g., at least 1 week, 2 weeks, 3 weeks, 4 weeks, 5 weeks, 6 weeks, 7 weeks, 8 weeks, 10 weeks, 12 weeks. 16 weeks, 20 weeks, 1 month, 2 months, 3 months, 4 months, 5 months, 6 months, 8 months, 12 months, 1 year, 2 years, 3 years, 4 years, 5 years, or more than 5 years after administration. In certain embodiments the method results in long-term expression of at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, at least 95%, or at least 99% of total hemoglobin as compared to a healthy subject. In certain embodiments, the method results in a reconstitution of all hematopoietic cell lineages, e.g., without any differentiation bias, e.g., without an erythroid lineage differentiation bias.

[0037] Provided herein in certain embodiments is a population of cells comprising a plurality of red blood cells (RBCs) cultured from a plurality of modified CD34+ cells from a subject having sickle cell disease, each modified cell comprising an indel in an HBG gene promoter. In certain embodiments, the plurality of modified cells as a whole includes the 108 deletions present in all 14 samples in Table 25. In certain embodiments, the population of cells is generated by delivering an RNP complex including a gRNA comprising a first targeting domain and a Cpf1 RNA-guided

nuclease or a modified Cpf1 RNA-guided nuclease to a plurality of unmodified CD34+ cells from a subject having sickle cell disease to generate the indels; and culturing the RBCs from the plurality of modified CD34+ cells.

[0038] Provided herein in certain embodiments is a population of cells comprising a plurality of red blood cells (RBCs) cultured from a plurality of modified CD34+ cells from a subject having sickle cell disease, each modified cell comprising an indel in an HBG gene promoter. In certain embodiments, 60% or more, 65% or more, 70% or more, 75% or more, 80% or more, 85% or more, 90% or more, or 95% or more of the indels in the HBG gene promoter are deletions of at least 4 base pairs. In certain embodiments, the population of cells is generated by delivering an RNP complex including a gRNA comprising a first targeting domain and a Cpf1 RNA-guided nuclease or a modified Cpf1 RNA-guided nuclease to a plurality of unmodified CD34+ cells from a subject having sickle cell disease to generate the indels; and culturing the RBCs from the plurality of modified CD34+ cells.

[0039] In certain embodiments of the population of cells comprising a plurality of RBCs provided herein, the plurality of RBCs sickle at a significantly lower oxygen tension, e.g., as measured by relative oxygen pressure, than a population of RBCs cultured from unmodified CD34+ cells, of the subject having sickle cell disease, have a significantly higher minimum elongation index upon deoxygenation than a population of RBCs cultured from unmodified CD34+ cells of the subject having sickle cell disease, and/or have a significantly higher velocity upon deoxygenation than a population of RBCs cultured from unmodified CD34+ cells of the subject having sickle cell disease.

[0040] Provided herein are genome editing systems, ribonucleoprotein (RNP) complexes, guide RNAs, Cpf1 proteins, including modified Cpf1 proteins (Cpf1 variants), and CRISPR-mediated methods for altering the promoter region of one or more γ–globin genes (e.g., *HBG1*, *HBG2*, or *HBG1* and *HBG2*) and increasing expression of fetal hemoglobin (HbF). In certain embodiments, an RNP complex may include a guide RNA (gRNA) complexed to a wild-type Cpf1 or modified Cpf1 RNA-guided nuclease (modified Cpf1 protein). In certain embodiments, a gRNA may comprise a sequence set forth in **Table 6**, **Table 12**, or **Table 13**. In certain embodiments, a gRNA may comprise a gRNA targeting domain. In certain embodiments, a gRNA targeting domain may comprise a sequence selected from the group consisting of SEQ ID NOs:1002, 1254, 1258, 1260, 1262, and 1264. In certain embodiments, a gRNA may comprise a gRNA sequence set forth in **Table 13**. In certain embodiments, a gRNA may comprise a sequence selected from the group consisting of SEQ ID NOs:1022, 1023, 1041-1105. In certain embodiments, the RNP complex may comprise an RNP complex set forth in **Table 15**. For example, an RNP complex may include a gRNA comprising the sequence set forth in SEQ ID NO:1051 and a modified Cpf1 protein encoded by the sequence set forth in SEQ ID NO:1097 (RNP32, **Table 15**).

[0041] The inventors have discovered herein that delivery of an RNP complex including a gRNA complexed to a modified Cpf1 protein may result in increased editing of a target nucleic acid. In certain embodiments, the modified Cpf1 protein may contain one or more modifications. In certain embodiments, the one or more modifications may include, without limitation, one or more mutations in a wild-type Cpf1 amino acid sequence, one or more mutations in a wild-type Cpf1 nucleic acid sequence, one or more nuclear localization signals (NLS), one or more purification tags (e.g., His tag), or a combination thereof. In certain embodiments, a modified Cpf1 may be encoded by a sequence set forth in SEQ ID NOs:1000, 1001, 1008-1018, 1032, 1035-39, 1094-1097, 1107-09 (Cpf1 polypeptide sequences) or SEQ ID NOs:1019-1021, 1110-17 (Cpf1 polynucleotide sequences). In certain embodiments, the gRNA may be a modified or unmodified gRNA. In certain embodiments, the gRNA may comprise a sequence set forth in **Table 6**, **Table 12**, or **Table 13**. In certain embodiments, the RNP complex may comprise an RNP complex set forth in **Table 15**. For example, the RNP complex may include a gRNA comprising the sequence set forth in SEQ ID NO:1051 and a modified Cpf1 protein encoded by the sequence set forth in SEQ ID NO:1097 (RNP32, **Table 15**). In certain embodiments, an RNP complex comprising a modified Cpf1 protein may increase editing of a target nucleic acid. In certain embodiments, an RNP complex comprising a modified Cpf1 protein may increase editing resulting in an increase of productive indels. In various embodiments, an increase in editing of the target nucleic acid may be assessed by any means known to skilled artisans, such as, but not limited to, PCR amplification of the target nucleic acid and subsequent sequencing analysis (e.g., Sanger sequencing, next generation sequencing).

[0042] The inventors have also discovered herein that delivery of an RNP complex including a modified gRNA complexed to an unmodified or modified Cpf1 protein may result in increased editing of a target nucleic acid. In certain embodiments, the modified gRNA may comprise one or more modifications including a phosphorothioate linkage modification, a phosphorodithioate (PS2) linkage modification, a 2'-O-methyl modification, one or more or a stretch of deoxyribonucleic acid (DNA) bases (also referred to herein as a "DNA extension"), one or more or a stretch of ribonucleic acid (RNA) bases (also referred to herein as a "RNA extension"), or combinations thereof. In certain embodiments, the DNA extension may comprise a sequence set forth in **Table 18**. For example, in certain embodiments, the DNA extension may comprise a sequence set forth in SEQ ID NOs:1235-1250. In certain embodiments, the RNA extension may comprise a sequence set forth in **Table 18**. For example, in certain embodiments, the RNA extension may comprise a sequence set forth in SEQ ID NOs:1231-1234, 1251-1253. In certain embodiments, the gRNA may comprise a sequence set forth in **Table 6**, **Table 12**, or **Table 13**. In certain embodiments, the RNP complex may comprise an RNP complex set forth in **Table 15**. For example, the RNP complex may include a gRNA comprising the sequence set forth in SEQ ID NO:1051 and a modified Cpf1 protein encoded by the sequence set forth in SEQ ID NO:1097 (RNP32, **Table 15**). In certain embodiments, an RNP complex comprising

a modified gRNA may increase editing of a target nucleic acid. In certain embodiments, an RNP complex comprising a modified gRNA may increase editing resulting in an increase of productive indels.

[0043] In certain embodiments, an RNP complex comprising a modified gRNA and a modified Cpf1 protein may increase editing of a target nucleic acid. In certain embodiments, an RNP complex comprising a modified gRNA and a modified Cpf1 protein may increase editing resulting in an increase of productive indels.

[0044] The inventors have also discovered that codelivery of an RNP complex comprising a gRNA complexed to a Cpf1 molecule (e.g., "gRNA-Cpf1-RNP") with a "booster element" may result in increased editing of a target nucleic acid. In certain embodiments, the RNP complex may comprise an RNP complex set forth in **Table 15**. For example, the RNP complex may include a gRNA comprising the sequence set forth in SEQ ID NO:1051 and a modified Cpf1 protein encoded by the sequence set forth in SEQ ID NO:1097 (RNP32, **Table 15**). As used herein, the term "booster element" refers to an element which, when co-delivered with a RNP complex comprising a gRNA complexed to an RNA-guided nuclease ("gRNA-nuclease-RNP"), increases editing of a target nucleic acid compared with editing of the target nucleic acid without the booster element. In certain embodiments, one or more booster elements may be codelivered with a gRNA-nuclease-RNP complex to increase editing of a target nucleic acid. In certain embodiments, codelivery of a booster element may increase editing resulting in an increase of productive indels. In various embodiments, an increase in editing of the target nucleic acid may be assessed by any means known to skilled artisans, such as, but not limited to, PCR amplification of the target nucleic acid and subsequent sequencing analysis (e.g., Sanger sequencing, next generation sequencing).

[0045] In certain embodiments, a gRNA-nuclease-RNP may comprise a gRNA-Cpf1-RNP. In certain embodiments, a Cpf1 molecule of the gRNA-Cpf1-RNP complex may be a wild-type Cpf1 or modified Cpf1. In certain embodiments, the Cpf1 molecule of the gRNA-Cpf1-RNP may be encoded by a sequence set forth in SEQ ID NOs:1000, 1001, 1008-1018, 1032, 1035-39, 1094-1097, 1107-09 (Cpf1 polypeptide sequences) or SEQ ID NOs:1019-1021, 1110-17 (Cpf1 polynucleotide sequences). In certain embodiments, the gRNA-Cpf1-RNP complex may comprise a gRNA comprising a targeting domain set forth in **Table 6** or **Table 12**. In certain embodiments, the gRNA-Cpf1-RNP complex may comprise a gRNA comprising a sequence set forth in **Table 13**. In certain embodiments, the gRNA may be a modified or unmodified gRNA.

[0046] In certain embodiments, a booster element may comprise a dead RNP comprising a dead gRNA molecule complexed with an RNA-guided nuclease molecule ("dead gRNA-nuclease-RNP"). In certain embodiments, the dead gRNA-nuclease-RNP may comprise a dead gRNA complexed with a wild-type (WT) Cas9 molecule ("dead gRNA-Cas9-RNP"), a dead gRNA complexed with a Cas9 nickase molecule ("dead gRNA-nickase-RNP") or a dead gRNA complexed with an enzymatically

inactive (ei) Cas9 molecule ("dead gRNA-eiCas9-RNP"). In certain embodiments, the dead gRNA-nuclease-RNP complex may have decreased activity or lack nuclease activity. In certain embodiments, the dead gRNA of the dead gRNA-nuclease-RNP complex may comprise any of the dead gRNAs set forth herein. For example, the dead gRNA may comprise a targeting domain may be the same as or may differ by no more than 3 nucleotides from a dead gRNA targeting domain set forth in **Table 8** or **Table 9**. In certain embodiments, the dead gRNA may include 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 8** or **Table 9**. In certain embodiments, the dead gRNA may be a modified or unmodified dead gRNA. As shown herein, codelivery of a gRNA-Cpf1-RNP with a dead gRNA-Cas9-RNP (i.e., an RNP comprising a dead gRNA complexed to a WT Cas9) or codelivery of a gRNA-Cpf1-RNP with a dead gRNA-nickase-RNP (i.e., an RNP comprising a dead gRNA complexed to a Cas9 nickase (i.e., the Cas9 D10A nickase)) resulted in an increase in total editing above levels observed following delivery of gRNA-Cpf1-RNP alone (see, e.g., **Examples 5, 7, 8**). 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, one or more booster elements may be comprised of one or more dead gRNA-nuclease-RNPs, e.g., dead gRNA-Cas9-RNP, dead gRNA-nickase-RNP, dead gRNA-eiCas9-RNP, to be codelivered with a gRNA-Cpf1-RNP. In certain embodiments, codelivery of a dead gRNA-nuclease-RNP does not alter the indel profile of a gRNA-Cpf1-RNP.

[0047] In certain embodiments, a booster element may comprise an RNP complex comprising a gRNA molecule complexed with an RNA-guided nuclease nickase molecule ("gRNA-nickase-RNP"). In certain embodiments, the RNA-guided nuclease nickase molecule may be a Cas9 nickase molecule, e.g., Cas9 D10A nickase. In certain embodiments, the gRNA of the gRNA-nickase-RNP may comprise any of the gRNAs set forth herein. For example, the gRNA may comprise a gRNA targeting domain set forth in **Table 8** or **Table 9**. In certain embodiments, the gRNA may be a modified or unmodified gRNA. As shown herein, codelivery of gRNA-Cpf1-RNP with a gRNA-nickase-RNP complex (RNP comprising a guide RNA complexed to a Cas9D10A nickase molecule) resulted in an increase in total editing above levels observed following delivery of gRNA-Cpf1-RNP alone (see, e.g., **Examples 4, 5**). Additionally, codelivery of a gRNA-nickase-RNP complex with a gRNA-Cpf1-RNP complex altered the directionality, length, and/or position of the indel profile of gRNA-Cpf1-RNP. In certain embodiments, a booster enhancer may be used to provide a desired editing outcome, for example, to increase the rate of productive indels. In certain embodiments,

codelivery of a gRNA-nickase-RNP complex with a gRNA-Cpf1-RNP complex may alter the indel profile of the gRNA-Cpf1-RNP.

[0048] In certain embodiments, a booster element may comprise a single stranded oligodeoxynucleotide (ssODN) or a double stranded oligodeoxynucleotide (dsODN). In certain embodiments, the ssODN may be any ssODN disclosed herein. In certain embodiments, an ssODN may comprise a sequence set forth in **Table 7**. For example, in certain embodiments, an ssODN may comprise the sequence set forth in SEQ ID NO:1040.

[0049] In one aspect, the disclosure relates to an RNP complex comprising a CRISPR from Prevotella and Franciscella 1 (Cpf1) RNA-guided nuclease or a variant thereof and a gRNA, wherein the gRNA is capable of binding to a target site in a promoter of an *HBG* gene in a cell. In certain embodiments, the gRNA may be modified or unmodified. In certain embodiments, the gRNA may comprise one or more modifications including a phosphorothioate linkage modification, a phosphorodithioate (PS2) linkage modification, a 2'-O-methyl modification, a DNA extension, an RNA extension, or combinations thereof. In certain embodiments, the DNA extension may comprise a sequence set forth in **Table 18**. In certain embodiments, the RNA extension may comprise a sequence set forth in **Table 18**. In certain embodiments, the gRNA may comprise a sequence set forth in **Table 6**, **Table 12**, or **Table 13**. In certain embodiments, the RNP complex may comprise an RNP complex set forth in **Table 15**. For example, the RNP complex may include a gRNA comprising the sequence set forth in SEQ ID NO:1051 and a Cpf1 variant protein encoded by the sequence set forth in SEQ ID NO:1097 (RNP32, **Table 15**). In certain embodiments, the Cpf1 variant protein may contain one or more modifications. In certain embodiments, the one or more modifications may include, without limitation, one or more mutations in a wild-type Cpf1 amino acid sequence, one or more mutations in a wild-type Cpf1 nucleic acid sequence, one or more nuclear localization signals (NLS), one or more purification tags (e.g., His tag), or a combination thereof. In certain embodiments, a Cpf1 variant protein may be encoded by a sequence set forth in SEQ ID NOs:1000, 1001, 1008-1018, 1032, 1035-39, 1094-1097, 1107-09 (Cpf1 polypeptide sequences) or SEQ ID NOs:1019-1021, 1110-17 (Cpf1 polynucleotide sequences).

[0050] In one aspect, the disclosure relates to a method of altering a promoter of an *HBG* gene in a cell comprising contacting the cell with an RNP complex disclosed herein. In certain embodiments, the alteration may comprise an indel within one or more regions set forth in **Table 11**. In certain embodiments, the alteration may comprise an indel within a CCAAT box target region of the promoter of an *HBG* gene. For example, in certain embodiments, the alteration may comprise an indel within Chr 11 (NC_000011.10): 5,249,955 – 5,249,987 (**Table 11**, Region 6), Chr 11 (NC_000011.10): 5,254,879 – 5,254,909 (**Table 11**, Region 16), or a combination thereof. In certain embodiments, the RNP complex may comprise a gRNA and a Cpf1 protein. In certain embodiments, the gRNA may comprise an RNA targeting domain set forth in **Table 13**. In certain embodiments, the

gRNA targeting domain may comprise a sequence selected from the group consisting of SEQ ID NOs:1002, 1254, 1258, 1260, 1262, and 1264. In certain embodiments, the gRNA may comprise a gRNA sequence set forth in **Table 13**. In certain embodiments, the gRNA may comprise a sequence selected from the group consisting of SEQ ID NOs:1022, 1023, 1041-1105. In certain embodiments, a gRNA may be configured to provide an editing event at Chr11:5249973, Chr11:5249977 (HBG1); Chr11:5250042, Chr11:5250046 (HBG1); Chr11:5250055, Chr11:5250059 (HBG1); Chr11:5250179, Chr11:5250183 (HBG1); Chr11:5254897, Chr11:5254901 (HBG2); Chr11:5254897, Chr11:5254901 (HBG2); Chr11:5254966, 5254970 (HBG2); Chr11:5254979, 5254983 (HBG2) (**Table 16**, **Table 17**). In certain embodiments, the cell may be further contacted with a booster element. In certain embodiment, a booster element may comprise a single stranded oligodeoxynucleotide (ssODN) or a double stranded oligodeoxynucleotide (dsODN). In certain embodiments, the ssODN may be any ssODN disclosed herein. In certain embodiments, an ssODN may comprise a sequence set forth in **Table 7**. For example, in certain embodiments, an ssODN may comprise the sequence set forth in SEQ ID NO:1040.

[0051] In one aspect, the disclosure relates to an isolated cell comprising an alteration in a promoter of *HBG* gene generated by the delivery of an RNP complex to the cell. In certain embodiments, the RNP complex may comprise a gRNA and a Cpf1 protein. In certain embodiments, the gRNA may be modified or unmodified. In certain embodiments, the gRNA may comprise one or more modifications including a phosphorothioate linkage modification, a phosphorodithioate (PS2) linkage modification, a 2'-O-methyl modification, a DNA extension, an RNA extension, or combinations thereof. In certain embodiments, the DNA extension may comprise a sequence set forth in **Table 18**. In certain embodiments, the RNA extension may comprise a sequence set forth in **Table 18**. In certain embodiments, the gRNA may comprise a sequence set forth in **Table 6**, **Table 12**, or **Table 13**. In certain embodiments, the RNP complex may comprise an RNP complex set forth in **Table 15**. For example, the RNP complex may include a gRNA comprising the sequence set forth in SEQ ID NO:1051 and a Cpf1 variant protein encoded by the sequence set forth in SEQ ID NO:1097 (RNP32, **Table 15**). In certain embodiments, the Cpf1 variant protein may contain one or more modifications. In certain embodiments, the one or more modifications may include, without limitation, one or more mutations in a wild-type Cpf1 amino acid sequence, one or more mutations in a wild-type Cpf1 nucleic acid sequence, one or more nuclear localization signals (NLS), one or more purification tags (e.g., His tag), or a combination thereof. In certain embodiments, a Cpf1 variant protein may be encoded by a sequence set forth in SEQ ID NOs:1000, 1001, 1008-1018, 1032, 1035-39, 1094-1097, 1107-09 (Cpf1 polypeptide sequences) or SEQ ID NOs:1019-1021, 1110-17 (Cpf1 polynucleotide sequences). In certain embodiments, a booster element may be co-delivered with the RNP complex. In certain embodiment, a booster element may comprise a single stranded oligodeoxynucleotide (ssODN) or a double stranded oligodeoxynucleotide (dsODN). In certain embodiments, the ssODN

may be any ssODN disclosed herein. In certain embodiments, an ssODN may comprise a sequence set forth in **Table 7**. For example, in certain embodiments, an ssODN may comprise the sequence set forth in SEQ ID NO:1040.

[0052] In one aspect, the disclosure relates to an *ex vivo* method of increasing the level of fetal hemoglobin (HbF) in a human cell by genome editing using an RNP complex comprising a gRNA and a Cpf1 RNA-guided nuclease or a variant thereof to affect an alteration in a promoter of an *HBG* gene, thereby to increase expression of HbF. In certain embodiments, the gRNA may be modified or unmodified. In certain embodiments, the gRNA may comprise one or more modifications including a phosphorothioate linkage modification, a phosphorodithioate (PS2) linkage modification, a 2'-O-methyl modification, a DNA extension, an RNA extension, or combinations thereof. In certain embodiments, the DNA extension may comprise a sequence set forth in **Table 18**. In certain embodiments, the RNA extension may comprise a sequence set forth in **Table 18**. In certain embodiments, the gRNA may comprise a sequence set forth in **Table 6**, **Table 12**, or **Table 13**. In certain embodiments, the RNP complex may comprise an RNP complex set forth in **Table 15**. For example, the RNP complex may include a gRNA comprising the sequence set forth in SEQ ID NO:1051 and a Cpf1 variant protein encoded by the sequence set forth in SEQ ID NO:1097 (RNP32, **Table 15**). In certain embodiments, the Cpf1 variant protein may contain one or more modifications. In certain embodiments, the one or more modifications may include, without limitation, one or more mutations in a wild-type Cpf1 amino acid sequence, one or more mutations in a wild-type Cpf1 nucleic acid sequence, one or more nuclear localization signals (NLS), one or more purification tags (e.g., His tag), or a combination thereof. In certain embodiments, a Cpf1 variant protein may be encoded by a sequence set forth in SEQ ID NOs:1000, 1001, 1008-1018, 1032, 1035-39, 1094-1097, 1107-09 (Cpf1 polypeptide sequences) or SEQ ID NOs:1019-1021, 1110-17 (Cpf1 polynucleotide sequences). In certain embodiments, a booster element may be co-delivered with the RNP complex. In certain embodiment, a booster element may comprise a single stranded oligodeoxynucleotide (ssODN) or a double stranded oligodeoxynucleotide (dsODN). In certain embodiments, the ssODN may be any ssODN disclosed herein. In certain embodiments, an ssODN may comprise a sequence set forth in **Table 7**. For example, in certain embodiments, an ssODN may comprise the sequence set forth in SEQ ID NO:1040.

[0053] In one aspect, the disclosure relates to a population of CD34+ or hematopoietic stem cells, wherein one or more cells in the population comprises an alteration in a promoter of an *HBG* gene, which alteration is generated by delivering an RNP complex comprising a gRNA and a Cpf1 RNA-guided nuclease or a variant thereof to the population of CD34+ or hematopoietic stem cells. In certain embodiments, the gRNA may be modified or unmodified. In certain embodiments, the gRNA may comprise one or more modifications including a phosphorothioate linkage modification, a phosphorodithioate (PS2) linkage modification, a 2'-O-methyl modification, a DNA extension, an

RNA extension, or combinations thereof. In certain embodiments, the DNA extension may comprise a sequence set forth in **Table 18**. In certain embodiments, the RNA extension may comprise a sequence set forth in **Table 18**. In certain embodiments, the gRNA may comprise a sequence set forth in **Table 6**, **Table 12**, or **Table 13**. In certain embodiments, the RNP complex may comprise an RNP complex set forth in **Table 15**. For example, the RNP complex may include a gRNA comprising the sequence set forth in SEQ ID NO:1051 and a Cpf1 variant protein encoded by the sequence set forth in SEQ ID NO:1097 (RNP32, **Table 15**). In certain embodiments, the Cpf1 variant protein may contain one or more modifications. In certain embodiments, the one or more modifications may include, without limitation, one or more mutations in a wild-type Cpf1 amino acid sequence, one or more mutations in a wild-type Cpf1 nucleic acid sequence, one or more nuclear localization signals (NLS), one or more purification tags (e.g., His tag), or a combination thereof. In certain embodiments, a Cpf1 variant protein may be encoded by a sequence set forth in SEQ ID NOs:1000, 1001, 1008-1018, 1032, 1035-39, 1094-1097, 1107-09 (Cpf1 polypeptide sequences) or SEQ ID NOs:1019-1021, 1110-17 (Cpf1 polynucleotide sequences). In certain embodiments, a booster element may be co-delivered with the RNP complex. In certain embodiment, a booster element may comprise a single stranded oligodeoxynucleotide (ssODN) or a double stranded oligodeoxynucleotide (dsODN). In certain embodiments, the ssODN may be any ssODN disclosed herein. In certain embodiments, an ssODN may comprise a sequence set forth in **Table 7**. For example, in certain embodiments, an ssODN may comprise the sequence set forth in SEQ ID NO:1040.

[0054] In one aspect, the disclosure relates to a method of alleviating one or more symptoms of sickle cell disease in a subject in need thereof, the method comprising: a) isolating a population of CD34+ or hematopoietic stem cells from the subject; b) modifying the population of isolated cells ex vivo by delivering an RNP complex comprising a gRNA and a Cpf1 RNA-guided nuclease or a variant thereof to the population of isolated cells, thereby to affect an alteration in a promoter of an *HBG* gene in one or more cells in the population; and c) administering the modified population of cells to the subject, thereby to alleviate one or more symptoms of sickle cell disease in the subject. In certain embodiments, the method may further comprise detecting progeny/daughter cells of the administered modified cells in the subject, e.g., in the form of BM-engrafted CD34+ hematopoietic stem cells or blood cells derived from those (e.g., myeloid progenitor or differentiated myeloid cells (e.g., erythrocyte, mast cells, myoblast); or lymphoid progenitors or differentiated lymphoid cells (e.g., T- or B- lymphocyte, or NK cell), e.g., at least [1, 2, 3, 4, 5, 6, 7, 8, 12, 16, or 20] weeks, or at least [1, 2, 3, 4, 5, or 6] months, or at least [1, 2, 3, 4, or 5] years after administration. In certain embodiments, the method may result in a reconstitution of all hematopoietic cell lineages, e.g., without any differentiation bias, e.g., without an erythroid lineage differentiation bias. In certain embodiments, the method may comprise administering a plurality of edited cells, and the method may result in long-term engraftment [e.g., at least [1, 2, 3, 4, 5, 6, 7, 8, 12, 16, or 20] weeks, or at least [1,

2, 3, 4, 5, or 6] months, or at least [1, 2, 3, 4, or 5] years after administration] of a plurality of [at least 5, 10, 15, 20, 25, … 100] different HSC clones in the BM. In certain embodiments, the method may further comprise detecting the level of total hemoglobin expression in the subject, at least [1, 2, 3, 4, 5, 6, 7, 8, 12, 16, or 20] weeks, or at least [1, 2, 3, 4, 5, or 6] months, or at least [1, 2, 3, 4, or 5] years after administration. In certain embodiments, the method may result in long-term expression [e.g., at least [1, 2, 3, 4, 5, 6, 7, 8, 12, 16, or 20] weeks, or at least [1, 2, 3, 4, 5, or 6] months, or at least [1, 2, 3, 4, or 5] years after administration] of [at least 50%, at least 60%... at least 99%] of total hemoglobin as compared to a healthy subject (e.g., as total Hb (e.g., HbA and HbF (if any) combined)). In certain embodiments, the alteration may comprise an indel within a CCAAT box target region of the promoter of the HBG gene. In certain embodiments, the RNP complex may be delivered using electroporation. In certain embodiments, at least about 5%, at least about 10%, at least about 20%, at least about 30%, at least about 40%, at least about 50%, at least about 60%, at least about 70%, at least about 80% or at least about 90% of the cells in the population of cells comprise a productive indel.

[0055] In one aspect, the disclosure relates to a gRNA comprising a 5' end and a 3' end, and comprising a DNA extension at the 5' end and a 2'-O-methyl-3'-phosphorothioate modification at the 3' end, wherein the gRNA includes an RNA segment capable of hybridizing to a target site and an RNA segment capable of associating with a Cpf1 RNA-guided nuclease. In certain embodiments, the DNA extension may comprise a sequence set forth in SEQ ID NOs:1235-1250. In certain embodiments, the gRNA may be modified or unmodified. In certain embodiments, the gRNA may comprise one or more modifications including a phosphorothioate linkage modification, a phosphorodithioate (PS2) linkage modification, a 2'-O-methyl modification, a DNA extension, an RNA extension, or combinations thereof. In certain embodiments, the DNA extension may comprise a sequence set forth in **Table 18**. In certain embodiments, the RNA extension may comprise a sequence set forth in **Table 18**. In certain embodiments, the gRNA may comprise a sequence set forth in **Table 6**, **Table 12**, or **Table 13**.

[0056] In one aspect, the disclosure relates to an RNP complex comprising a Cpf1 RNA-guided nuclease as disclosed herein and a gRNA as disclosed herein.

[0057] Also provided herein are genome editing systems, guide RNAs, and CRISPR-mediated methods for altering one or more γ–globin genes (e.g., *HBG1*, *HBG2*, or *HBG1* and *HBG2*) and increasing expression of fetal hemoglobin (HbF). In certain embodiments, one or more gRNAs comprising a sequence set forth in **Table 12** or **Table 13** may be used to introduce alterations in the promoter region of the *HBG* gene. In certain embodiments, genome editing systems, guide RNAs, 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, genome editing systems, guide RNAs, and CRISPR-mediated methods may alter a

CCAAT box target region that is 5' of the transcription site of the *HBG1*, *HBG2*, or *HBG1* and *HBG2* gene ("CCAAT box target region"). In certain embodiments, the CCAAT box target region may be 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). In certain embodiments, the CCAAT box target region may be 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). In certain embodiments, the CCAAT box target region may comprise a 18 nt target region, a 13 nt target region, a 11 nt target region, a 4 nt target region, a 1 nt target region, a -117G>A target region, or a combination thereof as disclosed herein. In certain embodiments, the alteration may be a 18 nt deletion, 13 nt deletion, 11 nt deletion, 4 nt deletion, 1 nt deletion, a substitution from G to A at c.-117 of the *HBG1*, *HBG2*, or *HBG1* and *HBG2* gene, or a combination thereof. In certain embodiments, the alteration may be a non-naturally occurring alteration or a naturally occurring alteration.

[0058] Also provided herein in certain embodiments are the use of optional genome editing system components such as template nucleic acids (oligonucleotide donor templates). In certain embodiments, template nucleic acids for use in targeting the CCAAT target region may include, without limitation, template nucleic acids encoding alterations of the CCAAT box target region. In certain embodiments, the CCAAT box target region may comprise a 18 nt target region, a 11 nt target region, a 4 nt target region, a 1 nt target region, or a combination thereof. In certain embodiments, the template nucleic acid may be a single stranded oligodeoxynucleotide (ssODN) or a double stranded oligodeoxynucleotide (dsODN). In certain embodiments, 5' and 3' homology arms, and exemplary full-length donor templates encoding alterations at the CCAAT box target region are also presented below (e.g., SEQ ID NOS: 974-995, 1040). In certain embodiments, the template nucleic acid may be a positive strand or a negative strand. In certain embodiments, the ssODN may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm. In certain embodiments, the 5' homology arm may be about 25 to about 200 nucleotides or more in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length; the replacement sequence may comprise 0 nucleotides in length; and the 3' homology arm may be about 25 to about 200 nucleotides or more in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the ssODN may comprise one or more phosphorothioates.

[0059] In certain embodiments, the genome editing systems, guide RNAs, and CRISPR-mediated methods for altering one or more γ–globin genes (e.g., *HBG1*, *HBG2*, or *HBG1* and *HBG2*), may include an RNA-guided nuclease. In certain embodiments, the RNA-guided nuclease may a Cpf1 or modified Cpf1 as disclosed herein.

[0060] In one aspect, the disclosure relates to compositions including a plurality of cells generated by the methods 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 methods 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. 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.

[0061] 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.

[0062] 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 8** or **Table 9**, 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 8** or **Table 9**. In certain embodiments, the second targeting domain differs by no more than 3 nucleotides from a gRNA targeting domain set forth in **Table 8** or **Table 9**. 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.

[0063] 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.

[0064] 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.

[0065] 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.

[0066] 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.

[0067] 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.

[0068] 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.

[0069] 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.

[0070] 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. In certain embodiments, the method may further comprise detecting progeny/daughter cells of the administered modified cells in the subject, e.g., in the form of BM-engrafted CD34+ hematopoietic stem cells or blood cells derived from those (e.g., myeloid progenitor or differentiated myeloid cells (e.g., erythrocyte, mast cells, myoblast); or lymphoid progenitors or differentiated lymphoid cells (e.g., T- or B- lymphocyte, or NK cell), e.g., at least [1, 2, 3, 4, 5, 6, 7, 8, 12, 16, or 20] weeks, or at least [1, 2, 3, 4, 5, or 6] months, or at least [1, 2, 3, 4, or 5] years after administration. In certain embodiments, the method may result in a reconstitution of all hematopoietic cell lineages, e.g., without any differentiation bias, e.g., without an erythroid lineage differentiation bias. In certain embodiments, the method may comprise administering a plurality of edited cells, and the method may result in long-term engraftment [e.g., at least [1, 2, 3, 4, 5, 6, 7, 8, 12, 16, or 20] weeks, or at least [1, 2, 3, 4, 5, or 6] months, or at least [1, 2, 3, 4, or 5] years after administration] of a plurality of [at least 5, 10, 15, 20, 25, … 100] different HSC clones in the BM. In certain embodiments, the method may further comprise detecting the level of total hemoglobin expression in the subject, at least [1, 2, 3, 4, 5, 6, 7, 8, 12, 16, or 20] weeks, or at least [1, 2, 3, 4, 5, or 6] months, or at least [1, 2, 3, 4, or 5] years after administration. In certain embodiments, the method may result in long-term expression [e.g., at least [1, 2, 3, 4, 5, 6, 7, 8, 12, 16, or 20] weeks, or at least [1, 2, 3, 4, 5, or 6] months, or at least [1, 2, 3, 4, or 5] years after administration] of [at least 50%, at least 60%... at least 99%] of total hemoglobin as compared to a healthy subject (e.g., as total Hb (e.g., HbA and HbF (if any) combined)). In certain embodiments, the alteration may comprise an indel within a CCAAT box target region of the promoter of the HBG gene.

[0071] In one aspect, the disclosure relates to a genome editing system, comprising: an RNA-guided nuclease; and a first guide RNA, in which the first guide RNA may comprise a first targeting domain that is complementary to a first sequence on a side of a CCAAT box target region of a 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 genome editing system may further comprise a template nucleic acid encoding an alteration of the CCAAT box target region of a human *HBG1, HBG2* gene, or a combination thereof. In certain embodiments, the template nucleic acid may be a single stranded oligodeoxynucleotide (ssODN) or a double stranded oligodeoxynucleotide (dsODN). In certain embodiments, the ssODN may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm. In certain embodiments, the homology arms may be symmetrical in length. In certain embodiments, the homology arms may be asymmetrical in length. In certain embodiments, the ssODN may comprise one or more phosphorothioate modifications. In certain embodiments, the one or more phosphorothioate modifications may be at the 5' end, the 3' end or a combination thereof. In certain embodiments, the ssODN may be a positive or negative strand. In certain embodiments, the alteration may be a non-naturally occurring alteration. In certain embodiments, the alteration may comprise a deletion of the CCAAT box target region. In certain embodiments, the deletion may comprise a 18 nt deletion, a 11 nt deletion, a 4 nt deletion, a 1 nt deletion, or a combination thereof. In certain embodiments, the CCAAT box target region may comprise a 18 nt target region, a 11 nt target region, a 4 nt target region, a 1 nt target region, or a combination thereof. In certain embodiments, the 5' homology arm may be about 25 to about 200 or more nucleotides in length, e.g., at about least 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length; the replacement sequence may comprise 0 nucleotides in length; and the 3' homology arm may be about 25 to about 200 or more nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 18 nt target region and the 3' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the 18 nt target region. In certain embodiments, the ssODN may comprise, may consist essentially of, or may consist of SEQ ID NO:974 or SEQ ID NO:975. In certain embodiments, the 5' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 11 nt target region and the 3' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the 11 nt target region. In certain embodiments, the ssODN may comprise, may consist essentially of, or may consist of SEQ ID NO:976 or SEQ ID NO:978. In certain embodiments, the 5' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 4 nt target region and the 3' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the 4 nt target region. In certain embodiments, the ssODN may comprise, may consist essentially of, or may consist of a sequence selected from the group consisting of SEQ ID NO:984, SEQ ID NO:985, SEQ ID NO:986, SEQ ID NO:987, SEQ ID NO:988, SEQ ID NO:989, SEQ ID NO:990, SEQ ID NO:991, SEQ ID NO:992, SEQ ID NO:993, SEQ ID NO:994, and SEQ ID NO:995. In certain embodiments,

the 5' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 1 nt target region and the 3' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the 1 nt target region. In certain embodiments, the homology arms may be symmetrical in length. In certain embodiments, the ssODN may comprise, may consist essentially of, or may consist of SEQ ID NO:982 or SEQ ID NO:983. In certain embodiments, the alteration may be a naturally occurring alteration. In certain embodiments, the alteration may comprise a deletion or mutation of the CCAAT box target region. In certain embodiments, the CCAAT box target region may comprise a 13 nt target region, -117G>A target region, or a combination thereof. In certain embodiments, the alteration may comprise a 13 nt deletion at the 13 nt target region or a substitution from G to A at the -117G>A target region, or a combination thereof. In certain embodiments, the 5' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 13 nt target region and the 3' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the 13 nt target region. In certain embodiments, the ssODN may comprise, may consist essentially of, or may consist of SEQ ID NO:977 or SEQ ID NO:979. In certain embodiments, the 5' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 13 nt target region and the 3' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the 13 nt target region. In certain embodiments, the ssODN may comprise, may consist essentially of, or may consist of SEQ ID NO:980 or SEQ ID NO:981. In certain embodiments, the RNA-guided nuclease may be an *S. pyogenes* Cas9. In certain embodiments, the RNA-guided nuclease may be a Cpf1 variant as disclosed herein. In certain embodiments, the first targeting domain may differ by no more than 3 nucleotides from a targeting domain listed in **Table 12, Table 13** or a gRNA in **Table 13**. In certain embodiments, the genome editing system may further comprise a second guide RNA, wherein the second guide RNA may comprise a second targeting domain that may be complementary to a second sequence on a side of a CCAAT box target region of a human *HBG1, HBG2* gene, or a combination thereof, wherein the second sequence optionally overlaps the CCAAT box target region of the human *HBG1, HBG2* gene, or a combination thereof. In certain embodiments, the RNA-guided nuclease may be a nickase, and optionally lacks RuvC activity. In certain embodiments, the genome editing system may comprise first and second RNA-guided nucleases. In certain embodiments, the first and second RNA-guided nucleases may be complexed with the first and second guide RNAs, respectively, forming first and second ribonucleoprotein complexes. In certain embodiments, the genome editing system may further comprise a third guide RNA; and optionally a fourth guide RNA, wherein the third and fourth guide RNAs may comprise 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, wherein one or both of the third and fourth sequences

optionally overlaps the GATA1 binding motif in *BCL11Ae* of the human *BCL11A* gene. In certain embodiments, the genome editing system may further comprise 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 RNA-guided nuclease may be a nickase, and 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, genome editing system may comprise 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.

[0072] In one aspect, the disclosure relates to a method of altering a cell comprising contacting a cell with a genome editing system. In certain embodiments, the step of contacting the cell with the genome editing system 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 may further comprise electroporating the cells, thereby introducing the first and second ribonucleoprotein complexes into the cell. In certain embodiments, the method of altering a cell may further comprise contacting the cell with a genome editing system, wherein the step of contacting the cell with the genome editing system may comprise contacting the cell with a solution comprising first, second, third, and optionally, fourth ribonucleoprotein complexes. In certain embodiments, the step of contacting the cell with the solution may further comprise 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, erythrocyte, or a precursor of an erythrocyte or erythroblast. In certain embodiments, the cell may be a $CD34^+$ cell.

[0073] In one aspect, the disclosure relates to a CRISPR-mediated method of altering a cell, comprising: introducing a first DNA single strand break (SSB) or double strand break (DSB) within a genome of the cell between positions c.-106 to -120 of a human *HBG1* or *HBG2* gene; and optionally introducing a second SSB or DSB within the genome of the cell between positions c.-106 to -120 of the human *HBG1* or *HBG2* gene, wherein the first and second SSBs or DSBs may be repaired by the cell in a manner that alters a CCAAT box 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 alteration of a CCAAT box target region of the human *HBG1* or *HBG2* gene. In certain

embodiments, the CRISPR-mediated method may further comprise a template nucleic acid encoding the alteration of the CCAAT box target region of a human *HBG1, HBG2* gene, or a combination thereof. In certain embodiments, the template nucleic acid may be a single stranded oligodeoxynucleotide (ssODN). In certain embodiments, the ssODN may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm. In certain embodiments, the ssODNs may be a positive or negative strand. In certain embodiments, the alteration may be a non-naturally occurring alteration. 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 CCAAT box target region of the human *HBG1* or *HBG2* gene. In certain embodiments, the CCAAT box target region may comprise a 18 nt target region, a 11 nt target region, a 4 nt target region, a 1 nt target region, or a combination thereof. In certain embodiments, the 5' homology arm may be about 25 to about 200 nucleotides or more in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length; the replacement sequence may comprise 0 nucleotides in length; and the 3' homology arm may be about 25 to about 200 nucleotides or more in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 18 nt target region, the 11 nt target region, the 4 nt target region, or the 1 nt target region and the 3' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of 18 nt target region, the 11 nt target region, the 4 nt target region, or the 1 nt target region. In certain embodiments, the ssODN may comprise, may consist essentially of, or may consist of a sequence selected from the group consisting of SEQ ID NO:974, SEQ ID NO:975, SEQ ID NO:976, SEQ ID NO:978, SEQ ID NO:984, SEQ ID NO:985, SEQ ID NO:986, SEQ ID NO:987, SEQ ID NO:988, SEQ ID NO:989, SEQ ID NO:990, SEQ ID NO:991, SEQ ID NO:992, SEQ ID NO:993, SEQ ID NO:994, SEQ ID NO:995, SEQ ID NO:982 and SEQ ID NO:983. In certain embodiments, the alteration may be a non-naturally occurring alteration. 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 CCAAT box target region of the human *HBG1* or *HBG2* gene. In certain embodiments, the CCAAT box target region may comprise a 13 nt target region, -117G>A target region, or a combination thereof. In certain embodiments, the alteration may comprise a 13 nt deletion at the 13 nt target region or a substitution from G to A at the -117G>A target region, or a combination thereof. In certain embodiments, the 5' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 13 nt target region or the -117G>A target region and the 3' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the 13 nt target region or the -117G>A target region. In certain embodiments, the ssODN may comprise, may consist essentially of,

or may consist of a sequence selected from the group consisting of SEQ ID NO:977 or SEQ ID NO:979. SEQ ID NO:980 or SEQ ID NO:981.

[0074] In one aspect, the disclosure relates to a composition that may comprise a plurality of cells generated by a method of altering a cell disclosed herein, wherein at least 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90% of the cells may comprise an alteration of a sequence of a CCAAT box target region of the human *HBG1* gene, *HBG2* gene, or a combination thereof. In certain embodiments, the alteration may comprise a 18 nt deletion, a 11 nt deletion, a 4 nt deletion, a 1 nt deletion, a 13 nt deletion, a substitution from G to A at the -117, of the human *HBG1* gene, *HBG2* gene, or a combination thereof. 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 composition may further comprise a pharmaceutically acceptable carrier.

[0075] In one aspect, the disclosure relates to a cell comprising a synthetic genotype generated by a method of altering a cell disclosed herein, wherein the cell may comprise a 18 nt deletion, a 11 nt deletion, a 4 nt deletion, a 1 nt deletion, a 13 nt deletion, a substitution from G to A at the -117, of the human *HBG1* gene, *HBG2* gene, or a combination thereof.

[0076] In one aspect, the disclosure relates to a cell comprising at least one allele of the *HBG* locus generated by a method of altering a cell disclosed herein, wherein the cell may encode a 18 nt deletion, a 11 nt deletion, a 4 nt deletion, a 1 nt deletion, a 13 nt deletion, a substitution from G to A at the -117, of the human *HBG1* gene, *HBG2* gene, or a combination thereof.

[0077] In one aspect, the disclosure relates to an AAV vector that may comprise a template nucleic acid encoding a non-naturally occurring alteration of a CCAAT box target region of a human *HBG1, HBG2* gene, or a combination thereof. In certain embodiments, the template nucleic acid may be a single stranded oligodeoxynucleotide (ssODN). In certain embodiments, the CCAAT box target region may comprise a 18 nt target region, a 11 nt target region, a 4 nt target region, a 1 nt target region, or a combination thereof. In certain embodiments, the ssODN may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm. In certain embodiments, the 5' homology arm may be about 25 to about 200 or more nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length; the replacement sequence may comprise 0 nucleotides in length; and the 3' homology arm may be about 25 to about 200 or more nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 18 nt target region, the 11 nt target region, the 4 nt target region, or the 1 nt target region and the 3' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90,

or 80 to 90 bp, homology 3' of 18 nt target region, the 11 nt target region, the 4 nt target region, or the 1 nt target region. In certain embodiments, the ssODN may comprise, may consist essentially of, or may consist of a sequence selected from the group consisting of SEQ ID NO:974-976, SEQ ID NO:978, SEQ ID NO:982-995.

[0078] In one aspect, the disclosure relates to a nucleotide sequence comprising a template nucleic acid encoding a non-naturally occurring alteration of a CCAAT box target region of a human *HBG1*, *HBG2* gene, or a combination thereof. In certain embodiments, the template nucleic acid may be a single stranded oligodeoxynucleotide (ssODN) or a double stranded oligodeoxynucleotide (dsODN) comprising the alteration. In certain embodiments, the CCAAT box target region may comprise a 18 nt target region, a 11 nt target region, a 4 nt target region, a 1 nt target region, or a combination thereof. In certain embodiments, the ssODN may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm. In certain embodiments, the 5' homology arm may be about 25 to about 200 or more nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length; the replacement sequence may comprise 0 nucleotides in length; and the 3' homology arm may be about 25 to about 200 or more nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 18 nt target region, the 11 nt target region, the 4 nt target region, or the 1 nt target region and the 3' homology arm may comprise about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of 18 nt target region, the 11 nt target region, the 4 nt target region, or the 1 nt target region. In certain embodiments, the ssODN may comprise, may consist essentially of, or may consist of a sequence selected from the group consisting of SEQ ID NO:974-976, SEQ ID NO:978, SEQ ID NO:982-995.

[0079] In one aspect, the disclosure relates to a cell comprising a synthetic genotype, wherein the cell may comprise a 18 nt deletion, a 11 nt deletion, a 4 nt deletion, a 1 nt deletion, a 13 nt deletion, a substitution from G to A at the -117, of the human *HBG1* gene, *HBG2* gene, or a combination thereof.

[0080] In one aspect, the disclosure relates to a composition, comprising a population of cells generated by a method of altering a cell disclosed herein, wherein the cells comprise a higher frequency of an alteration of a sequence of a CCAAT box target region of the human *HBG1* gene, *HBG2* gene, or a combination thereof relative to an unmodified population of cells. In certain embodiments, the higher frequency is at least about 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% or 90% higher. In certain embodiments, the alteration comprises a 18 nt deletion, a 11 nt deletion, a 4 nt deletion, a 1 nt deletion, a 13 nt deletion, a substitution from G to A at the -117, of the human *HBG1* gene, *HBG2* gene, or a combination thereof. In certain embodiments, at least a portion of the population of cells are within an erythroid lineage.

[0081] 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

[0082] 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.

[0083] **Fig. 1** depicts, in schematic form, *HBG1* and *HBG2* gene(s) in the context of the β-globin gene cluster on human chromosome 11. **Fig. 1**. Each gene in the β-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). β-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.

[0084] **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.

[0085] **Fig. 3** depicts gene editing of *HBG* of mPB CD34+ cells electroporated with variants of *Acidaminococcus sp.* Cpf1 ("AsCpf1"), namely His-AsCpf1-nNLS (SEQ ID NO:1000) and His-AsCpf1-sNLS-sNLS (SEQ ID NO:1001) complexed with the guide RNA HBG1-1 (OLI13620) (**Table 6**) ("His-AsCpf1-nNLS_HBG1-1 RNP" and "His-AsCpf1-sNLS-sNLS_HBG1-1 RNP"). The RNP were electroporated at 5 μM or 20 μM.

[0086] **Figs. 4A-4C** depict gene editing of *HBG* of mPB CD34+ cells electroporated with His-AsCpf1-sNLS-sNLS_HBG1-1 RNP alone or with various ssODNs. **Fig. 4A** depicts the percentage of indels detected by sequencing the *HBG* PCR product 72 hours after electroporation with His-

AsCpf1-sNLS-sNLS_HBG1-1 RNP alone or with OLI164324 ("-4nt - strand"), OLI16430 ("-4 nt + strand"), OLI16410 ("-18 nt - strand"), or OLI16409 ("-18 nt + strand") (**Table 7**). **Fig. 4B** depicts the percentage of the precise 18 nucleotide deletion indels detected by sequencing the *HBG* PCR product 72 hours after electroporation with His-AsCpf1-sNLS-sNLS_HBG1-1 RNP alone or with OLI16410 ("-18 nt - strand") or OLI16409 ("-18 nt + strand"). **Fig. 4C** depicts the percentage of the precise 18 nt deletion within all indels detected by sequencing the *HBG* PCR product 72 hours after electroporation with His-AsCpf1-sNLS-sNLS_HBG1-1 RNP alone or with OLI16410 ("-18 nt - strand") or OLI16409 ("-18 nt + strand").

[0087] **Figs. 5A-5F** depict schematics of the HBG1-1 target region and *S. Pyogenes* Cas9 gRNA pairs used in combination. **Fig. 5A** shows the target region of HBG1-1 gRNA (comprising the RNA targeting domain set forth in SEQ ID NO:1002, **Table 9**). The distal CCAAT box of *HBG* promoter (i.e., HBG1/2 c.-111 to -115) is indicated by a grey box. **Fig. 5B** shows the target region of HBG1-1, the distal CCAAT box of *HBG*, and the target region SpA gRNA (comprising the targeting domain of SEQ ID NO:941, **Table 9**). **Fig. 5C** shows the target region of HBG1-1, the distal CCAAT box of *HBG*, and the target region SpG gRNA (comprising the targeting domain of SEQ ID NO:359, **Table 9**). **Fig. 5D** shows the target region of HBG1-1, the distal CCAAT box of *HBG*, the target region of tSpA dead gRNA ("dgRNA") (comprising the targeting domain of SEQ ID NO:326, **Table 9**), and the target region of Sp182 dgRNA (comprising the targeting domain of SEQ ID NO:1028, **Table 9**). **Fig. 5E** shows the target region of HBG1-1, the distal CCAAT box of *HBG*, and tSpA dgRNA (comprising the targeting domain of SEQ ID NO:326, **Table 9**). **Fig. 5F** shows the target region of HBG1-1, the distal CCAAT box of *HBG*, and the target region of Sp182 dgRNA (comprising the targeting domain of SEQ ID NO:1028, **Table 9**).

[0088] **Figs. 6A-6B** depict HbF expression achieved by *ex vivo* editing of mPB CD34+ cells using HBG1-1-AsCpf1-RNP targeting the *HBG* promotor region. **Fig. 6A** shows results of editing at the *HBG* promoter region following delivery of 5 μM or 20 μM HBG1-1-AsCpf1-RNP ("HBG-1-1") via Amaxa electroporation in mPB CD34+ cells. Delivery of 20 μM HBG1-1-AsCpf1-RNP via Amaxa electroporation results in up to ~43% editing, and 21% HbF induction (above background levels). HbF levels are represented by black circles depicting expression levels of gamma-globin chains over total beta-like globin chains (gamma chains/[gamma chains + beta chain]) as measured by UPLC analysis on the erythroid progeny of mPB CD34+ cells. Grey bars depict the percentage of indels 72 hours post-electroporation detected by next generation sequencing (NGS) of the *HBG* PCR product. **Fig. 6B** shows results of editing at the *HBG* promoter region following delivery of 5 μM or 20 μM HBG1-1-AsCpf1-RNP ("HBG-1-1") via MaxCyte electroporation in mPB CD34+ cells. Delivery of 20 μM HBG1-1-AsCpf1-RNP via MaxCyte electroporation results in up to ~16% editing, and 7% HbF induction (above background levels). HbF levels are represented by black circles depicting expression levels of gamma-globin chains over total beta-like globin chains (gamma chains/[gamma

chains + beta chain]) as measured by UPLC analysis on the erythroid progeny of mPB CD34+ cells. Grey bars depict the percentage of indels 72 hours post-electroporation detected by NGS of the *HBG* PCR product.

[0089] **Fig. 7** depicts enhanced editing by HBG1-1-AsCpf1H800A-RNP at the *HBG* promotor region on the Maxcyte device by co-delivering various *S. Pyogenes* Cas9 WT or Cas9D10A RNPs. "S.Py D10A" represents the Cas9 D10A nickase protein and "S.Py WT" represents the Cas9 WT protein. RNPs tested include SpA-D10A-RNP, SpG-D10A-RNP, tSpA-Cas9-RNP, Sp182-Cas9-RNP, and tSpA-Cas9-RNP + Sp182-Cas9-RNP (**Table 8**). Total editing at the *HBG* promoter region (grey bars) and the associated HbF protein induction (black circles) following delivery of HBG1-1-AsCpf1H800A-RNP alone ("HBG1-1"), or in combination with *S. Pyogenes* Cas9 RNPs or pairs of RNP, is depicted. HbF levels are represented by black circles depicting expression levels of gamma-globin chains over total beta-like globin chains (gamma chains/[gamma chains + beta chain]) as measured by UPLC analysis on the erythroid progeny of mPB CD34+ cells. Grey bars depict the percentage of indels detected by NGS of the *HBG* PCR product.

[0090] **Fig. 8** depicts viability of mPB CD34+ cells following MaxCyte delivery of HBG1-1-AsCpf1H800A-RNP alone ("HBG1-1") or in combination with various *S. Pyogenes* Cas9 WT or Cas9D10A RNPs. "S.Py D10A" represents the Cas9 D10A nickase protein and "S.Py WT" represents the Cas9 WT protein. RNPs tested include SpA-D10A-RNP, SpG-D10A-RNP, tSpA-Cas9-RNP, Sp182-Cas9-RNP, and tSpA-Cas9-RNP + Sp182-Cas9-RNP (**Table 8**). Viability was measured by DAPI staining and flow cytometry analysis at 24h post electroporation.

[0091] **Figs. 9A-9B** depict the cleavage sites of HBG1-1-AsCpf1H800A-RNP and D10A-Cas9 RNP at the target region and the editing profile resulting from the co-delivery of the HBG1-1-AsCpf1H800A-RNP with a D10A RNP. **Fig. 9A** depicts the position of the HBG1-1-AsCpf1H800A-RNP cut sites on each strand of the target region (light grey arrows), as well as position of the nicking site targeted by the SpG-D10A-RNP and SpA-D10A-RNP (dark arrows) (**Table 8**). **Fig. 9B** depicts the editing profile resulting from the co-delivery of the HBG1-1-AsCpf1H800A-RNP ("HBG1-1 RNP") with either SpG-D10A-RNP ("spG RNP") or SpA-D10A-RNP ("spA RNP") in mPB CD34+ at 72h post-electroporation as detected by NGS analysis of the HBG PCR product (**Table 8**). The X-axis represents genomic position of the center of the indel relative to the HBG1-1-AsCpf1H800A-RNP positive strand cleavage site. The Y axis represents the length of the indel, where deletions are represented as negative values and insertions are represented as positive values. The total frequency of each indel is represented by the area of the symbol. Indels occurring at frequency equal or above 0.1% are depicted. The SpG and SpA target sites are indicated by dotted lines.

[0092] **Figs. 10A-10B** depict that the co-delivery of Sp182-Cas9-RNP with HBG1-1-AsCpf1H800A-RNP results in a boost in total indels and in distal CCAAT box disrupting indels with no substantial alteration of the indel profile, as detected by NGS analysis of the HBG PCR product at 72h post-

electroporation. **Fig. 10A** shows the indel profiles following editing with HBG1-1-AsCpf1H800A-RNP ("HBG1-1 RNP") alone, or in combination with Sp182-Cas9-RNP ("sp182 RNP"). The X-axis represents genomic position of the center of the indel relative to the HBG1-1-AsCpf1H800A-RNP positive strand cleavage site. The Y axis represents the length of the indel, where deletions are represented as negative values and insertions are represented as positive values. The total frequency of each indel is represented by the area of the symbol. Indels occurring at frequency equal or above 0.1% are depicted. The Sp182 target site is indicated by a dotted line. **Fig. 10B** depicts the frequency of indels disrupting either none, 1nt, 2nt, 3nt, 4nt, or the entire 5nt of the distal CCAAT box sequence.

[0093] **Fig. 11** depicts that optimal doses of HBG1-1-AsCpf1H800A-RNP co-delivered with Sp182-Cas9-RNP result in an increase in total editing, and HbF production (**Table 8**). Delivery of HBG1-1-AsCpf1H800A-RNP ("HBG1-1") as an RNP pair alongside Sp182-Cas9-RNP ("Sp182"), achieved >92% editing (grey bars) at the HBG promoter region with up to 34% HbF induction (above background) (black circles). No editing was observed when Sp182-Cas9-RNP was delivered alone at 12μM. HbF levels are represented by black circles depicting expression levels of gamma-globin chains over total beta-like globin chains (gamma chains/[gamma chains + beta chain]) as measured by UPLC analysis on the erythroid progeny of mPB CD34+ cells. Grey bars depict the percentage of indels detected by NGS of the *HBG* PCR product.

[0094] **Fig. 12** depicts the distribution of levels of gamma chain expression over total beta-like chains (gamma chains/[gamma chains + beta chain]) as measured by UPLC in the clonal erythroid progeny of single human mPB CD34+ cells edited at the *HBG* promotor region with HBG1-1-AsCpf1H800A-RNP in combination with Sp182-Cas9-RNP (**Table 8**). Each black circle represents the gamma-globin protein level detected in a clonal erythroid population derived from a single cell, isolated by FACS sorting at 48h post electroporation.

[0095] **Figs. 13A-13C** depicts total editing, HbF production, viability, and colony forming potential after co-delivery of RNP containing modified HBG1-1 gRNA (SEQ ID NO:1041, **Table 8**) complexed to His-AsCpf1-sNLS-sNLS H800A (SEQ ID NO:1032, **Table 8**) ("His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP," represented as "HBG1-1" in **Figs. 13A-C**) with increasing concentrations of ssODN OLI16431 (SEQ ID NO:1040, **Table 7**) (represented as "OLI16431" in **Figs. 13A-C**) (**Table 7**). **Fig. 13A** depicts editing at the distal CAATT box (grey bars) and HbF induction (black circles) after co-delivery of 6 μM His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP and increasing concentrations of ssODN OLI16431. HbF levels are represented by black circles depicting expression levels of gamma-globin chains over total beta-like globin chains (gamma chains/[gamma chains + beta chain]) as measured by UPLC analysis on the erythroid progeny of mPB CD34+ cells. Grey bars depict the percentage of indels detected by NGS of the *HBG* PCR product. **Fig. 13B** depicts viability of mPB CD34+ cells following delivery of His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP alone or in combination with increasing doses of ssODN OLI16431. Viability

was measured by DAPI exclusion at 72 hours post electroporation. **Fig. 13C** depicts the hematopoietic activity of the "HBG1-1" RNP and ssODN OLI16431 treated and donor matched untreated control $CD34^+$ cells in colony forming cell (CFC) assays. CFCs shown are per 800 $CD34^+$ cells plated. The number and subtype of colonies are indicated (GEMM: granulocyte-erythroid-monocyte-macrophage colony (black), GM: granulocyte-macrophage colony (dark grey), E: erythroid colony (light grey)).

[0096] **Figs. 14A-14B** depicts total editing, HbF production, and viability results using different concentrations of RNP containing unmodified HBG1-1 gRNA (SEQ ID NO:1022, **Table 8**) complexed to His-AsCpf1-sNLS-sNLS H800A (SEQ ID NO:1032, **Table 8**) ("His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP," represented as "HBG1-1" in **Figs. 14A-B**) co-delivered with different concentrations of ssODN OLI16431 (SEQ ID NO:1040, **Table 7**) (represented as "OLI16431" in **Figs. 14A-B**) (**Table 7**). **Fig. 14A** depicts editing at the distal CAATT box (black bars (48 hours) and light grey bars (14 days of erythroid culture)) and HbF induction (black circles) after co-delivery of His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP ("HBG1-1") and ssODN OLI16431 at varying concentrations. HbF levels are represented by black circles depicting expression levels of gamma-globin chains over total beta-like globin chains (gamma chains/[gamma chains + beta chain]) as measured by UPLC analysis on the erythroid progeny of mPB CD34+ cells. Bars depict the percentage of indels detected by NGS of the *HBG* PCR product at 48 hours (black) and 14 day erythroid culture (light grey) timepoints. **Fig. 14B** depicts viability of mPB CD34+ cells following delivery of His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP alone ("HBG1-1"), ssODN OLI16431 alone, or in combination with varying doses of HBG1-1 and OLI16431. Viability was measured by DAPI exclusion at 48 hours post electroporation and after 14 days in erythroid culture. Following the editing of mPB CD34+ cells, *ex vivo* differentiation into the erythroid linage was performed for 18 days (Giarratana 2011). At day 14 of culture, a subset of cells were isolated and viability (DAPI exclusion) and editing (NGS of the HBG PCR product) measurements were taken.

[0097] **Fig. 15** depicts editing by RNP in mPB CD34+ cells. RNPs included gRNA complexed with Cpf1 protein as set forth in **Table 15**. Illumina sequencing was performed on isolated genomic DNA at 72 hours post electroporation.

[0098] **Figs. 16A-16B** depict editing in bulk CD34+ cell population (black bars), progenitor cells (light grey bars), and HSCs (dark grey bars), as determined by Illumina sequencing 48 hours post electroporation. **Fig. 16A** depicts RNP33 (**Table 15**) delivered alone or co-delivered with ssODN OLI16431 (SEQ ID NO:1040, **Table 7**). **Fig. 16B** depicts RNP33 (**Table 15**) delivered alone or co-delivered with Sp182 RNP (dead gRNA comprising SEQ ID NO:1027 (**Table 8**) complexed with *S. pyogenes* Cas9 (SEQ ID NO:1033)).

[0099] **Fig. 17** depicts editing in bulk CD34+ cell population (black bars), progenitor cells (light grey bars), and HSCs (dark grey bars), as determined by Illumina sequencing 48 hours post

electroporation. RNP34, RNP33, and RNP43 (**Table 15**) were delivered alone or in combination with Sp182 RNP (dead gRNA comprising SEQ ID NO:1027 (**Table 8**) complexed with *S. pyogenes* Cas9 (SEQ ID NO:1033)) or ssODN OLI16431 (SEQ ID NO:1040, **Table 7**).

[0100] **Fig. 18** depicts editing in CD34+ cells as determined by Illumina sequencing 72 hours post electroporation. RNP64, RNP63, and RNP45 (**Table 15**) were delivered at a stoichiometry (gRNA:Cpf1 complexation ratio) of either 2, or 4, where the gRNA is in a molar excess.

[0101] **Fig. 19** depicts editing in CD34+ cells as determined by Illumina sequencing 72 hours post electroporation. RNP33, RNP64, RNP63, and RNP45 (**Table 15**) were delivered alone or in combination with Sp182 RNP (dead gRNA comprising SEQ ID NO:1027 (**Table 8**) complexed with *S. pyogenes* Cas9 (SEQ ID NO:1033)) or ssODN OLI16431 (SEQ ID NO:1040, **Table 7**).

[0102] **Fig. 20** depicts editing in CD34+ cells as determined by Illumina sequencing. RNPs comprising Cpf1 (SEQ ID:1094) complexed to gRNAs with various 5' DNA extensions (**Table 15**), were delivered alone or in combination with 8 μM OLI16431 (SEQ ID NO:1040, **Table 7**).

[0103] **Fig. 21** depicts editing in CD34+ cells, as determined by Illumina sequencing. RNPs comprising gRNAs with matched 5' ends (RNP49 vs RNP58 and RNP59 vs RNP60, **Table 15**) were delivered to CD34+ cells to assess the impact of 3' modifications. In both comparisons, gRNAs with 3' PS-OMe outperformed the unmodified 3' version at 24 hours post electroporation.

[0104] **Figs. 22A-22B** depict editing in CD34+ cells as determined by Illumina sequencing 24 and 48 hours post electroporation. RNP58 (**Table 15**) was delivered to CD34+ cells at a stoichiometry (gRNA:Cpf1 complexation ratio) of either 2:1, 1:1 or 0.5:1 molar ratios. At all doses tested, editing was best when RNP was complexed at 2:1 ratio.

[0105] **Figs. 23A-23B** depict editing in CD34+ cells as determined by Illumina sequencing. Figs. 45A and 45B depict RNPs comprising gRNAs with matched 5' ends, but different 3' modifications (**Table 15**) delivered to CD34+ cells to assess the impact of 3' modifications or extensions.

[0106] **Figs. 24A-24C** depict editing in CD34+ cells and their erythroid progeny, and HbF levels in the erythroid progeny following delivery of RNPs targeting various cut sites within the HBG locus. **Fig 24A** depicts RNPs comprising guide RNAs containing an unmodified 5' and 1 x PS-Ome at the 3' end (**Table 15**). **Fig 24B** depicts RNPs comprising guide RNAs containing 2PS +20 DNA extension at the 5' and 1 x PS-Ome at the 3' end (**Table 15**). **Fig 24C** depicts RNPs comprising guide RNAs containing a 25 DNA extension at the 5' and 1 x PS-Ome at the 3' end (**Table 15**).

[0107] **Fig. 25** depicts editing and HbF levels in erythroid progeny of CD34+ cells following delivery of RNP58 at 1 μM, 2 μM, and 4 μM.

[0108] **Fig. 26** depicts editing in CD34+ cells following Maxcyte electroporation of RNPs. RNP58, RNP26, RNP27, and RNP28 comprising gRNA SEQ ID:1051 complexed to different Cpf1 proteins (SEQ IDs: 1094, 1096, 1107, 1108) (**Table 15**) were delivered into CD34+ cells. Editing was determined by Illumina-sequencing 24 and 48 hours post electroporation.

[0109] **Fig. 27** depicts editing in CD34+ cells following Maxcyte electroporation of RNPs. RNP58, RNP29, RNP30, and RNP31 comprising Cpf1 protein SEQ ID: 1094 complexed to guide RNAs with various 5' extensions (**Table 15**) were delivered into CD34+ cells. Editing was determined by Illumina-seq 24 and 48 hours post electroporation. RNP30 was not tested (nt) at 1 μM due to limiting cell numbers.

[0110] **Fig. 28** depicts editing in bulk CD34+ cell population (black bars), progenitor cells (dark grey bars), and HSCs (light grey bars), as determined by Illumina sequencing 48 hours post electroporation. RNP58, RNP27, and RNP26 (**Table 15**) were delivered to CD34+ cells at 2 μM or 4 μM.

[0111] **Fig. 29** depicts editing in bulk CD34+ cell population (black bars), progenitor cells (dark grey bars), and HSCs (light grey bars), as determined by Illumina sequencing 48 hours post electroporation. RNP61, RNP62, and RNP34 (**Table 15**) (8 μM) were co-delivered to CD34+ cells with ssODN OLI16431 (SEQ ID NO:1040, **Table 7**) (8 μM).

[0112] **Fig. 30** depicts editing in bulk CD34+ cell population (black bars), progenitor cells (dark grey bars), and HSCs (light grey bars), as determined by Illumina sequencing 48 hours post electroporation. RNP58 and RNP32 (**Table 15**) were delivered to CD34+ cells at 2 μM.

[0113] **Fig. 31** depicts editing in bulk CD34+ cell population (black bars), progenitor cells (dark grey bars), and HSCs (light grey bars), as determined by Illumina sequencing 48 hours post electroporation. RNP58 and RNP1 (**Table 15**) were delivered to CD34+ cells at 2 μM, 4 μM, or 8 μM. The cells edited with RNP1 at 2 μM were not sorted (N.S), and thus editing data is not available.

[0114] **Fig. 32** depicts the indels of engrafted mPB CD34+ cells from BM of "NBSGW" mice 8 weeks post infusion of electroporated cells. RNP34 and RNP33 (**Table 15**) (8 μM) were co-delivered to CD34+ cells with ssODN OLI16431 (SEQ ID NO:1040, **Table 7**) (6 μM).

[0115] **Figs. 33A-33B** depict the indels of engrafted mPB CD34+ cells and HbF expression by erythroid cells derived from chimeric BM of "NBSGW" mice 8 weeks post infusion of electroporated cells. RNP33 or RNP34 (**Table 15**) was co-delivered with Sp182 RNP (dead gRNA comprising SEQ ID NO:1027 (**Table 8**) complexed with *S. pyogenes* Cas9 (SEQ ID NO:1033)) (16 μM total RNP) to CD34+ cells. **Fig. 33A** depicts the indel frequency in unfractionated bone marrow or flow-sorted individual populations of CD15+, CD19+, GlyA+, and Lin-CD34+ cells in mock-transfected (no RNP added) or RNP transfected cells. Lin-CD34+ cells are defined as CD34+ cells that are negative for CD3, CD14, CD15, CD16, CD19, CD20, and CD56) from bone marrow (BM) of nonirradiated NOD,B6.SCID Il2rγ-/- Kit(W41/W41) ("NBSGW") mice infused with mock (no RNP) or RNP transfected mPB CD34+ cells. Indels were determined for each cell population by Illumina sequencing. **Fig. 33B** depicts the HbF expression, calculated by UPLC as gamma/beta-like (%) from erythroid cell lysates following an 18-day erythroid differentiation culture from total chimeric BM.

[0116] **Figs. 34A-34B** depict the indels of engrafted mPB CD34+ cells and HbF expression by erythroid cells derived from chimeric BM of "NBSGW" mice 8 weeks post infusion of electroporated cells. RNP61 or RNP62 (**Table 15**) (8 μM) was co-delivered with ssODN OLI16431 (SEQ ID NO:1040, **Table 7**) (8 μM) to CD34+ cells. **Fig. 34A** depicts the indels of unfractionated bone marrow or flow-sorted individual populations of CD15+, CD19+, GlyA+, and Lin-CD34+ cells in mock-transfected (no RNP added) or RNP transfected cells. Lin-CD34+ cells are defined as CD34+ cells that are negative for CD3, CD14, CD15, CD16, CD19, CD20, and CD56) from bone marrow (BM) of nonirradiated NBSGW mice infused with mock (no RNP) or RNP transfected mPB CD34+ cells. Indels were determined for each cell population by Illumina sequencing. **Fig. 34B** depicts the HbF expression, calculated by UPLC as gamma/beta-like (%) by erythroid cells following an 18-day erythroid differentiation culture from total chimeric BM.

[0117] **Fig. 35** depicts human chimerism within bone marrow 8 weeks post infusion with mock (no RNP added) mPB CD34+ cells, or mPB CD34+ cells edited with RNP1 (4 or 8 μM) or RNP58 (2, 4 or 8 μM) (**Table 15**). Human chimerism and lineage reconstitution (CD45+, CD14+, CD19+, glycophorin A (GlyA, CD235a+), lineage, and CD34+, and mouse CD45+ marker expression) in BM was determined by flow cytometry.

[0118] **Fig. 36** depicts indels within unsorted bulk bone marrow 8 weeks post infusion with mock (no RNP added) mPB CD34+ cells, or mPB CD34+ cells edited with RNP1 (4 or 8 μM) or RNP58 (2, 4 or 8 μM) (**Table 15**). Indels were determined by Illumina sequencing.

[0119] **Fig. 37** depicts the indel frequency in unfractionated bone marrow or flow-sorted individual populations of CD15+, CD19+, GlyA+, and Lin-CD34+ cells in mock-transfected (no RNP added) or RNP transfected cells. Lin-CD34+ cells are defined as CD34+ cells that are negative for CD3, CD14, CD15, CD16, CD19, CD20, and CD56 from bone marrow (BM) of nonirradiated NOD,B6.SCID Il2rγ-/- Kit(W41/W41) ("NBSGW") mice infused with mock (no RNP) or RNP transfected mPB CD34+ cells. Indels were determined for each cell population by Illumina sequencing.

[0120] **Fig. 38** depicts HbF from GlyA+ fraction isolated from bone marrow at 8 weeks post infusion with mock (no RNP added) mPB CD34+ cells, or mPB CD34+ cells edited with RNP58 (2, 4 or 8 μM) (**Table 15**).

[0121] **Fig. 39** depicts colony forming potential of cells from bone marrow flushes taken 8 weeks post infusion of mock or edited human mobilized CD34+ cells. The number and subtype of colonies are indicated (GEMM: granulocyte-erythroid-monocyte-macrophage colony (black), GM: granulocyte-macrophage colony (dark grey), E: erythroid colony (light grey)).

[0122] **Fig. 40** depicts the sequences of Cpf1 protein variants set forth in **Table 14**. Nuclear localization sequences are shown as bolded letters, six-histidine sequences are shown as underlined letters. Additional permutations of the identity and N-terminal/C-terminal positions of NLS sequences, e.g., appending two or more nNLS sequences or combinations of nNLS and sNLS

sequences (or other NLS sequences) to either the N-terminal/C-terminal positions, as well as sequences with and without purification sequences, e.g., six-histidine sequences, are within the scope of the instantly disclosed subject matter.

[0123] **Figs. 41A-41E** depict editing in the HBG distal CCAAT box region. **Fig. 41A** shows a schematic of the Cpf1 (RNP34, **Table 15**) and SpCas9 (Sp35 RNP) cleavage sites at the HBG distal CCAAT box region. Sp35 RNP comprises Sp35 gRNA (comprising the targeting domain of SEQ ID NO:339 (i.e., CUUGUCAAGGCUAUUGGUCA (RNA)); SEQ ID NO:917 (i.e., CTTGTCAAGGCTATTGGTCA (DNA)) complexed with *S. pyogenes* wildtype (Wt) Cas9 protein. The dark grey jagged line with arrows marks the expected 4 nucleotide 5' overhang after cutting with RNP34. The grey dotted line marks the expected cut site of RNP34. This is the expected cut site for any RNP containing a gRNA comprising the gRNA targeting domain sequence UAAUUUCUACUCUUGUAGAUCCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1022). The Sp35 RNP expected cut site is indicated by the grey straight line with arrows. **Fig. 41B** depicts the percentage of indels derived from NHEJ and MMEJ repair in the CD34+ cell population, progenitor cells, and HSCs. MMEJ indels are represented by the black striped bar and NHEJ indels are represented by the white bar. **Fig. 41C** and **Fig. 41D** depicts the G-gamma chain expression levels (percentage of G-gamma chain/[total beta-like chains]) in erythroid cells derived from mPB CD34+ cells electroporated with Sp35 RNP or RNP34+ Sp182RNP; carrying indels ≤ 3 bp or > 3 bp in length on their HBG allele encoding for G-gamma. Only clones with a monoalleleic 4.9kb deletion (resulting in no g-gamma expression from one of the chromosome) were analyzed to ensure a single HBG allele was driving the g-gamma expression. Thus, the expression level shown represents the level of g-gamma expression by a single HBG gene, depending on the indel carried at the promoter, in cells that have the g-gamma gene on the other chromosome deleted. **Fig 41C** shows the results grouped by indel ≤ 3 bp or > 3 bp in length. **Fig. 41D** depicts the gamma chain expression levels (percentage of gamma chains/[total beta-like chains]) in erythroid cells derived from mPB CD34+ cells electroporated with Sp35 RNP or RNP34 + Sp182 RNP producing indels ≤ 3 bp or > 3 bp in length. The position of the deletion is indicated on the X axis. Xs indicate indels generated by Cas9 (Sp35 RNP) and circles indicate edits generated by Cpf1 (RNP34). **Fig. 41E** depicts the distribution of levels of gamma chain expression over total beta-like chains (gamma chains/[gamma chains + beta chain]) as measured by UPLC in the clonal erythroid progeny of single human mPB CD34+ cells edited at the *HBG* promotor region with spCas9 RNP "sp35" or Cpf1 RNP34 (HBG1-1-AsCpf1H800A-RNP) in combination with "booster element" Sp182-Cas9-RNP (**Table 8**). Each black circle represents the gamma-globin protein level detected in a clonal erythroid population derived from a single cell, isolated by FACS sorting at 48h post electroporation. The error bars show the median and interquartile range.

[0124] **Figs. 42A-42J** depict editing via Cpf1 or SpCas9 enzyme cleavage. **Fig. 42A** depicts the percentage of NHEJ mediated indels at the distal CCAAT box resulting from RNP58 (Cpf1) (**Table 15**) or Sp35 RNP (SpCas9) editing. Indel size is indicated by the x-axis. **Fig. 42B** depicts the percentage of > 3 bp NHEJ mediated indels, > 3 bp MMEJ mediated indels, and ≤ 3 bp indels at the distal CCAAT box resulting from editing by Cpf1 or SpCas9 RNP. For Sp35 RNP Cas9 indels in **Fig. 42B** (left pie chart): NHEJ > 3 bp = 21.42%; ≤ 3 bp = 48.84%; MMEJ > 3 bp = 29.74%. For RNP58 Cpf1 indels in **Fig. 42B** (right pie chart): NHEJ > 3 bp = 64.86%; ≤ 3 bp = 11.11%; MMEJ > 3 bp = 24.02%. **Fig. 42C** shows the wild type (wt) allele and top 18 indels, together with their average percentages in indel for all the samples, and their corresponding final sequences after the edit. Dashes represent deleted bases and the vertical line is the cut site (see also **Fig. 55A**). The long dark grey line above the wt allele marks the DNA sequence of the gRNA targeting domain of the gRNA of RNP58 (i.e., CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)), and the shorter black line marks the distal CCAAT box (appearing reverse complemented here). Base distances with respect to TSS are marked by arrows. The grey boxes respresent regions of homology which indicate MMEJ repair. **Fig. 42D** shows the wild type (wt) allele and top 18 indels, together with their average percentages in indel for all the samples, and their corresponding final sequences after the edit. Dashes represent deleted bases and the vertical line is the cut site (see also **Fig. 55A**). The long dark grey line above the wt allele marks the DNA sequence of the gRNA targeting domain of the gRNA of Sp35 RNP (i.e., CUUGUCAAGGCUAUUGGUCA SEQ ID NO:339), and the short black line marks the distal CCAAT box (appearing reverse complemented here). Base distances with respect to TSS are marked by arrows. The grey boxes respresent regions of homology which indicate MMEJ repair. **Fig. 42E** shows the percentage of bases deleted along the target region in samples edited with RNP58 or Sp35 RNP. The black line represents RNP58 editing and the dashed line represents Sp35 RNP editing. The target region DNA sequence is shown on the x axis. The grey line marks the DNA sequence of the gRNA targeting domain of the gRNA of Sp35 RNP (i.e., CUUGUCAAGGCUAUUGGUCA SEQ ID NO:339), and the short black line marks the distal CCAAT box (appearing reverse complemented here). **Fig. 42F** shows the normalized profiles (each with a maximum value of 1) for the detected deletions at each base on the target region in samples edited with RNP58 or Sp35 RNP. The black line represents RNP58 editing and the dashed line represents Sp35 RNP editing. The target region DNA sequence is shown on the x axis. The grey line marks the DNA sequence of the gRNA targeting domain of the gRNA of Sp35 RNP (i.e., CUUGUCAAGGCUAUUGGUCA SEQ ID NO:339), and the short black line marks the distal CCAAT box (appearing reverse complemented here). **Fig. 42G** shows the percentage of detected indels as a function of the deletion size (negative values on the x axis) or insertion (positive values in the x axis). The black line represents RNP58 editing and the dashed line represents Sp35 RNP editing. **Fig. 42H** shows the percentage of detected indels at each position in the target region. Sp35

RNP is represented by the dashed line. RNP58 is represented by the solid line. The target region DNA sequence is shown on the x axis. The gray line marks the DNA sequence of the gRNA targeting domain of the gRNA of Sp35 RNP (i.e., CUUGUCAAGGCUAUUGGUCA SEQ ID NO:339), and the short black line marks the distal CCAAT box (appearing reverse complemented here). **Fig. 42I** shows the frequency correlation for shared indels detected at ≥ 0.1% in either Sp35 RNP or RNP58 edited cells. **Fig. 42J** frequency correlation for all indels detected in at least one of the samples. Indels not detected in Sp35 RNP-edited cells are shown to the left of the vertical line at 1E-5. Indels not detected in RNP58-edited cells are shown to the bottom of the horizontal line at 1E-5. Triangles represent deletions ≤ 3 bp and circles represent deletions > 3bp.

[0125] **Figs. 43A-43O** depict RNP32 editing resulting in long term engraftment, indel maintenance, and high HbF induction *in vivo*. **Fig. 43A** depicts the percentage of > 3 bp NHEJ mediated indels, > 3 bp MMEJ mediated indels, and ≤ 3 bp indels at the distal CCAAT box resulting from editing by RNP32 24 hours following electroporation: NHEJ > 3 bp = 66.10%; ≤ 3 bp = 12.60%; MMEJ > 3 bp = 27.10%. **Fig. 43B** depicts the percentage of > 3 bp NHEJ mediated indels, > 3 bp MMEJ mediated indels, and < 3 bp indels at the distal CCAAT box resulting from editing by RNP32 48 hours following electroporation: NHEJ > 3 bp = 65.30%; ≤ 3 bp = 12.70%; MMEJ > 3 bp = 27.80%. **Fig. 43C** depicts the percentage of > 3 bp NHEJ mediated indels, > 3 bp MMEJ mediated indels, and < 3 bp indels at the distal CCAAT box resulting from editing by RNP32 72 hours following electroporation: NHEJ > 3 bp = 64.70%; ≤ 3 bp = 11.90%; MMEJ > 3 bp = 29.20%. **Fig. 43D** depicts the percentage of indels at the distal CCAAT box mediated via NHEJ repair resulting from editing by RNP32 24, 48, and 72 hours following electroporation and preinfusion. **Fig. 43E** depicts the percentage of indels at the distal CCAAT box mediated via NHEJ repair resulting from editing by RNP32 24 hours following electroporation and preinfusion. **Fig. 43F** depicts the percentage of indels at the distal CCAAT box mediated via NHEJ repair resulting from editing by RNP32 48 hours following electroporation and preinfusion. **Fig. 43G** depicts the percentage of indels at the distal CCAAT box mediated via NHEJ repair resulting from editing by RNP32 72 hours following electroporation and preinfusion. **Fig. 43H** depicts the percentage of all indels at the distal CCAAT box resulting from editing by RNP32 24, 48, and 72 hours following electroporation and preinfusion. **Fig. 43I** depicts the percentage of all indels at the distal CCAAT box from editing by RNP32 24 hours following electroporation and preinfusion. **Fig. 43J** depicts the percentage of all indels at the distal CCAAT box resulting from editing by RNP32 48 hours following electroporation and preinfusion. **Fig. 43K** depicts the percentage of all indels at the distal CCAAT box resulting from editing by RNP32 72 hours following electroporation and preinfusion. **Fig. 43L** depicts the percentage of indels in pre-infused RNP32 edited mPB CD34+ cells ("Preinfusion") and long-term repopulating CD34+ cells from NBSGW mice following infusion of RNP32 edited mPB CD34+ cells and 16 weeks engraftment ("BM"). **Fig. 43M** depicts human chimerism within bone marrow 16 weeks post

infusion with mock (no RNP added) mPB CD34+ cells or mPB CD34+ cells edited with RNP32 (**Table 15**). Human chimerism and lineage reconstitution (CD19+, CD15+, CD235A+), lineage, and CD34+, and mouse CD45+ marker expression) in BM was determined by flow cytometry. **Fig. 43N** depicts the percentage of F positive cells in mock (no RNP added) and CD235a+ (GlyA+) erythroid cells, derived from RNP32 edited CD34+ cells. **Fig. 43O** depicts the percentage of HbF shown by expression levels of gamma-globin chains over total beta-like globin chains (gamma chains/[gamma chains + beta chain]) as measured by UPLC analysis of CD235a+ (GlyA+) erythroid cells.

[0126] **Figs. 44A-44B** depict high polyclonality in NBSGW mice infused with RNP32 edited CD34+ cells. **Fig. 44A** shows high polyclonality over a 20 week period (8, 12, 16, and 20 weeks post-infusion) in the blood of mice (Mouse A, Mouse B, Mouse C, Mouse D) infused with RNP32 edited CD34+ cells. "0" represents data from pre-infused RNP32 edited mPB CD34+ cells. **Fig. 44B** shows high polyclonality in the bone marrow (BM) of NBSGW mice infused with RNP32 edited CD34+ cells at 20 weeks post-engraftment. In **Figs. 44A** and **44B**, each shade of grey within the graphical depiction represents a different indel signature, with the most frequent indel type located near the x-axis and the least frequent indel type located near the top of the plot.

[0127] **Fig. 45** shows a schematic of the β-globin locus. "CD34+ cells[a]" means CD34+ hematopoietic stem and progenitor cells, "HS" means hypersensitive site, "LCR" means locus control region, "TSS" means transcriptional start site.

[0128] **Fig. 46A** shows viability of normal and SCD CD34+ cells at Days 1 to 3 post electroporation with 6 μM of RNP32 compiled from two independent experiments. Individual data points and mean for each treatment group are shown. N=4 for normal donors; N=5 for sickle donors (CEL238-001 [normal CD34+ cells] and CEL211-001 [SCD CD34+ cells] were each tested twice). **Fig. 46B** shows CD34+ cell viability post-electroporation with RNP32. [a] indicates only applicable to RNP32-electroporated cells. [b] indicates insufficient number of cells available and thus viability was not assessed.

[0129] **Fig. 47A** shows indel levels of normal and SCD CD34+ cells at Days 1 to 3 post electroporation with 6 μM of RNP32 compiled from two independent experiments. Individual data points and mean for each treatment group are shown. N=4 for normal donors; N=5 for sickle donors. (CEL238-001 [normal CD34+ cells] and CEL211-001 [SCD CD34+ cells] were each tested twice. Indel=insertions and/or deletions. **Fig. 47B** shows comparable and efficient editing in RNP32 edited CD34+ cells from normal donors and patients with SCD. The percentage of indels was evaluated in unedited and RNP32 edited CD34+ cells from normal donors (n=3) and from unedited and RNP32 edited CD34+ cells from patients with SCD (n=4). Unedited cells did not undergo electroporation. **Fig. 47C** shows indel levels of CD34+ cells post-electroporation with RNP32. ND = Not Determined (Insufficient sample to run sequencing analysis).

[0130] **Figs. 48A-48D** depict PCR Primers and mixes for digital droplet polymerase chain reaction (ddPCR). **Fig. 48A** depicts a schematic representation of ddPCR primers and probe positions relative to RNP32 cut sites. **Fig. 48B** shows the primers and probe sequences for 4.9 kb fragment deletion assessment. **Fig. 48C** shows the Master Mix 1 and 2 compositions for ddPCR Assay 6 and 9. **Fig. 48D** shows the Master Mix 3 composition.

[0131] **Fig. 49A** shows the frequency of 4.9 kb fragment deletion of normal and SCD CD34+ cells at Day 1 post electroporation with 6 μM of RNP32. Data are from 1 study (SCD014). Each dot represents 1 sample. Untreated cells did not undergo electroporation. **Fig. 49B** shows the frequency of 4.9 kb deletion in CD34+ cells post-electroporation with RNP32. NA=data not available as digital droplet polymerase chain reaction assays failed. Untreated cells did not undergo electroporation.

[0132] **Fig. 50A** shows fold expansion of erythroid progeny of CD34+ cells. Normal and SCD CD34+ cells, either untreated or electroporated with 6 μM of RNP32, were placed in erythroid-inducing conditions for 18 days. Data were compiled from 2 independent experiments. Each dot represents 1 sample. **Fig. 50B** shows enucleation frequency of erythroid progeny of CD34+ cells. Normal and SCD CD34+ cells, either untreated or electroporated with 6 μM of RNP32, were placed in erythroid-inducing conditions for 18 days. Data were compiled from 2 independent experiments. Each dot represents 1 sample.

[0133] **Fig. 51A** shows an assessment of HbF induction in erythroid progeny in two experiments. Normal and SCD CD34+ cells, either untreated or electroporated with 6 μM of RNP32, were placed in erythroid-inducing conditions for 18 days. The relative abundance of globin chains in erythroid lysate was analyzed by RP-UPLC and HbF levels were calculated as HbF (%) = (Aγ + Gγ)/(Aγ + Gγ + β) (%). **Fig. 51B** shows the frequency of HbF+ RBCs evaluated by flow cytometry in one independent experiment. For **Fig. 51A** and **Fig. 51B**, each dot represents 1 sample. Paired T test was performed to determine whether the differences between RNP32-treated samples and untreated samples were statistically significant. * $p<0.05$, ** $p<0.01$, **** $p<0.0001$. RP-UPLC=reverse phase ultra-performance liquid chromatography. Untreated cells did not undergo electroporation. **Fig. 51C** shows comparable and robust *ex vivo* HbF expression in RNP32 edited CD34+ cells from normal donors and patients with SCD. The percentage of HbF is shown by expression levels of gamma-globin chains over total beta-like globin chains (gamma chains/[gamma chains + beta chain]) in unedited and RNP32 edited CD34+ cells from normal donors (n=3) and in unedited and RNP32 edited CD34+ cells from patients with SCD (n=4). Unedited cells did not undergo electroporation.

[0134] **Fig. 52** shows RNP32 edited CD34+ derived red blood cells (RBCs) from SCD patients have reduced sickling versus unedited RBCs from SCD patients when exposed to sodium metabisulfite and examined under a microscope. The percentage of sickled RBCs is shown for unedited SCD-derived RBCs and RNP32 edited SCD-derived RBCs. The mean HbF percentage is also shown for unedited SCD-derived RBCs and RNP32 edited SCD-derived RBCs.

[0135] **Fig. 53A** shows a graphic representation of the measured loss of deformability, the point-of-sickling of SCD RBCs, as a result of oxygen-depletion in time, followed by subsequent gain of deformability of RBC's during reoxygenation, as is visualized on the Oxygenscan. EImax represents RBC deformability at normoxia and EImin represents deformability upon deoxygenation. The point of sickling (PoS) reflects pO2 at which sickling begins and a >5% decrease in EI is observed during deoxygenation. EI=elongation index. **Figs. 53B-53E** shows the assessment of RBC deformability. Cultured RBCs from three batches of untreated normal CD34+ cells (**Fig. 53B**) and four batches of untreated or RNP32-edited SCD CD34+ cells (**Fig. 53C**) were analyzed on the Lorrca ektacytometer to measure deformability under shear stress, expressed as elongation index, when subjected to decreasing level of O2. In **Fig. 53C**, RBCs cultured from CD34+ cells that did not undergo electroporation with RNP32 (untreated) are represented by the black line (bottom line in all plots until at least 25 mmHg) and RBCs cultured from RNP32-edited SCD CD34+ cells are repsented by a dark gray line (top line in all plots until at least 25 mmHg). The point of sickling, representing the relative oxygen pressure when the SCD RBCs started to sickle during deoxygenation was plotted for each SCD RBC sample (**Fig. 53D**). The minimum elongation index of each SCD RBC sample, approximating the flexibility of RBC when deoxygenated is shown in (**Fig. 53E**). Each line connects the RBCs cultured from untreated and RNP32 edited cells from the same donor in the same experiment. CEL211-001 was tested twice in two independent experiments. A paired T test was performed to determine whether the differences between RBCs cultured from RNP32-treated samples and untreated samples were statistically significant. ** $p<0.01$, *** $p<0.001$.

[0136] **Fig. 54A** shows a rheology assessment of cultured RBCs. Untreated CD34+ cells and untreated or RNP32-electroporated SCD CD34+ cells were placed in erythroid-inducing conditions for 18 days to generate erythroid cells. Rheological behavior of cultured RBCs under varying concentrations of oxygen were evaluated using a microfluidic platform. The percentage velocity drop was calculated based on the differences between velocity at specified oxygen concentration and at atmospheric levels of oxygen (21%). Data shown are mean ± standard deviation. N= 5: untreated SCD samples, N= 5: RNP32 treated SCD samples, N= 4: untreated normal samples. Paired T test was used to compare the means between RBCs cultured from untreated and corresponding RNP32-edited SCD samples. *$p<0.05$, ***$p<0.001$, ****$p<0.0001$. **Fig. 54B** shows a summary of percentage velocity drop by varying oxygen concentration. [b] indicates samples clogged during assessment and experiments were aborted. Only data collected prior to the blockage formation are reported. Data are not plotted in **Fig. 54A**. **Fig. 54C** shows that RBCs cultured from RNP32-edited CD34+ cells from SCD patients have improved rheological properties, closer to RBCs from normal donors, compared to RBCs cultured from unedited CD34+ cells from SCD patients when placed in microfluidic channels that mimic blood flow in capillaries, in a range of oxygen levels. The percentage of normalized velocity was evaluated for RBCs cultured from unedited normal donor-derived CD34+ cells

(diamonds), RBCs cultured from unedited SCD-donor derived CD34+ cells (triangles), and RBCs cultured from RNP32-edited SCD-derived CD34+ cells (circles) at varying percentages of oxygen. Typical oxygen levels observed in the venous circulation are between ~4% to 6 % oxygen. **Fig. 54D** shows the correlation between HbF induction and rheology behavior. The level of HbF expressed by SCD samples (x-axis) was plotted against the corresponding percentage velocity drop (y-axis) at 0%, 2%, 4% and 6% oxygen concentration. Simple linear regression analysis was performed and the coefficient of determination is shown in each panel. **Fig. 54E** shows HbF levels correlate with velocity for RBCs cultured from unedited SCD-derived CD34+ cells (triangles) and RBCs from RNP32-edited SCD-derived CD34+ cells (circles) when placed in microfluidic channels mimicking blood flow in capillaries at an oxygen level of 4%. The percentage of HbF is shown by expression levels of gamma-globin chains over total beta-like globin chains (gamma chains/[gamma chains + beta chain]).

[0137] **Fig. 55A** shows a schematic of the distal CCAAT box region with 0-based hg38 coordinates chr11:5,249,949-5,249,987 (+) in *HBG1* and chr11:5,254,873-5,254,911 (+) in *HBG2*. The black line marks the DNA sequence of the gRNA targeting domain of the gRNA of RNP32 (i.e., CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)). The black box marks the distal CCAAT box. The dark grey jagged line with arrows marks the expected 4 nucleotide 5' overhang after cutting with RNP32. The grey dotted line marks the expected cut site of RNP32. The base before the cut site is marked with a black arrow at position -118 bases with respect to the TSS. The distances for the bases at the boundaries to the TSS are marked by black arrows at the end of the sequences (TSS: -93 and -130). **Fig. 55B** shows the oligonucleotides used for sequencing the indel profiles generated by RNP32. **Fig. 55C** shows the amplicon used for analysis of RNP32 editing. **Fig. 55D** shows an example of an indel_id used to characterize the indel profile of RNP32. The Indel_id is a string identifying the indel, shown as indel_start_position + _ + indel_length + _ + ID. Where ID is NA for deletions and for insertions is the sequence inserted.

[0138] **Fig. 56** depicts the percentage of > 3 bp NHEJ mediated indels, > 3 bp MMEJ mediated indels, and ≤ 3 bp indels at the distal CCAAT box resulting from editing by RNP32: NHEJ > 3 bp = 65.9%; ≤ 3 bp = 8.6%; MMEJ > 3 bp =25.5%.

[0139] **Fig. 57** shows the wild type (wt) allele and top 20 indels, together with their average percentages in indel for all the samples, and their corresponding final sequences after the edit. Dashes represent deleted bases and the vertical line is the cut site (see also **Fig. 55A**). The black line marks the DNA sequence of the gRNA targeting domain of the gRNA of RNP32 (i.e., CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)), and shorter black line marks the distal CCAAT box (appearing reverse complemented here). Base distances with respect to TSS are marked by arrows.

[0140] **Fig. 58** shows the percentage of detected deletions as a function of the distance between the deletion center and the cut site, represented by the vertical dashed grey line taken as the middle of the 5' overhang. The highest peak is at position –6 bp relative to the cut site (TSS: –113) zand towards the TSS (see also **Fig. 55A**). Profiles are shown as black for normal donor samples (Normal, N = 4,), light grey for sickle cell donor samples (SCD, N = 5) and dark grey for an additional set of normal donor samples (Normal = 5, SCD1) samples. The second set of Normal samples are from study SCD1 and were mobilized with G-CSF and plerixafor and generated using the large-scale process. The black line marks the DNA sequence of the gRNA targeting domain of the gRNA of RNP32 (i.e., CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)), and short black line marks the distal CCAAT box (appearing reverse complemented here).

[0141] **Fig. 59** shows the percentage of detected indels as a function of the deletion size (negative values on the x axis) or insertion (positive values in the x axis). The vertical line separates insertions and deletions. The highest peak is the deletion of size 18 corresponding to the indel 159_-18_NA. Profiles are shown as black for normal donor samples (Normal, N = 4), light grey for sickle cell donor samples (SCD, N = 5) and dark grey for an additional set of normal donor samples (Normal = 5, SCD1) samples. The second set of Normal samples are from study SCD1 and were mobilized with G-CSF and plerixafor and generated using the large-scale process.

[0142] **Fig. 60** shows the average percentages for indel lengths between 50 and -50. Positive values indicate insertions. Negative values indicate deletions.

[0143] **Fig. 61** shows the percentage of bases deleted along the target region in samples edited with RNP32. Profiles are shown as black for normal donor samples (Normal, N = 4), light grey for sickle cell donor samples (SCD, N = 5) and dark grey for an additional set of normal donor samples (Normal = 5, SCD1) samples. The second set of Normal samples are from study SCD1 and were mobilized with G-CSF and plerixafor and generated using the large-scale process. The target region DNA sequence is shown on the x axis. The black line marks the DNA sequence of the gRNA targeting domain of the gRNA of RNP32 (i.e., CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)), and short black line marks the distal CCAAT box (appearing reverse complemented here). Vertical dashed grey line represents the cut site, taken as the middle of the 5' overhang (see also **Fig. 55A**).

[0144] **Fig. 62** shows the normalized profiles (each with a maximum value of 1) for the detected deletions at each base on the target region in samples edited with RNP32. Profiles are shown as black for normal donor samples (Normal, N = 4), light grey for sickle cell donor samples (SC The second set of Normal samples are from study SCD1 and were mobilized with G-CSF and plerixafor and generated using the large-scale process. The target region sequence is shown in the x axis. The black line marks the DNA sequence of the gRNA targeting domain of the gRNA of RNP32 (i.e., CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)), and short black line marks the distal

CCAAT box (appearing reverse complemented here). Vertical dashed grey line represents the cut site, taken as the middle of the 5' overhang (see also **Fig. 55A**).

[0145] **Fig. 63** shows the number of indels detected across multiple samples. Shown are the counts for the number of indels as a function of the number of samples in which the indel was detected.

[0146] **Fig. 64** shows the indel reproducibility across all 14 samples as a function of their average percentage in indels. Shown are the average percentages in indel (y-axis) grouped by the number of samples in which the indel was detected (x-axis). Grey horizontal line at 0.22% marks the lowest percentage in indel of the top 55 indels.

[0147] **Figs. 65A** and **65B** show on-target indel levels of CD34+ cells based on RNP32 concentration at electroporation. Total on-target editing indel level was determined via Illumina sequencing at day 1 (**Fig. 65A**) and day 2 (**Fig. 65B**) after CD34+ cells were electroporated with RNP32 at the concentrations indicated. Data were compiled from eight independent experiments. A total of five RNP batches and four lots of normal donor CD34+ cells were tested across eleven concentrations of RNP32 ranging from 0.125 μM to 8 μM. Each dot represents an individual electroporation. Indel=insertions and/or deletions.

[0148] **Fig. 66** shows a summary of cell viability at day 1 electroporation with RNP32. Data marked with an * indicates RNP32 found to be poorly complexed based on differential scanning fluorimetry and thus data were excluded.

[0149] **Fig. 67** shows a summary of on-target indel levels at day 1 post-electroporation with RNP32. Data marked with an * indicates RNP32 found to be poorly complexed based on differential scanning fluorimetry and thus data were excluded.

[0150] **Fig. 68** shows a summary of on-target indels at day 2 post-electroporation. Data marked with an * indicates RNP32 found to be poorly complexed based on differential scanning fluorimetry and thus data were excluded.

[0151] **Figs. 69A** and **69B** shows the frequency of 4.9 kb fragment deletion in edited CD34+ cells at day 1 and day 2 post-electroporation. Frequency of 4.9kb fragment deletion between two RNP32 cut sites in CD34+ cells was assessed by ddPCR assays at Day 1 and Day 2 post electroporation with RNP at the concentration indicated (**Fig. 69A**). The 4.9 kb deletion to indel ratio was calculated by dividing levels of deletion with levels of indels for each sample (**Fig. 69B**). Data were compiled from three independent experiments. A total of two RNP batches and two lots of normal donor CD34+ cells were tested across eight concentrations of RNP32 ranging from 0.125 μM to 8 μM. Each dot represents an individual electroporation. Indel = insertions and/or deletions.

[0152] **Fig. 70** shows the summary of 4.9 kb fragment deletion frequency at day 1 and day 2 post-electroporation with RNP32.

[0153] **Figs. 71A** and **71B** show on-target indel levels of subpopulations of hematopoietic stem and progenitor cells. CD34+ cells were sorted 2 days post-electroporation into subpopulations of CMP,

MPP, and LT-HSC based on surface immunophenotype. **Fig. 71A** shows on-target indel levels determined via Illumina sequencing. The ratio of on-target indel levels within the phenotypic LT-HSC population to the total CD34+ cells is depicted in **Fig. 71B**. A ratio of 1 (dotted line) illustrates that the LT-HSCs have the same on-target indel levels as total CD34+ cells. CMP=common myeloid progenitors; Indel=insertions and/or deletions; LT-HSC=long-term hematopoietic stem cells; MPP=multipotent progenitors.

[0154] **Fig. 72** shows a summary of on-target indel levels in total CD34+ cells and sorted subpopulations at day 2 post-electroporation. CMP=common myeloid progenitors; Indel=insertions and/or deletions; LT-HSC=long-term hematopoietic stem cells; MPP=multipotent progenitors.

[0155] **Figs. 73A** and **73B** show the frequency of 4.9 kb fragment deletion in total CD34+ cells and sorted hematopoietic stem and progenitor cell subpopulations. CD34+ cells were sorted 2 days post electroporation into subpopulations of CMP, MPP, and LT HSC based on surface immunophenotype. **Fig. 73A** shows the frequency of 4.9 kb deletion determined via ddPCR. The 4.9 kb deletion to indel ratio was calculated by dividing levels of deletion with levels of indels for each sample (**Fig. 73B**). A lower ratio depicts a lower propensity of 4.9kb deletions to occur within a population. Multiple comparison was performed using Friedman test. *p=0.01, ****p <0.0001. CMP=common myeloid progenitors; Indel=insertions and/or deletions; LT HSC=long-term hematopoietic stem cells; MPP=multipotent progenitors.

[0156] **Fig. 74** shows a summary of 4.9 kb fragment deletion frequency in total CD34+ cells and sorted subpopulations at day 2 post-electroporation with RNP32. CD34+=cluster of differentiation 34; CMP=common myeloid progenitors; Indel=insertions and/or deletions; LT-HSC=long-term hematopoietic stem cells; MPP=multipotent progenitors; RNP=ribonucleoprotein.

DETAILED DESCRIPTION

Definitions and Abbreviations

[0157] Unless otherwise specified, each of the following terms has the meaning associated with it in this section.

[0158] 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.

[0159] The conjunctions “or” and “and/or” are used interchangeably as non-exclusive disjunctions.

[0160] “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.

[0161] 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.

[0162] The term "booster element" refers to an element which, when co-delivered with a ribonucleoprotein (RNP) complex comprising a gRNA complexed to an RNA-guided nuclease ("gRNA-nuclease-RNP"), increases editing of a target nucleic acid compared with editing of the target nucleic acid without the booster element. In certain embodiments, co-delivery may be sequential or simultaneous. In certain embodiments, a booster element may be an RNP complex comprised of a dead guide RNA complexed with a WT Cas9 protein, a Cas9 nickase protein (*e.g.,* Cas9 D10A protein), or an enzymatically inactive Cas9 (eiCas9) protein. In certain embodiments, a booster element may be an RNP complex comprised of a guide RNA complexed with a Cas9 nickase protein (*e.g.,* Cas9 D10A protein) or an enzymatically inactive Cas9 (eiCas9) protein. In certain embodiments, a booster element may be a single-or double stranded donor template DNA. In certain embodiments, one or more booster elements may be codelivered with a gRNA-nuclease-RNP to increase editing of a target nucleic acid. In certain embodiments, a booster element may be co-delivered with an RNP comprising a gRNA complexed to a Cpf1 molecule ("gRNA-Cpf1-RNP") to increase editing of a target nucleic acid.

[0163] "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.

[0164] 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.

[0165] "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.

[0166] 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.

[0167] "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.

[0168] "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.

[0169] Unless indicated otherwise, the term "HDR" as used herein encompasses both canonical HDR and alt-HDR.

[0170] "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).

[0171] "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.

[0172] "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.

[0173] "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.

[0174] "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.

[0175] 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 guide RNA 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 may include a booster element. 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.

[0176] 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.

[0177] Conventional IUPAC notation is used in nucleotide sequences presented herein, as shown in **Table 1**, below (*see also* Cornish-Bowden A, Nucleic Acids Res. 1985 May 10; 13(9):3021-30, 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 or Uracil
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

[0178] 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.

[0179] 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 other 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, 11nt deletion, 13 nt deletion, and 18 nt deletion, and -117 G>A alteration. As used herein, the terms "CCAAT box" and "CAAT box" can be used interchangeably.

[0180] 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.

[0181] 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.

[0182] 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.

[0183] 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.

[0184] 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.

[0185] 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.

[0186] 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.

[0187] 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.

[0188] Where ranges are provided herein, endpoints are included. Furthermore, it is to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values that are expressed as ranges can assume any specific value within the stated ranges in different embodiments of the invention, to the tenth of the unit of the lower limit of the range, unless the context clearly dictates otherwise. It is also to be understood that unless otherwise indicated or otherwise evident from the context and/or the understanding of one of ordinary skill in the art, values expressed as ranges can assume any subrange within the given range, wherein the endpoints of the subrange are expressed to the same degree of accuracy as the tenth of the unit of the lower limit of the range.

Overview

[0189] 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 Aγ and Gγ subunits of hemoglobin, respectively. In certain embodiments, increased expression of one or more γ–globin genes (e.g., HBG1, HBG2) using the methods provided herein results in preferential formation of HbF over HbA and/or increased HbF levels as a percentage of total hemoglobin. In certain embodiments, the disclosure generally relates to the use of RNP complexes comprising a gRNA complexed to a Cpf1 molecule. In certain embodiments, the gRNA may be unmodified or modified, the Cpf1 molecule may be a wild-type Cpf1 protein or a modified Cpf1 protein. In certain embodiments, the gRNA may comprise a sequence set forth in **Table 6**, **Table 12**, or **Table 13**. In certain embodiments, a modified Cpf1 may be encoded by a sequence set forth in SEQ ID NOs:1000,

1001, 1008-1018, 1032, 1035-39, 1094-1097, 1107-09 (Cpf1 polypeptide sequences) or SEQ ID NOs:1019-1021, 1110-17 (Cpf1 polynucleotide sequences). In certain embodiments, the RNP complex may comprise an RNP complex set forth in **Table 15**. For example, the RNP complex may include a gRNA comprising the sequence set forth in SEQ ID NO:1051 and a modified Cpf1 protein encoded by the sequence set forth in SEQ ID NO:1097 (RNP32, **Table 15**).

[0190] It has previously been shown that patients with the condition Hereditary Persistence of Fetal Hemoglobin (HPFH) contain mutations in an γ–globin regulatory element that results in fetal γ–globin expression throughout life, rather than being repressed around the time of birth (Martyn 2017). This results in elevated fetal hemoglobin (HbF) expression. HPFH mutations may be deletional or non-deletional (e.g., point mutations). Subjects with HPFH exhibit lifelong expression of HbF, i.e., they do not undergo or undergo only partial globin switching, with no symptoms of anemia.

[0191] HbF expression can be induced through point mutations in an γ–globin regulatory element that is associated with a naturally occurring HPFH variant, including, for example, HBG1 c.-114 C>T; c.-117 G>A; c.-158 C>T; c.-167 C>T; c.-170 G>A; c.-175 T>G; c.-175 T>C; c.-195 C>G; c.-196 C>T; c.-197 C>T; c.-198 T>C; c.-201 C>T; c.-202 C>T; c.-211 C>T, c.-251 T>C; or c.-499 T>A; or HBG2 c.-109 G>T; c.-110 A>C; c.-114 C>A; c.-114 C>T; c.-114 C>G; c.-157 C>T; c.-158 C>T; c.-167 C>T; c.-167 C>A; c.-175 T>C; c.-197 C>T; c.-200+C; c.-202 C>G; c.-211 C>T; c.-228 T>C; c.-255 C>G; c.-309 A>G; c.-369 C>G; or c.-567 T>G.

[0192] Naturally occurring mutations at the distal CCAAT box motif found within the promoter of the *HBG1* and/or *HBG2* genes (i.e., HBG1/2 c.-111 to -115) have also been shown to result in continued γ–globin expression and the HPFH condition. It is thought that alteration (mutation or deletion) of the CCAAT box may disrupt the binding of one or more transcriptional repressors, resulting in continued expression of the γ–globin gene and elevated HbF expression (Martyn 2017). For example, a naturally occurring 13 base pair del c.-114 to -102 ("13 nt deletion") has been shown to be associated with elevated levels of HbF (Martyn 2017). The distal CCAAT box likely overlaps with the binding motifs within and surrounding the CCAAT box of negative regulatory transcription factors that are expressed in adulthood and repress *HBG* (Martyn 2017).

[0193] A gene editing strategy disclosed herein is to increase HbF expression by disrupting one or more nucleotides in the distal CCAAT box and/or surrounding the distal CCAAT box. In certain embodiments, the "CCAAT box target region" may be 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). In other embodiments, the "CCAAT box target region" may be 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). Unique, non-naturally occurring alterations of the CCAAT box target region are disclosed herein that induce *HBG*

expression including, without limitation, HBG del c. -104 to -121 ("18 nt deletion"), HBG del c.-105 to -115 ("11 nt deletion"), HBG del c.-112 to -115 ("4 nt deletion"), and HBG del c.-116 ("1 nt deletion"). In certain embodiments, genome editing systems disclosed herein may be used to introduce alterations into the CCAAT box target region of *HBG1* and/or *HBG2*. In certain embodiments, the genome editing systems may include one or more of a DNA donor template that encodes an alteration (such as a deletion, insertion, or mutation) in the CCAAT box target region. In certain embodiments, the alterations may be non-naturally occurring alterations or naturally occurring alterations. In certain embodiments, the donor templates may encode the 1 nt deletion, 4 nt deletion, 11 nt deletion, 13 nt deletion, 18 nt deletion, or c.-117 G>A alteration. In certain embodiments, the genome editing systems may include an RNA guided nuclease including a Cas9, modified Cas 9, a Cpf1, or modified Cpf1. In certain embodiments, the genome editing systems may include an RNP comprising a gRNA and a Cpf1 molecule. In certain embodiments, a gRNA may be unmodified or modified, the Cpf1 molecule may be a wild-type Cpf1 protein or a modified Cpf1 protein, or a combination thereof. In certain embodiments, the gRNA may comprise a sequence set forth in **Table 6**, **Table 12**, or **Table 13**. In certain embodiments, a modified Cpf1 may be encoded by a sequence set forth in SEQ ID NOs:1000, 1001, 1008-1018, 1032, 1035-39, 1094-1097, 1107-09 (Cpf1 polypeptide sequences) or SEQ ID NOs:1019-1021, 1110-17 (Cpf1 polynucleotide sequences). In certain embodiments, the RNP complex may comprise an RNP complex set forth in **Table 15**. For example, the RNP complex may include a gRNA comprising the sequence set forth in SEQ ID NO:1051 and a modified Cpf1 protein encoded by the sequence set forth in SEQ ID NO:1097 (RNP32, **Table 15**).

[0194] The genome editing systems of this disclosure can include an RNA-guided nuclease such as 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.

[0195] A variety of approaches to the introduction of mutations into the CCAAT box target region, 13 nt target region, and/or proximal HBG1/2 promoter target sequence 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, and/or proximal HBG1/2 promoter target sequence, 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 and/or 13 nt target region.

[0196] 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 Aγ or Gγ 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.

[0197] In certain embodiments, alterations that result in induction of Aγ and/or Gγ 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 CCAAT box 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 CCAAT box 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 CCAAT box target region or a region proximate thereto. Examples of suitable gRNAs and gRNA targeting domains directed to the CCAAT box target region of *HBG1* and/or *HBG2* or proximate thereto for use in the embodiments disclosed herein include those set forth herein.

[0198] In certain embodiments, alterations that result in induction of Aγ and/or Gγ 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 gRNAs and gRNA 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 those set forth herein.

[0199] 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.

[0200] 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.

[0201] 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).

[0202] 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, the 13 nt target region, and/or proximal HBG1/2 promoter target sequence. 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.

[0203] In certain embodiments, CD34+ hematopoietic stem and progenitor cells (HSPCs) that have been edited using the genome editing methods disclosed herein may be used for the treatment of a hemoglobinopathy in a subject in need thereof. In certain embodiments, the hemoglobinopathy may be severe sickle cell disease (SCD) or thalassemia, such as β-thalassemia, δ-thalassemia, or β/δ-thalassemia. In certain embodiments, an exemplary protocol for treatment of a hemoglobinopathy

may include harvesting CD34+ HSPCs from a subject in need thereof, *ex vivo* editing of the autologous CD34+ HSPCs using the genome editing methods disclosed herein, followed by reinfusion of the edited autologous CD34+ HSPCs into the subject. In certain embodiments, treatment with edited autologous CD34+ HSPCs may result in increased HbF induction.

[0204] Prior to harvesting CD34+ HSPCs, in certain embodiments, a subject may discontinue treatment with hydroxyurea, if applicable, and receive blood transfusions to maintain sufficient hemoglobin (Hb) levels. In certain embodiments, a subject may be administered intravenous plerixafor (*e.g.*, 0.24 mg/kg) to mobilize CD34+ HSPCs from bone marrow into peripheral blood. In certain embodiments, a subject may undergo one or more leukapheresis cycles (*e.g.*, approximately one month between cycles, with one cycle defined as two plerixafor-mobilized leukapheresis collections performed on consecutive days). In certain embodiments, the number of leukapheresis cycles performed for a subject may be the number required to achieve a dose of edited autologous CD34+ HSPCs (*e.g.*, $\geq 2 \times 10^6$ cells/kg, $\geq 3 \times 10^6$ cells/kg, $\geq 4 \times 10^6$ cells/kg, $\geq 5 \times 10^6$ cells/kg, 2×10^6 cells/kg to 3×10^6 cells/kg, 3×10^6 cells/kg to 4×10^6 cells/kg, 4×10^6 cells/kg to 5×10^6 cells/kg) to be reinfused back into the subject, along with a dose of unedited autologous CD34+ HSPCs/kg for backup storage (*e.g.*, $\geq 1.5 \times 10^6$ cells/kg). In certain embodiments, the CD34+ HSPCs harvested from the subject may be edited using any of the genome editing methods discussed herein. In certain embodiments, any one or more of the gRNAs and one or more of the RNA-guided nucleases disclosed herein may be used in the genome editing methods.

[0205] In certain embodiments, the treatment may include an autologous stem cell transplant. In certain embodiments, a subject may undergo myeloablative conditioning with busulfan conditioning (*e.g.*, dose-adjusted based on first-dose pharmacokinetic analysis, with a test dose of 1 mg/kg). In certain embodiments, conditioning may occur for four consecutive days. In certain embodiments, following a three-day busulfan washout period, edited autologous CD34+ HSPCs (*e.g.*, $\geq 2 \times 10^6$ cells/kg, $\geq 3 \times 10^6$ cells/kg, $\geq 4 \times 10^6$ cells/kg, $\geq 5 \times 10^6$ cells/kg, 2×10^6 cells/kg to 3×10^6 cells/kg, 3×10^6 cells/kg to 4×10^6 cells/kg, 4×10^6 cells/kg to 5×10^6 cells/kg) may be reinfused into the subject (*e.g.*, into peripheral blood). In certain embodiments, the edited autologous CD34+ HSPCs may be manufactured and cryopreserved for a particular subject. In certain embodiments, a subject may attain neutrophil engraftment following a sequential myeloablative conditioning regimen and infusion of edited autologous CD34+ cells. Neutrophil engraftment may be defined as three consecutive measurements of ANC $\geq 0.5 \times 10^9$/L.

[0206] 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-napthoflavone, and CH-223191, and/or an innate immune response antagonist such as cyclosporin A, dexamethasone, reservatrol, 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.

[0207] 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.

[0208] Functionally, alteration of the CCAAT box target region, the 13 nt target region, and/or proximal HBG1/2 promoter target sequence using the compositions, methods and genome editing systems of this disclosure results in significant induction, among hemoglobin-expressing cells, of Aγ and/or Gγ 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 Aγ and/or Gγ 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, and/or proximal HBG1/2 promoter target sequence (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, and/or proximal HBG1/2 promoter target sequence.

[0209] 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 in 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 Aγ and Gγ globin chains as is known in the art.

[0210] It should be noted that the rate at which the CCAAT box target region (e.g., 18 nt, 11 nt, 4 nt, 1 nt, c.-117 G>A target regions), 13 nt target region, and/or proximal HBG1/2 promoter target sequence 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. In certain embodiments, donor templates for use in targeting the 18 nt target region may include, without limitation, donor templates encoding alterations (e.g., deletions) of HBG1 c.-104 to -121, HBG2 c.-104 to -121, or a combination thereof. Exemplary full-length donor templates encoding deletions such as c.-104 to -121 include SEQ ID NOs:974 and 975. In certain embodiments, donor templates for use in targeting the 11 nt target region may include, without limitation, donor templates encoding alterations (e.g., deletions) of HBG1 c.-105 to -115, HBG2 c.-105 to -115, or a combination thereof. Exemplary full-length donor templates encoding deletions such as c.-105 to -115 include SEQ ID NOs:976 and 978. In certain embodiments, donor templates for use in targeting the 4 nt target region may include, without limitation, donor templates encoding alterations (e.g., deletions) of HBG1 c.-112 to -115, HBG2 c.-112 to -115, or a combination thereof. Exemplary full-length donor templates encoding deletions such as c.-112 to -115 include SEQ ID NOs:984-995. In certain embodiments, donor templates for use in targeting the 1 nt target region may include, without limitation, donor templates encoding alterations (e.g., deletions) of HBG1 c.-116, HBG2 c.-116, or a combination thereof. Exemplary full-length donor templates encoding deletions such as c.-116 include SEQ ID NOs:982 and 983. In certain embodiments, donor templates for use in targeting the c.-117 G>A target region may include, without limitation, donor templates encoding alterations (e.g., deletions) of HBG1 c.-117 G>A, HBG2 c.-117 G>A, or a combination thereof. Exemplary full-length donor templates encoding deletions such as c.-117 G>A include SEQ ID NOs:980 and 981. In certain embodiments, the donor template may be a positive strand or a negative strand.

[0211] 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).

[0212] The embodiments described herein may be used in all classes of vertebrate including, but not limited to, primates, mice, rats, rabbits, pigs, dogs, and cats.

[0213] 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.

RNA-guided helicases, guide RNAs and dead guide RNAs

[0214] 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 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.

[0215] 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, and/or proximal HBG1/2 promoter target sequence). 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, and/or proximal HBG1/2 promoter target sequence).

[0216] 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. 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*).

[0217] 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

[0218] 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 guide RNA (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.

[0219] 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.

[0220] 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.

[0221] 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, 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.

[0222] 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* 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*.

[0223] 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 & Maizels 2014 (describing Alt-HDR); Frit 2014 (describing Alt-NHEJ); Iyama & Wilson 2013 (describing canonical HDR and NHEJ pathways generally)).

[0224] 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.

[0225] 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

[0226] 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 Cpf1 molecule 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 (see, e.g., Briner 2014, which is incorporated by reference; 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.

[0227] In bacteria and archea, 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).

[0228] 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 2013, 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.

[0229] 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 2014; 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.

[0230] 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.

[0231] 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 Franciscella 1") is a recently discovered RNA-guided nuclease that does not require a tracrRNA to function. (Zetsche 2015, 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). Exemplary targeting domains of Cpf1 gRNAs are set forth in **Table 6** and **Table 12**.

[0232] 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.

[0233] 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.

gRNA design

[0234] 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). Each of these references is 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.

[0235] Targeting domain sequences of gRNAs that were designed to target disruption of the CCAAT box target region include, but are not limited to, SEQ ID NO:1002. In certain embodiments, gRNAs comprising the sequence set forth in SEQ ID NO:1002 may be complexed with a Cpf1 protein or modified Cpf1 protein to generate alterations at the CCAAT box target region. In certain

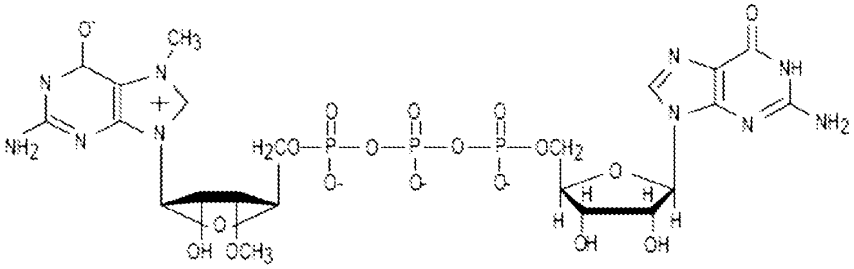
embodiments, gRNAs comprising any of the Cpf1 gRNAs set forth in **Table 9**, **Table 12,** or **Table 13** may be complexed with a Cpf1 protein or modified Cpf1 protein forming an RNP ("gRNA-Cpf1-RNP") to generate alterations at the CCAAT box target region. In certain embodiments, the modified Cpf1 protein may be His-AsCpf1-nNLS (SEQ ID NO: 1000) or His-AsCpf1-sNLS-sNLS (SEQ ID NO:1001). In certain embodiments, the Cpf1 molecule of the gRNA-Cpf1-RNP may be encoded by a sequence set forth in SEQ ID NOs:1000, 1001, 1008-1018, 1032, 1035-39 (Cpf1 polypeptide sequences) or SEQ ID NOs:1019-1021 (Cpf1 polynucleotide sequences).

gRNA modifications

[0236] 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.

[0237] 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.

[0238] 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:



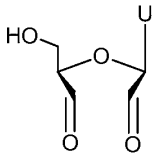
The cap or cap analog can be included during either chemical synthesis or *in vitro* transcription of the gRNA.

[0239] 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.

[0240] 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.

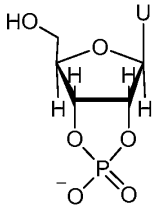
[0241] It should be noted that the modifications described herein can be combined in any suitable manner, e.g. a gRNA, whether transcribed *in vivo* from a DNA vector, or *in vitro* transcribed gRNA, can include either or both of a 5' cap structure or cap analog and a 3' polyA tract.

[0242] 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.

[0243] 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.

[0244] 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.

[0245] 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., NH_2; 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 phosphorothioate (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.

[0246] 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., NH_2; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino) and aminoalkoxy or $O(CH_2)_n$-amino (wherein amino can be, e.g., NH_2; alkylamino, dialkylamino, heterocyclyl, arylamino, diarylamino, heteroarylamino, or diheteroarylamino, ethylenediamine, or polyamino).

[0247] 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')).

[0248] 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.

[0249] 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.

[0250] In certain embodiments, gRNAs as used herein may be modified or unmodified gRNAs. In certain embodiments, a gRNA may include one or more modifications. In certain embodiments, the one or more modifications may include a phosphorothioate linkage modification, a phosphorodithioate (PS2) linkage modification, a 2'-O-methyl modification, or combinations thereof. In certain embodiments, the one or more modifications may be at the 5' end of the gRNA, at the 3' end of the gRNA, or combinations thereof.

[0251] In certain embodiments, a gRNA modification may comprise one or more phosphorodithioate (PS2) linkage modifications.

[0252] In some embodiments, a gRNA used herein includes one or more or a stretch of deoxyribonucleic acid (DNA) bases, also referred to herein as a "DNA extension." In some embodiments, a gRNA used herein includes a DNA extension at the 5' end of the gRNA, the 3' end of the gRNA, or a combination thereof. In certain embodiments, the DNA extension may be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 DNA bases long. For example, in certain embodiments, the DNA extension may be 1, 2, 3, 4, 5, 10, 15, 20, or 25 DNA bases long. In certain embodiments, the DNA extension may include one or more DNA bases selected from adenine (A), guanine (G), cytosine (C), or thymine (T). In certain embodiments, the DNA extension includes the same DNA bases. For example, the DNA extension may include a stretch of adenine (A) bases. In certain embodiments, the DNA extension may include a stretch of thymine (T) bases. In certain embodiments, the DNA extension includes a combination of different DNA bases. In certain embodiments, a DNA extension may comprise a sequence set forth in **Table 18**. For example, a DNA extension may comprise a sequence set forth in SEQ ID NOs:1235-1250. In certain embodiments, a gRNA used herein includes a DNA extension as well as one or more phosphorothioate linkage modifications, one or more phosphorodithioate (PS2) linkage modifications, one or more 2'-O-methyl modifications, or combinations thereof. In certain embodiments, the one or more modifications may be at the 5' end of the gRNA, at the 3' end of the gRNA, or combinations thereof. In certain

embodiments, a gRNA including a DNA extension may comprise a sequence set forth in **Table 13** that includes a DNA extension. In a particular embodiment, a gRNA including a DNA extension may comprise the sequence set forth in SEQ ID NO:1051. In certain embodiments, a gRNA including a DNA extension may comprise a sequence selected from the group consisting of SEQ ID NOs:1046-1060, 1067, 1068, 1074, 1075, 1078, 1081-1084, 1086-1087, 1089-1090, 1092-1093, 1098-1102, and 1106. Without wishing to be bound by theory, it is contemplated that any DNA extension may be used herein, so long as it does not hybridize to the target nucleic acid being targeted by the gRNA and it also exhibits an increase in editing at the target nucleic acid site relative to a gRNA which does not include such a DNA extension.

[0253] In some embodiments, a gRNA used herein includes one or more or a stretch of ribonucleic acid (RNA) bases, also referred to herein as an "RNA extension." In some embodiments, a gRNA used herein includes an RNA extension at the 5' end of the gRNA, the 3' end of the gRNA, or a combination thereof. In certain embodiments, the RNA extension may be 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, or 100 RNA bases long. For example, in certain embodiments, the RNA extension may be 1, 2, 3, 4, 5, 10, 15, 20, or 25 RNA bases long. In certain embodiments, the RNA extension may include one or more RNA bases selected from adenine (rA), guanine (rG), cytosine (rC), or uracil (rU), in which the "r" represents RNA, 2'-hydroxy. In certain embodiments, the RNA extension includes the same RNA bases. For example, the RNA extension may include a stretch of adenine (rA) bases. In certain embodiments, the RNA extension includes a combination of different RNA bases. In certain embodiments, an RNA extension may comprise a sequence set forth in **Table 18**. For example, an RNA extension may comprise a sequence set forth in 1231-1234, 1251-1253. In certain embodiments, a gRNA used herein includes an RNA extension as well as one or more phosphorothioate linkage modifications, one or more phosphorodithioate (PS2) linkage modifications, one or more 2'-O-methyl modifications, or combinations thereof. In certain embodiments, the one or more modifications may be at the 5' end of the gRNA, at the 3' end of the gRNA, or combinations thereof. In certain embodiments, a gRNA including a RNA extension may comprise a sequence set forth in **Table 13** that includes an RNA extension. gRNAs including an RNA extension at the 5' end of the gRNA may comprise a sequence selected from the group consisting of SEQ ID NOs:1042-1045, 1103-1105. gRNAs including an RNA extension at the 3' end of the gRNA may comprise a sequence selected from the group consisting of SEQ ID NOs:1070-1075, 1079, 1081, 1098-1100.

[0254] It is contemplated that gRNAs used herein may also include an RNA extension and a DNA extension. In certain embodiments, the RNA extension and DNA extension may both be at the 5' end of the gRNA, the 3' end of the gRNA, or a combination thereof. In certain embodiments, the RNA extension is at the 5' end of the gRNA and the DNA extension is at the 3' end of the gRNA. In certain embodiments, the RNA extension is at the 3' end of the gRNA and the DNA extension is at the 5' end of the gRNA.

[0255] In some embodiments, a gRNA which includes both a phosphorothioate modification at the 3' end as well as a DNA extension at the 5' end is complexed with a RNA-guided nuclease, e.g., Cpf1, to form an RNP, which is then employed to edit a hematopoietic stem cell (HSC) or a CD34+ cell *ex vivo* (i.e., outside the body of a subject from whom such a cell is derived), at the HBG locus.

[0256] An example of a gRNA as used herein comprises the sequence set forth in SEQ ID NO:1051.

Dead gRNA molecules

[0257] 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, for example, **Tables 8**, **9**, or **13** 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) to be edited. 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). 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.

[0258] 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.

[0259] As used herein, the terms "dead gRNA" and "truncated gRNA" are used interchangeably.

RNA-guided nucleases

[0260] RNA-guided nucleases according to the present disclosure include, but are not limited to, naturally-occurring Class 2 CRISPR nucleases such as Cpf1, and Cas9, as well as other nucleases derived or obtained therefrom. 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.

[0261] 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. For example, in certain embodiments, the RNA-guided nuclease may be Cas-Φ (Pausch 2020).

[0262] 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. Cpf1, on the other hand, generally recognizes PAM sequences that are 5' of the protospacer.

[0263] 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.

[0264] 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 and in Ran & Hsu 2013, incorporated by reference herein), or that do not cut at all.

Cas9

[0265] 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).

[0266] 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.

[0267] 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 HNN 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).

[0268] 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, and 14.

Cpf1

[0269] 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 by 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.

[0270] 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.

[0271] In certain embodiments, a Cpf1 protein may be a modified Cpf1 protein. In certain embodiments, a modified Cpf1 protein may include one or more modifications. In certain embodiments the modifications may be, without limitation, one or more mutations in a Cpf1 nucleotide sequence or Cpf1 amino acid sequence, one or more additional sequences such as a His tag or a nuclear localization signal (NLS), or a combination thereof. In certain embodiments, a modified Cpf1 may also be referred to herein as a Cpf1 variant.

[0272] In certain embodiments, the Cpf1 protein may be derived from a Cpf1 protein selected from the group consisting of *Acidaminococcus sp.* strain BV3L6 Cpf1 protein (AsCpf1), *Lachnospiraceae bacterium* ND2006 Cpf1 protein (LbCpf1), and *Lachnospiraceae bacterium* MA2020 (Lb2Cpf1). In certain embodiments, the Cpf1 protein may comprise a sequence selected from the group consisting of SEQ ID NOs:1016-1018, having the codon-optimized nucleic acid sequences of SEQ ID NOs:1019-1021, respectively.

[0273] In certain embodiments, the modified Cpf1 protein may comprise a nuclear localization signal (NLS). For example, but not by way of limitation, NLS sequences useful in connection with the methods and compositions disclosed herein will comprise an amino acid sequence capable of facilitating protein import into the cell nucleus. NLS sequences useful in connection with the methods and compositions disclosed herein are known in the art. Examples of such NLS sequences include the nucleoplasmin NLS having the amino acid sequence: KRPAATKKAGQAKKKK (SEQ ID NO:1006) and the simian virus 40 "SV40" NLS having the amino acid sequence PKKKRKV (SEQ ID NO:1007).

[0274] In certain embodiments, the NLS sequence of the modified Cpf1 protein is positioned at or near the C-terminus of the Cpf1 protein sequence. For example, but not by way of limitation, the modified Cpf1 protein can be selected from the following: His-AsCpf1-nNLS (SEQ ID NO:1000); His-AsCpf1-sNLS (SEQ ID NO:1008) and His-AsCpf1-sNLS-sNLS (SEQ ID NO:1001), where "His" refers to a six-histidine purification sequence, "AsCpf1" refers to the *Acidaminococcus sp.* Cpf1 protein sequence, "nNLS" refers to the nucleoplasmin NLS, and "sNLS" refers to the SV40 NLS. Additional permutations of the identity and C-terminal positions of NLS sequences, e.g., appending two or more nNLS sequences or combinations of nNLS and sNLS sequences (or other NLS sequences), as well as sequences with and without purification sequences, e.g., six-histidine sequences, are within the scope of the instantly disclosed subject matter.

[0275] In certain embodiments, the NLS sequence of the modified Cpf1 protein may be positioned at or near the N-terminus of the Cpf1 protein sequence. For example, but not by way of limitation, the modified Cpf1 protein may be selected from the following: His-sNLS-AsCpf1 (SEQ ID NO:1009), His-sNLS-sNLS-AsCpf1 (SEQ ID NO:1010), and sNLS-sNLS-AsCpf1 (SEQ ID NO:1011). Additional permutations of the identity and N-terminal positions of NLS sequences, e.g., appending two or more nNLS sequences or combinations of nNLS and sNLS sequences (or other NLS sequences), as well as sequences with and without purification sequences, e.g., six-histidine sequences, are within the scope of the instantly disclosed subject matter.

[0276] In certain embodiments, the modified Cpf1 protein may comprise NLS sequences positioned at or near both the N-terminus and C-terminus of the Cpf1 protein sequence. For example, but not by way of limitation, the modified Cpf1 protein may be selected from the following: His-sNLS-AsCpf1-sNLS (SEQ ID NO:1012) and His-sNLS-sNLS-AsCpf1-sNLS-sNLS (SEQ ID NO:1013). Additional permutations of the identity and N-terminal/C-terminal positions of NLS sequences, e.g., appending two or more nNLS sequences or combinations of nNLS and sNLS sequences (or other NLS sequences) to either the N-terminal/C-terminal positions, as well as sequences with and without purification sequences, e.g., six-histidine sequences, are within the scope of the instantly disclosed subject matter.

[0277] In certain embodiments, the modified Cpf1 protein may comprise an alteration (e.g., a deletion or substitution) at one or more cysteine residues of the Cpf1 protein sequence. For example, but not by way of limitation, modified Cpf1 protein may comprise an alteration at a position selected from the group consisting of: C65, C205, C334, C379, C608, C674, C1025, and C1248. In certain embodiments, the modified Cpf1 protein may comprise a substitution of one or more cysteine residues for a serine or alanine. In certain embodiments, the modified Cpf1 protein may comprise an alteration selected from the group consisting of: C65S, C205S, C334S, C379S, C608S, C674S, C1025S, and C1248S. In certain embodiments, the modified Cpf1 protein may comprise an alteration selected from the group consisting of: C65A, C205A, C334A, C379A, C608A, C674A, C1025A, and C1248A. In certain embodiments,

the modified Cpf1 protein may comprise alterations at positions C334 and C674 or C334, C379, and C674. In certain embodiments, the modified Cpf1 protein may comprise the following alterations: C334S and C674S, or C334S, C379S, and C674S. In certain embodiments, the modified Cpf1 protein may comprise the following alterations: C334A and C674A, or C334A, C379A, and C674A. In certain embodiments, the modified Cpf1 protein may comprise both one or more cysteine residue alteration as well as the introduction of one or more NLS sequences, e.g., His-AsCpf1-nNLS Cys-less (SEQ ID NO:1014) or His-AsCpf1-nNLS Cys-low (SEQ ID NO:1015). In various embodiments, the Cpf1 protein comprising a deletion or substitution in one or more cysteine residues exhibits reduced aggregation.

[0278] In certain embodiments, other modified Cpf1 proteins known in the art may be used with the methods and systems described herein. For example, in certain embodiments, the modified Cpf1 may be Cpf1 containing the mutation S542R/K548V/N552R ("Cpf1 RVR"). Cpf1 RVR has been shown to cleave target sites with a TATV PAM. In certain embodiments, the modified Cpf1 may be Cpf1 containing the mutation S542R/K607R ("Cpf1 RR"). Cpf1 RR has been shown to cleave target sites with a TYCV/CCCC PAM.

[0279] In some embodiments, a Cpf1 variant is used herein, wherein the Cpf1 variant comprises mutations at one or more residues of AsCpf1 (Acidaminococcus sp. BV3L6) selected from the group consisting of 11, 12, 13, 14, 15, 16, 17, 34, 36, 39, 40, 43, 46, 47, 50, 54, 57, 58, 111, 126, 127, 128, 129, 130, 131, 132, 133, 134, 135, 136, 157, 158, 159, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 532, 533, 534, 535, 536, 537, 538, 539, 540, 541, 542, 543, 544, 545, 546, 547, 548, 549, 550, 551, 552, 553, 554, 555, 556, 565, 566, 567, 568, 569, 570, 571, 572, 573, 574, 575, 592, 593, 594, 595, 596, 597, 598, 599, 600, 601, 602, 603, 604, 605, 606, 607, 608, 609, 610, 611, 612, 613, 614, 615, 616, 617, 618, 619, 620, 626, 627, 628, 629, 630, 631, 632, 633, 634, 635, 636, 637, 638, 642, 643, 644, 645, 646, 647, 648, 649, 651, 652, 653, 654, 655, 656, 676, 679, 680, 682, 683, 684, 685, 686, 687, 688, 689, 690, 691, 692, 693, 707, 711, 714, 715, 716, 717, 718, 719, 720, 721, 722, 739, 765, 768, 769, 773, 777, 778, 779, 780, 781, 782, 783, 784, 785, 786, 870, 871, 872, 873, 874, 875, 876, 877, 878, 879, 880, 881, 882, 883, 884, or 1048 or the corresponding position of an AsCpf1 orthologue, homologue, or variant.

[0280] In certain embodiments, a Cpf1 variant as used herein may include any of the Cpf1 proteins described in International Publication Number WO 2017/184768 A1 by Zhang et al. ("'768 Publication"), which is incorporated by reference herein.

[0281] In certain embodiments, a modified Cpf1 protein (also referred to as a Cpf1 variant) used herein may be encoded by any of the sequences set forth in SEQ ID NOs:1000, 1001, 1008-1018, 1032, 1035-39, 1094-1097, 1107-09 (Cpf1 polypeptide sequences) or SEQ ID NOs:1019-1021, 1110-17 (Cpf1 polynucleotide sequences). **Table 14** sets forth exemplary Cpf1 variant amino acid and nucleotide

sequences. These sequences are set forth in **Fig. 40**, which details the positions of six-histidine sequences (underlined letters) and NLS sequences (bolded letters). Additional permutations of the identity and N-terminal/C-terminal positions of NLS sequences, e.g., appending two or more nNLS sequences or combinations of nNLS and sNLS sequences (or other NLS sequences) to either the N-terminal/C-terminal positions, as well as sequences with and without purification sequences, e.g., six-histidine sequences, are within the scope of the instantly disclosed subject matter.

[0282] In certain embodiments, any of the Cpf1 proteins or modified Cpf1 proteins disclosed herein may be complexed with one or more gRNA comprising the targeting domain set forth in SEQ ID NOs 1002 and/or 1004 to alter a CCAAT box target region. In certain embodiments, any of the Cpf1 proteins or modified Cpf1 proteins disclosed herein may be complexed with one or more gRNA comprising a sequence set forth in **Table 12** or **Table 13**. In certain embodiments, the modified Cpf1 protein may be His-AsCpf1-nNLS (SEQ ID NO:1000) or His-AsCpf1-sNLS-sNLS (SEQ ID NO:1001). In certain embodiments, a modified Cpf1 protein used herein may be encoded by any of the sequences set forth in SEQ ID NOs:1000, 1001, 1008-1018, 1032, 1035-39, 1094-1097, 1107-09 (Cpf1 polypeptide sequences) or SEQ ID NOs:1019-1021, 1110-17 (Cpf1 polynucleotide sequences). In certain embodiments, the modified Cpf1 protein may comprise the sequence set forth in SEQ ID NO:1097.

[0283] In certain embodiments, the modified Cpf1 protein may include a Cpf1 variant described in Kleinstiver 2019. For example, without limitation, in certain embodiments, the modified Cpf1 protein may be enAsCas12a. In certain embodiments, the modified Cpf1 protein may cleave target sites with a TTTV PAM. In certain embodiments, the modified Cpf1 protein may cleave target sites with a NWYN PAM.

Modifications of RNA-guided nucleases

[0284] 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.

[0285] 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 & Hsu 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 as shown below (where C denotes the site of cleavage).

[0286] On the other hand, inactivation of a Cas9 HNH domain results in a nickase that cleaves the bottom or non-complementary strand.

[0287] Modifications of PAM specificity relative to naturally occurring Cas9 reference molecules has been described by Kleinstiver et al. for both *S. pyogenes* (Kleinstiver 2015a) and *S. aureus* (Kleinstiver 2015b). Kleinstiver et al. have also described modifications that improve the targeting fidelity of Cas9 (Kleinstiver 2016). Kleinstiver et al. have also described modifications of Cpf1 that provide increased activity and improved targeting ranges (Kleinstiver 2019). Each of these references is incorporated by reference herein.

[0288] RNA-guided nucleases have been split into two or more parts, as described by Zetsche 2015 and Fine 2015 (both incorporated by reference herein).

[0289] 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, incorporated by reference herein for all purposes.

[0290] 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.

[0291] 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.

Nucleic acids encoding RNA-guided nucleases

[0292] Nucleic acids encoding RNA-guided nucleases, e.g., Cas9, Cpf1 or functional fragments thereof, are provided herein. Exemplary nucleic acids encoding RNA-guided nucleases have been described previously (see, e.g., Cong 2013; Wang 2013; Mali 2013; Jinek 2012).

[0293] 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.

[0294] 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.

[0295] 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

[0296] 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)

[0297] 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.

[0298] 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.

[0299] Two non-limiting examples of DSF assay conditions are set forth below:

[0300] 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.

[0301] 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

[0302] 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.

[0303] 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.

[0304] 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.

[0305] 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 & Hsu 2013 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).

[0306] 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.

[0307] 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.

[0308] 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.

[0309] 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

[0310] 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. In certain embodiments, multiple gRNAs and an RNA-guided nuclease may be used in genome editing systems to introduce alterations (e.g., deletions, insertions) into the CCAAT box target region of *HBG1* and/or *HBG2*. In certain embodiments, the RNA-guided nuclease may be a Cpf1 or modified Cpf1.

Donor template design

[0311] 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.

[0312] 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].

[0313] 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.

[0314] Replacement sequences in donor templates have been described elsewhere, including in Cotta-Ramusino. 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.

[0315] 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 80-200 nucleotides (e.g., 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, or 200 nucleotides).

[0316] 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.

[0317] 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.

[0318] In certain embodiments, silent, non-pathogenic SNPs may be included in the ssODN donor template to allow for identification of a gene editing event.

[0319] 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.

[0320] A donor template or template nucleic acid, as that term is used herein, refers to a nucleic acid sequence which can be used in conjunction with an RNA nuclease molecule and one or more gRNA molecules to alter (e.g., delete, disrupt, or modify) a target DNA sequence. In certain embodiments, the template nucleic acid results in an alteration (e.g., deletion) at the CCAAT box target region of *HBG1* and/or *HBG2*. In certain embodiments, the alteration is a non-naturally occurring alteration. In certain embodiments, the non-naturally occurring alteration at the CCAAT box target region of

HBG1 and/or *HBG2* may comprise the 18 nt target region, the 11 nt target region, the 4 nt target region, or the 1 nt target region, or a combination thereof. In certain embodiments, the alteration is a naturally occurring alteration. In certain embodiments, the naturally occurring alteration at the CCAAT box target region of *HBG1* and/or *HBG2* may comprise the 13 nt target region, the c.-117 G>A target region, or a combination thereof. In certain embodiments, the template nucleic acid is an ssODN. In certain embodiments, the ssODN is a positive strand or a negative strand.

[0321] For example, a template nucleic acid for introducing the 18 nt deletion at the 18 nt target region (HBG1 c.-104 to -121, HBG2 c.-104 to -121, or a combination thereof) may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm, where the replacement sequence is 0 nucleotides or 0 bp. In certain embodiments, the 5' homology arm may be about 25 to about 200 nucleotides or more in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 18 nt target region. In certain embodiments, the 3' homology arm may be about 25 to about 200 nucleotides or more in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 3' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the 18 nt target region. In certain embodiments, the 5' and 3' homology arms are symmetrical in length. In certain embodiments, the 5' and 3' homology arms are asymmetrical in length. In certain embodiments, the template nucleic acid is an ssODN. In certain embodiments, the ssODN is a positive strand. In certain embodiments, the ssODN is a negative strand. In certain embodiments, the ssODN comprises, consists essentially of, or consists of SEQ ID NO:974 (OLI16409) or SEQ ID NO:975 (OLI16410).

[0322] In certain embodiments, a template nucleic acid for introducing the 11 nt deletion at the 11 nt target region (HBG1 c.-105 to -115, HBG2 c.-105 to -115, or a combination thereof) may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm, where the replacement sequence is 0 nucleotides or 0 bp. In certain embodiments, the 5' homology arm may be about 25 to about 200 nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 11 nt target region. In certain embodiments, the 3' homology arm may be about 25 to about 200 nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 3' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the 11 nt target region. In certain embodiments, the 5' and 3' homology arms are symmetrical in length. In certain embodiments, the 5' and 3' homology arms are asymmetrical in length. In certain embodiments, the template nucleic acid is an ssODN. In certain embodiments, the ssODN is a positive strand. In

certain embodiments, the ssODN is a negative strand. In certain embodiments, the ssODN comprises, consists essentially of, or consists of SEQ ID NO:976 (OLI16411) or SEQ ID NO:978 (OLI16413).

[0323] In certain embodiments, a template nucleic acid for introducing the 4 nt deletion at the 4 nt target region (HBG1 c.-112 to -115, HBG2 c.-112 to -115, or a combination thereof) may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm, where the replacement sequence is 0 nucleotides or 0 bp. In certain embodiments, the 5' homology arm may be about 25 to about 200 nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 4 nt target region. In certain embodiments, the 3' homology arm may be about 25 to about 200 nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 3' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the 4 nt target region. In certain embodiments, the 5' and 3' homology arms are symmetrical in length. In certain embodiments, the 5' and 3' homology arms are asymmetrical in length. In certain embodiments, the template nucleic acid is an ssODN. In certain embodiments, the ssODN is a positive strand. In certain embodiments, the ssODN is a negative strand. In certain embodiments, the ssODN comprises, consists essentially of, or consists of SEQ ID NO:984 (OLI16419), SEQ ID NO:985 (OLI16420), SEQ ID NO:986 (OLI16421), SEQ ID NO:987 (OLI16422), SEQ ID NO:988 (OLI16423), SEQ ID NO:989 (OLI16424), SEQ ID NO:990 (OLI16425), SEQ ID NO:991 (OLI16426), SEQ ID NO:992 (OLI16427), SEQ ID NO:993 (OLI16428), SEQ ID NO:994 (OLI16429), or SEQ ID NO:995 (OLI16430).

[0324] In certain embodiments, a template nucleic acid for introducing the 1 nt deletion at the 1 nt target region (HBG1 c.-116, HBG2 c.-116, or a combination thereof) may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm, where the replacement sequence is 0 nucleotides or 0 bp. In certain embodiments, the 5' homology arm may be about 25 to about 200 nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 1 nt target region. In certain embodiments, the 3' homology arm may be about 25 to about 200 nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 3' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the 1 nt target region. In certain embodiments, the 5' and 3' homology arms are symmetrical in length. In certain embodiments, the 5' and 3' homology arms are asymmetrical in length. In certain embodiments, the template nucleic acid is an ssODN. In certain embodiments, the ssODN is a positive strand. In

certain embodiments, the ssODN is a negative strand. In certain embodiments, the ssODN comprises, consists essentially of, or consists of SEQ ID NO:982 (OLI16417) or SEQ ID NO:983 (OLI16418).

[0325] In certain embodiments, the alteration at the CCAAT box target region recapitulates or is similar to a naturally occurring alteration, such as a 13 nt deletion. In certain embodiments, a template nucleic acid for introducing the 13 nt deletion at the 13 nt target region (HBG1 c.-116, HBG2 c.-116, or a combination thereof) may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm, where the replacement sequence is 0 nucleotides or 0 bp. In certain embodiments, the 5' homology arm may be about 25 to about 200 nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the 13 nt target region. In certain embodiments, the 3' homology arm may be about 25 to about 200 nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 3' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the 13 nt target region. In certain embodiments, the 5' and 3' homology arms are symmetrical in length. In certain embodiments, the 5' and 3' homology arms are asymmetrical in length. In certain embodiments, the template nucleic acid is an ssODN. In certain embodiments, the ssODN is a positive strand. In certain embodiments, the ssODN is a negative strand. In certain embodiments, the ssODN comprises, consists essentially of, or consists of SEQ ID NO:979 (OLI16414) or SEQ ID NO:977 (OLI16412).

[0326] In certain embodiments, the alteration at the CCAAT box target region recapitulates or is similar to a naturally occurring alteration, such as a substitution from G to A at the -117G>A target region. In certain embodiments, a template nucleic acid for introducing the -117G>A substitution at the -117G>A target region (HBG1 c.-117 G>A, HBG2 c.-117 G>A, or a combination thereof) may comprise a 5' homology arm, a replacement sequence, and a 3' homology arm, where the replacement sequence is 0 nucleotides or 0 bp. In certain embodiments, the 5' homology arm may be about 100 to about 200 nucleotides in length, e.g., at least about 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 5' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 5' of the -117G>A target region. In certain embodiments, the 3' homology arm may be about 25 to about 200 nucleotides in length, e.g., at least about 25, 50, 75, 100, 125, 150, 175, or 200 nucleotides in length. In certain embodiments, the 3' homology arm comprises about 50 to 100 bp, e.g., 55 to 95, 60 to 90, 70 to 90, or 80 to 90 bp, homology 3' of the -117G>A target region. In certain embodiments, the 5' and 3' homology arms are symmetrical in length. In certain embodiments, the 5' and 3' homology arms are asymmetrical in length. In certain embodiments, the template nucleic acid is an ssODN. In certain embodiments, the ssODN is a positive strand. In certain embodiments, the ssODN is a negative strand. In certain embodiments, the

ssODN comprises, consists essentially of, or consists of SEQ ID NO:980 (OLI16415) or SEQ ID NO:981 (OLI16416).

[0327] In certain embodiments, the 5' homology arm comprises a 5' phosphorothioate (PhTx) modification. In certain embodiments, the 3' homology arm comprises a 3' PhTx modification. In certain embodiments, the template nucleic acid comprises a 5' and 3' PhTx modification.

[0328] In certain embodiments, the ssODNs for introducing alterations (e.g., deletions) at the CCAAT box target region may be used in conjunction with an RNA nuclease and one or more gRNAs that target the CCAAT target region, for example, the gRNAs disclosed in **Table 6, Table 12**, **Table 13**.

Target cells

[0329] 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*.

[0330] 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. 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.

[0331] 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.

[0332] 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

[0333] 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 2** and **3** 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 2** 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 2

Genome Editing System 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 DNA		A first DNA or DNA vector encoding an RNA-guided nuclease and a donor template, and a second DNA or DNA vector encoding a gRNA
DNA RNA		A DNA or DNA vector encoding an RNA-guided nuclease and a donor template, and a gRNA
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.

[0334] **Table 3** 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 3

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
Viral	**Retrovirus**	NO	Stable	YES	RNA
	Lentivirus	YES	Stable	YES/NO with modifications	RNA
	Adenovirus	YES	Transient	NO	DNA
	Adeno-Associated Virus (AAV)	YES	Stable	NO	DNA

	Vaccinia Virus	YES	Very Transient	NO	DNA
	Herpes Simplex Virus	YES	Stable	NO	DNA
Non-Viral	**Cationic Liposomes**	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
	Polymeric Nanoparticles	YES	Transient	Depends on what is delivered	Nucleic Acids and Proteins
Biological Non-Viral Delivery Vehicles	**Attenuated Bacteria**	YES	Transient	NO	Nucleic Acids
	Engineered Bacteriophages	YES	Transient	NO	Nucleic Acids
	Mammalian Virus-like Particles	YES	Transient	NO	Nucleic Acids
	Biological liposomes: Erythrocyte Ghosts and Exosomes	YES	Transient	NO	Nucleic Acids

Nucleic acid-based delivery of genome editing systems

[0335] 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.

[0336] 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 3**, can also be used.

[0337] 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).

[0338] 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.

[0339] Nucleic acid vectors according to this disclosure include recombinant viral vectors. Exemplary viral vectors are set forth in **Table 3**, 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.

[0340] 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) nonparticles 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 4** and **Table 5** lists exemplary polymers for use in gene transfer and/or nanoparticle formulations.

Table 4: Lipids Used for Gene Transfer

Lipid	Abbreviation	Feature
1,2-Dioleoyl-sn-glycero-3-phosphatidylcholine	DOPC	Helper
1,2-Dioleoyl-sn-glycero-3-phosphatidylethanolamine	DOPE	Helper
Cholesterol		Helper
N-[1-(2,3-Dioleyloxy)propyl]*N,N,N*-trimethylammonium chloride	DOTMA	Cationic
1,2-Dioleoyloxy-3-trimethylammonium-propane	DOTAP	Cationic
Dioctadecylamidoglycylspermine	DOGS	Cationic
N-(3-Aminopropyl)-*N,N*-dimethyl-2,3-bis(dodecyloxy)-1-propanaminium bromide	GAP-DLRIE	Cationic

Cetyltrimethylammonium bromide	CTAB	Cationic
6-Lauroxyhexyl ornithinate	LHON	Cationic
1-(2,3-Dioleoyloxypropyl)-2,4,6-trimethylpyridinium	2Oc	Cationic
2,3-Dioleyloxy-*N*-[2(sperminecarboxamido-ethyl]-*N,N*-dimethyl-1-propanaminium trifluoroacetate	DOSPA	Cationic
1,2-Dioleyl-3-trimethylammonium-propane	DOPA	Cationic
N-(2-Hydroxyethyl)-*N,N*-dimethyl-2,3-bis(tetradecyloxy)-1-propanaminium bromide	MDRIE	Cationic
Dimyristooxypropyl dimethyl hydroxyethyl ammonium bromide	DMRI	Cationic
3β-[*N*-(*N',N'*-Dimethylaminoethane)-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
Dioctadecylamidoglicylspermidin	DSL	Cationic
rac-[(2,3-Dioctadecyloxypropyl)(2-hydroxyethyl)]-dimethylammonium chloride	CLIP-1	Cationic
rac-[2(2,3-Dihexadecyloxypropyl-oxymethyloxy)ethyl]trimethylammonium bromide	CLIP-6	Cationic
Ethyldimyristoylphosphatidylcholine	EDMPC	Cationic
1,2-Distearyloxy-*N,N*-dimethyl-3-aminopropane	DSDMA	Cationic
1,2-Dimyristoyl-trimethylammonium propane	DMTAP	Cationic
O,O'-Dimyristyl-N-lysyl aspartate	DMKE	Cationic
1,2-Distearoyl-sn-glycero-3-ethylphosphocholine	DSEPC	Cationic
N-Palmitoyl D-erythro-sphingosyl carbamoyl-spermine	CCS	Cationic
N-t-Butyl-*N*0-tetradecyl-3-tetradecylaminopropionamidine	diC14-amidine	Cationic
Octadecenolyoxy[ethyl-2-heptadecenyl-3 hydroxyethyl] imidazolinium chloride	DOTIM	Cationic
*N*1-Cholesteryloxycarbonyl-3,7-diazanonane-1,9-diamine	CDAN	Cationic
2-(3-[Bis(3-amino-propyl)-amino]propylamino)-*N*-ditetradecylcarbamoylme-ethyl-acetamide	RPR209120	Cationic
1,2-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 5: Polymers Used for Gene Transfer

Polymer	Abbreviation
Poly(ethylene)glycol	PEG
Polyethylenimine	PEI
Dithiobis(succinimidylpropionate)	DSP
Dimethyl-3,3'-dithiobispropionimidate	DTBP
Poly(ethylene imine) biscarbamate	PEIC
Poly(L-lysine)	PLL
Histidine modified PLL	
Poly(*N*-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(*N*-ethyl-4-vinylpyridinium bromide)	
Poly(phosphazene)s	PPZ
Poly(phosphoester)s	PPE
Poly(phosphoramidate)s	PPA
Poly(*N*-2-hydroxypropylmethacrylamide)	pHPMA
Poly (2-(dimethylamino)ethyl methacrylate)	pDMAEMA
Poly(2-aminoethyl propylene phosphate)	PPE-EA
Chitosan	
Galactosylated chitosan	
N-Dodacylated chitosan	
Histone	
Collagen	

Dextran-spermine	D-SPM

[0341] 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.

[0342] 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

[0343] 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.

[0344] *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

[0345] 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.

[0346] 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.

[0347] 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.

[0348] 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.

[0349] 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 pyrolidone), 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.

[0350] 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

[0351] 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 non-simultaneously. 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.

[0352] 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.

[0353] 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.

[0354] 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.

[0355] 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.

[0356] 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.

[0357] 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.

[0358] 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.

[0359] 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.

[0360] In certain embodiments, the second mode of delivery comprises a relatively transient element, e.g., an RNA or protein.

[0361] 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.

[0362] 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.

[0363] 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.

[0364] 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.

[0365] 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

[0366] The principles and embodiments described above are further illustrated by the non-limiting examples that follow:

Example 1: Cpf1 RNP containing gRNA targeting the CCAAT box supports gene editing in human hematopoietic stem/progenitor cells

[0367] A Cpf1 guide RNA, HBG1-1 (i.e., OLI13620 (**Table 6**)), was designed to target the HBG distal CCAAT box. To determine optimal nuclear localization signal configuration for *Acidaminococcus sp.* Cpf1 ("AsCpf1") delivery in CD34+ cells, HBG1-1 gRNA was complexed to two nuclear localization signal (NLS) variants of AsCpf1, namely His-AsCpf1-nNLS (SEQ ID NO: 1000) and His-AsCpf1-sNLS-sNLS (SEQ ID NO:1001). "His" refers to a six-histidine purification sequence, "AsCpf1" refers to the *Acidaminococcus sp.* Cpf1 protein sequence, "nNLS" refers to the nucleoplasmin NLS, and "sNLS" refers to the SV40 NLS.

[0368] Briefly, mPB CD34+ cells were pre-stimulated for 2 days with human cytokines in X-Vivo-10 and then electroporated with the RNPs at 5 µM or 20 µM. The genomic DNA was extracted three days post electroporation and next-generation sequencing was performed on the *HBG* PCR products. HBG1-1 gRNA complexed to either of the Cpf1 NLS variants tested ("His-AsCpf1-sNLS-sNLS_HBG1-1 RNP" or "His-AsCpf1-nNLS_HBG1-1 RNP"), supported editing of CD34+ cells at the 13 nt target site. His-AsCpf1-sNLS-sNLS_HBG1-1 RNP generated 60.6% edited alleles and His-AsCpf1-nNLS_HBG1-1 RNP generated 51.1% edited alleles at the highest dose tested (**Fig. 3A**).

Table 6: HBG1-1 gRNA sequence for targeting the CCAAT box in CD34$^+$ cells

gRNA ID	OLI-ID	Targeting domain sequence (RNA)	Targeting domain sequence (DNA)	Targeting domain sequence plus PAM (UUUG) (RNA)	Targeting domain sequence plus PAM (TTTG) (DNA)	Sense
HBG1-1	OLI13620	CCUUGUC AAGGCUA UUGGUC (SEQ ID NO:1002)	CCTTGTC AAGGCTA TTGGTC (SEQ ID NO:1003)	UUUGCCUUGU CAAGGCUAUU GGUC (SEQ ID NO:1004)	TTTGCCTTG TCAAGGCT ATTGGTC (SEQ ID NO:1005)	Antisense

Example 2: Co-delivery of Cpf1 RNP targeting the CCAAT box with ssODN donors supports gene editing in human hematopoietic stem/progenitor cells

[0369] RNP comprising HBG1-1 gRNA complexed to the His-AsCpf1-sNLS-sNLS variant ("His-AsCpf1-sNLS-sNLS_HBG1-1 RNP") were co-delivered by electroporation with single stranded oligodeoxynucleotide donor repair templates (ssODNs) to mPB CD34+ cells. OLI16430 and OLI16424 ssODNs were designed to "encode" a 4 nucleotide deletion and OLI16409 and OLI16410 ssODNs were designed to "encode" a 18 nucleotide deletion (**Table 7**). Both the 4 nt and 18 nt deletions disrupt the HBG distal CCAAT box and are associated with induction of *HBG* expression.

The ssODNs include 90 nucleotide-long homology arms flanking the encoded absent sequence to create perfect deletion. The ssODNs were modified to contain phosphorothioates (PhTx) at the 5' and 3' ends (OLI16430, OLI16424, OLI16409, and OLI16410, **Table 7**). Briefly, human adult mPB CD34+ cells pre-stimulated for two days in medium supplemented with human cytokines were electroporated with 5 µM RNP comprising the His-AsCpf1-sNLS-sNLS protein complexed to HBG1-1 gRNA ("His-AsCpf1-sNLS-sNLS_HBG1-1 RNP") either alone, or in combination with 2.5 µM of one of the ssODN donors (OLI16430, OLI164324, OLI16409, or OLI16410). Co-delivery of the RNP and ssODN donor encoding the 18 nt deletion with positive strand homology arms (OLI16409) enhanced the editing frequency from 39.5% without donor to 73.3%, as determined by sequencing analysis of the *HBG* PCR product from genomic DNA extracted at 72 hours post-electroporation (**Fig. 4A**).

Table 7: Single strand deoxynucleotide donor repair templates "encoding" deletions or mutations at or near the CCAAT box

ssODN ID	OLI-ID	Sequence
Ptx ssODN - Positive Strand -18nt (180)	OLI16409	G*GTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGGCCCC TGGCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTT GTCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGGACC GTTTCAGACAGATATTTGCATTGAGATAGTGTGGGGAAGGGG CCCCCAAGA*G (SEQ ID NO:974)
Ptx ssODN - Negative Strand -18nt (180)	OLI16410	C*TCTTGGGGGCCCCTTCCCCACACTATCTCAATGCAAATATC TGTCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCC AGCCTTGACAAGGCAAACTTGACCAATAGTCTTAGAGTATCC AGTGAGGCCAGGGGCCGGCGGCTGGCTAGGGATGAAGAATA AAAGGAAGCAC*C (SEQ ID NO:975)
Ptx ssODN - Negative Strand -11nt (180)	OLI16411	G*GCCCCTTCCCCACACTATCTCAATGCAAATATCTGTCTGAA ACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGCCTTGC CTTGAGACAAGGCAAACTTGACCAATAGTCTTAGAGTATCCA GTGAGGCCAGGGGCCGGCGGCTGGCTAGGGATGAAGAATAA AAGGAAGCACC*C (SEQ ID NO:976)
Ptx ssODN - Negative Strand -13nt (180)	OLI16412	G*GCCCCTTCCCCACACTATCTCAATGCAAATATCTGTCTGAA ACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGCCTTGC CTTGACAAGGCAAACTTGACCAATAGTCTTAGAGTATCCAGT GAGGCCAGGGGCCGGCGGCTGGCTAGGGATGAAGAATAAAA GGAAGCACCCT*T (SEQ ID NO:977)
Ptx ssODN - Positive Strand -11nt (180)	OLI16413	G*GGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGGCCC CTGGCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCT TGTCTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGC CAGGGACCGTTTCAGACAGATATTTGCATTGAGATAGTGTGG GGAAGGGGC*C (SEQ ID NO:978)
Ptx ssODN - Positive Strand -13nt (180)	OLI16414	A*AGGGTGCTTCCTTTTATTCTTCATCCCTAGCCAGCCGCCGGC CCCTGGCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTGC CTTGTCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGC CAGGGACCGTTTCAGACAGATATTTGCATTGAGATAGTGTGG GGAAGGGGC*C (SEQ ID NO:979)

Ptx ssODN - Negative Strand 117: G>A (180)	OLI16415	G*GCCCCTTCCCCACACTATCTCAATGCAAATATCTGTCTGAA ACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGCCTTGC CTTAACCAATAGCCTTGACAAGGCAAACTTGACCAATAGTCTT AGAGTATCCAGTGAGGCCAGGGGCCGGCGGCTGGCTAGGGAT GAAGAATAA*A (SEQ ID NO:980)
Ptx ssODN - Positive Strand 117: G>A (180)	OLI16416	T*TTATTCTTCATCCCTAGCCAGCCGCCGGCCCCTGGCCTCACT GGATACTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTA TTGGTTAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGC CAGGGACCGTTTCAGACAGATATTTGCATTGAGATAGTGTGG GGAAGGGGC*C (SEQ ID NO:981)
Ptx ssODN - Negative Strand -1nt (180)	OLI16417	G*GCCCCTTCCCCACACTATCTCAATGCAAATATCTGTCTGAA ACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGCCTTGC CTTGCCAATAGCCTTGACAAGGCAAACTTGACCAATAGTCTTA GAGTATCCAGTGAGGCCAGGGGCCGGCGGCTGGCTAGGGATG AAGAATAAA*A (SEQ ID NO:982)
Ptx ssODN - Positive Strand -1nt (180)	OLI16418	T*TTTATTCTTCATCCCTAGCCAGCCGCCGGCCCCTGGCCTCAC TGGATACTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCT ATTGGCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGC CAGGGACCGTTTCAGACAGATATTTGCATTGAGATAGTGTGG GGAAGGGGC*C (SEQ ID NO:983)
Ptx ssODN - Negative Strand -4nt 40/80 (120)	OLI16419	T*GGCTAAACTCCACCCATGGGTTGGCCAGCCTTGCCTTGATA GCCTTGACAAGGCAAACTTGACCAATAGTCTTAGAGTATCCA GTGAGGCCAGGGGCCGGCGGCTGGCTAGGGATGAA*G (SEQ ID NO:984)
Ptx ssODN - Negative Strand -4nt 30/70 (100)	OLI16420	C*CACCCATGGGTTGGCCAGCCTTGCCTTGATAGCCTTGACAA GGCAAACTTGACCAATAGTCTTAGAGTATCCAGTGAGGCCAG GGGCCGGCGGCTGGC*T (SEQ ID NO:985)
Ptx ssODN - Negative Strand -4nt (100)	OLI16421	A*AACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGCCT TGCCTTGATAGCCTTGACAAGGCAAACTTGACCAATAGTCTTA GAGTATCCAGTGAG*G (SEQ ID NO:986)
Ptx ssODN - Negative Strand -4nt (120)	OLI16422	T*ATCTGTCTGAAACGGTCCCTGGCTAAACTCCACCCATGGGT TGGCCAGCCTTGCCTTGATAGCCTTGACAAGGCAAACTTGACC AATAGTCTTAGAGTATCCAGTGAGGCCAGGGGCC*G (SEQ ID NO:987)
Ptx ssODN - Negative Strand -4nt (140)	OLI16423	T*CAATGCAAATATCTGTCTGAAACGGTCCCTGGCTAAACTCC ACCCATGGGTTGGCCAGCCTTGCCTTGATAGCCTTGACAAGGC AAACTTGACCAATAGTCTTAGAGTATCCAGTGAGGCCAGGGG CCGGCGGCTGGC*T (SEQ ID NO:988)
Ptx ssODN - Negative Strand -4nt (180)	OLI16424	G*GCCCCTTCCCCACACTATCTCAATGCAAATATCTGTCTGAA ACGGTCCCTGGCTAAACTCCACCCATGGGTTGGCCAGCCTTGC CTTGATAGCCTTGACAAGGCAAACTTGACCAATAGTCTTAGA GTATCCAGTGAGGCCAGGGGCCGGCGGCTGGCTAGGGATGAA GAATAAAAGG*A (SEQ ID NO:989)
Ptx ssODN - Positive Strand -4nt 40/80 (120)	OLI16425	T*ACTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAGGCTATC AAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGGGA CCGTTTCAGACAGATATTTGCATTGAGATAGTGTG*G (SEQ ID NO:990)
Ptx ssODN - Positive Strand -4nt 30/70 (100)	OLI16426	C*TATTGGTCAAGTTTGCCTTGTCAAGGCTATCAAGGCAAGGC TGGCCAACCCATGGGTGGAGTTTAGCCAGGGACCGTTTCAGA CAGATATTTGCATTG*A (SEQ ID NO:991)

Ptx ssODN - Positive Strand -4nt (100)	OLI16427	C*CTCACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTTGTC AAGGCTATCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTT AGCCAGGGACCGTT*T (SEQ ID NO:992)
Ptx ssODN - Positive Strand -4nt (120)	OLI16428	C*GGCCCCTGGCCTCACTGGATACTCTAAGACTATTGGTCAAG TTTGCCTTGTCAAGGCTATCAAGGCAAGGCTGGCCAACCCATG GGTGGAGTTTAGCCAGGGACCGTTTCAGACAGAT*A (SEQ ID NO:993)
Ptx ssODN - Positive Strand -4nt (140)	OLI16429	A*GCCAGCCGCCGGCCCCTGGCCTCACTGGATACTCTAAGACT ATTGGTCAAGTTTGCCTTGTCAAGGCTATCAAGGCAAGGCTGG CCAACCCATGGGTGGAGTTTAGCCAGGGACCGTTTCAGACAG ATATTTGCATTG*A (SEQ ID NO:994)
Ptx ssODN - Positive Strand -4nt (180)	OLI16430	T*CCTTTTATTCTTCATCCCTAGCCAGCCGCCGGCCCCTGGCCT CACTGGATACTCTAAGACTATTGGTCAAGTTTGCCTTGTCAAG GCTATCAAGGCAAGGCTGGCCAACCCATGGGTGGAGTTTAGC CAGGGACCGTTTCAGACAGATATTTGCATTGAGATAGTGTGG GGAAGGGGC*C (SEQ ID NO:995)
Ptx ssODN - Positive Strand -18nt (100)	OLI16431	C*CCTGGCCTCACTGGATACTCTAAGACTATTGGTCAAGTTTG CCTTGTCAAGGCTGGCCAACCCATGGGTGGAGTTTAGCCAGG GACCGTTTCAGACAG*A (SEQ ID NO:1040)
* : Represents phosphorothioate modification		

[0370] Further analysis of the specific type and size of deletions at the target site revealed that in the presence of the 18 nt positive strand donor (OLI16409), 57.3% of alleles carried the 18 nt deletion compared to 7.5% of alleles when the His-AsCpf1-sNLS-sNLS_HBG1-1 RNP was delivered alone (**Fig. 4B**). The ssODN co-delivery mediated precise repair of the DNA DSB because the 18 nt deletion represented 71.2% of all the indels generated with co-delivery of ssODN OLI16409 and His-AsCpf1-sNLS-sNLS_HBG1-1 RNP, whereas the 18 nt deletion represented only 19.0% of all the indels were generated when delivering His-AsCpf1-sNLS-sNLS_HBG1-1 RNP alone (**Fig. 4C**).

[0371] Although the data in **Figures 4A-4C** originally suggested that codelivery of Cpf1 RNP with ssODN donors increased precise gene editing of genomic DNA resulting in the 18 nt deletion, subsequent testing repeating the experiment in **Example 2** confirmed that these data were an artifact of the experiment. Nonetheless, data acquired from the subsequent repeated experiments testing the AsCpf1 RNP using the guide RNA HBG1-1 and ssODN OLI16409 indicated that codelivery of an ssODN donor supports increased total editing in human mPB-CD34+ cells, although not associated with precise repair of the DNA DSB toward the "-18nt deletion" (see **Example 10**).

Example 3: Cpf1 RNP containing gRNA targeting the distal CCAAT box region of the *HBG* promoter supports gene editing in human hematopoietic stem/progenitor cells which promotes induction of HbF protein expression in the erythroid progeny

[0372] Guide RNA HBG1-1 (**Table 8**) having a targeting domain comprising SEQ ID NO:1002 (**Table 9**), targets a site within the *HBG* promoter (**Fig. 5A**). The HBG1-1 gRNA (SEQ ID NO:1022)

was complexed to wild-type AsCpf1 (AsCpf1-sNLS-sNLS, SEQ ID NO:1001) to form an RNP ("AsCpf1-HBG1-1-RNP", **Table 8**). This complex (5μM or 20μM) was then electroporated into mobilized peripheral blood (mPB) derived CD34+ cells using either the Amaxa nucleofector (Lonza), or the MaxCyte GT (MaxCyte, Inc.) electroporation device. The level of insertions / deletions at the target site was analyzed by Illumina sequencing (NGS) of the PCR amplified target site three days after electroporation. AsCpf1-HBG1-1-RNP supported on-target editing of 43% and 17%, respectively, on the Amaxa and MaxCyte electroporation systems (**Figs. 6A** (Amaxa) **and 6B** (MaxCyte)). Following the editing of mPB CD34+ cells, *ex vivo* differentiation into the erythroid linage was performed for 18 days (Giarratana 2011). Then, relative expression levels of gamma-globin chains (over total beta-like globin chains) was measured by UPLC (**Figs. 6A** (Amaxa) and **6B** (MaxCyte)). AsCpf1-HBG1-1-RNP led to an increase of gamma-globin expression by up to 21% above levels detected in cells derived from mock electroporated mPB CD34+ cells (**Fig. 6A**).

Table 8: RNP components

gRNA ID	Protein variant	RNP components	RNP ID	gRNA modification	gRNA sequence (RNA)
HBG1-1 gRNA	AsCpf1-sNLS-sNLS (SEQ ID NO:1001)	gRNA + AsCpf1 WT	HBG1-1-AsCpf1-RNP	Synthetic unmodified	UAAUUUCUACUC UUGUAGAUCCUU GUCAAGGCUAUU GGUC (SEQ ID NO:1022)
HBG1-1 gRNA	His-AsCpf1-sNLS-sNLS H800A (SEQ ID NO:1032)	gRNA + AsCpf1 H800A	HBG1-1-AsCpf1H800A-RNP	Synthetic unmodified	UAAUUUCUACUC UUGUAGAUCCUU GUCAAGGCUAUU GGUC (SEQ ID NO:1022)
HBG1-1 gRNA	His-AsCpf1-sNLS-sNLS H800A (SEQ ID NO:1032)	gRNA + AsCpf1 H800A	HBG1-1-AsCpf1H800A-RNP	5'OMe-PS2OMe	UAAUUUCUACUC UUGUAGAUCCUU GUCAAGGCUAUU GGmU/PS2/mC (SEQ ID NO:1041)
tSpA dead gRNA	SpCas9WT (SEQ ID NO:1033)	dead gRNA + SpCas9 WT	tSpA-Cas9-RNP	Synthetic 5'-3' 3xPSOMe	mG*mG*mC*UGG CCAACCCAUGU UUUAGAGCUAG AAAUAGCAAGU UAAAAUAAGGC UAGUCCGUUAU CAACUUGAAAA AGUGGCACCGA GUCGGUGCmU* mU*mU*U (SEQ ID NO:1024)
SpA gRNA	His-NLS-SpCas9D10A (SEQ ID NO:1034)	gRNA + SpCas9 nickase	SpA-D10A-RNP	Synthetic 5'-3' 3xPSOMe	mG*mG*mC*AAG GCUGGCCAACC CAUGUUUUAGA GCUAGAAAUAG CAAGUUAAAAU AAGGCUAGUCC GUUAUCAACUU GAAAAAGUGGC ACCGAGUCGGU GCmU*mU*mU*U (SEQ ID NO:1025)
SpG gRNA	His-NLS-SpCas9D10A (SEQ ID NO:1034)	gRNA + SpCas9 nickase	SpG-D10A-RNP	Synthetic 5'-3' 3xPSOMe	mU*mA*mG*UCU UAGAGUAUCCA GUGGUUUUAGA GCUAGAAAUAG CAAGUUAAAAU AAGGCUAGUCC GUUAUCAACUU GAAAAAGUGGC ACCGAGUCGGU GCmU*mU*mU*U (SEQ ID NO:1026)

Sp182 dead gRNA	SpCas9WT (SEQ ID NO:1033)	dead gRNA + SpCas9 WT	Sp182-Cas9-RNP	Synthetic 5'-3' 3xPSOMe	mU*mU*mA*GAG UAUCCAGUGGU UUUAGAGCUAG AAAUAGCAAGU UAAAAUAAGGC UAGUCCGUUAU CAACUUGAAAA AGUGGCACCGA GUCGGUGCmU* mU*mU*U (SEQ ID NO:1027)
"His" refers to a six-histidine purification sequence "AsCpf1" refers to the *Acidaminococcus sp.* Cpf1 protein sequence "nNLS" refers to the nucleoplasmin NLS, and "sNLS" refers to the SV40 NLS "SpCas9" refers to the *S. Pyogenes* Cas9 protein sequence "*" represents phosphorothioate modification "PS" represents phosphorothioate modification "PS2" represents phosphorodithioate modification "OMe" represents a 2-o-methyl modification "m" represents a 2-o-methyl modification					

Table 9: guide RNA targeting domain sequences

gRNA ID	gRNA Targeting domain sequence (RNA)	gRNA Targeting domain sequence (DNA)
HBG1-1 gRNA	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)	CCTTGTCAAGGCTATTGGTC (SEQ ID NO:1003)
Sp182 dgRNA	UUAGAGUAUCCAGUG (SEQ ID NO:1028)	TTAGAGTATCCAGTG (SEQ ID NO:1029)
SpG gRNA	UAGUCUUAGAGUAUCCAGUG (SEQ ID NO:359)	TAGTCTTAGAGTATCCAGTG (SEQ ID NO:1030)
SpA gRNA	GGCAAGGCUGGCCAACCCAU (SEQ ID NO:941)	GGCAAGGCTGGCCAACCCAT (SEQ ID NO:944)
tSpA dgRNA	GGCUGGCCAACCCAU (SEQ ID NO:326)	GGCTGGCCAACCCAT (SEQ ID NO:1031)

Example 4: Cpf1 RNP editing efficiency at the CCAAT box region of the HBG promoter is enhanced when co-delivered with *S. Pyogenes* Cas9 RNP with no impact on viability

[0373] It was hypothesized that the co-delivery of Cpf1-RNP with proximally targeting *S. Pyogenes* Cas9 RNP, either catalytically inactive or introducing single nicks, could enhance editing levels at the target site. In an attempt to enhance Cpf1-RNP mediated editing at the distal CCAAT box region of the HBG promoter, HBG1-1-AsCpf1H800A-RNP (composed of His-AsCpf1-sNLS-sNLS-H800A (SEQ ID NO:1032) complexed to HBG1-1 gRNA with a 3' modification as shown in **Table 8** (SEQ ID NO:1041) was co-delivered into mPB CD34+ cells with: (1) *S. Pyogenes* Cas9 D10A RNP containing a Cas9 D10A protein (His-NLS-SpCas9D10A, SEQ ID NO:1034) complexed to a full length guide RNA (100mer) selected from: (a) SpA gRNA (**Tables 8**, **9**; **Fig. 5B**) ("SpA-D10A-

RNP", **Table 8**) or (b) SpG gRNA (**Tables 8**, **9**; **Fig. 5C**) ("SpG-D10A-RNP", **Table 8**)); (2) *S. Pyogenes* WT Cas9 RNP containing a WT Cas9 protein (SpCas9WT, SEQ ID NO:1033) complexed to a truncated dead guide RNA (95mer - with a shortened protospacer region) selected from (a) tSpA dgRNA (**Tables 8, 9**; **Fig. 5E**) ("tSpA-Cas9-RNP", **Table 8**); (b) Sp182 dgRNA (**Tables 8**, **9**; **Fig. 5F**) ("Sp182-Cas9-RNP", **Table 8**); or (c) tSpA-Cas9-RNP and Sp182-Cas9-RNP (**Tables 8**, **9**, **Fig. 5D**). The RNP complexes were electroporated using the MaxCyte GT device (MaxCyte, Inc). The level of insertions / deletions at the target site was then analyzed by Illumina sequencing (NGS) of the PCR amplified target site three days after electroporation. In all combinations tested, irrespective of the *S. Pyogenes* Cas9 enzyme used (D10A or WT) and of PAM orientation, total editing was increased above levels observed following Maxcyte delivery of HBG1-1-AsCpf1H800A-RNP alone (**Fig. 7** and **Table 10**). In addition there was no detrimental effect on viability of mPB CD34+ cells when *S. Pyogenes* Cas9 RNP were co-delivered (**Fig. 8**).

Table 10: Summary of indels and gamma globin expression with each RNP combination at the maximum RNP dose tested

Cpf1 RNP	Cas9 RNP1	Cas9 RNP2	Indels (NGS)	Gamma Globin (over total beta-like chains)
-	-	-	0%	13.5%
HBG1-1-AsCpf1H800A-RNP	-	-	8.0%	20.0%
HBG1-1-AsCpf1H800A-RNP	SpG-D10A-RNP	-	88.6%	30.2%
HBG1-1-AsCpf1H800A-RNP	SpA-D10A-RNP	-	42.6%	32.5%
HBG1-1-AsCpf1H800A-RNP	tSpA-Cas9-RNP	-	32.2%	19.8%
HBG1-1-AsCpf1H800A-RNP	Sp182-Cas9-RNP	-	85.4%	42.8%
HBG1-1-AsCpf1H800A-RNP	tSpA-Cas9-RNP	Sp182-Cas9-RNP	80.1%	33.7%

Example 5: Cpf1 editing profile can be manipulated with co-delivery of *S. Pyogenes* Cas9-D10A-RNP

[0374] In addition to the increase in total editing observed when HBG1-1-AsCpf1H800A-RNP was co-delivered to mPB CD34+ cells with *S. Pyogenes* Cas9-D10A-RNP (**Fig. 7**), changes to the indel profile were also observed. The introduction of a single strand break proximal to the Cpf1-RNP target site by the D10A enzyme (**Fig. 9A**) altered the directionality, length and/or position of the indels (**Fig. 9B**). For example, Sp182-D10A-RNP strongly shifted the profile toward the nicking site introduced by the D10A RNP, with deletions of various length extending from the Cpf1 cut site toward the upstream nicking site (Sp182-D10A-RNP, **Fig. 9B**). While the SpA-D10A-RNP also promoted the generation of deletions extending from the HBG1-1-AsCpf1H800A-RNP cut site to the nicking site introduced by the D10A enzyme (here located downstream), in this case, additional deletions apparently originating from the nicking site, but not extending fully to the HBG1-1-AsCpf1H800A-RNP cut site, were also produced (SpA-D10A-RNP, **Fig. 9B**).

[0375] It is probable that the orientation, target strand, and distance of the *S. Pyogenes* Cas9-D10A-RNP target site to the Cpf1-RNP target site leads to differences in the position and length of the mutations promoted by the additional DNA nick (**Fig. 9A**). It should be noted that in certain applications this directional manipulation of the indel profile, *i.e.*, an increase in the frequency of indels occurring between the Cpf1-RNP and *S. Pyogenes* Cas9 D10A-RNP binding sites, could be used to favor a desired editing outcome, to increase the rate of productive indels (*e.g.*, indels disrupting a targeted site). In the case set-out within, to disrupt the dCCAAT target region and induce HbF protein expression, the co-delivery of HBG1-1-AsCpf1H800A-RNP with *S. Pyogenes* D10A-RNP led to an increase of gamma globin levels by 16.7% (SpG-D10A-RNP) or 19.0% (SpA-D10A-RNP) above levels detected in mock electroporated cells, as detected by UPLC analysis after 18 days of erythroid differentiation post electroporation (**Fig. 7** and **Table 10**). The frequency of productive indels was higher when pairing HBG1-1-AsCpf1H800A-RNP with SpA-D10A-RNP (versus SpG-D10A-RNP) as a higher increase in HbF levels (19% with Cas9D10A-SpA-RNP vs 16.7% with SpG-D10A-RNP) was achieved from a lower editing level (42.6% with SpA-D10A-RNP vs 88.6% with SpG-D10A-RNP) (**Fig. 7** and **Table 10**).

Example 6: Cpf1 editing is increased without altering the editing profile in the presence of *S. Pyogenes* Cas9 WT RNP containing a dead guide RNA

[0376] When co-delivering HBG1-1-AsCpf1H800A-RNP (at a fixed dose of 6 μM) to mPB CD34+ cells with increasing doses of Sp182-Cas9-RNP (complexed with a truncated, 95mer dead gRNA, Sp182), using the MaxCyte electroporator device, an editing boost has been observed up to greater than 10-fold (85.4% editing with the highest dose of Sp182-Cas9-RNP versus 8.0% when HBG1-1-AsCpf1H800A-RNP was delivered alone) (**Fig. 7**). The levels of editing achieved here by co-delivering Sp182-Cas9-RNP also greatly surpassed the levels of editing obtained using the MaxCyte electroporator device when HBG1-1-AsCpf1H800A-RNP was delivered alone at doses as high as

20uM (**Fig. 6B**). In contrast to the results observed by co-delivering a D10A-RNP (in **Example 5**), the increase in editing achieved by co-delivering HBG1-1-AsCpf1H800A-RNP with Sp182-Cas9-RNP (Cas9-WT complexed with a truncated gRNA) was not associated with an apparent impact on the indel profile. The indels detected in electroporated cells were centered around the Cpf1 cut site in an approximal symmetrical fashion and no indels were detected at the Sp182-Cas9-RNP target site (**Fig. 10A**). Of note, 96% of the total indels were disrupting the distal-CCAAT box, and 79.5% of total indels disrupted 3 or more nucleotides of the CCAAT box motif (**Fig. 10B**). Further dosing optimization using HBG1-1-AsCpf1H800A-RNP (composed of His-AsCpf1-sNLS-sNLS-H800A (SEQ ID NO:1032) complexed to HBG1-1 gRNA (SEQ ID NO:1022)) and Sp182-Cas9-RNP enabled editing levels of up to 92% in mPB-CD34+ cells at 120 hours post-electroporation using the MaxCyte device, and lead to gamma chain expression levels 32% above background in erythroid derived cells (41% of gamma chains in treated cells, 8% in mock treated cells) (**Fig. 11**). Those results demonstrate that an RNP composed of WT *S. Pyogenes* Cas9 protein complexed with a truncated gRNA can increase the editing from a proximally binding Cpf1-RNP without introducing detectable levels of editing at its own binding site nor noticeably affecting the length nor directionality of indels generated by the Cpf1-RNP. Here, the editing enhancement provided by the Sp182-Cas9-RNP enables high editing of the HBG1-1-AsCpf1H800A-RNP at the HBG promoter, with a high frequency of indels disrupting the distal CCAAT box target motif and leading to therapeutically relevant levels of gamma chain expression in the bulk erythroid progeny of the electroporated cells.

Example 7: Clonal HbF distribution within cell population edited with Cpf1 gRNA targeting the distal CCAAT box

[0377] A single cell experiment was next performed to evaluate the distribution of gamma chain expression levels in erythroid cells derived from mPB CD34+ cells electroporated with the HBG1-1-AsCpf1H800A-RNP (composed of His-AsCpf1-sNLS-sNLS-H800A (SEQ ID NO:1032) complexed to HBG1-1 gRNA with a 3' modification as shown in **Table 8** (SEQ ID NO:1041)) + Sp182-Cas9-RNP combination (**Fig. 5F**, **Table 8**). After being left to recover during 48 hours post-electroporation, the cells were sorted by Fluorescence Activated Cell Sorting (FACS) at 1-cell / well in non-tissue culture treated 384-well plates. The cells were then differentiated and expanded clonally for 18 days into the erythroid lineage (adapted from Giarratana 2011). UPLC analysis was performed to determine the distribution of gamma chain expression levels (percentage of gamma chains/[total beta-like chains]) in the clonal erythroid progeny of cells derived from the total population of mPB-CD34+ cells initially edited. It is considered that in order to achieve a functional benefit and alleviate symptoms associated with sickle cell disease, ~30% of the erythroid cells should have fetal hemoglobin levels greater than 30%. Of the clones analyzed, 83.1% had gamma chains levels above the median level detected in erythroid clones derived from mock electroporated mPB CD34+ cells

(median value = 18.8%), 64.2% had gamma chains levels exceeding 30% of total beta-like and 35.8% had gamma chains levels exceeding 48.8% of total beta-like (30% +18.8% median level in control cells) (**Fig. 12**).

Example 8: Screen of Cpf1 gRNAs targeting the HBG promoter region

[0378] To identify other AsCpf1 gRNA that could be used as a component of a single RNP or in combination with a "booster element" to increase editing of the HBG promoter region in CD34+ cells and induce fetal globin expression in the erythroid progeny of modified cells, His-AsCpf1-NLS-NLS ("AsCpf1," SEQ ID NO:1000); AsCpf1 S542R/K607R ("AsCpf1 RR," SEQ ID NO:1036); or AsCpf1 S542R/K548V/N552R ("AsCpf1 RVR," SEQ ID NO:1037) gRNA sequences targeting several domains of the HBG promoter (**Table 11**) were designed (listed in **Table 12**). AsCpf1 RR and AsCpf1 RVR are engineered AsCpf1 variants which recognize TYCV/ACCC/CCCC and TATV/RATR PAMs, respectively (Gao 2017).

Table 11: Subdomains of the HBG genomic region

Genomic Coordinate of HbG*	Nucleotides	Name of Region
Chr 11 (NC_000011.10): 5,247,883-5,248,186	TCCTAAAGCT TGGAACACTT TCCCTTCCTT AAGAACCATC CTTGCTACTC AGCTGCAATC AATCCAGCCC CCAGGTCTTC ACTGAACCTT TTCCCATCTC TTCCAAAACA TCTGTTTCTG AGAAGTCCTG TCCTATAGAG GTCTTTCTTC CCACCGGATT TCTCCTACAC CATTTACTCC CACTTGCAGA ACTCCCGTGT ACAAGTGTCT TTACTGCTTT TATTTGCTCA TCAAAATGCA CATCTCATAT AAAAATAAAT GAGGAGCATG CACACACCAC AAACACAAAC AGGCATGCAG AAAT (SEQ ID NO:1118)	Region 1: Downstream of *HBG1*
Chr 11 (NC_000011.10): 5,248,509 – 5,249,173	ATAAAGATGA ACCCATAGTG AGCTGAGAGC TCCAGCCTGG CCTCCAGATA ACTACACACC AAGCTTCCAC CCAGAATCAA GCCTATGTTA ACTTCCCTCA AAGCCTGAGA TTTTGCCTTC CCATTAAATG CAGGTAGTTG TTCCCCTTCA AGCACTAGTC ACTGGCCATA ATTTAAATCT TGCTATCTTC TTGCCACCAT GAACCCTGTA TGTTGTAGGC TGAAGACGTT AAAAGAAACA	Region 2: *HBG1* Intron 2 - A

	CACGCTGACA CACACACACA CACGCGCGCG CGCACACACA CACACACACA CAGAGCTGAC TTTCAAAATC TACTCCAGCC CAAATGTTTC AATTGTTCCT CACCCCTGGA CATACTTTGC CCCCATCTGG AATTAAAGGA TATAAGTTTG TAATGAAGCA TTAGCAGCAT TTTATATGTG TCCAGCTGAT ATAGGAATAG CCTTAGCAAT GTATGTTTGG CCACCAAAGT TCCCCACTTT GACTGAGCCA ATATATGCCT TCTGCCTGCA TCTTTTTAAC GACCATACTT GTCCTGCCTC CAGATAGATG TTTTAAAACA ACAAAAATGA GGGAAAGATG AAAGTTCTTT CTACTGGAAT CTAATAAAGA AAAGTCATTT TCCTCATTTC CACCTCTCTT TTCTCAAAGT CAAAATTGTC CATCT (SEQ ID NO:1119)	
Chr 11 (NC_000011.10): 5,249,198 – 5,249,362	CCCTAAAACA TTACCACTGG GTCTCAGCCC AGTTAGTCCT CTGCAGTTTC TTCACCCCCA ACCCCAGTAT CTTCAAACAG CTCACACCCT GCTGTGCTCA GATCAATACT CCGTTGTCTA AGTTGCCTCG AGACTAAAGG CAACAGGGCT GAAACATCTC CTGGA (SEQ ID NO:1120)	Region 3: *HBG1* Intron 2 - B
Chr 11 (NC_000011.10): 5,249,591– 5,249,712	CTGTGAGATT GACAAGAACA GTTTGACAGT CAGAAGGTGC CACAAATCCT GAGAAGCGAC CTGGACTTTT GCCAGGCACA GGGTCCTTCC TTCCCTCCCT TGTCCTGGTC ACCAGAGCCT AC (SEQ ID NO:1121)	Region 4: *HBG1* Intron 1
Chr 11 (NC_000011.10): 5,249,904 – 5,249,927	GCCGCCGGCC CCTGGCCTCA CTGG (SEQ ID NO:1122)	Region 5: *HBG1* -60 nt region from Transcription Start Site (TSS)
Chr 11 (NC_000011.10): 5,249,955 – 5,249,987	CCTTGTCAAG GCTATTGGTC AAGGCAAGGC TGG (SEQ ID NO:1123)	Region 6: *HBG1* -110 nt region from TSS
Chr 11 (NC_000011.10): 5,250,040 – 5,250,075	TGAGATAGTG TGGGGAAGGG GCCCCC AAGAGGATAC (SEQ ID NO:1124)	Region 7: *HBG1* -200 nt region from TSS
Chr 11 (NC_000011.10): 5,250,089 – 5,250,129	TATAGCCTTT GCCTTGTTCC GATTCAGTCA TTCCAGTTTT T (SEQ ID NO:1125)	Region 8: *HBG1* -250 nt region from TSS

Chr 11 (NC_000011.10): 5,250,141 – 5,250,254	TCTTCCCTTT AGCTAGTTTC CTTCTCCCAT CATAGAGGAT ACCAGGACTT CTTTTGTCAG CCGTTTTTTA CCTTCTTGTC TCTAGCTCCA GTGAGGCCTG TAGTTTAAAG CTAA (SEQ ID NO:1126)	Region 9: *HBG1* -333 nt region from TSS
Chr 11 (NC_000011.10): 5,250,464 – 5,250,549	CCACAGTTTC AGCGCAGTAA TAGATTAGTG TTACATAATA TAAGACCTAA TGCTTACCTC AATATCTACT TATCCGTACC TATTTG (SEQ ID NO:1127)	Region 10: *HBG1* -650 nt region from TSS
Chr 11 (NC_000011.10): 5,250,594 – 5,250,735	TATTCAGGTA TGTATGTATA CACCAGATGA TGTGTATTTA CCACTGGATA AGTGTGTGTG CTGGCTGATG ACCCAGGGTT TTGGCGTAGC TCTTCTATGC TCAGTAAAGA TGATGGTAGA ATGTTCTTTG GCAGGTACTG TG (SEQ ID NO:1128)	Region 11: *HBG1* -800 nt region from TSS
Chr 11 (NC_000011.10): 5,253,425 – 5,254,121	CAATAAAGAT GAACCCATAG TGAGCTGAGA GCTCCAGCCT GGCCTCCAGA TAACTACACA CCAAGCTTCC ACCCAGAATC AAGCCTATGT TAACTTCCCT CAAAGCCTGA GATTTTGCTT TCCCATTAAA TGCAGGTAGT TGTTCTTCTT GCAGCACTAG TCACTGGCCA TAATTTAAAT CTTGTTATCT TCTTGCCACC ATGAACCCTG TATGCTGTAG GCTGAAAACG TTAAAAGAAA CACACGCTCT CACACACACA CAAACACACG CGCGCACACA CACACACACA CACACAGAGC TGACTTTCAA AATCTACTCC AGCCCAAATG TTTCAATTGT TCCTCACCCC TGGACATACT TTGCCCCCAT CTGGAATTAA AGGATATAAG TTTGTAATGA AGCATTAGCA GCATTTTATA TGTGTCCAGC TGATATAGGA ATAGCCTTAG CAATGTATGT TTGGCCACCA AAGTTCCCCA CTTTGACTGA GCCAATATAT GCCTTCTGCC TGCATCTTTT TAATGACCAT ACTTGTCCTG CCTCCAGATA GATGTTTTAA AACGAATAAC AAAAATAGGG GAAAGGTGAA AGTTCTTTCT ACCGAAATCT AATAAAGAAA AGTCATTTTC CTCATTTCCA CCTCTCTTTT CTCAAAGTCA AAGTTGTCCA TCTAGATTTT	Region 12: *HBG2* Intron 2 - A

	CAGAGGCACT CCTTAGG (SEQ ID NO:1129)	
Chr 11 (NC_000011.10): 5,254,122 – 5,254,306	CCCTAAAACA TTGCCACTGG GTCTCAGCCC AGTTAGTCCT CTGCAGTTTC TTCACTCCCA ACCCCAGTAT CTTCAAACAG CTCACACCCT GCTGTGCTCA GATCAATACT CAGTTGTCTA AGTTGCCTCG AGACTAAAGG CAACAGTGCT GAAACATCTC CTGGACTCAC CTTGAAGTTC TCAGG (SEQ ID NO:1130)	Region 13: *HBG2* Intron 2 - B
Chr 11 (NC_000011.10): 5,254,511 – 5,254,648	AGCCTGTGAG ATTGACAAGA ACAGTTTGAC AGTCAGAAGG TGCCACAAAT CCTGAGAAGC GACCTGGACT TTTGCCAGGC ACAGGGTCCT TCCTTCCCTC CCTTGTCCTG GTCACCAGAG CCTACCTTCC CAGGGTT (SEQ ID NO:1131)	Region 14: *HBG2* Intron 1
Chr 11 (NC_000011.10): 5,254,829 – 5,254,866	CCGCCGGCCC CTGGCCTCAC TGGATACTCT AAGACTAT (SEQ ID NO:1132)	Region 15: *HBG2* -60 nt region from TSS
Chr 11 (NC_000011.10): 5,254,879 – 5,254,909	CCTTGTCAAG GCTATTGGTC AAGGCAAGGC T (SEQ ID NO:1133)	Region 16: *HBG2* -110 nt region from TSS
Chr 11 (NC_000011.10): 5,254,935 5,255,009	CAGGGACCGT TTCAGACAGA TATTTGCATT GAGATAGTGT GGGGAAGGGG CCCCCAAGAG GATACTGCTG CTTAA (SEQ ID NO:1134)	Region 17: *HBG2* -200 nt region from TSS
Chr 11 (NC_000011.10): 5,255,025 – 5,255,053	TTGCCTTGTT CCGATTCAGT CATTCCAAT (SEQ ID NO:1135)	Region 18: *HBG2* -250 nt region from TSS
Chr 11 (NC_000011.10): 5,255,076 – 5,255,179	TTTAGCTAGT TTTCTTCTCC CACCATAGAA GATACCAGGA CTTCTTTTGT CAGCCGTTTT TCACCTTCTT GTCTGTAGCT CCAGTGAGGC CTGTAGTTTA AAGT (SEQ ID NO:1136)	Region 19: *HBG2* -330 nt region from TSS
Chr 11 (NC_000011.10): 5,255,255 – 5,255,292	GGACACGTCT TAGTCTCATT TAGTAAGCAT TGGTTTCC (SEQ ID NO:1137)	Region 20: *HBG2* -500 nt region from TSS
Chr 11 (NC_000011.10): 5,255,518 – 5,255,641	TTTTTTATAT TCAGGTATGT ATGTAGGCAC CCGATGATGT GTATTTATCA CTGGATAAGT GTATGTGCTG GCTGATGACC CAGGGTTTTG GTGTAGCTCT TCTATGCTCG GTAAAGATGA TGGT (SEQ ID NO:1138)	Region 21: *HBG2* -800 nt region from TSS

*NCBI Reference Sequence NC_000011, the coordinates are reported using the One-based coordinate system, "Homo sapiens chromosome 11, GRCh38.p12 Primary Assembly," (Version NC_000011.10).

Table 12: Cpf1 guide RNAs

gRNA targeting domain sequence ID*	gRNA Targeting domain sequence (RNA)	gRNA Targeting domain sequence (DNA)	Genomic coordinates at HbG1**	Genomic coordinates at HbG2**	% Editing	Strand
AsCpf1 HBG1 Promoter-1	AGACAGAUAU UUGCAUUGAG (SEQ ID NO:1139)	AGACAGATAT TTGCATTGAG (SEQ ID NO:1140)	Chr11:525002 4:5250043	Chr11:52549 48:5254967	5.43	+
AsCpf1 HBG1 Promoter-2	CAUUGAGAUA GUGUGGGGA A (SEQ ID NO:1141)	CATTGAGATA GTGTGGGGAA (SEQ ID NO:1142)	Chr11:525003 7:5250056	Chr11:52549 61:5254980	8.30	+
AsCpf1 HBG1 Promoter-3	UAGCCUUUGC CUUGUUCCGA (SEQ ID NO:1143)	TAGCCTTTGC CTTGTTCCGA (SEQ ID NO:1144)	Chr11:525009 1:5250110	Chr11:52550 19:5255038	0.23	+
AsCpf1 HBG1 Promoter-4	CCUUGUUCCG AUUCAGUCAU (SEQ ID NO:1145)	CCTTGTTCCG ATTCAGTCAT (SEQ ID NO:1146)	Chr11:525010 0:5250119	Chr11:52550 28:5255047	1.15	+
AsCpf1 HBG1 Promoter-5	UCUAAUUUAU UCUUCCCUUU (SEQ ID NO:1147)	TCTAATTTATT CTTCCCTTT (SEQ ID NO:1148)	Chr11:525013 1:5250150	Chr11:52550 59:5255078	0.16	+
AsCpf1 HBG1 Promoter-6	CUUCUCCCAU CAUAGAGGAU (SEQ ID NO:1149)	CTTCTCCCATC ATAGAGGAT (SEQ ID NO:1150)	Chr11:525016 1:5250180		12.73	+
AsCpf1 HBG2 Promoter-7	UUCUCCCACC AUAGAAGAU A (SEQ ID NO:1151)	TTCTCCCACC ATAGAAGATA (SEQ ID NO:1152)		Chr11:52550 90:5255109	8.11	+
AsCpf1 HBG1 Promoter-8	CCACUGGAUA AGUGUGUGU G (SEQ ID NO:1153)	CCACTGGATA AGTGTGTGTG (SEQ ID NO:1154)	Chr11:525063 4:5250653		13.33	+
AsCpf1 HBG1 Promoter-9	GCGUAGCUCU UCUAUGCUCA (SEQ ID NO:1155)	GCGTAGCTCT TCTATGCTCA (SEQ ID NO:1156)	Chr11:525067 7:5250696		13.48	+
AsCpf1 HBG1 Promoter-10	CUGAGCAUAG AAGAGCUACG (SEQ ID NO:1157)	CTGAGCATAG AAGAGCTACG (SEQ ID NO:1158)	Chr11:525067 8:5250697		10.73	-
AsCpf1 HBG2 Promoter-11	UCACUGGAUA AGUGUAUGU	TCACTGGATA AGTGTATGTG		Chr11:52555 65:5255584	0.43	+

	G (SEQ ID NO:1159)	(SEQ ID NO:1160)				
AsCpf1 HBG2 Promoter-12	GUGUAGCUCU UCUAUGCUCG (SEQ ID NO:1161)	GTGTAGCTCT TCTATGCTCG (SEQ ID NO:1162)		Chr11:52556 08:5255627	5.78	+
AsCpf1 HBG2 Promoter-13	CCGAGCAUAG AAGAGCUACA (SEQ ID NO:1163)	CCGAGCATAG AAGAGCTACA (SEQ ID NO:1164)		Chr11:52556 09:5255628	3.24	-
HBG1-1 AsCpf1	CCUUGUCAAG GCUAUUGGUC (SEQ ID NO:1002)	CCTTGTCAAG GCTATTGGTC (SEQ ID NO:1003)	Chr11:524995 5:5249974	Chr11:52548 79:5254898	17.96	+
AsCpf1 RR HBG1 Promoter-1	GACAGAUAUU UGCAUUGAGA (SEQ ID NO:1167)	GACAGATATT TGCATTGAGA (SEQ ID NO:1168)	Chr11:525002 5:5250044	Chr11:52549 49:5254968	8.48	+
AsCpf1 RR HBG1 Promoter-2	ACACUAUCUC AAUGCAAAUA (SEQ ID NO:1169)	ACACTATCTC AATGCAAATA (SEQ ID NO:1170)	Chr11:525003 1:5250050	Chr11:52549 55:5254974	0.09	-
AsCpf1 RR HBG1 Promoter-3	CACACUAUCU CAAUGCAAAU (SEQ ID NO:1171)	CACACTATCT CAATGCAAAT (SEQ ID NO:1172)	Chr11:525003 2:5250051	Chr11:52549 56:5254975	2.10	-
AsCpf1 RR HBG1 Promoter-4	CCACACUAUC UCAAUGCAAA (SEQ ID NO:1173)	CCACACTATC TCAATGCAAA (SEQ ID NO:1174)	Chr11:525003 3:5250052	Chr11:52549 57:5254976	2.52	-
AsCpf1 RR HBG1 Promoter-5	UUCCCCACAC UAUCUCAAUG (SEQ ID NO:1175)	TTCCCCACAC TATCTCAATG (SEQ ID NO:1176)	Chr11:525003 7:5250056	Chr11:52549 61:5254980	0.05	-
AsCpf1 RR HBG1 Promoter-6	GAUUCAGUCA UUCCAGUUUU (SEQ ID NO:1177)	GATTCAGTCA TTCCAGTTTT (SEQ ID NO:1178)	Chr11:525010 9:5250128		0.77	+
AsCpf1 RR HBG1 Promoter-7	AUUCAGUCAU UCCAGUUUUU (SEQ ID NO:1179)	ATTCAGTCAT TCCAGTTTTT (SEQ ID NO:1180)	Chr11:525011 0:5250129		0.24	+
AsCpf1 RR HBG1 Promoter-8	GUCAUUCCAG UUUUUCUCUA (SEQ ID NO:1181)	GTCATTCCAG TTTTTCTCTA (SEQ ID NO:1182)	Chr11:525011 5:5250134		1.00	+
AsCpf1 RR HBG1 Promoter-9	AGUUUUUCUC UAAUUUAUUC (SEQ ID NO:1183)	AGTTTTTCTCT AATTTATTC (SEQ ID NO:1184)	Chr11:525012 3:5250142		0.15	+
AsCpf1 RR HBG1 Promoter-10	GUUUUUCUCU AAUUUAUUCU (SEQ ID NO:1185)	GTTTTTCTCTA ATTTATTCT (SEQ ID NO:1186)	Chr11:525012 4:5250143		0.15	+

AsCpf1 RR HBG2 Promoter-11	GAUUCAGUCA UUCCAAUUUU (SEQ ID NO:1187)	CAAGAGGATA CTGCTGCTTA (SEQ ID NO:1188)		Chr11:52549 89:5255008	***	+
AsCpf1 RR HBG2 Promoter-12	AUUCAGUCAU UCCAAUUUUU (SEQ ID NO:1189)	AAGAGGATAC TGCTGCTTAA (SEQ ID NO:1190)		Chr11:52549 90:5255009	***	+
AsCpf1 RR HBG2 Promoter-13	GUCAUUCCAA UUUUUCUCUA (SEQ ID NO:1191)	GATTCAGTCA TTCCAATTTT (SEQ ID NO:1192)		Chr11:52550 37:5255056	0.25	+
AsCpf1 RR HBG2 Promoter-14	AAUUUUUCUC UAAUUUAUUC (SEQ ID NO:1193)	ATTCAGTCAT TCCAATTTTT (SEQ ID NO:1194)		Chr11:52550 38:5255057	0.13	+
AsCpf1 RR HBG2 Promoter-15	AUUUUUCUCU AAUUUAUUCU (SEQ ID NO:1195)	GTCATTCCAA TTTTTCTCTA (SEQ ID NO:1196)		Chr11:52550 43:5255062	0.32	+
AsCpf1 RR HBG2 Promoter-16	UUCUCCCAUC AUAGAGGAU A (SEQ ID NO:1197)	AATTTTTCTCT AATTTATTC (SEQ ID NO:1198)		Chr11:52550 51:5255070	0.14	+
AsCpf1 RR HBG2 Promoter-17	AUCAUAGAGG AUACCAGGAC (SEQ ID NO:1199)	ATTTTTCTCTA ATTTATTCT (SEQ ID NO:1200)		Chr11:52550 52:5255071	0.10	+
AsCpf1 RR HBG1 Promoter-18	ACCAUAGAAG AUACCAGGAC (SEQ ID NO:1201)	TTCTCCCATC ATAGAGGATA (SEQ ID NO:1202)	Chr11:525016 2:5250181		1.40	+
AsCpf1 RR HBG1 Promoter-19	CAGUACCUGC CAAAGAACAU (SEQ ID NO:1203)	ATCATAGAGG ATACCAGGAC (SEQ ID NO:1204)	Chr11:525016 9:5250188		7.88	+
AsCpf1 RR HBG2 Promoter-20	UAGUAUCUGG UAAAGAGCAU (SEQ ID NO:1205)	ACCATAGAAG ATACCAGGAC (SEQ ID NO:1206)		Chr11:52550 97:5255116	13.03	+
AsCpf1 RR HBG1 Promoter-21	UCAAUGCAAA UAUCUGUCUG (SEQ ID NO:1207)	CAGTACCTGC CAAAGAACAT (SEQ ID NO:1208)	Chr11:525071 4:5250733		13.31	-
AsCpf1 RR HBG2 Promoter-22	CUCUUGGGGG CCCCUUCCCC (SEQ ID NO:1209)	TAGTATCTGG TAAAGAGCAT (SEQ ID NO:1210)		Chr11:52556 45:5255664	4.07	-
AsCpf1 RVR HBG1 Promoter-1	GAUUCAGUCA UUCCAAUUUU (SEQ ID NO:1211)	TCAATGCAAA TATCTGTCTG (SEQ ID NO:1212)	Chr11:525002 3:5250042	Chr11:52549 47:5254966	0.15	-

AsCpf1 RVR HBG1 Promoter-2	AUUCAGUCAU UCCAAUUUUU (SEQ ID NO:1213)	CTCTTGGGGG CCCCTTCCCC (SEQ ID NO:1214)	Chr11:5250051:5250070	Chr11:5254975:5254994	1.09	-
AsCpf1 RVR HBG1 Promoter-3	AAAAAAAUU AGCAGUAUCC U (SEQ ID NO:1215)	AAAAAAATTA GCAGTATCCT (SEQ ID NO:1216)	Chr11:52500669:5250088		***	-
AsCpf1 RVR HBG1 Promoter-4	GCCUUUGCCU UGUUCCGAUU (SEQ ID NO:1217)	GCCTTTGCCTT GTTCCGATT (SEQ ID NO:1218)	Chr11:52500933:5250112	Chr11:5255021:5255040	3.96	+
AsCpf1 RVR HBG2 Promoter-5	AAAAAAAUU AAGCAGCAGU A (SEQ ID NO:1219)	AAAAAAATTA AGCAGCAGTA (SEQ ID NO:1220)		Chr11:5254997:5255016	***	-
AsCpf1 RVR HBG1 Promoter-6	CUCAGUAAAG AUGAUGGUA G (SEQ ID NO:1221)	CTCAGTAAAG ATGATGGTAG (SEQ ID NO:1222)	Chr11:52506933:5250712		5.32	+
AsCpf1 RVR HBG2 Promoter-7	ACUGGAUAAG UGUAUGUGCU (SEQ ID NO:1223)	ACTGGATAAG TGTATGTGCT (SEQ ID NO:1224)		Chr11:5255567:5255586	9.78	+
AsCpf1 RVR HBG2 Promoter-8	UGCUGGCUGA UGACCCAGGG (SEQ ID NO:1225)	TGCTGGCTGA TGACCCAGGG (SEQ ID NO:1226)	Chr11:52506652:5250671	Chr11:5255583:5255602	0.24	+
AsCpf1 RVR HBG2 Promoter-9	CUCGGUAAAG AUGAUGGUA G (SEQ ID NO:1227)	CTCGGTAAAG ATGATGGTAG (SEQ ID NO:1228)		Chr11:5255624:5255643	5.75	+
AsCpf1 RVR HBG2 Promoter-10	UGGUAAAGA GCAUUCUACC A (SEQ ID NO:1229)	TGGTAAAGAG CATTCTACCA (SEQ ID NO:1230)		Chr11:5255638:5255657	8.55	-

* the gRNA ID name provides the particular Cpf1 molecule used in the RNP complex
**NCBI Reference Sequence NC_000011, the coordinates are reported using the One-based coordinate system, "Homo sapiens chromosome 11, GRCh38.p12 Primary Assembly," (Version NC_000011.10).
*** represents gRNAs that were not tested.

[0379] RNPs (5 μM) containing AsCpf1 protein (SEQ ID NO:1000), AsCpf1 RR protein (SEQ ID NO:1036), or AsCpf1 RVR (SEQ ID NO:1037) complexed with single gRNAs comprising gRNA targeting domains from **Table 12** (see gRNA ID name for the particular Cpf1 molecule used) were delivered to mobilized peripheral blood (mPB) CD34+ cells using the Amaxa electroporator device (Lonza). After 72 hours, genomic DNA was extracted from cells and the level of insertions / deletions at the target site was then analyzed by Illumina sequencing (NGS) of the PCR amplified target site. The percentage of editing (indels= deletions and insertions) for each gRNA is shown in **Table 12** above. In certain embodiments, Cpf1 RNPs comprising one or more of the gRNAs set forth

in **Table 12** may be used to target the regions listed in **Table 11** to induce HbF expression and may be co-delivered with a "booster element" to achieve higher editing levels compared to the editing level of the Cpf1 RNP alone.

Example 9: Co-delivery of HBG1-1-Cpf1 RNP targeting the CCAAT box with ssODN supports an increase in gene editing in human hematopoietic stem/progenitor cells

[0380] A 100 nt ssODN generating the "18 nt deletion" (HBG Δ-104:-121) (*i.e.,* ssODN OLI16431 (SEQ ID NO:1040), **Table 7**) was co-delivered with Cpf1 RNP to further investigate the editing outcome.

[0381] Briefly, human adult mPB CD34+ cells pre-stimulated for two days in medium supplemented with human cytokines were electroporated with RNP comprising the His-AsCpf1-sNLS-sNLS H800A protein (SEQ ID NO:1032, **Table 8**) complexed to modified HBG1-1 gRNA (SEQ ID NO:1041, **Table 8**) ("His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP") either alone at 6 µM or in combination with OLI16431 (0.5 – 6 µM). Co-delivery of the RNP and ssODN donor encoding the 18 nt deletion with positive strand homology arms (OLI16431) enhanced the editing frequency from 8.0% without donor to 75.6%, as determined by sequencing analysis of the *HBG* PCR product from genomic DNA extracted at 72 hours post-electroporation (**Fig. 13A**). This level of editing enabled an increase in HbF levels of ~24% above background (**Fig. 13A**). Additionally, it should be noted that there was no decrease observed in cell viability (**Fig. 13B**) as measured by DAPI exclusion or of colony forming potential (**Fig. 13C**) at any of the doses tested.

Example 10: Optimization of HBG1-1-Cpf1 RNP and OLI16431 dosing to maximize editing at the RNP targeting distal CCAAT box site

[0382] Having demonstrated increased editing when co-delivering ssODN with RNP (*see, e.g.,* **Example 9**), the same methodology was used to optimize the dosing of each component in order to maximize total editing. Briefly, a dosing matrix was set up with RNP comprising the His-AsCpf1-sNLS-sNLS H800A protein (SEQ ID NO:1032, **Table 8**) complexed to unmodified HBG1-1 gRNA (SEQ ID NO:1022, **Table 8**) ("His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP") being co-delivered at 0 – 12 µM with OLI16431 (SEQ ID NO:1040, **Table 7**) at 0 – 12 µM. Using a dose of 8 µM His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP co-delivered with 8 µM OLI16431, a maximum of 89.4% indels was achieved when measured by sequencing analysis of the *HBG* PCR product from genomic DNA extracted from cells following 14 days erythroid culture (**Fig. 14A**). This coincided with an increase in HbF levels to >32% above background under these dosing conditions (**Fig. 14A**). The viability of this sample (His-AsCpf1-sNLS-sNLS H800A_HBG1-1 RNP [8 µM] + OLI16431 [8 µM]) was comparable to that of the untreated control sample at both 48 hours post electroporation and after 14 days in erythroid culture (**Fig. 14B**).

[0383] However, an artifact was observed when > 8 μM OLI16431 was electroporated into the cells alone, without RNP. Under these conditions, there was a false positive result when PCR amplification was performed 48 hour post electroporation, likely due to excessive ssODN within the system (see **Fig. 14A**, 8 μM and 12 μM for OLI16431 alone at 48 hours (black bars)). At lower doses of ssODN, and at later timepoints, this false positive was no longer apparent (**Fig. 14A**, 4 μM and 6 μM for OLI16431 alone at 48 hours (black bars) and 4 μM, 6 μM, 8 μM, and 12 μM for OLI16431 alone at 14 days (light grey bars)). In addition, the decrease in cell viability observed with the ssODN alone groups (**Fig. 14B**, dropping to ~57% with 12 μM OLI16431 alone at 48 hours (black bars)) may be a compounding factor. That there was no false positive result at ssODN doses of 4 μM and 6 μM suggests that this artifact may be due to excessive ssODN. These factors, along with the baseline HbF levels observed with the 8 μM and 12 μM OLI16431 alone groups (**Fig. 14A**, grey dots as compared to negative control), provide confidence that the substantial and sustained boost in RNP editing with the addition of ssODN is real and not an artifact.

Example 11: RNPs containing various Cpf1 and gRNA targeting the *HBG* promoter region support gene editing in human hematopoietic stem/progenitor cells which promotes induction of HbF protein expression in the erythroid progeny

[0384] Guide RNAs comprising SEQ ID NOs:1022, 1023, 1041-1093, 1098-1106 (**Table 13**) were complexed to various Cpf1 variant enzymes (SEQ ID NOs:1032, 1094-1097, 1107-1109, **Table 14**) to form various RNP complexes (**Table 15**). The RNPs contained gRNAs with modifications to the 5' end and/or modifications to the 3' end of the gRNA (**Table 13**). Guide RNAs comprising SEQ ID NOs:1022, 1023, 1041-1084, 1098-1106 (**Table 13**) have the same expected cleavage site at the distal CCAAT box target region, the related targeting domains contain the sequences set forth in SEQ ID NO:1002 (HBG1-1), SEQ ID NO:1254 (HBG1-1 – 21mer), SEQ ID NO:1256 (HBG1-1 – 22mer), SEQ ID NO:1258 (HBG1-1 – 23mer), for gRNA comprising a 20 mer, 21 mer, 22 mer, or 23 mer protospacer sequence respectively (**Table 15, Table 16**). In some cases gRNA targeting other positions within the HBG promoter were also tested, including guide RNAs SEQ ID NOs:1085-1096, comprising targeting domains containing the sequences set forth in SEQ ID NOs:1260 (AsCpf1 HBG1 Promoter-1 (21mer)), SEQ ID NO:1262 (AsCpf1 HBG1 Promoter-2 (21mer)) or SEQ ID NO:1264 (AsCpf1 HBG1 Promoter-6 (21mer)) (**Table 16, Table 17**). **Table 15** provides a listing of each RNP tested in Examples 11-13 and the SEQ ID NO of the gRNA and the SEQ ID NO of the Cpf1 variant that form each RNP complex. Additional information about each gRNA and Cpf1 variant may be found in **Table 13** and **Table 14**, respectively.

[0385] The gRNAs used in **Examples 11** and **12** were chemically synthesized. Chemicals for oligonucleotide synthesis were purchased from BioAutomation, Glen Research, Millipore Sigma, Sigma-Aldrich, ChemGenes, and Thermo Fisher Scientific. The solid support used for synthesis was either a Unylinker 2000 Å CPG resin, a 2'-TBDMS rU 2000 Å CPG resin or a 2'-O-methyl adenosine

(N-Bz) Icaa 2000 Å CPG resin from ChemGenes. RNA (TBDMS-protected) and DNA phosphoramidites were obtained from Thermo Fisher Scientific. In certain embodiments, phosphorothioates were introduced during a sulfurization step with a solution of DDTT (3-((dimethylamino-methylidene)amino)-3H-1,2,4-dithiazole-3-thione) from Glen Research. Oligonucleotides were synthesized using standard RNA and DNA phosphoramidite chemistry on either a BioAutomation MerMade 12 synthesizer or on a GE Äkta OligoPilot 100 synthesizer. Following synthesis, the oligonucleotides were cleaved from the solid support and deprotected in a two-step process using ammonium hydroxide/methylamine and TEA-3HF. After desalting, the oligonucleotides were purified using reversed-phase chromatography on a preparative HPLC.

[0386] First, the effect of editing using RNPs comprising gRNAs with modifications to the 5' end of the gRNAs was tested. Briefly, RNP complexes (6.0 µM and 12 µM) were delivered to 1 x 10^6 mPB CD34+ cells via MaxCyte electroporation, following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. The level of insertions / deletions at the target site was analyzed by Illumina sequencing (NGS) of the PCR amplified target site at 72 hours after electroporation. Results demonstrate that RNPs comprising gRNAs with no modification (RNP33, also referred to as "HBG1-1-AsCpf1 RNP," **Table 8**) or modifications to the 5' end of the gRNA including the addition of 5 nt RNA (RNP37), 10 nt RNA (RNP38), 25 nt RNA (RNP39), 60 nt RNA (RNP40), 5 nt DNA (RNP41), 10 nt DNA (RNP42), 25 nt DNA (RNP43), and 60 nt DNA (RNP44) (**Table 15**) supports on-target editing (**Fig. 15**).

[0387] Next, the effect of co-delivery of Cpf1 RNP with Sp182 dead RNP (dead gRNA comprising SEQ ID NO:1027 (**Table 8**) complexed with *S. pyogenes* Cas9 (SEQ ID NO:1033)) or ssODN OLI16431 (SEQ ID NO:1040, **Table 7**) was tested. Briefly, a dosing matrix with RNP33 (no 5' or 3' gRNA modification**, Table 15**) complexes at varying concentrations (6 µM, 8 µM, 8 µM, and 12 µM) were co-delivered with Sp182 RNP (8 µM, 8 µM, 6 µM, and 4 µM) or ssODN OLI16431 (8 µM, 8 µM, 6 µM, and 4 µM) to 5.25 x 10^6 mPB CD34+ cells via MaxCyte electroporation, following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. Following electroporation, cells were placed back to culture for a further 48 hours. Then, a fraction of the CD34 cells were split for gDNA extraction and indel quantification in the bulk cell population. In addition, to investigate the editing in phenotypic progenitors and phenotypic hematopoietic stem cells (HSCs), HSPC subpopulations were characterized (amongst the remainder CD34 cells) by immune-phenotyping (Notta 2011) and separated by fluorescence Activated Cell Sorting (FACS). Immune phenotyping at 48 hours post electroporation was performed by staining cells with antibodies against hCD34, hCD38, hCD45RA, hCD90, and hCD123. Phenotypic HSCs were defined as hCD34bright hCD38 hCD90+ hCD45RA-, and progenitors were defined as hCD34bright CD38+. These two populations were sorted by FACS and DNA was extracted to determine the editing levels in these sub-populations. The level of insertions / deletions at the target site was analyzed by Illumina

sequencing (NGS) of the PCR amplified target. Results demonstrate that RNP33 co-delivered with the "booster elements" Sp182 RNP or ssODN OLI16431 support on-target editing (**Figs. 16A-16B**) at levels higher than those observed when delivering RNP33 alone (**Fig. 15**).

[0388] Next, the effect of co-delivery of RNP33 (no 5' or 3' gRNA modification**, Table 15**), RNP43 (+25 DNA 5' gRNA modification, **Table 15**) or RNP34 (1xPS2-OMe + 1xOMe 3' gRNA modifications, **Table 15**) with Sp182 dgRNA (**Table 9**) or ssODN OLI16431 (**Table 7**) was tested. Briefly, RNP33, RNP34, or RNP43 complexes (8 μM) were co-delivered with Sp182 RNP (8 μM) or ssODN OLI16431 (8 μM) to 5.25 x 10^6 mPB CD34+ cells via MaxCyte electroporation, following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. Following electroporation, cells were placed back to culture for a further 48 hours prior to sorting into phenotypic progenitor and phenotypic HSC fractions for indel quantification. The level of insertions / deletions at the target site was analyzed by Illumina sequencing (NGS) of the PCR amplified target site at 48 hours after electroporation. Results demonstrate that RNPs containing gRNAs with 5' modification or 3' modifications co-delivered with the "booster elements" Sp182 RNP or ssODN OLI16431 support on-target editing (**Figs. 17**). Both booster elements tested provided comparable editing levels to RNP34 and RNP33. When compared to editing levels obtained without the booster element, while RNP33 editing was enhanced by both booster elements, RNP43 editing was however only increased with the addition of Sp182 RNP (**Figs. 15, 17**).

[0389] Next the effect of different Cpf1 proteins on RNP editing was tested. A stoichiometric comparison (gRNA:Cpf1) with RNPs comprising various Cpf1 proteins was performed. Briefly, RNPs (RNP64, RNP63, RNP45, **Table 15**) were delivered at a stoichiometry (gRNA:Cpf1 complexation ratio) of either 2 or 4, where the gRNA is in a molar excess. All RNPs were delivered via MaxCyte electroporation at 8 μM to 1 x 10^6 CD34+ cells following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. Following electroporation, cells were placed back to culture for a further 72 hours prior to indel quantification. Results demonstrated that there was no difference in editing rates with altered stoichiometry at the concentration tested (**Fig. 18**). RNP45 (comprising Cpf1 protein SEQ ID NO:1094) outperforms RNP63 (comprising Cpf1 protein SEQ ID NO:1095) and RNP64 (comprising Cpf1 SEQ ID NO:1109).

[0390] Next, the effect of Sp182 RNP or ssODN OLI16431 co-delivery with various RNPs containing different Cpf1 proteins was tested. Briefly, RNPs (RNP33, RNP64, RNP63, RNP45, **Table 15**) were delivered alone or in combination with Sp182 RNP or ssODN OLI16431. All reagents were delivered via MaxCyte electroporation at 8 μM to 1 x 10^6 CD34+ cells following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. Following electroporation, cells were placed back to culture for a further 72 hours prior to indel quantification. Results indicate the "booster elements" Sp182 RNP or ssODN OLI16431 enhanced editing for all RNPs tested (**Fig. 19**).

[0391] The effect of editing using RNPs containing gRNAs with various 5' DNA extensions, or RNP without a 5' DNA extension (RNP45) co-delivered with or without ssODN OLI16431 was tested. Briefly, following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3, RNPs (RNP45 (no 5' modification), RNP46, RNP47, RNP48, RNP49, RNP50, RNP51, RNP52, RNP53, RNP54, RNP55, RNP56, and RNP57, **Table 15**) at a concentration of 8 µM were delivered alone or co-delivered with 8 µM ssODN OLI16431 via MaxCyte electroporation to 1 x 10^6 CD34+ cells. Following electroporation, cells were placed back to culture prior to indel quantification. Results demonstrate that all RNPs support on-target editing (**Fig. 20**).

[0392] The effect of editing using RNPs including gRNAs having the same 5' DNA extension but different 3' gRNA modifications was tested to assess the impact of 3' gRNA modifications. Briefly, RNPs comprising gRNAs with matched 5' ends, but different 3' gRNA modifications (RNP49 vs. RNP58 and RNP59 v. RNP60) were delivered at a concentration of 8 µM via MaxCyte electroporation to 1 x 10^6 CD34+ cells following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. Following electroporation, cells were placed back to culture prior to indel quantification. In both comparisons, RNPs containing gRNA with 3' PS-OMe outperformed the unmodified 3' version at 24h post electroporation (**Fig. 21**).

[0393] Next, different concentrations of Cpf1 and gRNA for RNP58 were tested. Briefly, RNP58 (+25 DNA 5' gRNA modification and 1xPS-OMe 3' gRNA modification, **Table 15**) was delivered via MaxCyte electroporation to 1 x 10^6 CD34+ cells following 48 hours pre-stimulation at a stoichiometry (gRNA:Cpf1 complexation ratio) of either 2:1, 1:1 or 0.5:1 molar ratios. Following electroporation, cells were placed back to culture prior to indel quantification. At all doses tested, editing was best when RNP was complexed at a 2:1 ratio (**Figs. 22A-22B**).

[0394] The effect of editing using RNPs including gRNAs having the same 5' DNA extension but different 3' gRNA modifications was further tested to assess the impact of 3' gRNA modifications. Briefly, RNPs comprising gRNAs with matched 5' ends, but different 3' gRNA modifications (RNP58, RNP2, RNP3, RNP4, RNP5, RNP6, RNP7, RNP8, RNP9, RNP10) were delivered at varying concentrations (1 µM, 2 µM, 4 µM) via MaxCyte electroporation to 1 x 10^6 CD34+ cells following 48 hours pre-stimulation. Following electroporation, cells were placed back to culture prior to indel quantification. Results indicate that all RNPs support on-target editing (**Figs. 23A-23B**).

[0395] Next, editing by RNPs that include gRNAs with 3' gRNA modifications or 5' and 3' modifications that target various regions of the HBG promoter were tested. Those include guide RNAs SEQ ID NOs :1085-1096, comprising targeting domains SEQ ID NOs:1260 (AsCpf1 HBG1 Promoter-1 (21mer)), 1262 (AsCpf1 HBG1 Promoter-2 (21mer)), 1264 (AsCpf1 HBG1 Promoter-6 (21mer)) (**Table 16**, **Table 17**). Instead of the distal CCAAT box target region, those gRNAs are configured to provide an editing event within regions selected from **Table 11**. Briefly, RNPs including gRNAs containing an unmodified 5' gRNA and a 1xPS-OMe 3' gRNA modification

(RNP11, RNP16, RNP19, and RNP22, **Table 15**), RNPs including gRNAs containing a +20 DNA +2xPS 5' gRNA modification and a 1xPS-OMe 3' gRNA modification (RNP12, RNP21, and RNP24, **Table 15**), RNPs including gRNAs containing a +25 DNA 5' gRNA modification and a 1xPS-OMe 3' gRNA modification (RNP58 and RNP20, **Table 15**) were delivered at a concentration of 8 µM via MaxCyte electroporation to 1 x 10^6 CD34+ cells following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. Following the editing of mPB CD34+ cells, *ex vivo* differentiation into the erythroid lineage was performed for 18 days (Giarratana 2011), with gDNA being isolated on day 14 of culture for indel analysis. Briefly, RNP was delivered to CD34+ cells, as described above. Following 48 hours recovery in X-Vivo 10 media supplemented with SCF, TPO and FLT3, the treated cells were counted and transferred to erythroid differentiation media, with cell counts and feeds occurring on days 4, 7, 10 and 14, with erythroid collection at day 18. These CD34+ derived erythroid cells were then counted and lysed in HPLC grade water, before filtering to removed cell debris. Then, relative expression levels of gamma-globin chains (over total beta-like globin chains) was measured for each sample by UPLC, with protein separation being achieved by gradually increasing the ratio of acetonitrile with 0.1% trifluoroacetic acid, to water with 0.1% trifluoroacetic acid (**Figs. 24A-24C**). **Figs 24A-24C** showed all RNP supported on-target editing. However only editing at certain target sites give rise to increased HBF expression. The same experiment was performed with varying concentrations (0, 1 µM, 2 µM, 4 µM) of RNP58 (**Fig. 25**).

[0396] Next the effect of different Cpf1 proteins on RNP editing was tested. Briefly, RNP58, RNP26, RNP27, and RNP28 (**Table 15**) including gRNAs comprising SEQ ID NO:1051 (+25 DNA 5' gRNA modification and a 1xPS-OMe 3' gRNA modification, **Table 13**) complexed with varying Cpf1 proteins were delivered via MaxCyte electroporation at 8 µM to 1 x 10^6 CD34+ cells following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. Results demonstrated that all RNPs support on-target editing (**Fig. 26**).

[0397] Next the effect of different RNPs containing gRNAs with various 5' gRNA modifications and the same 3' modification was tested. Briefly, RNPs (RNP58 (+25 DNA 5' gRNA modification and a 1xPS-OMe 3' gRNA modification), RNP29 (+25 DNA + 2xPS 5' gRNA modification and a 1xPS-OMe 3' gRNA modification), RNP30 (PolyA RNA + 2xPS 5' gRNA modification and a 1xPS-OMe 3' gRNA modification), and RNP31 (PolyU RNA + 2xPS 5' gRNA modification and a 1xPS-OMe 3' gRNA modification) (**Table 15**)) were delivered via MaxCyte electroporation at 1 µM, 2 µM, or 4 µM to 1 x 10^6 CD34+ cells following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3 (RNP30 was not tested at 1 µM due to availability of cells). Results demonstrated that all RNPs support on-target editing (**Fig. 27**).

[0398] Next the effect of different Cpf1 proteins on RNP editing was tested. Briefly, RNP58, RNP27, and RNP26 (**Table 15**) including gRNAs comprising SEQ ID NO:1051 (+25 DNA 5' gRNA modification and a 1xPS-OMe 3' gRNA modification, **Table 13**) complexed with varying Cpf1

proteins were delivered via MaxCyte electroporation at 2 μM or 4 μM to 1 x 10^6 CD34+ cells following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. To investigate the editing in bulk CD34, phenotypic progenitors, and phenotypic hematopoietic stem cells (HSCs), HSPC subpopulations were characterized. Thus, following electroporation, cells were placed back to culture for a further 48 hours prior to sorting into progenitor and HSC fractions for indel quantification. The level of insertions / deletions at the target site was analyzed by Illumina sequencing (NGS) of the PCR amplified target site at 48 hours after electroporation. Results demonstrated that all RNPs support on-target editing in bulk CD34, phenotypic progenitors, and phenotypic hematopoietic stem cells (**Fig. 28**).

[0399] Next, the effect of co-delivery of RNP with Sp182 RNP (dead gRNA comprising SEQ ID NO:1027 (**Table 8**) complexed with *S. pyogenes* Cas9 (SEQ ID NO:1033)) or ssODN OLI16431 (SEQ ID NO:1040, **Table 7**) on RNP containing different Cpf1 proteins was tested. Briefly, RNP61, RNP62, RNP34 (**Table 15**) were co-delivered at 8 μM with Sp182 RNP (8 μM) or ssODN OLI16431 (8 μM) to 25 x 10^6 mPB CD34+ cells via MaxCyte electroporation, following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. To investigate the editing in bulk CD34, phenotypic progenitors, and phenotypic hematopoietic stem cells (HSCs), HSPC subpopulations were characterized. Thus, following electroporation, cells were placed back to culture for a further 48 hours prior to sorting into progenitor and HSC fractions for indel quantification. The level of insertions / deletions at the target site was analyzed by Illumina sequencing (NGS) of the PCR amplified target site at 48 hours after electroporation. Results demonstrate that RNP co-delivered with the "booster elements" Sp182 RNP or ssODN OLI16431 support on-target editing (**Fig. 29**). A fraction of these cells were also cryopreserved 24 hours post electroporation to be further characterized in an *in vivo* engraftment model (see **Example 12**).

[0400] Next, the effect of RNP58 and RNP32 editing was tested. Briefly, RNP58 and RNP32 (**Table 15**) were delivered at 2 μM to 6 x 10^6 mPB CD34+ cells via MaxCyte electroporation, following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. To investigate the editing in bulk CD34, phenotypic progenitors, and phenotypic HSCs, HSPC subpopulations were characterized. Thus, following electroporation, cells were placed back to culture for a further 48 hours prior to sorting into progenitor and HSC fractions for indel quantification. The level of insertions / deletions at the target site was analyzed by Illumina sequencing (NGS) of the PCR amplified target site at 48 hours after electroporation. Results demonstrate that RNP58 and RNP32 support on-target editing (**Fig. 30**).

[0401] RNP58 and RNP1 (**Table 15**) were delivered at concentrations of 2 μM to 8 μM to 25 x 10^6 mPB CD34+ cells via MaxCyte electroporation, following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. To investigate the editing in bulk CD34, phenotypic progenitors, and phenotypic hematopoietic stem cells (HSCs), HSPC subpopulations were

characterized. Thus, following electroporation, cells were placed back to culture for a further 48 hours prior to sorting into progenitor and HSC fractions for indel quantification. The level of insertions / deletions at the target site was analyzed by Illumina sequencing (NGS) of the PCR amplified target site at 48 hours after electroporation. Results demonstrate that the RNP tested support on-target editing (**Fig. 31**). A fraction of these cells were also cryopreserved 24 hours post electroporation for further characterized in an *in vivo* engraftment model (see **Example 12**).

Example 12: Infusion of edited mPB CD34+ cells into NOD,B6.SCID Il2rγ-/- Kit(W41/W41) mice results in long term engraftment and HbF induction

[0402] To determine whether delivery of RNP34 and RNP33 (**Table 15)** co-delivered with ssODN OLI16431 (SEQ ID NO:1040, **Table 7**) achieves edits in long term repopulating hematopoietic stem cells, human adult CD34+ cells from mobilized peripheral blood (mPB) were infused into nonirradiated NOD,B6.SCID Il2rγ-/- Kit(W41/W41) (Jackson lab stock name: NOD.Cg-Kit<W-41J> Tyr<+> Prkdc<scid> Il2rg<tm1Wjl>/ThomJ) ("NBSGW") mice. Briefly, mPB CD34+ cells at 62.5 x 10^6/mL were electroporated via MaxCyte electroporation with RNP at a dose of 8 µM and 6 µM ssODN OLI16431 following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. After 24 hours, mCD34+ cells were cryopreserved. Five days later, mPB CD34+ cells were thawed and infused into NBSGW mice at 1 million cells per mouse via intravenous tail vein injection. Eight weeks later, mice were euthanized and bone marrow (BM) was collected from femurs, tibias, and pelvic bones. Bone marrow sub-populations of cells respectively identified as CD15+, CD19+, glycophorin A (GlyA, CD235a+), and lineage-negative CD34+ were isolated by FACS, and DNA was extracted to determine the editing levels in each of these fractions. **Fig. 32** depicts the frequency of indels, as determined by next generation sequencing, of unfractionated BM, or flow-sorted individual BM sub-populations.

[0403] Next, to determine whether delivery of RNP34 or RNP33 (**Table 15)** co-delivered with Sp182 RNP (dead gRNA comprising SEQ ID NO:1027 (**Table 8**) complexed with *S. pyogenes* Cas9 (SEQ ID NO:1033)) achieves edits in long term repopulating hematopoietic stem cells, human adult CD34+ cells from mobilized peripheral blood (mPB) were infused into nonirradiated NBSGW mice. Briefly, mPB CD34+ cells at 62.5 x 10^6/mL were electroporated via MaxCyte electroporation with RNP at varying doses and varying doses of Sp182 RNP following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. After 24 hours, mCD34+ cells were cryopreserved. Five days later, mPB CD34+ cells were thawed and infused into NBSGW mice at 1 million cells per mouse via intravenous tail vein injection. Eight weeks later, mice were euthanized and bone marrow (BM) was collected from femurs, tibias, and pelvic bones. Bone marrow sub-populations of cells respectively identified as CD15+, CD19+, glycophorin A (GlyA, CD235a+), and lineage-negative CD34+ were isolated by FACS, and DNA was extracted to determine the editing levels in each of

these fractions. **Fig. 33A** depicts the frequency of indels, as determined by next generation sequencing, of unfractionated BM, or flow-sorted individual BM sub-populations.

[0404] Lastly, long term HbF induction by BM-derived CD34+ cells was analyzed. An aliquot of BM cells were cultured under erythroid differentiation conditions for 18 days (Giarratana 2011), and evaluated for HbF expression by UPLC. Briefly, unfractionated BM cells extracted from mice 8 weeks after infusion were placed in erythroid culture conditions for 18 days. Cell counts and feeds occurred on days 7, 10 and 14, with erythroid collection at day 18. These bone marrow derived erythroid cells were then counted and lysed in HPLC grade water before filtering to removed cell debris. Then, relative expression levels of gamma-globin chains (over total beta-like globin chains) was measured for each sample by UPLC, with protein separation being achieved by gradually increasing the ratio of acetonitrile with 0.1% trifluoroacetic acid, to water with 0.1% trifluoroacetic acid (**Fig. 33B**). These data demonstrate that robust long-term HbF induction is achieved by RNP34 and RNP33 co-delivered with Sp182 RNP editing of human CD34+ cells.

[0405] Next, to determine whether delivery of RNP61 or RNP62 (**Table 15)** co-delivered with ssODN OLI16431 (SEQ ID NO:1040, **Table 7**) achieves edits in long term repopulating hematopoietic stem cells, human adult CD34+ cells from mPB were infused into nonirradiated NBSGW mice. Briefly, mPB CD34+ cells at 62.5 x 10^6/mL were electroporated via MaxCyte electroporation with 8 µM RNP and 8 µM ssODN OLI16431 following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. After 24 hours, mCD34+ cells were cryopreserved. Four days later, mPB CD34+ cells were thawed and infused into NBSGW mice at 1 million cells per mouse via intravenous tail vein injection. Eight weeks later, mice were euthanized and bone marrow (BM) was collected from femurs, tibias, and pelvic bones. Bone marrow sub-populations of cells respectively identified as CD15+, CD19+, glycophorin A (GlyA, CD235a+), and lineage-negative CD34+ were isolated by FACS, and DNA was extracted to determine the editing levels in each of these fractions. **Fig. 34A** depicts the frequency of indels, as determined by next generation sequencing, of unfractionated BM, or flow-sorted individual BM sub-populations.

[0406] Lastly, long term HbF induction by BM-derived CD235a+ (GlyA+) erythroid cells was analyzed. An aliquot of BM cells were cultured under erythroid differentiation conditions for 18 days and evaluated for HbF expression by UPLC. Briefly, unfractionated BM cells extracted from mice 8 weeks after infusion were placed in erythroid culture conditions for 18 days. **Fig. 34B** depicts the HbF expression, calculated as gamma/beta-like chains (%) by erythroid cells. These data demonstrate that robust long-term HbF induction is achieved by RNP61 and RNP62 co-delivered with ssODN OLI16431 editing of human CD34+ cells.

[0407] Next, to determine whether delivery of RNP1 or RNP58 (**Table 15)** achieves edits in long term repopulating hematopoietic stem cells, human adult CD34+ cells from mPB were infused into nonirradiated NBSGW mice. Briefly, mPB CD34+ cells at 62.5 x 10^6/mL were electroporated via

MaxCyte electroporation with 4 μM or 8 μM RNP1 or 2 μM, 4 μM, or 8 μM RNP58 following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. After 24 hours, mCD34+ cells were cryopreserved. Four days later, mPB CD34+ cells were thawed and infused into NBSGW mice at 1 million cells per mouse via intravenous tail vein injection. Eight weeks later, mice were euthanized and bone marrow (BM) was collected from femurs, tibias, and pelvic bones. Human chimerism and lineage reconstitution (CD45+, CD14+, CD19+, glycophorin A (GlyA, CD235a+), lineage, and CD34+, and mouse CD45+ marker expression) in BM was determined by flow cytometry and analyzed (**Fig. 35**). The frequency of GlyA+ cells was calculated as GlyA+ cells/total cells in BM. Human chimerism was defined as human CD45/(human CD45+mCD45).

[0408] Similar chimerism and lineage distributions were achieved 8-weeks post-transplant by RNP-transfected mPB CD34+ cells compared to mock-transfected mPB CD34+ cells demonstrating that editing is comparable with retaining the engraftment potential of hematopoietic stem cells. **Fig. 36** depicts the indels, as determined by next generation sequencing, of unfractionated BM, at 8 weeks post-infusion. **Fig. 37** shows the frequency of indel rate in BM, CD15+, CD19+, glycophorin A (GlyA, CD235a+), and lin-CD34+ cells with 2 μM, 4 μM, or 8 μM of RNP58.

[0409] Long term HbF induction by CD235a+ (GlyA+) erythroid cells, derived from edited CD34+ cells was also analyzed. GlyA+ cells obtained from bone marrow, were isolated by fluorescence Activated Cell Sorting (FACS), collected and lysed in HPLC grade water. Lysates were then evaluated for HbF expression by UPLC. **Fig. 38** depicts the HbF expression, calculated as gamma/beta-like chains (%) by GlyA+ cells. These data demonstrate that robust long-term HbF induction is achieved RNP58 editing of human CD34+ cells at various RNP58 concentrations.

[0410] Importantly, $CD34^+$ cells that were electroporated with varying concentrations of RNP1 or RNP58 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. 39**).

Example 13: NHEJ mediated deletions of 4 nucleotides or larger at the distal CCAAT box region of HBG promoter are maintained long-term and promote elevated HbF expression.

[0411] Delivery of RNP containing the Cas9 or Cpf1 enzyme targeting the HBG promoter region (**Fig. 41A**) results in the generation of a multitude of insertions and deletions (indels). These include indels derived from microhomology mediated end joining (MMEJ) and non-homologous end joining (NHEJ) repair mechanisms. As shown below, the MMEJ repair mechanism is not well utilized during editing of the long-term stem cell population (HSC) and this type of edit is lost over time.

[0412] Briefly, mPB CD34+ cells at 62.5 x 10^6/mL were electroporated via MaxCyte electroporation with Sp35 RNP following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. Sp35 RNP comprises Sp35 gRNA (comprising the targeting domain of SEQ ID NO:339 (i.e., CUUGUCAAGGCUAUUGGUCA (RNA)); SEQ ID NO:917 (i.e.,

CTTGTCAAGGCTATTGGTCA (DNA)) complexed with *S. pyogenes* wildtype (Wt) Cas9 protein. After 24 hours, an aliquot of mPB CD34+ cells was collected for pre-infusion (CD34) indel analysis, another aliquot was placed in erythroid differentiation for indel analysis in the progeny of erythroid progenitors. The rest of the cells were cryopreserved and stored in liquid nitrogen until the initiation of the engraftment study. At the time of infusion, mPB CD34+ cells were thawed and infused into nonirradiated NOD,B6.SCID Il2rγ-/- Kit(W41/W41) (Jackson lab stock name: NOD.Cg-Kit<W-41J> Tyr<+> Prkdc<scid> Il2rg<tm1Wjl>/ThomJ) ("NBSGW") mice at 1 million cells per mouse via intravenous tail vein injection. Sixteen weeks later, mice were euthanized and bone marrow (BM) was collected from femurs, tibias, and pelvic bones and lineage-negative CD34+ were isolated by FACS, and DNA was extracted from cells collected at pre-infusion, in their erythroid progeny (derived *in vitro*) and in cells collected at 16 weeks *in vivo* to determine the indels (insertions/deletions) using the following primers: Forward = CATGGCGTCTGGACTAGGAG (SEQ ID NO:1266) and Reverse = AAACACATTTCACAATCCCTGAAC (SEQ ID NO:1267). The read sequence surrounding every detected deletion was analyzed to identify whether it was likely the result of DSB repair by MMEJ or NHEJ pathways. MMEJ is distinguished from NHEJ by its use of microhomology sequences. MMEJ repair relies on strand resection and annealing of proximally located repeated stretches of nucleotides (microhomologies), surrounding the DSB. The resulting deletion removes one of the microhomology sequences together with the entire intervening sequence between the two microhomologies. Thus, probable MMEJ-mediated deletion can be identified by searching for the presence of a nucleotide sequence at either end of the deleted sequence that is repeated in the region immediately flanking the other end of the deletion. Based on this, deletions were classified as "MMEJ" if stretches of 2 base pairs (bp) or more were detected at the deletion boundary and repeated in the region immediately flanking the other end of the deletion. All other deletions were classified as NHEJ.

[0413] Prior to infusion, ~30% of indels were derived from MMEJ repair (**Fig. 41B**, CD34, black striped bar = MMEJ, white bar = NHEJ). In erythroid progenitors (based on indels detected after erythroid differentiation of CD34 cells), MMEJ indels represented ~36% of indels (**Fig. 41B**, Prog, black striped bar = MMEJ, white bar = NHEJ). Following 16 weeks engraftment, the majority of the MMEJ indels were lost (**Fig. 41B**, HSC, black striped bar = MMEJ, white bar = NHEJ) with the NHEJ derived edits being maintained. This observation has also been observed by others (Bauer 2019, Wu 2019, Weber 2020), and does not appear to be site specific. It was attributed to the transition from a predominant population of actively cycling, and MMEJ-prone short-lived progenitors that constitute most of the CD34+ cells pre-infusion, to the output of more quiescent (and less MMEJ-prone) self-renewing hematopoietic stem cells (HSCs) at later timepoints *in vivo*. While HSCs provide hematopoietic reconstitution post-engraftment long-term, they represent a very rare population within the CD34+ bulk population pre-infusion. This finding highlights the need to use an

editing tool that generates HbF-inducing indels through NHEJ-mediated repair rather than MMEJ repair to achieve durable HbF expression.

[0414] In addition, genotype to phenotype analysis at the distal CCAAT-box region of the *HBG1/2* promoters identified mutations leading to most elevated HbF expression. A single cell experiment was performed to evaluate the distribution of gamma chain expression levels in erythroid cells derived from mPB CD34+ cells edited at the distal CCAAT box region. Indels were generated at this site using either Sp35 RNP or RNP34 (**Table 15**) + Sp182-Cas9-RNP (**Fig. 5F**, **Table 8**). Briefly, mPB CD34+ cells at 6.25 x 10^6/mL (100 µl) were electroporated via MaxCyte electroporation with 4 µM Sp35 RNP or 8 µM HBG1-1-AsCpf1H800A-RNP co-delivered with 8 µM Sp182-Cas9-RNP (all RNPs at a molar ratio of 4:1 gRNA:RNA guided nuclease) following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. Sp35 RNP comprises Sp35 gRNA (comprising the targeting domain of SEQ ID NO:339 (i.e., CUUGUCAAGGCUAUUGGUCA (RNA)); SEQ ID NO:917 (i.e., CTTGTCAAGGCTATTGGTCA (DNA)) complexed with *S. pyogenes* wildtype (Wt) Cas9 protein. RNP34 comprises HBG1-1-AsCpf1H800A-RNP (composed of His-AsCpf1-sNLS-sNLS-H800A (SEQ ID NO:1032) complexed to HBG1-1 gRNA with a 3' modification as shown in **Table 8** (SEQ ID NO:1041)). After being left to recover for 48 hours post-electroporation, the cells were sorted by Fluorescence Activated Cell Sorting (FACS) at 1-cell / well in non-tissue culture treated 384-well plates. The cells were then differentiated and expanded clonally for 18 days into the erythroid lineage (adapted from Giarratana 2011). UPLC analysis was performed to determine the distribution of G-gamma chain expression levels (percentage of G-gamma/[total beta-like chains]) in the clonal erythroid progeny of cells derived from the total population of mPB-CD34+ cells initially edited. Sequencing analysis (allele identification and detection of indels using the following primers: Forward = CATGGCGTCTGGACTAGGAG (SEQ ID NO:1266) and Reverse = AAACACATTTCACAATCCCTGAAC (SEQ ID NO:1267)) and ddPCR (loss of 4.9kb fragment deleting g-Gamma coding sequence) was performed on gDNA from an aliquot of each clonal culture collected on day 14, prior to enucleation, and enabled identification of clones containing only one allele coding for G-gamma, in association with a specific indel. Indels disrupting more than 3 nucleotides were generally associated with more elevated HbF expression compared to smaller indels (**Fig. 41C**). This finding highlights that using an editing tool that generates larger deletions at the distal CCAAT box region will provide more elevated levels of HbF expression, likely by more effectively disrupting the binding at this site of regulatory factors repressing HBG expression. Notably, most of the larger deletions evaluated were detected in clones derived from the CD34 cells edited with the Cpf1 RNP, while most of the smaller deletions evaluated were detected from clones derived from the CD34 cells edited with the Cas9 RNP (**Fig. 41D**), suggesting that larger indels occur more frequently with the Cpf1, leading to a potentially higher fraction of cells with elevated HbF in the erythroid progeny of CD34 cells edited with Cpf1 compared to Cas9. Consistent with this result,

distribution of gamma chain expression levels (percentage of gamma chains/[total beta-like chains]) in the clonal erythroid progeny of cells derived from the total population of mPB-CD34+ cells initially edited by either the Cpf1 or Cas9 RNP showed a higher frequency of cells with high HbF in the Cpf1 sample compared to the spCas9 sample (**Fig. 41E**). Of the clones analyzed derived from the CD34+ cells edited with the Cpf1 RNP, 83.1% had gamma chains levels above the median level detected in erythroid clones derived from mock electroporated mPB CD34+ cells (median value = 18.8%), 64.2% had gamma chains levels exceeding 30% of total beta-like and 35.8% had gamma chains levels exceeding 48.8% of total beta-like (30% +18.8% median level in control cells). Of the clones analyzed derived from the CD34+ cells edited with the Cas9 RNP, 71.4% had gamma chains levels above the median level detected in erythroid clones derived from mock electroporated mPB CD34+ cells (median value = 18.8%), 49.5% had gamma chains levels exceeding 30% of total beta-like and 23.1% had gamma chains levels exceeding 48.8% of total beta-like (30% +18.8% median level in control cells).

Example 14: Cpf1 RNP targeting the distal CCAAT box region of the HBG promoter is efficient at generating durable NHEJ mediated deletions of 4 nucleotides or larger

[0415] RNP, consisting of guide RNA targeting the distal CCAAT box region of the HBG promoter complexed with either Cpf1 or Cas9 enzyme was electroporated into mobilized peripheral blood (mPB) derived CD34+ cells using the MaxCyte GT (MaxCyte, Inc.) electroporation device. Briefly, mPB CD34+ cells at 62.5 x 10^6/mL were electroporated with 4 µM Sp35 RNP (at a molar ratio of 2:1 gRNA:RNA-guided nuclease) or 8 µM RNP58 (**Table 15**) (at a molar ratio of 4:1 gRNA:RNA-guided nuclease) following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. Sp35 RNP comprises Sp35 gRNA (comprising the targeting domain of SEQ ID NO:339 (i.e., CUUGUCAAGGCUAUUGGUCA) (RNA)); SEQ ID NO:917 (i.e., CTTGTCAAGGCTATTGGTCA) (DNA)) complexed with *S. pyogenes* wildtype (Wt) Cas9 protein.

[0416] The level and profile of insertions / deletions at the target site was analyzed 24 hours after electroporation. The fraction of small indels present in the samples was assessed by Illumina amplicon sequencing (ILL-seq), following library preparation method and analysis. A 15 bp window around the expected cut site was used in the analysis to calculate editing rates. The oligonucleotides and amplicon used to generate the targeted amplicon sequencing products are provided in **Fig. 55A** and **Fig. 55B**.

[0417] To characterize editing by RNP32, the indels generated in each sample were analyzed and processed. The results were summarized by looking at 1) percentage deletion for each base within the target region, 2) distribution of insertion and deletion center positions, and 3) distribution of insertion and deletion lengths. To perform these analyses, the cigar strings for the reads in the alignment bam files were processed, using the following steps:

1) Group indels appearing in the same read.

2) Create indel_id for indels from each read. Indel_id is a string identifying the indel, shown as indel_start_position + _ + indel_length + _ + ID. Where ID is NA for deletions and for insertions is the sequence inserted. See example in **Fig. 55D**.
3) Group reads with the same indel_id.
4) Count number of reads with the same indels and calculate their total fractions (including wt) and fractions in indels (excluding wt). Fractions in indels allow for a comparison between samples when they have different amounts of total editing.

[0418] Deletions were classified as "MMEJ" if stretches of 2 base pairs or more were deletion boundary and repeated in the region immediately flanking the other end of the deletion. All other deletions were classified as NHEJ. **Table 24** shows the indels detected in the samples edited by RNP58 or Sp35 RNP.

[0419] Delivery of RNP containing the Cas9 or Cpf1 enzyme targeting the HBG promoter region (**Fig. 41A**) results in the generation of a multitude of insertions and deletions (indels). Different enzymes have different mechanistic properties; including relative to their engagement with the DNA target site and their cleavage activity, and leave different DNA ends that impact the repair outcome (Sternberg 2014, Strohkendl 2018, Swarts 2018). Notably, Cpf1 editing results in a four nucleotide 5' overhang, which is quite different from the blunt ends of SpCas9 edits (see **Fig. 41A**).
Evaluation of the editing profile mediated by a Cpf1 and Cas9 RNP targeting the distal box (RNP58 and Sp35RNP, respectfully) showed that indels generated by both enzymes were generally centered at the distal CCAAT box region (**Fig. 42H**), and the most commonly deleted base was at position TSS:-113 for both RNPs (**Fig. 42E** and **42F**). However, evaluation of individual indels, as defined by their position relative the HBG-TSS, length, and sequence in the case of insertions, demonstrates differences in the editing outcome after editing this site with the SpCas9 or Cpf1 RNP. **Figs. 42C** and **42D** show the top 20 indels generated by the Cpf1 and Cas9 RNP, respectively. The most abundant indel generated by Cas9 was the MMEJ-meditated 13 bp deletion 157_-13_NA (HBG1/2 c.-102 to -114, **Table 24**), more commonly known as the 13 bp HPFH deletion, detected at 31.88% in the Cas9 sample compared to 3.36% in the Cpf1 sample (**Figs. 42D** and **42C**, respectively). The most abundant indel generated by Cpf1, was the MMEJ mediated 18 nt deletion 159_-18_NA (HBG1/2 c.-104 to -121, **Table 24**), which was detected at 18.53% in the Cpf1 sample compared to 1.35% in the Cas9 sample (**Figs. 42C** and **42D**, respectively). The five most common NHEJ indels detected in each sample were deletions of 3 to 7 bp in length for the Cpf1 sample and were deletions of 1 to 4 bp and a 1 bp insertion for Cas9 (**Figs. 42C** and **42D**, respectively). The full list of indels detected at > 0.1% in any of the two samples is shown in **Table 24**.

[0420] While most indels detected at > 0.1% in any of the two samples were also detected in the other sample (above or below 0.1%), their relative frequencies amongst all indels were different depending on the enzyme (**Table 24** and **Fig. 42I**). Notably, insertions were detected at 14.09% in the

Cas9 (Sp35 RNP) sample but were rarely detected in the Cpf1 (RNP58) sample (0.47 %). Also, larger indels were enriched in CD34 cells edited with the Cpf1 RNP (RNP58) compared to the Cas9 RNP (Sp35 RNP) (**Fig. 42J**). **Fig. 42G** shows the distribution of indel length generated by each RNP, and **Fig. 42A** shows the distribution of indel length generated by each RNP for indels categorized as NHEJ-mediated. Most indels generated by Cpf1 were larger than 3 bp (**Fig. 42G**), which was shown to be associated with more elevated HbF in erythroid cells (**Fig. 41C**). Cas9 also generated deletions larger than 3 bp, however, those were primarily generated via the MMEJ repair pathway, which are expected to be less frequently observed *in vivo* after infusion, as this repair pathway is poorly utilized by the rare population of HSC within the CD34+ cells. NHEJ is the major repair pathway utilized by HSC (**Fig. 41B**), thus evaluation of the NHEJ-mediated indel profile provides information on the type of indels that are expected to be maintained *in vivo* at long-term after infusion of CD34+ cells.

[0421] When compared to Cas9 RNP (Sp35 RNP), the Cpf1 RNP (RNP58) showed a shift to larger NHEJ mediated indels (**Fig. 42A**), with the most commonly observed NHEJ indel length being 1 bp for Cas9 compared to 5 bp for Cpf1, indicating a more favorable CD34+ cells editing profile by Cpf1 for both robust and durable induction of HbF.

[0422] The percentage of > 3 bp NHEJ mediated indels, > 3 bp MMEJ mediated indels, and ≤ 3 bp indels at the distal CCAAT box resulting from editing by RNP58 (Cpf1) or Sp35 RNP (Cas9) is shown in **Fig. 42B** (indels generated by SpCas9 (left pie chart): NHEJ > 3 bp = 21.42%; ≤ 3 bp = 48.84%; MMEJ > 3 bp = 29.74%; indels generated by RNP58 (right pie chart): NHEJ > 3 bp = 64.86%; ≤ 3 bp = 11.11%; MMEJ > 3 bp = 24.02%). Cas9 (Sp35 RNP)-edited cells comprised a large fraction of small indels less than 3 bp in length, and indels> 3 bp in length (associated with highest levels of HbF) were largely mediated by the MMEJ pathway (**Fig. 42B**). Instead, close to 90% of indels mediated by Cpf1 (RNP58) were > 3 bp in length, approximately two third of those mediated by NHEJ (**Fig. 42B**), suggesting that the high fraction of indels > 3 bp in length is expected to be maintained at long term *in vivo*, and the ratio of MMEJ-mediated indels is reduced in favor of NHEJ-mediated indels (**Fig. 41B**). Taken together, this highlights that the most durable NHEJ mediated indels are productive (> 3 bp) with RNP58, whereas Sp35 RNP produces a much lower fraction of productive indels, many of which are not durable (MMEJ meditated). This demonstrates that RNP58 is an advantageous editing tool that generates HbF-inducing indels through NHEJ-mediated repair rather than solely via MMEJ repair to achieve durable HbF expression.

Example 15: Infusion of RNP32 edited mPB CD34+ cells into NOD.B6.SCID Il2rγ-/- Kit (W41/W41) mice results in long term engraftment, indel maintenance, and high HbF induction

[0423] To determine whether RNP32 achieves edits in long term repopulating hematopoietic stem cells, human adult CD34+ cells from mobilized peripheral blood (mPB) were infused into nonirradiated NOD,B6.SCID Il2rγ-/- Kit(W41/W41) (Jackson lab stock name: NOD.Cg-Kit<W-41J> Tyr<+> Prkdc<scid> Il2rg<tm1Wjl>/ThomJ) ("NBSGW") mice. Briefly, mPB CD34+ cells at 62.5 x

10^6/mL were electroporated via MaxCyte electroporation with 6 μM RNP32 (**Table 15**) (at a molar ratio of 2:1 gRNA:RNA guided nuclease) or buffer only ("Mock") following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. After a further 24 hours, mCD34+ cells were cryopreserved. A portion of the cells was placed in *in vitro* culture for up to 72h post-electroporation and gDNA collected every day for indel analysis. On the day of infusion, mPB CD34+ cells were thawed and infused into NBSGW mice at 1 million cells per mouse via intravenous tail vein injection. Sixteen weeks later, mice were euthanized and bone marrow (BM) was collected from femurs, tibias, and pelvic bones.

[0424] Genomic DNA was isolated from both the pre-infused CD34+ cells (24-72 hours post electroporation) and from the bone marrow following 16 weeks engraftment. At 24 hours post electroporation, approximately 92% indels were achieved, as measured by Illumina sequencing of the on-target PCR product using the following primers: Forward = CATGGCGTCTGGACTAGGAG (SEQ ID NO:1266) and Reverse = AAACACATTTCACAATCCCTGAAC (SEQ ID NO:1267).

[0425] Analysis of the indel length distribution and classification of indels of length ≤3bp, and of indels of length >3bp either categorized as MMEJ or NHEJ mediated indels, was conducted on the preinfusion cell population after 24h to 72h of in vitro culture post-electroporation. Deletions were classified as "MMEJ" if stretches of 2 base pairs (bp) or more were detected at the deletion boundary and repeated in the region immediately flanking the other end of the deletion. All other deletions were classified as NHEJ.

[0426] The distribution of indel length, shown in **Figs. 43 H-K**, and for NHEJ-indels only in **Figs. 43D-G**), was very similar across all timepoints (**Figs. 43D** and **43K**). Insertions were rarely detected, and the most commonly observed deletion size was -18, which notably corresponds to the size of the 18 bp MMEJ deletion HBG1/2 c.-104 to -121 and is the most frequent repair outcome after editing this site with Cpf1 (**Figs. 42C, 57, Table 25**). The most commonly observed deletion size amongst NHEJ indels was 5 bp as shown in **Figs. 43 D-G**.

[0427] At all timepoints, indels of size larger than 3 bp, which were observed to be associated with the highest levels of HbF induction (**Fig. 41C**), represented close to 90% of all indels (**Fig. 43A-C**). About two thirds of those were mediated by NHEJ, suggesting that the high fraction of indels > 3 bp length is expected to be maintained at long term *in vivo*, where the ratio of MMEJ-mediated indels is reduced in favor of NHEJ-mediated indels (**Fig. 41B**).

[0428] Following 16 weeks engraftment, ~96% indels were measured demonstrating indel maintenance within the long term repopulating CD34+ cells (**Fig. 43A**).

[0429] Next, sub-populations of cells within the bone marrow were identified via FACS analysis to determine levels of multilineage engraftment. Human chimerism and lineage reconstitution (CD45+, CD15+, CD19+, glycophorin A (GlyA, CD235a+), lineage, and CD34+, and mouse CD45+ marker expression) in BM was determined by flow cytometry and analyzed. The frequency of GlyA+ cells

was calculated as GlyA+ cells/total cells in BM. Human chimerism was defined as human CD45/(human CD45+mCD45). Analysis revealed a high level of human chimerism, with no difference compared to mock electroporated cells. Greater than 90% human chimerism was achieved in both the mock and RNP32 edited groups of mice (**Fig. 43M**). In addition, comparable reconstitution of the B-cell, neutrophil, erythroid, and CD34+ cell lineages was observed in mice infused with RNP32 or Mock electroporated cells (**Fig. 43M**).

[0430] Long term HbF induction by CD235a+ (GlyA+) erythroid cells, derived from RNP32 edited CD34+ cells was also analyzed. GlyA+ cells obtained from bone marrow were stained for gamma globin expression and analyzed via flow cytometry. The edited cells demonstrated pancellular distribution with ~90% of cells being F positive (**Fig. 43N**). In addition, the GlyA+ cell population was isolated by Fluorescence Activated Cell Sorting (FACS), collected and lysed in HPLC grade water. Lysates were then evaluated for HbF expression by UPLC. **Fig. 43O** depicts the HbF expression within the GlyA+ population, which exceeded 50%, calculated as gamma/beta-like chains (%) by GlyA+ cells.

[0431] These data demonstrate that robust long-term HbF induction is achieved with RNP32 editing of human CD34+ cells. In addition, RNP32 editing does not negatively impact the engraftment and lineage reconstitution capacity of CD34+ cells.

Example 16: Infusion of RNP32 edited mPB CD34+ cells into NSG (NOD-SCID Il2rγ^{Null}) mice demonstrates a highly polyclonal and stable engraftment of edited cells over 20 weeks with no clonal outgrowth

[0432] To determine whether indels derived from RNP32 (**Table 15**) maintain a polyclonal profile longitudinally, human adult CD34+ cells from mobilized peripheral blood (mPB) were infused into nonirradiated NSG (NOD-SCID Il2rγ^{Null}) (Jackson lab stock name: NOD.Cg-Prkdc^{scid} Il2rg^{tm1Wjl}/SzJ ("NSG") mice. Briefly, mPB CD34+ cells at 62.5 x 10^6/mL were electroporated via MaxCyte electroporation with 6 μM RNP32 or buffer only ("Mock") following 48 hours pre-stimulation in X-Vivo 10 media supplemented with SCF, TPO and FLT3. After a further 24 hours, mCD34+ cells were cryopreserved. On the day of infusion, mPB CD34+ cells were thawed and infused into busulfan treated NSG mice at 5 million cells per mouse via intravenous tail vein injection. Blood was collected from mice via tail snips at 8, 12, 16, and 20 weeks post infusion and gDNA was isolated. Indel profile analysis was performed on these samples, along with the pre-infusion CD34+ cell product (24 hours post electroporation) to track maintenance of polyclonality. The following primers were used to detect indels: Forward = CATGGCGTCTGGACTAGGAG (SEQ ID NO:1266) and Reverse = AAACACATTTCACAATCCCTGAAC (SEQ ID NO:1267). No notable clonal outgrowth occurred in any of the mice analyzed demonstrating maintenance of a diverse indel profile over a 20 week period (**Fig. 44A**). In addition, at 20 weeks post engraftment, mice were euthanized and bone marrow (BM) was collected from femurs and tibias. Genomic DNA was isolated and indel profile

analysis was performed using the primers described above. As seen in the peripheral blood, a diverse array of indels was observed (**Fig. 44B**). This demonstrates that hematopoiesis was reconstituted by a large number of edited hematopoietic stem cells, indicating that the RNP mediated gene modification did not impair the capacity of HSC within the edited CD34 cell population to engraft and self renew *in vivo*. In addition, no clonal outgrowth was observed suggesting that no cells with uncontrolled proliferation of tumorigenic potential was generated following the modification of the CD34 cells by the RNP. Emergence of a dominant clone would be a potential indication of uncontrolled proliferation or tumorigenesis, which is not seen with RNP32 editing.

Example 17: Use of RNP32 for the treatment of sickle-cell disease

[0433] As described herein, an autologous cell therapy for sickle cell disease (SCD) can be developed that comprises CD34+ cells from patients with SCD that are edited with a AsCas12a (Cpf1) RNP at the *HBG1* and *HBG2* promoters to induce the expression of anti-sickling fetal hemoglobin. This autologous cell therapy is a therapeutic approach to SCD to promote the expression of anti-sickling fetal hemoglobin by directly targeting the promoter of the *HBG1* and *HBG2* genes which encode for the fetal gamma globin chains (**Fig. 45**).

[0434] The efficiency of RNP32 editing for such a cell therapy was determined using CD34+ cells obtained from three independent normal adult donors and four SCD patient donors and compared. Seven batches of mobilized peripheral blood CD34+ cells were used in two independent experiments (**Table**).

Table 19: CD34+ cell batches

Cell Batch ID	Donor Type	Donor Mobilization Regimen
CEL021-014	Normal	G-CSF and plerixafor
CEL042-001	Normal	G-CSF
CEL238-001	Normal	G-CSF and plerixafor
CEL211-001	SCD	Plerixafor
CEL239-001	SCD	Plerixafor
CEL240-001	SCD	Plerixafor
CEL241-001	SCD	Plerixafor
Normal donor type cells were obtained as leukopaks from Hemacare and enriched using the CliniMACS Plus system. SCD CD34+ cells were shipped frozen on dry ice from NIH and University of California, San Francisco. All vials were stored in liquid nitrogen vapor phase until ready for experiments.		

[0435] Briefly, CD34+ cells from normal or SCD donors were pre-stimulated in media consisting of X-Vivo 10, supplemented with 1 X Glutamax, 100 ng/mL stem cell factor (SCF), 100 ng/mL thrombopoietin (TPO), and 100 ng/mL FMS-like tyrosine kinase 3 ligand (Flt3L) for 2 days in a humidified incubator at 37°C, 5% carbon dioxide (CO2). After 2 days of culture, cells were collected and resuspended in MaxCyte electroporation buffer. RNP32 (6 µM, at a gRNA/protein molar ratio of 2) was delivered to CD34+ cells via a MaxCyte GT electroporation device. 1 x 10^6 to 6.25 x 10^6 cells

can be used per OC-100 cartridge for electroporation. However, due to the limited number of cells available from SCD donors, 0.4 x 10^6 to 1 x 10^6 cells from each batch were used for electroporation (**Fig. 46B**). Pre-warmed complete media was then added to the cells to give a final cell density of approximately 1 x 10^6 cells/mL (assuming approximately 10% cell loss during processing). The electroporated cells, along with untreated control cells (cells that did not undergo electroporation), were then placed in a humidified incubator at 37 °C, 5% CO_2. At Days 1, 2, and 3 post-electroporation, an aliquot of cells was counted to determine viability and cells were harvested for further analyses.

[0436] Comparable viabilities were obtained between normal and SCD CD34+ cells treated with RNP32 at Days 1 to 3 post-electroporation (**Fig. 46A**, **Fig. 46B**). The two samples with the lowest viability (one SCD and one normal donor CD34+ cells) coincided with low cell numbers used at electroporation (**Fig. 46B**).

[0437] Electroporated CD34+ cells were resuspended in Quick Extract at a concentration of 2,000 to 4,000 cells/µL. Crude genomic deoxyribonucleic acid (gDNA) extraction was conducted by subjecting the lysate to the following conditions in a thermocycler: 15 min at 65°C followed by 10 min at 95°C. Crude gDNA was then analyzed for indels by next generation sequencing using the following primers: Forward = CATGGCGTCTGGACTAGGAG (SEQ ID NO:1266) and Reverse = AAACACATTTCACAATCCCTGAAC (SEQ ID NO:1267). RNP32 edited normal donor CD34+ cells as efficiently as SCD CD34+ cells (**Fig. 47A**, **Fig. 47C**). The indel levels were variable at Day 1 post electroporation (normal, 40.07% to 84.39%; SCD, 46.78% to 79.50%) but increased to approximately 90% by Day 3 post electroporation in both normal (83.13% to 97.17%) and SCD CD34+ cells (83.04% to 95.78%) (**Fig. 47C**, **Fig. 47B** (Day 3)). The variable levels of editing observed at Day 1 was likely related to the low number of cells used in this experiment, which was limited by the availability of cells from SCD donors as described above.

[0438] RNP32 recognizes both *HBG1* and *HBG2* promoters. Cleavage at both sites can lead to the deletion of the 4.9 kb intervening sequence. To assess the frequency of the 4.9 kb deletion, two ddPCR assays were designed, on-target amplicon and reference amplicon (**Fig. 48A**). The relative positions of the primers and probes are shown in **Fig. 48A** and the sequences are provided in **Fig. 48B**. For each sample, approximately 1.1 ng of gRNA was added to a droplet generator (BioRad) in a PCR plate and droplets were generated following manufacturers instruction. The plate was then moved to the thermocycler and the following protocol was run: 1 cycle of 10 min at 98 °C, 40 cycles of 30 seconds at 94 °C and 2 min at 94 °C, followed by 1 cycle of 10 min at 98 °C and hold at 4 °C. Afterward, the plate was moved to the droplet reader (BioRad) for quantification and counting of droplets that are; a) Negative, b) Positive for both the on-target amplicon and reference amplicon, c) Positive for on-target only, d) Positive for reference only. To determine the percentage of 4.9 kb deletions, the following equation was applied to each sample.

$$\%\ 4.9kb\ deletion = \frac{conc.reference - \frac{conc.on\ target}{Reference\ correction\ factor\ *}}{conc.reference} \times 100$$

[0439] *Reference correction factor was calculated as an average of [(on target concentration)/ (reference concentration)] of the control samples.

[0440] The on-target editing by RNP32 also resulted in the deletion of the 4.9 kb fragment between the *HBG1* and *HBG2* promoters, which occurred at a frequency of approximately 27% of the beta globin loci in both normal (16.4% to 38.5%) and SCD CD34+ cells (20.7% to 32.4%) at Day 1 post-electroporation (**Fig. 49A**, **Fig. 49B**).

[0441] Erythroid progeny of normal and SCD CD34+ cells were also characterized and the ability of RNP32-edited CD34+ cells to differentiate into erythroid cells was determined. Briefly, at Day 1 post-electroporation (described above), 120,000 cells were cultured in erythroid-inducing media to generate erythroid cells using a modified three-step differentiation protocol developed by Giarratana and colleagues (Giarratana, 2005). CD34+ cells were cultured for 7 days in Step-1 media consisting of Iscove's modified Dulbecco's medium (IMDM) supplemented with 1X GlutaMAX (Gibco), 100 U/mL penicillin, 100 mg/mL streptomycin, 5% human AB+ plasma, 330 µg/mL human holo transferrin, 20 mg/mL human insulin, 2 U/mL heparin, 1 µM hydrocortisone, 3 U/mL recombinant human erythropoietin (EPO), 100 ng/mL SCF, and 5 ng/mL interleukin (IL)-3. On Day 7, cells were transferred to Step-2 media, which was identical to Step-1 media except the absence of hydrocortisone and IL-3 and cultured for 4 days. Next, cells were cultured for 7 days in Step-3 media, which was similar to Step-2 media but without the addition of SCF, and 5% human AB+ plasma was replaced with 5% KnockOut Serum Replacement (Gibco). At the end of the 18-day culture, 10 µL of cells were stained with 90 µL of erythroid fluorescence activated cell sorting (FACS) master mix for 15 min at 4°C and acquired on a Guava easyCyte 12HT flow cytometer to determine enucleation frequency. Up to 120 mL cell suspension per sample was spun down at 500 x g for 5 min to reduce the total volume of each sample down to 20 mL. Cell suspension was then passed through Acrodisc WBC syringe filters to remove nucleated cells and the resultant RBCs were used for further analyses in the experiments set forth in this example.

[0442] When placed under erythroid-inducing conditions, normal and SCD CD34+ cells (RNP32-edited cells and unedited cells), underwent robust expansion, averaging 23,000 fold across all experimental groups in 18 days (**Figs. 50A**, **Table 20**). By Day 18, the culture consisted of erythroblasts, RBCs, and extruded nuclei (pyrenocytes) (data not shown). Approximately 80% of the erythroid cells across all groups reached terminal maturation and lost their nucleus (**Fig. 50B**). In

summary, comparable fold expansion and terminal maturation were achieved by normal and SCD CD34+ cells with or without RNP32 electroporation.

Table 20: Phenotype characterization of erythroid progeny of CD34+ cells

Experiment ID	Donor Lot	Donor Type	Treatment	Fold Expansion	Enucleation (%)	HbF (%)	HbF+ cells (%)
SCD011	CEL211-001	SCD	untreated	12465	77.42	16.54	
			RNP32	15118	75.61	48.63	
	CEL238-001	Normal	untreated	25814	83.62	17.97	
			RNP32	16674	81.77	39.24	
SCD014	CEL240-001	SCD	untreated	11686	80.57	25.84	71.27
			RNP32	12856	78.30	58.99	94.50
	CEL241-001	SCD	untreated	16793	83.21	17.26	58.65
			RNP32	16937	84.72	53.70	93.45
	CEL239-001	SCD	untreated	19732	76.43	18.25	50.09
			RNP32	20540	76.79	51.03	93.29
	CEL211-001	SCD	untreated	27390	80.06	27.92	64.51
			RNP32	35845	81.21	58.71	93.79
	CEL238-001	normal	untreated	16214	78.15	16.36	58.23
			RNP32	29239	81.70	51.60	92.92
	CEL021-014	normal	untreated	21498	87.12	11.08	37.76
			RNP32	27659	86.71	38.38	90.61
	CEL042-001	normal	untreated	26615	69.14	10.49	45.00
			RNP32	53774	80.08	43.83	93.85
The enucleation rate was calculated as proportion of NucRed-cells (RBC) within CD235+ (erythroid) population.							

[0443] HbF induction in the erythroid progeny of normal and SCD CD34+ cells post-editing by RNP32 was also characterized and a globin chain analysis was performed. Expression of various hemoglobin subunits was analyzed using reverse phase ultra-performance liquid chromatography (RP-UPLC) modified from a method described by Masala and Manca (Masala 1994)). 1 x 10^6 cells post filtration from erythroid differentiation described above were washed once with phosphate buffered saline (PBS)-0.5% bovine serum albumin (BSA), and then lysed in 50 µL liquid chromatography-mass spectrometry (LC-MS) grade water. Samples were loaded onto the Agilent 1290 UPLC system. Elution was followed at 220 nm with no reference wavelength. The globin chains were eluted in the following order: β, α, AγT (a common Aγ variant, gene product of *HBG1*), Gγ (gene product of *HBG2*), and Aγ (gene product of *HBG1*). Area under the curve under each peak approximated the relative abundance of each globin chain and was used to calculate the level of HbF. As HbA is composed of α2β2 and HbF is composed of α2γ2, the level of HbF expression was calculated as (Aγ + Gγ)/(Aγ + Gγ + β) (%), or (AγT + Aγ + Gγ)/(AγT + Aγ + Gγ + β) (%), labelled as γ/β-like (%).

[0444] The frequency of HbF+ cells (HbF-expressing cells) in RBCs following filtration was determined using a method described by Thorpe (Thorpe 1994). In brief, the RBCs from the erythroid differentiation described above were fixed with 4% formaldehyde, permeabilized with ice-cold acetone, washed with PBS-0.5% BSA, and stained with HbF+ cell FACS master mix. Cells were

washed with PBS-0.5% BSA, resuspended in PBS 0.5% with NucRed (2 drops/mL) and acquired on the Guava flow cytometer.

[0445] Erythroid cells derived from normal and SCD CD34+ cells electroporated with RNP32 demonstrated elevated, and pancellular HbF expression. An average of 43.26% HbF was measured in the treated normal donor group (38.38% to 51.60%), and 54.21% HbF in the treated SCD donor group (48.63% to 58.99%), significantly increased from a background of 13.98% (10.49% to 17.97%) and 21.16% (16.54% to 27.92%) in the untreated normal donor group and SCD donor group, respectively (**Fig. 51A**, **Table 20**). In addition, on average > 92% of RBCs derived from RNP32-electroporated CD34+ cells of either donor type (normal or SCD donors) were HbF+ (normal, 90.61% to 93.85%; SCD, 93.29% to 94.50%) (**Fig. 51B**, **Table 20**), which was significantly elevated compared to RBCs derived from respective untreated CD34+ cells (normal, 37.76% to 58.23%; SCD, 50.09% to 71.27%). Thus, high RNP32 editing levels resulted in elevated HbF levels after *in vitro* erythroid differentiation, with over 50% of HbF measured in RNP32-derived red blood cells (**Fig. 51C**).

[0446] Several assays were conducted to demonstrate the clinical relevance of the elevated fetal hemoglobin in untreated sickle cell patient cells, and to a lesser extent the untreated cells from normal donors, as shown in **Fig. 51B**. Due to the *in vitro* culture used to differentiate the CD34+ cells into RBCs to conduct those assays, there is a high background of HbF in untreated cells derived from SCD patients. This is expected with these culture conditions, and has been previously reported by Métais et al. (Métais 2019). This likely partially masks the full extent of the benefit achieved in modified cells when compared to unmodified cells.

[0447] To evaluate the impact of HbF induction on the sickling of RBCs as a result of hypoxia-induced HbS polymerization, untreated RBCs derived from unedited SCD CD34+ cells (unedited control cells) and RBCs derived from edited SCD CD34+ cells were incubated with oxygen scavenger sodium metabisulfite solution to remove oxygen from the cell suspension, thereby placing the cells under extremely reduced oxygen tension. The percentage of sickled RBCs derived from unedited SCD CD34+ cells and sickled RBCs derived from RNP-32 edited SCD CD34+ cells was then determined and compared.

[0448] Briefly, cells were first induced to generate erythroid cells. One million RNP32-edited RBCs were collected from the erythroid differentiation assay as described above and washed with PBS-0.5% BSA. The cell pellet was then resuspended in 20 µL PBS and mixed with 20 µL of 2% sodium metabisulfite weight/volume in water. One drop (approximately 20 µL) of this cell mix was then placed on a microscopic slide, covered with a coverslip and the edges were sealed. Slides were stored at room temperature for 1 to 4 hours before imaging and analyzing for morphological changes. Average sickling frequency reduced from 38.3% in untreated SCD RBCs (having a mean HbF of 19.9%) to 10.6% in RNP32-edited SCD RBCs (having a mean HbF of 53.8%), representing an approximate 4-fold decrease in sickling morphology for RNP32-edited RBCs from SCD patients

compared with unedited RBCs from SCD patients (**Fig. 52**). Sodium meta-bisulfite leads to extreme hypoxia, below levels commonly observed physiologically, thus even lower percentage of sickling RBC is expected in vivo from the edited cells.

[0449] The deformability of SCD RBCs was also evaluated. To evaluate whether HbF induction decreases the rigidity and improves the deformability of SCD RBCs when deoxygenated, untreated RBCs from patients with SCD and RNP32-edited RBCs from SCD patients were analyzed on a Lorrca ektacytometer where shear stress was applied under decreasing oxygen tension (**Fig. 53A**). Fifteen million RBCs following filtration from erythroid differentiation (as described above) were collected and washed with PBS-0.5% BSA. Cells were centrifuged at 500 x g for 5 minutes, resuspended in 1.5 ml of OXY ISO solution (Lorrca), and mixed gently by inversion. Cell suspension was then loaded onto the Lorrca ektacytometer where Oxygenscan was performed per manufacturer's instructions to measure the deformability of RBCs under shear stress when deoxygenated. RBC deformability was expressed as elongation index (EI) (**Fig. 53A**) and "point of sickling" represented the relative oxygen pressure when the SCD RBCs started to sickle and lost > 5% of the EI during deoxygenation. Normal RBCs containing healthy adult hemoglobin (HbA and HbF) from three different donors were not impacted by the reduction in oxygen tension (**Fig. 53B**). In contrast, SCD RBCs from four donors became rigid when deoxygenated, as illustrated by the gradual decrease in elongation index corresponding with oxygen depletion (**Fig. 53C** (black line)). SCD RBCs derived from RNP32-edited CD34+ cells started to sickle at lower oxygen tension (represented as the point of sickling) compared to RBCs from untreated CD34+ cells (**Fig. 53C** (grey line), **53D, Table 21**). In addition, deoxygenated RNP32-treated SCD RBCs remained more flexible than untreated SCD RBCs as demonstrated by the higher minimum elongation index observed for RNP32-treated SCD RBCs relative to untreated SCD RBCs (**Fig. 53E**, **Table 21**).

Table 21: **Summary of deformability assessment of RBCs**

Experiment ID	Donor Lot	Donor Type	Treatment	Point of Sickling (pO2, mmHg)	Minimum Elongation Index
SCD011	CEL211-001	SCD	untreated	60.17	0.085
			RNP32	35.11	0.260
SCD014	CEL240-001	SCD	untreated	38.55	0.187
			RNP32	21.80	0.340
	CEL241-001	SCD	untreated	31.18	0.292
			RNP32	11.80	0.399
	CEL239-001	SCD	untreated	53.60	0.104
			RNP32	50.09	0.238
	CEL211-001	SCD	untreated	41.62	0.156
			RNP32	21.27	0.330

[0450] Next, to assess whether the reduced sickling and increased flexibility of SCD RBCs derived from RNP32-edited CD34+ cells would lead to improved rheological behavior, RBCs were evaluated

using a microfluidic platform. The microfluidic platform replicated blood flow in the microvasculature for direct observation of the bulk flow of cultured SCD RBCs under varying oxygen conditions. When deoxygenated, the intracellular polymerization of HbS makes RBCs inflexible and rigid which translates to a drop in the velocity through the microchannel. Briefly, one hundred and fifty million RBCs following filtration after erythroid induction as described above were spun down and resuspended in fresh Step-3 media to a final concentration of 10 x 10^6 cells/mL. Cultured RBCs were washed with PBS, and resuspended in PBS to achieve a 20% target hematocrit. A 15 µL volume of cultured RBCs was loaded onto the device under constant pressure and subjected to varying levels of oxygen controlled via a gas mixing system. The flow was captured by a high-speed camera and the blood velocity through the channel was determined using the Kanade-Lucus-Tomasi feature tracking in MATLAB. The velocity of the cells was then measured under a range of oxygen levels. Typical oxygen levels observed in the venous circulation are approximately 4% to 6% oxygen.

[0451] RBCs from normal donors showed no oxygen-dependent rheology impairment (**Fig. 54A**, **Fig. 54B**). While the rheological behavior of RBCs derived from RNP32-edited SCD CD34+ cells did not completely normalize, they exhibited lower magnitudes of impairment during hypoxia and required more extreme hypoxia before the onset of a decrease in velocity compared with RBCs derived from untreated SCD CD34+ cells (10% oxygen for RBCs derived from untreated SCD CD34+ cells and 6% oxygen for RBCs derived from RNP32-edited SCD CD34+ cells). As shown in **Fig. 54C**, RBCs derived from RNP32-edited SCD CD34+ cells (circles) demonstrated a drastically improved rheological behavior compared to RBCs derived from untreated SCD CD34+ cells (triangles) at physiological oxygen levels. Unedited RBCs from normal donors (diamonds) showed no oxygen-dependent rheology impairment. While the rheological behavior of RBCs derived from RNP32-edited SCD CD34+ cells did not completely normalize, they exhibited lower magnitudes of impairment during hypoxia and required more extreme hypoxia for the onset of decrease in velocity compared with RBCs derived from unedited SCD patient CD34+ cells. Notably, at oxygen levels typically observed in the venous circulation, the velocity of red blood cells derived from RNP32-edited CD34 cells was almost fully normalized compared to RBC derived normal donor cells. The reduction in rheology impairment under hypoxia was also strongly correlated with increased levels of HbF in the RBCs (**Fig. 54D**, **Fig. 54E**). For example **Fig. 54E** shows that RBCs that had the highest HbF levels exhibited the highest velocity at a physiological oxygen level of 4%. This indicates that the phenotypic correction of RBCs from sickle cell patients is greatest with very high levels of HbF expression, as observed with RBCs derived from RNP32-edited SCD CD34+ cells (**Fig. 54D**, **Fig. 54E**).

[0452] In sum, two studies were conducted to interrogate editing of SCD and normal CD34+ cells with RNP32 and determine the functional outcome in the erythroid progeny. RNP32 efficiently edited normal and SCD CD34+ cells, achieving approximately 90% insertions and/or deletions

(indels) at Day 3 post-electroporation. Editing of CD34+ cells with RNP32 did not negatively impact the erythroid differentiation or maturation. Robust HbF expression was obtained, averaging 43.26% and 54.21% of total hemoglobin in RBCs derived from RNP32-edited normal and SCD CD34+ cells respectively, distributed in a pancellular fashion (averaging > 92% RBCs). The high level of HbF expression in RBCs derived from RNP32-edited SCD CD34+ cells coincided with decreased sickling, improved deformability under shear stress, and improved flow through microfluidic channels when deoxygenated compared to RBCs derived from untreated SCD CD34+ cells. Potentially therapeutically relevant levels of HbF were achieved through highly efficient RNP32 editing of CD34+ cells at the *HBG1* and *HBG2* promoter region. This translates to a reduction in sickling and improved rheological properties in red blood cells, which is advantageous for an autologous cell therapy using RNP32 for the treatment of sickle-cell disease.

Example 18: Meta-analysis of on-target indels generated by RNP32 in CD34+ cells

[0453] Following CRISPR/Cas editing, DNA repair mechanisms like non-homologous end joining (NHEJ) and microhomology-mediated end joining (MMEJ) result in highly heterogeneous repair outcomes comprising hundreds of different genotypes. The editing outcomes produced when using Cpf1 (also referred to as AsCas12a), the enzyme used for RNP32, have not been characterized in detail. Cpf1 editing results in a four nucleotide 5' overhang (see **Fig. 55A**), which is quite different from the blunt ends of SpCas9 edits (see **Fig. 41A**). As used in this example, "cut site" is used to refer to the position at the middle of the expected overhang resulting from editing with RNP32 (**Fig. 55A**, grey dashed line).

[0454] DNA repair mechanisms often create indels ranging in size from 1 to approximately 50 base pairs (bp), which are typically detected by PCR-based targeted sequencing assays. More complex genomic repair outcomes have also been described (**Error! Reference source not found.** 2018; **Error! Reference source not found.** 2018; **Error! Reference source not found.** 2020), and include deletions at the on-target locus larger than ~50 bp, called resections, and translocations. These more complex rearrangements require other methods for a precise quantification (see, e.g., PCT/US2018/012652).

[0455] To determine the indel profile generated by RNP32, indels ranging in size from 1 bp to approximately 50 bp at the distal CCAAT-box generated by RNP32 were characterized using PCR-based targeted sequencing assays. The indel patterns were studied in a variety of samples, spanning different genotypes (normal vs sickle cell disease) and mobilization regimens (plerixafor alone vs G-CSF alone vs G-CSF + plerixafor) (**Table 12**). This meta-analysis demonstrates that RNP editing at the CCAAT box results in a distinct indel profile that is reproducible across a variety of samples and levels of editing.

Table 22: Samples used in meta-analysis

Exp. # / Batch #	Donor Type	CEL ID (Donor)	Donor Mobilization Regimen	RNP32 [μM]
SCD011	Normal	CEL238-001	G-CSF & plerixafor	6
SCD011	SCD	CEL211-001	plerixafor	6
SCD014	Normal	CEL021-014	G-CSF & plerixafor	6
SCD014	Normal	CEL042-001	G-CSF	6
SCD014	Normal	CEL238-001	G-CSF & plerixafor	6
SCD014	SCD	CEL211-001	plerixafor	6
SCD014	SCD	CEL239-001	plerixafor	6
SCD014	SCD	CEL240-001	plerixafor	6
SCD014	SCD	CEL241-001	plerixafor	6
SCD1	Normal	CEL172-008	G-CSF & plerixafor	8
SCD1	Normal	CEL182-008	G-CSF & plerixafor	8
SCD1	Normal	CEL171-008	G-CSF & plerixafor	8
SCD1	Normal	CEL192-008	G-CSF & plerixafor	8
SCD1	Normal	CEL168-008	G-CSF & plerixafor	8
Samples were obtained from research-scale (SCD011, SCD014) or large-scale (SCD1) processes. No significant differences were observed in RNP32-edited cells produced using research-scale versus large-scale processes (data not shown). The donors from experiments SCD011 and SCD014 are the same as those in **Example 17** (**Table 20**).				

[0456] Mobilized peripheral blood CD34+ cells from 12 distinct donors (normal donors and SCD donors) were used in this meta-analysis (**Table 22**). Five samples from normal donors were generated using a large-scale process (**Table 22**, Experiment SCD1). Briefly, leukopaks (HemaCare or KeyBiologics) were obtained from normal donors mobilized with granulocyte colony stimulating factor (G-CSF) and plerixafor. CD34+ cells were enriched using the CliniMACS Plus system, aliquoted, cryopreserved in Cryostor CS10, and stored in liquid nitrogen vapor phase. CD34+ cells were thawed, cultured for 2 days in complete media consisting of X-Vivo 10, supplemented with 1 X Glutamax, 100 ng/mL stem cell factor (SCF), 100 ng/mL thrombopoietin (TPO), and 100 ng/mL FMS-like tyrosine kinase 3 ligand (Flt3L), mixed with RNP32 (at a gRNA/protein molar ratio of 2) to a final concentration of 8 μM, and electroporated with a Maxcyte GT electroporation device per manufacturer's instruction. The cellular materials were cultured overnight following electroporation,

aliquoted, and cryopreserved in Cryostor CS10, and stored in liquid nitrogen vapor phase until ready for experimentation.

[0457] Nine samples were generated using a research-scale process (**Table 22**, SCD011 and SCD014). Cells were thawed, washed, and cultured at 1 x 10^6 cells/mL in complete media consisting of X-Vivo 10, supplemented with 1 X Glutamax, 100 ng/mL stem cell factor (SCF), 100 ng/mL thrombopoietin (TPO), and 100 ng/mL FMS-like tyrosine kinase 3 ligand (Flt3L) for 2 days in a humidified incubator at 37°C, 5% carbon dioxide (CO2). After 2 days of culture, cells were collected and resuspended in Maxcyte electroporation buffer. RNP32 (at a gRNA/protein molar ratio of 2) was added to the cell suspension to a final concentration of 6 μM. The mixture was transferred to a Maxcyte OC-100 cartridge and electroporated with a Maxcyte GT electroporation device per manufacturer's instruction. After electroporation, cells were cultured in complete media for 1 day prior to harvesting for analysis.

[0458] The fraction of small indels present in the samples was assessed by Illumina amplicon sequencing (ILL-seq), following library preparation method and analysis. A 15 bp window around the expected cut site was used in the analysis to calculate editing rates. The oligonucleotides and amplicon used to generate the targeted amplicon sequencing products are provided in **Fig. 55A** and **Fig. 55B**.

[0459] To characterize editing by RNP32, the indels generated in each sample were analyzed and processed. The results were summarized by looking at 1) percentage deletion for each base within the target region, 2) distribution of insertion and deletion center positions, and 3) distribution of insertion and deletion lengths. To perform these analyses, the cigar strings for the reads in the alignment bam files were processed, using the following steps:

5) Group indels appearing in the same read.
6) Create indel_id for indels from each read. Indel_id is a string identifying the indel, shown as indel_start_position + _ + indel_length + _ + ID. Where ID is NA for deletions and for insertions is the sequence inserted. See example in **Fig. 55D**.
7) Group reads with the same indel_id.
8) Count number of reads with the same indels and calculate their total fractions (including wt) and fractions in indels (excluding wt). Fractions in indels allow for a comparison between samples when they have different amounts of total editing.

[0460] To analyze the reproducibility of the individual indels detected across samples, the number of times they were detected across samples and their average percentage in indels was estimated. When an indel was not detected in a particular sample it was assumed to have a percentage of 0%.

[0461] The total percentage of indels for each RNP32-edited sample in this analysis are shown in **Table 23**. All the samples (including normal and SCD donors) exhibited overall editing greater than 72.6% (ranging from 72.6% to 89.2%), except for three samples that exhibited lower editing between

about 40% to 53%. The samples exhibiting lower editing had low viability potentially associated with lower numbers of cells used for electroporation.

Table 23: Editing results for all the samples

Donor Type	CEL ID	EXP #	Total Indels %
Normal	CEL238-001	SCD011	84.39%
SCD	CEL211-001	SCD011	79.50%
Normal	CEL021-014	SCD014	40.07%
Normal	CEL042-001	SCD014	78.37%
Normal	CEL238-001	SCD014	80.70%
SCD	CEL211-001	SCD014	46.78%
SCD	CEL239-001	SCD014	52.92%
SCD	CEL240-001	SCD014	75.38%
SCD	CEL241-001	SCD014	72.63%
Normal	CEL172-008	SCD1	88.91%
Normal	CEL182-008	SCD1	87.11%
Normal	CEL171-008	SCD1	84.04%
Normal	CEL192-008	SCD1	89.24%
Normal	CEL168-008	SCD1	88.76%

[0462] In total, 385 unique indels were detected, counting only those indels that were found at a percentage above 0.1% (out of total indels) in at least one of the samples in **Table 25**.

[0463] The wt allele and the top 20 indels are shown in **Fig. 57**, alongside their relative average frequency (amongst all samples). The most abundant indel, 159_-18_NA, was present on average 15.15% among all the samples. Notably, indel 157_-13_NA, more commonly known as the 13bp HPFH deletion, was detected on average at 2.63%. The 13bp deletion is a naturally occurring mutation associated with elevated HbF.

[0464] Indel analysis of the bulk CD34+ population (progenitors) 24-72 hours after electroporation can determine whether editing at this site was derived from Cas12a, or Cas9. In the case of Cas12a, the most dominant indel detected at this timepoint is the 159_-18 deletion, however with Cas9 the most frequent indel is the 157_-13 deletion. In addition, the most common NHEJ indel generated by Cas9 at this site is a -1 bp deletion (169_-1_NA, **Table 24**), which occurred here at ~9.6%. In contrast, following editing with Cas12a, the most common NHEJ indels generated are 6 bp and 4 bp deletions (165_-6_NA, and 167_-4_NA; **Table 24**, RNP58; **Table 25**, RNP32). Cpf1 generally produces larger NHEJ deletions, when compared to Cas9 at this site.

[0465] Indel analysis of the bulk CD34+ population (progenitors) 24-72 hours after electroporation can determine whether edits (indels) at this site were derived from Cpf1 or Cas9. For example, in the

case of Cpf1, the most dominant indel detected at this timepoint is the 159_-18 deletion at an average percentage of 15.15% of the total indels (**Table 25**) compared to an average percentage of 1.35% for SpCas9, see **Table 24**. However, the most frequent indel for SpCas9 is the 157_-13 deletion at an average percentage of 31.88% of the total indels (**Table 24**) versus an average percentage of 2.63% of the total indels for Cpf1 (**Table 25)**.

[0466] Results for the distributions of indel center points and indel lengths are shown in **Figs. 58** and **59**, respectively. The most common indel center position (peak) was located 6 bp away from the cut site, at position TSS:–113. Peaks in the deletion length distribution were observed at 13 bp and 18 bp, corresponding mostly to the 157_-13_NA and 159_-18_NA MMEJ indels. Due to the resection and utilization of microhomology sequences during MMEJ repair, this pathway tends to generate larger size deletions than those mediated via the NHEJ pathway. Besides those, the most commonly observed deletion length was 5 bp with 9.30% (**Fig. 60)**. Insertions were rarely detected (total contribution of 0.50%). Deletions between 1 bp and 25 bp had a total contribution of 92.03% among all indels (deletions between 3 bp and 25 bp had a total contribution of 85.85% among all indels, deletions between 4 bp and 25 bp had a total contribution of 80.35% among all indels, and deletions between 5 bp and 25 bp had a total contribution of 72.04% among all indels) (**Table 25**). Overall, both the distribution of indel center points and indel lengths were very similar across the Normal and SCD samples. As shown in **Fig 41C** above, the generation of indels 4 bp and longer results in the highest induction of HbF.

[0467] The results for the percentage of bases deleted along the target region (deletion profile) are shown in **Fig. 61** and **Fig. 62** (normalized to the same maximum value of 1 for comparison between samples). The shapes of the profiles are very similar across all samples, despite having different total editing values (**Table 23**). No major differences were observed between Normal donor (Normal) and sickle cell disease donor (SCD) samples or different mobilization regimens. The deletion profiles show that the peak of the distribution (most commonly deleted base) is 6 bp toward the TSS from the cut site, at TSS: –113.

[0468] Of the total of 385 indels detected at 0.1% in any of the samples evaluated, 108 indels were detected in all 14 samples (**Fig. 63** and **Table 25**). All indels present at an average percentage in indel greater than 0.22% were detected in all 14 samples (a total of 55 indels, above the grey line **Fig. 64** and also in **Table 25**), and had a total indel contribution of 69.90%. Overall, the reproducibility of indel detection was very high.

[0469] To look in more detail at the consistency in each individual indel frequency between the indels detected across all samples, their pairwise correlation plots and R^2 were calculated. To avoid under sampling bias, when comparing two samples, only indels which had a minimum percentage of 0.1% (of all indels) in at least one of the two samples was considered. Excluding the three samples with editing below 60% due to the low number of cells used for electroporation (see **Table 22**), all R^2

values were greater than 0.8 (data not shown). Overall, the correlation between indel percentages among the samples with high editing was very high.

[0470] In sum, a total of 385 unique indels were detected at a percentage above 0.1% in at least one of the 14 edited samples tested in this Example. The most abundant indel, 159_-18_NA, was present between all the indels on average 15.15% among all the samples. Indel 157_-13_NA, more commonly known as the 13bp HPFH deletion, was detected on average at 2.63%. The shapes of the indel profiles for all edited samples were very similar across all studies, despite having different total editing values. No major differences were observed between Normal donor and SCD donor samples or different mobilization regimens.

[0471] Evaluation of individual indel positions showed that indels generated by RNP32 were mostly centered around position TSS: –113, which was also the most commonly deleted base observed. Evaluation of the distribution of indel lengths showed peaks at –18 and –13 corresponding primarily to the MMEJ indels, 159_-18_NA (most abundant indel, starting at TSS: –104 and of length 18 bases) and 157_-13_NA (starting at TSS: –102, 13bp HPFH deletion). Besides these, the most commonly observed deletion length was 5 bp at 9.30%. Insertions were rarely detected (total contribution of 0.50%). Deletions between 1 bp and 25 bp had a total contribution of 92.03% among all indels. Overall, the distribution of indel length and center position were very similar between the Normal and SCD samples. Of the total of 385 unique indels, 108 indels were detected in all 14 samples. All indels present at an average percentage in indel greater than 0.22% were detected in all 14 samples (a total of 55 indels), and had a total indel contribution of 69.90%. Pairwise correlations among the detected indels (excluding the three samples with editing below 60%) had R^2 values greater than 0.8.

[0472] This data demonstrate that RNP32 editing at the *HBG1/2* distal CCAAT box results in a unique indel signature that is reproducible across a variety of samples and levels of editing.

Example 19: Analysis of normal donor CD34+ cell RNP32 editing efficiency including 4.9 kb fragment deletion

[0473] The on-target editing efficiency of RNP32 in mobilized peripheral blood CD34+ cells obtained from four independent normal adult donors was assessed, as well as the frequency of deletion of the 4.9 kb fragment between the two RNP32 cut sites. In addition, the on-target indel levels and frequency of the 4.9 kb deletion were also measured in several subpopulations of HSPC sorted from total CD34+ cells to address whether phenotypic long-term hematopoietic stem cells (LT-HSC) could be edited efficiently with RNP32.

[0474] Leukopaks from four normal donors treated with granulocyte colony-stimulating factor (G-CSF) plus Mozobil or with G-CSF alone were obtained from HemaCare. CD34+ cells (Lot: CEL045-002, CEL046-004, CEL047-002, and CEL021-021) were enriched using the CliniMACS system (Miltenyi), aliquoted, and cryopreserved in Cryostor CS10. Cells were thawed, washed, and cultured at 1 x 10^6 cells/mL in complete media consisting of X-Vivo 10, supplemented with 1 X Glutamax,

100 ng/mL stem cell factor (SCF), 100 ng/mL thrombopoietin (TPO), and 100 ng/mL FMS-like tyrosine kinase 3 ligand (Flt3L) for 2 days in a humidified incubator at 37°C, 5% carbon dioxide (CO2). After 2 days of cultures, cells were collected and resuspended in Maxcyte electroporation buffer. RNP32 (gRNA/protein molar ratio of 2) was added to the cell suspension to a final concentration of 6 μM. The mixture was transferred to a Maxcyte OC-100 cartridge and electroporated with a Maxcyte GT electroporation device per manufacturer's instruction.

[0475] RNP32 recognizes both *HBG1* and *HBG2* promoters. Cleavage at both sites can lead to the deletion of the 4.9 kb intervening sequence. To assess the frequency of the 4.9 kb fragment deletion, the two ddPCR assays, reference amplicon and on-target amplicon, were performed as described in **Example 17** (see **Figs. 48A-48D**). Briefly, cells at 1 or 2 days post-electroporation with RNP32, and sorted cells were resuspended in Quick Extract at a concentration of 2,000 to 4,000 cells/μL. Crude genomic deoxyribonucleic acid (gDNA) extraction was conducted by subjecting the lysate to the following conditions in a thermocycler: 15 min at 65°C followed by 10 min at 95°C. Crude gDNA was then analyzed for indels (insertions and deletions) by next generation sequencing (using the primers set forth in **Fig. 55A**) and 4.9kb fragment deletion assessment by digital droplet polymerase chain reaction (ddPCR).

[0476] RNP32 edited the *HBG1* and *HBG2* promoters of normal donor CD34+ cells in an RNP concentration-, and time-dependent manner without significantly impacting the cell viability (**Fig. 65A**, **65B**, **Fig. 66**, and **Fig. 67**). At 8 μM, the highest concentration of RNP32 tested, an average of 86% on-target indel level was achieved in normal donor CD34+ cells by Day 1 and rose to 91% by Day 2 post-electroporation (**Fig. 68**). The increase was more evident at lower RNP concentration with average indel level at 1 μM rising from 62% to 77% by Day 1 and Day 2 post-electroporation. By Day 2 post-electroporation, ≥ 85% on-target indel level was routinely achieved when CD34+ cells were transfected with ≥ 3 μM RNP32.

[0477] Editing of CD34+ cells with RNP32 resulted in frequent deletion of the 4.9 kb intervening fragment between the two RNP32 cut sites at the *HBG1* and *HBG2* gene promoters, respectively. On average, approximately 35% of the beta globin loci in the CD34+ cells had the 4.9 kb fragment deleted both at Day 1 and Day 2 post electroporation with ≥ 2 μM of RNP32 (**Fig. 69A**, **Fig. 69B** and **Fig. 70**). The 4.9 kb fragment deletion to indel ratio in CD34+ cells was approximately 0.47 at Day 1 and decreased to approximately 0.40 at Day 2 post-electroporation with ≥ 2 μM of RNP32.

[0478] CD34+ cells are heterogeneous and comprise lineage-restricted progenitors, MPPs, as well as self-renewing LT-HSCs. As LT-HSCs will be responsible for providing long-term reconstitution of a patient's hematopoietic system, high levels of editing in this population are pertinent for the durability of the treatment. To address whether different subpopulations of CD34+ cells were edited similarly, the on-target indel levels in CMPs, MPPs, and phenotypic LT-HSCs, defined by surface

immunophenotype, was evaluated and compared to the on-target indel levels in total CD34+ cells across multiple RNP concentrations.

[0479] Briefly, CD34+ cells were sorted two days post electroporation with RNP32 to obtain three subpopulations of HSPCs including phenotypic LT-HSC, multipotent progenitor (MPP) cells, and common myeloid progenitor (CMP) cells using a BD fluorescence-activated cell sorting (FACS)Aria Fusion cell sorter (BD Biosciences). Gating was set to collect the following enriched CD34+ cell subpopulations: phenotypic LT-HSC (P7, CD34 bright, CD38 low/negative, CD90+, CD45RA-), MPP (P6, CD34 bright, CD38 low/negative, CD90-, CD45RA-), and CMP (P9, CD34 bright, CD38 high, CD123+, CD45RA-).

[0480] Of the three subpopulations, CMPs consistently had the highest on-target indel levels and phenotypic LT-HSCs had the lowest on-target indel levels (analyzed via NGS using the primers set forth in **Fig. 55B**) across all RNP concentrations evaluated (**Fig. 71A**, **Fig. 71B** and **Fig. 72**Error! Reference source not found.) (Study MAX072). Nevertheless, at ≥ 3 μM RNP32, on-target indel levels of ≥ 85% were achieved in phenotypic LT-HSCs, similar to the on-target indel levels detected in total CD34+ cells, suggesting that self-renewing LT-HSCs can be efficiently edited using RNP32 (**Fig. 71A**, **Fig. 71B** and **Fig. 72**).

[0481] RNP32 editing resulted in deletion of the 4.9 kb fragment in all three subpopulations of hematopoietic stem and progenitor cells tested (**Fig. 73A** and **Fig. 74**). Of the three subpopulations, CMPs consistently had the highest levels of deletion of the 4.9 kb fragment and phenotypic LT-HSCs had the lowest levels of deletion of the 4.9 kb fragment across all RNP concentrations evaluated. The 4.9 kb fragment to indel ratio averaged across all RNP concentrations was approximately 0.43, 0.49, 0.33, and 0.25 in total CD34+ cells, CMP, MPP, and phenotypic LT-HSC respectively (**Fig. 73B**), demonstrating that the LT-HSCs had less frequent deletion of the 4.9 kb fragment as a result of RNP32 editing than their short term counterparts. Retention of the 4.9 kb inter-edit region is beneficial because a higher deletion of the 4.9 kb fragment results in a lower production of HbF. Therefore, lower levels of deletion of the 4.9 kb fragment in LT-HSCs edited by RNP32 is advantageous. The results herein demonstrate that editing with RNP32 can generate cells with clinically relevant levels of healthy HbF.

[0482] In sum, on-target indel levels and the associated 4.9 kb fragment deletion between the *HBG1* and *HBG2* promotors were assessed following electroporation of normal human donor CD34+ cells with RNP32. RNP32 was highly efficient at editing CD34+ cells, with consistent editing achieved across multiple RNP batches and cell donor lots. Greater than 85% on target indels in CD34+ cells were routinely achieved when electroporated with ≥ 3 μM RNP32. Comparable levels of indels between total CD34+ cells and sorted phenotypic LT HSC were also observed when electroporated with ≥ 3 μM RNP32. Loss of the 4.9 kb fragment between two RNP32 cut sites occurred at a frequency that was cell subpopulation dependent. Total CD34+ cells lost the 4.9 kb fragment at a

frequency of approximately 0.4 per indel. This frequency dropped to approximately 0.25 per indel in the phenotypic LT-HSC subset of CD34+ cells.

Example 20: Treatment of β-hemoglobinopathy using edited hematopoietic stem cells

[0483] The methods and genome editing systems disclosed herein may be used for the treatment of a β-hemoglobinopathy, such as sickle cell disease or beta-thalassemia, in a patient in need thereof. For example, genome editing may be performed on cells derived from the patient in an autologous procedure. Correction of the patient's cells *ex-vivo* and reintroduction of the cells into the patient may result in increased HbF expression and treatment of the β-hemoglobinopathy.

[0484] For example, HSCs may be extracted from the bone marrow of a patient with a β-hemoglobinopathy using techniques that are well-known to skilled artisans. The HSCs may be modified using methods disclosed herein for genome editing. For example, RNPs comprised of guide RNAs (gRNA) that target one or more regions in the *HBG* gene complexed with an RNA-guided nuclease may be used to edit the HSCs. In certain embodiments, the RNA-guided nuclease may be a Cpf1 protein. In certain embodiments, the Cpf1 protein may be a modified Cpf1 protein. In certain embodiments, the modified Cpf1 protein may be encoded by a sequence set forth in SEQ ID NOs:1000, 1001, 1008-1018, 1032, 1035-39, 1094-1097, 1107-09 (Cpf1 polypeptide sequences) or SEQ ID NOs:1019-1021, 1110-17 (Cpf1 polynucleotide sequences). For example, the modified Cpf1 protein may be encoded by the sequence set forth in SEQ ID NO:1097. In certain embodiments, the gRNA may be a modified or unmodified gRNA. In certain embodiments, the gRNA may comprise a sequence set forth in **Table 6**, **Table 12**, or **Table 13**. For example, in certain embodiments, the gRNA may comprise the sequence set forth in SEQ ID NO:1051. In certain embodiments, the RNP complex may comprise an RNP complex set forth in **Table 15**. For example, the RNP complex may include a gRNA comprising the sequence set forth in SEQ ID NO:1051 and a modified Cpf1 protein encoded by the sequence set forth in SEQ ID NO:1097 (RNP32, **Table 15**). In certain embodiments, modified HSCs have an increase in the frequency or level of an indel in the human *HBG1* gene, *HBG2* gene, or both, relative to unmodified HSCs. In certain embodiments, the modified HSCs can differentiate into erythroid cells that express an increased level of HbF. A population of the modified HSCs may be selected for reintroduction into the patient via transfusion or other methods known to skilled artisans. The population of modified HSCs for reintroduction may be selected based on, for example, increased HbF expression of the erythroid progeny of the modified HSCs or increased indel frequency of the modified HSCs. In some embodiments, any form of ablation prior to reintroduction of the cells may be used to enhance engraftment of the modified HSCs. In other embodiments, peripheral blood stem cells (PBSCs) can be extracted from a patient with a β-hemoglobinopathy using techniques that are well-known to skilled artisans (e.g., apheresis or leukapheresis) and stem cells can be removed from the PBSCs. The genome editing methods described above can be performed on the stem cells and the modified stem cells can be reintroduced into the patient as described above.

Table 14: Cpf1 Protein Variants

Cpf1 Variant ID	Cpf1 Variant Amino Acid SEQ ID NO*	Cpf1 Variant Nucleotide SEQ ID NO*
His-AsCpf1-sNLS-sNLS H800A	SEQ ID NO:1032	SEQ ID NO:1110
Cpf1-1	SEQ ID NO:1094	SEQ ID NO:1111
Cpf1-2	SEQ ID NO:1095	SEQ ID NO:1112
Cpf1-3	SEQ ID NO:1096	SEQ ID NO:1113
Cpf1-4	SEQ ID NO:1097	SEQ ID NO:1114
Cpf1-5	SEQ ID NO:1107	SEQ ID NO:1115
Cpf1-6	SEQ ID NO:1108	SEQ ID NO:1116
Cpf1-7	SEQ ID NO:1109	SEQ ID NO:1117
*See **Fig. 40**		

Table 15: RNP complexes

RNP Name	gRNA SEQ ID NO*	Cpf1 Amino Acid SEQ ID NO**
RNP64	SEQ ID NO: 1022	SEQ ID NO: 1109
RNP1	SEQ ID NO: 1051	SEQ ID NO: 1095
RNP2	SEQ ID NO: 1098	SEQ ID NO: 1094
RNP3	SEQ ID NO: 1099	SEQ ID NO: 1094
RNP4	SEQ ID NO: 1100	SEQ ID NO: 1094
RNP5	SEQ ID NO: 1101	SEQ ID NO: 1094
RNP6	SEQ ID NO: 1068	SEQ ID NO: 1094
RNP7	SEQ ID NO: 1102	SEQ ID NO: 1094
RNP8	SEQ ID NO: 1081	SEQ ID NO: 1094
RNP9	SEQ ID NO: 1082	SEQ ID NO: 1094
RNP10	SEQ ID NO: 1083	SEQ ID NO: 1094
RNP11	SEQ ID NO: 1069	SEQ ID NO: 1094
RNP12	SEQ ID NO: 1084	SEQ ID NO: 1094
RNP16	SEQ ID NO: 1085	SEQ ID NO: 1094
RNP19	SEQ ID NO: 1088	SEQ ID NO: 1094
RNP20	SEQ ID NO: 1089	SEQ ID NO: 1094
RNP21	SEQ ID NO: 1090	SEQ ID NO: 1094
RNP22	SEQ ID NO: 1091	SEQ ID NO: 1094
RNP23	SEQ ID NO: 1048	SEQ ID NO: 1097
RNP24	SEQ ID NO: 1093	SEQ ID NO: 1094
RNP26	SEQ ID NO: 1051	SEQ ID NO: 1096
RNP27	SEQ ID NO: 1051	SEQ ID NO: 1107
RNP28	SEQ ID NO: 1051	SEQ ID NO: 1108
RNP29	SEQ ID NO: 1103	SEQ ID NO: 1094
RNP30	SEQ ID NO: 1104	SEQ ID NO: 1094
RNP31	SEQ ID NO: 1105	SEQ ID NO: 1094
RNP32	SEQ ID NO: 1051	SEQ ID NO: 1097
RNP33	SEQ ID NO: 1022	SEQ ID NO: 1032
RNP34	SEQ ID NO: 1041	SEQ ID NO: 1032
RNP37	SEQ ID NO: 1042	SEQ ID NO: 1032

RNP38	SEQ ID NO: 1043	SEQ ID NO: 1032
RNP39	SEQ ID NO: 1044	SEQ ID NO: 1032
RNP40	SEQ ID NO: 1045	SEQ ID NO: 1032
RNP41	SEQ ID NO: 1046	SEQ ID NO: 1032
RNP42	SEQ ID NO: 1047	SEQ ID NO: 1032
RNP43	SEQ ID NO: 1048	SEQ ID NO: 1032
RNP44	SEQ ID NO: 1049	SEQ ID NO: 1032
RNP45	SEQ ID NO: 1022	SEQ ID NO: 1094
RNP46	SEQ ID NO: 1046	SEQ ID NO: 1094
RNP47	SEQ ID NO: 1047	SEQ ID NO: 1094
RNP48	SEQ ID NO: 1052	SEQ ID NO: 1094
RNP49	SEQ ID NO: 1048	SEQ ID NO: 1094
RNP50	SEQ ID NO: 1053	SEQ ID NO: 1094
RNP51	SEQ ID NO: 1054	SEQ ID NO: 1094
RNP52	SEQ ID NO: 1055	SEQ ID NO: 1094
RNP53	SEQ ID NO: 1056	SEQ ID NO: 1094
RNP54	SEQ ID NO: 1057	SEQ ID NO: 1094
RNP55	SEQ ID NO: 1058	SEQ ID NO: 1094
RNP56	SEQ ID NO: 1059	SEQ ID NO: 1094
RNP57	SEQ ID NO: 1060	SEQ ID NO: 1094
RNP58	SEQ ID NO: 1051	SEQ ID NO: 1094
RNP59	SEQ ID NO: 1067	SEQ ID NO: 1094
RNP60	SEQ ID NO: 1068	SEQ ID NO: 1094
RNP61	SEQ ID NO: 1041	SEQ ID NO: 1095
RNP62	SEQ ID NO: 1041	SEQ ID NO: 1094
RNP63	SEQ ID NO: 1022	SEQ ID NO: 1095
*See **Table 13** **See **Table 14**		

Table 16: Cpf1 HBG1 targeting domains and expected cleavage sites

gRNA ID	gRNA Targeting Domain (RNA)	gRNA Targeting Domain (DNA)	Targeting Domain coordinates at HBG1 *	Expected cleavage site coordinates at HBG1#	Strand	Targeting Domain Length
HBG1-1	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)	CCTTGTCAAGGCTATTGGTC (SEQ ID NO:1003)	Chr11:5249954–5249974	Chr11:5249973, Chr11:5249977	+	20
HBG1-1 (21mer)	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)	CCTTGTCAAGGCTATTGGTCA (SEQ ID NO:1255)	Chr11:5249954–Chr11:5249975	Chr11:5249973, Chr11:5249977	+	21
HBG1-1 (22mer)	CCUUGUCAAGGCUAUUGGUCAA (SEQ ID NO:1256)	CCTTGTCAAGGCTATTGGTCAA (SEQ ID NO:1257)	Chr11:5249954–Chr11:5249976	Chr11:5249973, Chr11:5249977	+	22
HBG1-1 (23mer)	CCUUGUCAAGGCUAUUGGUCAAG (SEQ ID NO:1258)	CCTTGTCAAGGCTATTGGTCAAG (SEQ ID NO:1259)	Chr11:5249954–Chr11:5249976	Chr11:5249973, Chr11:5249977	+	23
AsCpf1 HBG1 Promoter-1	AGACAGAUAUAUUUGCAUUGAG (SEQ ID NO:1139)	AGACAGATATTTGCATTGAG (SEQ ID NO:1140)	Chr11:5250023-Chr11:5250043	Chr11:5250042, Chr11:5250046	+	20
AsCpf1 HBG1 Promoter-1 (21mer)	AGACAGAUAUAUUUGCAUUGAGA (SEQ ID NO:1260)	AGACAGATATTTGCATTGAGA (SEQ ID NO:1261)	Chr11:5250023-Chr11:5250044	Chr11:5250042, Chr11:5250046	+	21
AsCpf1 HBG1 Promoter-2	CAUUGAGAUAGUGUGGGGAA (SEQ ID NO:1141)	CATTGAGATAGTGTGGGGAA (SEQ ID NO:1142)	Chr11:5250036-Chr11:5250056	Chr11:5250055, Chr11:5250059	+	20
AsCpf1 HBG1 Promoter-2 (21mer)	CAUUGAGAUAGUGUGGGGAAG (SEQ	CATTGAGATAGTGTGGGGAAG (SEQ ID NO:1263)	Chr11:5250036-Chr11:5250057	Chr11:5250055, Chr11:5250059	+	21

	ID NO:1262)					
AsCpf1 HBG1 Promoter-6	CUUCUC CCAUCA UAGAGG AU (SEQ ID NO:1149)	CTTCTCCCA TCATAGAG GAT (SEQ ID NO:1150)	Chr11:5250160-Chr11:5250180	Chr11:5250179, Chr11:5250183	+	20
AsCpf1 HBG1 Promoter-6 (21mer)	CUUCUC CCAUCA UAGAGG AUA (SEQ ID NO:1264)	CTTCTCCCA TCATAGAG GATA (SEQ ID NO:1265)	Chr11:5250160-Chr11:5250181	Chr11:5250179, Chr11:5250183	+	21
*NCBI Reference Sequence NC_000011, the coordinates are reported using the One-based coordinate system, "Homo sapiens chromosome 11, GRCh38.p12 Primary Assembly," (Version NC_000011.10). # Expected cleavage sites based on Zetsche et al, 2015, coordinates are reported using zero-based coordinates.						

Table 17: Cpf1 HBG2 targeting domains and expected cleavage sites

gRNA ID	gRNA Targeting Domain (RNA)	gRNA Targeting Domain (DNA)	Targeting Domain coordinates at HBG2 *	Expected cleavage site coordinates at HBG2#	Strand	Targeting Domain Length
HBG1-1	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)	CCTTGTCAAGGCTATTGGTC (SEQ ID NO:1003)	Chr11:5254878 Chr11:5254898	Chr11:5254897, Chr11:5254901	+	20
HBG1-1 21mer	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)	CCTTGTCAAGGCTATTGGTCA (SEQ ID NO:1255)	Chr11:5254878 Chr11:5254899	Chr11:5254897, Chr11:5254901	+	21
HBG1-1 22mer	CCUUGUCAAGGCUAUUGGUCAA (SEQ ID NO:1256)	CCTTGTCAAGGCTATTGGTCAA (SEQ ID NO:1257)	Chr11:5254878 Chr11:5254900	Chr11:5254897, Chr11:5254901	+	22
HBG1-1 23mer	CCUUGUCAAGGCUAUUGGUCAAG (SEQ ID NO:1258)	CCTTGTCAAGGCTATTGGTCAAG (SEQ ID NO:1259)	Chr11:5254878 Chr11:5254901	Chr11:5254897, Chr11:5254901	+	23
AsCpf1 HBG1 Promoter-1	AGACAGAUAUUUGCAUUGAG (SEQ ID NO:1139)	AGACAGATATTTGCATTGAG (SEQ ID NO:1140)	Chr11:5254947 Chr11:5254967	Chr11:5254966 Chr11:5254970	+	20
AsCpf1 HBG1 Promoter-1 (21mer)	AGACAGAUAUUUGCAUUGAGA (SEQ ID NO:1260)	AGACAGATATTTGCATTGAGA (SEQ ID NO:1261)	Chr11:5254947 Chr11:5254968	Chr11:5254966 Chr11:5254970	+	21
AsCpf1 HBG1 Promoter-2	CAUUGAGAUAGUGUGGGGAA (SEQ ID NO:1141)	CATTGAGATAGTGTGGGGAA (SEQ ID NO:1142)	Chr11:5254960 Chr11:5254980	Chr11:5254979 Chr11:5254983	+	20
AsCpf1 HBG1 Promoter-2 (21mer)	CAUUGAGAUAGUGUGGGGAAG (SEQ ID NO:1262)	CATTGAGATAGTGTGGGGAAG (SEQ ID NO:1263)	Chr11:5254960 Chr11:5254981	Chr11:5254979 Chr11:5254983	+	21

*NCBI Reference Sequence NC_000011, the coordinates are reported using the One-based coordinate system, "Homo sapiens chromosome 11, GRCh38.p12 Primary Assembly," (Version NC_000011.10).
Expected cleavage sites based on (Zetsche et al, 2015), coordinates are reported using zero-based coordinates.

Table 18: gRNA 5' Extensions

5' extension Sequence ID No:	5' extension sequence	5' modification
1231	rCrUrUrUrU	+5 RNA
1232	rArArGrArCrCrUrUrUrU	+10 RNA
1233	rArUrGrUrGrUrUrUrUrUrGrUrCrArArArArGrArCrCrUrUrUrU	+25 RNA
1234	rArGrGrCrCrArGrCrUrUrGrCrCrGrGrUrUrUrUrUrUrArGrUrCrGrUrGrCrUrGrCrUrUrCrArUrGrUrGrUrUrUrUrUrGrUrCrArArArArGrArCrCrUrUrUrU	+60 RNA
1235	CTTTT	+5 DNA
1236	AAGACCTTTT	+10 DNA
1237	ATGTGTTTTTGTCAAAAGACCTTTT	+25 DNA
1238	AGGCCAGCTTGCCGGTTTTTTAGTCGTGCTGCTTCATGTGTTTTTGTCAAAAGACCTTTT	+60 DNA
1239	TTTTTGTCAAAAGACCTTTT	+20 DNA
1240	GCTTCATGTGTTTTTGTCAAAAGACCTTTT	+30 DNA
1241	GCCGGTTTTTTAGTCGTGCTGCTTCATGTGTTTTTGTCAAAAGACCTTTT	+50 DNA
1242	TAGTCGTGCTGCTTCATGTGTTTTTGTCAAAAGACCTTTT	+40 DNA
1243	C*C*GAAGTTTTCTTCGGTTTT	+20 DNA + 2xPS
1244	T*T*TTTCCGAAGTTTTCTTCGGTTTT	+25 DNA + 2xPS
1245	A*A*CGCTTTTTCCGAAGTTTTCTTCGGTTTT	+30 DNA + 2xPS
1246	G*C*GTTGTTTTCAACGCTTTTTCCGAAGTTTTCTTCGGTTTT	+41 DNA + 2xPS
1247	G*G*CTTCTTTTGAAGCCTTTTTGCGTTGTTTTCAACGCTTTTTCCGAAGTTTTCTTCGGTTTT	+62 DNA + 2xPS
1248	A*T*GTGTTTTTGTCAAAAGACCTTTT	+25 DNA + 2xPS
1249	AAAAAAAAAAAAAAAAAAAAAAAAA	+25 A
1250	TTTTTTTTTTTTTTTTTTTTTTTTT	+25 T
1251	mA*mU*rGrUrGrUrUrUrUrGrUrCrArArArArGrArCrCrUrUrUrU	+25 RNA + 2xPS
1252	mA*mA*rArArArArArArArArArArArArArArArArArArArArArArA	PolyA RNA + 2xPS
1253	mU*mU*rUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrUrU	PolyU RNA + 2xPS

All bases are in upper case
Lowercase "r" represents RNA, 2'-hydroxy; bases not modified by an "r" are DNA
All bases are linked via standard phosphodiester bonds except as noted:
"*" represents phosphorothioate modification
"PS" represents phosphorothioate modification

Table 13: Cpf1 guide RNAs

gRNA Sequence SEQ ID NO.	gRNA Sequence	5' mod.**	3' mod.	Length of crRNA + gRNA targeting domain	Length of gRNA targeting domain	gRNA Targeting Domain (RNA)
1022	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrGrCrUrArUrUrGrGrUrC	-	-	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1023	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrGrCrUrArUrUrGrGrUmC/PS2/mA	-	1xPS2-OMe + 1xOMe	41	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)
1041	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrGrCrUrArUrUrGrGmU/PS2/mC	-	1xPS2-OMe + 1xOMe	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1042	rCrUrUrUrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrGrCrUrArUrUrGrGrUrC	+5 RNA	-	45	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1043	rArArGrArCrCrUrUrUrUrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrGrCrUrArUrUrGrGrUrC	+10 RNA	-	50	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1044	rArUrGrUrGrUrUrUrUrUrGrUrCrArArArArGrArCrCrUrUrUrUrArArUr	+25 RNA	-	65	20	CCUUGUCAAGGCUAUUGGUC

	UrUrCrUrArCrUrC rUrUrGrUrArGrAr UrCrCrUrUrGrUrC rArArGrGrCrUrAr UrUrGrGrUrC					(SEQ ID NO:1002)
1045	rArGrGrCrCrArGr CrUrUrGrCrCrGrG rUrUrUrUrUrUrAr GrUrCrGrUrGrCrU rGrCrUrUrCrArUr GrUrGrUrUrUrUr UrGrUrCrArArArA rGrArCrCrUrUrUr UrUrArArUrUrUrC rUrArCrUrCrUrUr GrUrArGrArUrCrC rUrUrGrUrCrArAr GrGrCrUrArUrUrG rGrUrC	+60 RNA	-	100	20	CCUUGUC AAGGCUA UUGGUC (SEQ ID NO:1002)
1046	CTTTTrUrArArUr UrUrCrUrArCrUrC rUrUrGrUrArGrAr UrCrCrUrUrGrUrC rArArGrGrCrUrAr UrUrGrGrUrC	+5 DNA	-	45	20	CCUUGUC AAGGCUA UUGGUC (SEQ ID NO:1002)
1047	AAGACCTTTTrUr ArArUrUrUrCrUrA rCrUrCrUrUrGrUr ArGrArUrCrCrUrU rGrUrCrArArGrGr CrUrArUrUrGrGrU rC	+10 DNA	-	50	20	CCUUGUC AAGGCUA UUGGUC (SEQ ID NO:1002)
1048	ATGTGTTTTTGT CAAAAGACCTT TTrUrArArUrUrUr CrUrArCrUrCrUrU rGrUrArGrArUrCr CrUrUrGrUrCrArA rGrGrCrUrArUrUr GrGrUrC	+25 DNA	-	65	20	CCUUGUC AAGGCUA UUGGUC (SEQ ID NO:1002)
1049	AGGCCAGCTTG CCGGTTTTTTAG TCGTGCTGCTTC ATGTGTTTTTGT	+60 DNA	-	100	20	CCUUGUC AAGGCUA UUGGUC

	CAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC					(SEQ ID NO:1002)
1050	ATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUmC/PS2/mA	+25 DNA	1xPS2-OMe + 1xOMe	66	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)
1051	ATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC*mA	+25 DNA	1xPS-OMe	66	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)
1052	TTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+20 DNA	-	60	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1053	GCTTCATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+30 DNA	-	70	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1054	GCCGGTTTTTTAGTCGTGCTGCTTCATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrU	+50 DNA	-	90	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)

	rUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC					
1055	TAGTCGTGCTGCTTCATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+40 DNA	-	80	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1056	C*C*GAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+20 DNA + 2xPS	-	60	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1057	T*T*TTTCCGAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+25 DNA + 2xPS	-	65	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1058	A*A*CGCTTTTTCCGAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+30 DNA + 2xPS	-	70	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1059	G*C*GTTGTTTTCAACGCTTTTTCCGAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrU	+41 DNA + 2xPS	-	81	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)

	rCrArArGrGrCrUrArUrUrGrGrUrC					
1060	G*G*CTTCTTTTGAAGCCTTTTTGCGTTGTTTTCAACGCTTTTTCCGAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+62 DNA + 2xPS	-	102	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1061	mUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	1xOMe	-	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1062	mU*rArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	1xPS-OMe	-	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1063	mUmArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	2xOMe	-	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1064	mU*mA*rArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	2xPS-OMe	-	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1065	mUmAmArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArA	3xOMe	-	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)

	rGrGrCrUrArUrUrGrGrUrC					
1066	mU*mA*mA*rUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	3xPS-OMe	-	40	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1067	A*T*GTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC	+25 DNA + 2xPS	-	65	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1068	A*T*GTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC*mA	+25 DNA + 2xPS	1xPS-OMe	66	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)
1069	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrC*mA	-	1xPS-OMe	41	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)
1070	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrCmA*rU	-	1xPS-OMe + rU	42	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)
1071	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrCrU	-	rU	41	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)

1072	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrCrArU	-	rU	42	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)
1073	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrCrArArGrU	-	rU	44	23	CCUUGUCAAGGCUAUUGGUCAAG (SEQ ID NO:1258)
1074	A*T*GTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrUrGrGrUrCrU	+25 DNA + 2xPS	rU	66	20	CCUUGUCAAGGCUAUUGGUC (SEQ ID NO:1002)
1075	A*T*GTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrCrArArGrU	+25 DNA + 2xPS	rU	69	23	CCUUGUCAAGGCUAUUGGUCAAG (SEQ ID NO:1258)
1076	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrCrA	-	-	41	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)
1077	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrGrGrUrCrArArG	-	-	43	23	CCUUGUCAAGGCUAUUGGUCAAG (SEQ ID NO:1258)

1078	A*T*GTGTTTTT GTCAAAAGACC TTTTrUrArArUrUr UrCrUrArCrUrCrU rUrGrUrArGrArUr CrCrUrUrGrUrCrA rArGrGrCrUrArUr UrGrGrUrCrArArG	+25 DNA + 2xPS	-	68	23	CCUUGUC AAGGCUA UUGGUCA AG (SEQ ID NO:1258)
1079	rUrArArUrUrUrCr UrArCrUrCrUrUrG rUrArGrArUrCrCr UrUrGrUrCrArArG rGrCrUrArUrUrGr GrUrCrUrUrUrU	-	4xrU	44	20	CCUUGUC AAGGCUA UUGGUC (SEQ ID NO:1002)
1080	rUrArArUrUrUrCr UrArCrUrCrUrUrG rUrArGrArUrCrCr UrUrGrUrCrArArG rGrCrUrArUrUrGr GrUrCmU*mU*m U*rU	-	3xPS-OMe-U + rU	44	20	CCUUGUC AAGGCUA UUGGUC (SEQ ID NO:1002)
1081	A*T*GTGTTTTT GTCAAAAGACC TTTTrUrArArUrUr UrCrUrArCrUrCrU rUrGrUrArGrArUr CrCrUrUrGrUrCrA rArGrGrCrUrArUr UrGrGrUrCrUrUrU rU	+25 DNA + 2xPS	4xrU	69	20	CCUUGUC AAGGCUA UUGGUC (SEQ ID NO:1002)
1082	AAAAAAAAAAA AAAAAAAAAAA AAArUrArArUrUr UrCrUrArCrUrCrU rUrGrUrArGrArUr CrCrUrUrGrUrCrA rArGrGrCrUrArUr UrGrGrUrC*mA	+25 A	1xPS-OMe	66	21	CCUUGUC AAGGCUA UUGGUCA (SEQ ID NO:1254)
1083	TTTTTTTTTTTTT TTTTTTTTTTTTr UrArArUrUrUrCrU rArCrUrCrUrUrGr UrArGrArUrCrCrU rUrGrUrCrArArGr	+25 T	1xPS-OMe	66	21	CCUUGUC AAGGCUA UUGGUCA (SEQ ID NO:1254)

	GrCrUrArUrUrGrGrUrC*mA					
1084	C*C*GAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrCrUrUrGrUrCrArArGrGrCrUrArUrUrUrGrGrUrC*mA	+20 DNA + 2xPS	1xPS-OMe	61	21	CCUUGUCAAGGCUAUUGGUCA (SEQ ID NO:1254)
1085	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArGrArCrArGrArUrArUrUrUrGrCrArUrUrGrArG*mA	-	1xPS-OMe	41	21	AGACAGAUAUUUGCAUUGAGA (SEQ ID NO:1260)
1086	ATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArGrArCrArGrArUrArUrUrUrGrCrArUrUrGrArG*mA	+25 DNA	1xPS-OMe	66	21	AGACAGAUAUUUGCAUUGAGA (SEQ ID NO:1260)
1087	C*C*GAAGTTTTCTTCGGTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrArGrArCrArGrArUrArUrUrUrGrCrArUrUGrArG*mA	+20 DNA + 2xPS	1xPS-OMe	61	21	AGACAGAUAUUUGCAUUGAGA (SEQ ID NO:1260)
1088	rUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCrArUrUrGrArGrArUrArGrUrGrUrGrGrGrGrArA*mG	-	1xPS-OMe	41	21	CAUUGAGAUAGUGUGGGGAAG (SEQ ID NO:1262)
1089	ATGTGTTTTTGTCAAAAGACCTTTTrUrArArUrUrUrCrUrArCrUrCrUrUrGrUrArGrArUrCr	+25 DNA	1xPS-OMe	66	21	CAUUGAGAUAGUGUGGGGAAG

	ArUrUrGrArGrAr UrArGrUrGrUrGr GrGrGrArA*mG					(SEQ ID NO:1262)
1090	C*C*GAAGTTTT CTTCGGTTTTrUr ArArUrUrUrCrUrA rCrUrCrUrUrGrUr ArGrArUrCrArUrU rGrArGrArUrArGr UrGrUrGrGrGrGr ArA*mG	+20 DNA + 2xPS	1xPS-OMe	61	21	CAUUGAG AUAGUGU GGGGAAG (SEQ ID NO:1262)
1091	rUrArArUrUrUrCr UrArCrUrCrUrUrG rUrArGrArUrCrUr UrCrUrCrCrCrArU rCrArUrArGrArGr GrArU*mA	-	1xPS-OMe	41	21	CUUCUCC CAUCAUA GAGGAUA (SEQ ID NO:1264)
1092	ATGTGTTTTTGT CAAAAGACCTT TTrUrArArUrUrUr CrUrArCrUrCrUrU rGrUrArGrArUrCr UrUrCrUrCrCrCrA rUrCrArUrArGrAr GrGrArU*mA	+25 DNA	1xPS-OMe	66	21	CUUCUCC CAUCAUA GAGGAUA (SEQ ID NO:1264)
1093	C*C*GAAGTTTT CTTCGGTTTTrUr ArArUrUrUrCrUrA rCrUrCrUrUrGrUr ArGrArUrCrUrUrC rUrCrCrCrArUrCr ArUrArGrArGrGr ArU*mA	+20 DNA + 2xPS	1xPS-OMe	61	21	CUUCUCC CAUCAUA GAGGAUA (SEQ ID NO:1264)
1098	A*T*GTGTTTTT GTCAAAAGACC TTTTrUrArArUrUr UrCrUrArCrUrCrU rUrGrUrArGrArUr CrCrUrUrGrUrCrA rArGrGrCrUrArUr UrGrGrUrCmA*rU	+25 DNA + 2xPS	1xPS-OMe + rU	67	21	CCUUGUC AAGGCUA UUGGUCA (SEQ ID NO:1254)

1099	A*T*GTGTTTTT GTCAAAAGACC TTTTrUrArArUrUr UrCrUrArCrUrCrU rUrGrUrArGrArUr CrCrUrUrGrUrCrA rArGrGrCrUrArUr UrGrGrUrCmA*m A*rU	+25 DNA + 2xPS	2xPS-OMe + rU	68	22	CCUUGUC AAGGCUA UUGGUCA A (SEQ ID NO:1256)
1100	A*T*GTGTTTTT GTCAAAAGACC TTTTrUrArArUrUr UrCrUrArCrUrCrU rUrGrUrArGrArUr CrCrUrUrGrUrCrA rArGrGrCrUrArUr UrGrGrUmC*rU	+25 DNA + 2xPS	1xPS-OMe + rU	66	20	CCUUGUC AAGGCUA UUGGUC (SEQ ID NO:1002)
1101	A*T*GTGTTTTT GTCAAAAGACC TTTTrUrArArUrUr UrCrUrArCrUrCrU rUrGrUrArGrArUr CrCrUrUrGrUrCrA rArGrGrCrUrArUr UrGrGrU*mC	+25 DNA + 2xPS	1xPS-OMe	65	20	CCUUGUC AAGGCUA UUGGUC (SEQ ID NO:1002)
1102	A*T*GTGTTTTT GTCAAAAGACC TTTTrUrArArUrUr UrCrUrArCrUrCrU rUrGrUrArGrArUr CrCrUrUrGrUrCrA rArGrGrCrUrArUr UrGrGrUrC*mA* mA	+25 DNA + 2xPS	2xPS-OMe	67	22	CCUUGUC AAGGCUA UUGGUCA A (SEQ ID NO:1256)
1103	mA*mU*rGrUrGr UrUrUrUrUrGrUrC rArArArArGrArCr CrUrUrUrUrUrArA rUrUrUrCrUrArCr UrCrUrUrGrUrArG rArUrCrCrUrUrGr UrCrArArGrGrCrU rArUrUrGrGrUrC* mA	+25 RNA + 2xPS	1xPS-OMe	66	21	CCUUGUC AAGGCUA UUGGUCA (SEQ ID NO:1254)

1104	mA*mA*rArArAr ArArArArArArAr ArArArArArArAr ArArArArArArUr ArArUrUrUrCrUrA rCrUrCrUrUrGrUr ArGrArUrCrCrUrU rGrUrCrArArGrGr CrUrArUrUrGrGrU rC*mA	PolyA RNA + 2xPS	1xPS-OMe	66	21	CCUUGUC AAGGCUA UUGGUCA (SEQ ID NO:1254)
1105	mU*mU*rUrUrUr UrUrUrUrUrUrUr UrUrUrUrUrUrUr UrUrUrUrUrUrUr ArArUrUrUrCrUrA rCrUrCrUrUrGrUr ArGrArUrCrCrUrU rGrUrCrArArGrGr CrUrArUrUrGrGrU rC*mA	PolyU RNA + 2xPS	1xPS-OMe	66	21	CCUUGUC AAGGCUA UUGGUCA (SEQ ID NO:1254)
All bases are in upper case Lowercase "r" represents RNA, 2'-hydroxy; bases not modified by an "r" are DNA All bases are linked via standard phosphodiester bonds except as noted: "*" represents phosphorothioate modification "PS" represents phosphorothioate modification "PS2" represents phosphorodithioate modification "OMe" represents a 2'-o-methyl modification "m" represents a 2'-o-methyl modification **__Table 18__ provides a listing of the sequences of the gRNA 5' extensions						

Table 24: Sp35 RNP and RNP58 Indels

Indel_id	"% in indel" SpCas9 (Sp35 RNP)	"% in indel" Cpf1 (RNP58)	MMEJ	Genomic Coordinates at HBG1	Genomic Coordinates at HBG2	HBG1 and HBG2 Position
159_-18_NA	0.013503714	0.185309532	True	chr11:5,249,961-5,249,978	chr11:5,254,885-5,254,902	HBG1/2 c.-104 to -121
165_-6_NA	0.011778239	0.050633257	False	chr11:5,249,967-5,249,972	chr11:5,254,891-5,254,896	HBG1/2 c.-110 to -115
167_-4_NA	0.046087674	0.049159297	False	chr11:5,249,969-5,249,972	chr11:5,254,893-5,254,896	HBG1/2 c.-112 to -115
168_-3_NA	0.072269874	0.044737417	False	chr11:5,249,970-5,249,972	chr11:5,254,894-5,254,896	HBG1/2 c.-113 to -115
157_-13_NA	0.318812673	0.033628125	True	chr11:5,249,959-5,249,971	chr11:5,254,883-5,254,895	HBG1/2 c.-102 to -114

166_-7_NA	0.001100303	0.027923354	False	chr11:5,249,968-5,249,974	chr11:5,254,892-5,254,898	HBG1/2 c.-111 to -117
166_-5_NA	0.007076946	0.025412163	False	chr11:5,249,968-5,249,972	chr11:5,254,892-5,254,896	HBG1/2 c.-111 to -115
169_-5_NA	0.000775213	0.024948138	False	chr11:5,249,971-5,249,975	chr11:5,254,895-5,254,899	HBG1/2 c.-114 to -118
168_-5_NA	0.000475131	0.018288023	False	chr11:5,249,970-5,249,974	chr11:5,254,894-5,254,898	HBG1/2 c.-113 to -117
169_-4_NA	0.001400385	0.017605634	False	chr11:5,249,971-5,249,974	chr11:5,254,895-5,254,898	HBG1/2 c.-114 to -117
167_-5_NA	0.000700193	0.014548531	False	chr11:5,249,969-5,249,973	chr11:5,254,893-5,254,897	HBG1/2 c.-112 to -116
164_-8_NA	0.002275626	0.01375696	False	chr11:5,249,966-5,249,973	chr11:5,254,890-5,254,897	HBG1/2 c.-109 to -116
166_-8_NA	0.000625172	0.011982749	False	chr11:5,249,968-5,249,975	chr11:5,254,892-5,254,899	HBG1/2 c.-111 to -118
166_-12_NA	0.000300083	0.01031772	False	chr11:5,249,968-5,249,979	chr11:5,254,892-5,254,903	HBG1/2 c.-111 to -122
169_-2_NA	0.084173148	0.010263129	False	chr11:5,249,971-5,249,972	chr11:5,254,895-5,254,896	HBG1/2 c.-114 to -115
167_-6_NA	0.000500138	0.008952942	False	chr11:5,249,969-5,249,974	chr11:5,254,893-5,254,898	HBG1/2 c.-112 to -117
157_-18_NA	0.001650454	0.008024894	False	chr11:5,249,959-5,249,976	chr11:5,254,883-5,254,900	HBG1/2 c.-102 to -119
166_-13_NA	0.012028308	0.006960367	False	chr11:5,249,968-5,249,980	chr11:5,254,892-5,254,904	HBG1/2 c.-111 to -123
168_-10_NA	0.000050014	0.006359865	False	chr11:5,249,970-5,249,979	chr11:5,254,894-5,254,903	HBG1/2 c.-113 to -122
162_-22_NA	0.001050289	0.006086909	True	chr11:5,249,964-5,249,985	chr11:5,254,888-5,254,909	HBG1/2 c.-107 to -128
169_-8_NA	0.000100028	0.00589584	False	chr11:5,249,971-5,249,978	chr11:5,254,895-5,254,902	HBG1/2 c.-114 to -121
169_-9_NA	0.000050014	0.005868545	False	chr11:5,249,971-5,249,979	chr11:5,254,895-5,254,903	HBG1/2 c.-114 to -122
165_-7_NA	0.000425117	0.005704771	False	chr11:5,249,967-5,249,973	chr11:5,254,891-5,254,897	HBG1/2 c.-110 to -116
169_-7_NA	0.000325089	0.005677476	False	chr11:5,249,971-5,249,977	chr11:5,254,895-5,254,901	HBG1/2 c.-114 to -120
168_-8_NA	0.000200055	0.005349929	False	chr11:5,249,970-5,249,977	chr11:5,254,894-5,254,901	HBG1/2 c.-113 to -120
170_-2_NA	0.001025282	0.00515886	False	chr11:5,249,972-5,249,973	chr11:5,254,896-5,254,897	HBG1/2 c.-115 to -116
168_-9_NA	0	0.005022382	False	chr11:5,249,970-5,249,978	chr11:5,254,894-5,254,902	HBG1/2 c.-113 to -121
168_-6_NA	0.000150041	0.004967791	False	chr11:5,249,970-5,249,975	chr11:5,254,894-5,254,899	HBG1/2 c.-113 to -118

169_-10_NA	0.000050014	0.004858609	False	chr11:5,249,971-5,249,980	chr11:5,254,895-5,254,904	HBG1/2 c.-114 to -123
166_-11_NA	0.000025007	0.004749427	False	chr11:5,249,968-5,249,978	chr11:5,254,892-5,254,902	HBG1/2 c.-111 to -121
164_-23_NA	0.000125034	0.004722131	False	chr11:5,249,966-5,249,988	chr11:5,254,890-5,254,912	HBG1/2 c.-109 to -131
166_-10_NA	0.000050014	0.00417622	False	chr11:5,249,968-5,249,977	chr11:5,254,892-5,254,901	HBG1/2 c.-111 to -120
167_-11_NA	0	0.004039742	False	chr11:5,249,969-5,249,979	chr11:5,254,893-5,254,903	HBG1/2 c.-112 to -122
168_-7_NA	0.000125034	0.003957856	False	chr11:5,249,970-5,249,976	chr11:5,254,894-5,254,900	HBG1/2 c.-113 to -119
168_-11_NA	0.000125034	0.003794082	False	chr11:5,249,970-5,249,980	chr11:5,254,894-5,254,904	HBG1/2 c.-113 to -123
166_-9_NA	0.000025007	0.003766787	False	chr11:5,249,968-5,249,976	chr11:5,254,892-5,254,900	HBG1/2 c.-111 to -119
167_-1_NA	0.000975268	0.0036849	False	chr11:5,249,969-5,249,969	chr11:5,254,893-5,254,893	HBG1/2 c.-112 to -112
168_-2_NA	0.003325915	0.003575718	False	chr11:5,249,970-5,249,971	chr11:5,254,894-5,254,895	HBG1/2 c.-113 to -114
167_-9_NA	0.000075021	0.003521127	False	chr11:5,249,969-5,249,977	chr11:5,254,893-5,254,901	HBG1/2 c.-112 to -120
168_-12_NA	0.000200055	0.003521127	False	chr11:5,249,970-5,249,981	chr11:5,254,894-5,254,905	HBG1/2 c.-113 to -124
166_-15_NA	0.000050014	0.003493831	False	chr11:5,249,968-5,249,982	chr11:5,254,892-5,254,906	HBG1/2 c.-111 to -125
162_-7_NA	0.005901623	0.00343924	True	chr11:5,249,964-5,249,970	chr11:5,254,888-5,254,894	HBG1/2 c.-107 to -113
169_-12_NA	0.001775488	0.003384649	False	chr11:5,249,971-5,249,982	chr11:5,254,895-5,254,906	HBG1/2 c.-114 to -125
168_-13_NA	0.001000275	0.003357353	False	chr11:5,249,970-5,249,982	chr11:5,254,894-5,254,906	HBG1/2 c.-113 to -125
166_-4_NA	0.001000275	0.003111693	False	chr11:5,249,968-5,249,971	chr11:5,254,892-5,254,895	HBG1/2 c.-111 to -114
167_-10_NA	0.000050014	0.003057102	False	chr11:5,249,969-5,249,978	chr11:5,254,893-5,254,902	HBG1/2 c.-112 to -121
169_-13_NA	0.001500413	0.002756851	False	chr11:5,249,971-5,249,983	chr11:5,254,895-5,254,907	HBG1/2 c.-114 to -126
157_-19_NA	0.000075021	0.002483896	False	chr11:5,249,959-5,249,977	chr11:5,254,883-5,254,901	HBG1/2 c.-102 to -120
166_-14_NA	0.000025007	0.0024566	False	chr11:5,249,968-5,249,981	chr11:5,254,892-5,254,905	HBG1/2 c.-111 to -124
158_-14_NA	0.000050014	0.002429305	False	chr11:5,249,960-5,249,973	chr11:5,254,884-5,254,897	HBG1/2 c.-103 to -116
167_-3_NA	0.002700743	0.002402009	False	chr11:5,249,969-5,249,971	chr11:5,254,893-5,254,895	HBG1/2 c.-112 to -114

163_-9_NA	0.000200055	0.002265531	False	chr11:5,249,965-5,249,973	chr11:5,254,889-5,254,897	HBG1/2 c.-108 to -116
162_-8_NA	0.006076671	0.002156349	False	chr11:5,249,964-5,249,971	chr11:5,254,888-5,254,895	HBG1/2 c.-107 to -114
168_-14_NA	0	0.002129053	False	chr11:5,249,970-5,249,983	chr11:5,254,894-5,254,907	HBG1/2 c.-113 to -126
167_-16_NA	0.000050014	0.00196528	False	chr11:5,249,969-5,249,984	chr11:5,254,893-5,254,908	HBG1/2 c.-112 to -127
166_-16_NA	0	0.001910689	False	chr11:5,249,968-5,249,983	chr11:5,254,892-5,254,907	HBG1/2 c.-111 to -126
169_-14_NA	0.000125034	0.001910689	False	chr11:5,249,971-5,249,984	chr11:5,254,895-5,254,908	HBG1/2 c.-114 to -127
165_-3_NA	0.000925254	0.001883393	False	chr11:5,249,967-5,249,969	chr11:5,254,891-5,254,893	HBG1/2 c.-110 to -112
157_-23_NA	0.000100028	0.001856098	False	chr11:5,249,959-5,249,981	chr11:5,254,883-5,254,905	HBG1/2 c.-102 to -124
165_-14_NA	0.000025007	0.001856098	False	chr11:5,249,967-5,249,980	chr11:5,254,891-5,254,904	HBG1/2 c.-110 to -123
166_-27_NA	0.000225062	0.00171962	True	chr11:5,249,968-5,249,994	chr11:5,254,892-5,254,918	HBG1/2 c.-111 to -137
165_-9_NA	0.000200055	0.001692324	False	chr11:5,249,967-5,249,975	chr11:5,254,891-5,254,899	HBG1/2 c.-110 to -118
167_-18_NA	0.000250069	0.001665029	False	chr11:5,249,969-5,249,986	chr11:5,254,893-5,254,910	HBG1/2 c.-112 to -129
156_-27_NA	0	0.001610438	True	chr11:5,249,958-5,249,984	chr11:5,254,882-5,254,908	HBG1/2 c.-101 to -127
165_-2_NA	0.000425117	0.001610438	False	chr11:5,249,967-5,249,968	chr11:5,254,891-5,254,892	HBG1/2 c.-110 to -111
150_-21_NA	0.000225062	0.001583142	False	chr11:5,249,952-5,249,972	chr11:5,254,876-5,254,896	HBG1/2 c.-95 to -115
152_-24_NA	0.000225062	0.001583142	True	chr11:5,249,954-5,249,977	chr11:5,254,878-5,254,901	HBG1/2 c.-97 to -120
165_-12_NA	0.000225062	0.001583142	False	chr11:5,249,967-5,249,978	chr11:5,254,891-5,254,902	HBG1/2 c.-110 to -121
167_-2_NA	0.000850234	0.001583142	False	chr11:5,249,969-5,249,970	chr11:5,254,893-5,254,894	HBG1/2 c.-112 to -113
164_-9_NA	0.000150041	0.001555847	False	chr11:5,249,966-5,249,974	chr11:5,254,890-5,254,898	HBG1/2 c.-109 to -117
165_-11_NA	0.000025007	0.001528551	False	chr11:5,249,967-5,249,977	chr11:5,254,891-5,254,901	HBG1/2 c.-110 to -120
159_-28_NA	0.000125034	0.001501256	True	chr11:5,249,961-5,249,988	chr11:5,254,885-5,254,912	HBG1/2 c.-104 to -131
169_-6_NA	0.013753782	0.001501256	True	chr11:5,249,971-5,249,976	chr11:5,254,895-5,254,900	HBG1/2 c.-114 to -119
160_-12_NA	0.000050014	0.00147396	False	chr11:5,249,962-5,249,973	chr11:5,254,886-5,254,897	HBG1/2 c.-105 to -116

145_-32_NA	0.001125309	0.001446664	True	chr11:5,249,947-5,249,978	chr11:5,254,871-5,254,902	HBG1/2 c.-90 to -121
169_-1_NA	0.096276476	0.001446664	False	chr11:5,249,971-5,249,971	chr11:5,254,895-5,254,895	HBG1/2 c.-114 to -114
163_-10_NA	0.000125034	0.001419369	False	chr11:5,249,965-5,249,974	chr11:5,254,889-5,254,898	HBG1/2 c.-108 to -117
152_-33_NA	0.000200055	0.001392073	True	chr11:5,249,954-5,249,986	chr11:5,254,878-5,254,910	HBG1/2 c.-97 to -129
156_-15_NA	0.000750206	0.001392073	False	chr11:5,249,958-5,249,972	chr11:5,254,882-5,254,896	HBG1/2 c.-101 to -115
167_-14_NA	0.000075021	0.001337482	False	chr11:5,249,969-5,249,982	chr11:5,254,893-5,254,906	HBG1/2 c.-112 to -125
158_-16_NA	0	0.001310187	False	chr11:5,249,960-5,249,975	chr11:5,254,884-5,254,899	HBG1/2 c.-103 to -118
162_-9_NA	0.002200605	0.001310187	False	chr11:5,249,964-5,249,972	chr11:5,254,888-5,254,896	HBG1/2 c.-107 to -115
164_-10_NA	0.000325089	0.001310187	False	chr11:5,249,966-5,249,975	chr11:5,254,890-5,254,899	HBG1/2 c.-109 to -118
165_-5_NA	0.001975543	0.001310187	False	chr11:5,249,967-5,249,971	chr11:5,254,891-5,254,895	HBG1/2 c.-110 to -114
169_-11_NA	0.015704319	0.001310187	True	chr11:5,249,971-5,249,981	chr11:5,254,895-5,254,905	HBG1/2 c.-114 to -124
170_-4_NA	0.00040011	0.001310187	False	chr11:5,249,972-5,249,975	chr11:5,254,896-5,254,899	HBG1/2 c.-115 to -118
152_-29_NA	0.000450124	0.001282891	True	chr11:5,249,954-5,249,982	chr11:5,254,878-5,254,906	HBG1/2 c.-97 to -125
161_-11_NA	0	0.0012283	False	chr11:5,249,963-5,249,973	chr11:5,254,887-5,254,897	HBG1/2 c.-106 to -116
169_-18_NA	0.000100028	0.0012283	False	chr11:5,249,971-5,249,988	chr11:5,254,895-5,254,912	HBG1/2 c.-114 to -131
164_-27_NA	0.000275076	0.001201004	False	chr11:5,249,966-5,249,992	chr11:5,254,890-5,254,916	HBG1/2 c.-109 to -135
168_-18_NA	0.000100028	0.001201004	False	chr11:5,249,970-5,249,987	chr11:5,254,894-5,254,911	HBG1/2 c.-113 to -130
168_-23_NA	0	0.001201004	False	chr11:5,249,970-5,249,992	chr11:5,254,894-5,254,916	HBG1/2 c.-113 to -135
150_-33_NA	0.000150041	0.001173709	False	chr11:5,249,952-5,249,984	chr11:5,254,876-5,254,908	HBG1/2 c.-95 to -127
151_-20_NA	0.000225062	0.001173709	False	chr11:5,249,953-5,249,972	chr11:5,254,877-5,254,896	HBG1/2 c.-96 to -115
169_-22_NA	0.000025007	0.001173709	False	chr11:5,249,971-5,249,992	chr11:5,254,895-5,254,916	HBG1/2 c.-114 to -135
162_-14_NA	0.000125034	0.001146413	False	chr11:5,249,964-5,249,977	chr11:5,254,888-5,254,901	HBG1/2 c.-107 to -120
166_-21_NA	0.000025007	0.001146413	False	chr11:5,249,968-5,249,988	chr11:5,254,892-5,254,912	HBG1/2 c.-111 to -131

160_-11_NA	0.001675461	0.001119118	False	chr11:5,249,962-5,249,972	chr11:5,254,886-5,254,896	HBG1/2 c.-105 to -115
155_-16_NA	0.000350096	0.001091822	False	chr11:5,249,957-5,249,972	chr11:5,254,881-5,254,896	HBG1/2 c.-100 to -115
166_-26_NA	0	0.001091822	False	chr11:5,249,968-5,249,993	chr11:5,254,892-5,254,917	HBG1/2 c.-111 to -136
157_-24_NA	0.000125034	0.001064527	False	chr11:5,249,959-5,249,982	chr11:5,254,883-5,254,906	HBG1/2 c.-102 to -125
165_-15_NA	0.000050014	0.001064527	False	chr11:5,249,967-5,249,981	chr11:5,254,891-5,254,905	HBG1/2 c.-110 to -124
162_-33_NA	0.000025007	0.001037231	True	chr11:5,249,964-5,249,996	chr11:5,254,888-5,254,920	HBG1/2 c.-107 to -139
163_-12_NA	0.000100028	0.001037231	False	chr11:5,249,965-5,249,976	chr11:5,254,889-5,254,900	HBG1/2 c.-108 to -119
163_-6_NA	0.001050289	0.001009936	False	chr11:5,249,965-5,249,970	chr11:5,254,889-5,254,894	HBG1/2 c.-108 to -113
164_-28_NA	0.000175048	0.001009936	False	chr11:5,249,966-5,249,993	chr11:5,254,890-5,254,917	HBG1/2 c.-109 to -136
155_-12_NA	0.001150316	0.000928049	True	chr11:5,249,957-5,249,968	chr11:5,254,881-5,254,892	HBG1/2 c.-100 to -111
137_-27_NA	0.003901073	0.000846162	True	chr11:5,249,939-5,249,965	chr11:5,254,863-5,254,889	HBG1/2 c.-82 to -108
161_-10_NA	0.001625447	0.000655093	False	chr11:5,249,963-5,249,972	chr11:5,254,887-5,254,896	HBG1/2 c.-106 to -115
170_-10_NA	0.001050289	0.000545911	False	chr11:5,249,972-5,249,981	chr11:5,254,896-5,254,905	HBG1/2 c.-115 to -124
171_-1_NA	0.018029958	0.000464024	False	chr11:5,249,973-5,249,973	chr11:5,254,897-5,254,897	HBG1/2 c.-116 to -116
154_-18_NA	0.001325364	0.000327547	False	chr11:5,249,956-5,249,973	chr11:5,254,880-5,254,897	HBG1/2 c.-99 to -116
168_-15_NA	0.007727125	0.000300251	True	chr11:5,249,970-5,249,984	chr11:5,254,894-5,254,908	HBG1/2 c.-113 to -127
171_-2_NA	0.003325915	0.000191069	False	chr11:5,249,973-5,249,974	chr11:5,254,897-5,254,898	HBG1/2 c.-116 to -117
170_1_A	0.001025282	0.000163773	False	chr11:5,249,972-5,249,970	chr11:5,254,896-5,254,894	HBG1/2 c.-115 to -113
171_1_C	0.007877166	0.000054591	False	chr11:5,249,973-5,249,971	chr11:5,254,897-5,254,895	HBG1/2 c.-116 to -114
169_2_GG	0.004901348	0.000027296	False	chr11:5,249,971-5,249,968	chr11:5,254,895-5,254,892	HBG1/2 c.-114 to -111
169_1_G	0.033684263	0	False	chr11:5,249,971-5,249,969	chr11:5,254,895-5,254,893	HBG1/2 c.-114 to -112
170_4_CCTT	0.003876066	0	False	chr11:5,249,972-5,249,967	chr11:5,254,896-5,254,891	HBG1/2 c.-115 to -110
170_4_GCTT	0.001450399	0	False	chr11:5,249,972-5,249,967	chr11:5,254,896-5,254,891	HBG1/2 c.-115 to -110

170_5_GCCTT	0.001175323	0	False	chr11:5,249,972-5,249,966	chr11:5,254,896-5,254,890	HBG1/2 c.-115 to -109
170_-301_NA	0.001225337	0	True	chr11:5,249,972-5,250,272	chr11:5,254,896-5,255,196	HBG1/2 c.-115 to -415
171_1_A	0.005251444	0	False	chr11:5,249,973-5,249,971	chr11:5,254,897-5,254,895	HBG1/2 c.-116 to -114
171_1_T	0.027407537	0	False	chr11:5,249,973-5,249,971	chr11:5,254,897-5,254,895	HBG1/2 c.-116 to -114
171_2_CC	0.007151967	0	False	chr11:5,249,973-5,249,970	chr11:5,254,897-5,254,894	HBG1/2 c.-116 to -113
171_2_CT	0.002650729	0	False	chr11:5,249,973-5,249,970	chr11:5,254,897-5,254,894	HBG1/2 c.-116 to -113
171_2_GA	0.001050289	0	False	chr11:5,249,973-5,249,970	chr11:5,254,897-5,254,894	HBG1/2 c.-116 to -113
171_2_TC	0.003275901	0	False	chr11:5,249,973-5,249,970	chr11:5,254,897-5,254,894	HBG1/2 c.-116 to -113
171_2_TT	0.002025557	0	False	chr11:5,249,973-5,249,970	chr11:5,254,897-5,254,894	HBG1/2 c.-116 to -113
171_3_CCC	0.001125309	0	False	chr11:5,249,973-5,249,969	chr11:5,254,897-5,254,893	HBG1/2 c.-116 to -112

The Indel_id is a string identifying the indel, shown as indel_start_position + _ + indel_length + _ + ID. Where ID is NA for deletions and for insertions is the sequence inserted. Positive values indicate insertions and negative values indicate deletions.

The "% in indel" is the percentage of each indel_ID amongst all indels detected in a sample.

In the column labeled MMEJ, "True" indicates the repair was mediated via MMEJ, and "False" indicates the repair was mediated via NHEJ.

The genomic coordinates at *HBG1* or *HBG2* are reported using the One-based coordinate system, "Homo sapiens chromosome 11, GRCh38.p12 Primary Assembly," (Version NC_000011.10), NCBI Reference Sequence NC_000011.

The *HBG1/2* position indicates the deletion relative to the TSS.

Table 25: RNP32 Indels

Indel Id	Ave % in Indel	Cts	Cum. Sum	MMEJ	Genomic Coordinates at HBG1	Genomic Coordinates at HBG2	HBG1/2 position
159_-18_NA	15.15%	14	15.15%	True	chr11:5,249,961-5,249,978	chr11:5,254,885-5,254,902	HBG1/2 c.-104 to -121
165_-6_NA	5.92%	14	21.07%	False	chr11:5,249,967-5,249,972	chr11:5,254,891-5,254,896	HBG1/2 c.-110 to -115
167_-4_NA	5.53%	14	26.60%	False	chr11:5,249,969-5,249,972	chr11:5,254,893-5,254,896	HBG1/2 c.-112 to -115
168_-3_NA	4.18%	14	30.78%	False	chr11:5,249,970-5,249,972	chr11:5,254,894-5,254,896	HBG1/2 c.-113 to -115
166_-5_NA	3.19%	14	33.97%	False	chr11:5,249,968-5,249,972	chr11:5,254,892-5,254,896	HBG1/2 c.-111 to -115
166_-7_NA	3.10%	14	37.07%	False	chr11:5,249,968-5,249,974	chr11:5,254,892-5,254,898	HBG1/2 c.-111 to -117

157_-13_NA	2.63%	14	39.70%	True	chr11:5,249,959-5,249,971	chr11:5,254,883-5,254,895	HBG1/2 c.-102 to -114
169_-5_NA	2.07%	14	41.77%	False	chr11:5,249,971-5,249,975	chr11:5,254,895-5,254,899	HBG1/2 c.-114 to -118
167_-5_NA	1.72%	14	43.49%	False	chr11:5,249,969-5,249,973	chr11:5,254,893-5,254,897	HBG1/2 c.-112 to -116
168_-5_NA	1.54%	14	45.03%	False	chr11:5,249,970-5,249,974	chr11:5,254,894-5,254,898	HBG1/2 c.-113 to -117
169_-4_NA	1.43%	14	46.46%	False	chr11:5,249,971-5,249,974	chr11:5,254,895-5,254,898	HBG1/2 c.-114 to -117
166_-8_NA	1.41%	14	47.88%	False	chr11:5,249,968-5,249,975	chr11:5,254,892-5,254,899	HBG1/2 c.-111 to -118
164_-8_NA	1.32%	14	49.20%	False	chr11:5,249,966-5,249,973	chr11:5,254,890-5,254,897	HBG1/2 c.-109 to -116
166_-12_NA	1.28%	14	50.48%	False	chr11:5,249,968-5,249,979	chr11:5,254,892-5,254,903	HBG1/2 c.-111 to -122
166_-13_NA	1.10%	14	51.58%	False	chr11:5,249,968-5,249,980	chr11:5,254,892-5,254,904	HBG1/2 c.-111 to -123
167_-6_NA	0.90%	14	52.47%	False	chr11:5,249,969-5,249,974	chr11:5,254,893-5,254,898	HBG1/2 c.-112 to -117
169_-2_NA	0.83%	14	53.30%	False	chr11:5,249,971-5,249,972	chr11:5,254,895-5,254,896	HBG1/2 c.-114 to -115
168_-6_NA	0.80%	14	54.10%	False	chr11:5,249,970-5,249,975	chr11:5,254,894-5,254,899	HBG1/2 c.-113 to -118
166_-10_NA	0.72%	14	54.82%	False	chr11:5,249,968-5,249,977	chr11:5,254,892-5,254,901	HBG1/2 c.-111 to -120
169_-9_NA	0.69%	14	55.51%	False	chr11:5,249,971-5,249,979	chr11:5,254,895-5,254,903	HBG1/2 c.-114 to -122
168_-8_NA	0.65%	14	56.16%	False	chr11:5,249,970-5,249,977	chr11:5,254,894-5,254,901	HBG1/2 c.-113 to -120
169_-8_NA	0.65%	14	56.80%	False	chr11:5,249,971-5,249,978	chr11:5,254,895-5,254,902	HBG1/2 c.-114 to -121
168_-7_NA	0.59%	14	57.39%	False	chr11:5,249,970-5,249,976	chr11:5,254,894-5,254,900	HBG1/2 c.-113 to -119
165_-7_NA	0.57%	14	57.97%	False	chr11:5,249,967-5,249,973	chr11:5,254,891-5,254,897	HBG1/2 c.-110 to -116
169_-7_NA	0.57%	14	58.54%	False	chr11:5,249,971-5,249,977	chr11:5,254,895-5,254,901	HBG1/2 c.-114 to -120
157_-18_NA	0.55%	14	59.10%	False	chr11:5,249,959-5,249,976	chr11:5,254,883-5,254,900	HBG1/2 c.-102 to -119
168_-10_NA	0.55%	14	59.64%	False	chr11:5,249,970-5,249,979	chr11:5,254,894-5,254,903	HBG1/2 c.-113 to -122
169_-10_NA	0.52%	14	60.16%	False	chr11:5,249,971-5,249,980	chr11:5,254,895-5,254,904	HBG1/2 c.-114 to -123
168_-13_NA	0.52%	14	60.68%	False	chr11:5,249,970-5,249,982	chr11:5,254,894-5,254,906	HBG1/2 c.-113 to -125
168_-9_NA	0.49%	14	61.17%	False	chr11:5,249,970-5,249,978	chr11:5,254,894-5,254,902	HBG1/2 c.-113 to -121
162_-22_NA	0.49%	14	61.65%	True	chr11:5,249,964-5,249,985	chr11:5,254,888-5,254,909	HBG1/2 c.-107 to -128
166_-11_NA	0.48%	14	62.13%	False	chr11:5,249,968-5,249,978	chr11:5,254,892-5,254,902	HBG1/2 c.-111 to -121

169_-6_NA	0.46%	14	62.59%	True	chr11:5,249,971-5,249,976	chr11:5,254,895-5,254,900	HBG1/2 c.-114 to -119
168_-12_NA	0.46%	14	63.06%	False	chr11:5,249,970-5,249,981	chr11:5,254,894-5,254,905	HBG1/2 c.-113 to -124
166_-15_NA	0.43%	14	63.49%	False	chr11:5,249,968-5,249,982	chr11:5,254,892-5,254,906	HBG1/2 c.-111 to -125
167_-16_NA	0.43%	14	63.92%	False	chr11:5,249,969-5,249,984	chr11:5,254,893-5,254,908	HBG1/2 c.-112 to -127
166_-4_NA	0.42%	14	64.35%	False	chr11:5,249,968-5,249,971	chr11:5,254,892-5,254,895	HBG1/2 c.-111 to -114
167_-11_NA	0.42%	14	64.77%	False	chr11:5,249,969-5,249,979	chr11:5,254,893-5,254,903	HBG1/2 c.-112 to -122
162_-7_NA	0.42%	14	65.19%	True	chr11:5,249,964-5,249,970	chr11:5,254,888-5,254,894	HBG1/2 c.-107 to -113
170_-2_NA	0.42%	14	65.60%	False	chr11:5,249,972-5,249,973	chr11:5,254,896-5,254,897	HBG1/2 c.-115 to -116
168_-11_NA	0.40%	14	66.00%	False	chr11:5,249,970-5,249,980	chr11:5,254,894-5,254,904	HBG1/2 c.-113 to -123
167_-1_NA	0.37%	14	66.37%	False	chr11:5,249,969-5,249,969	chr11:5,254,893-5,254,893	HBG1/2 c.-112 to -112
166_-9_NA	0.33%	14	66.71%	False	chr11:5,249,968-5,249,976	chr11:5,254,892-5,254,900	HBG1/2 c.-111 to -119
167_-10_NA	0.32%	14	67.03%	False	chr11:5,249,969-5,249,978	chr11:5,254,893-5,254,902	HBG1/2 c.-112 to -121
166_-14_NA	0.32%	14	67.35%	False	chr11:5,249,968-5,249,981	chr11:5,254,892-5,254,905	HBG1/2 c.-111 to -124
167_-9_NA	0.30%	14	67.65%	False	chr11:5,249,969-5,249,977	chr11:5,254,893-5,254,901	HBG1/2 c.-112 to -120
168_-2_NA	0.29%	14	67.94%	False	chr11:5,249,970-5,249,971	chr11:5,254,894-5,254,895	HBG1/2 c.-113 to -114
169_-12_NA	0.28%	14	68.22%	False	chr11:5,249,971-5,249,982	chr11:5,254,895-5,254,906	HBG1/2 c.-114 to -125
169_-13_NA	0.27%	14	68.49%	False	chr11:5,249,971-5,249,983	chr11:5,254,895-5,254,907	HBG1/2 c.-114 to -126
169_-11_NA	0.26%	14	68.74%	True	chr11:5,249,971-5,249,981	chr11:5,254,895-5,254,905	HBG1/2 c.-114 to -124
165_-9_NA	0.24%	14	68.99%	False	chr11:5,249,967-5,249,975	chr11:5,254,891-5,254,899	HBG1/2 c.-110 to -118
167_-3_NA	0.23%	14	69.22%	False	chr11:5,249,969-5,249,971	chr11:5,254,893-5,254,895	HBG1/2 c.-112 to -114
163_-9_NA	0.23%	14	69.45%	False	chr11:5,249,965-5,249,973	chr11:5,254,889-5,254,897	HBG1/2 c.-108 to -116
167_-14_NA	0.23%	14	69.68%	False	chr11:5,249,969-5,249,982	chr11:5,254,893-5,254,906	HBG1/2 c.-112 to -125
152_-24_NA	0.22%	14	69.90%	True	chr11:5,249,954-5,249,977	chr11:5,254,878-5,254,901	HBG1/2 c.-97 to -120
157_-19_NA	0.22%	13	70.12%	False	chr11:5,249,959-5,249,977	chr11:5,254,883-5,254,901	HBG1/2 c.-102 to -120
163_-12_NA	0.20%	14	70.32%	False	chr11:5,249,965-5,249,976	chr11:5,254,889-5,254,900	HBG1/2 c.-108 to -119
166_-16_NA	0.20%	14	70.52%	False	chr11:5,249,968-5,249,983	chr11:5,254,892-5,254,907	HBG1/2 c.-111 to -126

162_-10_NA	0.19%	13	70.72%	False	chr11:5,249,964-5,249,973	chr11:5,254,888-5,254,897	HBG1/2 c.-107 to -116
158_-14_NA	0.19%	13	70.91%	False	chr11:5,249,960-5,249,973	chr11:5,254,884-5,254,897	HBG1/2 c.-103 to -116
159_-28_NA	0.19%	14	71.10%	True	chr11:5,249,961-5,249,988	chr11:5,254,885-5,254,912	HBG1/2 c.-104 to -131
162_-8_NA	0.19%	14	71.28%	False	chr11:5,249,964-5,249,971	chr11:5,254,888-5,254,895	HBG1/2 c.-107 to -114
165_-2_NA	0.19%	14	71.47%	False	chr11:5,249,967-5,249,968	chr11:5,254,891-5,254,892	HBG1/2 c.-110 to -111
152_-33_NA	0.19%	14	71.65%	True	chr11:5,249,954-5,249,986	chr11:5,254,878-5,254,910	HBG1/2 c.-97 to -129
168_-14_NA	0.18%	14	71.84%	False	chr11:5,249,970-5,249,983	chr11:5,254,894-5,254,907	HBG1/2 c.-113 to -126
161_-12_NA	0.18%	13	72.02%	False	chr11:5,249,963-5,249,974	chr11:5,254,887-5,254,898	HBG1/2 c.-106 to -117
170_-4_NA	0.18%	14	72.20%	False	chr11:5,249,972-5,249,975	chr11:5,254,896-5,254,899	HBG1/2 c.-115 to -118
162_-9_NA	0.18%	13	72.37%	False	chr11:5,249,964-5,249,972	chr11:5,254,888-5,254,896	HBG1/2 c.-107 to -115
164_-23_NA	0.18%	13	72.55%	False	chr11:5,249,966-5,249,988	chr11:5,254,890-5,254,912	HBG1/2 c.-109 to -131
164_-10_NA	0.18%	14	72.73%	False	chr11:5,249,966-5,249,975	chr11:5,254,890-5,254,899	HBG1/2 c.-109 to -118
166_-27_NA	0.17%	14	72.90%	True	chr11:5,249,968-5,249,994	chr11:5,254,892-5,254,918	HBG1/2 c.-111 to -137
160_-12_NA	0.16%	14	73.06%	False	chr11:5,249,962-5,249,973	chr11:5,254,886-5,254,897	HBG1/2 c.-105 to -116
165_-12_NA	0.16%	14	73.22%	False	chr11:5,249,967-5,249,978	chr11:5,254,891-5,254,902	HBG1/2 c.-110 to -121
170_-5_NA	0.16%	14	73.38%	False	chr11:5,249,972-5,249,976	chr11:5,254,896-5,254,900	HBG1/2 c.-115 to -119
165_-11_NA	0.16%	14	73.54%	False	chr11:5,249,967-5,249,977	chr11:5,254,891-5,254,901	HBG1/2 c.-110 to -120
137_-27_NA	0.15%	14	73.69%	True	chr11:5,249,939-5,249,965	chr11:5,254,863-5,254,889	HBG1/2 c.-82 to -108
165_-3_NA	0.15%	14	73.84%	False	chr11:5,249,967-5,249,969	chr11:5,254,891-5,254,893	HBG1/2 c.-110 to -112
169_-14_NA	0.15%	14	73.99%	False	chr11:5,249,971-5,249,984	chr11:5,254,895-5,254,908	HBG1/2 c.-114 to -127
172_-5_NA	0.15%	14	74.14%	True	chr11:5,249,974-5,249,978	chr11:5,254,898-5,254,902	HBG1/2 c.-117 to -121
164_-9_NA	0.15%	14	74.29%	False	chr11:5,249,966-5,249,974	chr11:5,254,890-5,254,898	HBG1/2 c.-109 to -117
152_-29_NA	0.14%	13	74.44%	True	chr11:5,249,954-5,249,982	chr11:5,254,878-5,254,906	HBG1/2 c.-97 to -125
166_-22_NA	0.14%	14	74.58%	False	chr11:5,249,968-5,249,989	chr11:5,254,892-5,254,913	HBG1/2 c.-111 to -132
169_-1_NA	0.14%	14	74.72%	False	chr11:5,249,971-5,249,971	chr11:5,254,895-5,254,895	HBG1/2 c.-114 to -114
161_-17_NA	0.13%	14	74.85%	False	chr11:5,249,963-5,249,979	chr11:5,254,887-5,254,903	HBG1/2 c.-106 to -122

162_-14_NA	0.13%	14	74.98%	False	chr11:5,249,964-5,249,977	chr11:5,254,888-5,254,901	HBG1/2 c.-107 to -120
153_-19_NA	0.13%	14	75.11%	False	chr11:5,249,955-5,249,973	chr11:5,254,879-5,254,897	HBG1/2 c.-98 to -116
151_-20_NA	0.13%	14	75.23%	False	chr11:5,249,953-5,249,972	chr11:5,254,877-5,254,896	HBG1/2 c.-96 to -115
161_-11_NA	0.13%	13	75.36%	False	chr11:5,249,963-5,249,973	chr11:5,254,887-5,254,897	HBG1/2 c.-106 to -116
165_-5_NA	0.13%	14	75.49%	False	chr11:5,249,967-5,249,971	chr11:5,254,891-5,254,895	HBG1/2 c.-110 to -114
156_-15_NA	0.12%	14	75.61%	False	chr11:5,249,958-5,249,972	chr11:5,254,882-5,254,896	HBG1/2 c.-101 to -115
167_-2_NA	0.12%	14	75.73%	False	chr11:5,249,969-5,249,970	chr11:5,254,893-5,254,894	HBG1/2 c.-112 to -113
167_-18_NA	0.12%	14	75.86%	False	chr11:5,249,969-5,249,986	chr11:5,254,893-5,254,910	HBG1/2 c.-112 to -129
145_-32_NA	0.12%	13	75.98%	True	chr11:5,249,947-5,249,978	chr11:5,254,871-5,254,902	HBG1/2 c.-90 to -121
166_-21_NA	0.12%	14	76.10%	False	chr11:5,249,968-5,249,988	chr11:5,254,892-5,254,912	HBG1/2 c.-111 to -131
169_-18_NA	0.12%	14	76.22%	False	chr11:5,249,971-5,249,988	chr11:5,254,895-5,254,912	HBG1/2 c.-114 to -131
153_-38_NA	0.12%	13	76.34%	True	chr11:5,249,955-5,249,992	chr11:5,254,879-5,254,916	HBG1/2 c.-98 to -135
163_-17_NA	0.12%	14	76.46%	False	chr11:5,249,965-5,249,981	chr11:5,254,889-5,254,905	HBG1/2 c.-108 to -124
165_-14_NA	0.12%	14	76.58%	False	chr11:5,249,967-5,249,980	chr11:5,254,891-5,254,904	HBG1/2 c.-110 to -123
165_-15_NA	0.12%	14	76.70%	False	chr11:5,249,967-5,249,981	chr11:5,254,891-5,254,905	HBG1/2 c.-110 to -124
160_-14_NA	0.12%	13	76.81%	False	chr11:5,249,962-5,249,975	chr11:5,254,886-5,254,899	HBG1/2 c.-105 to -118
166_-25_NA	0.11%	13	76.93%	False	chr11:5,249,968-5,249,992	chr11:5,254,892-5,254,916	HBG1/2 c.-111 to -135
170_-9_NA	0.11%	14	77.04%	False	chr11:5,249,972-5,249,980	chr11:5,254,896-5,254,904	HBG1/2 c.-115 to -123
152_-18_NA	0.11%	14	77.15%	False	chr11:5,249,954-5,249,971	chr11:5,254,878-5,254,895	HBG1/2 c.-97 to -114
168_-18_NA	0.11%	14	77.26%	False	chr11:5,249,970-5,249,987	chr11:5,254,894-5,254,911	HBG1/2 c.-113 to -130
163_-10_NA	0.11%	14	77.38%	False	chr11:5,249,965-5,249,974	chr11:5,254,889-5,254,898	HBG1/2 c.-108 to -117
160_-11_NA	0.11%	14	77.49%	False	chr11:5,249,962-5,249,972	chr11:5,254,886-5,254,896	HBG1/2 c.-105 to -115
150_-21_NA	0.11%	13	77.59%	False	chr11:5,249,952-5,249,972	chr11:5,254,876-5,254,896	HBG1/2 c.-95 to -115
157_-23_NA	0.11%	13	77.70%	False	chr11:5,249,959-5,249,981	chr11:5,254,883-5,254,905	HBG1/2 c.-102 to -124
155_-16_NA	0.11%	12	77.81%	False	chr11:5,249,957-5,249,972	chr11:5,254,881-5,254,896	HBG1/2 c.-100 to -115
167_-17_NA	0.11%	12	77.92%	False	chr11:5,249,969-5,249,985	chr11:5,254,893-5,254,909	HBG1/2 c.-112 to -128

164_-28_NA	0.11%	13	78.02%	False	chr11:5,249,966-5,249,993	chr11:5,254,890-5,254,917	HBG1/2 c.-109 to -136
164_-27_NA	0.10%	14	78.12%	False	chr11:5,249,966-5,249,992	chr11:5,254,890-5,254,916	HBG1/2 c.-109 to -135
156_-38_NA	0.10%	13	78.22%	True	chr11:5,249,958-5,249,995	chr11:5,254,882-5,254,919	HBG1/2 c.-101 to -138
154_-18_NA	0.10%	12	78.32%	False	chr11:5,249,956-5,249,973	chr11:5,254,880-5,254,897	HBG1/2 c.-99 to -116
156_-27_NA	0.10%	13	78.42%	True	chr11:5,249,958-5,249,984	chr11:5,254,882-5,254,908	HBG1/2 c.-101 to -127
157_-24_NA	0.10%	11	78.52%	False	chr11:5,249,959-5,249,982	chr11:5,254,883-5,254,906	HBG1/2 c.-102 to -125
168_-19_NA	0.10%	14	78.62%	False	chr11:5,249,970-5,249,988	chr11:5,254,894-5,254,912	HBG1/2 c.-113 to -131
165_-10_NA	0.10%	14	78.72%	False	chr11:5,249,967-5,249,976	chr11:5,254,891-5,254,900	HBG1/2 c.-110 to -119
171_-1_NA	0.10%	14	78.82%	False	chr11:5,249,973-5,249,973	chr11:5,254,897-5,254,897	HBG1/2 c.-116 to -116
165_-16_NA	0.10%	13	78.92%	False	chr11:5,249,967-5,249,982	chr11:5,254,891-5,254,906	HBG1/2 c.-110 to -125
164_-14_NA	0.10%	12	79.02%	False	chr11:5,249,966-5,249,979	chr11:5,254,890-5,254,903	HBG1/2 c.-109 to -122
167_-4_NA;175_-1_NA	0.10%	14	79.11%	False	chr11:5,249,969-5,249,972;chr11:5,249,977-5,249,977	chr11:5,254,893-5,254,896;chr11:5,254,901-5,254,901	HBG1/2 c.-112 to -115;HBG1/2 c.-120 to -120
160_-19_NA	0.10%	14	79.21%	False	chr11:5,249,962-5,249,980	chr11:5,254,886-5,254,904	HBG1/2 c.-105 to -123
166_-17_NA	0.10%	13	79.30%	False	chr11:5,249,968-5,249,984	chr11:5,254,892-5,254,908	HBG1/2 c.-111 to -127
168_-15_NA	0.10%	14	79.40%	True	chr11:5,249,970-5,249,984	chr11:5,254,894-5,254,908	HBG1/2 c.-113 to -127
162_-33_NA	0.09%	12	79.49%	True	chr11:5,249,964-5,249,996	chr11:5,254,888-5,254,920	HBG1/2 c.-107 to -139
168_-25_NA	0.09%	13	79.58%	False	chr11:5,249,970-5,249,994	chr11:5,254,894-5,254,918	HBG1/2 c.-113 to -137
166_-20_NA	0.09%	10	79.67%	False	chr11:5,249,968-5,249,987	chr11:5,254,892-5,254,911	HBG1/2 c.-111 to -130
170_-3_NA	0.09%	13	79.76%	False	chr11:5,249,972-5,249,974	chr11:5,254,896-5,254,898	HBG1/2 c.-115 to -117
157_-28_NA	0.09%	11	79.84%	False	chr11:5,249,959-5,249,986	chr11:5,254,883-5,254,910	HBG1/2 c.-102 to -129
166_-26_NA	0.09%	13	79.93%	False	chr11:5,249,968-5,249,993	chr11:5,254,892-5,254,917	HBG1/2 c.-111 to -136
160_-259_NA	0.08%	9	80.01%	True	chr11:5,249,962-5,250,220	chr11:5,254,886-5,255,144	HBG1/2 c.-105 to -363
160_-16_NA	0.08%	12	80.09%	False	chr11:5,249,962-5,249,977	chr11:5,254,886-5,254,901	HBG1/2 c.-105 to -120
159_-23_NA	0.08%	13	80.17%	False	chr11:5,249,961-5,249,983	chr11:5,254,885-5,254,907	HBG1/2 c.-104 to -126

167_-25_NA	0.08%	13	80.25%	False	chr11:5,249,969-5,249,993	chr11:5,254,893-5,254,917	HBG1/2 c.-112 to -136
165_-6_NA;175_-1_NA	0.08%	14	80.33%	False	chr11:5,249,967-5,249,972;chr11:5,249,977-5,249,977	chr11:5,254,891-5,254,896;chr11:5,254,901-5,254,901	HBG1/2 c.-110 to -115;HBG1/2 c.-120 to -120
157_-14_NA	0.08%	13	80.41%	False	chr11:5,249,959-5,249,972	chr11:5,254,883-5,254,896	HBG1/2 c.-102 to -115
154_-28_NA	0.08%	13	80.49%	True	chr11:5,249,956-5,249,983	chr11:5,254,880-5,254,907	HBG1/2 c.-99 to -126
168_-23_NA	0.08%	12	80.56%	False	chr11:5,249,970-5,249,992	chr11:5,254,894-5,254,916	HBG1/2 c.-113 to -135
169_-23_NA	0.08%	14	80.64%	False	chr11:5,249,971-5,249,993	chr11:5,254,895-5,254,917	HBG1/2 c.-114 to -136
167_-4_NA;174_-2_NA	0.07%	12	80.71%	False	chr11:5,249,969-5,249,972;chr11:5,249,976-5,249,977	chr11:5,254,893-5,254,896;chr11:5,254,900-5,254,901	HBG1/2 c.-112 to -115;HBG1/2 c.-119 to -120
167_-36_NA	0.07%	11	80.79%	True	chr11:5,249,969-5,250,004	chr11:5,254,893-5,254,928	HBG1/2 c.-112 to -147
165_-29_NA	0.07%	12	80.86%	False	chr11:5,249,967-5,249,995	chr11:5,254,891-5,254,919	HBG1/2 c.-110 to -138
163_-6_NA	0.07%	13	80.93%	False	chr11:5,249,965-5,249,970	chr11:5,254,889-5,254,894	HBG1/2 c.-108 to -113
159_-33_NA	0.07%	13	81.01%	True	chr11:5,249,961-5,249,993	chr11:5,254,885-5,254,917	HBG1/2 c.-104 to -136
162_-11_NA	0.07%	13	81.08%	False	chr11:5,249,964-5,249,974	chr11:5,254,888-5,254,898	HBG1/2 c.-107 to -117
167_-20_NA	0.07%	10	81.15%	False	chr11:5,249,969-5,249,988	chr11:5,254,893-5,254,912	HBG1/2 c.-112 to -131
169_-24_NA	0.07%	13	81.21%	False	chr11:5,249,971-5,249,994	chr11:5,254,895-5,254,918	HBG1/2 c.-114 to -137
167_-24_NA	0.07%	11	81.28%	False	chr11:5,249,969-5,249,992	chr11:5,254,893-5,254,916	HBG1/2 c.-112 to -135
164_-15_NA	0.07%	13	81.35%	False	chr11:5,249,966-5,249,980	chr11:5,254,890-5,254,904	HBG1/2 c.-109 to -123
166_-19_NA	0.07%	12	81.42%	False	chr11:5,249,968-5,249,986	chr11:5,254,892-5,254,910	HBG1/2 c.-111 to -129
166_-1_NA	0.07%	13	81.49%	False	chr11:5,249,968-5,249,968	chr11:5,254,892-5,254,892	HBG1/2 c.-111 to -111
169_-17_NA	0.07%	12	81.55%	False	chr11:5,249,971-5,249,987	chr11:5,254,895-5,254,911	HBG1/2 c.-114 to -130
154_-17_NA	0.07%	11	81.62%	False	chr11:5,249,956-5,249,972	chr11:5,254,880-5,254,896	HBG1/2 c.-99 to -115
164_-11_NA	0.07%	12	81.68%	False	chr11:5,249,966-5,249,976	chr11:5,254,890-5,254,900	HBG1/2 c.-109 to -119
166_-5_NA;175_-1_NA	0.07%	12	81.75%	False	chr11:5,249,968-5,249,972;chr11:5,249,977-5,249,977	chr11:5,254,892-5,254,896;chr11:5,254,901-5,254,901	HBG1/2 c.-111 to -115;HBG1/2 c.-120 to -120

170_-10_NA	0.07%	14	81.81%	False	chr11:5,249,972-5,249,981	chr11:5,254,896-5,254,905	HBG1/2 c.-115 to -124
152_-19_NA	0.06%	11	81.88%	False	chr11:5,249,954-5,249,972	chr11:5,254,878-5,254,896	HBG1/2 c.-97 to -115
166_-23_NA	0.06%	12	81.94%	False	chr11:5,249,968-5,249,990	chr11:5,254,892-5,254,914	HBG1/2 c.-111 to -133
162_-19_NA	0.06%	14	82.01%	False	chr11:5,249,964-5,249,982	chr11:5,254,888-5,254,906	HBG1/2 c.-107 to -125
167_-38_NA	0.06%	12	82.07%	False	chr11:5,249,969-5,250,006	chr11:5,254,893-5,254,930	HBG1/2 c.-112 to -149
161_-10_NA	0.06%	13	82.14%	False	chr11:5,249,963-5,249,972	chr11:5,254,887-5,254,896	HBG1/2 c.-106 to -115
158_-15_NA	0.06%	12	82.20%	False	chr11:5,249,960-5,249,974	chr11:5,254,884-5,254,898	HBG1/2 c.-103 to -117
163_-11_NA	0.06%	12	82.26%	False	chr11:5,249,965-5,249,975	chr11:5,254,889-5,254,899	HBG1/2 c.-108 to -118
168_-17_NA	0.06%	11	82.33%	False	chr11:5,249,970-5,249,986	chr11:5,254,894-5,254,910	HBG1/2 c.-113 to -129
151_-43_NA	0.06%	12	82.39%	True	chr11:5,249,953-5,249,995	chr11:5,254,877-5,254,919	HBG1/2 c.-96 to -138
158_-16_NA	0.06%	12	82.45%	False	chr11:5,249,960-5,249,975	chr11:5,254,884-5,254,899	HBG1/2 c.-103 to -118
169_-19_NA	0.06%	13	82.51%	False	chr11:5,249,971-5,249,989	chr11:5,254,895-5,254,913	HBG1/2 c.-114 to -132
170_-17_NA	0.06%	12	82.57%	False	chr11:5,249,972-5,249,988	chr11:5,254,896-5,254,912	HBG1/2 c.-115 to -131
158_-21_NA	0.06%	12	82.63%	False	chr11:5,249,960-5,249,980	chr11:5,254,884-5,254,904	HBG1/2 c.-103 to -123
152_-28_NA	0.06%	13	82.69%	False	chr11:5,249,954-5,249,981	chr11:5,254,878-5,254,905	HBG1/2 c.-97 to -124
150_-33_NA	0.06%	13	82.74%	False	chr11:5,249,952-5,249,984	chr11:5,254,876-5,254,908	HBG1/2 c.-95 to -127
170_-13_NA	0.05%	13	82.80%	False	chr11:5,249,972-5,249,984	chr11:5,254,896-5,254,908	HBG1/2 c.-115 to -127
165_-19_NA	0.05%	13	82.85%	False	chr11:5,249,967-5,249,985	chr11:5,254,891-5,254,909	HBG1/2 c.-110 to -128
158_-17_NA	0.05%	11	82.91%	False	chr11:5,249,960-5,249,976	chr11:5,254,884-5,254,900	HBG1/2 c.-103 to -119
159_-11_NA	0.05%	7	82.96%	False	chr11:5,249,961-5,249,971	chr11:5,254,885-5,254,895	HBG1/2 c.-104 to -114
158_-20_NA	0.05%	12	83.01%	False	chr11:5,249,960-5,249,979	chr11:5,254,884-5,254,903	HBG1/2 c.-103 to -122
155_-12_NA	0.05%	10	83.07%	True	chr11:5,249,957-5,249,968	chr11:5,254,881-5,254,892	HBG1/2 c.-100 to -111
168_-3_NA;175_-1_NA	0.05%	12	83.12%	False	chr11:5,249,970-5,249,972;chr11:5,249,977-5,249,977	chr11:5,254,894-5,254,896;chr11:5,254,901-5,254,901	HBG1/2 c.-113 to -115;HBG1/2 c.-120 to -120
169_-16_NA	0.05%	14	83.17%	False	chr11:5,249,971-5,249,986	chr11:5,254,895-5,254,910	HBG1/2 c.-114 to -129

157_-40_NA	0.05%	13	83.22%	True	chr11:5,249,959-5,249,998	chr11:5,254,883-5,254,922	HBG1/2 c.-102 to -141
168_-24_NA	0.05%	11	83.27%	False	chr11:5,249,970-5,249,993	chr11:5,254,894-5,254,917	HBG1/2 c.-113 to -136
165_-6_NA;174_-2_NA	0.05%	12	83.33%	False	chr11:5,249,967-5,249,972;chr11:5,249,976-5,249,977	chr11:5,254,891-5,254,896;chr11:5,254,900-5,254,901	HBG1/2 c.-110 to -115;HBG1/2 c.-119 to -120
158_-36_NA	0.05%	11	83.38%	False	chr11:5,249,960-5,249,995	chr11:5,254,884-5,254,919	HBG1/2 c.-103 to -138
169_-22_NA	0.05%	12	83.43%	False	chr11:5,249,971-5,249,992	chr11:5,254,895-5,254,916	HBG1/2 c.-114 to -135
157_-16_NA	0.05%	14	83.48%	False	chr11:5,249,959-5,249,974	chr11:5,254,883-5,254,898	HBG1/2 c.-102 to -117
165_-291_NA	0.05%	7	83.53%	True	chr11:5,249,967-5,250,257	chr11:5,254,891-5,255,181	HBG1/2 c.-110 to -400
173_-1_NA	0.05%	12	83.58%	False	chr11:5,249,975-5,249,975	chr11:5,254,899-5,254,899	HBG1/2 c.-118 to -118
167_-26_NA	0.05%	11	83.63%	False	chr11:5,249,969-5,249,994	chr11:5,254,893-5,254,918	HBG1/2 c.-112 to -137
148_-22_NA	0.05%	10	83.68%	True	chr11:5,249,950-5,249,971	chr11:5,254,874-5,254,895	HBG1/2 c.-93 to -114
152_-23_NA	0.05%	10	83.72%	False	chr11:5,249,954-5,249,976	chr11:5,254,878-5,254,900	HBG1/2 c.-97 to -119
165_-4_NA	0.05%	11	83.77%	False	chr11:5,249,967-5,249,970	chr11:5,254,891-5,254,894	HBG1/2 c.-110 to -113
165_-17_NA	0.05%	12	83.82%	False	chr11:5,249,967-5,249,983	chr11:5,254,891-5,254,907	HBG1/2 c.-110 to -126
165_-298_NA	0.05%	6	83.87%	False	chr11:5,249,967-5,250,264	chr11:5,254,891-5,255,188	HBG1/2 c.-110 to -407
154_-23_NA	0.05%	11	83.92%	False	chr11:5,249,956-5,249,978	chr11:5,254,880-5,254,902	HBG1/2 c.-99 to -121
168_-3_NA;174_-2_NA	0.05%	11	83.96%	False	chr11:5,249,970-5,249,972;chr11:5,249,976-5,249,977	chr11:5,254,894-5,254,896;chr11:5,254,900-5,254,901	HBG1/2 c.-113 to -115;HBG1/2 c.-119 to -120
165_-5_NA;174_-12_NA	0.05%	1	84.01%	False	chr11:5,249,967-5,249,971;chr11:5,249,976-5,249,987	chr11:5,254,891-5,254,895;chr11:5,254,900-5,254,911	HBG1/2 c.-110 to -114;HBG1/2 c.-119 to -130
165_-271_NA	0.05%	6	84.06%	True	chr11:5,249,967-5,250,237	chr11:5,254,891-5,255,161	HBG1/2 c.-110 to -380
164_-4_NA;174_-2_NA	0.05%	4	84.11%	False	chr11:5,249,966-5,249,969;chr11:5,249,976-5,249,977	chr11:5,254,890-5,254,893;chr11:5,254,900-5,254,901	HBG1/2 c.-109 to -112;HBG1/2 c.-119 to -120
160_-8_NA;170_-251_NA	0.05%	7	84.15%	False	chr11:5,249,962-5,249,969;chr11:5,249,972-5,250,222	chr11:5,254,886-5,254,893;chr11:5,254,896-5,255,146	HBG1/2 c.-105 to -112;HBG1/2

							c.-115 to -365
167_-31_NA	0.05%	11	84.20%	False	chr11:5,249,969-5,249,999	chr11:5,254,893-5,254,923	HBG1/2 c.-112 to -142
159_-24_NA	0.05%	9	84.24%	False	chr11:5,249,961-5,249,984	chr11:5,254,885-5,254,908	HBG1/2 c.-104 to -127
151_-32_NA	0.05%	11	84.29%	True	chr11:5,249,953-5,249,984	chr11:5,254,877-5,254,908	HBG1/2 c.-96 to -127
170_-8_NA	0.04%	13	84.34%	False	chr11:5,249,972-5,249,979	chr11:5,254,896-5,254,903	HBG1/2 c.-115 to -122
151_-29_NA	0.04%	12	84.38%	False	chr11:5,249,953-5,249,981	chr11:5,254,877-5,254,905	HBG1/2 c.-96 to -124
151_-22_NA	0.04%	12	84.42%	False	chr11:5,249,953-5,249,974	chr11:5,254,877-5,254,898	HBG1/2 c.-96 to -117
153_-21_NA	0.04%	11	84.47%	False	chr11:5,249,955-5,249,975	chr11:5,254,879-5,254,899	HBG1/2 c.-98 to -118
167_-19_NA	0.04%	10	84.51%	False	chr11:5,249,969-5,249,987	chr11:5,254,893-5,254,911	HBG1/2 c.-112 to -130
168_-22_NA	0.04%	13	84.56%	False	chr11:5,249,970-5,249,991	chr11:5,254,894-5,254,915	HBG1/2 c.-113 to -134
164_-26_NA	0.04%	10	84.60%	False	chr11:5,249,966-5,249,991	chr11:5,254,890-5,254,915	HBG1/2 c.-109 to -134
153_-34_NA	0.04%	14	84.64%	False	chr11:5,249,955-5,249,988	chr11:5,254,879-5,254,912	HBG1/2 c.-98 to -131
162_-15_NA	0.04%	11	84.68%	False	chr11:5,249,964-5,249,978	chr11:5,254,888-5,254,902	HBG1/2 c.-107 to -121
159_-27_NA	0.04%	11	84.72%	False	chr11:5,249,961-5,249,987	chr11:5,254,885-5,254,911	HBG1/2 c.-104 to -130
153_-37_NA	0.04%	9	84.76%	True	chr11:5,249,955-5,249,991	chr11:5,254,879-5,254,915	HBG1/2 c.-98 to -134
175_-1_NA	0.04%	12	84.80%	False	chr11:5,249,977-5,249,977	chr11:5,254,901-5,254,901	HBG1/2 c.-120 to -120
160_-10_NA	0.04%	11	84.84%	False	chr11:5,249,962-5,249,971	chr11:5,254,886-5,254,895	HBG1/2 c.-105 to -114
155_-19_NA	0.04%	11	84.88%	False	chr11:5,249,957-5,249,975	chr11:5,254,881-5,254,899	HBG1/2 c.-100 to -118
164_-5_NA	0.04%	9	84.92%	False	chr11:5,249,966-5,249,970	chr11:5,254,890-5,254,894	HBG1/2 c.-109 to -113
163_-15_NA	0.04%	12	84.96%	False	chr11:5,249,965-5,249,979	chr11:5,254,889-5,254,903	HBG1/2 c.-108 to -122
161_-22_NA	0.04%	9	85.00%	False	chr11:5,249,963-5,249,984	chr11:5,254,887-5,254,908	HBG1/2 c.-106 to -127
155_-17_NA	0.04%	12	85.04%	False	chr11:5,249,957-5,249,973	chr11:5,254,881-5,254,897	HBG1/2 c.-100 to -116
165_-6_NA;176_-2_NA	0.04%	11	85.07%	False	chr11:5,249,967-5,249,972;chr11:5,249,978-5,249,979	chr11:5,254,891-5,254,896;chr11:5,254,902-5,254,903	HBG1/2 c.-110 to -115;HBG1/2 c.-121 to -122
150_-32_NA	0.04%	11	85.11%	False	chr11:5,249,952-5,249,983	chr11:5,254,876-5,254,907	HBG1/2 c.-95 to -126

169_-25_NA	0.04%	13	85.15%	False	chr11:5,249,971-5,249,995	chr11:5,254,895-5,254,919	HBG1/2 c.-114 to -138
148_-33_NA	0.04%	9	85.19%	False	chr11:5,249,950-5,249,982	chr11:5,254,874-5,254,906	HBG1/2 c.-93 to -125
166_-36_NA	0.04%	10	85.22%	False	chr11:5,249,968-5,250,003	chr11:5,254,892-5,254,927	HBG1/2 c.-111 to -146
164_-19_NA	0.04%	10	85.26%	False	chr11:5,249,966-5,249,984	chr11:5,254,890-5,254,908	HBG1/2 c.-109 to -127
171_-3_NA	0.04%	12	85.29%	False	chr11:5,249,973-5,249,975	chr11:5,254,897-5,254,899	HBG1/2 c.-116 to -118
169_-21_NA	0.03%	10	85.33%	False	chr11:5,249,971-5,249,991	chr11:5,254,895-5,254,915	HBG1/2 c.-114 to -134
152_-22_NA	0.03%	10	85.36%	False	chr11:5,249,954-5,249,975	chr11:5,254,878-5,254,899	HBG1/2 c.-97 to -118
165_-6_NA;175_-2_NA	0.03%	10	85.40%	False	chr11:5,249,967-5,249,972;chr11:5,249,977-5,249,978	chr11:5,254,891-5,254,896;chr11:5,254,901-5,254,902	HBG1/2 c.-110 to -115;HBG1/2 c.-120 to -121
162_-311_NA	0.03%	1	85.43%	True	chr11:5,249,964-5,250,274	chr11:5,254,888-5,255,198	HBG1/2 c.-107 to -417
166_-5_NA;174_-2_NA	0.03%	12	85.47%	False	chr11:5,249,968-5,249,972;chr11:5,249,976-5,249,977	chr11:5,254,892-5,254,896;chr11:5,254,900-5,254,901	HBG1/2 c.-111 to -115;HBG1/2 c.-119 to -120
168_-3_NA;174_-4_NA	0.03%	10	85.50%	False	chr11:5,249,970-5,249,972;chr11:5,249,976-5,249,979	chr11:5,254,894-5,254,896;chr11:5,254,900-5,254,903	HBG1/2 c.-113 to -115;HBG1/2 c.-119 to -122
161_-15_NA	0.03%	12	85.54%	False	chr11:5,249,963-5,249,977	chr11:5,254,887-5,254,901	HBG1/2 c.-106 to -120
167_-1_NA;172_-5_NA	0.03%	9	85.57%	False;True	chr11:5,249,969-5,249,969;chr11:5,249,974-5,249,978	chr11:5,254,893-5,254,893;chr11:5,254,898-5,254,902	HBG1/2 c.-112 to -112;HBG1/2 c.-117 to -121
157_-29_NA	0.03%	9	85.60%	False	chr11:5,249,959-5,249,987	chr11:5,254,883-5,254,911	HBG1/2 c.-102 to -130
163_-313_NA	0.03%	8	85.63%	True	chr11:5,249,965-5,250,277	chr11:5,254,889-5,255,201	HBG1/2 c.-108 to -420
152_-25_NA	0.03%	10	85.67%	False	chr11:5,249,954-5,249,978	chr11:5,254,878-5,254,902	HBG1/2 c.-97 to -121
159_-15_NA	0.03%	10	85.70%	False	chr11:5,249,961-5,249,975	chr11:5,254,885-5,254,899	HBG1/2 c.-104 to -118
156_-20_NA	0.03%	8	85.73%	False	chr11:5,249,958-5,249,977	chr11:5,254,882-5,254,901	HBG1/2 c.-101 to -120
157_-21_NA	0.03%	10	85.76%	False	chr11:5,249,959-5,249,979	chr11:5,254,883-5,254,903	HBG1/2 c.-102 to -122
168_-30_NA	0.03%	11	85.79%	True	chr11:5,249,970-5,249,999	chr11:5,254,894-5,254,923	HBG1/2 c.-113 to -142

161_-40_NA	0.03%	8	85.82%	True	chr11:5,249,963-5,250,002	chr11:5,254,887-5,254,926	HBG1/2 c.-106 to -145
172_-18_NA	0.03%	10	85.85%	False	chr11:5,249,974-5,249,991	chr11:5,254,898-5,254,915	HBG1/2 c.-117 to -134
167_-11_NA;180_-2_NA	0.03%	4	85.88%	False	chr11:5,249,969-5,249,979;chr11:5,249,982-5,249,983	chr11:5,254,893-5,254,903;chr11:5,254,906-5,254,907	HBG1/2 c.-112 to -122;HBG1/2 c.-125 to -126
167_-4_NA;175_-2_NA	0.03%	9	85.91%	False	chr11:5,249,969-5,249,972;chr11:5,249,977-5,249,978	chr11:5,254,893-5,254,896;chr11:5,254,901-5,254,902	HBG1/2 c.-112 to -115;HBG1/2 c.-120 to -121
166_-5_NA;175_-2_NA	0.03%	6	85.94%	False	chr11:5,249,968-5,249,972;chr11:5,249,977-5,249,978	chr11:5,254,892-5,254,896;chr11:5,254,901-5,254,902	HBG1/2 c.-111 to -115;HBG1/2 c.-120 to -121
161_-8_NA	0.03%	10	85.97%	False	chr11:5,249,963-5,249,970	chr11:5,254,887-5,254,894	HBG1/2 c.-106 to -113
162_-23_NA	0.03%	9	86.00%	False	chr11:5,249,964-5,249,986	chr11:5,254,888-5,254,910	HBG1/2 c.-107 to -129
156_-30_NA	0.03%	8	86.03%	False	chr11:5,249,958-5,249,987	chr11:5,254,882-5,254,911	HBG1/2 c.-101 to -130
46_-338_NA	0.03%	1	86.05%	True	chr11:5,249,848-5,250,185	chr11:5,254,772-5,255,109	HBG1/2 c.9 to -328
150_-41_NA	0.03%	6	86.08%	False	chr11:5,249,952-5,249,992	chr11:5,254,876-5,254,916	HBG1/2 c.-95 to -135
154_-20_NA	0.03%	11	86.11%	False	chr11:5,249,956-5,249,975	chr11:5,254,880-5,254,899	HBG1/2 c.-99 to -118
156_-17_NA	0.03%	12	86.14%	False	chr11:5,249,958-5,249,974	chr11:5,254,882-5,254,898	HBG1/2 c.-101 to -117
158_-25_NA	0.03%	9	86.16%	False	chr11:5,249,960-5,249,984	chr11:5,254,884-5,254,908	HBG1/2 c.-103 to -127
161_-32_NA	0.03%	10	86.19%	False	chr11:5,249,963-5,249,994	chr11:5,254,887-5,254,918	HBG1/2 c.-106 to -137
168_-32_NA	0.03%	10	86.22%	False	chr11:5,249,970-5,250,001	chr11:5,254,894-5,254,925	HBG1/2 c.-113 to -144
151_-30_NA	0.03%	8	86.24%	False	chr11:5,249,953-5,249,982	chr11:5,254,877-5,254,906	HBG1/2 c.-96 to -125
156_-40_NA	0.03%	5	86.27%	False	chr11:5,249,958-5,249,997	chr11:5,254,882-5,254,921	HBG1/2 c.-101 to -140
168_-29_NA	0.03%	8	86.29%	False	chr11:5,249,970-5,249,998	chr11:5,254,894-5,254,922	HBG1/2 c.-113 to -141
165_-38_NA	0.03%	10	86.32%	False	chr11:5,249,967-5,250,004	chr11:5,254,891-5,254,928	HBG1/2 c.-110 to -147
165_-2_NA;170_-14_NA;188_1_A	0.03%	1	86.35%	False	chr11:5,249,967-5,249,968;chr11:5,249,972-5,249,985;chr11:5,249,990-5,249,988	chr11:5,254,891-5,254,892;chr11:5,254,896-5,254,909;chr11:5,254,914-5,254,912	HBG1/2 c.-110 to -111;HBG1/2 c.-115 to -128;HBG1/2 c.-133 to -131

163_-16_NA	0.02%	10	86.37%	False	chr11:5,249,965-5,249,980	chr11:5,254,889-5,254,904	HBG1/2 c.-108 to -123
172_-2_NA	0.02%	9	86.39%	False	chr11:5,249,974-5,249,975	chr11:5,254,898-5,254,899	HBG1/2 c.-117 to -118
174_-2_NA	0.02%	11	86.42%	False	chr11:5,249,976-5,249,977	chr11:5,254,900-5,254,901	HBG1/2 c.-119 to -120
166_-7_NA;176_-2_NA	0.02%	6	86.44%	False	chr11:5,249,968-5,249,974;chr11:5,249,978-5,249,979	chr11:5,254,892-5,254,898;chr11:5,254,902-5,254,903	HBG1/2 c.-111 to -117;HBG1/2 c.-121 to -122
160_-21_NA	0.02%	11	86.47%	False	chr11:5,249,962-5,249,982	chr11:5,254,886-5,254,906	HBG1/2 c.-105 to -125
155_-24_NA	0.02%	10	86.49%	False	chr11:5,249,957-5,249,980	chr11:5,254,881-5,254,904	HBG1/2 c.-100 to -123
164_-20_NA	0.02%	7	86.52%	False	chr11:5,249,966-5,249,985	chr11:5,254,890-5,254,909	HBG1/2 c.-109 to -128
169_-8_NA;180_-5_NA	0.02%	2	86.54%	False	chr11:5,249,971-5,249,978;chr11:5,249,982-5,249,986	chr11:5,254,895-5,254,902;chr11:5,254,906-5,254,910	HBG1/2 c.-114 to -121;HBG1/2 c.-125 to -129
167_-28_NA	0.02%	10	86.56%	False	chr11:5,249,969-5,249,996	chr11:5,254,893-5,254,920	HBG1/2 c.-112 to -139
158_-22_NA	0.02%	10	86.59%	False	chr11:5,249,960-5,249,981	chr11:5,254,884-5,254,905	HBG1/2 c.-103 to -124
164_-16_NA	0.02%	11	86.61%	False	chr11:5,249,966-5,249,981	chr11:5,254,890-5,254,905	HBG1/2 c.-109 to -124
169_-4_NA;175_-2_NA	0.02%	7	86.63%	False	chr11:5,249,971-5,249,974;chr11:5,249,977-5,249,978	chr11:5,254,895-5,254,898;chr11:5,254,901-5,254,902	HBG1/2 c.-114 to -117;HBG1/2 c.-120 to -121
151_-25_NA	0.02%	10	86.66%	False	chr11:5,249,953-5,249,977	chr11:5,254,877-5,254,901	HBG1/2 c.-96 to -120
165_-6_NA;174_-4_NA	0.02%	11	86.68%	False	chr11:5,249,967-5,249,972;chr11:5,249,976-5,249,979	chr11:5,254,891-5,254,896;chr11:5,254,900-5,254,903	HBG1/2 c.-110 to -115;HBG1/2 c.-119 to -122
157_-13_NA;180_-5_NA	0.02%	8	86.70%	False;True	chr11:5,249,959-5,249,971;chr11:5,249,982-5,249,986	chr11:5,254,883-5,254,895;chr11:5,254,906-5,254,910	HBG1/2 c.-102 to -114;HBG1/2 c.-125 to -129
156_-36_NA	0.02%	6	86.73%	False	chr11:5,249,958-5,249,993	chr11:5,254,882-5,254,917	HBG1/2 c.-101 to -136
164_-10_NA;177_-2_NA	0.02%	1	86.75%	False	chr11:5,249,966-5,249,975;chr11:5,249,979-5,249,980	chr11:5,254,890-5,254,899;chr11:5,254,903-5,254,904	HBG1/2 c.-109 to -118;HBG1/2 c.-122 to -123
166_-293_NA	0.02%	2	86.77%	False	chr11:5,249,968-5,250,260	chr11:5,254,892-5,255,184	HBG1/2 c.-111 to -403

156_-31_NA	0.02%	8	86.79%	False	chr11:5,249,958-5,249,988	chr11:5,254,882-5,254,912	HBG1/2 c.-101 to -131
169_-8_NA;180_-4_NA	0.02%	4	86.82%	False;True	chr11:5,249,971-5,249,978;chr11:5,249,982-5,249,985	chr11:5,254,895-5,254,902;chr11:5,254,906-5,254,909	HBG1/2 c.-114 to -121;HBG1/2 c.-125 to -128
155_-8_NA;165_-5_NA	0.02%	1	86.84%	False	chr11:5,249,957-5,249,964;chr11:5,249,967-5,249,971	chr11:5,254,881-5,254,888;chr11:5,254,891-5,254,895	HBG1/2 c.-100 to -107;HBG1/2 c.-110 to -114
167_1_C;171_-4_NA	0.02%	1	86.86%	False	chr11:5,249,969-5,249,967;chr11:5,249,973-5,249,976	chr11:5,254,893-5,254,891;chr11:5,254,897-5,254,900	HBG1/2 c.-112 to -110;HBG1/2 c.-116 to -119
165_-7_NA;175_-1_NA	0.02%	6	86.88%	False	chr11:5,249,967-5,249,973;chr11:5,249,977-5,249,977	chr11:5,254,891-5,254,897;chr11:5,254,901-5,254,901	HBG1/2 c.-110 to -116;HBG1/2 c.-120 to -120
160_-20_NA	0.02%	11	86.91%	False	chr11:5,249,962-5,249,981	chr11:5,254,886-5,254,905	HBG1/2 c.-105 to -124
152_-21_NA	0.02%	9	86.93%	False	chr11:5,249,954-5,249,974	chr11:5,254,878-5,254,898	HBG1/2 c.-97 to -117
165_-39_NA	0.02%	9	86.95%	False	chr11:5,249,967-5,250,005	chr11:5,254,891-5,254,929	HBG1/2 c.-110 to -148
164_-4_NA	0.02%	7	86.97%	False	chr11:5,249,966-5,249,969	chr11:5,254,890-5,254,893	HBG1/2 c.-109 to -112
165_-3_NA;171_-3_NA	0.02%	10	86.99%	False	chr11:5,249,967-5,249,969;chr11:5,249,973-5,249,975	chr11:5,254,891-5,254,893;chr11:5,254,897-5,254,899	HBG1/2 c.-110 to -112;HBG1/2 c.-116 to -118
165_-2_NA;170_-4_NA	0.02%	9	87.02%	False	chr11:5,249,967-5,249,968;chr11:5,249,972-5,249,975	chr11:5,254,891-5,254,892;chr11:5,254,896-5,254,899	HBG1/2 c.-110 to -111;HBG1/2 c.-115 to -118
160_-11_NA;174_-9_NA	0.02%	1	87.04%	False	chr11:5,249,962-5,249,972;chr11:5,249,976-5,249,984	chr11:5,254,886-5,254,896;chr11:5,254,900-5,254,908	HBG1/2 c.-105 to -115;HBG1/2 c.-119 to -127
165_-4_NA;171_1_C	0.02%	1	87.06%	False	chr11:5,249,967-5,249,970;chr11:5,249,973-5,249,971	chr11:5,254,891-5,254,894;chr11:5,254,897-5,254,895	HBG1/2 c.-110 to -113;HBG1/2 c.-116 to -114
168_1_A	0.02%	8	87.08%	False	chr11:5,249,970-5,249,968	chr11:5,254,894-5,254,892	HBG1/2 c.-113 to -111
37_-362_NA	0.02%	3	87.10%	True	chr11:5,249,839-5,250,200	chr11:5,254,763-5,255,124	HBG1/2 c.18 to -343

152_-54_NA	0.02%	7	87.12%	True	chr11:5,249,954-5,250,007	chr11:5,254,878-5,254,931	HBG1/2 c.-97 to -150
161_-237_NA	0.02%	5	87.14%	True	chr11:5,249,963-5,250,199	chr11:5,254,887-5,255,123	HBG1/2 c.-106 to -342
167_-5_NA;175_-1_NA	0.02%	6	87.16%	False	chr11:5,249,969-5,249,973;chr11:5,249,977-5,249,977	chr11:5,254,893-5,254,897;chr11:5,254,901-5,254,901	HBG1/2 c.-112 to -116;HBG1/2 c.-120 to -120
165_-30_NA	0.02%	7	87.18%	False	chr11:5,249,967-5,249,996	chr11:5,254,891-5,254,920	HBG1/2 c.-110 to -139
167_-17_NA;186_-5_NA	0.02%	1	87.20%	False;True	chr11:5,249,969-5,249,985;chr11:5,249,988-5,249,992	chr11:5,254,893-5,254,909;chr11:5,254,912-5,254,916	HBG1/2 c.-112 to -128;HBG1/2 c.-131 to -135
166_-11_NA;180_-1_NA	0.02%	10	87.22%	False	chr11:5,249,968-5,249,978;chr11:5,249,982-5,249,982	chr11:5,254,892-5,254,902;chr11:5,254,906-5,254,906	HBG1/2 c.-111 to -121;HBG1/2 c.-125 to -125
162_-303_NA	0.02%	3	87.24%	True	chr11:5,249,964-5,250,266	chr11:5,254,888-5,255,190	HBG1/2 c.-107 to -409
166_-32_NA	0.02%	10	87.26%	False	chr11:5,249,968-5,249,999	chr11:5,254,892-5,254,923	HBG1/2 c.-111 to -142
165_-32_NA	0.02%	9	87.28%	False	chr11:5,249,967-5,249,998	chr11:5,254,891-5,254,922	HBG1/2 c.-110 to -141
144_-30_NA	0.02%	3	87.30%	False	chr11:5,249,946-5,249,975	chr11:5,254,870-5,254,899	HBG1/2 c.-89 to -118
158_-10_NA;171_-2_NA	0.02%	3	87.32%	False	chr11:5,249,960-5,249,969;chr11:5,249,973-5,249,974	chr11:5,254,884-5,254,893;chr11:5,254,897-5,254,898	HBG1/2 c.-103 to -112;HBG1/2 c.-116 to -117
165_-4_NA;172_-7_NA	0.02%	3	87.34%	False	chr11:5,249,967-5,249,970;chr11:5,249,974-5,249,980	chr11:5,254,891-5,254,894;chr11:5,254,898-5,254,904	HBG1/2 c.-110 to -113;HBG1/2 c.-117 to -123
167_-30_NA	0.02%	8	87.36%	False	chr11:5,249,969-5,249,998	chr11:5,254,893-5,254,922	HBG1/2 c.-112 to -141
169_-2_NA;173_1_A	0.02%	3	87.37%	False	chr11:5,249,971-5,249,972;chr11:5,249,975-5,249,973	chr11:5,254,895-5,254,896;chr11:5,254,899-5,254,897	HBG1/2 c.-114 to -115;HBG1/2 c.-118 to -116
167_-2_NA;170_4_GCAA	0.02%	1	87.39%	False	chr11:5,249,969-5,249,970;chr11:5,249,972-5,249,967	chr11:5,254,893-5,254,894;chr11:5,254,896-5,254,891	HBG1/2 c.-112 to -113;HBG1/2 c.-115 to -110
166_1_C;167_-4_NA	0.02%	9	87.41%	False	chr11:5,249,968-5,249,966;chr11:5,249,969-5,249,972	chr11:5,254,892-5,254,890;chr11:5,254,893-5,254,896	HBG1/2 c.-111 to -109;HBG1/2

							c.-112 to -115
170_-24_NA	0.02%	11	87.43%	False	chr11:5,249,972-5,249,995	chr11:5,254,896-5,254,919	HBG1/2 c.-115 to -138
167_-2_NA;171_-3_NA	0.02%	10	87.44%	False	chr11:5,249,969-5,249,970;chr11:5,249,973-5,249,975	chr11:5,254,893-5,254,894;chr11:5,254,897-5,254,899	HBG1/2 c.-112 to -113;HBG1/2 c.-116 to -118
162_-10_NA;177_1_T	0.02%	1	87.46%	False	chr11:5,249,964-5,249,973;chr11:5,249,979-5,249,977	chr11:5,254,888-5,254,897;chr11:5,254,903-5,254,901	HBG1/2 c.-107 to -116;HBG1/2 c.-122 to -120
124_-288_NA	0.02%	1	87.48%	True	chr11:5,249,926-5,250,213	chr11:5,254,850-5,255,137	HBG1/2 c.-69 to -356
159_-20_NA	0.02%	10	87.49%	False	chr11:5,249,961-5,249,980	chr11:5,254,885-5,254,904	HBG1/2 c.-104 to -123
161_-21_NA	0.02%	8	87.51%	False	chr11:5,249,963-5,249,983	chr11:5,254,887-5,254,907	HBG1/2 c.-106 to -126
160_-7_NA;169_-2_NA;173_-2_NA	0.02%	1	87.53%	False	chr11:5,249,962-5,249,968;chr11:5,249,971-5,249,972;chr11:5,249,975-5,249,976	chr11:5,254,886-5,254,892;chr11:5,254,895-5,254,896;chr11:5,254,899-5,254,900	HBG1/2 c.-105 to -111;HBG1/2 c.-114 to -115;HBG1/2 c.-118 to -119
156_-14_NA	0.02%	7	87.55%	False	chr11:5,249,958-5,249,971	chr11:5,254,882-5,254,895	HBG1/2 c.-101 to -114
159_-272_NA	0.02%	1	87.56%	True	chr11:5,249,961-5,250,232	chr11:5,254,885-5,255,156	HBG1/2 c.-104 to -375
166_-11_NA;180_-4_NA	0.02%	6	87.58%	False;True	chr11:5,249,968-5,249,978;chr11:5,249,982-5,249,985	chr11:5,254,892-5,254,902;chr11:5,254,906-5,254,909	HBG1/2 c.-111 to -121;HBG1/2 c.-125 to -128
167_8_GAAAAATT;170_-8_NA	0.02%	1	87.59%	False	chr11:5,249,969-5,249,960;chr11:5,249,972-5,249,979	chr11:5,254,893-5,254,884;chr11:5,254,896-5,254,903	HBG1/2 c.-112 to -103;HBG1/2 c.-115 to -122
166_-11_NA;178_1_G	0.02%	2	87.61%	False	chr11:5,249,968-5,249,978;chr11:5,249,980-5,249,978	chr11:5,254,892-5,254,902;chr11:5,254,904-5,254,902	HBG1/2 c.-111 to -121;HBG1/2 c.-123 to -121
169_-8_NA;180_-1_NA	0.02%	8	87.63%	False	chr11:5,249,971-5,249,978;chr11:5,249,982-5,249,982	chr11:5,254,895-5,254,902;chr11:5,254,906-5,254,906	HBG1/2 c.-114 to -121;HBG1/2 c.-125 to -125
158_-11_NA	0.02%	6	87.64%	False	chr11:5,249,960-5,249,970	chr11:5,254,884-5,254,894	HBG1/2 c.-103 to -113
173_1_T	0.02%	3	87.66%	False	chr11:5,249,975-5,249,973	chr11:5,254,899-5,254,897	HBG1/2 c.-118 to -116

168_-34_NA	0.02%	6	87.68%	False	chr11:5,249,970-5,250,003	chr11:5,254,894-5,254,927	HBG1/2 c.-113 to -146
172_-20_NA	0.02%	8	87.69%	True	chr11:5,249,974-5,249,993	chr11:5,254,898-5,254,917	HBG1/2 c.-117 to -136
166_-275_NA	0.02%	1	87.71%	False	chr11:5,249,968-5,250,242	chr11:5,254,892-5,255,166	HBG1/2 c.-111 to -385
156_-47_NA	0.02%	3	87.72%	False	chr11:5,249,958-5,250,004	chr11:5,254,882-5,254,928	HBG1/2 c.-101 to -147
168_-8_NA;178_-2_NA	0.02%	2	87.74%	False	chr11:5,249,970-5,249,977;chr11:5,249,980-5,249,981	chr11:5,254,894-5,254,901;chr11:5,254,904-5,254,905	HBG1/2 c.-113 to -120;HBG1/2 c.-123 to -124
164_-31_NA	0.02%	6	87.75%	False	chr11:5,249,966-5,249,996	chr11:5,254,890-5,254,920	HBG1/2 c.-109 to -139
159_-318_NA	0.02%	2	87.77%	True	chr11:5,249,961-5,250,278	chr11:5,254,885-5,255,202	HBG1/2 c.-104 to -421
166_-1_NA;168_-230_NA	0.01%	1	87.78%	False	chr11:5,249,968-5,249,968;chr11:5,249,970-5,250,199	chr11:5,254,892-5,254,892;chr11:5,254,894-5,255,123	HBG1/2 c.-111 to -111;HBG1/2 c.-113 to -342
111_-260_NA	0.01%	1	87.80%	False	chr11:5,249,913-5,250,172	chr11:5,254,837-5,255,096	HBG1/2 c.-56 to -315
166_-272_NA	0.01%	6	87.81%	True	chr11:5,249,968-5,250,239	chr11:5,254,892-5,255,163	HBG1/2 c.-111 to -382
166_-4_NA;171_2_AC	0.01%	4	87.83%	False	chr11:5,249,968-5,249,971;chr11:5,249,973-5,249,970	chr11:5,254,892-5,254,895;chr11:5,254,897-5,254,894	HBG1/2 c.-111 to -114;HBG1/2 c.-116 to -113
164_-3_NA;171_3_GGC	0.01%	1	87.84%	False	chr11:5,249,966-5,249,968;chr11:5,249,973-5,249,969	chr11:5,254,890-5,254,892;chr11:5,254,897-5,254,893	HBG1/2 c.-109 to -111;HBG1/2 c.-116 to -112
166_-10_NA;178_-1_NA	0.01%	4	87.86%	False	chr11:5,249,968-5,249,977;chr11:5,249,980-5,249,980	chr11:5,254,892-5,254,901;chr11:5,254,904-5,254,904	HBG1/2 c.-111 to -120;HBG1/2 c.-123 to -123
167_-307_NA	0.01%	1	87.87%	False	chr11:5,249,969-5,250,275	chr11:5,254,893-5,255,199	HBG1/2 c.-112 to -418
165_-308_NA	0.01%	2	87.88%	False	chr11:5,249,967-5,250,274	chr11:5,254,891-5,255,198	HBG1/2 c.-110 to -417
158_-2_NA;162_-9_NA	0.01%	1	87.90%	False	chr11:5,249,960-5,249,961;chr11:5,249,964-5,249,972	chr11:5,254,884-5,254,885;chr11:5,254,888-5,254,896	HBG1/2 c.-103 to -104;HBG1/2 c.-107 to -115
155_-22_NA;182_-1_NA	0.01%	1	87.91%	False	chr11:5,249,957-5,249,978;chr11:5,249,984-5,249,984	chr11:5,254,881-5,254,902;chr11:5,254,908-5,254,908	HBG1/2 c.-100 to -121;HBG1/2 c.-127 to -127

138_-32_NA;174_-3_NA	0.01%	1	87.92%	False	chr11:5,249,940-5,249,971;chr11:5,249,976-5,249,978	chr11:5,254,864-5,254,895;chr11:5,254,900-5,254,902	HBG1/2 c.-83 to -114;HBG1/2 c.-119 to -121
140_-52_NA	0.01%	2	87.94%	False	chr11:5,249,942-5,249,993	chr11:5,254,866-5,254,917	HBG1/2 c.-85 to -136
155_-22_NA	0.01%	5	87.95%	False	chr11:5,249,957-5,249,978	chr11:5,254,881-5,254,902	HBG1/2 c.-100 to -121
166_-5_NA;174_-8_NA;184_-1_NA	0.01%	1	87.96%	False	chr11:5,249,968-5,249,972;chr11:5,249,976-5,249,983;chr11:5,249,986-5,249,986	chr11:5,254,892-5,254,896;chr11:5,254,900-5,254,907;chr11:5,254,910-5,254,910	HBG1/2 c.-111 to -115;HBG1/2 c.-119 to -126;HBG1/2 c.-129 to -129
164_-8_NA;175_-3_NA	0.01%	5	87.97%	False	chr11:5,249,966-5,249,973;chr11:5,249,977-5,249,979	chr11:5,254,890-5,254,897;chr11:5,254,901-5,254,903	HBG1/2 c.-109 to -116;HBG1/2 c.-120 to -122
150_-10_NA;162_-7_NA	0.01%	1	87.99%	False;True	chr11:5,249,952-5,249,961;chr11:5,249,964-5,249,970	chr11:5,254,876-5,254,885;chr11:5,254,888-5,254,894	HBG1/2 c.-95 to -104;HBG1/2 c.-107 to -113
151_-37_NA	0.01%	6	88.00%	False	chr11:5,249,953-5,249,989	chr11:5,254,877-5,254,913	HBG1/2 c.-96 to -132
156_-270_NA	0.01%	1	88.01%	True	chr11:5,249,958-5,250,227	chr11:5,254,882-5,255,151	HBG1/2 c.-101 to -370
144_-47_NA	0.01%	4	88.02%	False	chr11:5,249,946-5,249,992	chr11:5,254,870-5,254,916	HBG1/2 c.-89 to -135
79_-340_NA	0.01%	1	88.03%	True	chr11:5,249,881-5,250,220	chr11:5,254,805-5,255,144	HBG1/2 c.-24 to -363
166_2_CA;167_-4_NA	0.01%	3	88.05%	False	chr11:5,249,968-5,249,965;chr11:5,249,969-5,249,972	chr11:5,254,892-5,254,889;chr11:5,254,893-5,254,896	HBG1/2 c.-111 to -108;HBG1/2 c.-112 to -115
147_-25_NA	0.01%	4	88.06%	False	chr11:5,249,949-5,249,973	chr11:5,254,873-5,254,897	HBG1/2 c.-92 to -116
168_-7_NA;176_1_A	0.01%	3	88.07%	False	chr11:5,249,970-5,249,976;chr11:5,249,978-5,249,976	chr11:5,254,894-5,254,900;chr11:5,254,902-5,254,900	HBG1/2 c.-113 to -119;HBG1/2 c.-121 to -119
168_-1_NA;170_-66_NA	0.01%	1	88.08%	False	chr11:5,249,970-5,249,970;chr11:5,249,972-5,250,037	chr11:5,254,894-5,254,894;chr11:5,254,896-5,254,961	HBG1/2 c.-113 to -113;HBG1/2 c.-115 to -180
162_-3_NA;167_-9_NA	0.01%	2	88.09%	False	chr11:5,249,964-5,249,966;chr11:5,249,969-5,249,977	chr11:5,254,888-5,254,890;chr11:5,254,893-5,254,901	HBG1/2 c.-107 to -109;HBG1/2

							c.-112 to -120
124_-267_NA	0.01%	1	88.10%	True	chr11:5,249,926-5,250,192	chr11:5,254,850-5,255,116	HBG1/2 c.-69 to -335
156_-9_NA;169_-1_NA;175_-1_NA	0.01%	1	88.11%	False	chr11:5,249,958-5,249,966;chr11:5,249,971-5,249,971;chr11:5,249,977-5,249,977	chr11:5,254,882-5,254,890;chr11:5,254,895-5,254,895;chr11:5,254,901-5,254,901	HBG1/2 c.-101 to -109;HBG1/2 c.-114 to -114;HBG1/2 c.-120 to -120
159_4_TGAC;162_-3_NA;167_3_GCC	0.01%	1	88.12%	False	chr11:5,249,961-5,249,956;chr11:5,249,964-5,249,966;chr11:5,249,969-5,249,965	chr11:5,254,885-5,254,880;chr11:5,254,888-5,254,890;chr11:5,254,893-5,254,889	HBG1/2 c.-104 to -99;HBG1/2 c.-107 to -109;HBG1/2 c.-112 to -108
169_-1_NA;173_-3_NA	0.01%	5	88.13%	False	chr11:5,249,971-5,249,971;chr11:5,249,975-5,249,977	chr11:5,254,895-5,254,895;chr11:5,254,899-5,254,901	HBG1/2 c.-114 to -114;HBG1/2 c.-118 to -120
152_-19_NA;176_-2_NA	0.01%	1	88.14%	False	chr11:5,249,954-5,249,972;chr11:5,249,978-5,249,979	chr11:5,254,878-5,254,896;chr11:5,254,902-5,254,903	HBG1/2 c.-97 to -115;HBG1/2 c.-121 to -122
42_-176_NA	0.01%	2	88.15%	False	chr11:5,249,844-5,250,019	chr11:5,254,768-5,254,943	HBG1/2 c.13 to -162
169_-37_NA	0.01%	1	88.16%	True	chr11:5,249,971-5,250,007	chr11:5,254,895-5,254,931	HBG1/2 c.-114 to -150
159_-301_NA	0.01%	1	88.17%	True	chr11:5,249,961-5,250,261	chr11:5,254,885-5,255,185	HBG1/2 c.-104 to -404
169_-64_NA	0.01%	5	88.18%	True	chr11:5,249,971-5,250,034	chr11:5,254,895-5,254,958	HBG1/2 c.-114 to -177
147_-41_NA	0.01%	2	88.19%	False	chr11:5,249,949-5,249,989	chr11:5,254,873-5,254,913	HBG1/2 c.-92 to -132
140_-31_NA;177_-1_NA	0.01%	1	88.20%	False	chr11:5,249,942-5,249,972;chr11:5,249,979-5,249,979	chr11:5,254,866-5,254,896;chr11:5,254,903-5,254,903	HBG1/2 c.-85 to -115;HBG1/2 c.-122 to -122
170_-1_NA;172_-303_NA	0.01%	2	88.21%	False	chr11:5,249,972-5,249,972;chr11:5,249,974-5,250,276	chr11:5,254,896-5,254,896;chr11:5,254,898-5,255,200	HBG1/2 c.-115 to -115;HBG1/2 c.-117 to -419
166_-4_NA;172_-21_NA	0.01%	1	88.22%	False	chr11:5,249,968-5,249,971;chr11:5,249,974-5,249,994	chr11:5,254,892-5,254,895;chr11:5,254,898-5,254,918	HBG1/2 c.-111 to -114;HBG1/2 c.-117 to -137

166_-7_NA;177_2_CC	0.01%	1	88.23%	False	chr11:5,249,968-5,249,974;chr11:5,249,979-5,249,976	chr11:5,254,892-5,254,898;chr11:5,254,903-5,254,900	HBG1/2 c.-111 to -117;HBG1/2 c.-122 to -119
152_-254_NA	0.01%	1	88.23%	True	chr11:5,249,954-5,250,207	chr11:5,254,878-5,255,131	HBG1/2 c.-97 to -350
165_-4_NA;172_-5_NA	0.01%	3	88.24%	False;True	chr11:5,249,967-5,249,970;chr11:5,249,974-5,249,978	chr11:5,254,891-5,254,894;chr11:5,254,898-5,254,902	HBG1/2 c.-110 to -113;HBG1/2 c.-117 to -121
174_-13_NA	0.01%	5	88.25%	False	chr11:5,249,976-5,249,988	chr11:5,254,900-5,254,912	HBG1/2 c.-119 to -131
154_-17_NA;175_-6_NA	0.01%	1	88.26%	False	chr11:5,249,956-5,249,972;chr11:5,249,977-5,249,982	chr11:5,254,880-5,254,896;chr11:5,254,901-5,254,906	HBG1/2 c.-99 to -115;HBG1/2 c.-120 to -125
161_-9_NA;172_-20_NA	0.01%	1	88.27%	False;True	chr11:5,249,963-5,249,971;chr11:5,249,974-5,249,993	chr11:5,254,887-5,254,895;chr11:5,254,898-5,254,917	HBG1/2 c.-106 to -114;HBG1/2 c.-117 to -136
132_-42_NA	0.01%	1	88.28%	False	chr11:5,249,934-5,249,975	chr11:5,254,858-5,254,899	HBG1/2 c.-77 to -118
162_-259_NA	0.01%	2	88.29%	False	chr11:5,249,964-5,250,222	chr11:5,254,888-5,255,146	HBG1/2 c.-107 to -365
166_-18_NA	0.01%	2	88.29%	True	chr11:5,249,968-5,249,985	chr11:5,254,892-5,254,909	HBG1/2 c.-111 to -128
145_-25_NA	0.01%	1	88.30%	False	chr11:5,249,947-5,249,971	chr11:5,254,871-5,254,895	HBG1/2 c.-90 to -114
148_-29_NA	0.01%	1	88.31%	False	chr11:5,249,950-5,249,978	chr11:5,254,874-5,254,902	HBG1/2 c.-93 to -121
144_-327_NA	0.01%	1	88.32%	False	chr11:5,249,946-5,250,272	chr11:5,254,870-5,255,196	HBG1/2 c.-89 to -415
169_-274_NA	0.01%	1	88.33%	False	chr11:5,249,971-5,250,244	chr11:5,254,895-5,255,168	HBG1/2 c.-114 to -387

The Indel_id is a string identifying the indel, shown as indel_start_position + _ + indel_length + _ + ID. Where ID is NA for deletions and for insertions is the sequence inserted. Positive values indicate insertions and negative values indicate deletions.

The average percentage ("AVE % in indel") is the average percentage of each indel_ID amongst all indels detected in a sample. The average is calculated from n = 14 samples.

"Cts" = counts and provides the number of samples in which the indel was identified.

"Cum. Sum" = Cumulative sum and the cumulative percentage of indels, e.g. the top ten most abuntant indels account for 45.03% of all indels.

In the column labeled MMEJ, "True" indicates the repair was mediated via MMEJ, and "False" indicates the repair was mediated via NHEJ.

The genomic coordinates at *HBG1* or *HBG2* are reported using the One-based coordinate system, "Homo sapiens chromosome 11, GRCh38.p12 Primary Assembly," (Version NC_000011.10), NCBI Reference Sequence NC_000011.

The *HBG1/2* position indicates the deletion relative to the TSS.

SEQUENCES

[0485] 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.

INCORPORATION BY REFERENCE

[0486] 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

[0487] 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

Ahern et al., Br J Haematol 25(4):437-444 (1973)
Akinbami Hemoglobin 40:64-65 (2016)
Aliyu et al. Am J Hematol 83:63-70 (2008)
Anders et al. Nature 513(7519):569-573 (2014)
Angastiniotis & Modell Ann N Y Acad Sci 850:251-269 (1998)
Bae et al. Bioinformatics 30(10):1473-1475 (2014)
Barbosa et al. Braz J Med Bio Res 43(8):705-711 (2010)
Bauer et al. Nat. Med. 25(5):776-783 (2019)
Bothmer et al. CRISPR J 3(3):177-187 (2020)
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)
Caldecott Nat Rev Genet 9(8):619-631 (2008)
Canvers et al. Nature 527(12):192-197 (2015)

Chang et al. Mol Ther Methods Clin Dev 4:137-148 (2017)
Chassanidis Ann Hematol 88(6):549-555 (2009)
Chylinski et al. RNA Biol 10(5):726–737 (2013)
Cong et al. Science 399(6121):819-823 (2013)
Costa et al., Cad Saude Publica 18(5):1469-1471 (2002)
Davis & Maizels 2 Proc Natl Acad Sci USA 111(10):E924-932 (2014)
Fine et al. Sci Rep. 5:10777 (2015)
Frit et al. DNA Repair (Amst) 17:81-97 (2014)
Fu et al. Nat Biotechnol 32:279-284 (2014)
Gao et al. Nat Biotechnol.: 35(8):789-792 (2017)
Giarratana et al. Nat Biotechnol. 23(1):69-74 (2005)
Giarratana et al. Blood:118, 5071-5079 (2011)
Giannoukos et al. BMC Genomics 19(1):212 (2018)
Guilinger et al. Nat Biotechnol 32:577–582 (2014)
Heigwer et al. Nat Methods 11(2):122-123 (2014)
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)
Kleinstiver et al. Nat Biotechnol 37(3):276-282 (2019)
Komor et al. Nature 533(7603):420–424 (2016)
Kosicki et al. Nat Biotechnol 36(8): 765-771 (2018)
Lee et al. Nano Lett 12(12):6322-6327 (2012)
Lewis "Medical-Surgical Nursing: Assessment and Management of Clinical Problems" (2014)
Li Cell Res 18(1):85-98 (2008)
Makarova et al. Nat Rev Microbiol 9(6):467–477 (2011)
Mali et al. Science 339(6121):823-826 (2013)
Mantovani et al. Nucleic Acids Res 16(16):7783-7797 (1988)
Masala Methods Enzymol. 231:21-44 (1994)
Marteijn et al. Nat Rev Mol Cell Biol 15(7):465-481 (2014)
Martyn et al., Biochim Biophys Acta 1860(5):525-536 (2017)
Métais et al. Blood Adv. 3(21):3379-92 (2019)

Nishimasu et al. Cell 156(5):935-949 (2014)
Nishimasu et al. Cell 162:1113-1126 (2015)
Notta et al. Science 333(6039):218-21 (2011)
Pausch et al. Science 369(6501):333-337 (2020)
Ran & Hsu Cell 154(6):1380-1389 (2013)
Richardson et al. Nat Biotechnol 34:339–344 (2016)
Swarts et al. May 22:e1481. doi: 10.1002/wrna.1481. Epub ahead of print. PMID: 29790280 (2018)
Shmakov et al. Molecular Cell 60(3):385–397 (2015)
Sternberg et al. Nature 507(7490):62-67 (2014)
Strohkendl et al. Mol. Cell 71:816-824 (2018)
Superti-Furga et al. EMBO J 7(10):3099-3107 (1988)
Thein Hum Mol Genet 18(R2):R216-223 (2009)
Thorpe et al. Br J Haematol. 87(1):125-132 (1994)
Tsai et al. Nat Biotechnol 34(5): 483 (2016)
Waber et al. Blood 67(2):551-554 (1986)
Wang et al. Cell 153(4):910-918 (2013)
Weber et al. Sci Adv. 6(7):eaay9392 (2020)
Wu et al. Nat. Med. 25(5): 776-83 (2019)
Xiao et al. Bioinformatics 30(8):1180-1182 (2014)
Xu et al. Genes Dev 24(8):783-798 (2010)
Yamano et al. Cell 165(4): 949–962 (2016)
Zetsche et al. Nat Biotechnol 33(2):139-42 (2015)

CLAIMS

1. A first population of modified cells comprising

a plurality of modified CD34+ or hematopoietic stem cells,

one or more of the plurality of modified cells having an indel in an *HBG* gene promoter,

the plurality of modified cells as a whole including a HBG1/2 c.-104 to -121 deletion in a *HBG1* promoter, a *HBG2* promoter, or a combination thereof, and

the HBG1/2 c.-104 to -121 deletion making up 2% or more of the indels in the plurality of modified cells as a whole.

2. The first population of modified cells of claim 1, wherein the HBG1/2 c.-104 to -121 deletion makes up less than 25% of the indels in the plurality of modified cells as a whole.

3. A first population of modified cells comprising

a plurality of modified CD34+ or hematopoietic stem cells,

one or more of the plurality of modified cells including an indel in an *HBG* gene promoter, and

60% or more of the indels in the plurality of modified cells as a whole are deletions of at least 4 base pairs.

4. The first population of modified cells of any of claims 1-3, produced by a process comprising delivering a first RNP complex including a first guide RNA (gRNA) and a Cpf1 RNA-guided nuclease or a modified Cpf1 RNA-guided nuclease to a first population of unmodified cells comprising a plurality of unmodified CD34+ or hematopoietic stem cells to generate the indels, the first gRNA including a first gRNA targeting domain.

5. A first population of modified cells comprising:

a plurality of modified CD34+ or hematopoietic stem cells,

one or more of the plurality of modified cells including an indel in an *HBG* gene promoter, and

60% or more of the indels in the plurality of modified cells as a whole are deletions of at least 4 base pairs;

the indels generated by delivering a first RNP complex including a first gRNA and a Cpf1 RNA-guided nuclease or a modified Cpf1 RNA-guided nuclease to a first population of unmodified

cells comprising a plurality of unmodified CD34+ or hematopoietic stem cells, the first gRNA including a first gRNA targeting domain.

6. The first population of modified cells of any of claims 1-5, wherein 25% or more of the indels in the plurality of modified cells as a whole are deletions of at least 4 base pairs and are introduced by a repair mechanism other than microhomology-mediated end joining (MMEJ) repair.

7. The first population of modified cells of any of claims 1-5, wherein 25% or more of the indels in the plurality of modified cells as a whole are deletions of at least 4 base pairs and are introduced by non-homologous end joining (NHEJ) repair.

8. The first population of modified cells of any of claims 1-7, wherein 50% or more of the indels in the plurality of modified cells as a whole are deletions between 1 base pair and 25 base pairs.

9. The first population of modified cells of any of claims 1-7, wherein 50% or more of the indels in the plurality of modified cells as a whole are deletions between 3 base pairs and 25 base pairs.

10. The first population of modified cells of any of claims 1-7, wherein 50% or more of the indels in the plurality of modified cells as a whole are deletions between 4 base pairs and 25 base pairs.

11. The first population of modified cells of any of claims 1-7, wherein 50% or more of the indels in the plurality of modified cells as a whole are deletions between 5 base pairs and 25 base pairs.

12. The first population of modified cells of any of claims 1-11, wherein the plurality of modified cells as a whole comprise the following deletions:

a HBG1/2 c.-104 to -121 deletion in a *HBG1* promoter, *HBG2* promoter, or a combination thereof, the HBG1/2 c.-104 to -121 deletion making up 2% or more of the indels in the plurality of modified cells as a whole; and

a HBG1/2 c.-110 to -115 deletion in a *HBG1* promoter, *HBG2* promoter, or a combination thereof, the HBG1/2 c. -110 to -115 deletion making up 2% or more of the indels in the plurality of modified cells as a whole.

13. The first population of modified cells of any of claims 1-12, wherein the plurality of modified cells as a whole comprise the following deletions:

a HBG1/2 c.-104 to -121 deletion in a *HBG1* promoter, *HBG2* promoter, or a combination thereof, the HBG1/2 c.-104 to -121 deletion making up 1% to 15.5% of the indels in the plurality of modified cells as a whole;

a HBG1/2 c.-110 to -115 deletion in a *HBG1* promoter, *HBG2* promoter, or a combination thereof, the HBG1/2 c. -110 to -115 deletion making up 1% to 6% of the indels in the plurality of modified cells as a whole;

a HBG1/2 c.-112 to -115 deletion in a *HBG1* promoter, *HBG2* promoter, or a combination thereof, the HBG1/2 c.-112 to -115 deletion making up 1% to 6% of the indels in the plurality of modified cells as a whole;

a HBG1/2 c.-113 to -115 deletion in a *HBG1* promoter, *HBG2* promoter, or a combination thereof, the HBG1/2 c.-113 to -115 deletion making up 1% to 6% of the indels in the plurality of modified cells as a whole;

a HBG1/2 c.-111 to -115 deletion in a *HBG1* promoter, *HBG2* promoter, or a combination thereof, the HBG1/2 c.-111 to -115 deletion making up 1% to 6% of the indels in the plurality of modified cells as a whole;

a HBG1/2 c.-111 to -117 deletion in a *HBG1* promoter, *HBG2* promoter, or a combination thereof, the HBG1/2 c.-111 to -117 deletion making up 1% to 6% of the indels in the plurality of modified cells as a whole;

a HBG1/2 c.-102 to -114 deletion in a *HBG1* promoter, *HBG2* promoter, or a combination thereof, the HBG1/2 c.-102 to -114 deletion making up 1% to 6% of the indels in the plurality of modified cells as a whole;

a HBG1/2 c.-114 to -118 deletion in a *HBG1* promoter, *HBG2* promoter, or a combination thereof, the HBG1/2 c.-114 to -118 deletion making up 1% to 6% of the indels in the plurality of modified cells as a whole;

a HBG1/2 c.-112 to -116 deletion in a *HBG1* promoter, *HBG2* promoter, or a combination thereof, the HBG1/2 c.-112 to -116 deletion making up 1% to 6% of the indels in the plurality of modified cells as a whole; and

a HBG1/2 c.-113 to -117 deletion in a *HBG1* promoter, *HBG2* promoter, or a combination thereof, the HBG1/2 c.-113 to -117 deletion making up 1% to 6% of the indels in the plurality of modified cells as a whole.

14. The first population of modified cells of any of claims 1-13, wherein the plurality of modified cells as a whole comprises 108 deletions present in all 14 samples set forth in Fig. 56A.

15. The first population of modified cells of any of claims 1-13, wherein the plurality of modified cells as a whole comprise the indels identified as having an Ave % in Indel of 0.1% or more in Table 25.

16. The first population of modified cells of any of claims 1-15, wherein the plurality of modified cells as a whole comprises the indels set forth in Table 25.

17. The first population of modified cells of any of claims 1-16, wherein the plurality of modified cells as a whole comprises at least 10% more deletions of at least 4 base pairs than a second population of modified cells comprising a plurality of modified CD34+ or hematopoietic stem cells,

one or more modified cells in the second population of modified cells having an indel in an *HBG* gene promoter; and

the indels of the second population of modified cells generated by delivering a second RNP complex comprising a second gRNA and a Cas9 RNA-guided nuclease to a second population of unmodified cells comprising a plurality of unmodified CD34+ or hematopoietic stem cells, the second gRNA including a second gRNA targeting domain comprising SEQ ID NO:339.

18. The first population of modified cells of any of claims 1-17, wherein the plurality of modified cells as a whole comprises at least 10% more deletions of at least 4 base pairs introduced by an NHEJ repair mechanism than a second population of modified cells comprising a plurality of modified CD34+ or hematopoietic stem cells,

one or more modified cells in the second population of modified cells comprising an indel in an *HBG* gene promoter; and

the indels of the second population of modified cells generated by delivering a second RNP complex comprising a second gRNA and a Cas9 RNA-guided nuclease to a second population of unmodified cells comprising a plurality of unmodified CD34+ or hematopoietic stem cells, the second gRNA including a second gRNA targeting domain comprising SEQ ID NO:339.

19. The first population of modified cells of claims 4-18, wherein the first RNP complex is delivered to the first population of unmodified cells by electroporation.

20. The first population of modified cells of claim 17 or 18, wherein the second RNP complex is delivered to the second population of unmodified cells by electroporation.

21. The first population of modified cells of any of claims 1-20, wherein the first population of unmodified cells is from a subject having sickle cell disease.

22. The first population of modified cells of claim 17 or 18, wherein the second population of unmodified cells is from a subject having sickle cell disease.

23. The first population of modified cells of any of claims 4-22, wherein the first population of modified cells comprises higher HbF levels than the first population of the unmodified cells.

24. The first population of modified cells of claim 17 or 18, wherein the first population of modified cells comprises higher HbF levels than the second population of modified cells.

25. The first population of modified cells of any of claims 4-24, the first gRNA including a 5' end and a 3' end, a DNA extension at the 5' end, and a 2'-O-methyl-3'-phosphorothioate modification at the 3' end.

26. The first population of modified cells of claim 25, wherein the DNA extension comprises a sequence set forth in SEQ ID NOs:1235-1250.

27. The first population of modified cells of any of claims 4-26, wherein the first gRNA targeting domain comprises SEQ ID NO:1254.

28. The first population of modified cells of any of claims 4-27, wherein the gRNA comprises SEQ ID NO:1051.

29. The first population of modified cells of any of claims 4-28, wherein the modified Cpf1 comprises SEQ ID NO:1097.

30. The first population of modified cells of any of claims 1-29, wherein the indel in an *HBG* gene promoter is in a CCAAT box target region.

31. A method of inducing expression of fetal hemoglobin (HbF) in a first population of modified cells comprising a plurality of modified CD34+ or hematopoietic stem cells, the method comprising:

delivering a first RNP complex including a first guide RNA (gRNA) and a Cpf1 RNA-guided nuclease or a modified Cpf1 RNA-guided nuclease to a first population of unmodified cells comprising a plurality of unmodified CD34+ or hematopoietic stem cells to generate indels, the first gRNA including a first gRNA targeting domain,

each modified CD34+ or hematopoietic stem cell comprising an indel in an *HBG* gene promoter,

60% or more of the indels in the plurality of modified cells as a whole are deletions of at least 4 base pairs;

the first population of modified cells comprising higher HbF levels than the first population of unmodified cells.

32. The method of inducing expression of HbF of claim 31, wherein the first population of modified cells comprises the first population of modified cells of any of claims 1-30.

33. A method of decreasing sickling in a first population of red blood cells (RBCs) cultured from a first population of modified cells comprising a plurality of modified CD34+ cells from a subject having sickle cell disease, the method comprising:

a) delivering a first RNP complex including a first gRNA and a Cpf1 RNA-guided nuclease or a modified Cpf1 RNA-guided nuclease to a first population of unmodified cells comprising a plurality of unmodified CD34+ cells from the subject to generate indels, the first gRNA including a first gRNA targeting domain, each modified CD34+ cell comprising an indel in an *HBG* gene promoter, 60% or more of the indels in the plurality of modified CD34+ cells as a whole are deletions of at least 4 base pairs; and

b) culturing the first population of RBCs from the first population of cells comprising the plurality of modified CD34+ cells,

the first population of RBCs exhibiting significantly decreased sickling upon deoxygenation than a second population of RBCs comprising a plurality of RBCs cultured from the first population of unmodified cells.

34. The method of decreasing sickling of claim 33, wherein the first population of RBCs sickle at a significantly lower oxygen tension than the second population of RBCs.

35. The method of decreasing sickling of claim 34, wherein the oxygen tension is measured by relative oxygen pressure.

36. The method of decreasing sickling of any of claims 33-35, wherein the first population of RBCs has a significantly higher minimum elongation index upon deoxygenation than the second population of RBCs.

37. The method of decreasing sickling of any of claims 33-35, wherein the first population of RBCs have a significantly higher velocity upon deoxygenation than a second population of RBCs comprising a plurality of RBCs cultured from the first population of unmodified cells.

38. The method of decreasing sickling of any of claims 33-35, wherein the first population of RBCs comprises higher HbF levels than a second population of RBCs comprising a plurality of RBCs cultured from the first population of unmodified cells.

39. The method of decreasing sickling of any of claims 33-38, wherein the first population of modified cells comprises the first population of modified cells of any of claims 1-30.

40. A method of alleviating one or more symptoms of sickle cell disease in a subject in need thereof comprising:

a) delivering a first RNP complex including a first gRNA and a Cpf1 RNA-guided nuclease or a modified Cpf1 RNA-guided nuclease to a first population of unmodified cells comprising a plurality of unmodified CD34+ or hematopoietic stem cells to generate indels, each modified CD34+ or hematopoietic stem comprising an indel in an *HBG* gene promoter, 60% or more of the indels in a plurality of the modified CD34+ or hematopoietic stem cells as a whole are deletions of at least 4 base pairs; and

b) administering a first population of modified cells comprising the plurality of modified CD34+ or hematopoietic stem cells to the subject to alleviate one or more symptoms of sickle cell disease.

41. The method of claim 40, further comprising detecting a population of modified erythroid progeny cells comprising a plurality of modified erythroid progeny cells cultured from the first population of modified cells at one or more periods between 1 week and 5 years after administration.

42. The method of claim 40 or 41, wherein the method results in long-term engraftment of a plurality of HSC clones in bone marrow.

43. The method of any of claim 40-42, wherein the method results in long-term expression of at least 50% of total hemoglobin as compared to a healthy subject.

44. The method of any of claims 40-43, wherein the first population of modified cells comprises the first population of modified cells of any of claims 1-30.