Innovation, Sciences et Développement économique Canada
Office de la Propriété Intellectuelle du Canada

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3079437 A1 2019/04/25

(21) 3 079 437

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2018/10/17
(87) Date publication PCT/PCT Publication Date: 2019/04/25
(85) Entrée phase nationale/National Entry: 2020/04/16
(86) N° demande PCT/PCT Application No.: US 2018/056271
(87) N° publication PCT/PCT Publication No.: 2019/079437
(30) Priorités/Priorities: 2017/10/18 (US62/574,163); 2018/01/24 (US62/621,102)

(51) Cl.Int./Int.Cl. *C12N 7/01* (2006.01), *A61K 35/76* (2015.01), *A61K 48/00* (2006.01), *A61P 7/00* (2006.01), *C07K 14/015* (2006.01), *C07K 14/805* (2006.01), *C12N 15/09* (2006.01), *C12N 15/12* (2006.01), *C12N 15/64* (2006.01), *C12N 15/864* (2006.01)

(71) Demandeurs/Applicants:
CITY OF HOPE, US;
HOMOLOGY MEDICINES, INC., US

(72) Inventeurs/Inventors:
CHATTERJEE, SASWATI, US;
WONG, KAMEHAMEHA K., US;
BENHAJSALAH, MARWA, US;

(54) Titre : COMPOSITIONS DE VIRUS ADENO-ASSOCIES POUR RESTAURER LA FONCTION DU GENE HBB ET LEURS PROCEDES D'UTILISATION

(54) Title: ADENO-ASSOCIATED VIRUS COMPOSITIONS FOR RESTORING HBB GENE FUNCTION AND METHODS OF USE THEREOF

(57) Abrégé/Abstract:
Provided herein are adeno-associated virus (AAV) compositions for correcting a mutation in a beta globin gene (HBB) gene and methods of using the same to correct an HBB gene mutation in a cell. Also provided are packaging systems for making the adeno-associated virus compositions.

(21) 3 079 437

(13) A1

(72) Inventeurs(suite)/Inventors(continued): SMITH, LAURA JANE, US; SEYMOUR, ALBERT BARNES, US; WRIGHT, JASON BOKE, US; MCSWIGGEN, JAMES ANTHONY, US; DOLLIVE, SERENA NICOLE, US; ST. MARTIN, THIA BABOVAL, US; PROUT, JAIME, US

(74) Agent: SMART & BIGGAR LLP

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
25 April 2019 (25.04.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/079437 A1

(51) International Patent Classification:
A61K 48/00 (2006.01) *A61K 35/76* (2015.01)
C12N 7/04 (2006.01) *C12N 15/90* (2006.01)
C12N 15/864 (2006.01)

(21) International Application Number:
PCT/US2018/056271

(22) International Filing Date:
17 October 2018 (17.10.2018)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/574,163 18 October 2017 (18.10.2017) US
62/621,102 24 January 2018 (24.01.2018) US

(71) Applicants: **CITY OF HOPE** [US/US]; 1500 East Duarte Road, Duarte, California 91010-3000 (US). **HOMOLOGY MEDICINES, INC.** [US/US]; 45 Wiggins Avenue, Bedford, Massachusetts 01730 (US).

(72) Inventors: **CHATTERJEE, Saswati**; 614 Church Canyon Place, Altadena, California 91001 (US). **WONG, Kamehameha K.**; 614 Church Canyon Place, Altadena, California 91001 (US). **BENHAJSALAH, Marwa**; 1412 Duarte Road, Apt. C, Duarte, California 91010 (US). **SMITH, Laura Jane**; 14131 Bay Drive, Westford, Massachusetts 01886 (US). **SEYMOUR, Albert Barnes**; 1 Brickyard Lane, Westborough, Massachusetts 01581 (US). **WRIGHT, Jason Boke**; 57 Whits End Road, Concord, MA 01742 (US). **MCSWIGGEN, James Anthony**; 15 Beacon Street, Unit A, Arlington, Massachusetts 02474 (US). **DOLLIVE, Serena Nicole**; 19 Everett St., Unit 1, Waltham, Massachusetts 02453 (US). **ST. MARTIN, Thia Baboval**; 488 Holman St., Lunenburg, Massachusetts 01462 (US). **PROUT, Jaime**; 3 Raven Dr., Hudson, NH 03051 (US).

(74) Agent: **WILKINS, Andrew T.** et al.; Lathrop Gage LLP, 28 State Street, Suite 700, Boston, Massachusetts 02109 (US).

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

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

Declarations under Rule 4.17:
— *as to the identity of the inventor (Rule 4.17(i))*

Published:
— *with international search report (Art. 21(3))*
— *with sequence listing part of description (Rule 5.2(a))*

WO 2019/079437 A1

(54) **Title:** ADENO-ASSOCIATED VIRUS COMPOSITIONS FOR RESTORING HBB GENE FUNCTION AND METHODS OF USE THEREOF

(57) **Abstract:** Provided herein are adeno-associated virus (AAV) compositions for correcting a mutation in a beta globin gene (HBB) gene and methods of using the same to correct an HBB gene mutation in a cell. Also provided are packaging systems for making the adeno-associated virus compositions.

ADENO-ASSOCIATED VIRUS COMPOSITIONS FOR RESTORING HBB GENE FUNCTION AND METHODS OF USE THEREOF

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application Nos.: 62/574,163, filed October 18, 2017; and 62/621,102, filed January 24, 2018, each of which is hereby incorporated by reference in its entirety.

SEQUENCE LISTING

[0002] The instant application contains a Sequence Listing which has been submitted electronically in ASCII format, created on October 12, 2018, named 606107_HMT-023PC_Sequence_Listing_ST25.txt, and is 200,630 bytes in size. The Sequence Listing is hereby incorporated by reference in its entirety.

STATEMENT OF GOVERNMENT INTEREST

[0003] This invention was made with government support under Grant No. P30CA033572 awarded by the National Institutes of Health. The Government has certain rights in the invention.

BACKGROUND

[0004] Hemoglobinopathies comprise a family of genetic disorders in which the production, structure, and/or function of a hemoglobin protein is abnormal. Hemoglobin protein makes up approximately 97% of the dry weight of erythrocytes and it increases the oxygen carrying ability of blood by about seventy-fold. The predominant adult hemoglobin protein is composed of two alpha globin (HBA) subunits, two beta globin (HBB) subunits, and a heme group associated with each subunit. Genetic defects in HBB on chromosome 11 can cause certain hemoglobinopathies such as sickle cell disease (SCD) and beta thalassemia.

[0005] Sickle cell disease, also called sickle cell anemia, is an autosomal recessive disease affecting approximately 100,000 Americans. It is prevalent among African Americans but also present in other ethnic groups. In West and Central Africa, 1-2% of all babies are born with SCD. SCD is caused by homozygous mutation at nucleotide 20 in the coding sequence of HBB. This mutation replaces the negatively charged amino acid glutamate (encoded by GAG) with a neutral, hydrophobic residue, valine (encoded by GUG) at the sixth amino acid of mature beta globin. Hemoglobin containing a beta globin chain having the SCD mutation tends to aggregate into multi-stranded polymers which distort the shape of erythrocytes, making these cells fragile and shaped like crescents or sickles. These

abnormal erythrocytes are more disposed towards hemolysis and deliver less oxygen to tissues and organs. Moreover, hemoglobin aggregation renders erythrocytes rigid and easily retained in small blood vessels, thereby decreasing the blood flow and causing vascular occlusion. As a result, SCD patients suffer from anemia and episodes of pain called "crisis", and organ damage during the crisis is the major cause of the mortality and morbidity associated with SCD. In particular, infarction (*i.e.*, necrosis of tissue due to insufficient blood supply) of bone, spleen, kidney, and lungs is particularly common. By contrast, people who are heterozygous for the sickle cell mutation are largely asymptomatic.

[0006] Beta thalassemia affects about 1 out of 100,000 individuals throughout the world and about 1 out of 10,000 people in the European Union. Beta thalassemia is a group of disorders caused by various mutations in HBB that reduce expression of beta globin. So far, 884 different mutations, including substitution, insertions and deletions, have been identified in beta thalassemia (HbVar database). These mutations are located throughout the genomic locus of HBB. Among the substitutions, insertions, and small deletions, pathogenic variants have been found upstream of the 5' UTR and into the 3' UTR. Besides these known 884 mutations, additional variants may be pathogenic only in conjunction with specific variation in the HBB sequence. The beta thalassemia mutations may affect gene transcription, RNA processing, post-transcriptional modification, translation of mRNA, *etc.* Beta thalassemia is highly variable in severity, with some HBB mutations leading to complete loss of beta globin production and other HBB mutations leading to only a reduction in the quantity of beta globin. Patients with severe beta thalassemia (*i.e.*, thalassemia major), who often have HBB mutation(s) in both alleles, suffer anemia, growth retardation and abnormal organ development. Patients with mild to moderate beta thalassemia (*i.e.*, thalassemia minor or thalassemia intermedia) manifest less severe symptoms.

[0007] Hemoglobinopathy can be managed by blood transfusion and supportive care, with SCD and beta thalassemia major requiring chronic transfusions. However, repeated transfusions result in iron overload and require iron chelation therapy to reduce the incidence of complications. Morbidity and mortality of SCD and beta thalassemia can also be attenuated by hydroxyurea, the only FDA approved drug to date for SCD. However, this treatment is not widely used due to its low prescription rate and poor compliance.

[0008] In order to cure SCD or beta thalassemia, patients need to receive hematopoietic stem cells carrying at least one copy of functional beta hemoglobin that can be adequately expressed. One approach is to obtain wild-type hematopoietic stem cells from an

allogeneic donor through bone marrow transplantation. However, the availability of matched donors is a major limiting factor, and bone marrow transplantation is often associated with serious complications that lead to a mortality rate of 5-10%. More recently, gene therapy approaches have been employed to introduce a beta globin expressing polynucleotide *ex vivo* into mutant hematopoietic stem cells isolated from the patient.

[0009] To date, all HBB gene therapy clinical trials have involved the use of retroviral vectors, such as lentiviral vectors. However, retrovirus-based gene therapy raises a number of safety and efficacy concerns. For example, because insertion of retroviral vectors into the human genome is non-targeted, there is a risk of the vector disrupting a tumor suppressor gene or activating an oncogene, thereby causing a malignancy. Indeed, in a clinical trial for treating X-linked severe combined immunodeficiency (SCID) by transducing $CD34^+$ bone marrow precursors with a gammaretroviral vector, four out of ten patients developed leukemia (Hacein-Bey-Abina *et al.*, J Clin Invest. (2008) 118(9):3132-42). Moreover, due to these safety concerns, lentiviral gene therapy can only be performed *ex vivo*. This *ex vivo* use reduces the efficacy of the therapy because the number of hematopoietic stem cells that can be extracted from a subject for *ex vivo* therapy is only a small fraction of those present in the subject, and there is no reliable method currently in clinical use to expand hematopoietic stem cells *ex vivo*.

[0010] It has also been speculated that nuclease-based gene editing technologies, such as meganucleases, zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered, regularly interspaced, short palindromic repeat (CRISPR) technology, may be used to correct defect in the HBB gene in SCD and beta thalassemia patients. However, each of these technologies raises safety concerns due to the potential for off-target mutation of sites in the human genome similar in sequence to the intended target site.

[0011] Accordingly, there is a need in the art for improved gene therapy compositions and methods that can efficiently and safely restore HBB gene function in SCD and beta thalassemia patients.

SUMMARY

[0012] Provided herein are adeno-associated virus (AAV) compositions for correcting a mutation in an HBB gene and methods of using the same to correct an HBB gene mutation in a cell. Also provided are packaging systems for making the adeno-associated virus compositions.

[0013] The AAV compositions and methods disclosed herein are particularly advantageous in that they allow for highly efficient correction of mutations in an HBB gene *in vivo*, without the need for cleavage of genomic DNA using an exogenous nuclease (*e.g.*, a meganuclease, a zinc finger nuclease, a transcriptional activator-like nuclease (TALEN), or an RNA-guided nuclease such as a Cas9).

[0014] Accordingly, in one aspect, the instant disclosure provides a replication-defective adeno-associated virus (AAV) comprising (a) an AAV capsid comprising an AAV Clade F capsid protein, and (b) a correction genome comprising (i) an editing element for editing a target locus in a target gene, (ii) a 5' homology arm nucleotide sequence 5' of the editing element having homology to a first genomic region 5' to the target locus, and (iii) a 3' homology arm nucleotide sequence 3' of the editing element having homology to a second genomic region 3' to the target locus.

[0015] In another aspect, the instant disclosure provides a method for correcting a mutation in a beta globin (HBB) gene in a cell, the method comprising transducing the cell with a replication-defective adeno-associated virus (AAV) comprising (a) an AAV capsid comprising an AAV Clade F capsid protein, and (b) a correction genome comprising: (i) an editing element for editing a target locus in a target gene; (ii) a 5' homology arm nucleotide sequence 5' of the editing element having homology to a first genomic region 5' to the target locus; and (iii) a 3' homology arm nucleotide sequence 3' of the editing element having homology to a second genomic region 3' to the target locus, wherein the cell is transduced without co-transducing or co-administering an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

[0016] In certain embodiments, the cell is a pluripotent stem cell. In certain embodiments, the cell is a hematopoietic stem cell. In certain embodiments, the cell is a $CD34^+$ hematopoietic stem cell. In certain embodiments, the cell is in a mammalian subject and the AAV is administered to the subject in an amount effective to transduce the cell in the subject.

[0017] In another aspect, the instant disclosure provides a method for treating a subject having a disease or disorder associated with an HBB gene mutation, the method comprising (a) transducing an erythrocyte progenitor cell from the subject *ex vivo* with a replication-defective AAV comprising an AAV capsid comprising an AAV Clade F capsid protein; and a correction genome comprising: (i) an editing element for editing a target locus in a target gene; (ii) a 5' homology arm nucleotide sequence 5' of the editing element having homology to a first genomic region 5' to the target locus; and (iii) a 3' homology arm

nucleotide sequence 3' of the editing element having homology to a second genomic region 3' to the target locus; and (b) administering the transduced cell to the subject, wherein the cell is transduced without co-transducing or co-administering an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

[0018] In certain embodiments, the erythrocyte progenitor cell is a pluripotent stem cell. In certain embodiments, the erythrocyte progenitor cell is a hematopoietic stem cell. In certain embodiments, the erythrocyte progenitor cell is $CD34^+$ hematopoietic stem cell.

[0019] In another aspect, the instant disclosure provides a method for treating a subject having a disease or disorder associated with an HBB gene mutation, the method comprising administering to the subject an effective amount of a replication-defective AAV comprising: (a) an AAV capsid comprising an AAV Clade F capsid protein, and (b) a correction genome comprising (i) an editing element for editing a target locus in a target gene; (ii) a 5' homology arm nucleotide sequence 5' of the editing element having homology to a first genomic region 5' to the target locus; and (iii) a 3' homology arm nucleotide sequence 3' of the editing element having homology to a second genomic region 3' to the target locus, wherein the cell is transduced without co-transducing or co-administering an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

[0020] In certain embodiments, the disease or disorder is thalassemia or sickle cell disease. In certain embodiments, the subject is a human subject.

[0021] The following embodiments apply to each of the foregoing aspects.

[0022] In certain embodiments, the target gene is the HBB gene. In certain embodiments, the target locus is at a nucleotide point mutation or deletion in the HBB gene. In certain embodiments, the nucleotide point mutation or deletion in the HBB gene is selected from the group consisting of G at position -87, G at position -31, A at position -30, G at position -29, G at position -28, T at position -10, C at position 1, A at position 1, G at position 2, deletion of C and T at positions 17 and 18, A at position 19, deletion of A at position 20, T at position 20, deletion of A and A at positions 25 and 26, addition of G after position 26, A at position 47, A at position 48, deletion of C at position 51, A at position 52, G at position 58, G at position 59, A at position 79, T at position 82, addition of C after position 84, T at position 93, A at position 93, C at position 97, C at position 98, G at position 202, G at position 208, C at position 222, deletion of T at position 241 or 242, deletion of T and T and C and T at positions 254 to 257, T at position 260, deletion of C at position 264 or 265, addition of A after position 343, deletion of G and T at positions 399 and 400, T at position 401, addition of A after position 417, A at position 446, T at position 1099, A at

position 1293, T at position 1344. In certain embodiments, the editing element comprises a portion of the wild-type HBB gene that corresponds to the mutation.

[0023] In certain embodiments, the editing element comprises the coding regions of one or more exons of an HBB gene. In certain embodiments, the editing element consists of the coding regions of one or more exons of an HBB gene.

[0024] In certain embodiments, the editing element comprises a portion of an HBB gene comprising the coding region of exon 1, the entire intron 1, the entire exon 2, the entire intron 2, and the coding region of exon 3. In certain embodiments, the coding regions have been silently altered to be less than 100%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50% identical to the corresponding exons of the wild-type HBB gene. In certain embodiments, the editing element comprises at least one of the nucleotide sequences selected from the group consisting of SEQ ID NOs: 43-46 and 105-107.

[0025] In certain embodiments, the target locus is AAVS1.

[0026] In certain embodiments, the editing element comprises a coding sequence of the HBB gene or a portion thereof. In certain embodiments, the editing element comprises a nucleotide sequence encoding SEQ ID NO: 48. In certain embodiments, the nucleotide sequence encoding SEQ ID NO: 48 consists of nucleotides 4 to 444 of SEQ ID NO: 27. In certain embodiments, the nucleotide sequence encoding SEQ ID NO: 48 is silently altered to be less than 70%, 75%, 80%, 85%, or 90% identical to nucleotides 4 to 444 of SEQ ID NO: 27. In certain embodiments, the nucleotide sequence encoding SEQ ID NO: 48 consists of the sequence of SEQ ID NO: 47 or 100. In certain embodiments, the editing element comprises a stuffer-inserted coding sequence of the HBB gene.

[0027] In certain embodiments, the target locus is the internucleotide bond between nucleotide 3 and nucleotide 4 of the target gene, whereby integration of the editing element at the target locus results in the target locus comprising an HBB coding sequence or stuffer-inserted coding sequence starting with the start codon of the target gene. In certain embodiments, the editing element comprises an HBB coding sequence or stuffer-inserted coding sequence consisting of 5' to 3' a start codon and the nucleotide sequence encoding SEQ ID NO: 48, or a portion of the HBB coding sequence or stuffer-inserted coding sequence. In certain embodiments, the target locus is in an intron of the target gene, and wherein the editing element comprises 5' to 3' a splice acceptor site, a ribosomal skipping element, and an HBB coding sequence or stuffer-inserted coding sequence. In certain embodiments, the target locus is in intron 1 of the HBB gene. In certain embodiments, the target locus is adjacently 3' to a coding nucleotide of the target gene, and wherein the editing

element comprises 5' to 3' a ribosomal skipping element and an HBB coding sequence or stuffer-inserted coding sequence. In certain embodiments, the target locus is the stop codon of a wild-type target gene (e.g., HBB gene) or the corresponding nucleotides of a mutant target gene (e.g., HBB gene). In certain embodiments, the target locus is the internucleotide bond adjacently 5' to the stop codon of a wild-type target gene (e.g., HBB gene) or the corresponding internucleotide bond of a mutant target gene (e.g., HBB gene).

[0028] In certain embodiments, the 5' homology arm nucleotide sequence is at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the first genomic region. In certain embodiments, the 3' homology arm nucleotide sequence is at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the second genomic region. In certain embodiments, the first genomic region is located in a first editing window, and the second genomic region is located in a second editing window. In certain embodiments, the first and second editing windows are different. In certain embodiments, the first and second editing windows are the same. In certain embodiments, the first editing window consists of the nucleotide sequence set forth in SEQ ID NO: 101, 102, or 103. In certain embodiments, the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 101, 102, or 103. In certain embodiments, the first genomic region consists of the nucleotide sequence set forth in SEQ ID NO: 101. In certain embodiments, the second genomic region consists of the nucleotide sequence set forth in SEQ ID NO: 102. In certain embodiments, the 5' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 101. In certain embodiments, the 3' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 102.

[0029] In certain embodiments, the editing element further comprises an exogenous polyadenylation sequence 3' to the nucleotide sequence encoding SEQ ID NO: 48. In certain embodiments, the editing element further comprises a restriction endonuclease site not present in the target gene. In certain embodiments, the editing element comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 23-28.

[0030] In certain embodiments, each of the 5' and 3' homology arm nucleotide sequences independently has a length of about 100 to about 2000 nucleotides.

[0031] In certain embodiments, the correction genome further comprises a 5' inverted terminal repeat (5' ITR) nucleotide sequence 5' of the 5' homology arm nucleotide sequence, and a 3' inverted terminal repeat (3' ITR) nucleotide sequence 3' of the 3' homology arm nucleotide sequence. In certain embodiments, the 5' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO:18, and the 3' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO:19. In certain embodiments, the 5' ITR nucleotide

sequence has at least 95% sequence identity to SEQ ID NO:20, and the 3' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO:21. In certain embodiments, the correction genome comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 29-42 and 104.

[0032] In certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 95% sequence identity with the amino acid sequence of amino acids 203-736 of SEQ ID NO:2, optionally wherein: the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G.

[0033] In certain embodiments, (a) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G; (b) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; (c) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; (d) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; or (e) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2

is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C.

[0034] In certain embodiments, the capsid protein comprises the amino acid sequence of amino acids 203-736 of SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17.

[0035] In certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 95% sequence identity with the amino acid sequence of amino acids 138-736 of SEQ ID NO:2, optionally wherein: the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 2 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G.

[0036] In certain embodiments, (a) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G; (b) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; (c) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; (d) the

amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; or (e) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C.

[0037] In certain embodiments, the capsid protein comprises the amino acid sequence of amino acids 138-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17.

[0038] In certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 95% sequence identity with the amino acid sequence of amino acids 1-736 of SEQ ID NO:2, optionally wherein: the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 2 is T; the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 68 of SEQ ID NO: 2 is V; the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 2 is L; the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 2 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G.

[0039] In certain embodiments, (a) the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 2 is T, and the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; (b) the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 2 is I, and the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is Y; (c) the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; (d) the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 2 is L, and the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; (e) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G; (f) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; (g) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; (h) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; or (i) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C.

[0040] In certain embodiments, the capsid protein comprises the amino acid sequence of amino acids 1-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

[0041] In certain embodiments, the integration efficiency of the editing element into the target locus is at least 1% when the AAV is contacted in vitro in the absence of an exogenous nuclease with a population of $CD34^+$ human stem cells under standard AAV transduction conditions. In certain embodiments, the allelic frequency of integration of the editing element into the target locus is at least 0.5% when the AAV is contacted in vitro in the absence of an exogenous nuclease with a population of $CD34^+$ human stem cells under standard AAV transduction conditions.

[0042] In another aspect, the instant disclosure provides a pharmaceutical composition comprising an AAV as disclosed herein.

[0043] In another aspect, the instant disclosure provides a packaging system for recombinant preparation of an AAV, wherein the packaging system comprises (a) a Rep nucleotide sequence encoding one or more AAV Rep proteins, (b) a Cap nucleotide sequence encoding one or more AAV Clade F capsid proteins as described herein, and (c) a correction genome as disclosed herein, wherein the packaging system is operative in a cell for enclosing the correction genome in the capsid to form the AAV.

[0044] In certain embodiments, the packaging system comprises a first vector comprising the Rep nucleotide sequence and the Cap nucleotide sequence, and a second vector comprising the correction genome. In certain embodiments, the Rep nucleotide sequence encodes an AAV2 Rep protein. In certain embodiments, the AAV2 Rep protein is 78/68 or Rep 68/52. In certain embodiments, the AAV2 Rep protein comprises an amino acid sequence having a minimum percent sequence identity to the AAV2 Rep amino acid sequence of SEQ ID NO:22, wherein the minimum percent sequence identity is at least 70% across the length of the amino acid sequence encoding the AAV2 Rep protein.

[0045] In certain embodiments, the packaging system further comprises a third vector, wherein the third vector is a helper virus vector. In certain embodiments, the helper virus vector is an independent third vector. In certain embodiments, the helper virus vector is integral with the first vector. In certain embodiments, the helper virus vector is integral with the second vector. In certain embodiments, the third vector comprises genes encoding helper virus proteins.

[0046] In certain embodiments, the helper virus is selected from the group consisting of adenovirus, herpes virus, vaccinia virus, and cytomegalovirus (CMV). In certain embodiments, the helper virus is adenovirus. In certain embodiments, the adenovirus genome comprises one or more adenovirus RNA genes selected from the group consisting of E1, E2, E4 and VA. In certain embodiments, the helper virus is herpes simplex virus (HSV). In certain embodiments, the HSV genome comprises one or more of HSV genes selected from the group consisting of UL5/8/52, ICPO, ICP4, ICP22 and UL30/UL42.

[0047] In certain embodiments, the first vector and the third vector are contained within a first transfecting plasmid. In certain embodiments, the nucleotides of the second vector and the third vector are contained within a second transfecting plasmid. In certain embodiments, the nucleotides of the first vector and the third vector are cloned into a recombinant helper virus. In certain embodiments, the nucleotides of the second vector and the third vector are cloned into a recombinant helper virus.

[0048] In another aspect, the instant disclosure provides a method for recombinant preparation of an AAV, the method comprising introducing a packaging system as described herein into a cell under conditions operative for enclosing the correction genome in the capsid to form the AAV.

BRIEF DESCRIPTION OF THE DRAWINGS

[0049] **Figures 1A and 1B** are graphs showing flow cytometry results of integration of AAVS1-FP into the genome of the GM16265 cells, wherein the AAVS1-FP vector was packaged in AAVHSC7, AAVHSC15, and AAVHSC17.

[0050] **Figure 1C** is a graph showing the percentage of alleles having integration of the FP coding sequence in primary human $CD34^+$ hematopoietic stem cells (HSCs) transduced with the AAVS1-FP vector packaged in AAVHSC17 capsid.

[0051] **Figure 2** depicts the plasmid map of the HBB correction vector hHBB-hL-014 containing a 12 bp linker. In this figure, the black regions of HBB denote nucleotide sequences in the exons encoding a human HBB protein, and the dashed lines between the black regions denote introns between the exons.

[0052] **Figure 3A** is an image of DNA electrophoresis showing editing-specific size of DNA amplified from the genomic DNA of GM16265 cells transduced with the hHBB-hL-014 vector packaged in AAVHSC15 and AAVHSC17 capsids.

[0053] **Figure 3B** is an image of DNA electrophoresis showing editing-specific size of DNA amplified from the genomic DNA of GM16265, GM16266, and GM16267 LCLs transduced with the hHBB-hL-014 vector packaged in AAVHSC17 capsid.

[0054] **Figures 4A, 4B, 4C, and 4D** are vector maps showing the genetic elements between the two AAV ITRs of HBB correction vectors hHBB-hL-001, hHBB-hLW-013, hHBB-hL-011, and hHBB-hLW-012, respectively. In these figures, the black regions of HBB denote nucleotide sequences in the exons encoding a human HBB protein, and the dashed lines between the black regions denote introns between the exons.

[0055] **Figures 5A, 5B, and 5C** are images of DNA electrophoresis showing editing-specific size of DNA amplified from the genomic DNA of primary human $CD34^+$ HSCs transduced with the vectors indicated in the figures packaged in AAVHSC17 capsid.

[0056] **Figure 6** is a graph showing the fraction of edited CD34+ cells across samples as indicated.

[0057] **Figures 7A, 7B, 7C, 7D, and 7E** are vector maps showing the genetic elements between the two AAV ITRs of HBB correction vectors hHBB-hA-009, hHBB-hAW-002, hHBB-h1-010, hHBB-h1W-008, and hHBB-hE3C-001, respectively. In Figures

7A-D, the black regions labeled as "HBB coding region" or "HBB coding region (66%)" denote nucleotide sequences encoding a human HBB protein, either from the start codon to the stop codon (Figures 7C and 7D), or from the second codon to the stop codon (Figures7 and 7B).

[0058] **Figure 8** is an image of DNA electrophoresis showing editing-specific size (1,874 bp) and non-specific size (1,180 bp) of DNA amplified from genomic DNA of RKO and GM16265 LCL cells transduced with the vectors indicated in the figures packaged in AAVHSC7 capsid.

[0059] **Figure 9** is a graph showing the percentage of alleles having integration of the FP coding sequence in blood, bone marrow ("BM"), and spleen cells from NSG mice xenografted with human HSCs following administration of the AAVS1-FP vector packaged in AAVHSC7 and AAVHSC17 capsids.

DETAILED DESCRIPTION

[0060] The instant disclosure provides adeno-associated virus (AAV) compositions for correcting a mutation in an HBB gene and methods of using the same to correct an HBB gene mutation in a cell. Also provided are packaging systems for making the adeno-associated virus compositions.

I. Definitions

[0061] As used herein, the term "replication-defective adeno-associated virus" refers to an AAV comprising a genome lacking Rep and Cap genes.

[0062] As used herein, the term "HBB gene" refers to a wild-type or mutant human beta globin gene, including but not limited to the coding regions, exons, introns, 5' UTR, 3' UTR, and transcriptional regulatory regions of the HBB gene.

[0063] As used herein, the term "correcting a mutation in an HBB gene" refers to the insertion, deletion, or substitution of one or more nucleotides at a target locus in a target gene (e.g., a mutant HBB gene) to create a locus that is capable of expressing a wild-type HBB protein or a functional equivalent thereof. In certain embodiments, "correcting a mutation in an HBB gene" involves reverting a mutation in an HBB gene back to the wild-type sequence. In certain embodiments, "correcting a mutation in an HBB gene" involves inserting a nucleotide sequence encoding at least a portion of a wild-type beta globin protein or a functional equivalent thereof into the target gene (e.g., the mutant HBB gene), such that a wild-type beta globin protein or a functional equivalent thereof is expressed from the locus of the target gene (e.g., the mutant HBB gene locus), optionally under the control of an endogenous target gene promoter (e.g., HBB gene promoter). As used herein, a "functional

equivalent" refers to a product of a gene or fragment thereof that can function as a wild-type beta globin. In certain embodiments, a functional equivalent of HBB can include other globin genes or pseudogenes, such as epsilon globin (HBE), delta globin (HBD), gamma globin 1 (HBG1), gamme globin 2 (HBG2), and HBB pseudogene HBBP. In certain embodiments, a functional equivalent of HBB can be a modified beta globin protein, wherein the modification confers at least one characteristic not found in wild-type beta globin, *e.g.*, the ability to inhibit aggregation of beta globin carrying an SCD mutation.

[0064] As used herein, the term "correction genome" refers to a recombinant AAV genome that is capable of integrating an editing element (*e.g.*, one or more nucleotides or an internucleotide bond) via homologous recombination into a target locus to correct a genetic defect in an HBB gene. In certain embodiments, the target locus is in the human HBB gene. The skilled artisan will appreciate that the portion of a correction genome comprising the 5' homology arm, editing element, and 3' homology arm can be in the sense or antisense orientation relative to the target locus (*e.g.*, the human HBB gene).

[0065] As used herein, the term "editing element" refers to the portion of a correction genome that when integrated at a target locus modifies the target locus. An editing element can mediate insertion, deletion or substitution of one or more nucleotides at the target locus.

[0066] As used herein, the term "target locus" refers to a region of a chromosome or an internucleotide bond that is modified by an editing element. In certain embodiments, the target locus is a region or an internucleotide bond in the HBB gene, optionally wherein the target locus comprises at least one genetic mutation that compromises the expression or function of beta globin protein. In certain embodiments, the target locus is AAVS1. The AAVS1 locus is on chromosome 19 qter13.3-13.4, nucleotides 55,090,913 to 55,117,600 of NCBI Reference Sequence No. NC_000019.10, as described in Giraud et al., Proc Natl Acad Sci U S A. (1994) 91(21):10039-43; Linden et al., Proc Natl Acad Sci U S A. (1996) 93(21):11288-94; and Linden et al., Proc Natl Acad Sci U S A. (1996) 93(15):7966-72, each of which is incorporated herein by reference in its entirety. In certain embodiments, the target locus is a safe harbor locus. Safe harbor loci are sites in the genome able to accommodate the integration of new genetic material in a manner that ensures that the newly inserted genetic material: (1) function predictably; and (2) do not cause alterations of the host genome that may pose a risk to the host cell or organism. Accordingly, in certain embodiments, the target locus may be any safe harbor locus known in the art that can support predictable transgene expression while minimizing the risk of unwanted interactions with the host genome.

[0067] As used herein, the term "target gene" refers to a gene in which a target locus or a portion thereof is located. In certain embodiments, the target locus is fully in a target gene. In certain embodiments, the target gene is HBB. In certain embodiments, the target gene is human PPP1R12C. In certain embodiments, the target gene is expressed in an erythrocyte progenitor.

[0068] As used herein, the term "homology arm" refers to a portion of a correction genome positioned 5' or 3' of an editing element that is substantially identical to the genome flanking a target locus. In certain embodiments, the target locus is in a human HBB gene, and the homology arm comprises a sequence substantially identical to the genome flanking the target locus.

[0069] As used herein, the term "Clade F capsid protein" refers to an AAV VP1, VP2, or VP3 capsid protein that comprises an amino acid sequence having at least 90% identity with the VP1, VP2, or VP3 amino acid sequences set forth, respectively, in amino acids 1-736, 138-736, and 203-736 of SEQ ID NO:1 herein. As used herein, the identity between two nucleotide sequences or between two amino acid sequences is determined by the number of identical nucleotides or amino acids in alignment divided by the full length of the longer nucleotide or amino acid sequence.

[0070] As used herein, the term "a disease or disorder associated with an HBB gene mutation" refers to any disease or disorder caused by, exacerbated by, or genetically linked with variation of an HBB gene. In certain embodiments, the disease or disorder associated with an HBB gene mutation is a hemoglobinopathy, such as sickle cell disease or beta thalassemia.

[0071] As used herein, the term "silently altered" refers to alteration of a coding sequence or a stuffer-inserted coding sequence of a gene (*e.g.*, by nucleotide substitution) without changing the amino acid sequence of the polypeptide encoded by the coding sequence or stuffer-inserted coding sequence. Such silent alteration is advantageous in that it reduces the likelihood of integration of the correction genome into loci of other genes or pseudogenes paralogous to the target gene (*e.g.*, another globin gene locus or a beta globin pseudogene locus). Such silent alteration also reduces the homology between the editing element and the target gene, thereby reducing undesired integration mediated by the editing element rather than by a homology arm.

[0072] As used herein, the term "coding sequence" refers to the portion of a complementary DNA (cDNA), or a silently altered sequence thereof, that encodes a polypeptide, starting at a start codon and ending at a stop codon. A gene may have one or

more wild-type coding sequences due to alternative splicing and/or alternative translation initiation. An exemplary wild-type HBB coding sequence is set forth in nucleotides 51-494 of the NCBI Reference Sequence: NM_000518.4.

[0073] As used herein, the term "coding nucleotide" refers to a nucleotide of a gene that corresponds to a nucleotide in a coding sequence of the gene, except the 3' nucleotide of the stop codon. Accordingly, in certain embodiments, a coding nucleotide of the HBB gene is any one of nucleotides 1-443 of the HBB gene.

[0074] As used herein, the term "stuffer-inserted coding sequence" of a gene refers to a nucleotide sequence comprising one or more introns inserted in a coding sequence of the gene. In certain embodiments, at least one of the introns is a nonnative intron, i.e., having a sequence different from a native intron of the gene. In certain embodiments, all of the introns in the stuffer-inserted coding sequence are nonnative introns. A nonnative intron can have the sequence of an intron from a different species or the sequence of an intron in a different gene from the same species. Alternatively or additionally, at least a portion of a nonnative intron sequence can be synthetic. A skilled worker will appreciate that nonnative intron sequences can be designed to mediate RNA splicing by introducing any consensus splicing motifs known in the art. Exemplary consensus splicing motifs are provided in Sibley et al., (2016) Nature Reviews Genetics, 17, 407-21, which is incorporated by reference herein in its entirety. Insertion of a nonnative intron may promote the efficiency and robustness of vector packaging, as stuffer sequences allow adjustments of the vector to reach an optimal size (e.g., 4.5-4.8 kb). In certain embodiments, at least one of the introns is a native intron of the gene. In certain embodiments, all of the introns in the stuffer-inserted coding sequence are native introns of the gene. The nonnative or native introns can be inserted at any internucleotide bonds in the coding sequence. In certain embodiments, one or more nonnative or native introns are inserted at internucleotide bonds predicted to promote efficient splicing (*see e.g.*, Zhang (1998) Human Molecular Gentics, 7(5):919-32, which is incorporated by reference herein in its entirety). In certain embodiments, one or more nonnative or native introns are inserted at internucleotide bonds that link two endogenous exons.

[0075] As used herein, the term "ribosomal skipping element" refers to a nucleotide sequence encoding a short peptide sequence capable of causing generation of two peptide chains from translation of one mRNA molecule. As used herein, the term "ribosomal skipping peptide" refers to a peptide encoded by a ribosomal skipping element. In certain embodiments, the ribosomal skipping peptide comprises a consensus motif of $X_1X_2EX_3NPGP$, wherein X_1 is D or G, X_2 is V or I, and X_3 is any amino acid (SEQ ID NO:

49). In certain embodiments, the ribosomal skipping peptide is selected from the group consisting of thosea-asigna virus 2A peptide (T2A), porcine teschovirus-1 2A peptide (P2A), foot-and-mouth disease virus 2A peptide (F2A), equine rhinitis A virus 2A peptide (E2A), cytoplasmic polyhedrosis virus 2A peptide (BmCPV 2A), and flacherie virus of *B. mori* 2A peptide (BmIFV 2A). Exemplary amino acid sequences of T2A peptide and P2A peptide are set forth in SEQ ID NOs: 71 and 73, respectively. Exemplary nucleotide sequences of T2A element and P2A element are set forth in SEQ ID NOs: 72 and 74, respectively. In certain embodiments, the ribosomal skipping element encodes a peptide that further comprises a sequence of Gly-Ser-Gly at the N terminus, optionally wherein the sequence of Gly-Ser-Gly is encoded by the nucleotide sequence of GGCAGCGGA (SEQ ID NO: 75). While not wishing to be bound by theory, it is hypothesized that ribosomal skipping elements function by: terminating translation of the first peptide chain and re-initiating translation of the second peptide chain; or by cleavage of a peptide bond in the ribosomal skipping peptide by an intrinsic protease activity of the encoded peptide or by another protease in the environment (e.g., cytosol).

[0076] As used herein, the term "polyadenylation sequence" refers to a DNA sequence that when transcribed into RNA constitutes a polyadenylation signal sequence.

[0077] In the instant disclosure, nucleotide positions in a gene are specified relative to the first nucleotide of the start codon. The first nucleotide of a start codon is position 1, nucleotides 5' to the first nucleotide of a start codon have negative numbers, and the nucleotides 3' to the first nucleotide of a start codon have positive numbers. For example, nucleotide 1 of the HBB gene as used herein is nucleotide 70,595 of the NCBI Reference Sequence: NG_000007.3. The nucleotide adjacently 5' to the start codon is nucleotide -1.

[0078] In the instant disclosure, exons and introns in a gene are specified relative to the exon encompassing the first nucleotide of the start codon. The exon encompassing the first nucleotide of the start codon is exon 1. Exons 3' to exon 1 are from 5' to 3': exon 2, exon 3, etc. Introns 3' to exon 1 are from 5' to 3': intron 1, intron 2, etc. Accordingly, a gene comprises from 5' to 3': exon 1, intron 1, exon 2, intron 2, exon 3, etc. An exemplary exon 1 of the human HBB gene is nucleotides 70,545 to 70,686 of the NCBI Reference Sequence: NG_000007.3. An exemplary intron 1 of the human HBB gene is nucleotides 70,687 to 70,816 of the NCBI Reference Sequence: NG_000007.3. A skilled artisan will appreciate that a gene may be transcribed into multiple different mRNAs. As such, a gene (*e.g.*, HBB) may have multiple different sets of exons and introns.

[0079] As used herein, the term "integration" refers to introduction of an editing element into a target locus by homologous recombination between a correction genome and the target gene. Integration of an editing element can result in substitution, insertion and/or deletion of one or more nucleotides in the target gene.

[0080] As used herein, the term "integration efficiency of the editing element into the target locus" refers to the percentage of cells in a transduced population in which integration of the editing element into the target locus has occurred.

[0081] As used herein, the term "allelic frequency of integration of the editing element into the target locus" refers to the percentage of alleles in a population of transduced cells in which integration of the editing element into the target locus has occurred.

[0082] As used herein, the term "standard AAV transduction conditions" refers to transduction of 2 x 10^5 $CD34^+$ human stem cells with an AAV at a multiplicity of infection (MOI) of 1.5 x 10^5, wherein the cells are cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 20% Fetal Calf Serum (FCS), 100 μg/mL streptomycin, 100 U/mL penicillin, 2mmol/L L-glutamine, 10 ng/mL human IL-3, 10 ng/mL human IL-6, and 1 ng/mL human SCF, at 37 °C in an incubator environment of 5% carbon dioxide (CO_2), wherein the AAV is formulated in phosphate buffered saline (PBS), and wherein the AAV is added to the cell culture medium containing the $CD34^+$ cells in a volume that is less than or equal to 1/9th of the volume of the culture medium.

[0083] As used herein, the term "effective amount" in the context of the administration of an AAV to a subject refers to the amount of the AAV that achieves a desired prophylactic or therapeutic effect.

[0084] As used herein, the term "an erythrocyte progenitor" refers to a cell capable of differentiating into an erythrocyte. In certain embodiments, the erythrocyte progenitor is a pluripotent stem cell. In certain embodiments, the erythrocyte progenitor is an induced pluripotent stem cell. In certain embodiments, the erythrocyte progenitor is a hematopoietic stem cell. In certain embodiments, the erythrocyte progenitor is a $CD34^+$ hematopoietic stem cell. In certain embodiments, the erythrocyte progenitor is a myeloid progenitor cell. In certain embodiments, the erythrocyte progenitor is a megakaryocyte-erythroid progenitor cell. In certain embodiments, the erythrocyte progenitor is an erythroid precursor cell.

II. Adeno-Associated Virus Compositions

[0085] In one aspect, the disclosure provides novel replication-defective AAV compositions useful for correcting mutations in an HBB gene. The AAV disclosed herein

generally comprise: an AAV capsid comprising an AAV Clade F capsid protein; and a correction genome for editing a target locus in an HBB gene.

[0086] Any AAV Clade F capsid protein or derivative thereof can be used in the AAV compositions disclosed herein. For example, in certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 99% sequence identity with the amino acid sequence of amino acids 203-736 of SEQ ID NO:2. In certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 99% sequence identity with the amino acid sequence of amino acids 203-736 of SEQ ID NO:2, wherein: the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid

346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C. In certain embodiments, the AAV Clade F capsid protein comprises the amino acid sequence of amino acids 203-736 of SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17.

[0087] For example, in certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 99% sequence identity with the amino acid sequence of amino acids 138-736 of SEQ ID NO:2. In certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 80%, 85%, 90%, 95%, or 99% sequence identity with the amino acid sequence of amino acids 138-736 of SEQ ID NO:2, wherein: the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 2 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid

464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C. In certain embodiments, the AAV Clade F capsid protein comprises the amino acid sequence of amino acids 138-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17.

[0088] For example, in certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 90%, 95%, or 99% sequence identity with the amino acid sequence of amino acids 1-736 of SEQ ID NO:2. In certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 90%, 95%, or 99% sequence identity with the amino acid sequence of amino acids 1-736 of SEQ ID NO:2, wherein: the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 2 is T; the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 68 of SEQ ID NO: 2 is V; the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 2 is L; the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 2 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is

R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 2 is T, and the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 2 is I, and the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is Y. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 2 is L, and the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C. In certain embodiments, the AAV Clade F capsid protein comprises the amino acid sequence of amino acids 1-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

[0089] In certain embodiments, the AAV capsid comprises two or more of: (a) a Clade F capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17. In certain embodiments, the AAV capsid comprises: (a) a Clade F capsid protein having an amino acid sequence consisting of amino acids 203-736 of SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17; (b) a Clade F capsid protein having an amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17; and (c) a Clade F capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

[0090] In certain embodiments, the AAV capsid comprises one or more of: (a) a Clade F capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 8; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 8; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 8. In certain embodiments, the AAV capsid comprises two or more of: (a) a Clade F capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 8; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 8; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 8. In certain embodiments, the AAV capsid comprises: (a) a Clade F capsid protein having an amino acid sequence consisting of amino acids 203-736 of SEQ ID NO: 8; (b) a Clade F capsid protein having an amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 8; and (c) a Clade F capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 8.

[0091] In certain embodiments, the AAV capsid comprises one or more of: (a) a Clade F capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 13; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 13; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 13. In certain embodiments, the AAV capsid comprises two or more of: (a) a Clade F capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 13; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 13; and (c) a Clade F capsid

protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 13. In certain embodiments, the AAV capsid comprises: (a) a Clade F capsid protein having an amino acid sequence consisting of amino acids 203-736 of SEQ ID NO: 13; (b) a Clade F capsid protein having an amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 13; and (c) a Clade F capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 13.

[0092] In certain embodiments, the AAV capsid comprises one or more of: (a) a Clade F capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 16; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 16; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 16. In certain embodiments, the AAV capsid comprises two or more of: (a) a Clade F capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 16; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 16; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 16. In certain embodiments, the AAV capsid comprises: (a) a Clade F capsid protein having an amino acid sequence consisting of amino acids 203-736 of SEQ ID NO: 16; (b) a Clade F capsid protein having an amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 16; and (c) a Clade F capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 16.

[0093] Correction genomes useful in the AAV compositions disclosed herein generally comprise: (i) an editing element for editing a target locus in a target gene, (ii) a 5' homology arm nucleotide sequence 5' of the editing element having homology to a first genomic region 5' to the target locus, and (iii) a 3' homology arm nucleotide sequence 3' of the editing element having homology to a second genomic region 3' to the target locus. In certain embodiments, the correction genome comprises a 5' inverted terminal repeat (5' ITR) nucleotide sequence 5' of the 5' homology arm nucleotide sequence, and a 3' inverted terminal repeat (3' ITR) nucleotide sequence 3' of the 3' homology arm nucleotide sequence.

[0094] Editing elements used in the correction genomes disclosed herein can mediate insertion, deletion, or substitution of one or more nucleotides at the target locus. The target locus can locate fully or partially in a target gene, which can be the HBB gene or another gene expressed in an erythrocyte progenitor.

[0095] In certain embodiments, when correctly integrated by homologous recombination at the target locus, the editing element corrects a mutation in an HBB gene

back to the wild-type HBB sequence or to a silently altered sequence that encodes the wild-type HBB protein or a functional equivalent thereof. Most mutations in the HBB gene can be corrected by an editing element as disclosed herein. In certain embodiments, the editing element is one or more nucleotides that correct a substitution or deletion mutation in the HBB gene. In certain embodiments, the editing element is an internucleotide bond that deletes an insertion mutation in the HBB gene. In certain embodiments, the editing element comprises one or more coding exons of an HBB gene. For example, the editing element can comprise a portion of an HBB gene that encompasses the coding region of exon 1, the entire intron 1, the entire exon 2, the entire intron 2, and the coding region of exon 3. The exons can be wild-type or silently altered as disclosed herein.

[0096] In certain embodiments, the editing element comprises at least a portion of a coding sequence or stuffer-inserted coding sequence of an HBB gene. In certain embodiments, the editing element comprises all or substantially all of a coding sequence or stuffer-inserted coding sequence of an HBB gene. For example, in certain embodiments, the editing element comprises nucleotides 4 to 444 of an HBB coding sequence, or a portion of an HBB stuffer-inserted coding sequence from nucleotide 4 to the stop codon, optionally further comprising an exogenous polyadenylation sequence 3' to the portion of HBB coding sequence or stuffer-inserted coding sequence. In certain embodiments, the editing element comprises a nucleotide sequence encoding SEQ ID NO: 48, and can optionally further comprise an exogenous polyadenylation sequence 3' to the nucleotide sequence encoding SEQ ID NO: 48. In certain embodiments, such editing elements can be integrated into exon 1 immediately after the endogenous start codon of the HBB gene (e.g., between nucleotide 3 and nucleotide 4 of the HBB gene) by homologous recombination, whereby integration of the editing element results in generation of a complete HBB coding sequence in-frame with the start codon of the endogenous HBB gene. In certain embodiments, such editing elements can be integrated into exon 1 immediately after the endogenous start codon of a non-HBB target gene (e.g., between nucleotide 3 and nucleotide 4 of the target gene) by homologous recombination, whereby integration of the editing element results in generation of a complete HBB coding sequence in-frame with the start codon of the endogenous target gene. The portion of HBB coding sequence in the editing element can be wild-type or silently mutated as disclosed herein. The portion of HBB amino acid sequence encoded by the editing element can be wild-type or a functional equivalent thereof.

[0097] In certain embodiments, the editing element comprises at least a portion of an HBB coding sequence or stuffer-inserted coding sequence (e.g., a complete HBB coding

sequence or stuffer-inserted coding sequence), and a ribosomal skipping element or an exogenous polyadenylation sequence. In certain embodiments, the editing element comprises 5' to 3' a ribosomal skipping element, and at least a portion of an HBB coding sequence or stuffer-inserted coding sequence (e.g., a complete HBB coding sequence or stuffer-inserted coding sequence). In certain embodiments, the editing element comprises 5' to 3': a ribosomal skipping element; a complete HBB coding sequence or stuffer-inserted coding sequence; and an exogenous polyadenylation sequence. In certain embodiments, the aforementioned editing element can be integrated adjacently 3' to a coding nucleotide of a target gene (e.g., adjacently 5' to the stop codon of the native HBB gene) by homologous recombination to produce a recombinant target gene comprising 5' to 3': a 5' portion of the target gene ending at the coding nucleotide, a ribosomal skipping element, a complete HBB coding sequence or stuffer-inserted coding sequence, and an exogenous polyadenylation sequence, wherein the ribosomal skipping element is positioned such that it is in frame with the coding region of the target gene and the complete HBB coding sequence or stuffer-inserted coding sequence. Expression of this recombinant target gene produces a first polypeptide comprising the amino acid sequence encoded by an N-terminal portion of the target gene fused to a first portion of the encoded ribosomal skipping peptide, and a second polypeptide comprising a second portion of the encoded ribosomal skipping peptide (e.g., a single proline residue) fused to the complete HBB amino acid sequence. The HBB coding sequence or the coding regions of the HBB stuffer-inserted coding sequence in the editing element can be wild-type or silently mutated as disclosed herein. The HBB amino acid sequence encoded by the editing element can be wild-type or a functional equivalent thereof (e.g., lacking the first methionine). The target locus can be an internucleotide bond or a nucleotide sequence adjacently 3' to a coding nucleotide of the target gene. In certain embodiments, the target locus consists of the native stop codon of the HBB gene.

[0098] In certain embodiments, the editing element comprises at least a portion of an HBB coding sequence or stuffer-inserted coding sequence (e.g., a complete HBB coding sequence or stuffer-inserted coding sequence), and one or more of: a splice acceptor site; a splice donor site; a ribosomal skipping element; and an exogenous polyadenylation sequence. In certain embodiments, the editing element comprises 5' to 3': a splice acceptor site; a ribosomal skipping element; and at least a portion of an HBB coding sequence or stuffer-inserted coding sequence (e.g., a complete HBB coding sequence or stuffer-inserted coding sequence). In certain embodiments, the editing element comprises 5' to 3': a splice acceptor site; a ribosomal skipping element; a complete HBB coding sequence or stuffer-inserted

coding sequence; and an exogenous polyadenylation sequence. In certain embodiments, the aforementioned editing element can be integrated into an intron of a target gene (e.g., intron 1 of the endogenous HBB gene) by homologous recombination to produce a recombinant HBB gene comprising 5' to 3': the one or more exons 5' to the intron of the target gene; a 5' portion of the intron including the endogenous splice donor site; a splice acceptor site; a ribosomal skipping element, a complete HBB coding sequence or stuffer-inserted coding sequence; and an exogenous polyadenylation sequence, wherein the ribosomal skipping element is positioned such that it is in frame with the complete HBB coding sequence or stuffer-inserted coding sequence and such that splicing of the splice acceptor site to the endogenous splice donor site of the target gene places it in frame with the coding region of the target gene. Expression of this recombinant target gene produces a first polypeptide comprising the target gene amino acid sequence encoded by the endogenous exon(s) 5' to the insertion site fused to a first portion of the encoded ribosomal skipping peptide, and a second polypeptide comprising a second portion of the encoded ribosomal skipping peptide (e.g., a single proline residue) fused to a complete HBB amino acid sequence. The HBB coding sequence or stuffer-inserted coding sequence in the editing element can be wild-type or silently mutated as disclosed herein. The HBB amino acid sequence encoded by the editing element can be wild-type or a functional equivalent thereof (e.g., lacking the first methionine). The target locus can be an internucleotide bond or a nucleotide sequence adjacently 3' to a nucleotide in an intron of the target gene.

[0099] In certain embodiments, one or more portions of an HBB coding sequence or stuffer-inserted coding sequence within an editing element can be silently altered to be non-identical to the corresponding exons of the wild-type HBB gene. Such silent alteration is advantageous in that it reduces the likelihood of integration of the correction genome into loci of other globin genes or pseudogenes, *e.g.*, a beta globin pseudogene locus. Such silent alteration also reduces the homology between the editing element and the target gene, thereby reducing undesired integration mediated by the editing element rather than by a homology arm.

[00100] Accordingly, in certain embodiments, the editing element comprises coding regions of one or more exons of an HBB gene that have been silently altered to be less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding coding regions of exons of the wild-type HBB gene. In certain embodiments, the editing element comprises coding regions of one or more exons of an HBB gene that have been silently altered to be less than 70% identical to the corresponding coding

regions of exons of the wild-type HBB gene. In certain embodiments, the editing element comprises coding regions of one or more exons of an HBB gene that have been silently altered to be less than 85% identical to the corresponding coding regions of exons of the wild-type HBB gene.

[00101] In certain embodiments, the editing element comprises a portion of an HBB gene encompassing the coding region of exon 1, the entire intron 1, the entire exon 2, the entire intron 2, and the coding region of exon 3, wherein one or more of the coding region of exon 1, the entire exon 2, and the coding region of exon 3 has been silently altered to be less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding regions of exons of the wild-type HBB gene. In certain embodiments, the editing element comprises a portion of an HBB gene encompassing the coding region of exon 1, the entire intron 1, the entire exon 2, the entire intron 2, and the coding region of exon 3, wherein one or more of the coding region of exon 1, the entire exon 2, and the coding region of exon 3 has been silently altered to be less than 70% identical to the corresponding regions of exons of the wild-type HBB gene.

[00102] In certain embodiments, the editing element comprises a portion of an HBB gene encompassing the coding region of exon 1, the entire intron 1, the entire exon 2, the entire intron 2, and the coding region of exon 3, wherein each one of the coding region of exon 1, the entire exon 2, and the coding region of exon 3 has been silently altered to be less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding regions of exons of the wild-type HBB gene. In certain embodiments, the editing element comprises a portion of an HBB gene encompassing the coding region of exon 1, the entire intron 1, the entire exon 2, the entire intron 2, and the coding region of exon 3, wherein each one of the coding region of exon 1, the entire exon 2, and the coding region of exon 3 has been silently altered to be less than 70% identical to the corresponding regions of exons of the wild-type HBB gene.

[00103] In certain embodiments, the editing element comprises one or more of the nucleotide sequences selected from the group consisting of SEQ ID NOs: 43-46 and 105-107. In certain embodiments, the editing element comprises two or more of the nucleotide sequences selected from the group consisting of SEQ ID NOs: 43-46 and 105-107. In certain embodiments, the editing element comprises the nucleotide sequence set forth in SEQ ID NO: 46. In certain embodiments, the editing element comprises the nucleotide sequences set forth in SEQ ID NOs: 43, 44, and 45. In certain embodiments, the editing element comprises the nucleotide sequences set forth in SEQ ID NOs: 105, 106, and 107.

[00104] In certain embodiments, the editing element comprises at least a portion of an HBB coding sequence that has been silently altered to be less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding portion of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises at least a portion of an HBB coding sequence that has been silently altered to be less than 70% identical to the corresponding portion of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises at least a portion of an HBB coding sequence that has been silently altered to be less than 85% identical to the corresponding portion of the wild-type HBB coding sequence.

[00105] In certain embodiments, the editing element comprises at least a portion of an HBB stuffer-inserted coding sequence that has been silently altered, such that the at least a portion of HBB stuffer-inserted coding sequence can be transcribed and spliced into a coding sequence less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding portion of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises at least a portion of an HBB stuffer-inserted coding sequence that has been silently altered, such that the at least a portion of HBB stuffer-inserted coding sequence can be transcribed and spliced into a coding sequence less than 70% identical to the corresponding portion of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises at least a portion of an HBB stuffer-inserted coding sequence that has been silently altered, such that the at least a portion of HBB stuffer-inserted coding sequence can be transcribed and spliced into a coding sequence less than 85% identical to the corresponding portion of the wild-type HBB coding sequence.

[00106] In certain embodiments, the editing element comprises nucleotides 4 to 444 of an HBB coding sequence that has been silently altered to be less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding portion of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises nucleotides 4 to 444 of an HBB coding sequence that has been silently altered to be less than 70% identical to the corresponding portion of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises the nucleotide sequence set forth in SEQ ID NO: 47. In certain embodiments, the editing element comprises nucleotides 4 to 444 of an HBB coding sequence that has been silently altered to be less than 85% identical to the corresponding portion of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises the nucleotide sequence set forth in SEQ ID NO: 100. Such editing elements can further comprise an exogenous polyadenylation

sequence 3' to the HBB gene coding sequence. In certain embodiments, the editing element comprises 5' to 3': nucleotides 4 to 444 of an HBB coding sequence; and an exogenous polyadenylation sequence, wherein the nucleotides 4 to 444 of an HBB coding sequence have been silently altered to be less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding region of the wild-type HBB coding sequence.

[00107] In certain embodiments, the editing element comprises a portion of HBB stuffer-inserted coding sequence that has been silently altered, such that the portion of HBB stuffer-inserted coding sequence can be transcribed and spliced into nucleotides 4 to 444 of an HBB coding sequence that is less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding portion of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises a portion of HBB stuffer-inserted coding sequence that has been silently altered, such that the portion of HBB stuffer-inserted coding sequence can be transcribed and spliced into nucleotides 4 to 444 of an HBB coding sequence that is less than 70% identical to the corresponding portion of the wild-type HBB coding sequence. In certain embodiments, the portion of HBB stuffer-inserted coding sequence can be transcribed and spliced into the nucleotide sequence set forth in SEQ ID NO: 47. In certain embodiments, the editing element comprises a portion of HBB stuffer-inserted coding sequence that has been silently altered, such that the portion of HBB stuffer-inserted coding sequence can be transcribed and spliced into nucleotides 4 to 444 of an HBB coding sequence that is less than 85% identical to the corresponding portion of the wild-type HBB coding sequence. In certain embodiments, the portion of HBB stuffer-inserted coding sequence can be transcribed and spliced into the nucleotide sequence set forth in SEQ ID NO: 100. Such editing elements can further comprise an exogenous polyadenylation sequence 3' to the HBB gene coding sequence. In certain embodiments, the editing element comprises 5' to 3': a portion of HBB stuffer-inserted coding sequence that has been silently altered; and an exogenous polyadenylation sequence, wherein the portion of HBB stuffer-inserted coding sequence can be transcribed and spliced into nucleotides 4 to 444 of an HBB coding sequence that is less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding region of the wild-type HBB coding sequence.

[00108] In certain embodiments, the editing element comprises a complete HBB coding sequence that has been silently altered to be less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the wild-type HBB coding

sequence. In certain embodiments, the editing element comprises a complete HBB coding sequence that has been silently altered to be less than 70% identical to the wild-type HBB coding sequence. In certain embodiments, the editing element comprises 5' to 3' a start codon and the nucleotide sequence set forth in SEQ ID NO: 47. In certain embodiments, the editing element comprises the nucleotide sequence set forth in SEQ ID NO: 28. In certain embodiments, the editing element comprises a complete HBB coding sequence of an HBB gene that has been silently altered to be less than 85% identical to the wild-type HBB coding sequence. In certain embodiments, the editing element comprises 5' to 3' a start codon and the nucleotide sequence set forth in SEQ ID NO: 100. In certain embodiments, the editing element comprises the nucleotide sequence set forth in SEQ ID NO: 99. Such editing elements can further comprise one or more of: a splice acceptor site; a ribosomal skipping element; and an exogenous polyadenylation sequence 3' to the HBB gene coding sequence. In certain embodiments, the editing element comprises 5' to 3': a ribosomal skipping element; and a complete HBB coding sequence optionally lacking the start codon, wherein the complete HBB coding sequence has been silently altered to be less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding region of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises 5' to 3': a ribosomal skipping element; a complete HBB coding sequence optionally lacking the start codon; and an exogenous polyadenylation sequence, wherein the complete HBB coding sequence has been silently altered to be less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding region of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises 5' to 3': a splice acceptor site; a ribosomal skipping element; and a complete HBB coding sequence optionally lacking the start codon, wherein the complete HBB coding sequence has been silently altered to be less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding region of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises 5' to 3': a splice acceptor site; a ribosomal skipping element; a complete HBB coding sequence optionally lacking the start codon; and an exogenous polyadenylation sequence, wherein the complete HBB coding sequence has been silently altered to be less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding region of the wild-type HBB coding sequence.

[00109] In certain embodiments, the editing element comprises an HBB stuffer-inserted coding sequence that has been silently altered, such that the HBB stuffer-inserted

coding sequence can be transcribed and spliced into a complete HBB coding sequence that is less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the wild-type HBB coding sequence. In certain embodiments, the editing element comprises an HBB stuffer-inserted coding sequence that has been silently altered, such that the HBB stuffer-inserted coding sequence can be transcribed and spliced into a complete HBB coding sequence that is less than 70% identical to the wild-type HBB coding sequence. In certain embodiments, the complete HBB coding sequence consists of 5' to 3' a start codon and the nucleotide sequence set forth in SEQ ID NO: 47. In certain embodiments, the complete HBB coding sequence consists of the nucleotide sequence set forth in SEQ ID NO: 28. In certain embodiments, the HBB stuffer-inserted coding sequence comprises the nucleotide sequences set forth in SEQ ID NOs: 43-45. In certain embodiments, the editing element comprises an HBB stuffer-inserted coding sequence that has been silently altered, such that the HBB stuffer-inserted coding sequence can be transcribed and spliced into a complete HBB coding sequence that is less than 85% identical to the wild-type HBB coding sequence. In certain embodiments, the HBB stuffer-inserted coding sequence comprises 5' to 3' the nucleotide sequences set forth in SEQ ID NOs: 105, 106, and 107. Such editing elements can further comprise one or more of: a splice acceptor site; a ribosomal skipping element; and an exogenous polyadenylation sequence 3' to the HBB gene coding sequence. In certain embodiments, the editing element comprises 5' to 3': a ribosomal skipping element; and an HBB stuffer-inserted coding sequence that has been silently altered, wherein the HBB stuffer-inserted coding sequence can be transcribed and spliced into a complete HBB coding sequence that is less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding region of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises 5' to 3': a ribosomal skipping element; an HBB stuffer-inserted coding sequence that has been silently altered; and an exogenous polyadenylation sequence, wherein the HBB stuffer-inserted coding sequence can be transcribed and spliced into a complete HBB coding sequence that is less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding region of the wild-type HBB coding sequence. In certain embodiments, the editing element comprises 5' to 3': a splice acceptor site; a ribosomal skipping element; and an HBB stuffer-inserted coding sequence that has been silently altered, wherein the HBB stuffer-inserted coding sequence can be transcribed and spliced into a complete HBB coding sequence that is less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding region of the wild-type

HBB coding sequence. In certain embodiments, the editing element comprises 5' to 3': a splice acceptor site; a ribosomal skipping element; an HBB stuffer-inserted coding sequence that has been silently altered; and an exogenous polyadenylation sequence, wherein the HBB stuffer-inserted coding sequence can be transcribed and spliced into a complete HBB coding sequence that is less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to the corresponding region of the wild-type HBB coding sequence.

[00110] Any and all of the editing elements disclosed herein can further include a unique sequence not present in the target gene (e.g., wild-type or mutant HBB gene, or a gene encoding a functional equivalent thereof), thereby allowing for identification of cells that have integration of the editing element at the target locus. Such unique sequence can be a suitable sequence for nucleic acid sequencing analysis (e.g., by PCR or next-generation sequencing) of the target locus and its flanking regions or a nucleic acid amplified therefrom. Such unique sequence can also be a restriction endonuclease site that allows for identification of cells that have integration of the editing element at the target locus based upon restriction fragment length polymorphism analysis of the target locus and its flanking regions or a nucleic acid amplified therefrom.

[00111] Any and all of the editing elements disclosed herein can comprise one or more nucleotide alterations that cause one or more amino acid modifications (e.g., substitution, insertion, or deletion) in beta globin protein when integrated into the target locus. In certain embodiments, the modified beta globin protein is a functional equivalent of the wild-type beta globin, *i.e.*, can function as a wild-type beta globin. In certain embodiments, the functionally equivalent beta globin further comprises at least one characteristic not found in wild-type beta globin, *e.g.*, the ability to inhibit aggregation of beta globin carrying the SCD mutation.

[00112] In certain embodiments, an editing element as described herein comprises at least 0, 1, 2, 10, 100, 200, 500, 1000, 1500, 2000, 3000, 4000, or 5000 nucleotides. In certain embodiments, the editing element comprises or consists of 1 to 5000, 1 to 4500, 1 to 4000, 1 to 3000, 1 to 2000, 1 to 1000, 1 to 500, 1 to 200, 1 to 100, 1 to 50, or 1 to 10 nucleotides.

[00113] In certain embodiments, an editing element as described herein comprises or consists of an exon, an intron, a 5' untranslated region (UTR), a 3' UTR, a promoter, a splice donor, a splice acceptor, a ribosomal skipping element, a sequence encoding a non-coding RNA, an insulator, a gene, or a combination thereof.

[00114] In certain embodiments, the editing element comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 23-28. In certain embodiments, the editing element consists of the nucleotide sequence set forth in any one of SEQ ID NOs: 23-26.

[00115] Homology arms used in the correction genomes disclosed herein can be directed to any region of the target gene (e.g., the HBB gene) or a gene nearby on the genome. The precise identity and positioning of the homology arms are determined by the identity of the editing element and/or the target locus.

[00116] Homology arms employed in the correction genomes disclosed herein are substantially identical to the genome flanking a target locus (*e.g.*, a target locus in the HBB gene). In certain embodiments, the 5' homology arm has at least about 90% (*e.g.*, at least about 95%, 96%, 97%, 98%, 99%, or 99.5%) nucleotide sequence identity to a first genomic region 5' to the target locus. In certain embodiments, the 5' homology arm has 100% nucleotide sequence identity to the first genomic region. In certain embodiments, the 3' homology arm has at least about 90% (*e.g.*, at least about 95%, 96%, 97%, 98%, 99%, or 99.5%) nucleotide sequence identity to a second genomic region 3' to the target locus. In certain embodiments, the 3' homology arm has 100% nucleotide sequence identity to the second genomic region. In certain embodiments, the 5' and 3' homology arms are each at least about 90% (*e.g.*, at least about 95%, 96%, 97%, 98%, 99%, or 99.5%) identical to a first genomic region 5' to the target locus and a second genomic region 3' to the target locus, respectively. In certain embodiments, the 5' and 3' homology arms are each 100% identical to the first and second genomic regions, respectively. In certain embodiments, the differences in nucleotide sequences of the 5' homology arm and the first genomic region, and/or the difference in nucleotide sequences of the 3' homology arm and the second genomic region, comprise, consist essentially of or consist of non-coding differences in nucleotide sequences.

[00117] The skilled worker will appreciate that homology arms do not need to be 100% identical to the genomic sequence flanking the target locus to be able to mediate integration of an editing element into that target locus by homologous recombination. The skilled worker will further appreciate that in situations where a homology arm is not 100% identical to the genomic sequence flanking the target locus, homologous recombination between the homology arm and the genome may alter the genomic sequence flanking the target locus such that it becomes identical to the sequence of the homology arm used.

[00118] In certain embodiments, the first genomic region 5' to the target locus is located in a first editing window, wherein the nucleotide sequence of the first editing window

consists of a sequence selected from the group consisting of SEQ ID NOs: 101-103. In certain embodiments, the second genomic region 3' to the target locus is located in a second editing window, wherein the nucleotide sequence of the second editing window consists of a sequence selected from the group consisting of SEQ ID NOs: 101-103. In certain embodiments, the first genomic region 5' to the target locus is located in a first editing window, wherein the nucleotide sequence of the first editing window consists of a sequence selected from the group consisting of SEQ ID NOs: 101-103; and the second genomic region 3' to the target locus is located in a second editing window, wherein the nucleotide sequence of the second editing window consists of a nucleotide sequence set forth in SEQ ID NO: 101, 102, or 103.

[00119] In certain embodiments, the first and second editing windows are different. In certain embodiments, the first editing window is located 5' to the second editing window. In certain embodiments, the first editing window consists of the nucleotide sequence set forth in SEQ ID NO: 101. In certain embodiments, the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 102. In certain embodiments, the first editing window consists of the nucleotide sequence set forth in SEQ ID NO: 101, and the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 102. In certain embodiments, the first genomic region consists of a sequence that is shorter than the sequence of the first editing window. In certain embodiments, the first genomic region consists of the sequence of the first editing window. In certain embodiments, the second genomic region consists of a sequence that is shorter than the sequence of the second editing window. In certain embodiments, the second genomic region consists of the sequence of the second editing window.

[00120] In certain embodiments, the first and second editing windows are the same. In certain embodiments, the target locus is an internucleotide bond or a nucleotide sequence in the editing window, wherein the first genomic locus consists of a first portion of the editing window 5' to the target locus, and the second genomic locus consists of a second portion of the editing window 3' to the target locus. In certain embodiments, the first portion of the editing window consists of the sequence from the 5' end of the editing window to the nucleotide adjacently 5' to the target locus. In certain embodiments, the second portion of the editing window consists of the sequence from the nucleotide adjacently 3' to the target locus to the 3' end of the editing window. In certain embodiments, the first portion of the editing window consists of the sequence from the 5' end of the editing window to the nucleotide adjacently 5' to the target locus, and the second portion of the editing window consists of the

sequence from the nucleotide adjacently 3' to the target locus to the 3' end of the editing window. In certain embodiments, the editing window consists of the nucleotide sequence set forth in SEQ ID NO: 103. In certain embodiments, the first and second portions of the editing windows have substantially equal lengths (e.g., the ratio of the length of the shorter portion to the length of the longer portion is greater than 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.96, 0.97, 0.98, or 0.99).

[00121] In certain embodiments, the 5' homology arm has a length of about 50 to about 4000 nucleotides (*e.g.*, about 100 to about 3000, about 200 to about 2000, about 500 to about 1000 nucleotides). In certain embodiments, the 5' homology arm has a length of about 800 nucleotides. In certain embodiments, the 5' homology arm has a length of about 100 nucleotides. In certain embodiments, the 3' homology arm has a length of about 50 to about 4000 nucleotides (*e.g.*, about 100 to about 3000, about 200 to about 2000, about 500 to about 1000 nucleotides). In certain embodiments, the 3' homology arm has a length of about 800 nucleotides. In certain embodiments, the 3' homology arm has a length of about 100 nucleotides. In certain embodiments, each of the 5' and 3' homology arms independently has a length of about 50 to about 4000 nucleotides (*e.g.*, about 100 to about 3000, about 200 to about 2000, about 500 to about 1000 nucleotides). In certain embodiments, the 5' and 3'homology arm has a length of about 800 nucleotides.

[00122] In certain embodiments, the 5' and 3' homology arms have substantially equal nucleotide lengths. In certain embodiments, the 5' and 3 ' homology arms have asymmetrical nucleotide lengths. In certain embodiments, the asymmetry in nucleotide length is defined by a difference between the 5' and 3' homology arms of up to 90% in the length, such as up to an 80%, 70%, 60%, 50%, 40%, 30%, 20%, or 10% difference in the length.

[00123] In certain embodiments, the correction genome comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 29-42.

[00124] In certain embodiments, the correction genome comprises 5' to 3': the sequence set forth in SEQ ID NO: 101; a ribosomal skipping element; an HBB coding sequence that has been silently altered to be less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50% identical to the wild-type HBB coding sequence; an exogenous polyadenylation sequence (e.g., as set forth in SEQ ID NO: 76, 77, 78, or 79); and the sequence set forth in SEQ ID NO: 102. In certain embodiments, the correction genome comprises 5' to 3': the sequence set forth in SEQ ID NO: 101; a ribosomal skipping element; the sequence set forth in SEQ ID NO: 99; an exogenous polyadenylation sequence (e.g., as set forth in SEQ ID NO: 76, 77, 78, or 79); and the sequence set forth in SEQ ID NO: 102. In

certain embodiments, the correction genome comprises 5' to 3': the sequence set forth in SEQ ID NO: 101; a ribosomal skipping element; the coding region of the first exon of HBB that has been silently altered to be less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50% identical to wild-type coding region of the first exon of HBB; an optional first nonnative intron; the second exon of HBB that has been silently altered to be less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50% identical to wild-type second exon of HBB; an optional second intron; the coding region of the third exon of HBB that has been silently altered to be less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50% identical to wild-type coding region of the third exon of HBB; an exogenous polyadenylation sequence (e.g., as set forth in SEQ ID NO: 76, 77, 78, or 79); and the sequence set forth in SEQ ID NO: 102. In certain embodiments, the correction genome comprises 5' to 3': the sequence set forth in SEQ ID NO: 101; a ribosomal skipping element; the sequence set forth in SEQ ID NO: 105; an optional first nonnative intron; the sequence set forth in SEQ ID NO: 106; an optional second intron; the sequence set forth in SEQ ID NO: 107; an exogenous polyadenylation sequence (e.g., as set forth in SEQ ID NO: 76, 77, 78, or 79); and the sequence set forth in SEQ ID NO: 102. In certain embodiments, the correction genome comprises the nucleotide sequence set forth in SEQ ID NO: 104.

[00125] In certain embodiments, the correction genomes disclosed herein further comprise a 5' inverted terminal repeat (5' ITR) nucleotide sequence 5' of the 5' homology arm nucleotide sequence, and a 3' inverted terminal repeat (3' ITR) nucleotide sequence 3' of the 3' homology arm nucleotide sequence. ITR sequences from any AAV serotype or variant thereof can be used in the correction genomes disclosed herein. The 5' and 3' ITR can be from an AAV of the same serotype or from AAVs of different serotypes. Exemplary ITRs for use in the correction genomes disclosed herein are set forth in SEQ ID NO: 18-21 herein. In certain embodiments, the 5' ITR nucleotide sequence and the 3' ITR nucleotide sequence are substantially complementary to each other (*e.g.*, are complementary to each other except for mismatch at 1, 2, 3, 4 or 5 nucleotide positions in the 5' or 3' ITR).

[00126] In certain embodiments, the 5' ITR or 3' ITR is from AAV2. In certain embodiments, both the 5' ITR and the 3' ITR are from AAV2. In certain embodiments, the 5' ITR nucleotide sequence has at least 95% (*e.g.*, at least 96%, at least 97%, at least 98%, at least 99%, or 100%) sequence identity to SEQ ID NO:18, or the 3' ITR nucleotide sequence has at least 95% (*e.g.*, at least 96%, at least 97%, at least 98%, at least 99%, or 100%) sequence identity to SEQ ID NO:19. In certain embodiments, the 5' ITR nucleotide sequence has at least 95% (*e.g.*, at least 96%, at least 97%, at least 98%, at least 99%, or 100%)

sequence identity to SEQ ID NO:18, and the 3' ITR nucleotide sequence has at least 95% (*e.g.*, at least 96%, at least 97%, at least 98%, at least 99%, or 100%) sequence identity to SEQ ID NO:19. In certain embodiments, the correction genome comprises an editing element having the nucleotide sequence set forth in any one of SEQ ID NOs: 23-28, a 5' ITR nucleotide sequence having the sequence of SEQ ID NO:18, and a 3' ITR nucleotide sequence having the sequence of SEQ ID NO:19. In certain embodiments, the correction genome comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 29-42, a 5' ITR nucleotide sequence having the sequence of SEQ ID NO:18, and a 3' ITR nucleotide sequence having the sequence of SEQ ID NO:19. In certain embodiments, the correction genome consists of 5' to 3' a 5' ITR nucleotide sequence having the sequence of SEQ ID NO:18, the nucleotide sequence set forth in any one of SEQ ID NOs: 29-42, and a 3' ITR nucleotide sequence having the sequence of SEQ ID NO:19.

[00127] In certain embodiments, the 5' ITR or 3' ITR are from AAV5. In certain embodiments, both the 5' ITR and 3' ITR are from AAV5. In certain embodiments, the 5' ITR nucleotide sequence has at least 95% (*e.g.*, at least 96%, at least 97%, at least 98%, at least 99%, or 100%) sequence identity to SEQ ID NO:20, or the 3' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO:21. In certain embodiments, the 5' ITR nucleotide sequence has at least 95% (*e.g.*, at least 96%, at least 97%, at least 98%, at least 99%, or 100%) sequence identity to SEQ ID NO:20, and the 3' ITR nucleotide sequence has at least 95% (*e.g.*, at least 96%, at least 97%, at least 98%, at least 99%, or 100%) sequence identity to SEQ ID NO:21. In certain embodiments, the correction genome comprises an editing element having the nucleotide sequence set forth in any one of SEQ ID NOs: 23-28, a 5' ITR nucleotide sequence having the sequence of SEQ ID NO:20, and a 3' ITR nucleotide sequence having the sequence of SEQ ID NO:21. In certain embodiments, the correction genome comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 29-42, a 5' ITR nucleotide sequence having the sequence of SEQ ID NO:20, and a 3' ITR nucleotide sequence having the sequence of SEQ ID NO:21. In certain embodiments, the correction genome consists of 5' to 3' a 5' ITR nucleotide sequence having the sequence of SEQ ID NO:20, the nucleotide sequence set forth in any one of SEQ ID NOs: 29-42, and a 3' ITR nucleotide sequence having the sequence of SEQ ID NO:21.

[00128] In certain embodiments, the correction genome disclosed herein has a length of about 0.5 to about 8 kb, and any range enclosed therebetween (*e.g.*, about 1 to about 5, about 2 to about 5, about 3 to about 5, about 4 to about 5, about 4.5 to about 4.8, or about 4.7 kb).

[00129] The correction genomes disclosed herein can be configured to integrate an editing element into any desired target locus of the HBB gene. In certain embodiments, the target locus is a mutation (*e.g.*, insertion, deletion or substitution of one or more nucleotides) in an HBB gene sequence relative to the corresponding wild-type HBB gene sequence. In certain embodiments, the target locus is at a nucleotide point mutation or deletion in the HBB gene. Exemplary HBB point mutations or deletion include, without limitation, G at position -87, G at position -31, A at position -30, G at position -29, G at position -28, T at position -10, C at position 1, A at position 1, G at position 2, deletion of C and T at positions 17 and 18, A at position 19, deletion of A at position 20, T at position 20, deletion of A and A at positions 25 and 26, addition of G after position 26, A at position 47, A at position 48, deletion of C at position 51, A at position 52, G at position 58, G at position 59, A at position 79, T at position 82, addition of C after position 84, T at position 93, A at position 93, C at position 97, C at position 98, G at position 202, G at position 208, C at position 222, deletion of T at position 241 or 242, deletion of T and T and C and T at positions 254 to 257, T at position 260, deletion of C at position 264 or 265, addition of A after position 343, deletion of G and T at positions 399 and 400, T at position 401, addition of A after position 417, A at position 446, T at position 1099, A at position 1293, T at position 1344. In certain embodiments, the target locus is the sickle cell disease mutation (i.e., T at position 20 of the HBB gene). In certain embodiments, the target locus is a region of chromosome or an internucleotide bond in exon 1 of the HBB gene, *e.g.*, immediately following the endogenous start codon (e.g., the internucleotide bond between nucleotide 3 and nucleotide 4 of the HBB gene). In certain embodiments, the target locus is a region of chromosome or an internucleotide bond in intron 1 of the HBB gene. In certain embodiments, the target locus consists of the native stop codon of a wild-type HBB gene or the corresponding nucleotides of a mutant HBB gene. In certain embodiments, the target locus consists of the internucleotide bond adjacently 5' to the stop codon of a wild-type HBB gene or the corresponding internucleotide bond of a mutant HBB gene.

[00130] The AAV compositions disclosed herein are particular advantageous in that they are capable of correcting a mutation in an HBB gene in a cell with high efficiency both *in vivo* and *in vitro*. In certain embodiments, the integration efficiency of the editing element into the target locus is at least 0.1% (*e.g.* at least 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 1.5%, 2%, 2.5%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%) when the AAV is contacted *in vitro* in the absence of an exogenous nuclease with a population of $CD34^+$

human hematopoietic stem cells under standard AAV transduction conditions. In certain embodiments, the allelic frequency of integration of the editing element into the target locus is at least 0.05% (*e.g.* at least 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 1.5%, 2%, 2.5%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%) when the AAV is contacted *in vitro* in the absence of an exogenous nuclease with a population of $CD34^+$ human hematopoietic stem cells under standard AAV transduction conditions.

[00131] Any methods of determining the efficiency of gene editing can be employed. In certain embodiments, individual cells are separated from the population of transduced cells and subject to single-cell PCR using PCR primers that can identify the presence of an editing element correctly integrated into the target locus. Such a method can further comprise single-cell PCR of the same cells using PCR primers that selectively amplify an unmodified target locus. In this way, the genotype of the cells can be determined. For example, if the single cell PCR showed that a cell has both an edited target locus and an unmodified target locus, then the cell would be considered heterozygous for the edited HBB gene.

[00132] Additionally or alternatively, in certain embodiments, linear amplification mediated PCR (LAM-PCR), quantitative PCR (qPCR) or digital droplet PCR (ddPCR) can be performed on DNA extracted from a population of transduced cells (e.g., a tissue or organ) to assess the allelic frequency of integration. In certain embodiments, the extracted DNA is analyzed by digital droplet PCR (ddPCR) using at least two pair of primers that detect different sequences. For example, the ddPCR can employ a first pair of primers that detect the unedited and edited target loci, a second pair of primers that detect a sequence present in the unintegrated and integrated vectors, and optionally a third pair of primers that detect a sequence in the homology arm, which is present in the unedited and edited target loci, as well as the unintegrated and integrated vectors. After correction of the chance of co-partitioning of an unedited genomic DNA and an unintegrated vector, the percentage of droplets positive with both the first and the second pair of primers corresponds to the allelic frequency of integration. An example of this method is described herein in Example 1.

[00133] Additionally or alternatively, in certain embodiments, the HBB locus can be amplified from DNA extracted from a population of transduced cells (e.g., a tissue or organ) either by PCR using primers that bind to regions of the HBB gene flanking the genomic region encompassed by the correction genome, or by linear amplification mediated PCR (LAM-PCR) using a primer that binds a region within the correction genome (*e.g.*, a region comprising an exogenous sequence non-native to the locus. The resultant PCR amplicons

can be individually sequenced using single molecule next generation sequencing (NGS) techniques to determine the relative number of edited and unedited HBB alleles present in the population of transduced cells. These numbers can be used to determine the allelic frequency of integration of the editing element into the target locus.

[00134] In another aspect, the instant disclosure provides pharmaceutical compositions comprising an AAV as disclosed herein together with a pharmaceutically acceptable excipient, adjuvant, diluent, vehicle or carrier, or a combination thereof. A "pharmaceutically acceptable carrier" includes any material which, when combined with an active ingredient of a composition, allows the ingredient to retain biological activity and without causing disruptive physiological reactions, such as an unintended immune reaction. Pharmaceutically acceptable carriers include water, phosphate buffered saline, emulsions such as oil/water emulsion, and wetting agents. Compositions comprising such carriers are formulated by well-known conventional methods such as those set forth in Remington's Pharmaceutical Sciences, current Ed., Mack Publishing Co., Easton Pa. 18042, USA; A. Gennaro (2000) "Remington: The Science and Practice of Pharmacy", 20th edition, Lippincott, Williams, & Wilkins; Pharmaceutical Dosage Forms and Drug Delivery Systems (1999) H. C. Ansel et al, 7th ed., Lippincott, Williams, & Wilkins; and Handbook of Pharmaceutical Excipients (2000) A. H. Kibbe et al, 3rd ed. Amer. Pharmaceutical Assoc.

III. Method of Use

[00135] In another aspect, the instant disclosure provides methods for correcting a mutation in an HBB gene in a cell. The methods generally comprise transducing the cell with a replication-defective AAV as disclosed herein. Such methods are highly efficient at correcting mutations in an HBB gene and do not require cleavage of the genome at the target locus by the action of an exogenous nuclease (*e.g.*, a meganuclease, a zinc finger nuclease, a transcriptional activator-like nuclease (TALEN), or an RNA-guided nuclease such as a Cas9) to facilitate such correction. Accordingly, in certain embodiments, the methods disclosed herein involve transducing the cell with a replication-defective AAV as disclosed herein without co-transducing or co-administering an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

[00136] The methods disclosed herein can be applied to any cell harboring a mutation in an HBB gene. The skilled worker will appreciate that cells that are capable of differentiating into erythrocytes are of particular interest. Accordingly, in certain embodiments, the method is applied to stem cells, including without limitation, pluripotent stem cells, induced pluripotent stem cells (iPSCs), and hematopoietic stem cells (HSC).

Exemplary HSCs to which the methods can be applied include without limitation, $CD34^+$ HSCs.

[00137] The methods disclosed herein can be performed *in vitro* for research purposes or can be performed *ex vivo* or *in vivo* for therapeutic purposes.

[00138] In certain embodiments, the cell to be transduced is taken from a subject and is transduced to correct a mutation in the HBB gene *ex vivo* according to the methods disclosed herein, and the transduced cell is subsequently administered back to the subject. Accordingly, in certain embodiments, the instant disclosure provides a method for treating a subject having a disease or disorder associated with an HBB gene mutation, the method comprising: transducing a stem cell (*e.g.*, a $CD34^+$ hematopoietic stem cell) *ex vivo* with a replication-defective AAV as disclosed herein to obtain a transduced cell; and administering the transduced cell to the subject. The transduced cells can be selected for correct genetic integration and/or cultured for clonal expansion before administration to the subject. In certain embodiments, the stem cell to be transduced is obtained from bone marrow, cord blood, or peripheral blood, wherein the stem cell is optionally selected by a method based on one or more cell markers (*e.g.*, cell size, cell density, and surface markers such as CD34). In certain embodiments, the stem cell is autologous, *i.e.*, from the subject to which the cells after AAV transduction will be administered. In certain embodiments, the stem cell is allogeneic to the subject in need thereof, *i.e.*, the stem cell is obtained from a donor that is genetically non-identical to the recipient subject. Accordingly, in certain embodiments, the instant disclosure provides a method for treating a subject having a disease or disorder associated with an HBB gene mutation, the method comprising: transducing an allogeneic stem cell (*e.g.*, a CD34+ hematopoietic stem cell) *ex vivo* with a replication-defective AAV as disclosed herein to obtain a transduced cell; and administering the transduced cell to the subject. In certain embodiments, the allogeneic stem cell is derived from a matched donor. The skilled worker will recognize that for allogeneic applications, the transduced cell may require further modifications before administration, *e.g.*, genetic modifications to prevent and/or reduce the occurrence of graft-versus-host disease (GVHD). The subject can be a human subject or a rodent subject (*e.g.*, a mouse) containing human erythrocyte precursor cells. Suitable mouse subjects include with limitation, mice into which human stem cells (*e.g.*, human $CD34^+$ HSCs) have been engrafted. Any disease or disorder associated with an HBB gene mutation can be treated using the methods disclosed herein. Suitable diseases or disorders include, without limitation, beta thalassemia or sickle cell disease. In certain

embodiments, the cell is transduced without co-transducing an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

[00139] In certain embodiments, the cell to be transduced is in a mammalian subject and the AAV is administered to the subject in an amount effective to transduce the cell in the subject. Accordingly, in certain embodiments, the instant disclosure provides a method for treating a subject having a disease or disorder associated with an HBB gene mutation, the method generally comprising administering to the subject an effective amount of a replication-defect as disclosed herein. The subject can be a human subject, a non-human primate subject (e.g., *Macaca fascicularis*), or a rodent subject (*e.g.*, a mouse) containing human erythrocyte precursor cells. Suitable mouse subjects include without limitation, mice into which human stem cells (*e.g.*, human $CD34^+$ HSCs) have been engrafted. Any disease or disorder associated with an HBB gene mutation can be treated using the methods disclosed herein. Suitable diseases or disorders include, without limitation, beta thalassemia or sickle cell disease. In certain embodiments, the cell is transduced without co-transducing or co-administering an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

[00140] The methods disclosed herein are particularly advantageous in that they are capable of correcting an HBB gene in a cell with high efficiency both *in vivo* and *in vitro*. In certain embodiments, the integration efficiency of the editing element into the target locus is at least 0.1% (*e.g.* at least 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 1.5%, 2%, 2.5%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%) when the AAV is contacted *in vitro* in the absence of an exogenous nuclease with a population of $CD34^+$ human stem cells under standard AAV transduction conditions. In certain embodiments, the allelic frequency of integration of the editing element into the target locus is at least 0.05% (*e.g.* at least 0.1%, 0.2%, 0.3%, 0.4%, 0.5%, 0.6%, 0.7%, 0.8%, 0.9%, 1%, 1.5%, 2%, 2.5%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%) when the AAV is contacted *in vitro* in the absence of an exogenous nuclease with a population of $CD34^+$ human stem cells under standard AAV transduction conditions. Any methods of determining the efficiency of gene editing can be employed including, without limitation, those described herein.

[00141] In certain embodiments, transduction of a cell with an AAV composition disclosed herein can be performed as provided herein or by any method of transduction known to one of ordinary skill in the art. In certain embodiments, the cell may be contacted

with the AAV at a multiplicity of infection (MOI) of 50,000; 100,000; 150,000; 200,000; 250,000; 300,000; 350,000; 400,000; 450,000; or 500,000, or at any MOI that provides for optimal transduction of the cell. In certain embodiments, the subject may be administered with the AAV at a dose of about 10^{11}, 10^{12}, 10^{13}, 10^{14}, or 10^{15} vector genomes per kg of body weight.

[00142] An AAV composition disclosed herein can be administered to a subject by any appropriate route including, without limitation, intravenous, intraperitoneal, subcutaneous, intramuscular, intransal, topical or intradermal routes. In certain embodiments, the composition is formulated for administration via intravenous injection or subcutaneous injection.

IV. AAV Packaging Systems

[00143] In another aspect, the instant disclosure provides packaging systems for recombinant preparation of a replication–defective AAV disclosed herein. Such packaging systems generally comprise: a Rep nucleotide sequence encoding one or more AAV Rep proteins; a Cap nucleotide sequence encoding one or more AAV Clade F capsid proteins as disclosed herein; and a correction genome for correction of a mutation in the HBB gene as disclosed herein, wherein the packaging system is operative in a cell for enclosing the correction genome in the capsid to form the AAV.

[00144] In certain embodiments, the packaging system comprises a first vector comprising the Rep nucleotide sequence and the Cap nucleotide sequence, and a second vector comprising the correction genome. As used in the context of a packaging system as described herein, a "vector" refers to a nucleic acid molecule that is a vehicle for introducing nucleic acids into a cell (*e.g.*, a plasmid, a virus, a cosmid, an artificial chromosome, etc.).

[00145] Any AAV Rep protein can be employed in the packaging systems disclosed herein. In certain embodiments of the packaging system, the Rep nucleotide sequence encodes an AAV2 Rep protein. Suitable AAV2 Rep proteins include, without limitation, Rep 78/68 or Rep 68/52. In certain embodiments of the packaging system, the nucleotide sequence encoding the AAV2 Rep protein comprises a nucleotide sequence that encodes a protein having a minimum percent sequence identity to the AAV2 Rep amino acid sequence of SEQ ID NO:22, wherein the minimum percent sequence identity is at least 70% (*e.g.*, at least 75%, at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, at least 99%, or 100%) across the length of the amino acid sequence of the AAV2 Rep protein. In certain embodiments of the packaging system, the AAV2 Rep protein has the amino acid sequence set forth in SEQ ID NO:22.

[00146] In certain embodiments of the packaging system, the packaging system further comprises a third vector, *e.g.*, a helper virus vector. The third vector may be an independent third vector, integral with the first vector, or integral with the second vector. In certain embodiments, the third vector comprises genes encoding helper virus proteins.

[00147] In certain embodiments of the packaging system, the helper virus is selected from the group consisting of adenovirus, herpes virus (including herpes simplex virus (HSV)), poxvirus (such as vaccinia virus), cytomegalovirus (CMV), and baculovirus. In certain embodiments of the packaging system, where the helper virus is adenovirus, the adenovirus genome comprises one or more adenovirus RNA genes selected from the group consisting of El, E2, E4 and VA. In certain embodiments of the packaging system, where the helper virus is HSV, the HSV genome comprises one or more of HSV genes selected from the group consisting of UL5/8/52, ICPO, ICP4, ICP22 and UL30/UL42.

[00148] In certain embodiments of the packaging system, the first, second, and/or third vector are contained within one or more transfecting plasmids. In certain embodiments, the first vector and the third vector are contained within a first transfecting plasmid. In certain embodiments the second vector and the third vector are contained within a second transfecting plasmid.

[00149] In certain embodiments of the packaging system, the first, second, and/or third vector are contained within one or more recombinant helper viruses. In certain embodiments, the first vector and the third vector are contained within a recombinant helper virus. In certain embodiments, the second vector and the third vector are contained within a recombinant helper virus.

[00150] In a further aspect, the disclosure provides a method for recombinant preparation of an AAV as described herein, wherein the method comprises transfecting or transducing a cell with a packaging system as described under conditions operative for enclosing the correction genome in the capsid to form the AAV as described herein. Exemplary methods for recombinant preparation of an AAV include transient transfection (*e.g.*, with one or more transfection plasmids containing a first, and a second, and optionally a third vector as described herein), viral infection (*e.g.* with one or more recombinant helper viruses, such as a adenovirus, poxvirus (such as vaccinia virus), herpes virus (including HSV, cytomegalovirus, or baculovirus, containing a first, and a second, and optionally a third vector as described herein), and stable producer cell line transfection or infection (*e.g.*, with a stable producer cell, such as a mammalian or insect cell, containing a Rep nucleotide sequence encoding one or more AAV Rep proteins and/or a Cap nucleotide sequence

encoding one or more AAV Clade F capsid proteins as described herein, and with a correction genome as described herein being delivered in the form of a transfecting plasmid or a recombinant helper virus).

V. Examples

[00151] The recombinant AAV vectors disclosed herein mediate highly efficient gene editing *in vitro* and *in vivo* via a nuclease-free homology dependent repair based mechanism. The following examples demonstrate the efficient correction of an HBB gene which is mutated in certain human diseases, such as sickle cell disease (SCD) and beta thalassemia, using an AAV-based vector as disclosed herein. These examples are offered by way of illustration, and not by way of limitation.

Example 1: Selection of AAV capsid for gene editing of HBB-mutant cells

[00152] This example characterizes the integration efficiency of a gene editing AAV vector, AAVS1-FP, packaged in clade F AAV capsids, such as AAVHSC7, AAVHSC15, and AAVHSC17, in HBB-mutant cells. AAVHSC7, AAVHSC15, and AAVHSC17, also known as AAVF7, AAVF15, and AAVF17, respectively, are fully described in WO2016049230A1, which is incorporated herein by reference in its entirety.

[00153] AAVS1-FP, the gene editing vector employed herein, was fully described in WO2016049230A1. It comprises from 5' to 3': an AAV2 5' inverted terminal repeat (ITR), a 5' homology arm consisting of 800 nucleotides having the sequence of the DNA upstream from a target locus, a splice acceptor, a 2A element, a coding sequence of a fluorescent protein (FP), a 3' homology arm consisting of 800 nucleotides having the sequence of the DNA downstream from a target locus, and an AAV2 3' ITR, wherein the target locus is in intron 1 of human PPP1R12C in the AAVS1 locus on chromosome 19, and wherein after homologous recombination between the AAVS1-FP vector and the human genome, exon 1 of PPP1R12C, the 2A element, and the FP coding sequence are in frame. Because the FP coding sequence in the vector is promoterless, cells transduced with this vector will express the FP only when the vector is integrated into the genome. AAVHSC-scEGFP, a self-complementary AAV vector comprising AAV2 ITRs and a promoter operably linked to enhanced green fluorescent protein (EGFP), served as a control for transduction efficiency (see *e.g.*, US Patent No. 8,628,966, which is incorporated by reference herein in its entirety).

[00154] GM16265 cells, a lymphoblastoid cell line (LCL) having the A to T mutation at position 20 in intron 1 of HBB that is characteristic of sickle cell disease (SCD), were obtained from the Coriell Institute for Medical Research (Camden, NJ). The LCLs were

cultured in RPMI supplemented with 15% FCS and 2mM L-glutamine. Cells were seeded at approximately 200,000 cells per ml and split when reaching 500,000 to 1,000,000 cells per ml. Prior to plating cells, the amount of virus needed was calculated for each transduction. Transduction volume of the virus did not exceed 10% of the total volume of the well. On the day of transduction, log phase cells were counted and plated. Cells were transduced with viruses at a multiplicity of infection (MOI) of 1.5×10^5. Packaged AAV particles were thawed on ice and sonicated on ice if necessary prior to transduction, and were added to each well individually. The cells were harvested 48 hours after the transduction.

[00155] GM16265 cells were transduced with the AAVS1-FP vector packaged in AAVHSC7, AAVHSC15, or AAVHSC17 capsids. The integration efficiency was assessed by flow cytometry using the following method: cells were harvested using FACS Buffer (1X PBS, 2% FCS, 0.1% Sodium Azide) and centrifuged at 1200 RPM for 10 minutes. Excess supernatant was decanted so that approximately 200 µl remained. 4',6-Diamidino-2-Phenylindole (DAPI) was added from a 100µM working stock immediately before flow cytometry analysis to a final concentration of 3µM.

[00156] As shown in Figure 1A, the percentages of FP-positive cells among all live cells transduced by AAVHSC7-AAVS1-FP and AAVHSC17-AAVS1-FP were 24.3% (34.0% minus background level 9.7%) and 7.8% (17.5% minus background level 9.7%), respectively. As shown in Figure 1B, in another experiment, the percentages of FP-positive cells among all live cells transduced by AAVHSC15-AAVS1-FP and AAVHSC17-AAVS1-FP were 25.1% (29.8% minus background level 4.7%) and 37.6% (42.3% minus background level 4.7%), respectively. This data shows that GM16265 cells can be efficiently transduced by AAVS1-FP packaged in AAVHSC7, AAVHSC15, and AAVHSC17 capsids.

[00157] The integration efficiency of this AAVS1-FP vector packaged in an AAVHSC17 capsid was also examined in primary human $CD34^+$ hematopoietic stem cells (HSCs). Primary human $CD34^+$ HSCs were purified from human peripheral blood cells from donors with SCD by enriching with the Miltenyi CD34 MicroBeads twice, or were obtained from ReachBio Inc. that followed a similar double enrichment procedure. The cells were cultured in Iscove's Modified Dulbecco's Medium (IMDM) supplemented with 20% Fetal Calf Serum (FCS), 100 µg/mL streptomycin, 100 U/mL penicillin, 2mmol/L L-glutamine, 10 ng/mL human IL-3, 10 ng/mL human IL-6, and 1 ng/mL human SCF. About 200,000 cells were plated in a 500 µl medium. AAV particles were added directly to the medium. The cells were harvested 48 hours after the transduction.

[00158] The editing efficiency was measured by digital droplet PCR (ddPCR) using the BioRad QX200™ Droplet Digital™ PCR System. Two sets of primers and probes, as shown in Table 1, were designed for quantifying the integration by ddPCR. The AAVS1_Genomic set detected a sequence at the AAVS1 locus that was present outside the homology arm in the unedited genome and the edited genome after the targeted integration of the FP coding sequence into the AAVS1 locus of the genome. The AAVS1_FP set detected a sequence in the FP coding region, which was present only in the edited genome after the targeted integration of the FP coding sequence into the AAVS1 locus of the genome. The two probes were conjugated to fluorescent moieties of different wavelengths.

Table 1: Primers for ddPCR analysis

Primer Name	SEQ ID NO:	Nucleotide Sequence
AAVS1_Genomic, forward primer	90	GCGTTAGAGGGCAGAGTTC
AAVS1_Genomic, reverse primer	91	AGCTCCCATAGCTCAGTCT
AAVS1_Genomic, probe	92	CATTGTCACTTTGCGCTGCCCTC
AAVS1_FP forward primer	93	GCAATAGCATCACAAATTTCAC
AAVS1_FP, reverse primer	94	GATCCAGACATGATAAGATACATTG
AAVS1_FP, probe	95	TCACTGCATTCTAGTTGTGGTTTGTCCA

[00159] Samples having 100 pg/µl of DNA were partitioned into oil droplets. Most of the oil droplets contained no DNA molecule or a single DNA molecule—an unedited genome, positive for only the AAVS1_Genomic set; an unintegrated vector, positive for only the AAVS1_FP set; or an edited genome, positive for both primer/probe sets. The chance of co-partitioning was determined by a number of standard samples (see Regan et al., A rapid molecular approach for chromosomal phasing, PLoS One. (2015) 10(3):e0118270, incorporated herein by reference in its entirety). The standard samples contained 100 unedited genomes per µl, 1000 episomal vector per µl, and a range of cloned positive alleles at 1, 5, 10, 15, 20, and 25 edited alleles per µl, respectively. A standard curve of co-partitioning against the ratio of unedited to edited allele was plotted ($R2 = 0.972$, Pearson correlation $p<0.001$).

[00160] Each sample was analyzed by ddPCR in at least three experiments, and the amounts of AAVS1_Genomic positive, AAVS1_FP positive, and double positive droplets in each sample were measured and plotted against the known ratio of unedited to edited allele in

each sample. As shown in Figure 1C, integration of the FP coding sequence into the genome was detected in about 30% of all alleles from the primary human $CD34^+$ HSCs. Accordingly, AAVS1-FP packaged in the AAVHSC17 capsid efficiently transduced primary human $CD34^+$ HSCs.

Example 2: *In vitro* correction of an HBB mutation

[00161] An AAV-based HBB correction vector named hHBB-hL-014, as shown in Figure 2, was generated. This correction vector was designed to correct an HBB mutation, *e.g.*, the A to T mutation at nucleotide 20 in the coding region of exon 1 (starting from the start codon) of the HBB gene in sickle cell disease. The hHBB-hL-014 vector included 5' and 3' AAV2 ITRs, flanking a portion of HBB and its neighboring genomic sequence, wherein the A to T mutation at nucleotide 20 in exon 1 was reversed. The portion of HBB genomic sequence was obtained from the wild type HBB and its neighboring loci using the amplification primers as shown in Table 2, which were designed using NCBI Primer Blast (www.ncbi.nlm.nih.gov/tools/primer-blast/). The PCR product covered all exons and introns of HBB, and further included 1678 nucleotides upstream from the HBB transcription initiation site and 234 nucleotides downstream from the HBB polyadenylation sequence. Relative to the HBB mutation in sickle cell disease, this vector contained homology arms (the genomic sequences 5' and 3' to the mutation reversion site) of about 1.7 kb in length each. Integrity of the ITRs was confirmed by restriction digest screening using *Bgl*II, *Msc*I, and *Sma*I and sequencing using an ITR specific sequencing protocol (Mroske *et al.*, Hum Gene Ther Methods (2012) 23(2):128-36). The insert was verified by restriction digest and Sanger sequencing using the primers shown in Table 3 herein. This vector was able to correct not only mutations in the HBB exons and introns, but mutations in the 5' and 3' untranslated regions that affect HBB expression as observed in beta thalassemia.

Table 2: Human beta globin genomic region amplification primers

Primer Name	SEQ ID NO:	Nucleotide Sequence
HBB2M Gib 5' fwd	50	AGGGGTGGAGTCGTGACGTGCCAAATCAAGCCTC TACTTGAATCC
HBB2M Gib 3' rev	51	AATGATTAACCCGCCATGCTACTTATCTACGTAA ACCTAGGCTCCAGATAGCCA

Table 3: Human beta globin sequencing primers

Primer Name	SEQ ID NO:	Nucleotide Sequence
Fwd3 seq HBB2	52	GGAAGCAGAACTCTGCAC
Fwd4 seq HBB2	53	GCATTAAGAGGTCTCTAGTTTTTTATC
Fwd5 seq HBB2	54	GATGGTATGGGGCCAAGAGATATATC
Fwd6 seq HBB2	55	GTCTACCCTTGGACCCAGAG
Fwd7 seq HBB2	56	CAGTCTGCCTAGTACATTACTATTTG
Fwd8 seq HBB2	57	CATGTTCATACCTCTTATCTTCC
Fwd9 seq HBB2	58	GCAAACAGCTAATGCACATTGG
Rev3 seq HBB2	59	CAGAATCCAGATGCTCAAGGCC
Rev4 seq HBB2	60	CCCTGATTTGGTCAATATGT
Rev5 seq HBB2	61	CATCAAGCGTCCCATAGACTCAC
Rev6 seq HBB2	62	GCAGACTTCTCCTCAGGAGTC
Rev7 seq HBB2	63	CTTACAGGACAGAATGGATGAAAAC
Rev8 seq HBB2	64	GAAAAGGTCTTCTACTTGGCTC
Rev9 seq HBB2	65	GGTTAACCAAAAGAAACTGG

[00162] To facilitate detection of the corrected gene, a 12-bp Linker having the sequence of ACTAGTATCGAT (SEQ ID NO: 80), containing a *Cla*I restriction site and an *Spe*I restriction site, was inserted in the HBB gene. This Linker sequence was located in intron 1, 117 bp from the start codon and 97 bp from the mutation reversion site, established a strong genetic linkage between the Linker and the desired gene correction. Disruption of key donor and acceptor sites in intron 1 was avoided to maintain mRNA splicing of corrected HBB.

[00163] The hHBB-hL-014vector was packaged into AAVHSC15 or AAVHSC17 capsid proteins using the packaging method described in Chatterjee *et al.*, (1993) Methods 5:51-59, which is incorporated by reference herein in its entirety. The titer of the packaged virus was determined by qPCR using the primers and probes shown in Table 4.

Table 4: HBB correction vector qPCR primers and probe

Primer Name	SEQ ID NO:	Nucleotide Sequence
qHBB2M Forward Primer	66	TGCAGATTAGTCCAGGCAGAAA
qHBB2M Reverse Primer	67	GGGTAATCAGTGGTGTCAAATAGGA
qHBB2M Probe	68	AGTTAGATGTCCCCAGTTAA

[00164] The AAVHSC15-hHBB-hL-014 and AAVHSC17-hHBB-hL-014 viruses were tested for their ability to edit the HBB gene in GM16265 cells using a targeted integration (TI) assay. In this assay, cells were centrifuged at 4000 RPM for 10 minutes, washed with 1X PBS, and the pellets were frozen at 80 °C for subsequent use. Frozen cell pellets were re-suspended in 200 µl for 100,000 cells. 1 µl of DNase-free RNase was added and incubated at 37 °C for 1 hour. 10 µl of 10% SDS and 1.2 µl of proteinase K was added and incubated at 56 °C overnight. High molecular weight DNA was extracted by standard phenol and chloroform extraction. For a high DNA yield, back extraction was performed with a 0.5X volume of Tris-EDTA buffer (TE) (pH 8.0) and added to the final tube. DNA was precipitated with 10M ammonium acetate at a final concentration of 2.5M. An approximately 4X volume of ice cold 100% ethanol was added. DNA was allowed to precipitate at -80°C for at least an hour. DNA was washed with 70% ethanol, dried, re-suspended in approximately 30 to 50 µ of TE, and quantified by Nanodrop. After quantification, the DNA was subject to a PCR-based "targeted integration" (TI) assay using appropriate primers to confirm the correct recombination between the hHBB-hL-014 vector and the HBB gene.

[00165] The primers having the sequences set forth in Table 5 were used for the TI assay in this Example. The HBB2MTI100 primer targeted the Linker and its neighboring region, and the HBB350 primer targeted a genomic sequence outside the homology arm. The PCR reaction would generate a 2,219 bp amplicon from DNA isolated from edited genomes, but would not substantially amplify from DNA isolated from untransduced cells or from the hHBB-hL-014 vector alone.

[00166] The PCR reaction was set up as follows: up to 50 µl of PCR water; 10 µl of 5X Q5 Buffer; 5 µl of betaine; 1 µl of 10mM dNTPs; 1 µl of HBB2MTI100 Forward Primer (25µM concentration); 1 µl of HBB 350 Reverse Primer (25 µM concentration); 100 ng to 1 µg of genomic DNA; 1 µl of NEB Q5 High Fidelity Polymerase. The PCR machine was set

up as follows: initial denature at 95°C for 5 minutes; 15 cycles of denature at 95°C for 10 seconds, anneal at 70°C for 30 seconds, decreasing 0.5 degrees every cycle, and extension at 72°C for 2 minutes; 20 cycles of denature at 95°C for 10 seconds; anneal at 65°C for 30 seconds, and extension at 72°C for 2 minutes; and a final extension at 72°C for 5 minutes. The PCR products were analyzed by gel electrophoresis.

[00167] Amplicons of with the apparent size of 2.2 kb were isolated and blunt-end ligated into a pUC118 backbone. The resultant plasmids were analyzed by restriction digestion using *Cla*I or *Spe*I endonucleases to identify clones with correct insertion of the linker. Positive clones identified by restriction digestion were further analyzed by DNA sequencing using M13F and M13R oligonucleotide primers.

Table 5: Primers for HBB targeted integration assay

Primer Name	SEQ ID NO:	Nucleotide Sequence
HBB2MTI100 Forward Primer (TM 54.1 °C)	69	CTATTGGTCTCCTTAAAATCGATACTAGT
HBB350 Reverse Primer (TM 54.8 °C)	70	ATATTCAAACTTCCGCAGAACACT

[00168] As shown in Figure 3A, a 2.2 kb PCR band (indicative of correct editing by the hHBB-hL-014 vector) was detected in the TI assay in GM16265 cells transduced with AAVHSC15-hHBB-hL-014 and AAVHSC17-hHBB-hL-014 viruses, whereas this PCR product was not detected in untransduced GM16265 cells.

[00169] Correction of the A to T mutation at position 20 of the HBB gene in GM16265 cells was verified by sequencing of positive clones identified in the TI assay. Sequence analysis showed that after transduction with AAVHSC15-hHBB-hL-014 or AAVHSC17-hHBB-hL-014 virus, the T mutation was corrected to A, and asymptomatic mutations nearby were also corrected to wild-type. Sequencing of the opposite PCR strand confirmed correction of the mutations. Insertion of the linker into the HBB locus was detected in all clones. Moreover, none of the examined clones showed any undesired mutations (e.g., additional insertion, deletion, or inversion) in the genomic regions corresponding to the termini of the homology arms.

[00170] To further confirm the ability of the hHBB-hL-014 vector to edit the HBB gene, two additional LCLs, GM16266 and GM16267, were used. These LCLs were also obtained from the Coriell Institute for Medical Research (Camden, NJ), and were collected

from different donors from that of GM16265. Both LCLs had the A to T mutation at position 20 in intron 1 of HBB. Following the same methods for culture and transduction of GM16265, the cells were transduced with hHBB-hL-014 packaged in an AAVHSC17 capsid.

[00171] As shown in Figure 3B, a PCR amplicon of about 2.2kb was detected from cells transduced with AAVHSC17- hHBB-hL-014, but not from the untransduced cells. Sequencing results showed that the SCD mutation was corrected in the transduced cells from all three LCLs (including the GM16265 cells), and no undesired mutations (e.g., additional insertion, deletion, or inversion) were detected in the genomic regions corresponding to the termini of the homology arms. Sequencing results also confirmed appropriate linker insertion and seamless transition from the homology arm to the genome.

[00172] The above results show that the hHBB-hL-014 vector is able to revert a mutant HBB gene to the wild-type sequence in multiple SCD cell lines. Accordingly, mutations in the exons, introns, or regulatory sequences of the HBB gene in other genetic diseases, such as beta thalassemia, should also be correctable using the hHBB-hL-014 vector.

Example 3: HBB correction vectors comprising a genomic sequence of HBB

[00173] This example provides AAV-based HBB correction vectors comprising a genomic sequence of HBB capable of correcting mutations in the HBB gene.

a) HBB correction vector hHBB-hL-001

[00174] HBB correction vector hHBB-hL-001, as shown in Figure 4A, contains HBB genomic sequence including all exons, all introns, and the polyadenylation sequence. This vector additionally contains a 5' region encompassing 800 bp upstream from the HBB transcription initiation site (referred to as "HBB HAL" in Figure 4A), and a 3' region encompassing 800 bp downstream from the HBB polyadenylation sequence ("HBB HAR" in Figure 4A). The hHBB-hL-001 vector comprises the nucleotide sequence set forth in SEQ ID NO: 31 (with a TI RE Linker) or SEQ ID NO:32 (without an TI RE Linker), and further comprises a 5' ITR (e.g., having the sequence of SEQ ID NO: 18) and a 3' ITR (e.g., having the sequence of SEQ ID NO: 19). This vector corrects not only mutations in the HBB exons and introns, but mutations in the 5' and 3' untranslated regions that affect HBB expression as observed in beta thalassemia.

b) HBB correction vector hHBB-hLW-013

[00175] HBB correction vector hHBB-hLW-013, as shown in Figure 4B, contains the same genetic elements as in HBB correction vector hHBB-hL-001 except that the DNA sequences of the coding regions of exons 1, 2 and 3 are silently altered to be about 67% identical, rather than fully identical, to the corresponding wild-type sequences. The reduced

sequence identity results from codon alteration, wherein degenerate codons are substituted for original codons without changing the encoded amino acids. This silent codon alteration is not expected to significantly alter the expression level of HBB. Instead, it reduces the homology of HBB exons with other globin genes or pseudogenes, such as epsilon globin (HBE), delta globin (HBD), gamma globin 1 (HBG1), gamma globin 2 (HBG2), and HBB pseudogene HBBP, thereby reducing the possibility of undesired recombination of this vector at other genomic loci. In a specific example, the silently altered sequence of the coding region of exon 1, exon 2, and exon 3 are set forth in SEQ ID NOs: 43,44 and 45, respectively. The hHBB-hLW-013 vector comprises the nucleotide sequence set forth in SEQ ID NO: 33 (with a TI RE Linker) or SEQ ID NO: 34 (without a TI RE Linker), and further comprises a 5' ITR (e.g., having the sequence of SEQ ID NO: 18) and a 3' ITR (e.g., having the sequence of SEQ ID NO: 19).

c) HBB correction vector hHBB-hL-011

[00176] HBB correction vector hHBB-hL-011, as shown in Figure 4C, contains the same genetic elements as in HBB correction vector hHBB-hL-001 except that the 3' region downstream from the HBB polyadenylation sequence (referred to as "HBB HAR" in Figure 4C) is about 100 bp in length. This modification is made to minimize inclusion of transcription promoter sequences of other genes (*e.g.*, GATA1, MYC, etc.), which are located about 100 bp downstream from HBB polyadenylation sequence, because the promoter sequences may recruit transcriptional factors and co-factors, thereby reducing the efficiency of homologous recombination. Additionally, inclusion of transcription promoter sequences may increase aberrant expression of HBB from the vector. The hHBB-hL-011 vector comprises the nucleotide sequence set forth in SEQ ID NO: 35 (with a TI RE Linker) or SEQ ID NO: 36 (without a TI RE Linker), and further comprises a 5' ITR (e.g., having the sequence of SEQ ID NO: 18) and a 3' ITR (e.g., having the sequence of SEQ ID NO: 19).

d) HBB correction vector hHBB-hLW-012

[00177] HBB correction vector hHBB-hLW-012, as shown in Figure 4D, contains the same genetic elements as in HBB correction vector hHBB-hL-011 except that the DNA sequences of the coding regions of exons 1, 2 and 3 are silently altered to be 67% identical to the corresponding wild-type sequences. In a specific example, the silently altered sequence of the coding region of exon 1, exon 2, and exon 3 are set forth in SEQ ID NOs: 43, 44 and 45, respectively. The hHBB-hLW-012 vector comprises the nucleotide sequence set forth in SEQ ID NO: 37 (with a TI RE Linker) or 38 (without a TI RE Linker), and further comprises

a 5' ITR (e.g., having the sequence of SEQ ID NO: 18) and a 3' ITR (e.g., having the sequence of SEQ ID NO: 19).

[00178] Each of the four HBB correction vectors in this example contains a Linker sequence, which comprises recognition and cleavage sites for unique restriction endonucleases, to facilitate detection of the corrected gene. This Linker sequence is located in intron 1, 117 bp from the start codon. Disruption of key donor and acceptor sites in intron 1 is avoided to maintain mRNA splicing of corrected HBB.

[00179] Each of the four HBB correction vectors described above was generated and packaged in AAVHSC17. Primary human $CD34^+$ HSCs were transduced with the viruses using the method as described in Example 1, and integration was assessed by the TI assay as described in Example 2. As shown in Figures 5A and 5B, hHBB-hL-001, hHBB-hL-011, and hHBB-hLW-012 were all capable of editing the HBB gene.

[00180] The efficiency of genome editing was measured quantitatively by next-generation sequencing (NGS). PCR reactions were performed using primers specific to a region of the genome outside the homology arms, as shown in Table 6. These primers will amplify a product of 2,342 bp from unedited and edited alleles, but will not amplify the AAV vector.

Table 6: Primers for NGS sample preparation

Primer Name	SEQ ID NO:	Nucleotide Sequence
HBB350 Reverse Primer	70	ATATTCAAACTTCCGCAGAACACT
HBB L NGS S1	81	CCTCTGGGTCCAAGGGTAGA

[00181] The PCR reaction was set up as follows: up to 50 µl of PCR water; 10 µl of 5X Q5 Buffer; 5 µl of Betaine; 1 µl of 10mM dNTPs; 1 µl of HBB350 Reverse Primer (25 µM); 1 µl of HBB L NGS S1 primer (25 µM); 200 ng of genomic DNA; and 1 µl of Q5 Hifidelity Polymerase. The PCR machine was set up as follows: initial denature at 98°C for 30 minutes; 30 cycles of denature at 95°C for 10 seconds, anneal at 65°C for 30 seconds, and extension at 72°C for 2 minutes; and a final extension at 72°C for 5 minutes.

[00182] The PCR product of the correct size was isolated by gel electrophoresis, and was extracted using Qiagen Qiaquick Gel Extraction Kit according to standard protocol. The absence of vector genomes in the gel extracted amplicons was confirmed by PCR using vector specific primers and a known number of vector specific genome template as positive control. To confirm absence of vector genomes, the following PCR conditions were used:

initial denature at 98°C for 30 minutes; 30 cycles of denature at 98°C for 10 seconds, anneal at 66°C for 30 seconds, and extension at 72°C for 1 minute; and a final extension at 72°C for 2 minutes. The forward primer used was AAAGTCAGGGCAGAGCCATC (SEQ ID NO: 108) and the reverse primer used was AATGATTAACCCGCCATGCT (SEQ ID NO: 109), and will amplify an amplicon of 1,797 base pairs.

[00183] The extracted PCR product was used for NGS sequencing, and/or digital PCR quantitation as described below. In the case of NGS sequencing, the extracted PCR product was subjected to a nested round of PCR using the primers shown in Table 7. Each sample had a unique combination of forward and reverse primers, and the size of each correct PCR product was about 388 bp. The PCR machine was set up as follows: initial denature at 98°C for 30 minutes; 30 cycles of denature at 98°C for 10 seconds, anneal at 72°C for 30 seconds, and extension at 72°C for 30 seconds; and a final extension at 72°C for 2 minutes.

Table 7: Primers for NGS sample preparation

Primer Name	SEQ ID NO:	Nucleotide Sequence
HBB NGS Nest 5 Forward Adapter 1	82	AATGATACGGCGACCACCGAGATCTACACAAGTAGA GTCTTTCCCTACACGACGCTCTTCCGATCTGGGCATA AAAGTCAGGGCAGA
HBB NGS Nest 5 Forward Adapter 2	83	AATGATACGGCGACCACCGAGATCTACACCATGCTTA TCTTTCCCTACACGACGCTCTTCCGATCTtGGGCATAA AAGTCAGGGCAGA
HBB NGS Nest 5 Forward Adapter 3	84	AATGATACGGCGACCACCGAGATCTACACGCACATCT TCTTTCCCTACACGACGCTCTTCCGATCTatGGGCATA AAAGTCAGGGCAGA
HBB NGS Nest 5 Forward Adapter 4	85	AATGATACGGCGACCACCGAGATCTACACTGCTCGAC TCTTTCCCTACACGACGCTCTTCCGATCTgatGGGCATA AAAGTCAGGGCAGA
HBB NGS Nest 5 Reverse Adapter 1	86	CAAGCAGAAGACGGCATACGAGATCATGATCGGTGA CTGGAGTTCAGACGTGTGCTCTTCCGATCTGTCTCCA CATGCCCAGTTTCTA
HBB NGS Nest 5 Reverse Adapter 2	87	CAAGCAGAAGACGGCATACGAGATAGGATCTAGTGA CTGGAGTTCAGACGTGTGCTCTTCCGATCTtGTCTCCA CATGCCCAGTTTCTA
HBB NGS Nest 5 Reverse Adapter 3	88	CAAGCAGAAGACGGCATACGAGATGACAGTAAGTGA CTGGAGTTCAGACGTGTGCTCTTCCGATCTatGTCTCC ACATGCCCAGTTTCTA
HBB NGS Nest 5 Reverse Adapter 4	89	CAAGCAGAAGACGGCATACGAGATCCTATGCCGTGA CTGGAGTTCAGACGTGTGCTCTTCCGATCTgaGTCTCC ACATGCCCAGTTTCTA

[00184] Amplicon sizes were confirmed by gel electrophoresis, and the PCR products were purified using Qiagen Qiaquick PCR Purification Kit according to standard protocol. Amplicons from all tested samples were mixed at equal molarities, and the concentration was confirmed by an Advanced Analytical bioanalyzer. The samples were sequenced using a MiSeq V2 300 cycle kit.

[00185] Table 8 shows the number of acceptable reads with and without the linker sequence present as well as the number of reads with HBB sequences, as determined from the NGS analysis. Because the presence of the linker sequence indicated AAV vector integration, the percentages of alleles with linker represented the allelic frequencies of integration of these vectors, which are generally between 0.1% and 1%.

Table 8: Allelic frequency of integration of HBB correction vectors

Sample	# linkers	# no linkers	#linker adjacent sequence	% with linker	% reads captured*
HBB-hL-014 #1	4	2703	2710	0.14	99.88
HBB-hL-011 #1	3101	475533	479142	0.64	99.89
HBB-hL-011 #2	5	1917	1925	0.26	99.84
HBB-hL-001 #1	3	449	455	0.66	99.34
HBB-hL-011 #3	2	235	235	0.84	100.85
Negative control	0	1	1	0	1
Untransduced	0	0	0	0	0

*% reads captured is a quality control metric that indicates the percentage quality passing reads containing the sequence adjacent to the linker sequence divided by the sum on the reads containing the linker sequence and no linker sequence.

[00186] In the case of digital PCR, the extracted PCR product was subjected to digital PCR analysis using the BioRad QX200™ Droplet Digital™ PCR System. Editing was determined by calculating the linkage between the genomic target and the inserted vector payload (linker) and was measured by detecting the amount of partitioned droplets that contain both the vector and the genome in relation to expectation of coincidence by chance employing the methods used for genetic linkage between variants (see, e.g., Regan et al., A rapid molecular approach for chromosomal phasing, PLoS One. (2015) 10(3):e0118270, incorporated herein by reference in its entirety). A concentration of 0.1 ng/ul of genomic DNA was analyzed across a minimum of three experiments per sample for linkage and measured using multiplexed ddPCR with a vector specific probe set and a genomic specific probe set. The primer and probe set is as follows:

Table 9: Primer and probe set used for digital PCR

	SEQ ID NO:	Nucleotide sequence
Linker specific primer and probe set		
HBB Probe	110	AACTGGGCATGTGGAGACAGAGAA
HBB Linker F	111	GTTACAAGACAGGACTAGTATCGAT
HBB R	112	TAGACCAATAGGCAGAGAGAGT
Genomer specific primer and probe set		
HBB0101GDNA F	113	CTGAGCCAAGTAGAAGACCTTT
HBB0101GDNA R	114	CTGTTTCTGCCTGGACTAATCT
HBB0101GDNA Probe	115	CCCTACTTTCTAAGTCACAGAGGCT

[00187] In order to measure genome editing/linkage as per the method above against a known quantity of edited material, a standard DNA series was created. The standard consisted of 100 unedited genomes per ul, 1000 episomal vector per ul and a range of cloned positive alleles at 1 per ul, 5 per ul, 10 per ul, 15 per ul, 20 per ul, and 25 edited alleles per ul, respectively. The amount of genetic linkage in each sample was measured and plotted against the known ratio of unedited to edited allele in each sample ($R2 = 0.972$, pearson correlation $p<0.001$).

[00188] Editing of the HBB locus in primary CD34+ mixed cord blood cells transduced with various AAVHSC7 HBB editing vectors at a MOI of 1.5×10^5 were measured by digital PCR. Cells were harvested at 48 hours post-transduction and assayed for percentage of alleles edited by an out/out PCR and digital PCR analysis of PCR products. Figure 6 shows the fraction of edited loci across the samples as indicated.

Example 4: HBB correction vectors comprising an HBB coding sequence or a portion thereof

[00189] This example provides HBB correction vectors that are capable of inserting an HBB coding sequence or a portion thereof into the HBB gene, e.g., after the start codon or into intron 1. The inserted sequence can be transcribed and translated from the native locus under the control of the native transcription regulatory elements, thereby restoring the expression of a functional HBB protein.

[00190] Each of the HBB correction vectors contained an HBB coding sequence or a portion thereof from the second codon to the stop codon. The HBB coding sequence or portion thereof is followed by an SV40 polyadenylation sequence, which is strong enough to support proper expression and significantly reduce further transcription of the rest of the endogenous HBB gene. A targeted integration restriction cassette ("TI RE cassette") comprising recognition and cleavage sites for a unique restriction endonuclease is optionally

inserted downstream from the polyadenylation sequence, facilitating detection of the desired homologous recombination.

a) HBB correction vector hHBB-hA-009

[00191] HBB correction vector hHBB-hA-009, as shown in Figure 7A, contains a portion of a wild-type HBB coding sequence from the second codon to the stop codon (nucleotides 4-444 of SEQ ID NO: 27), followed by an SV40 polyadenylation sequence as described above. The vector further contains a 5' homology arm (referred to as "HBB HAL" in Figure 7A) comprising the wild-type genomic sequence upstream from and including the HBB start codon, and a 3' homology arm (referred to as "HBB HAR" in Figure 7A) comprising the wild-type genomic sequence downstream from but not including the HBB start codon. Each of the 5' homology arm and the 3' homology arm is about 800 bp in length. The hHBB-hA-009 vector comprises the nucleotide sequence set forth in SEQ ID NO: 39, and further comprises a 5' ITR (e.g., having the sequence of SEQ ID NO: 18) and a 3' ITR (e.g., having the sequence of SEQ ID NO: 19). The 5' homology arm has the ability to correct mutations in the start codon and/or 5' untranslated region (UTR) that affect HBB expression as observed in some beta thalassemia patients. As a result, the integration of HBB correction vector hHBB-hA-009 can restore the expression of wild-type HBB that has been impaired by mutations in 5' UTR, coding sequence, or 3' UTR.

b) HBB correction vector hHBB-hAW-002

[00192] HBB correction vector hHBB-hAW-002, as shown in Figure 7B, contains the same genetic elements as in HBB correction vector hHBB-hA-009, except that the portion of HBB coding sequence is silently altered to SEQ ID NO: 47, 67% identical to the corresponding region of wild-type cDNA sequence. As described in Example 3, this codon alteration is not expected to significantly alter the expression level of HBB. Instead, it reduces the homology of HBB exons with other globin genes and pseudogenes, thereby reducing undesired recombination of this vector at other genomic loci. The hHBB-hAW-002 vector comprises the nucleotide sequence set forth in SEQ ID NO: 40, and further comprises a 5' ITR (e.g., having the sequence of SEQ ID NO: 18) and a 3' ITR (e.g., having the sequence of SEQ ID NO: 19).

c) HBB correction vector hHBB-h1-010

[00193] HBB correction vector hHBB-h1-010, as shown in Figure 7C, is designed to insert a wild-type HBB coding sequence into intron 1 by homologous recombination. Specifically, the insertion site is between nucleotides 160 and 161 of the HBB gene, and the insertion avoids disruption of key splice donor sites in intron 1. The editing element (inserted

region) of this vector contains from 5' to 3' a splice acceptor site ("SA" in Figure 7C, e.g., SEQ ID NO: 14), a ribosomal skipping element ("T2A" in Figure 7C, e.g., SEQ ID NO: 72) in-frame with the HBB start codon upon integration, a wild-type HBB coding sequence (SEQ ID NO: 27), and an SV40 polyadenylation sequence. Upon integration, the pre-mRNA transcribed from the HBB locus contains from 5' to 3': exon 1 of endogenous HBB; the first 68 nucleotides of intron 1, including the endogenous splice donor at the 5' end of intron 1; the splice acceptor introduced by the hHBB-h1-010 vector; the ribosomal skipping element; the HBB coding sequence; and a poly(A) tail. After splicing, the mRNA contains from 5' to 3': exon 1 of endogenous HBB; the in-frame ribosomal skipping element; the HBB coding sequence; and a poly(A) tail. The ribosomal skipping element leads to generation of two polypeptides: a truncated HBB peptide terminated at the end of exon 1 fused with a partial ribosomal skipping peptide, and a proline from the ribosomal skipping peptide fused to the N-terminus of a full-length HBB polypeptide.

[00194] The hHBB-h1-010 vector further comprises a 5' homology arm (referred to as "HBB HAL" in Figure 7C) comprising the wild-type genomic sequence upstream from the insertion site, and a 3' homology arm (referred to as "HBB HAR" in Figure 7C) comprising the wild-type genomic sequence downstream from the insertion site. Each of the 5' homology arm and the 3' homology arm is about 800 bp in length. The hHBB-h1-010 vector comprises the nucleotide sequence set forth in SEQ ID NO: 41, and further comprises a 5' ITR (e.g., having the sequence of SEQ ID NO: 18) and a 3' ITR (e.g., having the sequence of SEQ ID NO: 19). The 5' homology arm has the ability to correct mutations in the 5' UTR that affect HBB expression as observed in some beta thalassemia patients, and restore expression of wild-type HBB that has been impaired by mutations in 5' UTR, coding sequence, or 3' UTR.

d) HBB correction vector hHBB-h1W-008

[00195] HBB correction vector hHBB-h1W-008, as shown in Figure 7D, contains the same genetic elements as in HBB correction vector hHBB-h1-010, except that the HBB coding sequence is silently altered to be 67% identical to the corresponding region of wild-type cDNA sequence. As described in Example 3, this sequence modification is not expected to significantly alter the expression level of HBB. Instead, it reduces the homology of HBB exons with other globin genes and pseudogenes, thereby reducing undesired of this vector to other genomic loci. In a specific example, the silently altered HBB coding sequence is set forth in SEQ ID NO: 47. The hHBB-h1W-008 vector comprises the nucleotide sequence set

forth in SEQ ID NO: 42, and further comprises a 5' ITR (e.g., having the sequence of SEQ ID NO: 18) and a 3' ITR (e.g., having the sequence of SEQ ID NO: 19).

e) HBB correction vector hHBB-hE3C-001

[00196] HBB correction vector hHBB-hE3C-001, as shown in Figure 7E, is designed to insert an HBB coding sequence into exon 3 of the HBB gene, directly after the stop codon, by homologous recombination. The editing element (inserted region) of this vector contains from 5' to 3' a ribosomal skipping element ("P2A" in Figure 7E, e.g., SEQ ID NO: 74) in frame with a silently altered HBB coding sequence (SEQ ID NO: 99, 85% identical to the wild-type HBB coding sequence), and an SV40 polyadenylation sequence (SEQ ID NO: 77). The silent alteration of the HBB coding sequence is designed to increase the level of protein expressed from the coding sequence, remove low-complexity sequences that could lead to off-targeting of the vector to undesired genomic loci, and/or reduce the homology between the editing element and the genome, thereby reducing undesired integration mediated by the editing element rather than by a homology arm.

[00197] Upon integration, the mRNA transcribed from the HBB locus contains from 5' to 3': a portion of the native HBB mRNA adjacently 5' to the stop codon, the ribosomal skipping element, the silently altered HBB coding sequence, and the SV40 polyadenylation sequence. The ribosomal skipping element leads to generation of two polypeptides: a native full-length HBB peptide fused with an N-terminal part of the ribosomal skipping peptide, and a proline residue from the ribosomal skipping peptide fused to the N-terminus of a full-length wild-type HBB polypeptide.

[00198] The hHBB-hE3C-001 vector further comprises a 5' homology arm (referred to as "HAL" in Figure 7E) comprising the sequence of SEQ ID NO: 101, a wild-type genomic sequence upstream from the insertion site, and a 3' homology arm (referred to as "HAR" in Figure 7E) comprising the sequence of SEQ ID NO: 102, a wild-type genomic sequence downstream from the insertion site.

[00199] The nucleotide sequence of hHBB-hE3C-001 is set forth in SEQ ID NO: 104. The vector further comprises a 5' ITR (e.g., having the sequence of SEQ ID NO: 18) and a 3' ITR (e.g., having the sequence of SEQ ID NO: 19).

[00200] The integration efficiency of hHBB-h1-010 and hHBB-h1W-008 was assessed in RKO and LCL cells. GM16265 LCL cells were transduced with HBB correction vectors packaged in AAVHSC7 using the method described in Example 1.

[00201] RKO cells were obtained from ATCC. The cells were cultured in DMEM supplemented with 10% FCS and 2mM L-glutamine. They were plated at a density of

750,000 cells per well in a 6 well plate. The cells were transfected using the following method: 24 hours after plating, the cells were transfected in OptiMEM media by adding a transfection mixture, which was prepared by mixing and incubating (a) 2 µg of HBB editing plasmid diluted in 250 µl OptiMEM and (b) 5 µl Lipofectamine 2000 diluted in 250 µl OptiMEM for 15 minutes. The cells were harvested 24 hours after the transfection.

[00202] Integration was assessed by a TI assay using the primers having the sequences set forth in Table 9. The SA-2A-FM1 and SA-2A-FM2 primers were specific to the splice acceptor and T2A element inserted into the genome by the hHBB-h1-010 or hHBB-h1W-008 vector, and the HBB-Out-RM2 primer was specific to a region of the genome downstream from the 3' homology arm. The primer pairs of HBB-Out-RM2 with SA-2A-FM1 or SA-2A-FM2 do not amplify a product in untransduced cells or a product from the correction vector alone. PCR reaction using SA-2A-FM1 and HBB-Out-RM2 would generate a 1,881 bp amplicon if the hHBB-h1-010 or hHBB-h1W-008 vector is integrated by homologous recombination through the 5' and 3' homology arms, and would generate a 1,188 bp amplicon if the vector is integrated by homologous recombination through the 5' homology arm and the exon 2 sequence in the editing element. PCR reaction using SA-2A-FM2 and HBB-Out-RM2 would also generate different sizes of amplicon from these two manners of integration.

[00203] The PCR reaction was set up as follows: up to 50 µl of PCR water; 5 µl of 10X PCR Buffer; 1 µl of 10mM dNTPs; 1 µl of 50mM $MgCl_2$; 10 µl of 5X Q Reagent; 2.5 µl of TI Forward Primer (5 µM concentration); 2.5 µl of TI Reverse Primer (5 µM concentration); 100 ng genomic DNA; and 0.5 µl of HotStarTaq Polymerase. The PCR machine was set up as follows: initial denature at 95°C for 15 minutes; 40 cycles of denature at 94°C for 10 seconds, anneal at 58°C for 30 seconds, and extension at 72°C for 3 minutes; and a final extension at 68°C for 10 minutes. The PCR products were analyzed by gel electrophoresis.

Table 9: Primers for HBB targeted integration assay

Primer Name	SEQ ID NO:	Nucleotide Sequence
SA-2A-FM1	96	GCTTCTGACCTCTTCTCTTCCTCCC
SA-2A-FM2	97	GCGGTGACGTGGAGGAGAATC
HBB-Out-RM2	98	GCAGAATGGTAGCTGGATTGTAGC

[00204] As shown in Figure 8, using the SA-2A-FM1 and HBB-Out-RM2 primers, targeted integration PCR product having a length of 1,874 nucleotides was detected in RKO cells transduced with hHBB-h1W-002, indicating successful integration of the vector in the

desired manner. By contrast, a shortened PCR product, generated by recombination of exon 2 in the editing element (rather than in the 3' homology arm) with the HBB locus, was detected in RKO and LCL cells transduced with hHBB-h1-010. Similar results were obtained from PCR using the SA-2A-FM2 and HBB-Out-RM2 primers. This result indicates that silent codon alteration in the editing element reduces or eliminates undesired recombination, thereby assuring accurate editing of the HBB gene.

Example 5: *In vivo* correction of HBB mutations

[00205] This example provides an animal model for examining HBB correction vectors, such as those described in the previous examples. An NSG mouse having the genotype of NOD.*Cg-Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ was sublethally irradiated and underwent hematopoietic reconstitution by transplantation with primary human wild-type CD34$^+$ HSCs. Engraftment levels of human CD34$^+$ HSCs in the NSG mice were determined 12 weeks after transplantation by assaying peripheral blood for the presence of human and murine CD45$^+$ cells by flow cytometry. Mice having greater than 25% circulating human cells in the peripheral blood were used for assessing the integration efficiency of specific AAV vectors to transduce primary human CD34$^+$ HSCs *in vivo*.

[00206] The AAVS1-FP vector was packaged in AAVHSC7 and AAVHSC17, and the viral particles were administered to the reconstituted NSG mice intravenously at a dose between 1.22×10^{13} and 1.54×10^{13} vector genomes per kg. Blood, bone marrow, and spleen samples were collected 4 weeks after the administration. DNA was purified from the samples by the phenol/chloroform extraction method known in the art, and the extracted DNA was analyzed by ddPCR using the method as described in Example 1.

[00207] Data from the AAVHSC7 and AAVHSC17 groups were pooled. As shown in Figure 9, in the transplanted NSG mice administered with AAVS1-FP vector, the allelic frequency of integration was about 3% in the blood, and about 1% in the bone marrow. This result suggested that the AAVHSC7 and AAVHSC17 capsids efficiently delivered a vector for integration into the AAVS1 locus, and can potentially be used for delivering HBB editing therapeutic vectors.

[00208] A modified animal model reconstituted with HBB defective primary human CD34$^+$ HSCs is useful for testing correction of HBB mutations using a correction vector or correction vector as described in the foregoing examples. For example, the correction vector packaged in an AAV clade F capsid, such as an AAVHSC7, AAVHSC15 or AAVHSC17 capsid, can be administered to the reconstituted animal. The integration efficiency can be measured by collecting blood or bone marrow samples and quantifying the percentage of

cells in which desired homologous recombination has occurred either in a broad population or in a specific type of cells, such as progenitors of erythrocytes.

[00209] The animal reconstituted with HBB defective primary human $CD34^+$ HSCs are expected to manifest hemoglobinopathy due to lack of an HBB gene, and can be used to determine the efficacy and safety of HBB correction vectors packaged in various AAV capsids. Efficacy is assessed by measuring reticulocyte count, complete blood counts (CBCs), blood smears and targeted integration of the vector sequence. Safety is assessed by measuring the levels of hepatic transaminases such as aspartate transaminase (AST) and alanine transaminase (ALT).

[00210] This model can also be used to assess the longevity of HBB correction after each administration, thereby optimizing the dosing regimen.

* * *

[00211] The invention is not to be limited in scope by the specific embodiments described herein. Indeed, various modifications of the invention in addition to those described will become apparent to those skilled in the art from the foregoing description and accompanying figures. Such modifications are intended to fall within the scope of the appended claims.

[00212] All references (*e.g.*, publications or patents or patent applications) cited herein are incorporated herein by reference in their entirety and for all purposes to the same extent as if each individual reference (*e.g.*, publication or patent or patent application) was specifically and individually indicated to be incorporated by reference in its entirety for all purposes. Other embodiments are within the following claims.

We claim:

1. A method for correcting a mutation in a beta globin (HBB) gene in a cell, the method comprising transducing the cell with a replication-defective adeno-associated virus (AAV) comprising:

a) an AAV capsid comprising an AAV Clade F capsid protein; and

b) a correction genome comprising: (i) an editing element for editing a target locus in a target gene; (ii) a 5' homology arm nucleotide sequence 5' of the editing element having homology to a first genomic region 5' to the target locus; and (iii) a 3' homology arm nucleotide sequence 3' of the editing element having homology to a second genomic region 3' to the target locus,

wherein the cell is transduced without co-transducing or co-administering an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

2. The method of claim 1, wherein the cell is a pluripotent stem cell.

3. The method of claim 1, wherein the cell is a hematopoietic stem cell.

4. The method of claim 1, wherein the cell is a $CD34^+$ hematopoietic stem cell.

5. The method of any one of claims 1-4, wherein the cell is in a mammalian subject and the AAV is administered to the subject in an amount effective to transduce the cell in the subject.

6. A method for treating a subject having a disease or disorder associated with an HBB gene mutation, the method comprising:

a) transducing an erythrocyte progenitor cell from the subject *ex vivo* with a replication-defective AAV comprising:

(i) an AAV capsid comprising an AAV Clade F capsid protein; and

(ii) a correction genome comprising: (A) an editing element for editing a target locus in a target gene; (B) a 5' homology arm nucleotide sequence 5' of the editing element having homology to a first genomic region 5' to the target locus; and (C) a 3' homology arm nucleotide sequence 3' of the editing element having homology to a second genomic

region 3' to the target locus, thereby obtaining a transduced cell having a corrected HBB gene, and;

b) administering the transduced cell to the subject,

wherein the cell is transduced without co-transducing an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

7. The method of claim 6, wherein the erythrocyte progenitor cell is a pluripotent stem cell.

8. The method of claim 6, wherein the erythrocyte progenitor cell is a hematopoietic stem cell.

9. The method of claim 6, wherein the erythrocyte progenitor cell is $CD34^+$ hematopoietic stem cell.

10. A method for treating a subject having a disease or disorder associated with an HBB gene mutation, the method comprising administering to the subject an effective amount of a replication-defective AAV comprising:

a) an AAV capsid comprising an AAV Clade F capsid protein; and

b) a correction genome comprising: (i) an editing element for editing a target locus in a target gene; (ii) a 5' homology arm nucleotide sequence 5' of the editing element having homology to a first genomic region 5' to the target locus; and (iii) a 3' homology arm nucleotide sequence 3' of the editing element having homology to a second genomic region 3' to the target locus,

without co-administering an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

11. The method of any one of claims 6-10, wherein the disease or disorder is thalassemia or sickle cell disease.

12. The method of any one of claims 6-10, wherein the subject is a human subject.

13. The method of any one of the preceding claims, wherein the target gene is the HBB gene.

14. The method of any one of the preceding claims, wherein the target locus is at a nucleotide point mutation, insertion or deletion in the HBB gene.

15. The method of claim 14, wherein the nucleotide point mutation or deletion in the HBB gene is selected from the group consisting of G at position -87, G at position -31, A at position -30, G at position -29, G at position -28, T at position -10, C at position 1, A at position 1, G at position 2, deletion of C and T at positions 17 and 18, A at position 19, deletion of A at position 20, T at position 20, deletion of A and A at positions 25 and 26, addition of G after position 26, A at position 47, A at position 48, deletion of C at position 51, A at position 52, G at position 58, G at position 59, A at position 79, T at position 82, addition of C after position 84, T at position 93, A at position 93, C at position 97, C at position 98, G at position 202, G at position 208, C at position 222, deletion of T at position 241 or 242, deletion of T and T and C and T at positions 254 to 257, T at position 260, deletion of C at position 264 or 265, addition of A after position 343, deletion of G and T at positions 399 and 400, T at position 401, addition of A after position 417, A at position 446, T at position 1099, A at position 1293, T at position 1344.

16. The method of claim 14 or 15, wherein the editing element comprises a portion of the wild-type HBB gene that corresponds to the mutation.

17. The method of any one of claims 1-16, wherein the editing element comprises the coding regions of one or more exons of an HBB gene.

18. The method of any one of claims 1-16, wherein the editing element comprises a portion of an HBB gene comprising the coding region of exon 1, the entire intron 1, the entire exon 2, the entire intron 2, and the coding region of exon 3.

19. The method of claim 17 or 18, wherein the coding regions have been silently altered to be less than 100%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50% identical to the corresponding exons of the wild-type HBB gene.

20. The method of claim 19, wherein the editing element comprises at least one of the nucleotide sequences selected from the group consisting of SEQ ID NOs: 43-46 and 105-107.

21. The method of any one of claims 1-12, wherein the target locus is AAVS1.

22. The method of any one of claims 1-16 and 21, wherein the editing element comprises a nucleotide sequence encoding SEQ ID NO: 48 or a portion thereof.

23. The method of claim 22, wherein the nucleotide sequence encoding SEQ ID NO: 48 consists of nucleotides 4 to 444 of SEQ ID NO: 27.

24. The method of claim 22, wherein the nucleotide sequence encoding SEQ ID NO: 48 is silently altered to be less than 70%, 75%, 80%, 85%, or 90% identical to nucleotides 4 to 444 of SEQ ID NO: 27.

25. The method of claim 24, wherein the nucleotide sequence encoding SEQ ID NO: 48 consists of the sequence of SEQ ID NO: 47 or 100.

26. The method of claim 22, wherein the nucleotide sequence encoding SEQ ID NO: 48 consists of a stuffer-inserted coding sequence of the HBB gene.

27. The method of any one of claims 22-26, wherein the target locus is the internucleotide bond between nucleotide 3 and nucleotide 4 of the target gene, whereby integration of the editing element at the target locus results in the target locus comprising an HBB coding sequence or stuffer-inserted coding sequence starting with the start codon of the target gene.

28. The method of any one of claims 22-26, wherein the editing element comprises an HBB coding sequence or stuffer-inserted coding sequence consisting of 5' to 3' a start codon and the nucleotide sequence encoding SEQ ID NO: 48, or a portion of the HBB coding sequence or stuffer-inserted coding sequence.

29. The method of claim 28, wherein the target locus is in an intron of the target gene, and wherein the editing element comprises 5' to 3' a splice acceptor site, a ribosomal skipping element, and an HBB coding sequence or stuffer-inserted coding sequence.

30. The method of claim 29, wherein the target locus is in intron 1 of the HBB gene.

31. The method of claim 28, wherein the target locus is adjacently 3' to a coding nucleotide of the target gene, and wherein the editing element comprises 5' to 3' a ribosomal skipping element and an HBB coding sequence or stuffer-inserted coding sequence.

32. The method of claim 31, wherein the target locus is the stop codon of a wild-type target gene or the corresponding nucleotides of a mutant target gene.

33. The method of any one of claims 22-32, wherein the editing element further comprises an exogenous polyadenylation sequence 3' to the nucleotide sequence encoding SEQ ID NO: 48.

34. The method of any one of the preceding claims, wherein the editing element further comprises a restriction endonuclease site not present in the target gene.

35. The method of any one of the preceding claims, wherein the editing element comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 23-28.

36. The method of any one of the preceding claims, wherein the 5' homology arm nucleotide sequence is at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the first genomic region.

37. The method of any one of the preceding claims, wherein the 3' homology arm nucleotide sequence is at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the second genomic region.

38. The method of any one of the preceding claims, wherein the first genomic region is located in a first editing window, and the second genomic region is located in a second editing window.

39. The method of claim 38, wherein the first and second editing windows are different.

40. The method of claim 38, wherein the first and second editing windows are the same.

41. The method of any one of claims 38-40, wherein the first editing window consists of the nucleotide sequence set forth in SEQ ID NO: 101, 102, or 103.

42. The method of any one of claims 38-41, wherein the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 101, 102, or 103.

43. The method of any one of claims 38-42, wherein the first genomic region consists of the nucleotide sequence set forth in SEQ ID NO: 101.

44. The method of any one of claims 38-43, wherein the second genomic region consists of the nucleotide sequence set forth in SEQ ID NO: 102.

45. The method of any one of the preceding claims, wherein the 5' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 101.

46. The method of any one of the preceding claims, wherein the 3' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 102.

47. The method of any one of the preceding claims, wherein each of the 5' and 3' homology arm nucleotide sequences independently has a length of about 100 to about 2000 nucleotides.

48. The method of any one of the preceding claims, wherein the correction genome further comprises a 5' inverted terminal repeat (5' ITR) nucleotide sequence 5' of the 5' homology arm nucleotide sequence, and a 3' inverted terminal repeat (3' ITR) nucleotide sequence 3' of the 3' homology arm nucleotide sequence.

49. The method of claim 48, wherein the 5' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO:18, and the 3' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO:19.

50. The method of claim 48, wherein the 5' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO:20, and the 3' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO:21.

51. The method of any one of the preceding claims, wherein the correction genome comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 29-42 and 104.

52. The method of any one of the preceding claims, wherein the AAV Clade F capsid protein comprises an amino acid sequence having at least 95% sequence identity with the amino acid sequence of amino acids 203-736 of SEQ ID NO:2, wherein: the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G.

53. The method of claim 52, wherein:

(a) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G;

(b) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M;

(c) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R;

(d) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; or

(e) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C.

54. The method of claim 52, wherein the capsid protein comprises the amino acid sequence of amino acids 203-736 of SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17.

55. The method of any one of the preceding claims, wherein the AAV Clade F capsid protein comprises an amino acid sequence having at least 95% sequence identity with the amino acid sequence of amino acids 138-736 of SEQ ID NO:2, wherein: the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 2 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G.

56. The method of claim 55, wherein:

(a) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G;

(b) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M;

(c) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R;

(d) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; or

(e) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C.

57. The method of claim 55, wherein the capsid protein comprises the amino acid sequence of amino acids 138-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17.

58. The method of any one of the preceding claims, wherein the AAV Clade F capsid protein comprises an amino acid sequence having at least 95% sequence identity with the amino acid sequence of amino acids 1-736 of SEQ ID NO:2, wherein: the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 2 is T; the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 68 of SEQ ID NO: 2 is V; the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 2 is L; the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 2 is D; the amino acid in the

capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G.

59. The method of claim 58, wherein:

(a) the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 2 is T, and the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q;

(b) the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 2 is I, and the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is Y;

(c) the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K;

(d) the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 2 is L, and the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S;

(e) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G;

(f) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is

N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M;

(g) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R;

(h) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; or

(i) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C.

60. The method of claim 58, wherein the capsid protein comprises the amino acid sequence of amino acids 1-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

61. The method of any one of the preceding claims, wherein the integration efficiency of the editing element into the target locus is at least 1% when the AAV is contacted *in vitro* in the absence of an exogenous nuclease with a population of $CD34^+$ human stem cells under standard AAV transduction conditions.

62. The method of any one of the preceding claims, wherein the allelic frequency of integration of the editing element into the target locus is at least 0.5% when the AAV is contacted *in vitro* in the absence of an exogenous nuclease with a population of $CD34^+$ human stem cells under standard AAV transduction conditions.

63. A replication-defective adeno-associated virus (AAV) comprising:

a) an AAV capsid comprising an AAV Clade F capsid protein; and

b) a correction genome comprising (i) an editing element for editing a target locus in a target gene, (ii) a 5' homology arm nucleotide sequence 5' of the editing element having homology to a first genomic region 5' to the target locus, and (iii) a 3' homology arm

nucleotide sequence 3' of the editing element having homology to a second genomic region 3' to the target locus.

64. The AAV of claim 63, wherein the target gene is the HBB gene.

65. The AAV of claim 63 or 64, wherein the target locus is at a nucleotide point mutation or deletion in the HBB gene.

66. The AAV of claim 65, wherein the nucleotide point mutation or deletion in the HBB gene is selected from the group consisting of G at position -87, G at position -31, A at position -30, G at position -29, G at position -28, T at position -10, C at position 1, A at position 1, G at position 2, deletion of C and T at positions 17 and 18, A at position 19, deletion of A at position 20, T at position 20, deletion of A and A at positions 25 and 26, addition of G after position 26, A at position 47, A at position 48, deletion of C at position 51, A at position 52, G at position 58, G at position 59, A at position 79, T at position 82, addition of C after position 84, T at position 93, A at position 93, C at position 97, C at position 98, G at position 202, G at position 208, C at position 222, deletion of T at position 241 or 242, deletion of T and T and C and T at positions 254 to 257, T at position 260, deletion of C at position 264 or 265, addition of A after position 343, deletion of G and T at positions 399 and 400, T at position 401, addition of A after position 417, A at position 446, T at position 1099, A at position 1293, T at position 1344.

67. The AAV of claim 65 or 66, wherein the editing element comprises a portion of the wild-type HBB gene that corresponds to the mutation.

68. The AAV of any one of claims 63-67, wherein the editing element comprises the coding regions of one or more exons of an HBB gene.

69. The AAV of any one of claims 63-68, wherein the editing element comprises a portion of an HBB gene comprising the coding region of exon 1, the entire intron 1, the entire exon 2, the entire intron 2, and the coding region of exon 3.

70. The AAV of claim 68 or 69, wherein the coding regions have been silently altered to be less than 100%, 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50% identical to the corresponding exons of the wild-type HBB gene.

71. The AAV of claim 70, wherein the editing element comprises at least one of the nucleotide sequences selected from the group consisting of SEQ ID NOs: 43-46 and 105-107.

72. The AAV of claim 63, wherein the target locus is AAVS1.

73. The AAV of any one of claims 63-66 and 72, wherein the editing element comprises a nucleotide sequence encoding SEQ ID NO: 48 or a portion thereof.

74. The AAV of claim 73, wherein the nucleotide sequence encoding SEQ ID NO: 48 consists of nucleotides 4 to 444 of SEQ ID NO: 27.

75. The AAV of claim 73, wherein the nucleotide sequence encoding SEQ ID NO: 48 is silently altered to be less than 70%, 75%, 80%, 85%, or 90% identical to nucleotides 4 to 444 of SEQ ID NO: 27.

76. The AAV of claim 75, wherein the nucleotide sequence encoding SEQ ID NO: 48 consists of the sequence of SEQ ID NO: 47 or 100.

77. The AAV of claim 73, wherein the nucleotide sequence encoding SEQ ID NO: 48 consists of a stuffer-inserted coding sequence of the HBB gene.

78. The AAV of any one of claims 73-77, wherein the target locus is the internucleotide bond between nucleotide 3 and nucleotide 4 of the target gene, whereby integration of the editing element at the target locus results in the target locus comprising an HBB coding sequence or stuffer-inserted coding sequence starting with the start codon of the target gene.

79. The AAV of any one of claims 73-77, wherein the editing element comprises an HBB coding sequence or stuffer-inserted coding sequence consisting of 5' to 3' a start codon and the nucleotide sequence encoding SEQ ID NO: 48, or a portion of the HBB coding sequence or stuffer-inserted coding sequence.

80. The AAV of claim 79, wherein the target locus is in an intron of the target gene, and wherein the editing element comprises 5' to 3' a splice acceptor site, a ribosomal skipping element, and an HBB coding sequence or stuffer-inserted coding sequence.

81. The AAV of claim 80, wherein the target locus is in intron 1 of the HBB gene.

82. The AAV of claim 79, wherein the target locus is adjacently 3' to a coding nucleotide of the target gene, and wherein the editing element comprises 5' to 3' a ribosomal skipping element and an HBB coding sequence or stuffer-inserted coding sequence.

83. The AAV of claim 82, wherein the target locus is the stop codon of a wild-type target gene or the corresponding nucleotides of a mutant target gene.

84. The AAV of any one of claims 73-83, wherein the editing element further comprises an exogenous polyadenylation sequence 3' to the nucleotide sequence encoding SEQ ID NO: 48.

85. The AAV of any one of claims 63-84, wherein the editing element further comprises a restriction endonuclease site not present in the target gene.

86. The AAV of any one of claims 63-85, wherein the editing element comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 23-28.

87. The AAV of any one of claims 63-86, wherein the 5' homology arm nucleotide sequence is at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the first genomic region.

88. The AAV of any one of claims 63-87, wherein the 3' homology arm nucleotide sequence is at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the second genomic region.

89. The AAV of any one of claims 62-88, wherein the first genomic region is located in a first editing window, and the second genomic region is located in a second editing window.

90. The AAV of claim 89, wherein the first and second editing windows are different.

91. The AAV of claim 89, wherein the first and second editing windows are the same.

92. The AAV of any one of claims 89-91, wherein the first editing window consists of the nucleotide sequence set forth in SEQ ID NO: 101, 102, or 103.

93. The AAV of any one of claims 89-92, wherein the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 101, 102, or 103.

94. The AAV of any one of claims 89-93, wherein the first genomic region consists of the nucleotide sequence set forth in SEQ ID NO: 101.

95. The AAV of any one of claims 89-94, wherein the second genomic region consists of the nucleotide sequence set forth in SEQ ID NO: 102.

96. The AAV of any one of claims 63-95, wherein the 5' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 101.

97. The AAV of any one of claims 63-96, wherein the 3' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 102.

98. The AAV of any one of the claims 63-97, wherein each of the 5’ and 3’ homology arm nucleotide sequences independently has a length of about 100 to about 2000 nucleotides.

99. The AAV of any one of the claims 63-98, wherein the correction genome further comprises a 5’ inverted terminal repeat (5’ ITR) nucleotide sequence 5’ of the 5’ homology arm nucleotide sequence, and a 3’ inverted terminal repeat (3’ ITR) nucleotide sequence 3’ of the 3’ homology arm nucleotide sequence.

100. The AAV of claim 99, wherein the 5’ ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO:18, and the 3’ ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO:19.

101. The AAV of claim 99, wherein the 5' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO:20, and the 3' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO:21.

102. The AAV of any one of claims 63-101, wherein the correction genome comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 29-42 and 104.

103. The AAV of any one of claims 63-102, wherein the AAV Clade F capsid protein comprises an amino acid sequence having at least 95% sequence identity with the amino acid sequence of amino acids 203-736 of SEQ ID NO:2, wherein: the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G.

104. The AAV of claim 103, wherein:

(a) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G;

(b) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is

R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M;

(c) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R;

(d) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; or

(e) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C.

105. The AAV of claim 103, wherein the capsid protein comprises the amino acid sequence of amino acids 203-736 of SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17.

106. The AAV of any one of claims 63-105, wherein the AAV Clade F capsid protein comprises an amino acid sequence having at least 95% sequence identity with the amino acid sequence of amino acids 138-736 of SEQ ID NO:2, wherein: the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 2 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the

capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G.

107. The AAV of claim 106, wherein:

(a) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G;

(b) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M;

(c) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R;

(d) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; or

(e) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C.

108. The AAV of claim 106, wherein the capsid protein comprises the amino acid sequence of amino acids 138-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17.

109. The AAV of any one of claims 63-108, wherein the AAV Clade F capsid protein comprises an amino acid sequence having at least 95% sequence identity with the amino acid sequence of amino acids 1-736 of SEQ ID NO:2, wherein: the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 2 is T; the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 68 of SEQ ID NO: 2 is V; the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 2 is R; the amino acid in the capsid

protein corresponding to amino acid 119 of SEQ ID NO: 2 is L; the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 2 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G.

110. The AAV of claim 109, wherein:

(a) the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 2 is T, and the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q;

(b) the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 2 is I, and the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is Y;

(c) the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K;

(d) the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 2 is L, and the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S;

(e) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G;

(f) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M;

(g) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R;

(h) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; or

(i) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C.

111. The AAV of claim 109, wherein the capsid protein comprises the amino acid sequence of amino acids 1-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

112. The AAV of any one of claims 63-111, wherein the integration efficiency of the editing element into the target locus is at least 1% when the AAV is contacted *in vitro* in the absence of an exogenous nuclease with a population of $CD34^+$ human stem cells under standard AAV transduction conditions.

113. The AAV of any one of claims 63-112, wherein the allelic frequency of integration of the editing element into the target locus is at least 0.5% when the AAV is contacted *in vitro* in the absence of an exogenous nuclease with a population of $CD34^+$ human stem cells under standard AAV transduction conditions.

114. A pharmaceutical composition comprising an AAV of any one of the claims 62-111.

115. A packaging system for recombinant preparation of an AAV, wherein the packaging system comprises

a) a Rep nucleotide sequence encoding one or more AAV Rep proteins;

b) a Cap nucleotide sequence encoding one or more AAV Clade F capsid proteins as set forth in any one of claims 103-111; and

c) a correction genome as set forth in any one of claims 63-102, wherein the packaging system is operative in a cell for enclosing the correction genome in the capsid to form the AAV.

116. The packaging system of claim 115, wherein the packaging system comprises a first vector comprising the Rep nucleotide sequence and the Cap nucleotide sequence, and a second vector comprising the correction genome.

117. The packaging system of 115 or 116, wherein the Rep nucleotide sequence encodes an AAV2 Rep protein.

118. The packaging system of claim 117, wherein the AAV2 Rep protein is 78/68 or Rep 68/52.

119. The packaging system of claim 117 or 118, wherein the AAV2 Rep protein comprises an amino acid sequence having a minimum percent sequence identity to the AAV2 Rep amino acid sequence of SEQ ID NO:22, wherein the minimum percent sequence identity is at least 70% across the length of the amino acid sequence encoding the AAV2 Rep protein.

120. The packaging system of any one of claims 115-119, further comprising a third vector, wherein the third vector is a helper virus vector.

121. The packaging system of claim 120, wherein the helper virus vector is an independent third vector.

122. The packaging system of claim 120, wherein the helper virus vector is integral with the first vector.

123. The packaging system of claim 120, wherein the helper virus vector is integral with the second vector.

124. The packaging system of any one of claims 120-123, wherein the third vector comprises genes encoding helper virus proteins.

125. The packaging system of any one of claims 120-124 wherein the helper virus is selected from the group consisting of adenovirus, herpes virus, vaccinia virus, and cytomegalovirus (CMV).

126. The packaging system of claim 125, wherein the helper virus is adenovirus.

127. The packaging system of claim 126, wherein the adenovirus genome comprises one or more adenovirus RNA genes selected from the group consisting of E1, E2, E4 and VA.

128. The packaging system of claim 125, wherein the helper virus is herpes simplex virus (HSV).

129. The packaging system of claim 128, wherein the HSV genome comprises one or more of HSV genes selected from the group consisting of UL5/8/52, ICPO, ICP4, ICP22 and UL30/UL42.

130. The packaging system of any one of claims 120-129, wherein the first vector and the third vector are contained within a first transfecting plasmid.

131. The packaging system of any one of claims 120-129, wherein the nucleotides of the second vector and the third vector are contained within a second transfecting plasmid.

132. The packaging system of any one of claims 120-129, wherein the nucleotides of the first vector and the third vector are cloned into a recombinant helper virus.

133. The packaging system of any one of claims 120-129, wherein the nucleotides of the second vector and the third vector are cloned into a recombinant helper virus.

134. A method for recombinant preparation of an AAV, the method comprising introducing the packaging system of any one of claims 115-133 into a cell under conditions operative for enclosing the correction genome in the capsid to form the AAV.

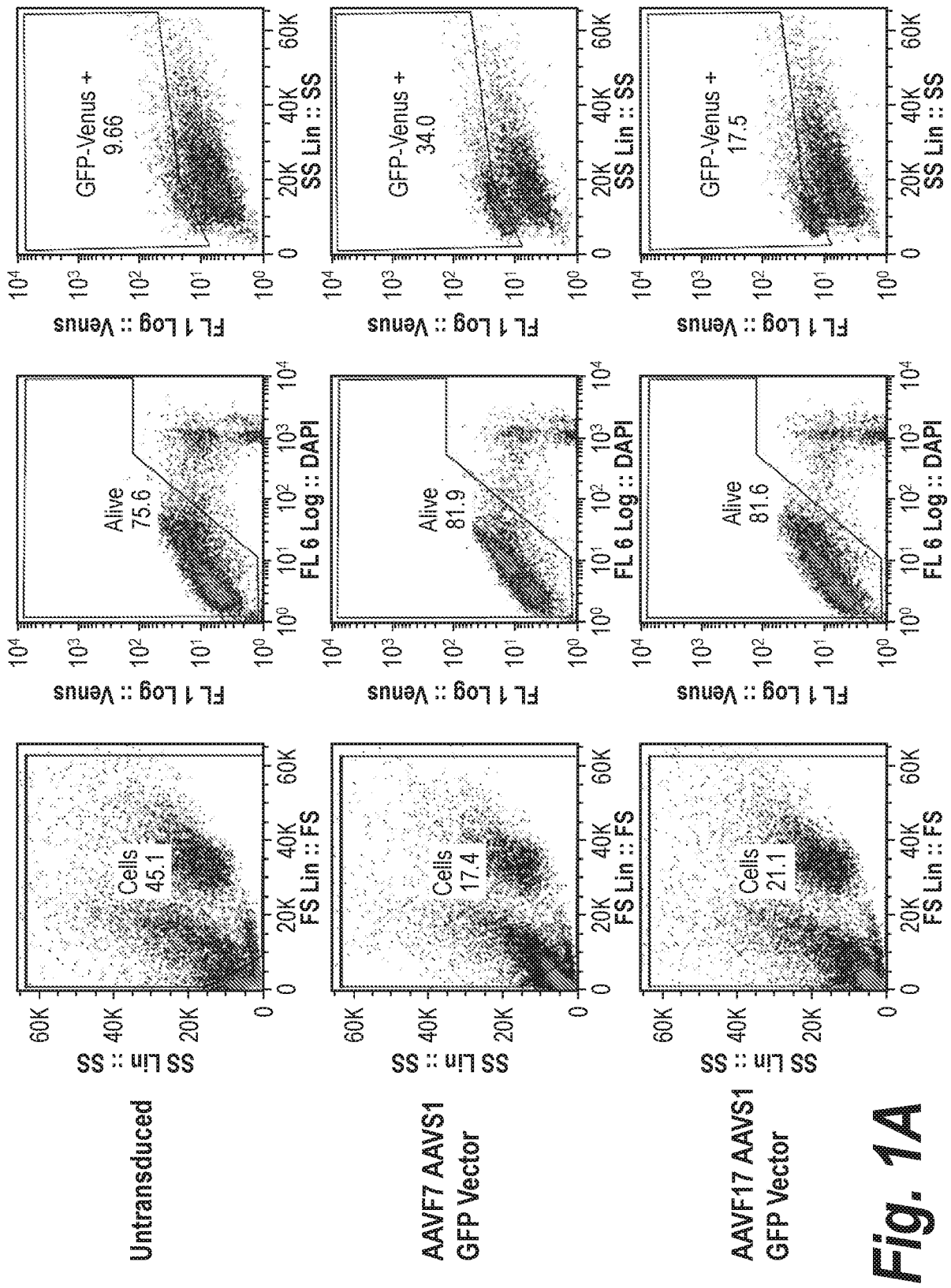

Untransduced
AAVF7 AAVS1 GFP Vector
AAVF17 AAVS1 GFP Vector
SS Lin :: SS
FS Lin :: FS
Cells 45.1
Cells 17.4
Cells 21.1
FL 6 Log :: DAPI
Alive 75.6
Alive 81.9
Alive 81.6
FL 1 Log :: Venus
GFP-Venus + 9.66
GFP-Venus + 34.0
GFP-Venus + 17.5
SS Lin :: SS

Fig. 1A

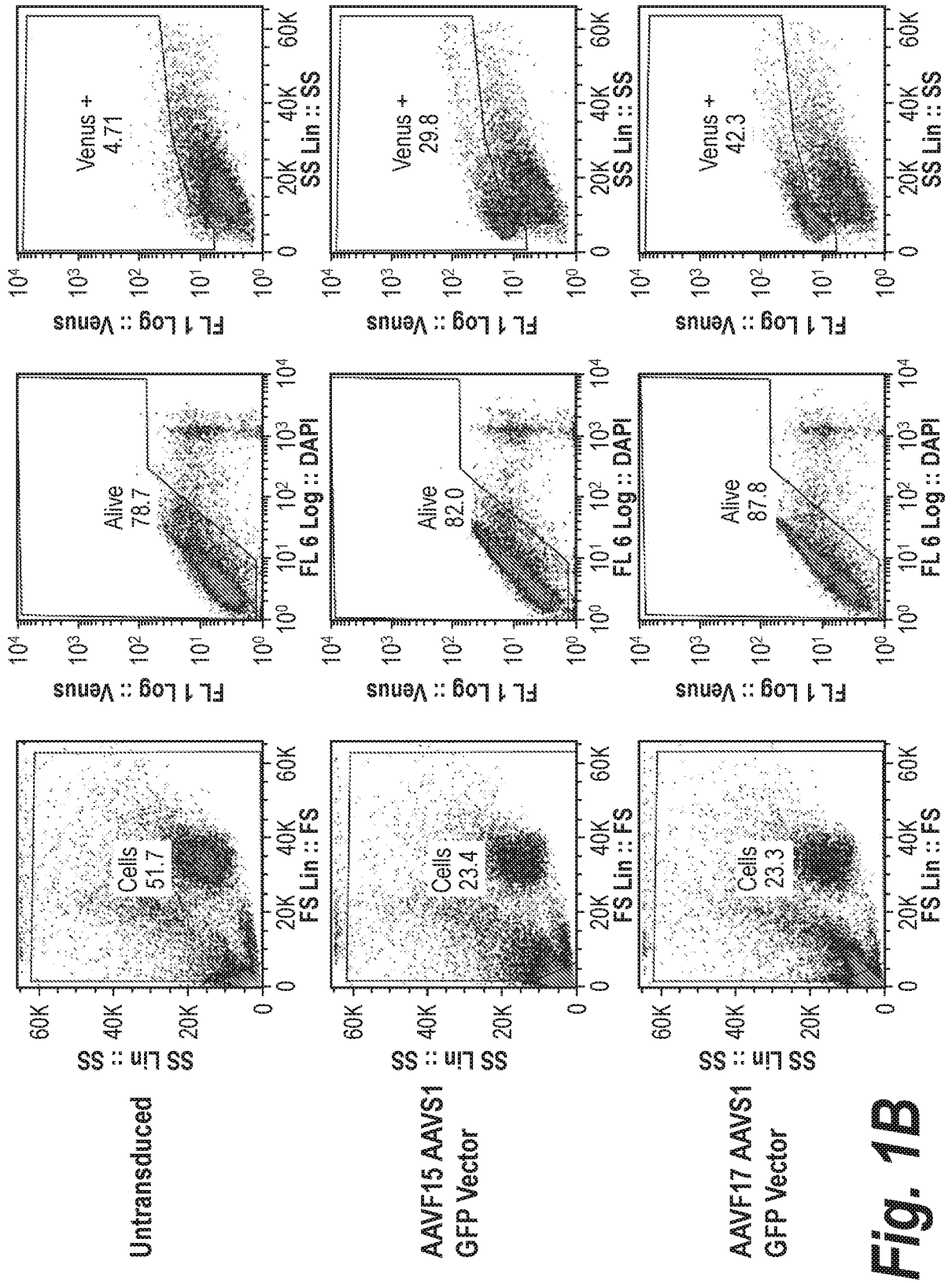

Untransduced
AAVF15 AAVS1 GFP Vector
AAVF17 AAVS1 GFP Vector
SS Lin :: SS
FS Lin :: FS
Cells 51.7
Cells 23.4
Cells 23.3
FL 6 Log :: DAPI
Alive 78.7
Alive 82.0
Alive 87.8
FL 1 Log :: Venus
Venus + 4.71
Venus + 29.8
Venus + 42.3
0
20K
40K
60K
10^0
10^1
10^2
10^3
10^4

Fig. 1B

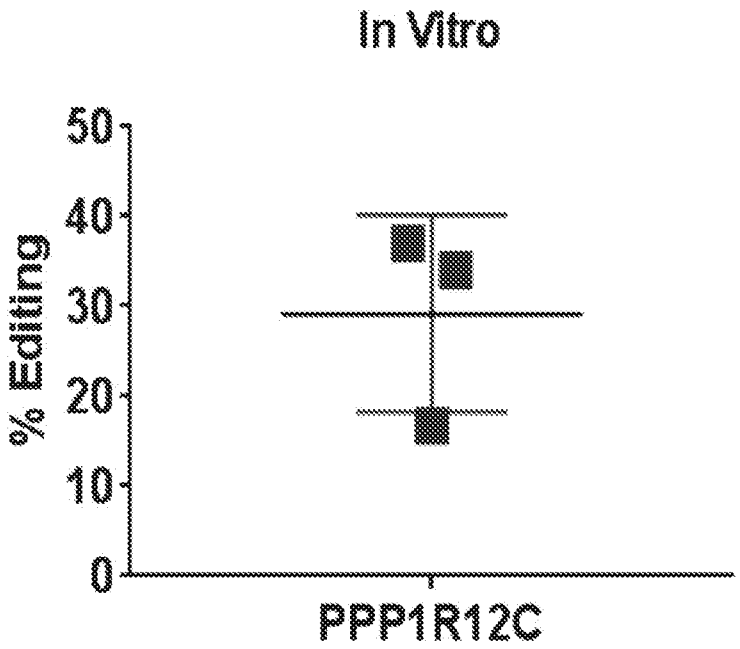


Fig. 1C

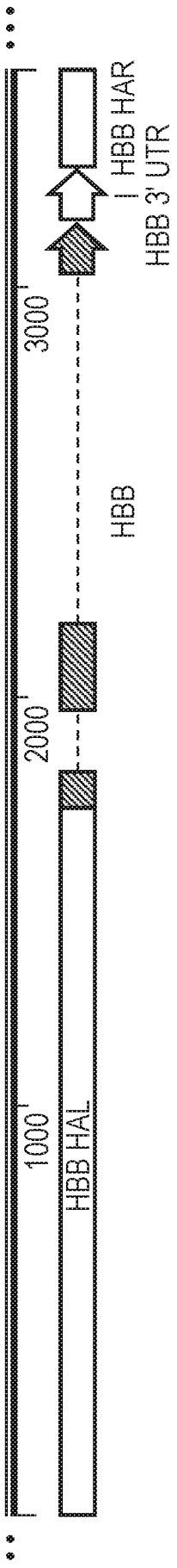


Fig. 2

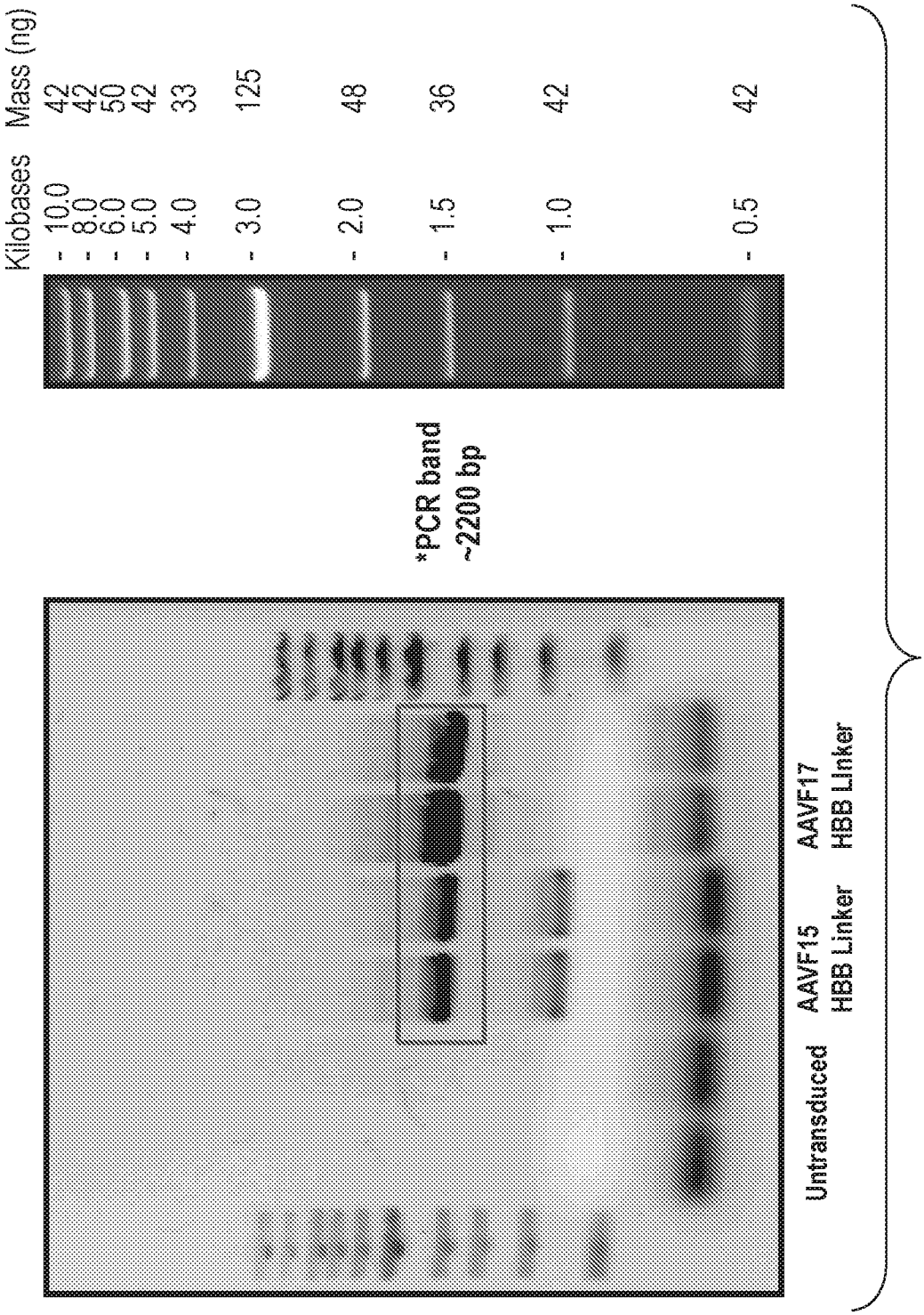
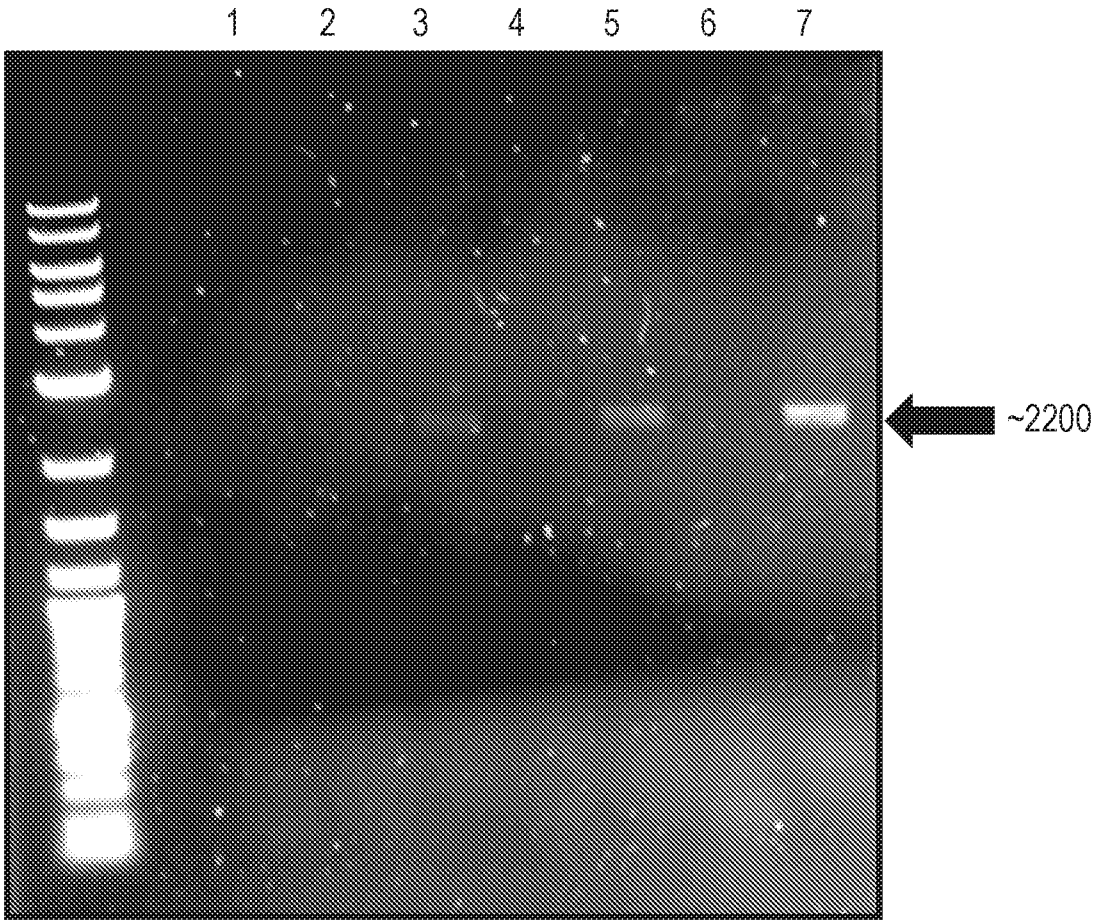

Kilobases
Mass (ng)
10.0
8.0
6.0
5.0
4.0
3.0
2.0
1.5
1.0
0.5
42
42
50
42
33
125
48
36
42
42
*PCR band
~2200 bp
Untransduced
AAVF15
HBB Linker
AAVF17
HBB Linker

Fig. 3A



Lanes:
1. Negative Control
2. 16265 SCD LCLs untransduced
3. 16265 SCD LCLs transduced with AAVHSC17-hHBB-ht-014 vector
4. 16266 SCD LCLs untransduced
5. 16266 SCD LCLs transduced with AAVHSC17-hHBB-hL-014 vector
6. 16267 SCD LCLs untransduced
7. 16267 SCD LCLs transduced with AAVHSC17-hHBB-hL-014 vector

Fig. 3B

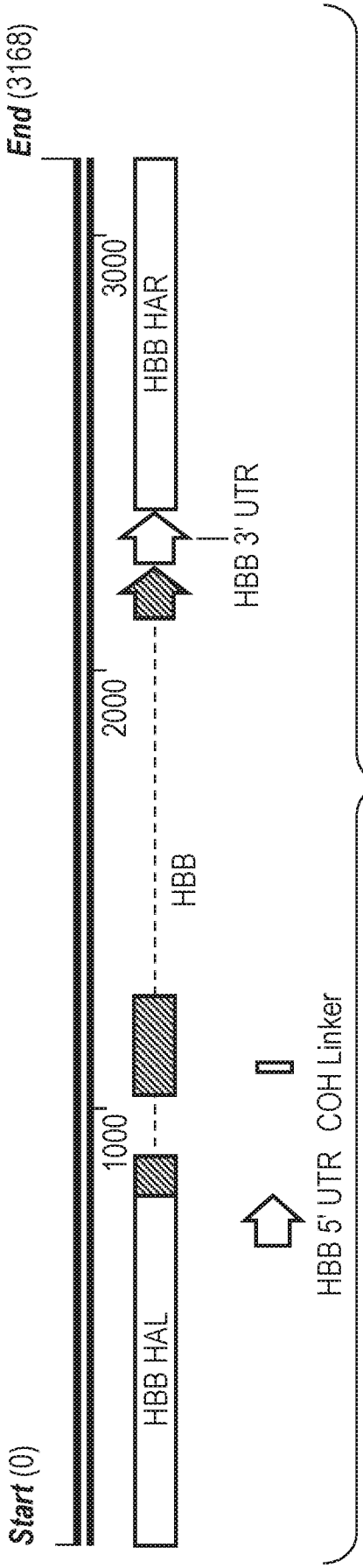

Start (0)
End (3168)
HBB HAL
HBB HAR
1000
2000
3000
HBB
HBB 5' UTR
COH Linker
HBB 3' UTR

Fig. 4A

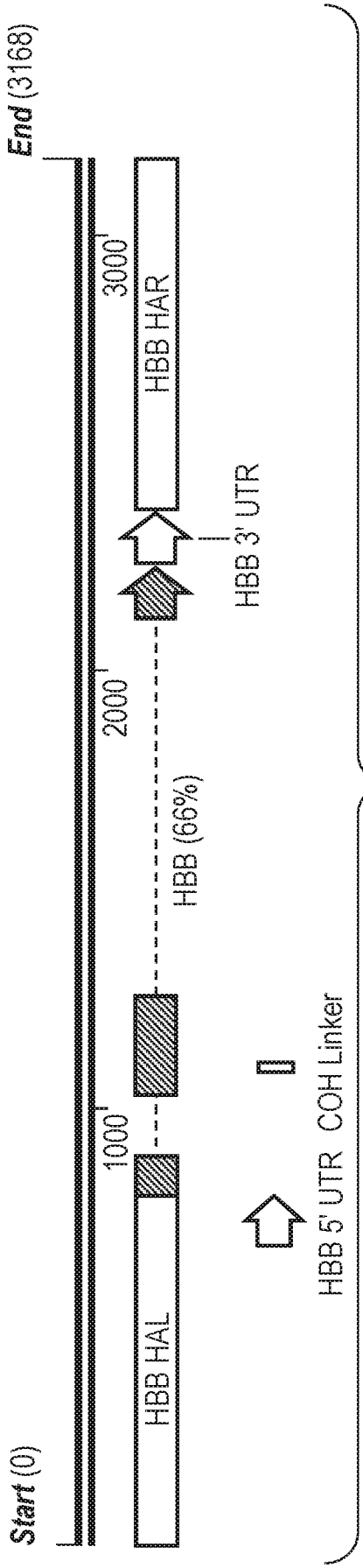

Start (0)
End (3168)
HBB HAL
HBB HAR
1000
2000
3000
HBB (66%)
HBB 5' UTR
COH Linker
HBB 3' UTR

Fig. 4B

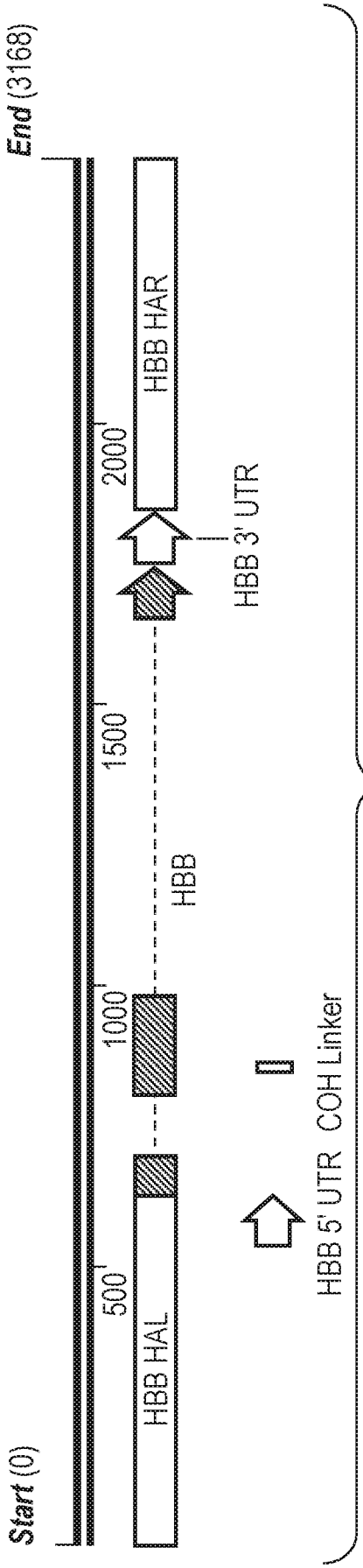

Start (0)
End (3168)
500
1000
1500
2000
HBB HAL
HBB
HBB HAR
HBB 5' UTR
COH Linker
HBB 3' UTR

Fig. 4C

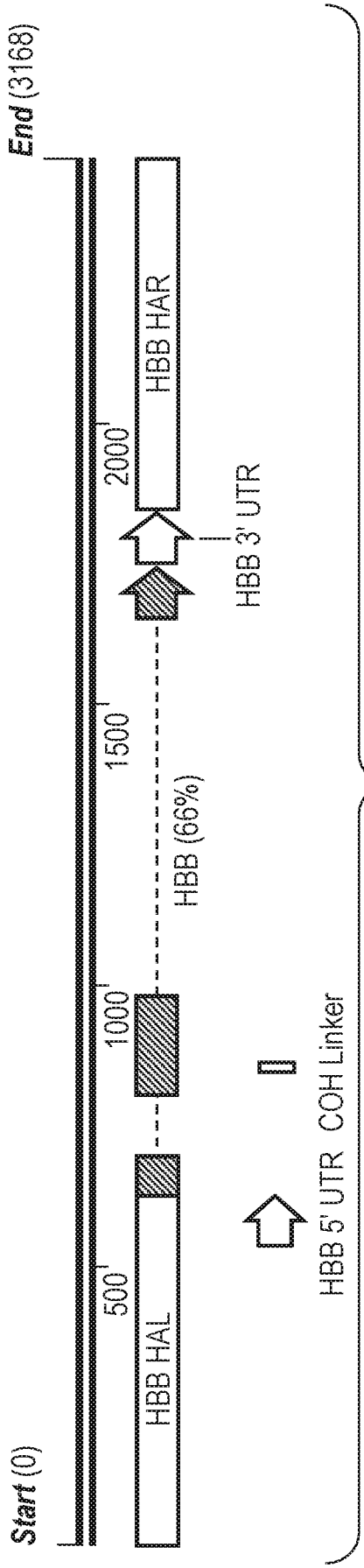

Start (0)
End (3168)
500
1000
1500
2000
HBB HAL
HBB (66%)
HBB HAR
HBB 5' UTR
COH Linker
HBB 3' UTR

Fig. 4D

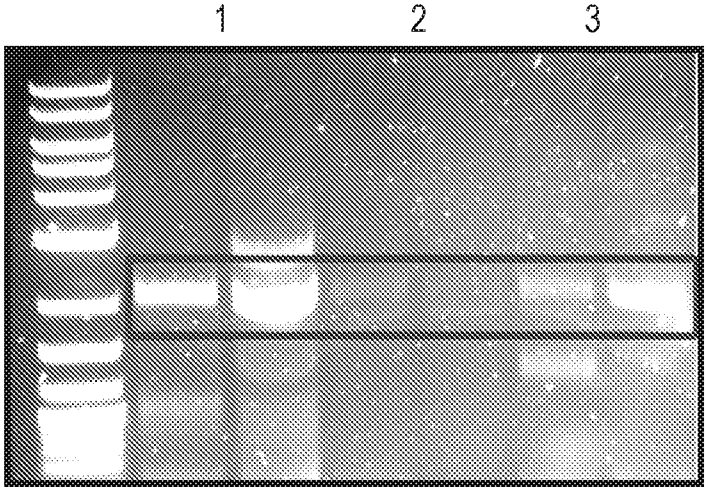


Fig. 5A

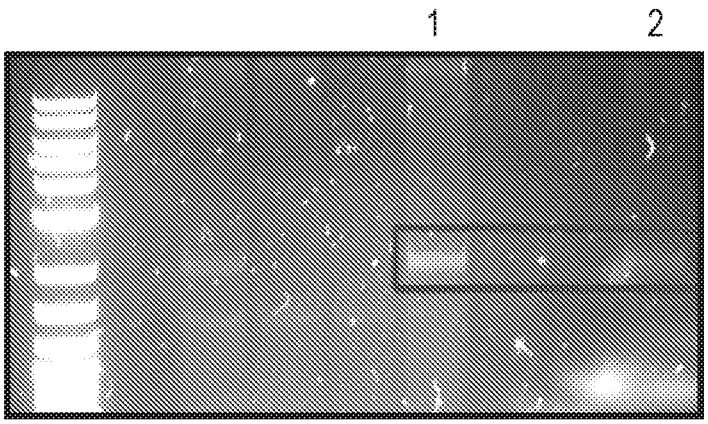


Fig. 5B

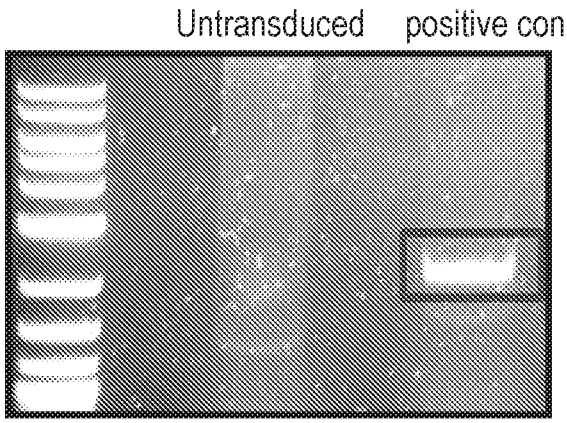


Fig. 5C

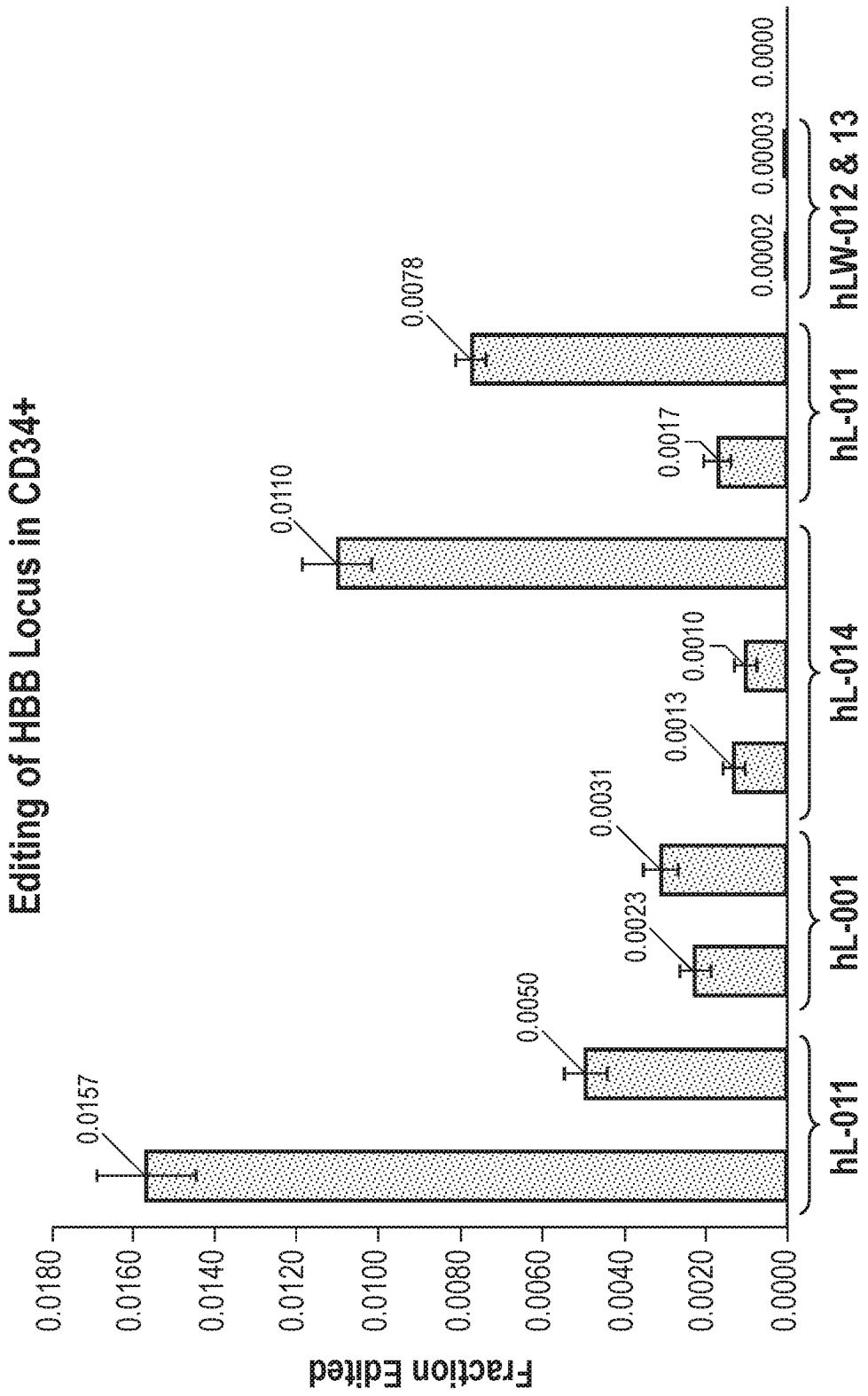

Editing of HBB Locus in CD34+
Fraction Edited
0.0180
0.0160
0.0140
0.0120
0.0100
0.0080
0.0060
0.0040
0.0020
0.0000
0.0157
0.0050
0.0023
0.0031
0.0013
0.0010
0.0110
0.0017
0.0078
0.00002
0.00003
0.0000
hL-011
hL-001
hL-014
hL-011
hLW-012 & 13

Fig. 6

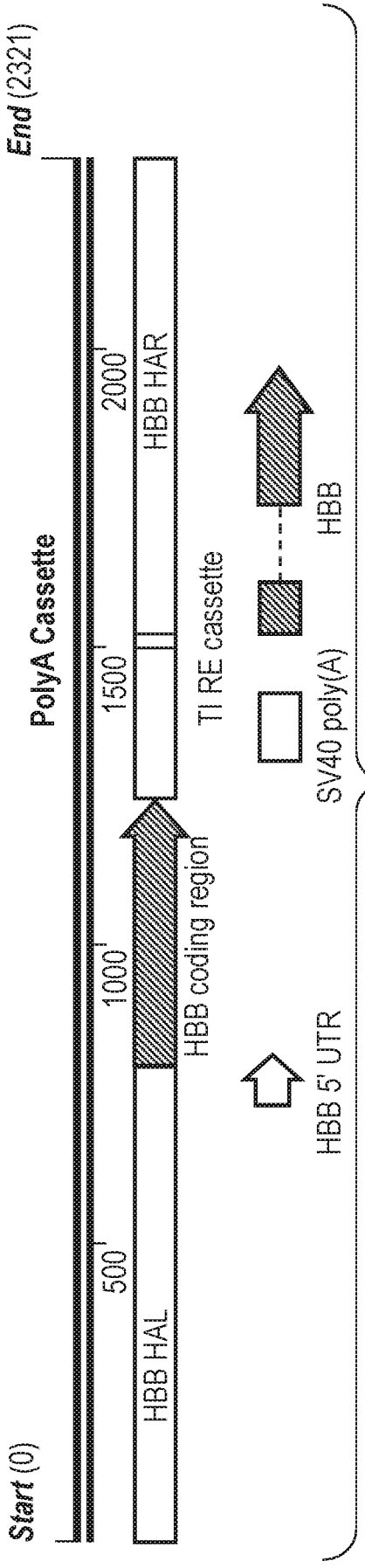

Start (0)
End (2321)
PolyA Cassette
500
1000
1500
2000
HBB HAL
HBB coding region
HBB HAR
TI RE cassette
HBB 5' UTR
SV40 poly(A)
HBB

Fig. 7A

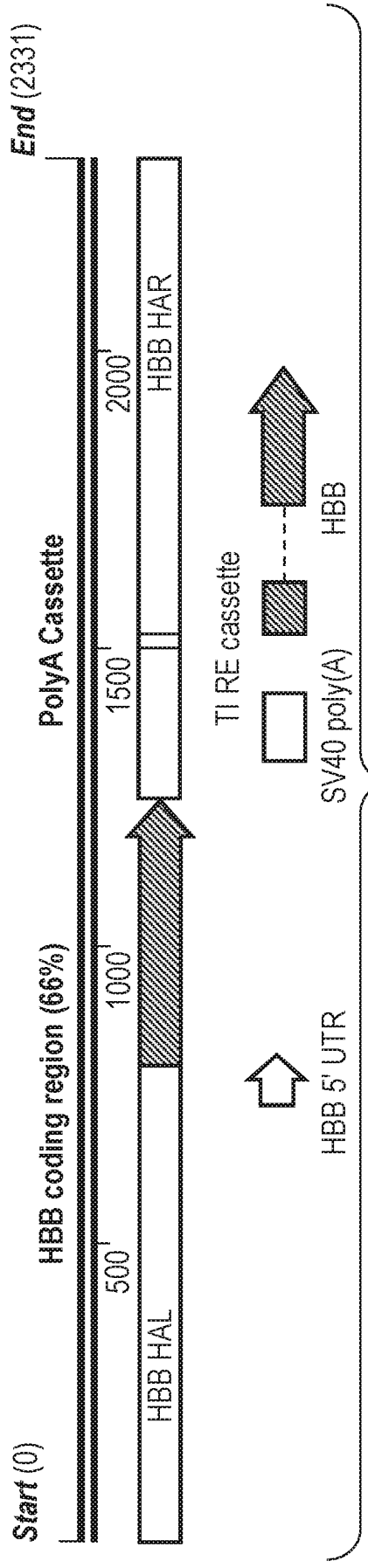

Start (0)
End (2331)
PolyA Cassette
HBB coding region (66%)
500
1000
1500
2000
HBB HAL
HBB HAR
TI RE cassette
HBB 5' UTR
SV40 poly(A)
HBB

Fig. 7B

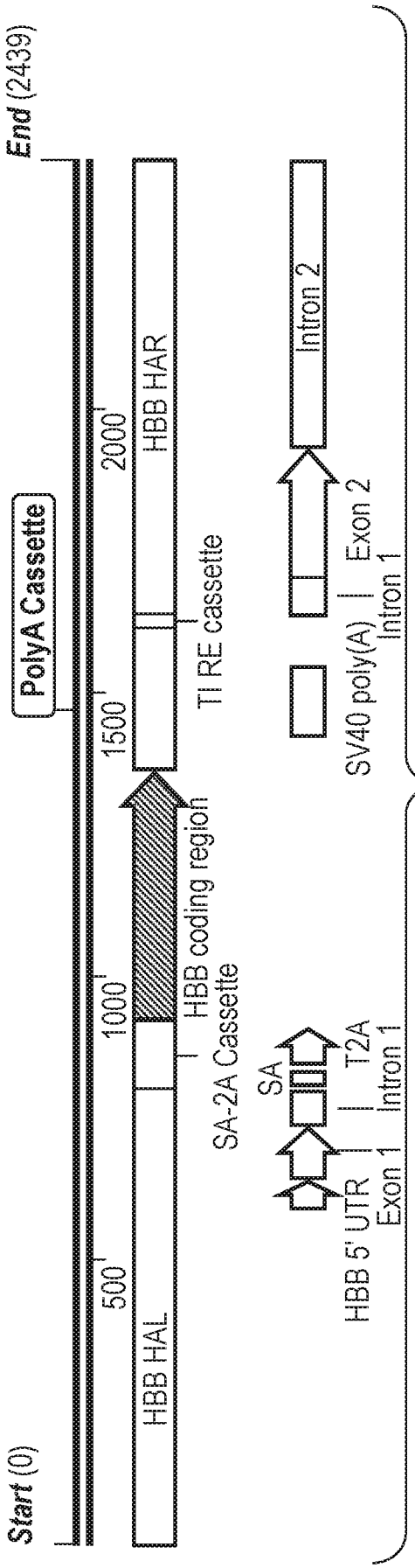

Start (0)
500
1000
1500
2000
End (2439)
HBB HAL
HBB coding region
SA-2A Cassette
PolyA Cassette
TI RE cassette
HBB HAR
HBB 5' UTR
Exon 1
SA
T2A
Intron 1
SV40 poly(A)
Intron 1
Exon 2
Intron 2

Fig. 7C

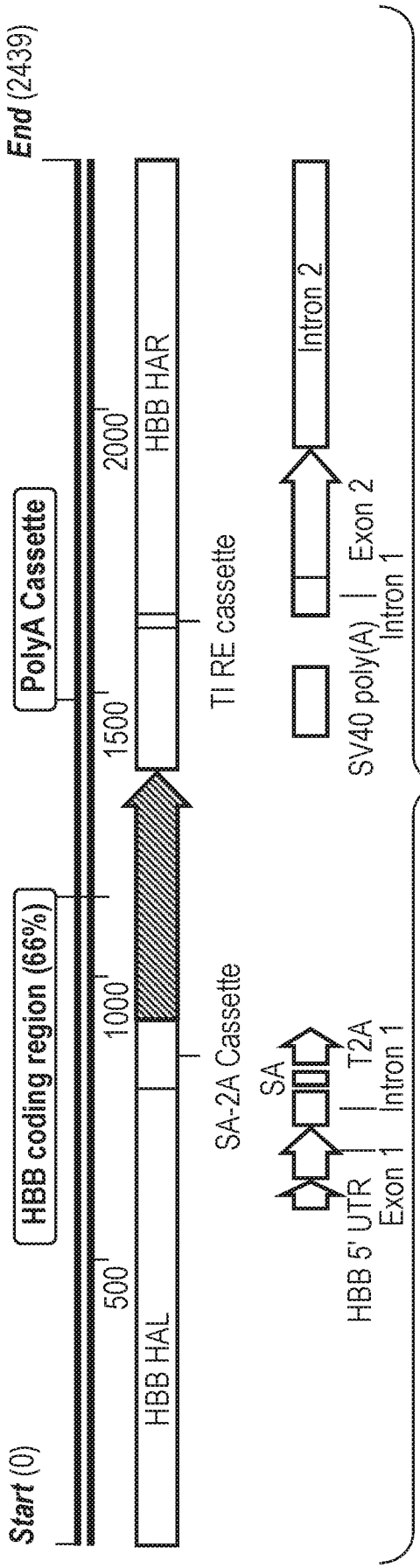
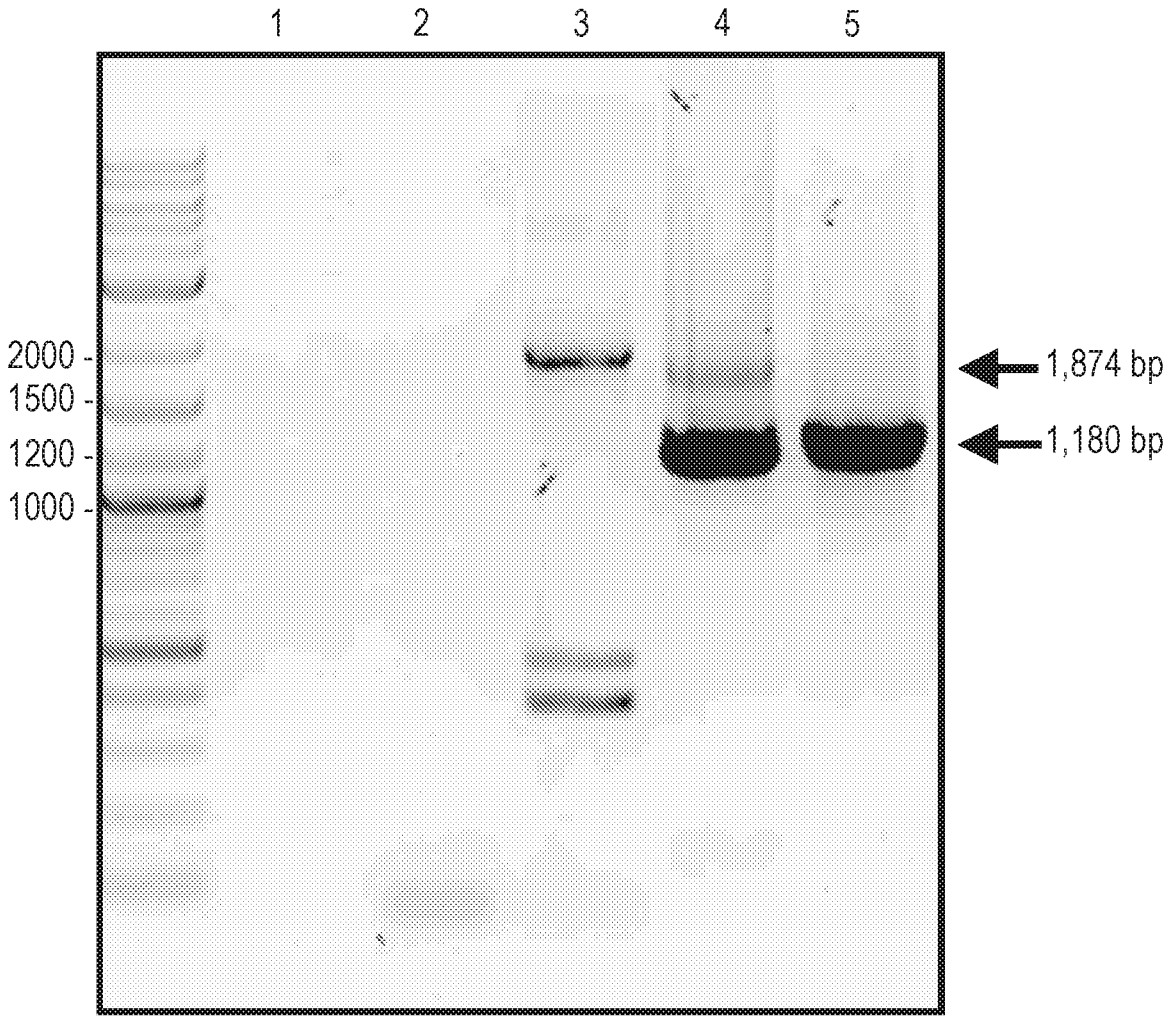

Start (0)
500
1000
1500
2000
End (2439)
HBB HAL
HBB coding region (66%)
SA-2A Cassette
PolyA Cassette
TI RE cassette
HBB HAR
HBB 5' UTR
Exon 1
SA
T2A
Intron 1
SV40 poly(A)
Intron 1
Exon 2
Intron 2

Fig. 7D


500
1000
1500
2000
HAL
P2A
Silently altered HBB coding sequence
HBB Exon 3
SV40_pA
HAR

Fig. 7E



Lanes:

1. negative control
2. untransduced cells
3. RKO cells transduced with hHBB-h1W-002
4. RKO cells transduced with hHBB-h1-010
5. LCL cells transduced with hHBB-h1-010

Fig. 8

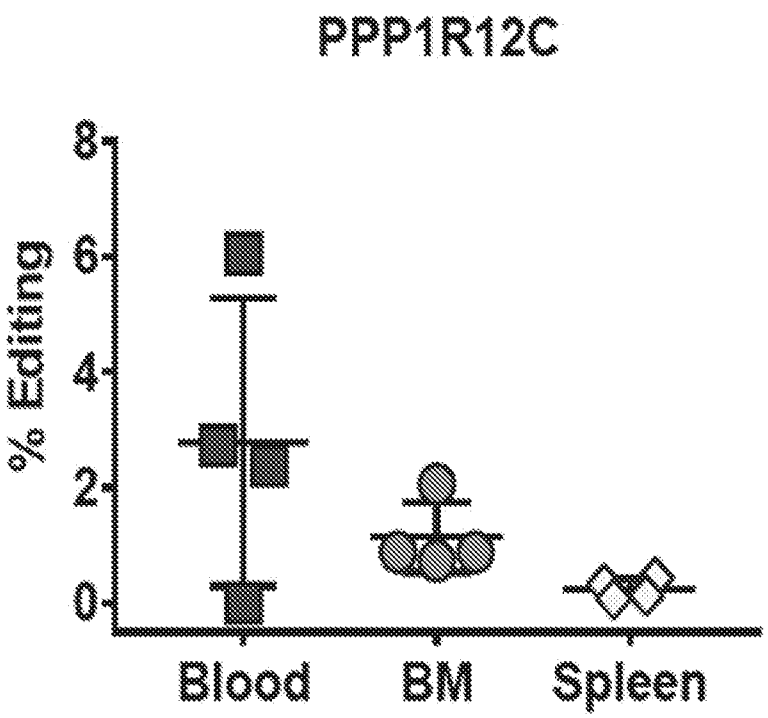


Fig. 9