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

Innovation, Science and Economic Development Canada
Canadian Intellectual Property Office

CA 3090226 A1 2019/08/08

(21) 3 090 226

(12) DEMANDE DE BREVET CANADIEN
CANADIAN PATENT APPLICATION

(13) A1

(86) Date de dépôt PCT/PCT Filing Date: 2019/02/01
(87) Date publication PCT/PCT Publication Date: 2019/08/08
(85) Entrée phase nationale/National Entry: 2020/07/31
(86) N° demande PCT/PCT Application No.: US 2019/016354
(87) N° publication PCT/PCT Publication No.: 2019/152843
(30) Priorités/Priorities: 2018/02/01 (US62/625,149); 2018/05/16 (US62/672,377)

(51) Cl.Int./Int.Cl. *A61K 35/76* (2015.01), *A61K 35/761* (2015.01), *A61K 48/00* (2006.01), *A61P 3/00* (2006.01), *C12N 15/86* (2006.01), *C12N 15/861* (2006.01)

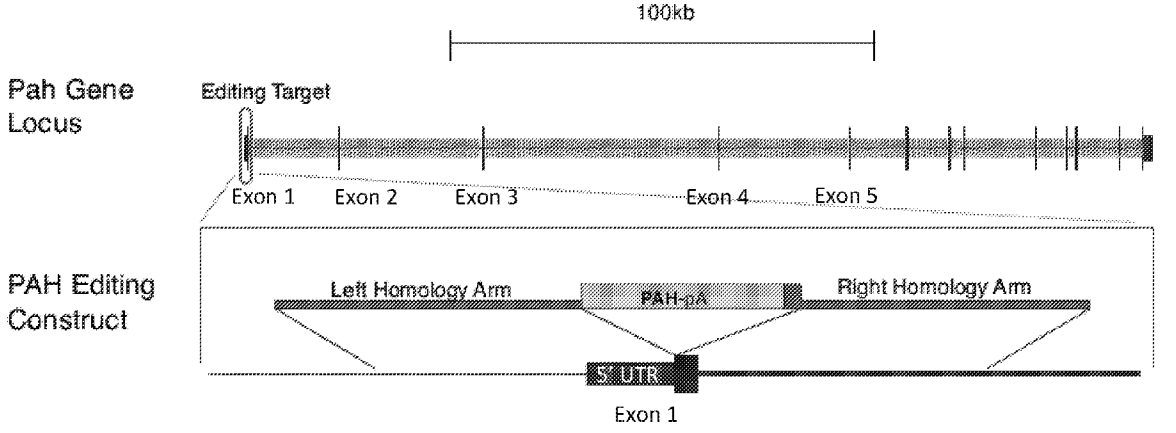
(71) Demandeur/Applicant:
HOMOLOGY MEDICINES, INC., US

(72) Inventeurs/Inventors:
SEYMOUR, ALBERT BARNES, US;
AHMED, SEEMIN SEHER, US;
WRIGHT, JASON BOKE, US;
DOLLIVE, SERENA NICOLE, US;
MCSWIGGEN, JAMES ANTHONY, US;
PROUT, JAIME MICHELLE, US; ...

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

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

Figure 4A



(57) Abrégé/Abstract:
Provided herein are adeno-associated virus (AAV) compositions that can restore phenylalanine hydroxylase (PAH) gene function in cell. Also provided are methods of use of the AAV compositions, and packaging systems for making the AAV compositions.

(21) 3 090 226

(13) A1

(72) Inventeurs(suite)/Inventors(continued): SOOKIASIAN, DANIELLE LAUREN, 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
08 August 2019 (08.08.2019)

WIPO | PCT

(10) International Publication Number
WO 2019/152843 A1

(51) International Patent Classification:
A61K 35/76 (2015.01) *A61P 3/00* (2006.01)
A61K 35/761 (2015.01) *C12N 15/86* (2006.01)
A61K 48/00 (2006.01) *C12N 15/861* (2006.01)

(21) International Application Number:
PCT/US2019/016354

(22) International Filing Date:
01 February 2019 (01.02.2019)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/625,149 01 February 2018 (01.02.2018) US
62/672,377 16 May 2018 (16.05.2018) US

(71) Applicant: **HOMOLOGY MEDICINES, INC.** [US/US]; 45 Wiggins Avenue, Bedford, MA 01730 (US).

(72) Inventors: **SEYMOUR, Albert, Barnes**; 1 Brickyard Lane, Westborough, MA 01581 (US). **AHMED, Seemin, Seher**; 81 Park Avenue, Apartment 10, Worcester, MA 01605 (US). **WRIGHT, Jason, Boke**; 57 Whits End Road, Concord, MA 01742 (US). **DOLLIVE, Serena, Nicole**; 19 Everett St., Unit 1, Waltham, MA 02453 (US). **MCSWIGGEN, James, Anthony**; 15 Beacon Street, Unit A, Arlington, MA 02474 (US). **PROUT, Jaime, Michelle**; 6 Raven Dr., Hudson, NH 03051 (US). **SOOKIASIAN, Danielle, Lauren**; 43 Estabrook Rd., Quincy, MA 02170 (US).

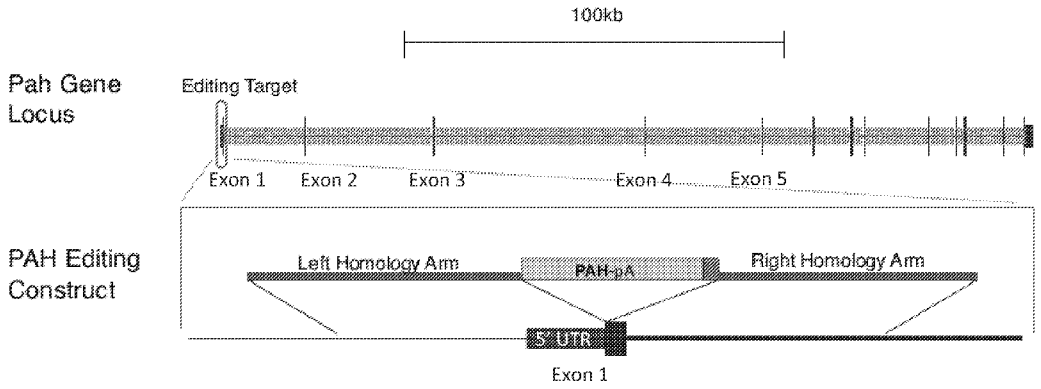
(74) Agent: **WILKINS, Andrew, T.** et al.; Lathrop Gage LLP, 28 State Street, Suite 700, Boston, MA 02109 (US).

(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).

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

Figure 4A



(57) Abstract: Provided herein are adeno-associated virus (AAV) compositions that can restore phenylalanine hydroxylase (PAH) gene function in cell. Also provided are methods of use of the AAV compositions, and packaging systems for making the AAV compositions.

[Continued on next page]

WO 2019/152843 A1

WO 2019/152843 A1

Declarations under Rule 4.17:

— *as to the identity of the inventor (Rule 4.17(i))*
— *as to applicant's entitlement to apply for and be granted a patent (Rule 4.17(ii))*

Published:

— *with international search report (Art. 21(3))*
— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*
— *with sequence listing part of description (Rule 5.2(a))*

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

RELATED APPLICATIONS

[0001] This application claims priority to U.S. Provisional Patent Application Serial Nos. 62/625,149, filed February 1, 2018, and 62/672,377, filed May 16, 2018, the entire disclosures of which are hereby incorporated herein by reference.

REFERENCE TO SEQUENCE LISTING SUBMITTED ELECTRONICALLY

[0002] The content of the electronically submitted sequence listing in ASCII text file (Name: HMT-024PC_SeqList_ST25.txt; Size: 367,287 bytes; and Date of Creation: January 30, 2019) is incorporated herein by reference in its entirety.

BACKGROUND

[0003] Phenylketonuria (PKU) is an autosomal recessive genetic disorder where the majority of cases are caused by mutations in the phenylalanine hydroxylase (PAH) gene. The PAH gene encodes a hepatic enzyme that catalyzes the hydroxylation of L-phenylalanine (Phe) to L-tyrosine (Tyr) upon multimerization. Reduction or loss of PAH activity leads to phenylalanine accumulation and its conversion into phenylpyruvate (also known as phenylketone). This abnormality in phenylalanine metabolism impairs neuronal maturation and the synthesis of myelin, resulting in mental retardation, seizures and other serious medical problems.

[0004] Currently, there is no cure for PKU. The standard of care is diet management by minimizing foods that contain high amounts of phenylalanine. Dietary management from birth with a low phenylalanine formula largely prevents the development of the neurological consequences of the disorder. However, even on a low-protein diet, children still suffer from growth retardation, and adults often have osteoporosis and vitamin deficiencies. Moreover, adherence to life-long dietary treatment is difficult, particularly beyond school age.

[0005] New treatment strategies have recently emerged, including large neutral amino acid (LNAA) supplementation, cofactor tetrahydrobiopterin therapy, enzyme replacement therapy, and genetically modified probiotic therapy. However, these strategies suffer from shortcomings. The LNAA supplementation is suitable only for adults not adhering to a low Phe diet. The cofactor tetrahydrobiopterin can only be used in some mild forms of PKU. Enzyme replacement by administration of a substitute for PAH, e.g., phenylalanine ammonia-lyase (PAL), can lead to immune responses that reduce the efficacy and/or cause side effects.

As to genetically modified probiotic therapy, the pathogenicity of PAL-expressing *E. coli* has been a concern.

[0006] Gene therapy provides a unique opportunity to cure PKU. Retroviral vectors, including lentiviral vectors, are capable of integrating nucleic acids into host cell genomes. However, these vectors may raise safety concerns due to their non-targeted insertion into the genome. For example, 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).

[0007] 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 defects in the PAH gene in PKU 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.

[0008] Accordingly, there is a need in the art for improved gene therapy compositions and methods that can efficiently and safely restore PAH gene function in PKU patients.

SUMMARY

[0009] Provided herein are adeno-associated virus (AAV) compositions that can restore PAH gene function in cells, and methods for using the same to treat diseases associated with reduction of PAH gene function (e.g., PKU). Also provided are packaging systems for making the adeno-associated virus compositions.

[0010] Accordingly, in one aspect, the instant disclosure provides a method for correcting a mutation in a phenylalanine hydroxylase (PAH) gene in a cell, the method comprising transducing the cell with a replication-defective adeno-associated virus (AAV) comprising:

(a) an AAV capsid; and

(b) a correction genome comprising: (i) an editing element for editing a target locus in the PAH 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.

[0011] In certain embodiments, the cell is a hepatocyte, a renal cell, or a cell in the brain, pituitary gland, adrenal gland, pancreas, urinary bladder, gallbladder, colon, small intestine, or breast. 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.

In another aspect, the instant disclosure provides a method for treating a subject having a disease or disorder associated with a PAH gene mutation, the method comprising administering to the subject an effective amount of a replication-defective AAV comprising:

(a) an AAV capsid; and

(b) a correction genome comprising: (i) an editing element for editing a target locus in the PAH 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 an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease is not co-administered to the subject.

[0012] In certain embodiments, the disease or disorder is phenylketonuria. In certain embodiments, the subject is a human subject.

[0013] In another aspect, the instant disclosure provides a replication-defective adeno-associated virus (AAV) comprising:

(a) an AAV capsid; and

(b) a correction genome comprising: (i) an editing element for editing a target locus in the PAH 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.

[0014] The following embodiments apply to each of the foregoing aspects.

[0015] In certain embodiments, the editing element comprises at least a portion of a PAH coding sequence. In certain embodiments, the editing element comprises a PAH coding sequence. In certain embodiments, the PAH coding sequence encodes an amino acid sequence set forth in SEQ ID NO: 23. In certain embodiments, the PAH coding sequence

comprises the sequence set forth in SEQ ID NO: 24. In certain embodiments, the PAH coding sequence is silently altered. In certain embodiments, the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25, 116, 131, 132, 138, 139, or 143.

[0016] In certain embodiments, the editing element comprises a PAH intron-inserted coding sequence, optionally wherein the PAH intron-inserted coding sequence comprises a nonnative intron inserted in a PAH coding sequence. In certain embodiments, the nonnative intron is selected from the group consisting of a first intron of a hemoglobin beta gene and a minute virus in mice (MVM) intron. In certain embodiments, the nonnative intron consists of a nucleotide sequence at least 90% identical to any one of SEQ ID NOs: 28-30, and 120-130. In certain embodiments, the nonnative intron consists of a nucleotide sequence set forth in any one of SEQ ID NOs: 28-30, and 120-130.

[0017] In certain embodiments, the PAH intron-inserted coding sequence encodes an amino acid sequence set forth in SEQ ID NO: 23. In certain embodiments, the PAH intron-inserted coding sequence comprises from 5' to 3': a first portion of a PAH coding sequence, the intron, and a second portion of a PAH coding sequence, wherein the first portion and the second portion, when spliced together, form a complete PAH coding sequence. In certain embodiments, the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 24. In certain embodiments, the PAH coding sequence is silently altered. In certain embodiments, the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25 or 116. In certain embodiments, the first portion of the PAH coding sequence comprises the amino acid sequence set forth in SEQ ID NO: 64 or 65, and/or the second portion of the PAH coding sequence comprises the amino acid sequence set forth in SEQ ID NO: 66 or 67. In certain embodiments, the first portion of the PAH coding sequence consist of the amino acid sequence set forth in SEQ ID NO: 64 or 65, and the second portion of the PAH coding sequence consists of the amino acid sequence set forth in SEQ ID NO: 66 or 67.

[0018] In certain embodiments, the editing element comprises from 5' to 3': a ribosomal skipping element, and the PAH coding sequence or the PAH intron-inserted coding sequence. In certain embodiments, the editing element further comprises a polyadenylation sequence 3' to the PAH coding sequence or the PAH intron-inserted coding sequence. In certain embodiments, the polyadenylation sequence is an exogenous polyadenylation sequence, optionally wherein the exogenous polyadenylation sequence is an SV40 polyadenylation sequence. In certain embodiments, the SV40 polyadenylation sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 31-34, and a sequence complementary thereto.

[0019] In certain embodiments, the nucleotide 5' to the target locus is in an exon of the PAH gene. In certain embodiments, the nucleotide 5' to the target locus is in exon 1 of the PAH gene.

[0020] In certain embodiments, the editing element further comprises a splice acceptor 5' to the ribosomal skipping element. In certain embodiments, the nucleotide 5' to the target locus is in an intron of the PAH gene. In certain embodiments, the nucleotide 5' to the target locus is in intron 1 of the PAH gene. In certain embodiments, the editing element comprises the nucleotide sequence set forth in SEQ ID NO: 35.

[0021] 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.

[0022] 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 editing window consists of the nucleotide sequence set forth in SEQ ID NO: 36 or 45. In certain embodiments, the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 36 or 45. In certain embodiments, the first editing window consists of the nucleotide sequence set forth in SEQ ID NO: 36, and the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 45.

[0023] In certain embodiments, the first genomic region consists of the nucleotide sequence set forth in SEQ ID NO: 36. In certain embodiments, the second genomic region consists of the nucleotide sequence set forth in SEQ ID NO: 45.

[0024] In certain embodiments, each of the 5' and 3' homology arm nucleotide sequences independently has a length of about 100 to about 2000 nucleotides.

[0025] In certain embodiments, the 5' homology arm comprises: C corresponding to nucleotide -2 of the PAH gene, G corresponding to nucleotide 4 of the PAH gene, G corresponding to nucleotide 6 of the PAH gene, G corresponding to nucleotide 7 of the PAH gene, G corresponding to nucleotide 9 of the PAH gene, A corresponding to nucleotide -467 of the PAH gene, A corresponding to nucleotide -465 of the PAH gene, A corresponding to nucleotide -181 of the PAH gene, G corresponding to nucleotide -214 of the PAH gene, C corresponding to nucleotide -212 of the PAH gene, A corresponding to nucleotide -211 of the PAH gene, G corresponding to nucleotide 194 of the PAH gene, C corresponding to nucleotide -433 of the PAH gene, C corresponding to nucleotide -432 of the PAH gene, ACGCTGTTCTTCGCC (SEQ ID NO: 68) corresponding to nucleotides -394 to -388 of the

PAH gene, A corresponding to nucleotide -341 of the PAH gene, A corresponding to nucleotide -339 of the PAH gene, A corresponding to nucleotide -225 of the PAH gene, A corresponding to nucleotide -211 of the PAH gene, and/or A corresponding to nucleotide -203 of the PAH gene. In certain embodiments, the 5' homology arm comprises:

(a) C corresponding to nucleotide -2 of the PAH gene, G corresponding to nucleotide 4 of the PAH gene, G corresponding to nucleotide 6 of the PAH gene, G corresponding to nucleotide 7 of the PAH gene, and G corresponding to nucleotide 9 of the PAH gene;

(b) A corresponding to nucleotide -467 of the PAH gene, and A corresponding to nucleotide -465 of the PAH gene;

(c) A corresponding to nucleotide -181 of the PAH gene;

(d) G corresponding to nucleotide -214 of the PAH gene, C corresponding to nucleotide -212 of the PAH gene, and A corresponding to nucleotide -211 of the PAH gene;

(e) G corresponding to nucleotide 194 of the PAH gene;

(f) C corresponding to nucleotide -433 of the PAH gene, and C corresponding to nucleotide -432 of the PAH gene;

(g) ACGCTGTTCTTCGCC (SEQ ID NO: 68) corresponding to nucleotides -394 to -388 of the PAH gene; and/or

(h) A corresponding to nucleotide -341 of the PAH gene, A corresponding to nucleotide -339 of the PAH gene, A corresponding to nucleotide -225 of the PAH gene, A corresponding to nucleotide -211 of the PAH gene, and A corresponding to nucleotide -203 of the PAH gene. In certain embodiments, the 5' homology arm comprises the modifications of (c) and (d), (f) and (g), and/or (b) and (h).

[0026] In certain embodiments, the 5' homology arm consists of a nucleotide sequence set forth in any one of SEQ ID NOs: 36-44, 111, 115, and 142. In certain embodiments, the 3' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 45, 112, 117, 144.

[0027] In certain embodiments, the correction genome comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 46-54, 113, 118, 134, 136, and 145.

[0028] 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 5' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO: 26, and the 3' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO: 27.

[0029] In certain embodiments, the correction genome comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 55-63, 114, 119, 135, 137, and 146. In certain embodiments, the correction genome consists of the nucleotide sequence set forth in any one of SEQ ID NOs: 55-63, 114, 119, 135, 137, and 146.

[0030] In certain embodiments, the AAV capsid comprises an AAV Clade F capsid protein.

[0031] 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, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17. In certain embodiments, 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,

(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.

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.

[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 138-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17. In certain embodiments, 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,

(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.

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.

[0033] 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, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17. In certain embodiments, 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,

(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.

[0034] 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.

[0035] In certain embodiments, the integration efficiency of the editing element into the target locus is at least 1% when the AAV is administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration 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 administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions.

[0036] In another aspect, the instant disclosure provides a pharmaceutical composition comprising an AAV disclosed herein.

[0037] 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) Cap nucleotide sequence encoding one or more AAV Clade F capsid proteins as disclosed herein; and

(c) a correction genome or transfer genome as disclosed herein, wherein the packaging system is operative in a cell for enclosing the correction genome or transfer genome in the capsid to form the AAV.

[0038] 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.

[0039] 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.

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.

[0040] 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.

[0041] 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 or the transfer genome in the capsid to form the AAV.

BRIEF DESCRIPTION OF THE DRAWINGS

[0042] **Figure 1A** is a map of the pHMI-hPAH-hAC-008 vector.

[0043] **Figure 1B** is a map of the pHMI-hPAH-h1C-007 vector.

[0044] **Figure 1C** is a map of the pHMIA-hPAH-hI1C-032.1 vector.

[0045] **Figure 2** is an image of Western blot showing the expression of human PAH from the pCOH-WT-PAH ("WT PAH"), pCOH-CO-PAH ("CO PAH pCOH"), and pHMI-CO-PAH ("CO PAH pHMI") vectors. $5x10^5$ HEK 293 cells were transfected with 1μg of

vector. Lysate of the cells was collected 48 hours after transfection. The expression of human PAH was detected by Western blotting with an anti-PAH antibody (Sigma HPA031642). The amount of GAPDH protein as detected by an anti-GAPDH antibody (Millipore MAB 374) was shown as a loading control.

[0046] **Figure 3A** is a graph showing quantitation of the PAH cDNA cassette following linear amplification ("LAM-Enriched") or PCR amplification ("Amplicon") of the editing target site.

[0047] **Figure 3B** is a graph showing quantitative analysis of integration of the pHMI-hPAH-hA-002 vector by droplet digital PCR (ddPCR).

[0048] **Figure 4A** shows the design of the pHMI-hPAH-mAC-006 vector and its expected integration into a mouse genome.

[0049] **Figure 4B** is a diagram illustrating a method for detecting by PCR an allele edited by the pHMI-hPAH-mAC-006 vector. Two pairs of primers were designed: the first pair could amplify a 867 bp DNA from an unedited allele ("Control PCR"); the second pair could specifically amplify a 2459 bp DNA from an edited allele ("Edited Allele PCR").

[0050] **Figure 4C** is an image of DNA electrophoresis showing the PCR product from the Control PCR ("Control PCR") and Edited Allele PCR ("Edit PCR") as illustrated in Figure 4A. The pHMI-hPAH-mAC-006 vector packaged in an AAVHSC capsid was injected to two wild-type neonatal mice intravenously via the tail vein at a dose of $2x10^{13}$ vector genomes per kg of body weight. Liver samples were collected after 2 weeks. A liver sample from a saline treated mouse and a cell sample of 3T3 mouse fibroblasts were used as negative control for the Edited Allele PCR.

[0051] **Figure 5A** is a diagram illustrating a method for quantifying an edited allele by ddPCR. A first pair of primers was designed to amplify a first sequence in the pHMI-hPAH-mAC-006 vector, and a first probe ("vector probe") was designed to hybridize to the first sequence. A second pair of primers was designed to amplify a second sequence on the mouse genome near the vector, and a second probe ("locus probe") was designed to hybridize to the second sequence. DNA samples were partitioned into oil droplets. The concentration of DNA was optimized to 600 pg per 20 μL in order to significantly reduce the probability that one oil droplet randomly contains a vector particle and a genomic DNA particle ($p<0.001$). Upon integration of the vector into the genome, the rate of double positivity of the vector probe and the locus probe in the same droplet increases.

[0052] **Figure 5B** is a diagram illustrating an expected result using the method described in Figure 5A. In this diagram, each dot represents a single oil droplet. The dots

with negative vector probe signal but positive locus probe signal represent the unedited alleles, whereas the dots with positive vector probe signal but positive locus probe signal represent the edited alleles.

[0053] **Figure 5C** is a graph showing the data generated from mouse liver using the method described in Figure 5A. The pHMI-hPAH-mAC-006 vector packaged in an AAVHSC capsid was injected to two wild-type neonatal mice intravenously via the tail vein at a dose of $2x10^{13}$ vector genomes per kg of body weight. Liver samples were collected after 2 weeks. One sample was analyzed using the method described in Figure 5A. Vector probe and locus probe double positive droplets were detected.

[0054] **Figure 5D** is a graph showing the data generated from a sample containing liver from a saline treated mouse and the pHMI-hPAH-mAC-006 plasmid. Few probe and locus probe double positive droplets were detected, suggesting that the sample has been sufficiently diluted so that the probability that one oil droplet randomly contains a vector particle and a genomic DNA particle is very low.

[0055] **Figure 5E** is a graph showing the quantification of the graph in Figure 5D and the graphs generated from other samples. The two control mice had 0% and 0.0395% edited alleles in the liver, respectively, and the two mice treated with the pHMI-hPAH-mAC-006 vector had 2.504% and 2.783% edited alleles in the liver, respectively.

[0056] **Figure 6** is a graph showing the mRNA expression of human PAH in the liver after administration of the pHMI-hPAH-mAC-006 vector. RNA was extracted and reverse transcribed. A pair of primers and a probe were designed to specifically detect PAH expression from the edited allele. Each PAH expression level is normalized to the expression level of endogenous Hprt.

[0057] **Figure 7A** is a graph showing the transduction efficiency of the pHMI-hPAH-mAC-006 vector packaged in AAVHSC capsids in mouse blood samples, measured by ddPCR using primer and probe sets to measure the vector and the mouse PAH genomic loci copy numbers. The numbers of vector genomes per cell ("VG per Cell") is calculated from the measured ratio of number of vectors versus the copy numbers of the genomic locus of mouse PAH.

[0058] **Figure 7B** is a graph showing the percentage editing efficiency in mouse blood samples measured by multiplexed ddPCR using primer probe sets to measure the frequency of the integrated DNA from the AAV vector ("payload") integrating into the mouse PAH locus and the human PAH locus. Editing frequency was calculated based on the detected co-partitioning of a payload and a target DNA in a single droplet in excess of

expected probability of co-partitioning of a payload and a target DNA in separate nucleic acid molecules.

[0059] **Figure 7C** is a graph showing the percentage levels of serum phenylalanine relative to the baseline in the mice after administration of the pHMI-hPAH-mAC-006 vector packaged in an AAVHSC capsid. The average levels in the treated animals and control animals (mice that did not receive AAV administration) are plotted.

[0060] **Figure 7D** is a graph showing the percentage levels of serum phenylalanine relative to the baseline in each individual mouse injected with the pHMI-hPAH-mAC-006 vector packaged in an AAVHSC capsid or in each control mice that did not receive AAV administration. The p values were calculated by ANOVA against the control distribution.

[0061] **Figure 7E** is a graph showing the correlation between the percentage levels of serum phenylalanine relative to the baseline and the percentage editing efficiency.

[0062] **Figure 7F** is a set of images showing in situ hybridization (ISH) of *Pah* mRNA and possibly virus DNA comprising PAH sequence in liver samples of mice injected with the hPAH-mAC-006 vector (middle panel), a non-integrating *Pah* transgene vector (right panel), or saline control (left panel).

[0063] **Figure 8A** is a graph showing the transduction efficiency of the hPAH-hAC-008 vector and hPAH-hAC-008-HBB vector in human and mouse hepatocytes in mice administered with the vector packaged in AAVHSC15 capsids, as measured by ddPCR using primers and probe sets specific for the vector. The y-axis represents the number of vectors measured relative to genomes of the mouse or human cells.

[0064] **Figure 8B** is a series of photos showing *in situ* hybridization of human *Pah* mRNA and possibly virus DNA comprising PAH sequence with silent codon alteration in liver samples from mice administered an unmodified or a modified hPAH-hAC-008 vector. The probe detected only the mRNA transcribed from a gene locus edited by the unmodified or modified hPAH-hAC-008 vector.

[0065] **Figure 8C** is a graph showing the percentage editing efficiency of the hPAH-hAC-008 vector in mouse and human hepatocytes from mice transplanted with human hepatocytes, as measured by multiplexed ddPCR. The left half of the figure refers to the editing efficiency of an animal treated with the hPAH-hAC-008-HBB vector, and the right half refers to that of an animal treated with the hPAH-hAC-008 vector. The p values were calculated by ANOVA.

[0066] **Figure 9A** depicts a schematic of the assay used to determine editing efficiency of the PAH gene in mice.

[0067] **Figure 9B** is a graph showing the PAH gene editing efficiency in cells from mice that have been administered either the pHMI-hPAH-mAC-006 vector or vehicle control.

[0068] **Figure 10A** is a graph showing the average percentage levels of serum phenylalanine relative to the baseline in mice after administration of either the pHMI-hPAH-mAC-006 vector packaged in AAVHSC15 capsids or a vehicle control.

[0069] **Figure 10B** is a graph showing the average percentage levels of serum tyrosine relative to the baseline in mice after administration of either the pHMI-hPAH-mAC-006 vector packaged in AAVHSC15 capsids or a vehicle control.

[0070] **Figure 10C** is a graph showing the ratio between serum phenylalanine and serum tyrosine levels in mice that received either the pHMI-hPAH-mAC-006 vector packaged in AAVHSC15 capsids or a vehicle control.

[0071] **Figure 11A** is a graph showing the average PAH gene editing efficiency and transduction efficiency in cells obtained from mice administered either the pHMI-hPAH-mAC-006 vector or a vehicle control.

[0072] **Figure 11B** depicts a graph showing the relative quantity of PAH mRNA expressed, normalized to the expression level of mouse GAPDH, in cells obtained from mice administered either the pHMI-hPAH-mAC-006 vector (AAVHSC15-mPAH) and or a vehicle control.

[0073] **Figure 12A** is a schematic showing the HuLiv humanized liver mouse model.

[0074] **Figure 12B** depicts the average PAH gene editing efficiency in cells obtained from mice 1 week and 6 weeks after administration of the pHMIK-hPAH-hI1C-032 vector packaged in AAVHSC15 capsids.

[0075] **Figure 13** is a graph showing the average PAH gene editing efficiency, as determined by ddPCR and next generation sequencing (NGS), in cells obtained from HuLiv mice administered the pHMIK-hPAH-hI1C-032 vector packaged in AAVHSC15 capsids.

[0076] **Figure 14** is a graph showing the average serum phenylalanine levels of PAH knock-out mouse model (*PAH*^{ENU2}) mice administered intravenously with either the pHMIK-hPAH-hI1C-032 vector (hPAH-032) or the pHMI-hPAH-mAC-006 vector (mPAH-006), packaged in AAVHSC15 capsids, compared to control mice.

[0077] **Figure 15A** is a graph showing the relationship between human PAH expression and serum Phe levels.

[0078] **Figure 15B** is a plot showing the expression of human PAH relative to human GAPDH in two different HuLiv mice treated with pHMIK-hPAH-hI1C-032 vector packaged in AAVHSC15 capsids.

[0079] **Figure 16** is a plot showing human PAH gene expression in HuLiv mice treated with (left) and mouse PAH gene expression in PAH^{ENU2} mice treated with pHMI-hPAH-mAC-006 vector (right) packaged in AAVHSC15 capsids.

[0080] **Figure 17A, 17B, 17C, 17D, 17E** depict vector maps of pKITR-hPAH-mAC-006-HCR, pKITR-hPAH-hI1C-032-HCR, pKITR-hPAH-mAC-006-SD.3, pHMIA2-hPAH-hI1C-032-SD.3, and pHMIA2-hPAH-mAC-006-HBB1, respectively.

DETAILED DESCRIPTION

[0081] The instant disclosure provided adeno-associated virus (AAV) compositions that can restore PAH gene function in a cell. Also provide are packaging systems for making the adeno-associated virus compositions.

I. Definitions

[0082] As used herein, the term "replication-defective adeno-associated virus" refers to an AAV comprising a genome lacking Rep and Cap genes.

[0083] As used herein, the term "PAH gene" refers to the phenylalanine hydroxylase (PAH) gene, including but not limited to the coding regions, exons, introns, 5' UTR, 3' UTR, and transcriptional regulatory regions of the PAH gene. The human PAH gene is identified by Entrez Gene ID 5053. An exemplary nucleotide sequence of a PAH mRNA is provided as SEQ ID NO: 24. An exemplary amino acid sequence of a PAH polypeptide is provided as SEQ ID NO: 23.

[0084] As used herein, the term "correcting a mutation in a PAH gene" refers to the insertion, deletion, or substitution of one or more nucleotides at a target locus in a mutant PAH gene to create a PAH gene that is capable of expressing a wild-type PAH polypeptide. In certain embodiments, "correcting a mutation in a PAH gene" involves inserting a nucleotide sequence encoding at least a portion of a wild-type PAH polypeptide or a functional equivalent thereof into the mutant PAH gene, such that a wild-type PAH polypeptide or a functional equivalent thereof is expressed from the mutant PAH gene locus (e.g., under the control of an endogenous PAH gene promoter).

[0085] 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 a PAH gene. In certain embodiments, the target locus is in the human PAH 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 PAH gene).

[0086] 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. As used herein, the term "target locus" refers to a region of a chromosome or an internucleotide bond (e.g., a region or an internucleotide bond of the human PAH gene) that is modified by an editing element.

[0087] 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 PAH gene, and the homology arm comprises a sequence substantially identical to the genome flanking the target locus.

[0088] 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.

[0089] 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.

[0090] As used herein, the term "a disease or disorder associated with a PAH gene mutation" refers to any disease or disorder caused by, exacerbated by, or genetically linked with variation of a PAH gene. In certain embodiments, the disease or disorder associated with a PAH gene mutation is phenylketonuria (PKU).

[0091] 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. Codon alteration can be conducted by any method known in the art (e.g., as described in Mauro & Chappell (2014) Trends Mol Med. 20(11):604-13, which is incorporated by reference herein in its entirety). 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. 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.

[0092] As used herein, the term "coding sequence" refers to the portion of a complementary DNA (cDNA) that encodes a polypeptide, starting at the start codon and ending at the stop codon. A gene may have one or more coding sequences due to alternative splicing and/or alternative translation initiation. A coding sequence may either be wild-type or silently altered. An exemplary wild-type PAH coding sequence is set forth in SEQ ID NO: 24.

[0093] As used herein, the term "intron-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 intron-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 intron-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 Genetics, 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.

[0094] 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. In certain embodiments, the ribosomal skipping element encodes a peptide comprising 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: 75). In certain

embodiments, the ribosomal skipping element encodes 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), or 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: 76 and 77, respectively. Exemplary nucleotide sequences of T2A element and P2A element are set forth in SEQ ID NOs: 78 and 79, 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. 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 peptide sequence encoded by the ribosomal skipping element by an intrinsic protease activity of the encoded peptide, or by another protease in the environment (e.g., cytosol).

[0095] As used herein, the term "ribosomal skipping peptide" refers to a peptide encoded by a ribosomal skipping element.

[0096] As used herein, the term "polyadenylation sequence" refers to a DNA sequence that when transcribed into RNA constitutes a polyadenylation signal sequence. The polyadenylation sequence can be native (e.g., from the PAH gene) or exogenous. The exogenous polyadenylation sequence can be a mammalian or a viral polyadenylation sequence (e.g., an SV40 polyadenylation sequence).

[0097] In the instant disclosure, nucleotide positions in a PAH gene are specified relative to the first nucleotide of the start codon. The first nucleotide of a start codon is position 1; the nucleotides 5' to the first nucleotide of the start codon have negative numbers; the nucleotides 3' to the first nucleotide of the start codon have positive numbers. As used herein, nucleotide 1 of the human PAH gene is nucleotide 5,473 of the NCBI Reference Sequence: NG_008690.1, and nucleotide -1 of the human PAH gene is nucleotide 5,472 of the NCBI Reference Sequence: NG_008690.1.

[0098] In the instant disclosure, exons and introns in a PAH gene are specified relative to the exon encompassing the first nucleotide of the start codon, which is nucleotide 5473 of the NCBI Reference Sequence: NG_008690.1. 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, the PAH gene

comprises from 5' to 3': exon 1, intron 1, exon 2, intron 2, exon 3, etc. As used herein, exon 1 of the human PAH gene is nucleotides 5001-5532 of the NCBI Reference Sequence: NG_008690.1, and intron 1 of the human PAH gene is nucleotides 5533-9704 of the NCBI Reference Sequence: NG_008690.1.

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

[0099] 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.

[00100] 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.

[00101] As used herein, the term "standard AAV administration conditions" refers to transduction of human hepatocytes implanted into a mouse following hepatocyte ablation, wherein the AAV is administered intravenously at a dose of 1×10^{13} vector genomes per kilogram of body weight, as provided by the method of Example 5, section b.

[00102] 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.

II. Adeno-Associated Virus Compositions

[00103] In one aspect, provided herein are novel replication-defective AAV compositions useful for restoring PAH expression in cells with reduced or otherwise defective PAH gene function. Such AAV compositions are highly efficient at correcting mutations in the PAH gene or restoring PAH expression, 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 AAV composition disclosed herein does not comprise an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

[00104] In certain embodiments, the AAV disclosed herein comprise: an AAV capsid; and a correction genome for editing a target locus in a PAH gene. The AAV

capsid proteins that can be used in the AAV compositions disclosed herein include without limitation AAV capsid proteins and derivatives thereof of Clade A AAVs, Clade B AAVs, Clade C AAVs, Clade D AAVs, Clade E AAVs, and Clade F AAVs. In certain embodiments, the AAV capsid protein is an AAV capsid protein or a derivative thereof of AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, or AAVrh10. In certain embodiments, the AAV capsid comprises an AAV Clade F capsid protein.

[00105] 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%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the amino acid sequence of amino acids 203-736 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 Clade F capsid protein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the amino acid sequence of amino acids 203-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17, 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.

[00106] For example, in certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the amino acid sequence of amino acids 138-736 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 Clade F capsid protein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the amino acid sequence of amino acids 138-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17, 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.

[00107] For example, in certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with 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 Clade F capsid protein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with 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, 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.

[00108] 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.

[00109] In certain embodiments, the AAV capsid comprises one or more of: (a) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 203-736 of SEQ ID NO: 8; (b) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 138-736 of SEQ ID NO: 8; and (c) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 1-736 of SEQ ID NO: 8. 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.

[00110] In certain embodiments, the AAV capsid comprises one or more of: (a) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 203-736 of SEQ ID NO: 11; (b) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 138-736 of SEQ ID NO: 11; and (c) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 1-736 of SEQ ID NO: 11. 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: 11; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 11; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 11. 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: 11; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 11; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 11. 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: 11; (b) a Clade F capsid protein having an

amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 11; and (c) a Clade F capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 11.

[00111] In certain embodiments, the AAV capsid comprises one or more of: (a) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 203-736 of SEQ ID NO: 13; (b) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 138-736 of SEQ ID NO: 13; and (c) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 1-736 of SEQ ID NO: 13. 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.

[00112] In certain embodiments, the AAV capsid comprises one or more of: (a) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 203-736 of SEQ ID NO: 16; (b) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 138-736 of SEQ ID NO: 16; and

(c) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 1-736 of SEQ ID NO: 16. 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.

[00113] Correction genomes useful in the AAV compositions disclosed herein generally comprise: (i) an editing element for editing a target locus in an PAH 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 portion of the correction genome comprising the 5' homology arm, editing element, and 3' homology arm can be in the sense or antisense orientation relative to the PAH gene 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.

[00114] Editing elements used in the correction genomes disclosed herein can mediate insertion, deletion or substitution of one or more nucleotides at the target locus.

[00115] In certain embodiments, when correctly integrated by homologous recombination at the target locus, the editing element inserts a nucleotide sequence comprising at least a portion of a PAH coding sequence into a mutant PAH gene, such that a wild-type PAH polypeptide or a functional equivalent thereof is expressed from the mutant

PAH gene locus. In certain embodiments, the editing element comprises a complete PAH coding sequence (e.g., a wild-type PAH coding sequence or a silently altered PAH coding sequence). In certain embodiments, the editing element comprises nucleotides 4 to 1359 of a PAH coding sequence. In certain embodiments, the editing element comprises a PAH intron-inserted coding sequence (e.g., comprising an intron inserted in a wild-type or silently altered PAH coding sequence).

[00116] In certain embodiments, the PAH coding sequence encodes a wild-type PAH polypeptide (e.g., having the amino acid sequence set forth in SEQ ID NO: 23). In certain embodiments, the PAH coding sequence is wild-type (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 24). In certain embodiments, the PAH coding sequence is 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 exons of the wild-type PAH gene. In certain embodiments, the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 25). In certain embodiments, the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 116).

In certain embodiments, the PAH intron-inserted coding sequence encodes a wild-type PAH polypeptide (e.g., having the amino acid sequence set forth in SEQ ID NO: 23). In certain embodiments, the PAH intron-inserted coding sequence comprises at least one (e.g., 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12) intron inserted in a PAH coding sequence. The intron can comprise a native intron sequence of the PAH gene, an intron sequence from a different species or a different gene from the same species, and/or a synthetic intron sequence. In certain embodiments, the nonnative intron is no more than 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000, 1,500, or 2,000 nucleotides in length. While not wishing to be bound by theory, it is hypothesized that introns can increase transgene expression, for example, by reducing transcriptional silencing and enhancing mRNA export from the nucleus to the cytoplasm. A skilled worker will appreciate that synthetic intron sequences can be designed to mediate RNA splicing by introducing any consensus splicing motifs known in the art (e.g., in Sibley et al., (2016) Nature Reviews Genetics, 17, 407-21, which is incorporated by reference herein in its entirety). Exemplary intron sequences are provided in Lu et al. (2013) Molecular Therapy 21(5): 954-63, and Lu et al. (2017) Hum. Gene Ther. 28(1): 125-34, which are incorporated by reference herein in their entirety. In certain embodiments, the editing element comprises a first intron of a hemoglobin beta gene in any species (e.g., human, mouse, or rabbit). In certain embodiments, the editing element comprises a first intron of a human HBB gene (e.g., comprising a nucleotide sequence at least 90%, 91%,

92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 28). In certain embodiments, the editing element comprises a first intron of a mouse HBB gene (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 29). In certain embodiments, the editing element comprises a minute virus of mouse (MVM) intron (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 30).

[00117] In certain embodiments, the editing element comprises a chimeric MVM intron (also referred to herein as ChiMVM), e.g., comprising or consisting of a nucleotide sequence of at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 120. In certain embodiments, the editing element comprises an SV40 intron, e.g., comprising or consisting of a nucleotide sequence of at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 121. In certain embodiments, the editing element comprises an adenovirus tripartite leader intron (also referred to herein as AdTPL), e.g., comprising or consisting of a nucleotide sequence of at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 122. In certain embodiments, the editing element comprises a mini β-globin intron (also referred to herein as MiniBGlobin), e.g., comprising or consisting of a nucleotide sequence of at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 123. In certain embodiments, the editing element comprises an AdV/Ig chimeric intron (also referred to herein as AdVIgG), e.g., comprising or consisting of a nucleotide sequence of at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 124. In certain embodiments, the editing element comprises a β-globin Ig heavy chain intron (also referred to herein as BglobinIg), which is a chimeric intron comprising a β-globin splice donor region and a IgG heavy chain splice acceptor region, e.g., comprising or consisting of a nucleotide sequence of at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 125. In certain embodiments, the editing element comprises a Wu MVM intron (also referred to herein as Wu MVM), which is a variant of the wild type MVM intron, e.g., comprising or consisting of a nucleotide sequence of at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 126. In certain embodiments, the editing element comprises an HCR1 element (also referred to herein as OptHCR), e.g., comprising or consisting of a nucleotide sequence of at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 127. In certain embodiments, the editing element comprises a β-globin intron (also referred to herein as Bglobin), e.g., comprising or consisting of a nucleotide sequence of at least 90%, 91%, 92%,

93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 128. In certain embodiments, the editing element comprises a Factor IX intron (also referred to herein as tFIX or FIX intron), e.g., comprising or consisting of a nucleotide sequence of at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 129. In certain embodiments, the editing element comprises a ch2BLood intron (also referred to herein as BloodEnh), e.g., comprising or consisting of a nucleotide sequence of at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 130.In certain embodiments, the PAH intron-inserted coding sequence encodes a wild-type PAH polypeptide (e.g., having the amino acid sequence set forth in SEQ ID NO: 23). In certain embodiments, the PAH intron-inserted coding sequence comprises portions of a PAH coding sequence that when spliced together, form a complete PAH coding sequence. In certain embodiments, the PAH coding sequence is wild-type (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 24). In certain embodiments, the PAH coding sequence is 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 exons of the wild-type PAH gene. In certain embodiments, the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 25). In certain embodiments, the PAH coding sequence comprises or consists of the nucleotide sequence set forth in SEQ ID NO: 116. In certain embodiments, an intron-inserted PAH coding sequence comprises a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 116. In certain embodiments, the PAH coding sequence consists of the nucleotide sequence set forth in SEQ ID NO: 116. In certain embodiments, an intron-inserted PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 80, 81, 82, 131, 132, or 143. In certain embodiments, an intron-inserted PAH coding sequence comprises a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 80, 81, 82, 131, 132, or 143. In certain embodiments, an intron-inserted PAH coding sequence consists of the nucleotide sequence set forth in SEQ ID NO: 80, 81, 82, 131, 132, or 143.

[00118] The intron can be inserted at any position in the PAH coding sequence. In certain embodiments, the intron is inserted at a position corresponding to an internucleotide bond that links two native exons. In certain embodiments, the intron is inserted at a position corresponding to an internucleotide bond that links native exon 8 and exon 9. In certain embodiments, the PAH intron-inserted coding sequence comprises from 5' to 3': a first portion of a PAH coding sequence, the intron, and a second portion of a PAH coding

sequence, wherein the first portion and the second portion, when spliced together, form a complete PAH coding sequence (e.g., wild-type PAH coding sequence, or silently altered PAH coding sequence). In certain embodiments, the first portion of the PAH coding sequence comprises the amino acid sequence set forth in SEQ ID NO: 64 or 65, and/or the second portion of the PAH coding sequence comprises the amino acid sequence set forth in SEQ ID NO: 66 or 67. In certain embodiments, the first portion of the PAH coding sequence consist of the amino acid sequence set forth in SEQ ID NO: 64 or 65, and the second portion of the PAH coding sequence consists of the amino acid sequence set forth in SEQ ID NO: 66 or 67. In certain embodiments, the first portion of the PAH coding sequence consist of the amino acid sequence set forth in SEQ ID NO: 65, and the second portion of the PAH coding sequence consists of the amino acid sequence set forth in SEQ ID NO: 67. In certain embodiments, the editing element comprises from 3' to 5': a first portion of a PAH coding sequence consist of the nucleotide sequence set forth in SEQ ID NO: 64, or a silently altered variant thereof (e.g., consisting of the nucleotide sequence set forth in SEQ ID NO: 65); an intron (e.g., consisting the nucleotide sequence set forth in SEQ ID NO: 28, 29, or 30); and a second portion of a PAH coding sequence consist of the nucleotide sequence set forth in SEQ ID NO: 66, or a silently altered variant thereof (e.g., consisting of the nucleotide sequence set forth in SEQ ID NO: 66).

[00119] In certain embodiments, the PAH coding sequence comprises a modified splice donor site. In certain embodiments, a splice donor site-modified PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 138 or 139. In certain embodiments, a splice donor site-modified PAH coding sequence comprises a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 138 or 139. In certain embodiments, a splice donor site-modified PAH coding sequence consists of the nucleotide sequence set forth in SEQ ID NO: 138 or 139.

[00120] In certain embodiments, the editing element further comprises a transcription terminator 3' to the PAH coding sequence or the PAH intron-inserted coding sequence. In certain embodiments, the transcription terminator comprises a polyadenylation sequence (e.g., an exogenous polyadenylation sequence). In certain embodiments, the exogenous polyadenylation sequence comprises an SV40 polyadenylation sequence (e.g., comprising a nucleotide sequence selected from the group consisting of SEQ ID NOs: 31-34, or a sequence complementary thereto). In certain embodiments, the SV40 polyadenylation sequence comprises the nucleotide sequence set forth in SEQ ID NO: 31. In certain embodiments, the editing element comprises from 5' to 3': a PAH coding sequence (e.g., comprising the

nucleotide sequence set forth in SEQ ID NO: 25) or a PAH intron-inserted coding sequence (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 80), and an SV40 polyadenylation sequence (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 31).

[00121] In certain embodiments, the editing element may further comprise an ID cassette 5' to an SV40 polyadenylation sequence (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 31). The ID cassette provides a sequence that can be used for identification purposes when performing next generation sequencing experiments. In certain embodiments, the ID cassette comprises the nucleotide sequence set forth in SEQ ID NO: 33. In certain embodiments, the ID cassette comprises a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 33. In certain embodiments, the ID cassette consists of the nucleotide sequence set forth in SEQ ID NO: 33. In certain embodiments, the editing element comprises from 5' to 3': a PAH coding sequence or PAH intron-inserted coding sequence, an ID cassette, and an SV40 polyadenylation sequence.

[00122] In certain embodiments, the editing element further comprises a ribosomal skipping element 5' to the PAH coding sequence or the PAH intron-inserted coding sequence. In certain embodiments, the editing element comprises from 5' to 3': a ribosomal skipping element; a PAH coding sequence or a PAH intron-inserted coding sequence; and optionally a transcription terminator (e.g., polyadenylation sequence). In certain embodiments, the aforementioned editing elements can be integrated into an exon of the PAH gene (e.g., the nucleotide 5' to the target locus is in an exon of the PAH gene) by homologous recombination to produce a recombinant sequence comprising from 5' to 3': a portion of the PAH gene 5' to the target locus; the ribosomal skipping element; the PAH coding sequence or PAH intron-inserted coding sequence; and the transcription terminator (e.g., polyadenylation sequence), wherein the ribosomal skipping element is positioned such that it is in frame with the portion of the PAH gene 5' to the target locus and the complete PAH coding sequence. Transcription and translation of this recombinant sequence produces a first polypeptide comprising the amino acid sequence encoded by the portion of the PAH gene 5' to the target locus fused to a 5' portion of the encoded ribosomal skipping peptide, and a second polypeptide comprising a 3' portion of the encoded ribosomal skipping peptide fused to the complete amino acid sequence of the PAH polypeptide.

[00123] In certain embodiments, the nucleotide 5' to the target locus is in an exon (e.g., exon 1, exon 2, exon 3, exon 4, exon 5, exon 6, exon 7, exon 8, exon 9, exon 10, exon 11,

exon 12, or exon 13) of the PAH gene. In certain embodiments, the target locus is an internucleotide bond in an exon (e.g., exon 1, exon 2, exon 3, exon 4, exon 5, exon 6, exon 7, exon 8, exon 9, exon 10, exon 11, exon 12, or exon 13) of the PAH gene. In certain embodiments, the target locus is a sequence in the PAH gene, wherein the 5' end of this sequence is in an exon (e.g., exon 1, exon 2, exon 3, exon 4, exon 5, exon 6, exon 7, exon 8, exon 9, exon 10, exon 11, exon 12, or exon 13) of the PAH gene or in the intergenic region between Achaete-scute homolog 1 (ASCL1) and PAH, and wherein the 3' end of this sequence can be any nucleotide in the PAH gene or in the intergenic region between PAH and insulin-like growth factor 1 (IGF1). In certain embodiments, the nucleotide 5' to the target locus is in exon 1, exon 2, or exon 3 of the PAH gene. In certain embodiments, the target locus is an internucleotide bond in exon 1, exon 2, or exon 3 of the PAH gene. In certain embodiments, the target locus is a sequence in the PAH gene wherein the 5' end of this sequence is in exon 1, exon 2, or exon 3 of the PAH gene, wherein the 3' end of this sequence can be any nucleotide in the PAH gene or in the intergenic region between PAH and IGF1.

[00124] In certain embodiments, the editing element comprises a splice acceptor 5' to the ribosomal skipping element. In certain embodiments, the editing element comprises from 5' to 3': a splice acceptor; a ribosomal skipping element; a PAH coding sequence or a PAH intron-inserted coding sequence; and optionally a transcription terminator (e.g., polyadenylation sequence). In certain embodiments, the aforementioned editing element can be integrated into an intron of the PAH gene (e.g., the nucleotide 5' to the target locus is in an intron of the PAH gene) by homologous recombination to produce a recombinant sequence comprising 5' to 3': a portion of the PAH gene 5' to the target locus including the endogenous splice donor site but not the endogenous splice acceptor of the intron; the splice acceptor; the ribosomal skipping element, the PAH coding sequence or PAH intron-inserted coding sequence; and the transcription terminator (e.g., polyadenylation sequence), wherein the ribosomal skipping element is positioned such that it is in frame with the PAH coding sequence or PAH intron-inserted coding sequence, and such that splicing of the splice acceptor to the endogenous splice donor of the intron of PAH places it in frame with the portion of the PAH gene 5' to the target locus. Expression of this recombinant sequence produces a first polypeptide comprising the amino acid sequence encoded by the portion of the PAH gene 5' to the target locus fused to a 5' portion of the encoded ribosomal skipping peptide, and a second polypeptide comprising the complete amino acid sequence of the PAH polypeptide fused to a 3' portion of the encoded ribosomal skipping peptide.

[00125] In certain embodiments, the nucleotide 5' to the target locus is in an intron (e.g., intron 1, intron 2, intron 3, intron 4, intron 5, intron 6, intron 7, intron 8, intron 9, intron 10, intron 11, or intron 12) of the PAH gene. In certain embodiments, the target locus is an internucleotide bond in an intron (e.g., intron 1, intron 2, intron 3, intron 4, intron 5, intron 6, intron 7, intron 8, intron 9, intron 10, intron 11, or intron 12) of the PAH gene. In certain embodiments, the target locus is a sequence in the PAH gene wherein the 5' end of this sequence is in an intron (e.g., intron 1, intron 2, intron 3, intron 4, intron 5, intron 6, intron 7, intron 8, intron 9, intron 10, intron 11, or intron 12) of the PAH gene, wherein the 3' end of this sequence can be any nucleotide in the PAH gene or in the intergenic region between PAH and IGF1. In certain embodiments, the nucleotide 5' to the target locus is in intron 1, intron 2, or intron 3 of the PAH gene. In certain embodiments, the target locus is an internucleotide bond in intron 1, intron 2, or intron 3 of the PAH gene. In certain embodiments, the target locus is a sequence in the PAH gene wherein the 5' end of this sequence is in intron 1, intron 2, or intron 3 of the PAH gene, wherein the 3' end of this sequence can be any nucleotide in the PAH gene or in the intergenic region between PAH and IGF1. In certain embodiments, the nucleotide 5' to the target locus is in intron 1 of the PAH gene. In certain embodiments, the target locus is a sequence in the PAH gene wherein the 5' end of this sequence is in intron 1 of the PAH gene, wherein the 3' end of this sequence can be any nucleotide in the PAH gene or in the intergenic region between PAH and IGF1.

[00126] Any and all of the editing elements disclosed herein can further comprise a restriction endonuclease site not present in the wild-type PAH gene. Such restriction endonuclease sites allow for identification of cells that have integration of the editing element at the target locus based upon restriction fragment length polymorphism analysis or by nucleic sequencing analysis of the target locus and its flanking regions, or a nucleic acid amplified therefrom.

[00127] Any and all of the editing elements disclosed herein can comprise one or more nucleotide alterations that cause one or more amino acid mutations in PAH polypeptide when integrated into the target locus. In certain embodiments, the mutant PAH polypeptide is a functional equivalent of the wild-type PAH polypeptide, i.e., can function as a wild-type PAH polypeptide. In certain embodiments, the functionally equivalent PAH polypeptide further comprises at least one characteristic not found in the wild-type PAH polypeptide, e.g., the ability to stabilize PAH protein (e.g., dimer or tetramer), or the ability to resist protein degradation.

[00128] 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.

[00129] In certain embodiments, an editing element as described herein comprises or consists of a PAH coding sequence or a portion thereof (e.g., the complete human PAH coding sequence, or nucleotides 4 to 1359 of the human PAH coding sequence), a 5' untranslated region (UTR), a 3' UTR, a promoter, a splice donor, a splice acceptor, a sequence encoding a non-coding RNA, an insulator, a gene, or a combination thereof.

[00130] In certain embodiments, the editing element comprises a nucleotide sequence at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to the sequence set forth in SEQ ID NO: 35, 83, or 84. In certain embodiments, the editing element comprises the nucleotide sequence set forth in SEQ ID NO: 35, 83, or 84. In certain embodiments, the editing element consists of the nucleotide sequence set forth in SEQ ID NO: 35, 83, or 84. In certain embodiments, the editing element comprises a nucleotide sequence at least 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5% identical to the sequence set forth in SEQ ID NO: 147, 148, 149, 150, 151, 152, or 153. In certain embodiments, the editing element comprises the nucleotide sequence set forth in SEQ ID NO: 147, 148, 149, 150, 151, 152, or 153. In certain embodiments, the editing element consists of the nucleotide sequence set forth in SEQ ID NO: 147, 148, 149, 150, 151, 152, or 153.

[00131] Homology arms used in the correction genomes disclosed herein can be directed to any region of the PAH 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.

[00132] 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 a PAH gene). In certain embodiments, the 5' homology arm has at least about 90% (e.g., at least about 91%, 92%, 93%, 94%, 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 91%, 92%, 93%, 94%, 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 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) identical to the first and second genomic regions flanking the target locus (e.g., a target locus in the PAH gene), respectively. In certain embodiments, the 5' and 3' homology arms are each 100% identical to the first and second genomic regions flanking the target locus (e.g., a target locus in the PAH gene), respectively. In certain embodiments, differences in nucleotide sequences of the 5' homology arm and/or the 3' homology arm and the corresponding regions the genome flanking a target locus comprise, consist essentially of or consist of non-coding differences in nucleotide sequences.

[00133] 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. For example, the homology arms can comprise one or more genetic variations in the human population, and/or one or more modifications (e.g., nucleotide substitutions, insertions, or deletions) designed to improve expression level or specificity. Human genetic variations include both inherited variations and *de novo* variations that are private to the target genome, and encompass simple nucleotide polymorphisms, insertions, deletions, rearrangements, inversions, duplications, micro-repeats, and combinations thereof. Such variations are known in the art, and can be found, for example, in the databases of dnSNP (see Sherry et al. Nucleic Acids Res. 2001; 29(1):308-11), the Database of Genomic Variants (see Nucleic Acids Res. 2014; 42(Database issue):D986-92), ClinVar (see Nucleic Acids Res. 2014; 42(Database issue): D980–D985), Genbank (see Nucleic Acids Res. 2016; 44(Database issue): D67–D72), ENCODE (genome.ucsc.edu/encode/terms.html), JASPAR (see Nucleic Acids Res. 2018; 46(D1): D260–D266), and PROMO (see Messeguer et al. Bioinformatics 2002; 18(2):333-334; Farré et al. Nucleic Acids Res. 2003; 31(13):3651-3653), each of which is incorporated herein by reference. 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.

[00134] In certain embodiments, the first genomic region 5' to the target locus is located in a first editing window, wherein the first editing window consists of the nucleotide sequence set forth in SEQ ID NO: 36. In certain embodiments, the second genomic region 3'

to the target locus is located in a second editing window, wherein the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 45. In certain embodiments, the first genomic region 5' to the target locus is located in a first editing window, wherein the first editing window consists of the nucleotide sequence set forth in SEQ ID NO: 36; and the second genomic region 3' to the target locus is located in a second PAH targeting locus, wherein the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 45.

[00135] 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 genomic region consists of a sequence shorter than the sequence of the first editing window in which the first genomic region is located. In certain embodiments, the first genomic region consists of the sequence of the first editing window in which the first genomic region is located. In certain embodiments, the second genomic region consists of a sequence shorter than the sequence of the second editing window in which the second genomic region is located. In certain embodiments, the second genomic region consists of the sequence of the second editing window in which the second genomic region is located. In certain embodiments, the first genomic region 5' to the target locus has the sequence set forth in SEQ ID NO: 36. In certain embodiments, the second genomic region 3' to the target locus has the sequence set forth in SEQ ID NO: 45. In certain embodiments, the first genomic region 5' to the target locus and the second genomic region 3' to the target locus have the sequences set forth in SEQ ID NOs: 36 and 45, respectively.

[00136] 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 region consists of a first portion of the editing window 5' to the target locus, and the second genomic region 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: 36 or 45. 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).

[00137] 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.

[00138] 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.

[00139] In certain embodiments, the 5' homology arm comprises: C corresponding to nucleotide -2 of the PAH gene, G corresponding to nucleotide 4 of the PAH gene, G corresponding to nucleotide 6 of the PAH gene, G corresponding to nucleotide 7 of the PAH gene, G corresponding to nucleotide 9 of the PAH gene, A corresponding to nucleotide -467 of the PAH gene, A corresponding to nucleotide -465 of the PAH gene, A corresponding to nucleotide -181 of the PAH gene, G corresponding to nucleotide -214 of the PAH gene, C corresponding to nucleotide -212 of the PAH gene, A corresponding to nucleotide -211 of the PAH gene, G corresponding to nucleotide 194 of the PAH gene, C corresponding to nucleotide -433 of the PAH gene, C corresponding to nucleotide -432 of the PAH gene, ACGCTGTTCTTCGCC (SEQ ID NO: 68) corresponding to nucleotides -394 to -388 of the PAH gene, A corresponding to nucleotide -341 of the PAH gene, A corresponding to nucleotide -339 of the PAH gene, A corresponding to nucleotide -225 of the PAH gene, A

corresponding to nucleotide -211 of the PAH gene, and/or A corresponding to nucleotide -203 of the PAH gene.

[00140] In certain embodiments, the 5' homology arm comprises:

(a) C corresponding to nucleotide -2 of the PAH gene, G corresponding to nucleotide 4 of the PAH gene, G corresponding to nucleotide 6 of the PAH gene, G corresponding to nucleotide 7 of the PAH gene, and G corresponding to nucleotide 9 of the PAH gene;

(b) A corresponding to nucleotide -467 of the PAH gene, and A corresponding to nucleotide -465 of the PAH gene;

(c) A corresponding to nucleotide -181 of the PAH gene;

(d) G corresponding to nucleotide -214 of the PAH gene, C corresponding to nucleotide -212 of the PAH gene, and A corresponding to nucleotide -211 of the PAH gene;

(e) G corresponding to nucleotide 194 of the PAH gene;

(f) C corresponding to nucleotide -433 of the PAH gene, and C corresponding to nucleotide -432 of the PAH gene;

(g) ACGCTGTTCTTCGCC (SEQ ID NO: 68) corresponding to nucleotides -394 to -388 of the PAH gene; and/or

(h) A corresponding to nucleotide -341 of the PAH gene, A corresponding to nucleotide -339 of the PAH gene, A corresponding to nucleotide -225 of the PAH gene, A corresponding to nucleotide -211 of the PAH gene, and A corresponding to nucleotide -203 of the PAH gene.

In certain embodiments, the 5' homology arm comprises:

(a) C corresponding to nucleotide -2 of the PAH gene, G corresponding to nucleotide 4 of the PAH gene, G corresponding to nucleotide 6 of the PAH gene, G corresponding to nucleotide 7 of the PAH gene, and G corresponding to nucleotide 9 of the PAH gene;

(b) A corresponding to nucleotide -467 of the PAH gene, and A corresponding to nucleotide -465 of the PAH gene;

(c) A corresponding to nucleotide -181 of the PAH gene;

(d) A corresponding to nucleotide -181 of the PAH gene, G corresponding to nucleotide -214 of the PAH gene, C corresponding to nucleotide -212 of the PAH gene, and A corresponding to nucleotide -211 of the PAH gene;

(e) G corresponding to nucleotide 194 of the PAH gene;

(f) C corresponding to nucleotide -433 of the PAH gene, and C corresponding to nucleotide -432 of the PAH gene;

(g) C corresponding to nucleotide -433 of the PAH gene, C corresponding to nucleotide -432 of the PAH gene, and ACGCTGTTCTTCGCC (SEQ ID NO: 68) corresponding to nucleotides -394 to -388 of the PAH gene; and/or

(h) A corresponding to nucleotide -467 of the PAH gene, A corresponding to nucleotide -465 of the PAH gene, A corresponding to nucleotide -341 of the PAH gene, A corresponding to nucleotide -339 of the PAH gene, A corresponding to nucleotide -225 of the PAH gene, A corresponding to nucleotide -211 of the PAH gene, and A corresponding to nucleotide -203 of the PAH gene.

[00141] In certain embodiments, the 5' homology arm has at least about 90% (e.g., at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) nucleotide sequence identity to the nucleotide sequence set forth in SEQ ID NO: 36, optionally comprising one or more of the nucleotides at the positions set forth above. In certain embodiments, the 5' homology arm further comprises one or more genetic variations in the human population. In certain embodiments, the 5' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 36, 37, 38, 39, 40, 41, 42, 43, or 44. In certain embodiments, the 5' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 36, 37, 38, 39, 40, 41, 42, 43, or 44.

[00142] In certain embodiments, the 3' homology arm has at least about 90% (e.g., at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) nucleotide sequence identity to the nucleotide sequence set forth in SEQ ID NO: 45. In certain embodiments, the 3' homology arm further comprises one or more genetic variations in the human population. In certain embodiments, the 3' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 45. In certain embodiments, the 3' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 45.

[00143] In certain embodiments, the 5' homology arm and the 3' homology arm each have at least about 90% (e.g., at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) nucleotide sequence identity to the nucleotide sequences set forth in SEQ ID NOs: 36 and 45, respectively, optionally wherein the 5' homology arm comprises one or more of the nucleotides at the positions set forth above. In certain embodiments, the 5' homology arm and the 3' homology arm comprise the nucleotide sequences set forth in SEQ ID NOs: 36 and 45, 37 and 45, 38 and 45, 39 and 45, 40 and 45, 41 and 45, 42 and 45, 43 and 45, or, 44 and 45, respectively. In certain embodiments, the 5' homology arm and the 3' homology arm consist of the nucleotide sequences set forth in SEQ ID NOs: 36 and 45, 37 and 45, 38 and 45, 39 and 45, 40 and 45, 41 and 45, 42 and 45, 43 and 45, or, 44 and 45, respectively.

[00144] In certain embodiments, the 5' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 69 or 72. In certain embodiments, the 5' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 69 or 72. In certain embodiments, the 3' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 70 or 73. In certain embodiments, the 3' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 70 or 73. In certain embodiments, the 5' homology arm and the 3' homology arm comprise the nucleotide sequences set forth in SEQ ID NOs: 69 and 70, or 72 and 73, respectively. In certain embodiments, the 5' homology arm and the 3' homology arm consist of the nucleotide sequences set forth in SEQ ID NOs: 69 and 70, or 72 and 73, respectively.

[00145] In certain embodiments, the 5' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 111, 115, or 142. In certain embodiments, the 5' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 111, 115, or 142. In certain embodiments, the 3' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 112, 117, or 144. In certain embodiments, the 3' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 112, 117, or 144. In certain embodiments, the 5' homology arm and the 3' homology arm comprise the nucleotide sequences set forth in SEQ ID NOs: 111 and 112, 115 and 117, or 142 and 144, respectively. In certain embodiments, the 5' homology arm and the 3' homology arm consist of the nucleotide sequences set forth in SEQ ID NOs: 111 and 112, 115 and 117, or 142 and 144, respectively.

[00146] In certain embodiments, the correction genome comprises a nucleotide sequence at least 90% (e.g., at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) identical to SEQ ID NO: 46, 47, 48, 49, 50, 51, 52, 53, 54, 85, 86, 113, 118, 134, 136, or 145. In certain embodiments, the correction genome comprises the nucleotide sequence set forth in SEQ ID NO: 46, 47, 48, 49, 50, 51, 52, 53, 54, 85, 86, 113, 118, 134, 136, or 145. In certain embodiments, the correction genome consists of the nucleotide sequence set forth in SEQ ID NO: 46, 47, 48, 49, 50, 51, 52, 53, 54, 85, 86, 113, 118, 134, 136, or 145

[00147] 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).

[00148] 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 SEQ ID NO: 35, 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: 46-54, 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: 46-54, and a 3' ITR nucleotide sequence having the sequence of SEQ ID NO:19.

[00149] 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 SEQ ID NO: 35, 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: 46-54, 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: 46-54, and a 3' ITR nucleotide sequence having the sequence of SEQ ID NO:21.

[00150] 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).

[00151] In certain embodiments, the 5' ITR or the 3' ITR is modified to reduce or abolish resolution by Rep protein ("non-resolvable ITR"). In certain embodiments, the non-resolvable ITR comprises an insertion, deletion, or substitution in the nucleotide sequence of the terminal resolution site. Such modification allows formation of a self-complementary, double-stranded DNA genome of the AAV after the transfer genome is replicated in an infected cell. Exemplary non-resolvable ITR sequences are known in the art (see e.g., those provided in U.S. Patent Nos. 7,790,154 and 9,783,824, which are incorporated by reference herein in their entirety). In certain embodiments, the 5' ITR comprises a nucleotide sequence at least 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 26. In certain embodiments, the 5' ITR consists of a nucleotide sequence at least 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 26. In certain embodiments, the 5' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 26. In certain embodiments, the 3' ITR comprises a nucleotide sequence at least 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 27. In certain embodiments, the 5' ITR consists of a nucleotide sequence at least 95%, 96%, 97%, 98%, 3or 99% identical to SEQ ID NO: 27. In certain embodiments, the 3' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 27. In certain embodiments, the 5' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 26, and the 3' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 27. In certain embodiments, the 5' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 26, and the 3' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 19.

[00152] In certain embodiments, the 3' ITR is flanked by an additional nucleotide sequence derived from a wild-type AAV2 genomic sequence. In certain embodiments, the 3' ITR is flanked by an additional 37 bp sequence derived from a wild-type AAV2 sequence that is adjacent to a wild-type AAV2 ITR. *See, e.g., Savy et al., Human Gene Therapy Methods (2017) 28(5): 277-289* (which is hereby incorporated by reference herein in its

entirety). In certain embodiments, the additional 37 bp sequence is internal to the 3' ITR. In certain embodiments, the 37 bp sequence consists of the sequence set forth in SEQ ID NO: 140. In certain embodiments, the 3' ITR comprises a nucleotide sequence at least 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 141. In certain embodiments, the 3' ITR comprises the nucleotide sequence set forth in SEQ ID NO: 141. In certain embodiments, the nucleotide sequence of the 3' ITR consists of a nucleotide sequence at least 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 141. In certain embodiments, the nucleotide sequence of the 3' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 141.

[00153] In certain embodiments, the correction genome disclosed herein has a length of about 0.5 to about 8 kb (*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).

[00154] In certain embodiments, the correction genome comprises a nucleotide sequence at least 90% (e.g., at least 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) identical to SEQ ID NO: 55, 56, 57, 58, 59, 60, 61, 62, 63, 87, 88, 114, 119, 135, 137, or 146. In certain embodiments, the correction genome comprises the nucleotide sequence set forth in SEQ ID NO: 55, 56, 57, 58, 59, 60, 61, 62, 63, 87, 88, 114, 119, 135, 137, or 146. In certain embodiments, the correction genome consists of the nucleotide sequence set forth in SEQ ID NO: 55, 56, 57, 58, 59, 60, 61, 62, 63, 87, 88, 114, 119, 135, 137, or 146.

[00155] In certain embodiments, the replication-defective AAV comprises: (a) an AAV capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 16, and a transfer genome comprising 5' to 3' the following genetic elements: a 5' ITR element (e.g., the 5' ITR of SEQ ID NOs: 18), a 5' homology arm (e.g., the 5' homology arm of SEQ ID NOs: 115), a splice acceptor (e.g., the splice acceptor of SEQ ID NOs: 14), a 2A element (e.g., the 2A element of SEQ ID NOs: 74), a silently altered human PAH coding sequence (e.g., the PAH coding sequence of SEQ ID NOs: 116), an SV40 polyadenylation sequence e.g., the SV40 polyadenylation sequence of SEQ ID NOs: 31), a 3' homology arm (e.g., the 3' homology arm of SEQ ID NOs: 117, and a 3' ITR element (e.g., the 3' ITR of SEQ ID NOs: 19); (b) an AAV capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 16, and a transfer genome comprising 5' to 3' the following genetic elements: a 5' ITR element (e.g., the 5' ITR of SEQ ID NOs: 18), a 5' homology arm (e.g., the 5' homology arm of SEQ ID NOs: 115), a splice acceptor (e.g., the splice acceptor of SEQ ID NOs: 14), a 2A element (e.g., the 2A element of SEQ ID NOs: 74), a silently altered human PAH coding sequence (e.g., the PAH coding sequence of SEQ ID NOs: 116), an SV40 polyadenylation sequence e.g., the SV40 polyadenylation sequence of SEQ ID NOs:

31), a 3' homology arm (e.g., the 3' homology arm of SEQ ID NOs: 117, and a 3' ITR element (e.g., the 3' ITR of SEQ ID NOs: 19); and/or (c) an AAV capsid protein comprising the amino acid sequence of SEQ ID NO: 16, and a transfer genome comprising 5' to 3' the following genetic elements: a 5' ITR element (e.g., the 5' ITR of SEQ ID NOs: 18), a 5' homology arm (e.g., the 5' homology arm of SEQ ID NOs: 115), a splice acceptor (e.g., the splice acceptor of SEQ ID NOs: 14), a 2A element (e.g., the 2A element of SEQ ID NOs: 74), a silently altered human PAH coding sequence (e.g., the PAH coding sequence of SEQ ID NOs: 116), an SV40 polyadenylation sequence e.g., the SV40 polyadenylation sequence of SEQ ID NOs: 31), a 3' homology arm (e.g., the 3' homology arm of SEQ ID NOs: 117, and a 3' ITR element (e.g., the 3' ITR of SEQ ID NOs: 19).

[00156] In certain embodiments, the replication-defective AAV comprises: (a) an AAV capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 16, and a correction genome comprising the nucleotide sequence set forth in any one of SEQ ID NOs: 25, 46-63, 113, 114, 116, 118, 119, 134-137, 145, and 146; (b) an AAV capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 16, and a correction genome comprising the nucleotide sequence set forth in any one of SEQ ID NOs: 25, 46-63, 113, 114, 116, 118, 119, 134-137, 145, and 146; and/or (c) an AAV capsid protein comprising the amino acid sequence of SEQ ID NO: 16, and a correction genome comprising the nucleotide sequence set forth in any one of SEQ ID NOs: 25, 46-63, 113, 114, 116, 118, 119, 134-137, 145, and 146.

[00157] The AAV compositions disclosed herein are particularly advantageous in that they are capable of correcting a PAH 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 1% (e.g. at least 2%, 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 administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions. In certain embodiments, the allelic frequency of integration of the editing element into the target locus is at least 0.5% (e.g. at least 1%, 2%, 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 administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions.

[00158] Any methods of determining the efficiency of editing of the PAH gene 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 of the PAH gene. Such 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 PAH gene.

[00159] 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 the population of transduced cells using primers and probes that only detect edited PAH alleles. Such methods can further comprise an additional qPCR or ddPCR (either in the same reaction or a separate reaction) to determine the number of total genomes in the sample and the number of unedited PAH alleles. These numbers can be used to determine the allelic frequency of integration of the editing element into the target locus.

[00160] Additionally or alternatively, in certain embodiments, the PAH locus can be amplified from DNA extracted from the population of transduced cells either by PCR using primers that bind to regions of the PAH gene flanking the target locus, or by 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 PAH 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.

[00161] 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

[00162] In another aspect, the instant disclosure provides methods for correcting a mutation in the PAH gene or expressing a PAH polypeptide 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 the PAH gene or restoring PAH expression, 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.

[00163] The methods disclosed herein can be applied to any cell harboring a mutation in the PAH gene. The skilled worker will appreciate that cells that actively express PAH are of particular interest. Accordingly, in certain embodiments, the method is applied to cells in the liver, kidney, brain, pituitary gland, adrenal gland, pancreas, urinary bladder, gallbladder, colon, small intestine, or breast. In certain embodiments, the method is applied to hepatocytes and/or renal cells.

[00164] The methods disclosed herein can be performed *in vitro* for research purposes or can be performed *ex vivo* or *in vivo* for therapeutic purposes.

[00165] 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 a PAH gene mutation, the method generally comprising administering to the subject an effective amount of a replication-defective AAV as disclosed herein. The subject can be a human subject or a rodent subject (e.g., a mouse) containing human liver cells. Suitable mouse subjects include without limitation, mice into which human liver cells (e.g., human hepatocytes) have been engrafted. Any disease or disorder associated with a PAH gene mutation can be treated using the methods disclosed herein. Suitable diseases or disorders include, without limitation,

phenylketonuria. 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.

[00166] The methods disclosed herein are particularly advantageous in that they are capable of correcting a PAH 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 1% (e.g. at least 2%, 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 administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions. In certain embodiments, the allelic frequency of integration of the editing element into the target locus is at least 0.5% (e.g. at least 1%, 2%, 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 administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions.

[00167] 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.

[00168] In certain embodiments, the foregoing methods employ a replication-defective AAV comprising: (a) an AAV capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 16, and a transfer genome comprising 5' to 3' the following genetic elements: a 5' ITR element (e.g., the 5' ITR of SEQ ID NOs: 18), a 5' homology arm (e.g., the 5' homology arm of SEQ ID NOs: 115), a splice acceptor (e.g., the splice acceptor of SEQ ID NOs: 14), a 2A element (e.g., the 2A element of SEQ ID NOs: 74), a silently altered human PAH coding sequence (e.g., the PAH coding sequence of SEQ ID NOs: 116), an SV40 polyadenylation sequence e.g., the SV40 polyadenylation sequence of SEQ ID NOs: 31), a 3' homology arm (e.g., the 3' homology arm of SEQ ID NOs: 117, and a 3' ITR element (e.g., the 3' ITR of SEQ ID NOs: 19); (b) an AAV capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 16, and a transfer genome comprising 5' to 3' the following genetic elements: a 5' ITR element (e.g., the 5' ITR of SEQ ID NOs: 18), a 5' homology arm (e.g., the 5' homology arm of SEQ ID NOs: 115), a splice acceptor (e.g., the

splice acceptor of SEQ ID NOs: 14), a 2A element (e.g., the 2A element of SEQ ID NOs: 74), a silently altered human PAH coding sequence (e.g., the PAH coding sequence of SEQ ID NOs: 116), an SV40 polyadenylation sequence e.g., the SV40 polyadenylation sequence of SEQ ID NOs: 31), a 3' homology arm (e.g., the 3' homology arm of SEQ ID NOs: 117, and a 3' ITR element (e.g., the 3' ITR of SEQ ID NOs: 19); and/or (c) an AAV capsid protein comprising the amino acid sequence of SEQ ID NO: 16, and a transfer genome comprising 5' to 3' the following genetic elements: a 5' ITR element (e.g., the 5' ITR of SEQ ID NOs: 18), a 5' homology arm (e.g., the 5' homology arm of SEQ ID NOs: 115), a splice acceptor (e.g., the splice acceptor of SEQ ID NOs: 14), a 2A element (e.g., the 2A element of SEQ ID NOs: 74), a silently altered human PAH coding sequence (e.g., the PAH coding sequence of SEQ ID NOs: 116), an SV40 polyadenylation sequence e.g., the SV40 polyadenylation sequence of SEQ ID NOs: 31), a 3' homology arm (e.g., the 3' homology arm of SEQ ID NOs: 117, and a 3' ITR element (e.g., the 3' ITR of SEQ ID NOs: 19).

[00169] In certain embodiments, the foregoing methods employ a replication-defective AAV comprising: (a) an AAV capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 16, and a correction genome comprising the nucleotide sequence set forth in any one of SEQ ID NOs: 25, 46-63, 113, 114, 116, 118, 119, 134-137, 145, and 146; (b) an AAV capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 16, and a correction genome comprising the nucleotide sequence set forth in any one of SEQ ID NOs: 25, 46-63, 113, 114, 116, 118, 119, 134-137, 145, and 146; and/or (c) an AAV capsid protein comprising the amino acid sequence of SEQ ID NO: 16, and a correction genome comprising the nucleotide sequence set forth in any one of SEQ ID NOs: 25, 46-63, 113, 114, 116, 118, 119, 134-137, 145, and 146.

[00170] An AAV composition disclosed herein can be administered to a subject by any appropriate route including, without limitation, intravenous, intraperitoneal, subcutaneous, intramuscular, intranasal, topical or intradermal routes. In certain embodiments, the composition is formulated for administration via intravenous injection or subcutaneous injection.

IV. AAV Packaging Systems

[00171] 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 the PAH gene or a transfer genome for expression of the PAH 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.

[00172] 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 or transfer 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.).

[00173] 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.

[00174] 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.

[00175] 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.

[00176] 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.

[00177] 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.

[00178] 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

[00179] The recombinant AAV vectors disclosed herein mediate highly efficient gene editing *in vitro* and *in vivo*. The following examples provide correction vectors that can be packaged with an AAV clade F capsid (e.g., AAVHSC7, AAVHSC15 or AAVHSC17, as disclosed in U.S. Patent No. 9,623,120, which is incorporated by reference herein in its entirety), and demonstrate the efficient restoration of the expression of the PAH gene which is mutated in certain human diseases, such as phenylketonuria. These examples are offered by way of illustration, and not by way of limitation.

Example 1: PAH correction vector pHMI-hPAH-hAC-008

a) PAH correction vector pHMI-hPAH-hAC-008

[00180] PAH correction vector pHMI-hPAH-hAC-008, as shown in Figure 1A, comprises 5' to 3' the following genetic elements: a 5' ITR element, a 5' homology arm, a silently altered human PAH coding sequence, an SV40 polyadenylation sequence, a targeted integration restriction cassette ("TI RE"), a 3' homology arm, and a 3' ITR element. The sequences of these elements are set forth in Table 1. The 5' homology arm comprises a wild-type genomic sequence 800 nucleotides upstream from the human PAH start codon, and thus has the ability to correct mutations in the start codon and/or 5' untranslated region (UTR) that affect PAH expression as observed in some PKU patients. The 3' homology arm comprises the wild-type genomic sequence 800 nucleotides downstream from the start codon. Integration of the PAH correction vector pHMI-hPAH-hAC-008 into the human genome inserts the silently altered human PAH coding sequence, the SV40 polyadenylation sequence, and the targeted integration restriction cassette at the PAH start codon target locus (i.e., replacing nucleotides 1-3 of the PAH gene), thereby restoring the expression of a wild-type PAH protein that has been impaired by mutations in 5' UTR, coding sequence, or 3' UTR of the PAH gene.

Table 1: Genetic elements in PAH correction vector pHMI-hPAH-hAC-008

Genetic Element	SEQ ID NO
5' ITR element	18
5' homology arm	69
silently altered human PAH coding sequence	25
SV40 polyadenylation sequence	31
targeted integration restriction cassette	71
3' homology arm	70
3' ITR element	19
Editing element	83
Correction genome from 5' homology arm to 3' homology arm	85
Correction genome from 5' ITR to 3' ITR	87

b) PAH correction vector pHMI-hPAH-h1C-007

[00181] PAH correction vector pHMI-hPAH-h1C-007, as shown in Figure 1B, comprises 5' to 3' the following genetic elements: a 5' ITR element, a 5' homology arm, a

splice acceptor, a 2A element, a silently altered human PAH coding sequence, an SV40 polyadenylation sequence, a targeted integration restriction cassette ("TI RE"), a 3' homology arm, and a 3' ITR element. The sequences of these elements are set forth in Table 2. The 5' homology arm comprises the wild-type genomic sequence of 800 nucleotides upstream from nucleotide 2128 of human PAH, which is located in intron 1. The 3' homology arm comprises the wild-type genomic sequence of 800 nucleotides downstream from nucleotide 2127 of human PAH. Integration of the PAH correction vector pHMI-hPAH-h1C-007 into the human genome allows transcription of the PAH locus into a pre-mRNA comprising 5' to 3' the following elements: exon 1 of endogenous PAH, part of intron 1 from its 5' splice donor to nucleotide 2127, the splice acceptor in the vector pHMI-hPAH-h1C-007, the 2A element, the silently altered human PAH coding sequence, and the SV40 polyadenylation sequence. Splicing of this pre-mRNA generates an mRNA comprising 5' to 3' the following elements: exon 1 of endogenous PAH, the 2A element (in frame with the PAH exon 1), the silently altered human PAH coding sequence (in frame with the 2A element), and the SV40 polyadenylation sequence. The 2A element leads to generation of two polypeptides: a truncated PAH peptide terminated at the end of exon 1 fused with an N-terminal part of the 2A peptide, and a proline from the 2A peptide fused with a full-length PAH polypeptide. Therefore, integration of the vector pHMI-hPAH-h1C-007 can restore the expression of wild-type PAH protein that has been impaired by mutations in the coding sequence or 3' UTR of the PAH gene.

Table 2: Genetic elements in PAH correction vector pHMI-hPAH-h1C-007

Genetic Element	**SEQ ID NO**
5' ITR element	18
5' homology arm	72
Splice acceptor	14
2A element	74
silently altered human PAH coding sequence	25
SV40 polyadenylation sequence	31
targeted integration restriction cassette	71
3' homology arm	73
3' ITR element	19
Editing element	84
Correction genome from 5' homology arm to 3' homology arm	86

Genetic Element	SEQ ID NO
Correction genome from 5' ITR to 3' ITR	88

[00182] The silent alteration adopted in the two vectors above significantly improved the expression of the PAH protein, as demonstrated by comparison of expression vectors pCOH-WT-PAH, pCOH-CO-PAH, and pHMI-CO-PAH. The pCOH-WT-PAH vector comprises a CBA promoter operably linked to a wild-type PAH coding sequence set forth in SEQ ID NO: 24. The pCOH-CO-PAH and pHMI-CO-PAH vectors each comprise a CBA promoter operably linked to a silently altered human PAH coding sequence as set forth in SEQ ID NO: 25. The pCOH-CO-PAH and pHMI-CO-PAH vectors were highly similar. Each vector was transfected in HEK 293 cells which is naturally deficient in PAH. As shown in Figure 2, VG-GT-CO-PAH ("CO-hPAH") gave rise to an expression level of human PAH notably higher than VG-GT-PAH ("WT-hPAH").

Example 2: PAH correction vector pHMIA-hPAH-hI1C-032.1 and its variants

[00183] In order to identify homology arm sequences that facilitate efficient gene editing, 130 correction vectors were designed, and 70 of them were tested in human hepatocellular carcinoma cells. The pHMIA-hPAH-hI1C-032.1 vector showed the highest editing efficiency *in vitro*. This example provides the structure of this vector and its variants.

a) PAH correction vector pHMIA-hPAH-hI1C-032.1

[00184] PAH correction vector pHMIA-hPAH-hI1C-032.1, as shown in Figure 1C, comprises 5' to 3' the following genetic elements: a 5' ITR element, a 5' homology arm, a splice acceptor, a P2A element, a silently altered human PAH coding sequence, an SV40 polyadenylation sequence, a 3' homology arm, and a 3' ITR element. The sequences of these elements are set forth in Table 3. The 5' homology arm comprises the wild-type genomic sequence of nucleotides -686 to 274 of human PAH, the 3' end of which is located in intron 1. The 3' homology arm comprises the wild-type genomic sequence of nucleotides 415 to 1325 of human PAH. Integration of the PAH correction vector pHMIA-hPAH-hI1C-032.1 into the human genome allows transcription of the PAH locus into a pre-mRNA comprising 5' to 3' the following elements: exon 1 of endogenous PAH, part of intron 1 from its 5' splice donor to nucleotide 274, the splice acceptor in the vector pHMIA-hPAH-hI1C-032.1, the P2A element, the silently altered human PAH coding sequence, and the SV40 polyadenylation sequence. Splicing of this pre-mRNA generates an mRNA comprising 5' to 3' the following elements: exon 1 of endogenous PAH, the P2A element (in frame with the PAH exon 1), the silently altered human PAH coding sequence (in frame with the P2A element), and the SV40

polyadenylation sequence. The P2A element leads to generation of two polypeptides: a truncated PAH peptide terminated at the end of exon 1 fused with an N-terminal part of the P2A peptide, and a proline from the P2A peptide fused with a full-length PAH polypeptide. Therefore, integration of the vector pHMIA-hPAH-hI1C-032.1 can restore the expression of wild-type PAH protein that has been impaired by mutations in the coding sequence or 3' UTR of the PAH gene.

Table 3: Genetic elements in PAH correction vector pHMIA-hPAH-hI1C-032.1

Genetic Element	SEQ ID NO
5' ITR element	18
5' homology arm	36
Splice acceptor	14
P2A element	79
silently altered human PAH coding sequence	25
SV40 polyadenylation sequence	31
3' homology arm	45
3' ITR element	19
Editing element	35
Correction genome from 5' homology arm to 3' homology arm	46
Correction genome from 5' ITR to 3' ITR	55

b) Variants of PAH correction vector pHMIA-hPAH-hI1C-032.1

[00185] Eight variants of the pHMIA-hPAH-hI1C-032.1 vector have been designed to improve the expression of the PAH gene locus. These variants, named pHMIA-hPAH-hI1C-032.2 to pHMIA-hPAH-hI1C-032.9, differ from pHMIA-hPAH-hI1C-032.1 only in the 5' homology arm. The sequences of the different elements are set forth in Table 4.

Table 4: Variants of the pHMIA-hPAH-hI1C-032.1 vector

Vector name	SEQ ID NO		
	5' homology arm (HA)	Correction genome from 5' HA to 3' HA	Correction genome from 5' ITR to 3' ITR
pHMIA-hPAH-hI1C-032.2	37	47	56
pHMIA-hPAH-hI1C-032.3	38	48	57
pHMIA-hPAH-hI1C-032.4	39	49	58

Vector name	SEQ ID NO		
	5' homology arm (HA)	Correction genome from 5' HA to 3' HA	Correction genome from 5' ITR to 3' ITR
pHMIA-hPAH-hI1C-032.5	40	50	59
pHMIA-hPAH-hI1C-032.6	41	51	60
pHMIA-hPAH-hI1C-032.7	42	52	61
pHMIA-hPAH-hI1C-032.8	43	53	62
pHMIA-hPAH-hI1C-032.9	44	54	63

[00186] The pHMIA-hPAH-hI1C-032.2 vector was designed to optimize the Kozak sequence for improved ribosome recruitment to the transcript. It differs from pHMIA-hPAH-hI1C-032.1 in having the nucleotides C, G, G, G, and G at positions -2, 4, 6, 7, and 9, respectively, of the PAH gene.

[00187] The pHMIA-hPAH-hI1C-032.3 vector was designed to remove a single quadruplex in 5' UTR of the PAH gene that might suppress expression. It differs from pHMIA-hPAH-hI1C-032.1 in having the nucleotides A and A at positions -467 and -465, respectively, of the PAH gene.

[00188] The pHMIA-hPAH-hI1C-032.4 vector was designed to optimize a cyclic AMP response element to increase expression. It differs from pHMIA-hPAH-hI1C-032.1 in having the nucleotide A at position -181 of the PAH gene.

[00189] The pHMIA-hPAH-hI1C-032.5 vector was designed to optimize two cyclic AMP response elements to increase expression. It differs from pHMIA-hPAH-hI1C-032.1 in having the nucleotides G, C, A, and A at positions -214, -212, -211, and -181, respectively, of the PAH gene.

[00190] The pHMIA-hPAH-hI1C-032.6 vector was designed to incorporate the minor allele of SNP rs1522295, which correlates with altered PAH expression in humans. It differs from pHMIA-hPAH-hI1C-032.1 in having the nucleotide G at position 194 of the PAH gene.

[00191] The pHMIA-hPAH-hI1C-032.7 vector was designed to optimize a glucocorticoid binding site in the 5' UTR to increase expression. It differs from pHMIA-hPAH-hI1C-032.1 in having the nucleotides C and C at positions -433 and -432, respectively, of the PAH gene.

[00192] The pHMIA-hPAH-hI1C-032.8 vector was designed to modify two glucocorticoid binding sites and a single AP2 binding site for improved expression. It differs

from pHMIA-hPAH-hI1C-032.1 in having the nucleotides C and C at positions -433 and -432, respectively, of the PAH gene, and having the nucleotide sequence ACGCTGTTCTTCGCC (SEQ ID NO: 68) at positions -394 to -388 of the PAH gene.

[00193] The pHMIA-hPAH-hI1C-032.9 vector was designed to disrupt three G-quadruplexes in the 5' UTR that might suppress expression. It differs from pHMIA-hPAH-hI1C-032.1 in having the nucleotide A at each of the nucleotide positions -467, -465, -341, -339, -225, -211, and -203 of the PAH gene.

Example 3: *in vitro* human PAH gene editing

[00194] This example provides an *in vitro* method for examining PAH correction vectors, such as those described in the previous examples.

[00195] PAH correction vector pHMI-hPAH-hA-002, a variant of pHMI-hPAH-hAC-008 wherein the PAH coding sequence is wild-type (i.e., not silently altered), and PAH correction vector pHMI-hPAH-h1-001, a variant of pHMI-hPAH-h1C-007 wherein the PAH coding sequence is wild-type (i.e., not silently altered), were examined for assessment of targeted integration. K562 cells were transduced with the pHMI-hPAH-hA-002 vector packaged in AAVHSC17 at an MOI of 150,000. The genomic DNA of the cells was collected after 48 hours. Single biotinylated primers with the sequences ccaaatcccaccagctcact (SEQ ID NO: 89) and tcccatgaaactgaggtgtga (SEQ ID NO: 90), each located outside the homology arms, were separately used to amplify the DNA samples by linear amplification. Both the edited and unedited alleles were amplified without bias. The amplified DNA samples were pooled and enriched by streptavidin pulldown. The number of alleles with pHMI-hPAH-hA-002 integration was measured by ddPCR using the PAH_Genomic Set 1 primer/probe set.

[00196] As shown in Figure 3A, left panel ("LAM-Enriched"), the desired integration was detected in a sample from cells transduced with the pHMI-hPAH-hA-002 vector ("R1 ATG"), but not detected in samples from cells transduced with the pHMI-hPAH-h1-001 vector ("R1 Intron") or untransduced cells ("R1 WT"). In the right panel of Figure 3A ("Amplicon"), the amount of vector integration was measured by ddPCR using the SV40_FAM Set 1 primer/probe set. Positive signals were detected in samples from both the cells transduced with the pHMI-hPAH-hA-002 vector ("T001 Frag") and the cells transduced with the pHMI-hPAH-h1-001 vector ("T002 Frag"), indicating that both cells underwent vector integration.

[00197] To quantify the targeted integration, three sets of primers and probes, as shown in Table 6, were designed for detection the integration by ddPCR. PAH_Genomic Set

1 detected the unedited genome and the edited genome after the targeted integration of pHMI-hPAH-hA-002. SV40_FAM Set 1 detected a sequence in the SV40 polyadenylation sequence, which was present in the edited genome and the unintegrated vectors. PAH_HA Set 1 detected a region in the homology arm, which was present in both edited and unedited genomes, as well as in the unintegrated vectors.

[00198] DNA samples were partitioned into oil droplets. The concentration of DNA was optimized to a concentration of 600 pg per 20 μL in order to significantly reduce the probability that one oil droplet randomly contains two DNA molecules (e.g., a vector particle and a genomic DNA particle) ($p<0.001$). The quantity of DNA identified by PAH_Genomic Set 1 (Quantity_genome) represented the total amount of unedited and edited genomes. The quantity of DNA identified by SV40_FAM Set 1 (Quantity_payload) represented the total amount of edited genomes and unintegrated vectors. The quantity of DNA identified by PAH_HA Set 1 (Quantity_HA) represented the total amount of unedited genomes, edited genomes, and unintegrated vectors. Thus, the quantity of edited genome can be calculated by the follow formula: Quantity_genome + Quantity_payload – Quantity_HA. The fraction of genome having the correct integration can be calculated as the quantity of edited genome divided by Quantity_genome.

Table 5: Primers and probes for quantifying integration of human PAH into the human genome

Primer or Probe	Sequence	SEQ ID NO
PAH_Genomic Set 1, primer F	GCTCCATCCTGCACATAGTT	91
PAH_Genomic Set 1, primer R	CCTATGCTTTCCTGATGAGATCC	92
PAH_Genomic Set 1, probe	TTGGTGCTGCTGGCAATACGGTC	93
SV40_FAM Set 1, primer F	GCAATAGCATCACAAATTTCAC	94
SV40_FAM Set 1, primer R	GATCCAGACATGATAAGATACATTG	95
SV40_FAM Set 1, probe	TCACTGCATTCTAGTTGTGGTTTGTCCA	96
PAH_HA Set 1, primer F	TCCAGTCACCAGACAGTTAGT	97
PAH_HA Set 1, primer R	GGAGAGAAATGGAGCAAGTGAA	98
PAH_HA Set 1, probe	ACAGCCTATATTTCACCATGCTGATCCC	99

[00199] As shown in Figure 3B, the percentage of genome having the correct integration of the pHMI-hPAH-hA-002 vector, as measured by the above primer/probe sets, was 17.86%. No integration was detected in the control cells which were not transduced with the pHMI-hPAH-hA-002 vector.

Example 4: *in vivo* PAH gene editing in mouse liver

[00200] This example provides animal models for examining PAH correction vectors that are capable of editing mouse PAH gene, and determining their editing efficiency in mouse liver.

a) Editing of the mouse PAH gene in wild-type mice

[00201] In a specific example, provided herein is *in vivo* editing of the mouse genome using the pHMI-hPAH-mAC-006 vector. The pHMI-hPAH-mAC-006 vector was similar to the pHMI-hPAH-hAC-008 vector, but was capable of editing the mouse PAH gene rather than the human PAH gene (Figure 4A). Specifically, pHMI-hPAH-mAC-006 comprised 5' to 3' the following genetic elements: a 5' ITR element, a 5' homology arm, a silently altered human PAH coding sequence, an SV40 polyadenylation sequence, a targeted integration restriction cassette ("TI RE"), a 3' homology arm, and a 3' ITR element. The sequences of these elements are set forth in Table 6. The 5' homology arm comprised the wild-type genomic sequence upstream from and including the mouse PAH start codon, and thus had the ability to correct mutations in the start codon and/or 5' untranslated region (UTR) of the mouse PAH gene. The 3' homology arm comprised the wild-type genomic sequence downstream from the start codon of mouse PAH. Integration of the PAH correction vector pHMI-hPAH-mAC-006 into the mouse genome could insert the silently altered human PAH coding sequence, the SV40 polyadenylation sequence, and the targeted integration restriction cassette at the start codon of the mouse PAH gene (i.e., replacing nucleotides 1-3 of the mouse PAH gene), thereby expressing a wild-type human PAH protein in a mouse cell. The vector alone did not include a promoter sequence, and could not drive independent PAH expression without genomic integration.

Table 6: Genetic elements in PAH correction vector pHMI-hPAH-mAC-006

Genetic Element	SEQ ID NO
5' ITR element	18
5' homology arm	100
Silently altered human PAH coding sequence	25
SV40 polyadenylation sequence	31

Genetic Element	SEQ ID NO
targeted integration restriction cassette	71
3' homology arm	101
3' ITR element	19

[00202] The pHMI-hPAH-mAC-006 vector was packaged in AAVHSC17 capsid and injected to two wild-type neonatal mice intravenously via the tail vein at a dose of $2x10^{13}$ vector genomes per kg of body weight. Two control mice received saline injection via the tail vein. Liver samples were collected after 2 weeks.

[00203] A PCR method was developed to detect the integration of the pHMI-hPAH-mAC-006 vector into the mouse genome. As shown in Figure 4B, a first pair of primers (SEQ ID NOs: 62 and 63) were designed to amplify an 867 bp DNA from an unedited allele ("Control PCR"); a second pair of primers (SEQ ID NOs: 64 and 65) were designed to specifically amplify a 2459 bp DNA from an edited allele ("Edited Allele PCR"). As shown in Figure 4C, a liver sample from a saline treated mouse and a cell sample of 3T3 mouse fibroblasts did not generate the PCR product corresponding to the edited allele, whereas the liver samples from the two mice injected with the pHMI-hPAH-mAC-006 vector generated the PCR product corresponding to the edited allele. All four samples generated similar levels of the PCR product corresponding to the unedited allele, suggesting that the samples were comparable in quality.

[00204] A ddPCR method was developed to quantify the integration of the pHMI-hPAH-mAC-006 vector into the mouse genome. Two sets of primers and probes, as shown in Table 7, were designed for detection the integration by ddPCR. mPAH_ATG_gDNA_FAM Set 1 detected the unedited genome and the edited genome after the targeted integration of pHMI-hPAH-mAC-006. SV40_FAM Set 1 detected a sequence in the SV40 polyadenylation sequence, which was present in the edited genome and the unintegrated vectors (Figure 5A).

[00205] DNA samples were partitioned into oil droplets. The concentration of DNA was optimized to 600 pg per 20 μL in order to significantly reduce the probability that one oil droplet randomly contains a vector particle and a genomic DNA particle ($p<0.001$) (Figure 5C and 5D). Upon integration of the vector into the genome, the rate of double positivity of the vector probe and the locus probe in the same droplet increases (Figure 5B). As shown in Figure 5E, the two control mice had 0% and 0.0395% edited alleles in the liver, respectively, and the two mice treated with the pHMI-hPAH-mAC-006 vector had 2.504% and 2.783%

edited alleles in the liver, respectively. Thus, the overall integration efficiency of the pHMI-hPAH-mAC-006 vector in the liver under the given conditions was about 2.6%. The integration efficiency for each individual cell is expected to be higher, because not all cells were transduced with the vector.

Table 7: Primers and probes for quantifying integration of human PAH into the mouse genome

Primer or Probe	Sequence	SEQ ID NO
mPAH_ATG_gDNA_FAM Set 1, primer F	CAGCATCAGAAGCAGAACATTT	102
mPAH_ATG_gDNA_FAM Set 1, primer R	AAAGCACATCAGCAGTTTCAA	103
mPAH_ATG_gDNA_FAM Set 1, probe	AGATGAAAGCAACTGAACATCGACTACGA	104
SV40_FAM Set 1, primer F	GCAATAGCATCACAAATTTCAC	105
SV40_FAM Set 1, primer R	GATCCAGACATGATAAGATACATTG	106
SV40_FAM Set 1, probe	TCACTGCATTCTAGTTGTGGTTTGTCCA	107
mPah_1C_LHA_FAM Set 3, primer F	gcaagctccagatcaccaata	108
mPah_1C_LHA_FAM Set 3, primer R	ctgagcaatgcattcagcaataa	109
mPah_1C_LHA_FAM Set 3, probe	CCCTGAACATCCCTTGACAGAGCA	110

[00206] The relative quantity of the mRNA expressed from the edited allele was determined by ddPCR. SV40_FAM Set 1 was used to specifically detect human PAH expression from the edited allele. Each PAH expression level was normalized to the expression level of endogenous Hprt. As shown in Figure 6, control mice showed no expression of human PAH, suggesting that the primers and probe did not cross react with the endogenous mouse PAH. The percent PAH expression relative to wild-type levels was calculated based on the human PAH signal relative to Hprt normalized against the endogenous mouse PAH signal relative to Hprt. The two mice treated with the pHMI-hPAH-mAC-006 vector had 5.378% and 4.846% mRNA levels relative to the endogenous mouse

PAH levels, respectively. Thus, the overall mRNA level of the pHMI-hPAH-mAC-006 vector in the liver under the given conditions was about 5%. The mRNA level for each individual cell is expected to be higher, because not all cells were transduced with the vector.

b) Editing of the mouse PAH gene in pah knockout mice

[00207] In one experiment, the efficacy of the pHMI-hPAH-mAC-006 vector in phenotypic correction was determined using a PAH knock-out mouse model (PAH^{ENU2}). Briefly, the hPAH-mAC-006 vector packaged in AAVHSC15 capsids was administered intravenously, in 5 consecutive days, to these mice at a dose of 1.16×10^{14} vector genomes per kilogram of body weight. Serum phenylalanine (Phe) was measured weekly for 5 months by mass spectrometry. After 5 months, DNA was extracted from liver samples, and the numbers of vector genomes per cell were analyzed by ddPCR using primer and probe sets to measure the vector and the human PAH genomic locus copy numbers.

[00208] Transduction efficiency (measured in number of vector genomes per cell ("VG per Cell")) was the determined by ddPCR using primer and probe sets to measure the vector, and the mouse and human PAH genomic loci copy numbers. Editing frequency was measured by multiplexed ddPCR using primer probe sets to measure the frequency of the editing element DNA from the AAV vector ("payload") integrated into the mouse PAH locus and the human PAH locus. Briefly, single DNA strands were partitioned into oil droplets. Each droplet was tested for the presence of either human or mouse PAH DNA along with the presence or absence of the payload. Editing frequency was calculated based on the detected co-partitioning of a payload and a target DNA in a single droplet in excess of expected probability of co-partitioning of a payload and a target DNA in separate nucleic acid molecules.

[00209] The PAH knock-out mice had a phenotype of increased phenylalanine (Phe) levels in the blood. To examine phenotypic changes, the serum levels of Phe after administration of the AAV vectors were measured, the percentage levels were calculated relative to the baseline at time zero, and the percentage levels were compared to the control mice that did not receive the AAV vectors.

[00210] The mice administered the hPAH-mAC-006 vector packaged in AAVHSC15 capsids showed a transduction efficiency of about 8 to 18 vector genomes per cell (Figure 7A), and an average editing efficiency of about 4.4% relative to the number of alleles (Figure 7B). This editing efficiency supported an expression level of PAH sufficient to reduce Phe levels in the serum of the mice by about 50% for at least 5 months (Figures 7C and 7D), and the phenotypic changes correlated with the editing efficiency (Figure 7E). The correct

homologous recombination of the vector at the *Pah* locus was verified by the length of the PCR product amplified from the edited genomic locus using a first primer that hybridized to the payload, and a second primer that hybridized to a genomic sequence downstream from the right homology arm (data not shown).

[00211] To determine whether the homologous recombination introduced any genomic alterations into the edited alleles, the DNA sequences in the genomic regions corresponding to the homology arms were further analyzed by deep sequencing (Illumina). The samples all had high quality sequence reads, and all the positions were sequenced with a depth of over 20,000 reads. Insertions and deletions (hereinafter "indels") were identified by Somatic Variant Callers with an indel quality filter and a strand bias filter. Specifically, a region in the right homology arm comprising 10 continuous G showed an elevated indel rate of about 0.02-0.05% in both control and treated animals. Indels at this locus, as well as several other loci, did not pass filters for bona fide changes, and were removed from further analysis. As shown in Table 8, the untreated control animals showed an indel rate of 0.002-0.006%. Treated animal 1 had an indel rate of 0.031%; treated animal 2 had no indels that passed the filters; treated animal 3 had an indel rate similar to those of the control animals. All the indels identified were located in untranslated regions.

Table 8. Deep sequencing data for individual animals

Animal	Total reads	Average depth per base	Number of mutations passing filter	Accumulative mutations passing filter in %
Control animal 1	4,218,356	341,291	1	0.002%
Control animal 2	5,599,928	453,069	2	0.006%
Treated animal 1	4,785,826	387,203	9	0.031%
Treated animal 2	3,353,288	271,302	0	0.000%
Treated animal 3	9,514,938	769,817	9	0.006%

[00212] The results above demonstrated the feasibility of reversing the phenotypes of PAH deficiency using correction vectors that insert a PAH coding sequence in a genome.

[00213] To detect expression of human PAH in individual mouse hepatocytes after the *in vivo* transduction, RNA in situ hybridization (ISH) was performed on liver tissue sections using a probe specific to >1kb of the human PAH RNA having the silent codon alteration as

described above (Advanced Cell Diagnostics, Inc., Hayward, CA). As shown in Figure 7F, this probe detected human PAH RNA and possibly virus DNA comprising PAH sequence in mouse hepatocytes transduced with the hPAH-mAC-006 vector, but did not cross-hybridize to endogenous mouse *Pah* RNA. A liver sample of a mouse transduced with a transgene construct comprising a CMV promoter driving the expression of a human *Pah* RNA having the same silent codon alteration was used as a positive control.

c) PAH correction vector pHMI-hPAH-mAC-006

[00214] The pHMI-hPAH-mAC-006 vector comprised 5' to 3' the following genetic elements: a 5' ITR element, a 5' homology arm, a silently altered human PAH coding sequence, an SV40 polyadenylation sequence, a 3' homology arm, and a 3' ITR element. The sequences of these elements are set forth in Table 9.

Table 9: Genetic elements in PAH correction vector pHMI-hPAH-mAC-006

Genetic Element	SEQ ID NO
5' ITR element	18
5' homology arm	111
Silently altered human PAH coding sequence	25
SV40 polyadenylation sequence	31
Targeted integration restriction cassette	71
3' homology arm	112
3' ITR element	19
correction genome (5' HA to 3' HA)	113
correction genome (5' ITR to 3' ITR)	114

d) PAH gene editing efficiency in mice administered pHMI-hPAH-mAC-006

[00215] Figure 9A depicts a schematic of the assay used to determine editing efficiency of the PAH gene in mice that received the pHMI-hPAH-mAC-006 vector. ddPCR and LAM-NGS (LAM-PCR followed by next generation sequencing (NGS)) was performed as described herein and as indicated in Figure 9A. Figure 9B shows a graph of PAH gene

editing efficiency as determined in cells of mice administered either the pHMI-hPAH-mAC-006 vector or vehicle control. As shown in Figure 9B, PAH gene editing efficiency in mice administered the HMI-hPAH-mAC-006 vector was determined to be about 8% relative to the number of alleles. No errors were detected in the edited regions.

e) Durable phenotypic correction of hyperphenylalaninemia in mouse models

[00216] In one experiment, the efficacy of the pHMI-hPAH-mAC-006 vector in phenotypic correction was determined using a PAH knock-out mouse model (PAH^{ENU2}). The pHMI-hPAH-mAC-006 vector was packaged in AAVHSC15 capsids and administered intravenously to mice at a dose of 1×10^{14} vector genomes per kilogram of body weight. To examine phenotypic changes, the serum levels of phenylalanine (Phe) and tyrosine (Tyr) after administration of the pHMI-hPAH-mAC-006 vector packaged in AAVHSC15 capsids was measured weekly beyond 7 weeks, the percentage levels were calculated relative to the baseline at time zero, and the percentage levels were compared to the control mice that received a vehicle control. A total of 4 mice were administered the pHMI-hPAH-mAC-006 vector packaged in AAVHSC15 capsids, and 2 mice were administered vehicle control. As shown in Figure 10, a significant reduction in serum levels of Phe (Figure 10A; * indicates $p<0.0001$ by repeated measures 2-way ANOVA vs vehicle; $p<0.0001$ by repeated measures 2-way ANOVA vs time) and a significant increase in serum levels of Tyr (Figure 10B; * indicates $p<0.05$ by repeated measures 2-way ANOVA vs vehicle; $p<0.0003$ by repeated measures 2-way ANOVA vs time) were observed in mice that received the vector. Figure 10C shows the ratio between serum Phe and serum Tyr in mice that received the vector or a vehicle control (* indicates $p< 0.002$ by repeated measures 2-way ANOVA vs vehicle; $p<0.0004$ by repeated measures 2-way ANOVA vs time).

[00217] Figure 11A depicts a graph showing the PAH gene editing efficiency and transduction efficiency of cells obtained from mice administered either the pHMI-hPAH-mAC-006 vector or a vehicle control. The left y-axis of Figure 11A indicates the percentage of editing efficiency and shows that mice administered the pHMI-hPAH-mAC-006 vector (AAVHSC15-mPAH) had about 5% editing efficiency relative to the number of alleles. The right y-axis of Figure 11A indicates the number of vector genomes per cell and shows that mice administered the pHMI-hPAH-mAC-006 vector (AAVHSC15-mPAH) had a transduction efficiency of about 140 vector genomes per cell.

[00218] Figure 11B depicts a graph showing the relative quantity of PAH mRNA expressed, normalized to the expression level of mouse GAPDH, of cells obtained from mice administered either the pHMI-hPAH-mAC-006 vector (AAVHSC15-mPAH) or a vehicle

control. As shown, cells obtained from mice administered the pHMI-hPAH-mAC-006 vector (AAVHSC15-mPAH) had significant levels of human PAH mRNA, as compared mice administered a vehicle control (* indicates $p<0.005$ by two-tailed Mann Whitney test vs vehicle).

Example 5: *in vivo* editing of the human PAH gene in a mouse model

[00219] This example provides animal models for examining PAH correction vectors, such as those described in the previous examples, in the editing of the human PAH gene in a mouse model.

a) Editing of human PAH in human blood cells in a mouse model

[00220] Briefly, NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ (NSG) mice were myeloblated through sublethal irradiation, and transplanted with human CD34$^+$ hematopoietic stem cells. Engraftment levels were determined after 12 weeks by identifying the amounts of human and murine CD45$^+$ cells in the peripheral blood by flow cytometry, and the mice having more than 50% of circulating human CD45$^+$ cells were selected. The hPAH-hAC-008 vector packaged with the AAVHSC17 capsid was administered intravenously to 12 such mice divided equally into four groups. The first and second groups of mice received a dose of 1.54×10^{13} vector genomes per kilogram of body weight, and the third and fourth groups received a dose of 2.1×10^{12} vector genomes per kilogram of body weight. The mice were euthanized 6 weeks after the injections. Samples of blood, bone marrow and spleen tissues were collected, and genomic DNA was extracted.

[00221] Editing frequency in mouse and human cells were measured by multiplexed droplet digital PCR (ddPCR) using primer probe sets to measure the frequency of the integrated DNA from the AAV vector ("payload") integrating into the mouse PAH locus and the human PAH locus. In short, single DNA strands were partitioned into oil droplets. Each droplet was tested for the presence of either human or mouse PAH DNA along with the presence or absence of the payload. Editing frequency was calculated based on the detected co-partitioning of a payload and a target DNA in a single droplet in excess of expected probability of co-partitioning of a payload and a target DNA in separate nucleic acid molecules.

[00222] As shown in Table 10, editing of human cells was detected in bone marrow samples in a dose-dependent manner. Notably, editing was specific to human genome, as no editing was detected in mouse cells.

Table 10: Editing efficiencies of hPAH-hAC-008 in mouse tissues

Group	% Editing in bone marrow	% Editing in spleen	% Editing in blood
1	0.16	0.0	0.0
2	0.25	0.01	0.0
3	0.09	0.09	0.0
4	0.02	0.013	0.001

[00223] Figure 8A shows the transduction efficiency of the hPAH-hAC-008 vector and hPAH-hAC-008-HBB vector in human and mouse hepatocytes in mice administered with the vector packaged in AAVHSC15 capsids.

b) Editing of human PAH in human hepatocytes in a mouse model using a vector comprising an HBB intron

[00224] The hPAH-hAC-008 vector comprises a complete human PAH coding sequence without any intron. A modified vector hPAH-hAC-008-HBB, wherein the first intron of the human HBB gene (having the nucleotide sequence of SEQ ID NO: 28) is added between nucleotides 912 and 913 of the human PAH coding sequence, was generated for improving the nuclear export and stability of RNA molecules transcribed from the vector. The internucleotide bond between nucleotides 912 and 913 corresponds to the splicing site between exon 8 and exon 9 of the native PAH gene, which was not disrupted by the silent alteration of the codons.

[00225] The vectors were packages with AAVHSC15 capsids, and were administered into mice intravenously at a dose of 1×10^{13} vector genomes per kilogram of body weight. Six weeks after the administration, liver samples were collected, and the localization of the silently altered human PAH mRNA and possibly virus DNA comprising PAH sequence was examined by *in situ* hybridization. As shown in Figure 8B, the addition of the HBB intron substantially improved the nuclear export of the mRNA. This result demonstrated that addition of an intron in the PAH coding sequence could potentially increase the expression level of the PAH gene, and this feature can be included in the design of PAH correction vectors.

c) Editing of human PAH in human hepatocytes in a mouse model

[00226] Briefly, *Fah*$^{-/-}$ *Rag2*$^{-/-}$ *Il2rg*$^{-/-}$ mice on the C57Bl/6 background, commonly referred to as the FRG® Knockout mice, were used as a model for liver humanization. The mice were immunodeficient and lacked the tyrosine catabolic enzyme fumarylacetoacetate

hydrolase (Fah). Ablation of mouse hepatocytes was induced by the withdrawal of the protective drug 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC). The mice were then engrafted with human hepatocytes, and a urokinase-expressing adenovirus was administered to enhance repopulation of the human hepatocytes. Engraftment was sustained over the life of the animal with an appropriate regimen of CuRx™ Nitisinone (20-0026) and prophylactic treatment of SMX/TMP antibiotics (20-0037). The animals weighed 22 grams on average and had a typical lifespan of 18–24 months.

[00227] The hPAH-hAC-008 or hPAH-hAC-008-HBB vector was packaged with AAVHSC15 capsids, and was administered into mice intravenously at a dose of 1×10^{13} vector genomes per kilogram of body weight. Six weeks after the administration, liver samples were collected, the human and mouse hepatocytes were separated and purified using Miltenyi autoMACS columns following liver perfusion. DNA was extracted, and the efficiency of gene editing was measured using the same ddPCR method as described above.

[00228] As shown in Figure 8C, the percentage editing efficiency in human hepatocytes, measured as the percentage of edited alleles out of all alleles, was 2.2% in an animal treated with the hPAH-hAC-008 vector, and 4.3% in an animal treated with the hPAH-hAC-008-HBB vector. Editing was not detected in mouse hepatocytes from either animal. The lack of detection of editing in mouse hepatocytes from either animal is unlikely to be due to lack of transduction efficiency as mouse hepatocytes were transduced well (Figure 8A). In a separate experiment, editing of the human genome by the hPAH-hAC-008 vector was detected at a rate of 2.131% relative to the number of alleles of human genome, whereas editing of the mouse genome in the liver sample was detected at a rate of 0.05% relative to the number of alleles of mouse genome. These results showed human-specific editing of the PAH gene by the hPAH-hAC-008 vector or a modified version thereof, and provided an *in vivo* model for examining the editing efficiency in hepatocytes.

d) Editing of human PAH in human hepatocytes in a mouse model

[00229] In one experiment, *Fah*$^{-/-}$ *Rag2*$^{-/-}$ *Il2rg*$^{-/-}$ mice on the C57Bl/6 background, commonly referred to as the FRG® Knockout mice (also referred to herein as HuLiv mice), were used as a model for liver humanization, as described above (see Figure 12A).

[00230] The pHMIK-hPAH-hI1C-032 vector comprised 5' to 3' the following genetic elements: a 5' ITR element, a 5' homology arm, a splicing acceptor, a 2A element, a silently altered human PAH coding sequence, an SV40 polyadenylation sequence, a 3' homology arm, and a 3' ITR element. The sequences of these elements are set forth in Table 11.

Table 11: Genetic elements in PAH correction vector pHMIK-hPAH-hI1C-032

Genetic Element	SEQ ID NO
5' ITR element	18
5' homology arm	115
Splice acceptor	14
2A element	74
Silently altered human PAH coding sequence	116
SV40 polyadenylation sequence	31
3' homology arm	117
3' ITR element	19
Correction genome (5' HA to 3' HA)	118
Correction genome (5' ITR to 3' ITR)	119

[00231] The pHMIK-hPAH-hI1C-032 vector was packaged with AAVHSC15 capsids, and was administered into mice intravenously at a dose of 1×10^{14} vector genomes per kilogram of body weight. Liver samples from 3 mice that received the pHMIK-hPAH-hI1C-032 vector packaged with AAVHSC15 capsids were collected, the human and mouse hepatocytes were separated and purified, and DNA was extracted. The efficiency of gene editing was measured using the same ddPCR method as described above.

[00232] The durability of PAH gene editing in human hepatocytes was measured by determining the percentage of edited alleles out of all alleles in cells obtained from treated mice 1 week and 6 weeks post-administration of vector. As shown in Figure 12B, about 4% PAH gene editing was measured in cells obtained from mice 1 week after administration of the vector, and about 7% editing was measured in cells obtained from mice 6 weeks after administration of the vector.

[00233] Genome editing mediated by the pHMIK-hPAH-hI1C-032 vector was found to be specific for human hepatocytes in the HuLiv mice. As shown in Figure 13, at 1 week after administration of the vector, PAH gene editing (as determined by ddPCR and NGS) was detected at a rate of about 3% to 3.5% relative to the number of alleles of human genome, whereas editing of the mouse genome in the liver sample was close to 0% relative to the number of alleles of mouse genome. At 6 weeks after administration of the vector, editing was detected at a rate of about 5% to 6.5% relative to the number of alleles of human genome. * indicates $p<0.0025$ compared to mouse values.

[00234] Further, the pHMIK-hPAH-hI1C-032 vector was found to be ineffective in non-human cells. As shown in Figure 14, when PAH knock-out mouse (*PAH*^{ENU2}) mice were administered intravenously the pHMIK-hPAH-hI1C-032 vector (hPAH-032) packaged with AAVHSC15 capsids at a dose of 1×10^{14} vector genomes per kilogram of body weight, the level of serum phenylalanine was similar to that of mice administered a control up to 3 weeks post-injection. In contrast, mice administered the pHMI-hPAH-mAC-006 vector (mPAH-006) showed reduction in serum Phe levels as soon as 1 week post-injection.

[00235] Figure 15A shows the relationship between human PAH expression and serum Phe levels. As shown, in data gleaned from experiments using the pHMI-hPAH-mAC-006 vector in PAH^{ENU2} mice, 10% of human PAH expression corrects the phenotype in PAH^{ENU2} mice. Thus, 10% of human PAH expression relative to endogenous levels was determined to be the level required to correct phenylalaninemia (e.g., a therapeutic level).

[00236] Therapeutic levels of expression were detected with the pHMIK-hPAH-hI1C-032 vector. Human PAH expression in human hepatocytes was measured relative to human GAPDH in HuLiv mice administered the pHMIK-hPAH-hI1C-032 vector (hPAH-032) at a dose of 1×10^{14} vector genomes per kilogram of body weight. As shown in Figure 15B, using two different expression probes to measure expression of human PAH in two different HuLiv mice treated with the vector, human PAH expression was determined to be greater than 10% in human hepatocytes. The PAH gene editing range in human hepatocytes of HuLiv mice administered the vector was measured to be about 5% to about 11% in 13 different mice across 3 different experiments.

[00237] The pHMIK-hPAH-hI1C-032 vector was found to target human PAH gene and resulted in corrected levels of edited mRNA in HuLiv mice. The PAH mRNA level required for phenotypic correction was first established in a murine model (using the PAH knock-out mouse model (*PAH*^{ENU2})). This was determined to be about 10% of PAH expression relative to endogenous levels (see Figure 15A). As shown in Figure 16, human PAH gene expression relative to GAPDH expression was determined to be about 44.9% (left), and mouse PAH gene expression relative to GAPDH expression was determined to be about 39.7% (right).

Example 6: Human PAH correction vectors

[00238] This example provides the human PAH correction vectors pKITR-hPAH-mAC-006-HCR, pKITR-hPAH-hI1C-032-HCR, pKITR-hPAH-mAC-006-SD.3, pHMIA2-hPAH-hI1C-032-SD.3, and pHMIA2-hPAH-mAC-006-HBB1. Schematics of the vectors are depicted in Figures 17A, 17B, 17C, 17D, and 17E, respectively.

a) pKITR-hPAH-mAC-006-HCR, pKITR-hPAH-hI1C-032-HCR, and pHMIA2-hPAH-mAC-006-HBB1

[00239] Vectors pKITR-hPAH-mAC-006-HCR and pKITR-hPAH-hI1C-032-HCR were generated by inserting an HCR intron into the PAH coding sequence. Vector pHMIA2-hPAH-mAC-006-HBB1 was generated by inserting an HBB1 intron into the PAH coding sequence. The HCR and HBB1 introns were selected based on their performance in intron screening experiments using a luciferase reporter to determine introns that exhibit high expression in liver and blood cell lines. The introns used in the screen are set forth in Table 12.

Table 12: Intron sequences used in luciferase reporter screen

Intron	SEQ ID NO
Chimeric MVM Intron (ChiMVM)	120
SV40 Intron	121
Adenovirus Tripartite Leader Intron (AdTPL)	122
Mini B-Globin Intron	123
AdV/Ig Chimeric Intron (AdVIgG)	124
B-Globin Ig Heavy Chain Intron (BGlobinIg)	125
Wu MVM Intron (Wu MVM)	126
HCR1 Intron (OptHCR)	127
B-Globin Intron	128
tFIX Intron (FIX)	129
ch2BLood Intron (BloodEnh)	130

[00240] pKITR-hPAH-mAC-006-HCR comprised 5' to 3' the following genetic elements: a 5' ITR element, a 5' homology arm, a silently altered human PAH coding sequence with HCR intron inserted therein, an SV40 polyadenylation sequence, a targeted integration restriction cassette ("TI RE"), a 3' homology arm, and a 3' ITR element. pKITR-hPAH-hI1C-032-HCR comprised 5' to 3' the following genetic elements: a 5' ITR element, a 5' homology arm, a splice acceptor, a 2A element, a silently altered human PAH coding sequence with HCR intron inserted therein, an SV40 polyadenylation sequence, a 3' homology arm, and a 3' ITR element. pHMIA2-hPAH-mAC-006-HBB1 comprised 5' to 3'

the following genetic elements: a 5' ITR element, a 5' homology arm, a silently altered human PAH coding sequence with HBB intron inserted therein, an SV40 polyadenylation sequence, a targeted integration restriction cassette ("TI RE"), a 3' homology arm, and a 3' ITR element. The sequences of these elements are set forth in Table 13.

Table 13: Genetic elements in PAH correction vectors pKITR-hPAH-mAC-006-HCR, pKITR-hPAH-hI1C-032-HCR, and pHMIA2-hPAH-mAC-006-HBB1

	SEQ ID NO		
Genetic Element	**-006-HCR**	**-032-HCR**	**-006-HBB1**
5' ITR element	18	18	18
5' homology arm	111	115	142
Splice acceptor	N/A	14	N/A
2A element	N/A	74	N/A
Human PAH coding sequence	131	132	143
SV40 polyadenylation sequence	31	31	31
Targeted integration restriction cassette	71	N/A	71
3' homology arm	112	117	144
3' ITR element	19	19	19
Correction genome (5' HA to 3' HA)	134	136	145
Correction genome (5' ITR to 3' ITR)	135	137	146

[00241] *b) pKITR-hPAH-mAC-006-SD.3 and pHMIA2-hPAH-hI1C-032-SD.3*Vectors pKITR-hPAH-mAC-006-SD.3 and pHMIA2-hPAH-hI1C-032-SD.3 were generated by modifying a splice donor site. The splice donor was modified as indicated in Figures 17C and 17D, respectively. pKITR-hPAH-mAC-006-SD.3 comprised 5' to 3' the following genetic elements: a 5' ITR element, a 5' homology arm, a silently altered human PAH coding sequence with splice donor modification, an SV40 polyadenylation sequence, a targeted integration restriction cassette ("TI RE"), a 3' homology arm, and a 3' ITR element. pHMIA2-hPAH-hI1C-032-SD.3 comprised 5' to 3' the following genetic elements: a 5' ITR element, a 5' homology arm, a splicing acceptor, a 2A element, a silently altered human PAH coding sequence with splice donor modification, an SV40 polyadenylation sequence, a 3'

homology arm, and a 3' ITR element. The sequences of these elements are set forth in Table 14.

Table 14: Genetic elements in PAH correction vectors pKITR-hPAH-mAC-006-SD.3 and pHMIA2-hPAH-hI1C-032-SD.3

	SEQ ID NO	
Genetic Element	**-006-SD.3**	**-032-SD.3**
5' ITR element	18	18
5' homology arm	111	115
Splice acceptor	N/A	14
2A element	N/A	74
Human PAH coding sequence	138	139
SV40 polyadenylation sequence	31	31
Targeted integration restriction cassette	71	N/A
3' homology arm	112	117
3' ITR element	19	19

* * *

[00242] 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.

[00243] 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 phenylalanine hydroxylase (PAH) gene in a cell, the method comprising transducing the cell with a replication-defective adeno-associated virus (AAV) comprising:

(a) an AAV capsid; and

(b) a correction genome comprising: (i) an editing element for editing a target locus in the PAH 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 hepatocyte, a renal cell, or a cell in the brain, pituitary gland, adrenal gland, pancreas, urinary bladder, gallbladder, colon, small intestine, or breast.

3. The method of claim 1 or 2, 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.

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

(a) an AAV capsid; and

(b) a correction genome comprising: (i) an editing element for editing a target locus in the PAH 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 an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease is not co-administered to the subject.

5. The method of claim 4, wherein the disease or disorder is phenylketonuria.

6. The method of claim 4 or 5, wherein the subject is a human subject.

7. The method of any one of the preceding claims, wherein the editing element comprises at least a portion of a PAH coding sequence.

8. The method of claim 7, wherein the editing element comprises a PAH coding sequence.

9. The method of claim 7 or 8, wherein the PAH coding sequence encodes an amino acid sequence set forth in SEQ ID NO: 23.

10. The method of claims 7-9, wherein the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 24.

11. The method of any one of claims 7-9, wherein the PAH coding sequence is silently altered.

12. The method of claim 11, wherein the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25, 116, 131, 132, 138, 139, or 143..

13. The method of claim 7, wherein the editing element comprises a PAH intron-inserted coding sequence, optionally wherein the PAH intron-inserted coding sequence comprises a nonnative intron inserted in a PAH coding sequence.

14. The method of claim 13, wherein the nonnative intron is selected from the group consisting of a first intron of a hemoglobin beta gene and a minute virus in mice (MVM) intron.

15. The method of claim 14, wherein the nonnative intron consists of a nucleotide sequence at least 90% identical to any one of SEQ ID NOs: 28-30, and 120-130.

16. The method of claim 15, wherein the nonnative intron consists of a nucleotide sequence set forth in any one of SEQ ID NOs: 28-30, and 120-130.

17. The method of any one of claims 13-16, wherein the PAH intron-inserted coding sequence encodes an amino acid sequence set forth in SEQ ID NO: 23.

18. The method of any one of claims 13-17, wherein the PAH intron-inserted coding sequence comprises from 5' to 3': a first portion of a PAH coding sequence, the intron, and a second portion of a PAH coding sequence, wherein the first portion and the second portion, when spliced together, form a complete PAH coding sequence.

19. The method of claims 18, wherein the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 24.

20. The method of claims 18, wherein the PAH coding sequence is silently altered.

21. The method of claim 20, wherein the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25, 116, 131, 132, 138, 139, or 143..

22. The method of any one of claims 18-21, wherein the first portion of the PAH coding sequence comprises the amino acid sequence set forth in SEQ ID NO: 64 or 65, and/or the second portion of the PAH coding sequence comprises the amino acid sequence set forth in SEQ ID NO: 66 or 67.

23. The method of any one of claims 18-22, wherein the first portion of the PAH coding sequence consist of the amino acid sequence set forth in SEQ ID NO: 64 or 65, and the second portion of the PAH coding sequence consists of the amino acid sequence set forth in SEQ ID NO: 66 or 67.

24. The method of any one of claims 7-23, wherein the editing element comprises from 5' to 3': a ribosomal skipping element, and the PAH coding sequence or the PAH intron-inserted coding sequence.

25. The method of claim 24, wherein the editing element further comprises a polyadenylation sequence 3' to the PAH coding sequence or the PAH intron-inserted coding sequence.

26. The method of claim 25, wherein the polyadenylation sequence is an exogenous polyadenylation sequence, optionally wherein the exogenous polyadenylation sequence is an SV40 polyadenylation sequence.

27. The method of claim 26, wherein the SV40 polyadenylation sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 31-34.

28. The method of any one of claims 24-27, wherein the nucleotide 5' to the target locus is in an exon of the PAH gene.

29. The method of any one of claims 24-28, wherein the nucleotide 5' to the target locus is in exon 1 of the PAH gene.

30. The method of any one of claims 24-27, wherein the editing element further comprises a splice acceptor 5' to the ribosomal skipping element.

31. The method of claim 30, wherein the nucleotide 5' to the target locus is in an intron of the PAH gene.

32. The method of claim 30 or 31, wherein the nucleotide 5' to the target locus is in intron 1 of the PAH gene.

33. The method of any one of the preceding claims, wherein the editing element comprises the nucleotide sequence set forth in SEQ ID NO: 35.

34. 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.

35. 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.

36. 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.

37. The method of claim 36, wherein the first editing window consists of the nucleotide sequence set forth in SEQ ID NO: 36 or 45.

38. The method of claim 36 or 37, wherein the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 36 or 45.

39. The method of any one of claims 36-38, wherein the first editing window consists of the nucleotide sequence set forth in SEQ ID NO: 36, and the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 45.

40. The method of any one of the preceding claims, wherein the first genomic region consists of the nucleotide sequence set forth in SEQ ID NO: 36.

41. The method of any one of the preceding claims, wherein the second genomic region consists of the nucleotide sequence set forth in SEQ ID NO: 45.

42. 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.

43. The method of any one of the preceding claims, wherein the 5' homology arm comprises: C corresponding to nucleotide -2 of the PAH gene, G corresponding to nucleotide 4 of the PAH gene, G corresponding to nucleotide 6 of the PAH gene, G corresponding to nucleotide 7 of the PAH gene, G corresponding to nucleotide 9 of the PAH gene, A corresponding to nucleotide -467 of the PAH gene, A corresponding to nucleotide -465 of the PAH gene, A corresponding to nucleotide -181 of the PAH gene, G corresponding to nucleotide -214 of the PAH gene, C corresponding to nucleotide -212 of the PAH gene, A corresponding to nucleotide -211 of the PAH gene, G corresponding to nucleotide 194 of the PAH gene, C corresponding to nucleotide -433 of the PAH gene, C corresponding to nucleotide -432 of the PAH gene, ACGCTGTTCTTCGCC (SEQ ID NO: 68) corresponding

to nucleotides -394 to -388 of the PAH gene, A corresponding to nucleotide -341 of the PAH gene, A corresponding to nucleotide -339 of the PAH gene, A corresponding to nucleotide -225 of the PAH gene, A corresponding to nucleotide -211 of the PAH gene, and/or A corresponding to nucleotide -203 of the PAH gene.

44. The method of any one of the preceding claims, wherein the 5' homology arm comprises:

(a) C corresponding to nucleotide -2 of the PAH gene, G corresponding to nucleotide 4 of the PAH gene, G corresponding to nucleotide 6 of the PAH gene, G corresponding to nucleotide 7 of the PAH gene, and G corresponding to nucleotide 9 of the PAH gene;

(b) A corresponding to nucleotide -467 of the PAH gene, and A corresponding to nucleotide -465 of the PAH gene;

(c) A corresponding to nucleotide -181 of the PAH gene;

(d) G corresponding to nucleotide -214 of the PAH gene, C corresponding to nucleotide -212 of the PAH gene, and A corresponding to nucleotide -211 of the PAH gene;

(e) G corresponding to nucleotide 194 of the PAH gene;

(f) C corresponding to nucleotide -433 of the PAH gene, and C corresponding to nucleotide -432 of the PAH gene;

(g) ACGCTGTTCTTCGCC (SEQ ID NO: 68) corresponding to nucleotides -394 to -388 of the PAH gene; and/or

(h) A corresponding to nucleotide -341 of the PAH gene, A corresponding to nucleotide -339 of the PAH gene, A corresponding to nucleotide -225 of the PAH gene, A corresponding to nucleotide -211 of the PAH gene, and A corresponding to nucleotide -203 of the PAH gene.

45. The method of claim 44, wherein the 5' homology arm comprises the modifications of (c) and (d), (f) and (g), and/or (b) and (h).

46. The method of any one of the preceding claims, wherein the 5' homology arm consists of a nucleotide sequence set forth in any one of SEQ ID NOs: 36-44, 111, 115, and 142.

47. 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: 45, 112, 117, 144.

48. 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: 46-54, 113, 118, 134, 136, and 145.

49. 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.

50. The method of claim 49, 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.

51. The method of claim 49, 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.

52. The method of claim 49, wherein the 5' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO: 26, and the 3' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO: 27.

53. 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: 55-63, 114, 119, 135, 137, and 146.

54. The method of any one of the preceding claims, wherein the correction genome consists of the nucleotide sequence set forth in any one of SEQ ID NOs: 55-63, 114, 119, 135, 137, and 146.

55. The method of any one of the preceding claims, wherein the AAV capsid comprises an AAV Clade F capsid protein.

56. 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, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17.

57. The method of claim 56, 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.

58. The method of claim 57, 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.

59. The method of claim 57, 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.

60. 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, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17.

61. The method of claim 60, 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.

62. The method of claim 61, 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.

63. The method of claim 61, 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.

64. 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, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

65. The method of claim 64, 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.

66. The method of claim 65, 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.

67. The method of claim 65, 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.

68. 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 administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions.

69. 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 administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions.

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

(a) an AAV capsid; and

(b) a correction genome comprising: (i) an editing element for editing a target locus in the PAH 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,

71. The AAV of claim 70, wherein the editing element comprises at least a portion of a PAH coding sequence.

72. The AAV of claim 70 or 71, wherein the editing element comprises a PAH coding sequence.

73. The AAV of claim 71 or 72, wherein the PAH coding sequence encodes an amino acid sequence set forth in SEQ ID NO: 23.

74. The AAV of claim 73, wherein the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 24.

75. The AAV of claim 73, wherein the PAH coding sequence is silently altered.

76. The AAV of claim 75, wherein the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25, 116, 131, 132, 138, 139, or 143.

77. The AAV of claim 71, wherein the editing element comprises a PAH intron-inserted coding sequence, optionally wherein the PAH intron-inserted coding sequence comprises a nonnative intron inserted in a PAH coding sequence.

78. The AAV of claim 77, wherein the nonnative intron is selected from the group consisting of a first intron of a hemoglobin beta gene and a minute virus in mice (MVM) intron.

79. The AAV of claim 78, wherein the nonnative intron consists of a nucleotide sequence at least 90% identical to any one of SEQ ID NOs: 28-30, and 120-130.

80. The AAV of claim 79, wherein the nonnative intron consists of a nucleotide sequence set forth in any one of SEQ ID NOs: 28-30, and 120-130.

81. The AAV of any one of claims 77-80, wherein the PAH intron-inserted coding sequence encodes an amino acid sequence set forth in SEQ ID NO: 23.

82. The AAV of any one of claims 77-81, wherein the PAH intron-inserted coding sequence comprises from 5' to 3': a first portion of a PAH coding sequence, the intron, and a second portion of a PAH coding sequence, wherein the first portion and the second portion, when spliced together, form a complete PAH coding sequence.

83. The AAV of claims 82, wherein the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 24.

84. The AAV of claims 82, wherein the PAH coding sequence is silently altered.

85. The AAV of claim 84, wherein the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25, 116, 131, 132, 138, 139, or 143..

86. The AAV of any one of claims 82-85, wherein the first portion of the PAH coding sequence comprises the amino acid sequence set forth in SEQ ID NO: 64 or 65, and/or the second portion of the PAH coding sequence comprises the amino acid sequence set forth in SEQ ID NO: 66 or 67.

87. The AAV of any one of claims 82-86, wherein the first portion of the PAH coding sequence consist of the amino acid sequence set forth in SEQ ID NO: 64 or 65, and the second portion of the PAH coding sequence consists of the amino acid sequence set forth in SEQ ID NO: 66 or 67.

88. The AAV of any one of claims 71-87, wherein the editing element comprises from 5' to 3': a ribosomal skipping element, and the PAH coding sequence or the PAH intron-inserted coding sequence.

89. The AAV of claim 88, wherein the editing element further comprises a polyadenylation sequence 3' to the PAH coding sequence or the PAH intron-inserted coding sequence.

90. The AAV of claim 89, wherein the polyadenylation sequence is an exogenous polyadenylation sequence, optionally wherein the exogenous polyadenylation sequence is an SV40 polyadenylation sequence.

91. The AAV of claim 90, wherein the SV40 polyadenylation sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 31-34.

92. The AAV of any one of claims 88-91, wherein the nucleotide 5' to the target locus is in an exon of the PAH gene.

93. The AAV of any one of claims 88-92, wherein the nucleotide 5' to the target locus is in exon 1 of the PAH gene.

94. The AAV of any one of claims 88-91, wherein the editing element further comprises a splice acceptor 5' to the ribosomal skipping element.

95. The AAV of claim 94, wherein the nucleotide 5' to the target locus is in an intron of the PAH gene.

96. The AAV of claim 94 or 95, wherein the nucleotide 5' to the target locus is in intron 1 of the PAH gene.

97. The AAV of any one of claims 70-96, wherein the editing element comprises the nucleotide sequence set forth in SEQ ID NO: 35.

98. The AAV of any one of claims 70-97, wherein the 5' homology arm nucleotide sequence is at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the first genomic region.

99. The AAV of any one of claims 70-98, wherein the 3' homology arm nucleotide sequence is at least 90%, 95%, 96%, 97%, 98%, or 99% identical to the second genomic region.

100. The AAV of any one of claims 70-99, wherein the first genomic region is located in a first editing window, and the second genomic region is located in a second editing window.

101. The AAV of claim 100, wherein the first editing window consists of the nucleotide sequence set forth in SEQ ID NO: 36 or 45.

102. The AAV of claim 100 or 101, wherein the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 36 or 45.

103. The AAV of any one of claims 100-102, wherein the first editing window consists of the nucleotide sequence set forth in SEQ ID NO: 36, and the second editing window consists of the nucleotide sequence set forth in SEQ ID NO: 45.

104. The AAV of any one of claims 70-103, wherein the first genomic region consists of the nucleotide sequence set forth in SEQ ID NO: 36.

105. The AAV of any one of claims 70-104, wherein the second genomic region consists of the nucleotide sequence set forth in SEQ ID NO: 45.

106. The AAV of any one of claims 70-105, wherein each of the 5' and 3' homology arm nucleotide sequences independently has a length of about 100 to about 2000 nucleotides.

107. The AAV of any one of claims 70-106, wherein the 5' homology arm comprises: C corresponding to nucleotide -2 of the PAH gene, G corresponding to nucleotide 4 of the PAH gene, G corresponding to nucleotide 6 of the PAH gene, G corresponding to nucleotide 7 of the PAH gene, G corresponding to nucleotide 9 of the PAH gene, A corresponding to nucleotide -467 of the PAH gene, A corresponding to nucleotide -465 of the PAH gene, A corresponding to nucleotide -181 of the PAH gene, G corresponding to nucleotide -214 of the PAH gene, C corresponding to nucleotide -212 of the PAH gene, A corresponding to nucleotide -211 of the PAH gene, G corresponding to nucleotide 194 of the PAH gene, C

corresponding to nucleotide -433 of the PAH gene, C corresponding to nucleotide -432 of the PAH gene, ACGCTGTTCTTCGCC (SEQ ID NO: 68) corresponding to nucleotides -394 to -388 of the PAH gene, A corresponding to nucleotide -341 of the PAH gene, A corresponding to nucleotide -339 of the PAH gene, A corresponding to nucleotide -225 of the PAH gene, A corresponding to nucleotide -211 of the PAH gene, and/or A corresponding to nucleotide -203 of the PAH gene.

108. The AAV of any one of claims 70-107, wherein the 5' homology arm comprises:
(a) C corresponding to nucleotide -2 of the PAH gene, G corresponding to nucleotide 4 of the PAH gene, G corresponding to nucleotide 6 of the PAH gene, G corresponding to nucleotide 7 of the PAH gene, and G corresponding to nucleotide 9 of the PAH gene;
(b) A corresponding to nucleotide -467 of the PAH gene, and A corresponding to nucleotide -465 of the PAH gene;
(c) A corresponding to nucleotide -181 of the PAH gene;
(d) G corresponding to nucleotide -214 of the PAH gene, C corresponding to nucleotide -212 of the PAH gene, and A corresponding to nucleotide -211 of the PAH gene;
(e) G corresponding to nucleotide 194 of the PAH gene;
(f) C corresponding to nucleotide -433 of the PAH gene, and C corresponding to nucleotide -432 of the PAH gene;
(g) ACGCTGTTCTTCGCC (SEQ ID NO: 68) corresponding to nucleotides -394 to -388 of the PAH gene; and/or
(h) A corresponding to nucleotide -341 of the PAH gene, A corresponding to nucleotide -339 of the PAH gene, A corresponding to nucleotide -225 of the PAH gene, A corresponding to nucleotide -211 of the PAH gene, and A corresponding to nucleotide -203 of the PAH gene.

109. The AAV of claim 108, wherein the 5' homology arm comprises the modifications of (c) and (d), (f) and (g), and/or (b) and (h).

110. The AAV of any one of claims 70-109, wherein the 5' homology arm consists of a nucleotide sequence set forth in any one of SEQ ID NOs: 36-44, 111, 115, and 142.

111. The AAV of any one of claims 70-110, wherein the 3' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 45, 112, 117, 144.

112. The AAV of any one of claims 70-111, wherein the correction genome comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 46-54, 113, 118, 134, 136, and 145.

113. The AAV of any one of claims 70-112, 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.

114. The AAV of claim 113, 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.

115. The AAV of claim 113, 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.

116. The AAV of claim 113, wherein the 5' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO: 26, and the 3' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO: 27.

117. The AAV of any one of claims 70-116, wherein the correction genome comprises the nucleotide sequence set forth in any one of SEQ ID NOs: 55-63, 114, 119, 135, 137, and 146.

118. The AAV of any one of claims 70-117, wherein the correction genome consists of the nucleotide sequence set forth in any one of SEQ ID NOs: 55-63, 114, 119, 135, 137, and 146.

119. The AAV of any one of claims 70-118, wherein the AAV capsid comprises an AAV Clade F capsid protein.

120. The AAV of any one of claims 70-119, 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, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17.

121. The AAV of claim 120, 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.

122. The AAV of claim 121, 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.

123. The AAV of claim 121, 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.

124. The AAV of any one of claims 70-123, 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, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17.

125. The AAV of claim 124, 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.

126. The AAV of claim 125, 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.

127. The AAV of claim 125, 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.

128. The AAV of any one of claims 70-127, 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, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

129. The AAV of claim 128, 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.

130. The AAV of claim 129, 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.

131. The AAV of claim 129, 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.

132. The AAV of any one of claims 70-131, wherein the integration efficiency of the editing element into the target locus is at least 1% when the AAV is administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions.

133. The AAV of any one of claims 70-132, wherein the allelic frequency of integration of the editing element into the target locus is at least 0.5% when the AAV is administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions.

134. A pharmaceutical composition comprising an AAV of any one of claims 70-133.

135. A packaging system for recombinant preparation of a replication-defective 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 119-131; and

(c) a correction genome as set forth in any one of claims 70-118, 132, and 133, wherein the packaging system is operative in a cell for enclosing the correction genome in the capsid to form the AAV.

136. The packaging system of claim 135, 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.

137. The packaging system of claims 135 or 136, wherein the Rep nucleotide sequence encodes an AAV2 Rep protein.

138. The packaging system of claim 137, wherein the AAV2 Rep protein is 78/68 or Rep 68/52.

139. The packaging system of claim 137 or 138, 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.

140. The packaging system of any one of claims 136-139, further comprising a third vector, wherein the third vector is a helper virus vector.

141. The packaging system of claim 140, wherein the helper virus vector is an independent third vector.

142. The packaging system of claim 140, wherein the helper virus vector is integral with the first vector.

143. The packaging system of claim 140, wherein the helper virus vector is integral with the second vector.

144. The packaging system of any one of claims 140-143, wherein the third vector comprises genes encoding helper virus proteins.

145. The packaging system of any one of claims 140-144, wherein the helper virus is selected from the group consisting of adenovirus, herpes virus, vaccinia virus, and cytomegalovirus (CMV).

146. The packaging system of claim 145, wherein the helper virus is adenovirus.

147. The packaging system of claim 146, wherein the adenovirus genome comprises one or more adenovirus RNA genes selected from the group consisting of E1, E2, E4 and VA.

148. The packaging system of claim 145, wherein the helper virus is herpes simplex virus (HSV).

149. The packaging system of claim 148, 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.

150. The packaging system of any one of claims 140-149, wherein the first vector and the third vector are contained within a first transfecting plasmid.

151. The packaging system of any one of claims 140-149, wherein the nucleotides of the second vector and the third vector are contained within a second transfecting plasmid.

152. The packaging system of any one of claims 140-149, wherein the nucleotides of the first vector and the third vector are cloned into a recombinant helper virus.

153. The packaging system of any one of claims 140-149, wherein the nucleotides of the second vector and the third vector are cloned into a recombinant helper virus.

154. A method for recombinant preparation of a replication-defective AAV, the method comprising introducing the packaging system of any one of claims 135-153 into a cell under conditions operative for enclosing the correction genome in the capsid to form the AAV.

Figure 1A

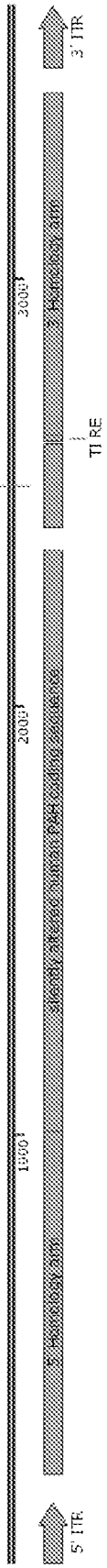


Figure 1B

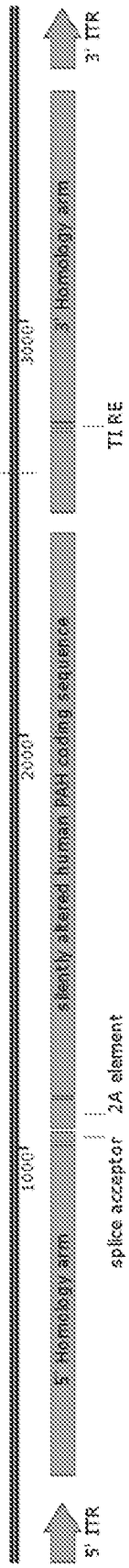


Figure 1C

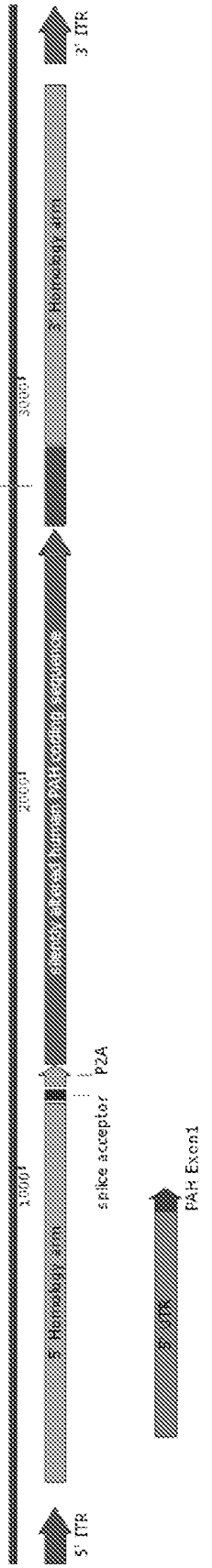


Figure 2

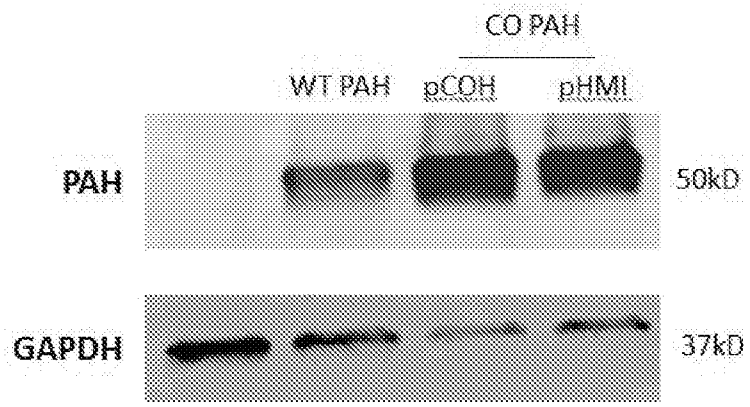


Figure 3A

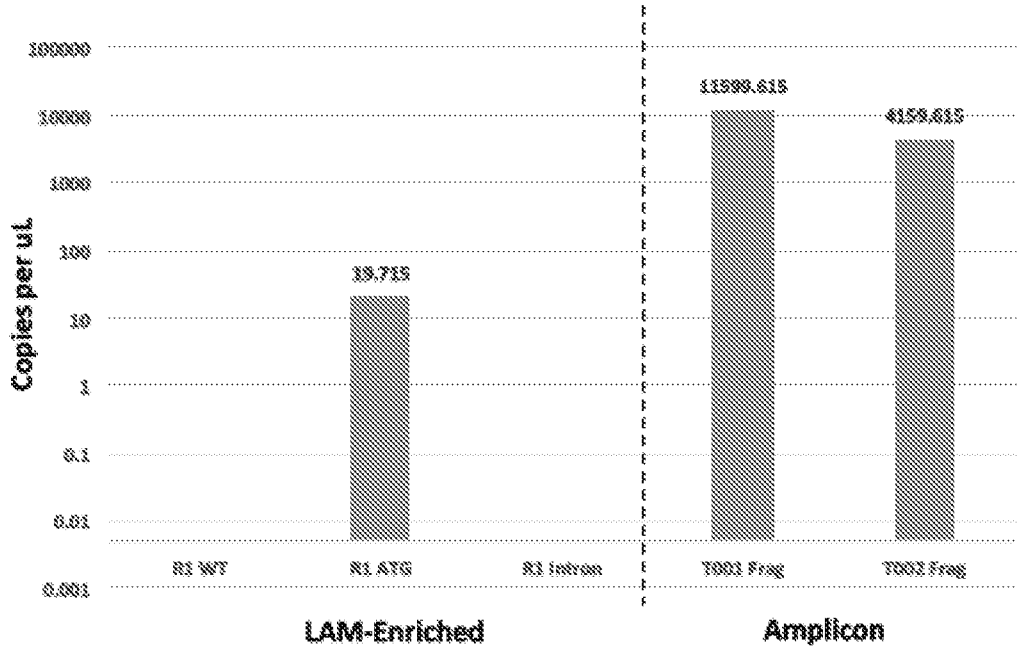


Figure 3B

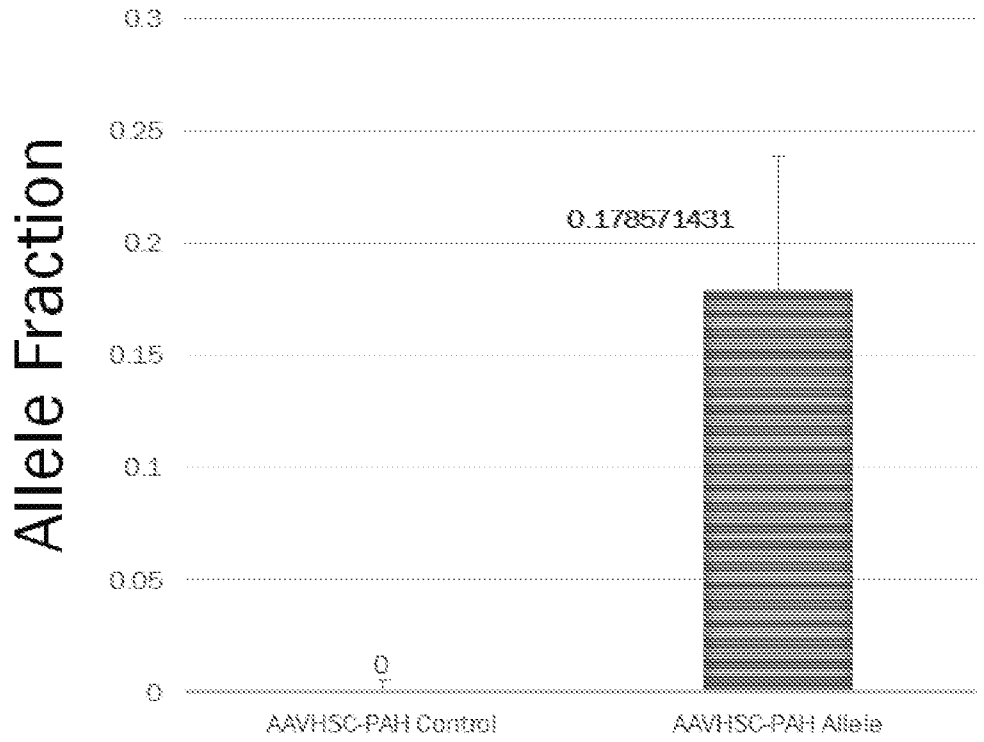


Figure 4A

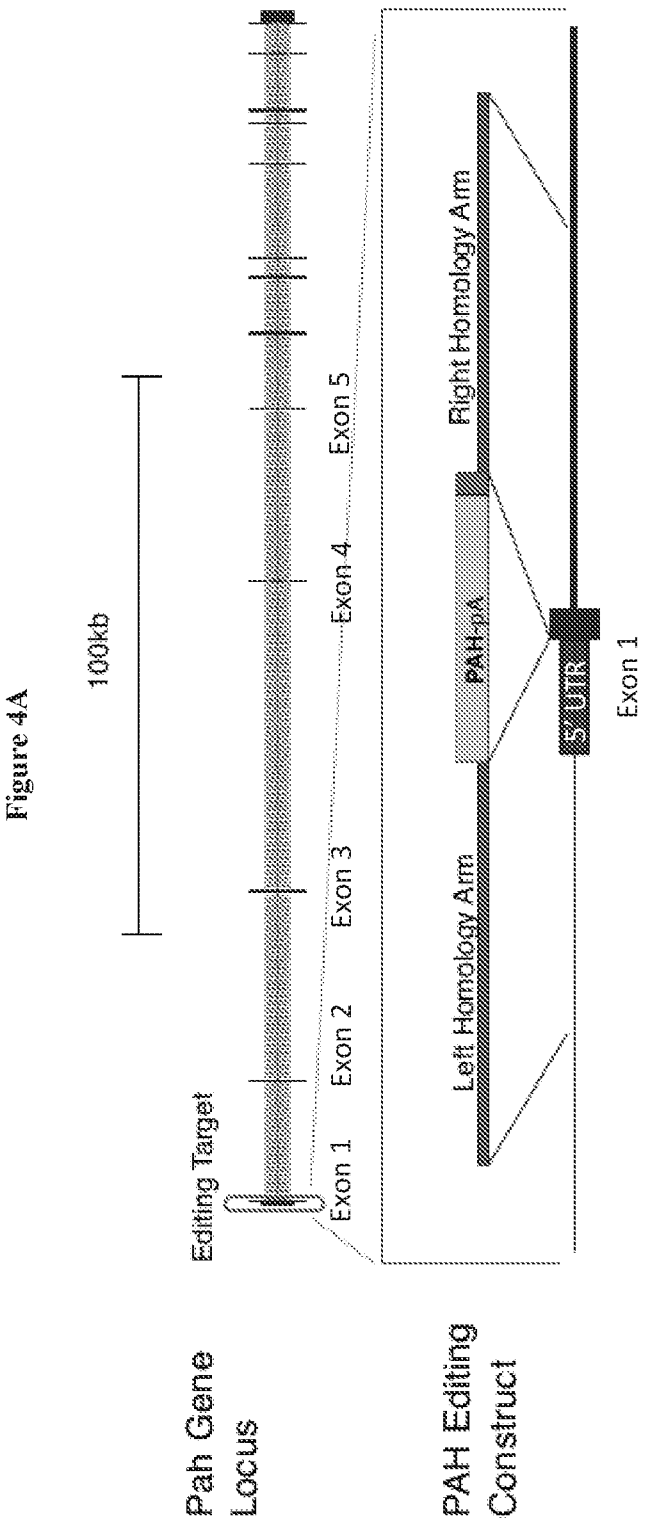

100kb
Pah Gene Locus
Editing Target
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
PAH Editing Construct
Left Homology Arm
PAHpA
Right Homology Arm
5' UTR
Exon 1

Figure 4B

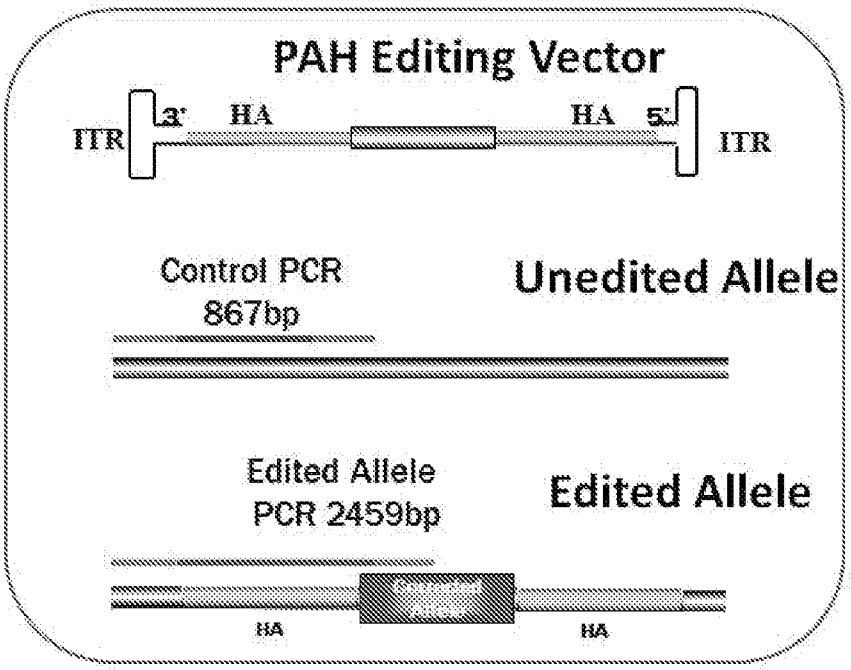


Figure 4C

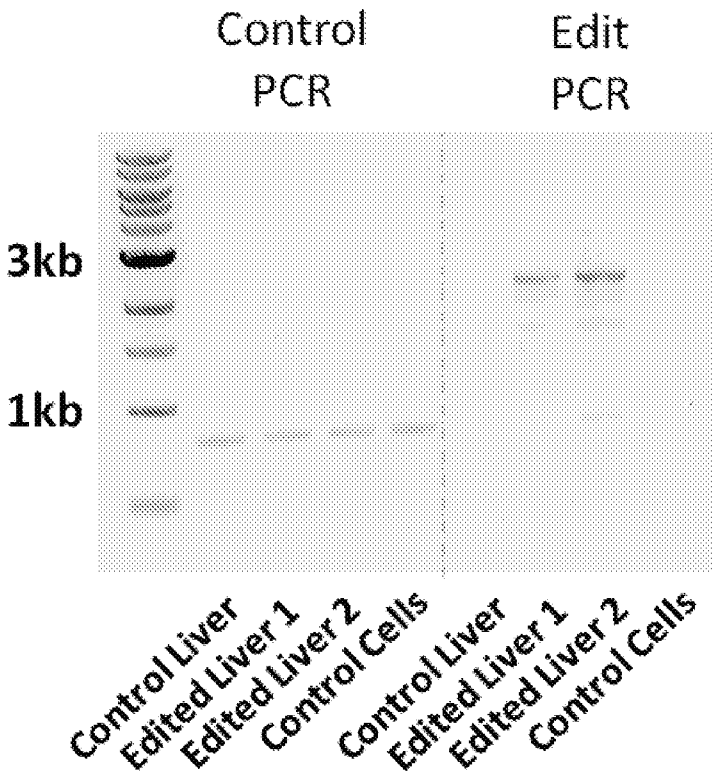
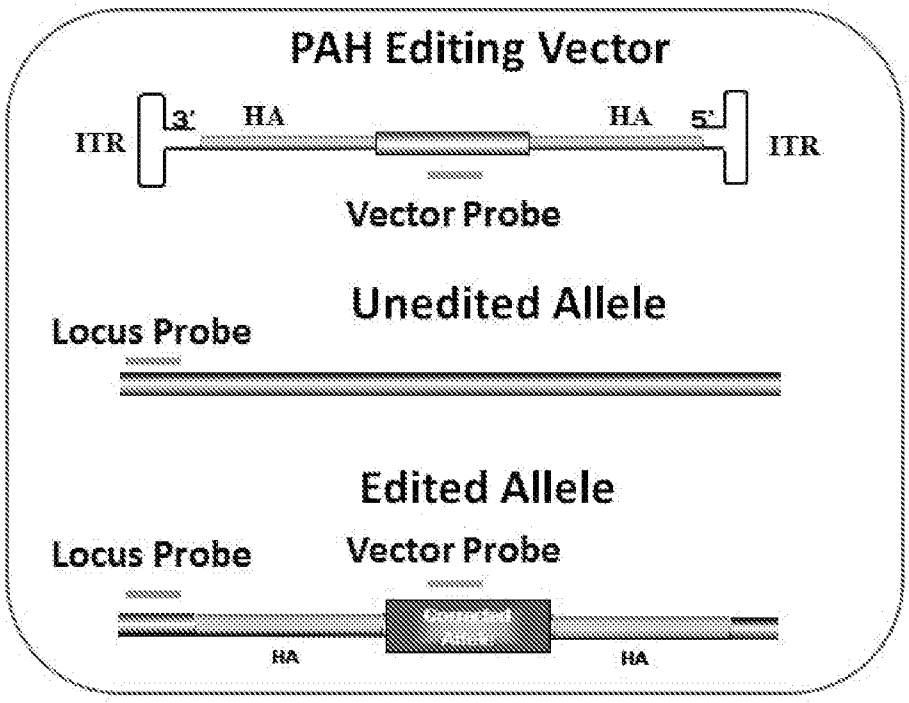


Figure 5A


PAH Editing Vector
ITR
3'
HA
HA
5'
ITR
Vector Probe
Unedited Allele
Locus Probe
Edited Allele
Locus Probe
Vector Probe
HA
HA

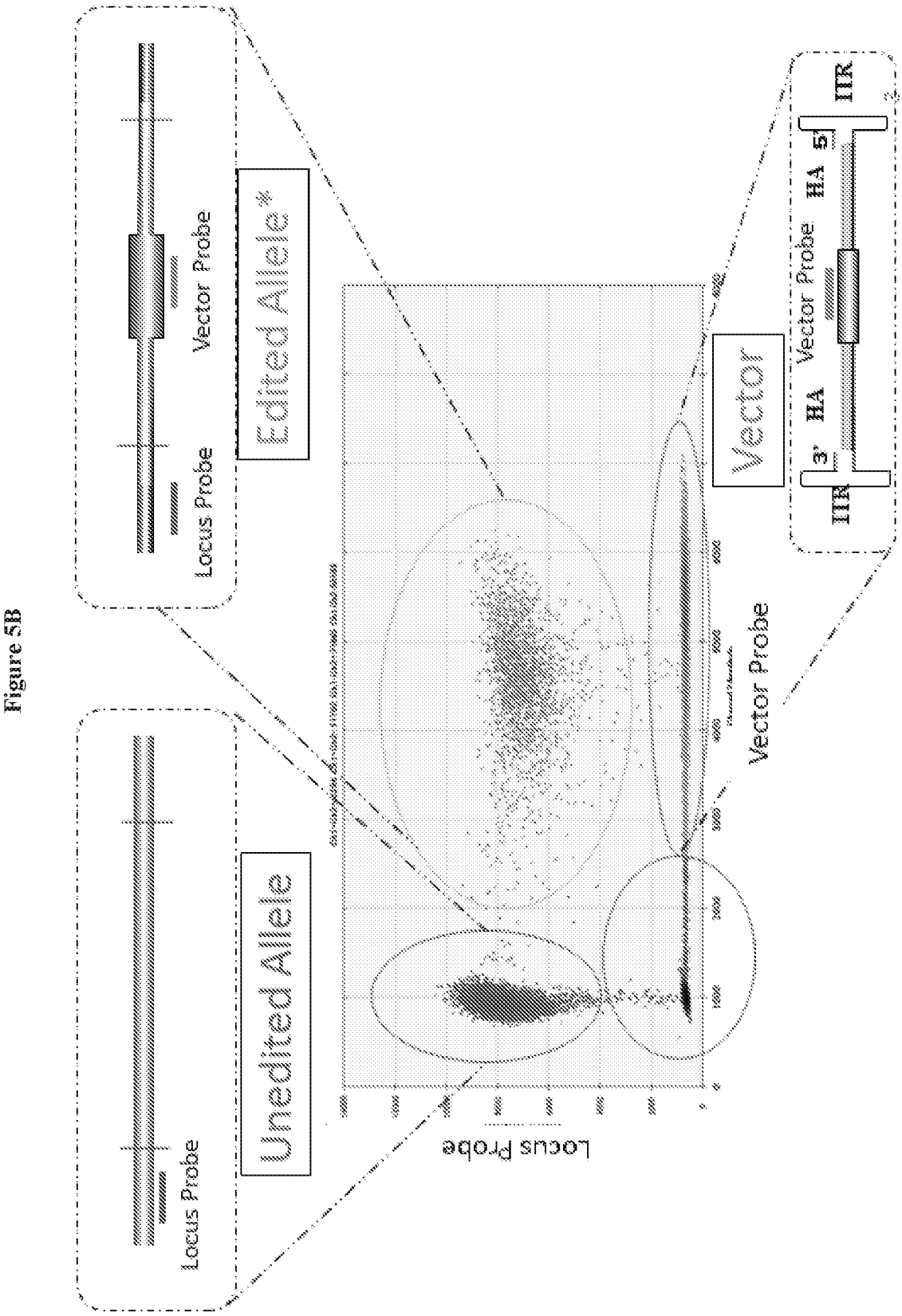

Edited Allele*
Locus Probe
Vector Probe
Unedited Allele
Locus Probe
Vector
ITR
HA
Vector Probe
HA
ITR
Vector Probe
Locus Probe

Figure 5B

Figure 5C

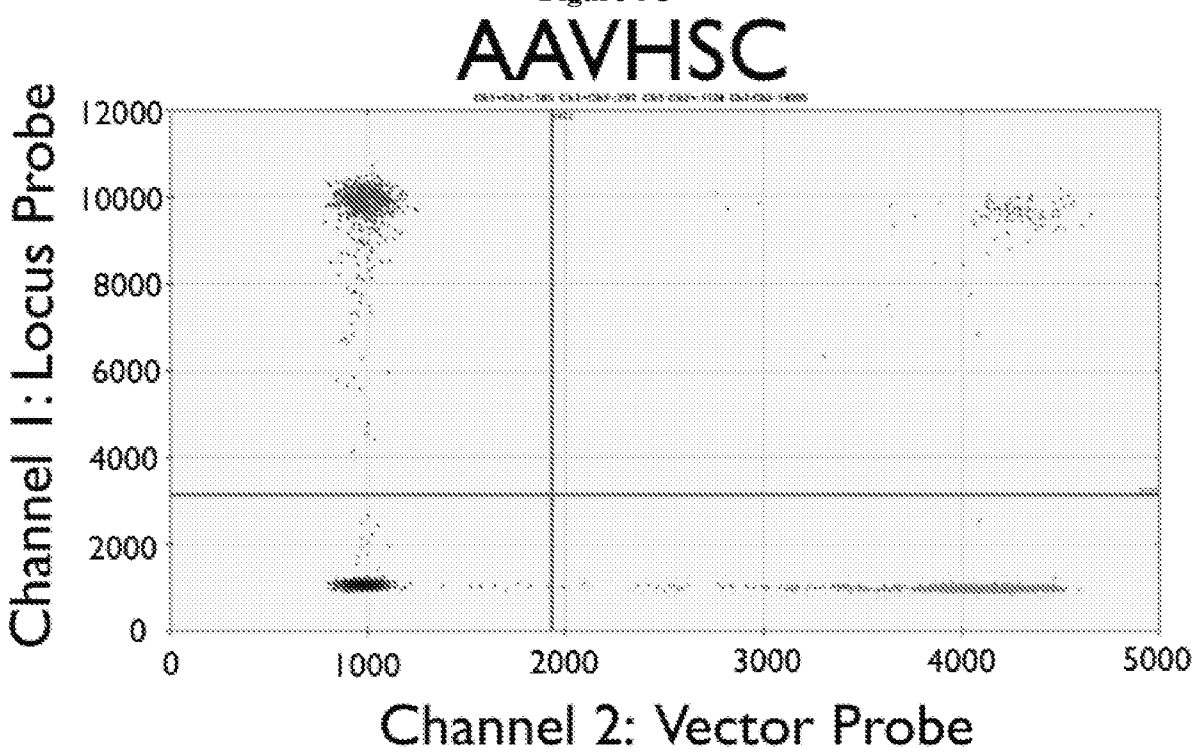


Figure 5D

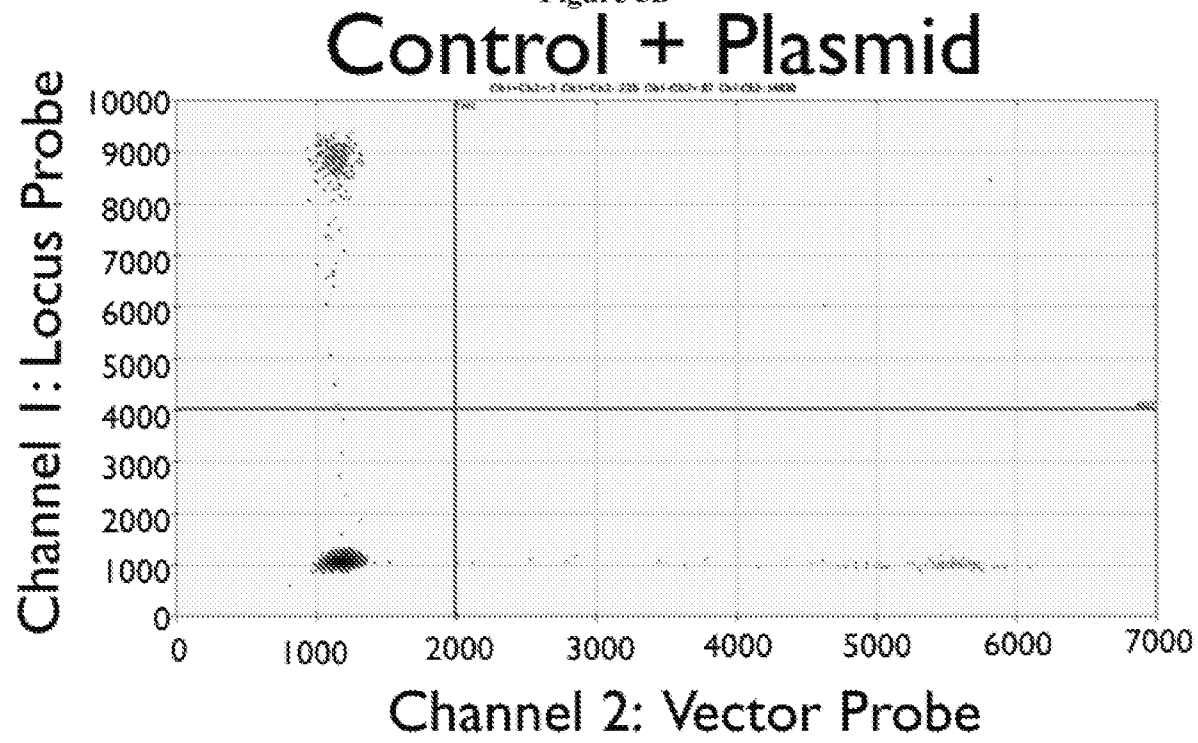


Figure 5E

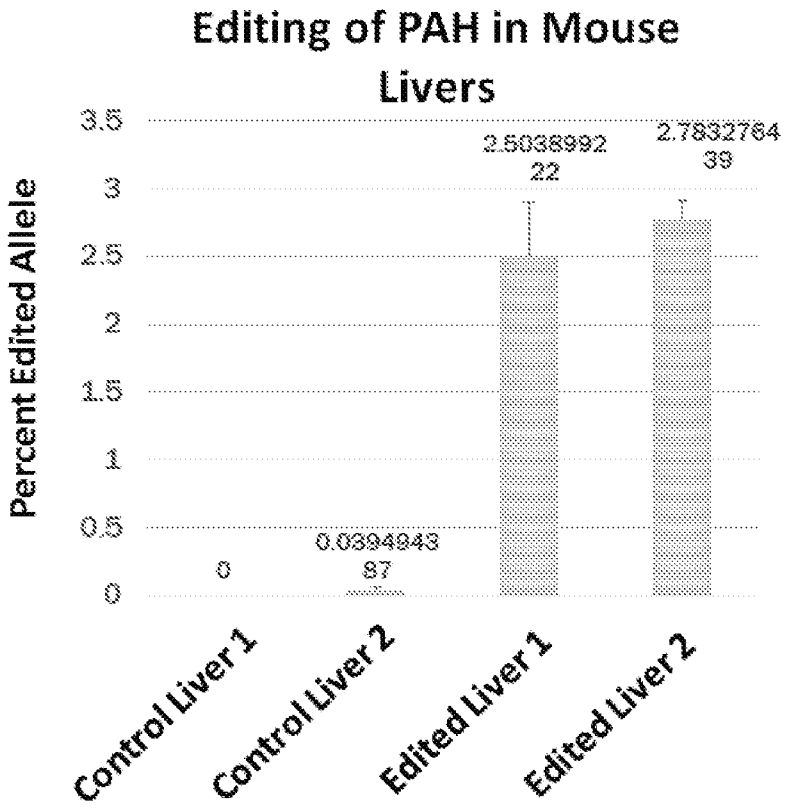

Editing of PAH in Mouse Livers
Percent Edited Allele
3.5
3
2.5
2
1.5
1
0.5
0
0
0.0394943 87
2.5038992 22
2.7832764 39
Control Liver 1
Control Liver 2
Edited Liver 1
Edited Liver 2

Figure 6

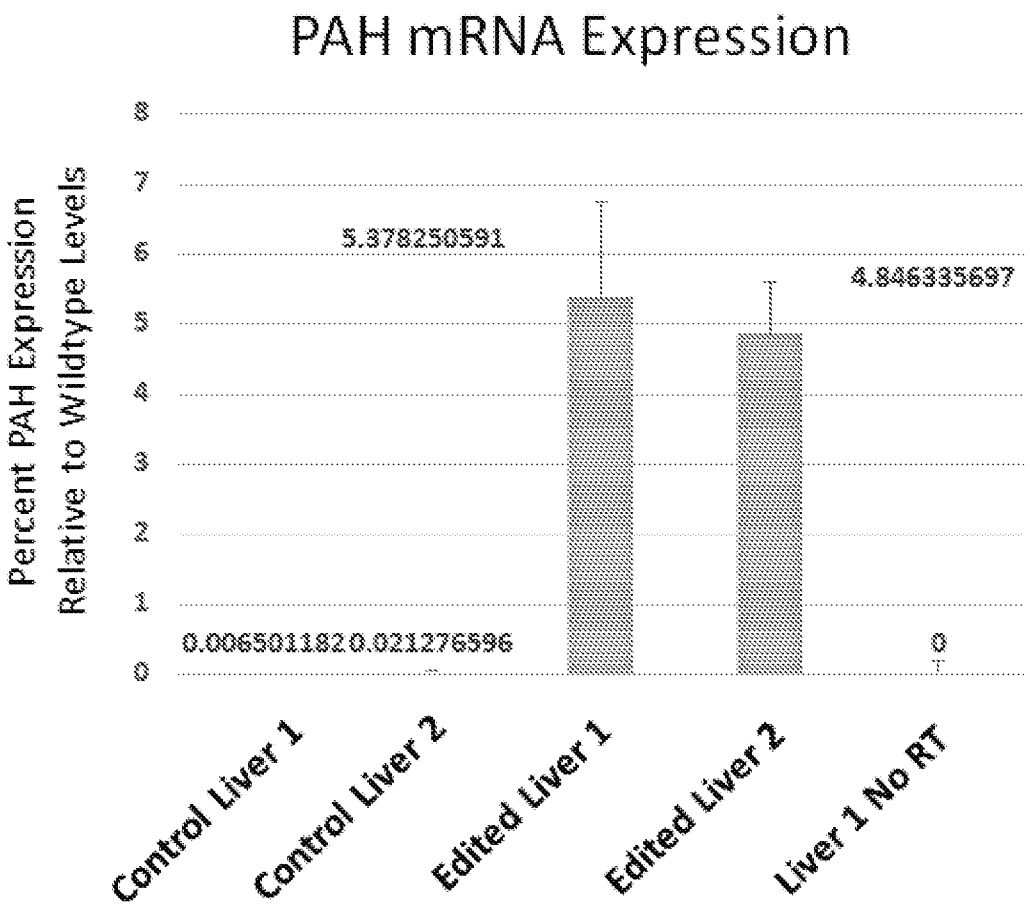

PAH mRNA Expression
Percent PAH Expression Relative to Wildtype Levels
0
1
2
3
4
5
6
7
8
0.006501182
0.021276596
5.378250591
4.846335697
0
Control Liver 1
Control Liver 2
Edited Liver 1
Edited Liver 2
Liver 1 No RT

Figure 7A

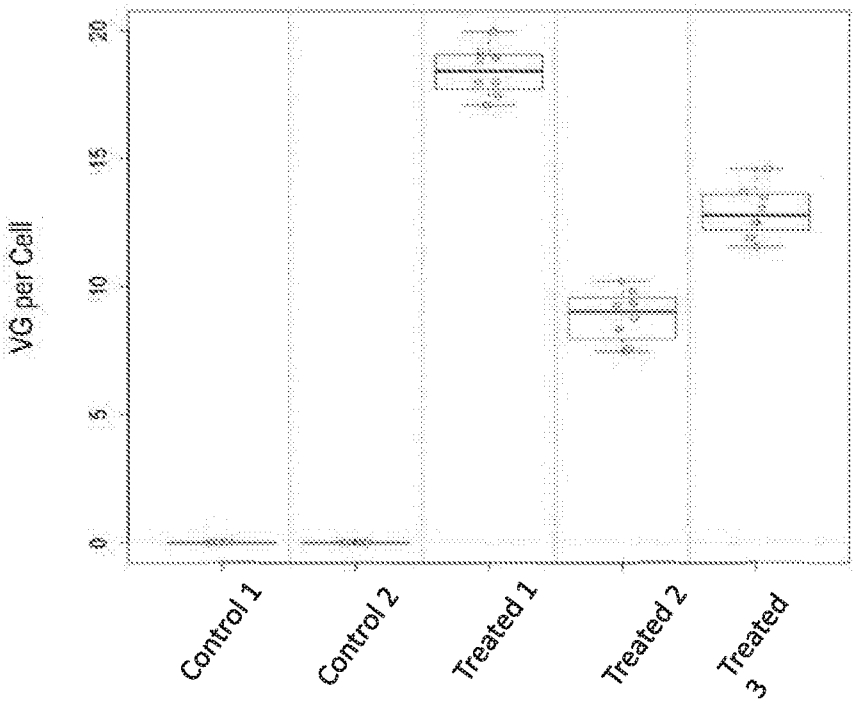


Figure 7B

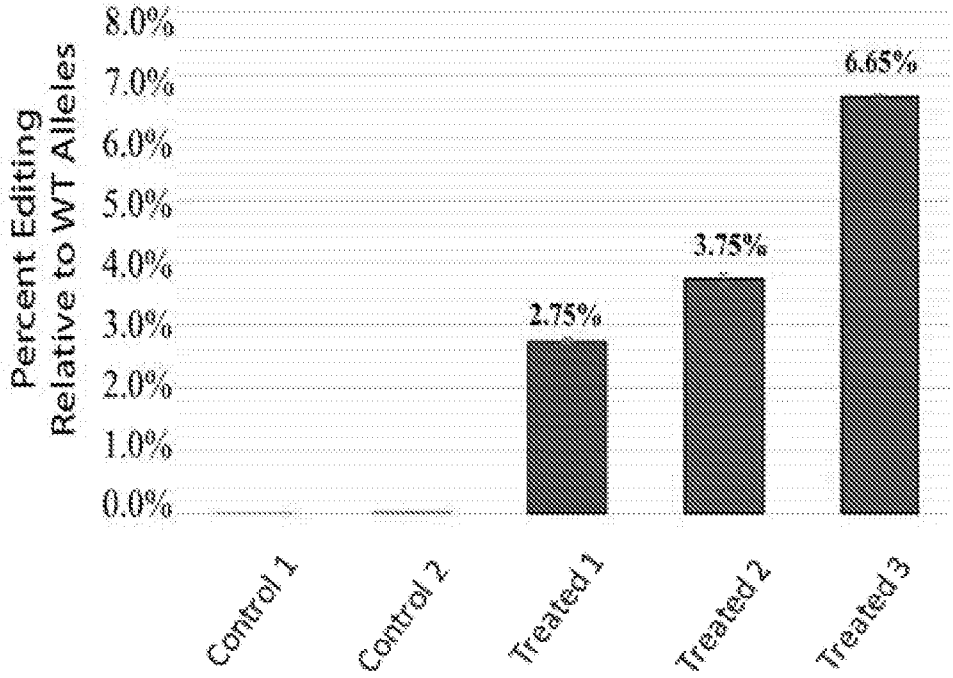


Figure 7C

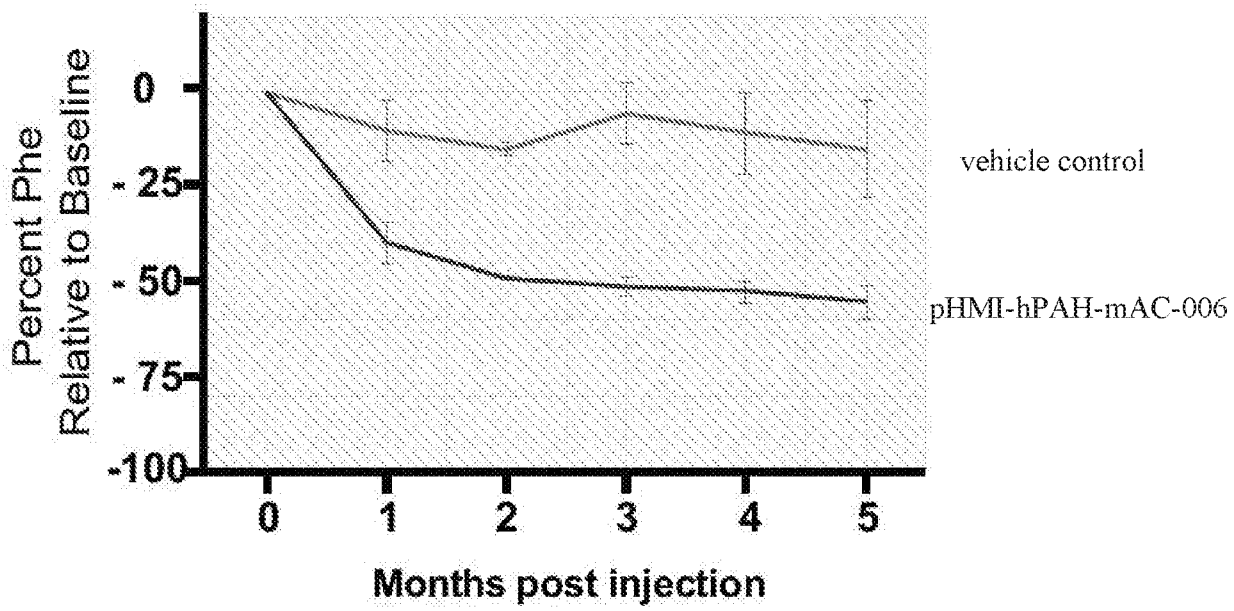


Figure 7D

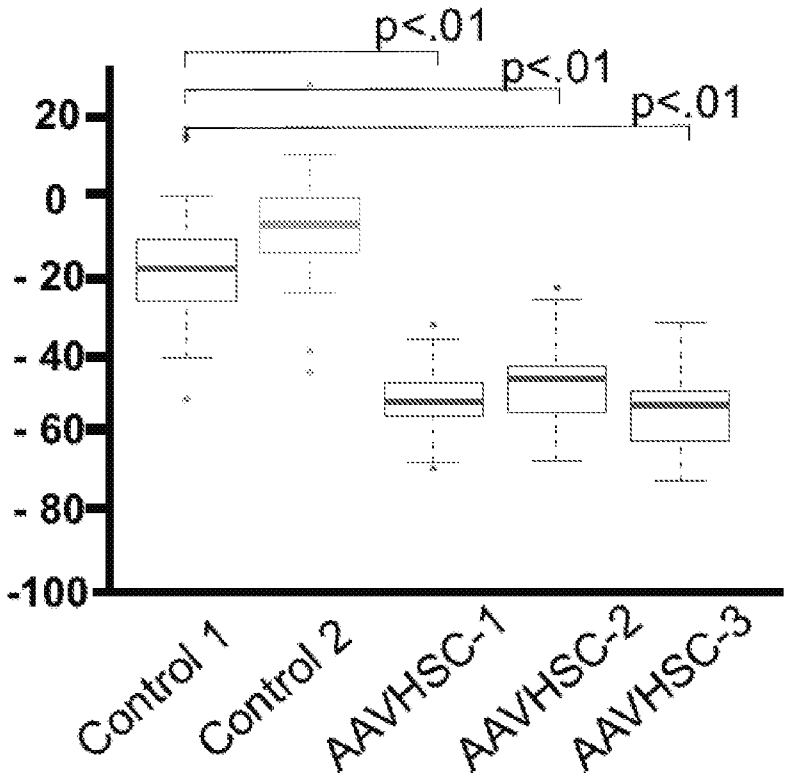


Figure 7E

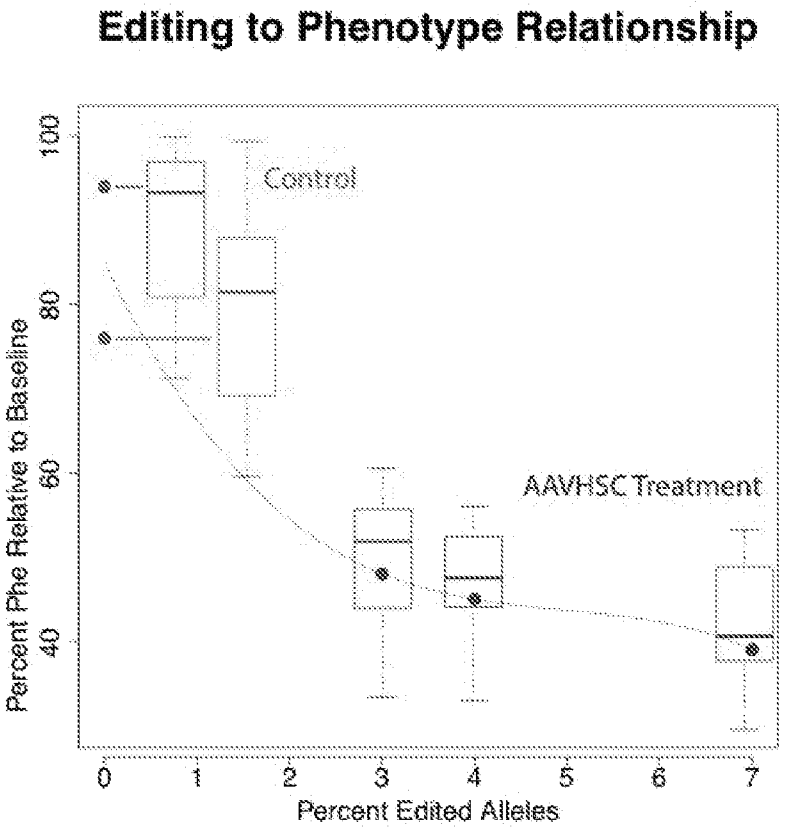

Editing to Phenotype Relationship
Control
AAVHSC Treatment
Percent Phe Relative to Baseline
100
80
60
40
0
1
2
3
4
5
6
7
Percent Edited Alleles

Figure 7F

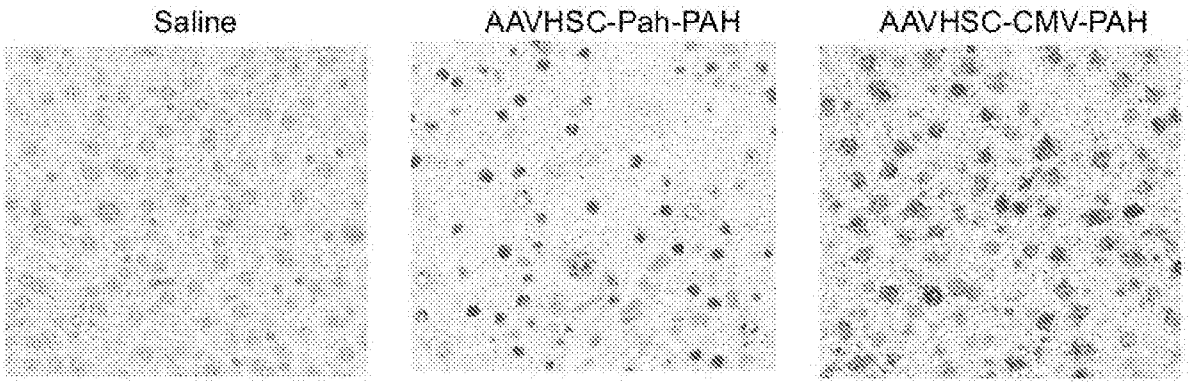


Figure 8A

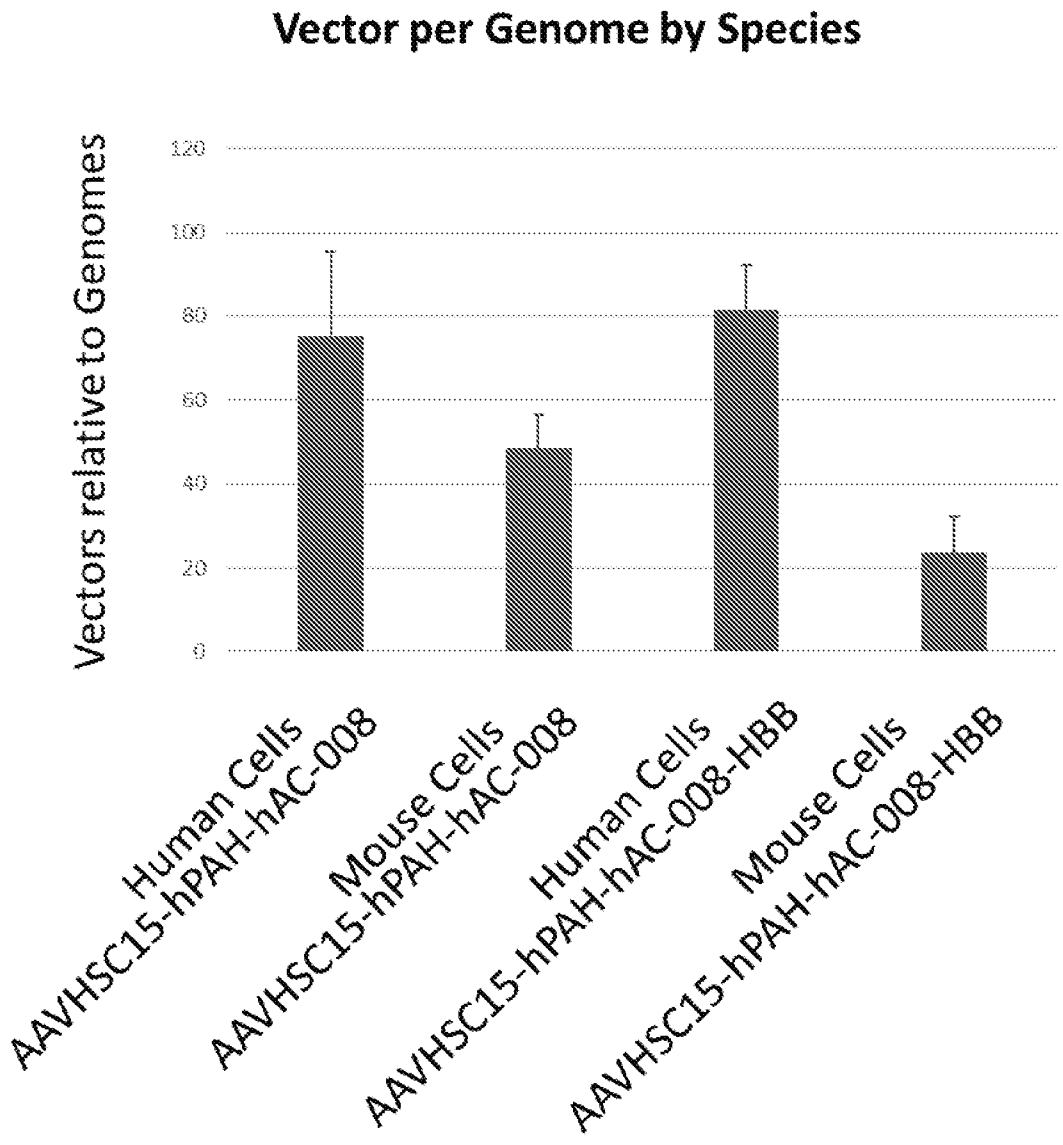

Vector per Genome by Species
Vectors relative to Genomes
120
100
80
60
40
20
0
Human Cells AAVHSC15-hPAH-hAC-008
Mouse Cells AAVHSC15-hPAH-hAC-008
Human Cells AAVHSC15-hPAH-hAC-008-HBB
Mouse Cells AAVHSC15-hPAH-hAC-008-HBB

Figure 8B

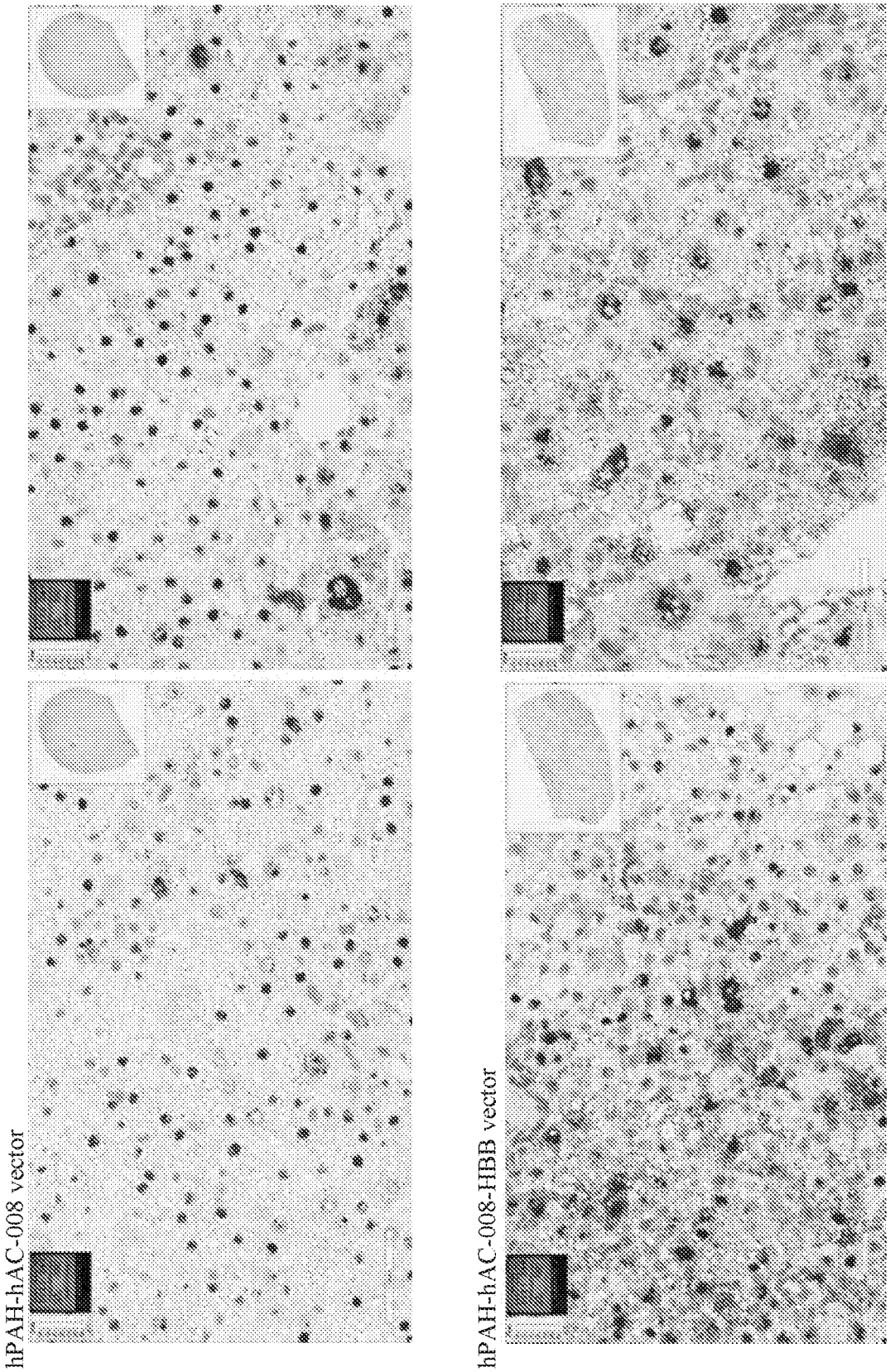


Figure 8C

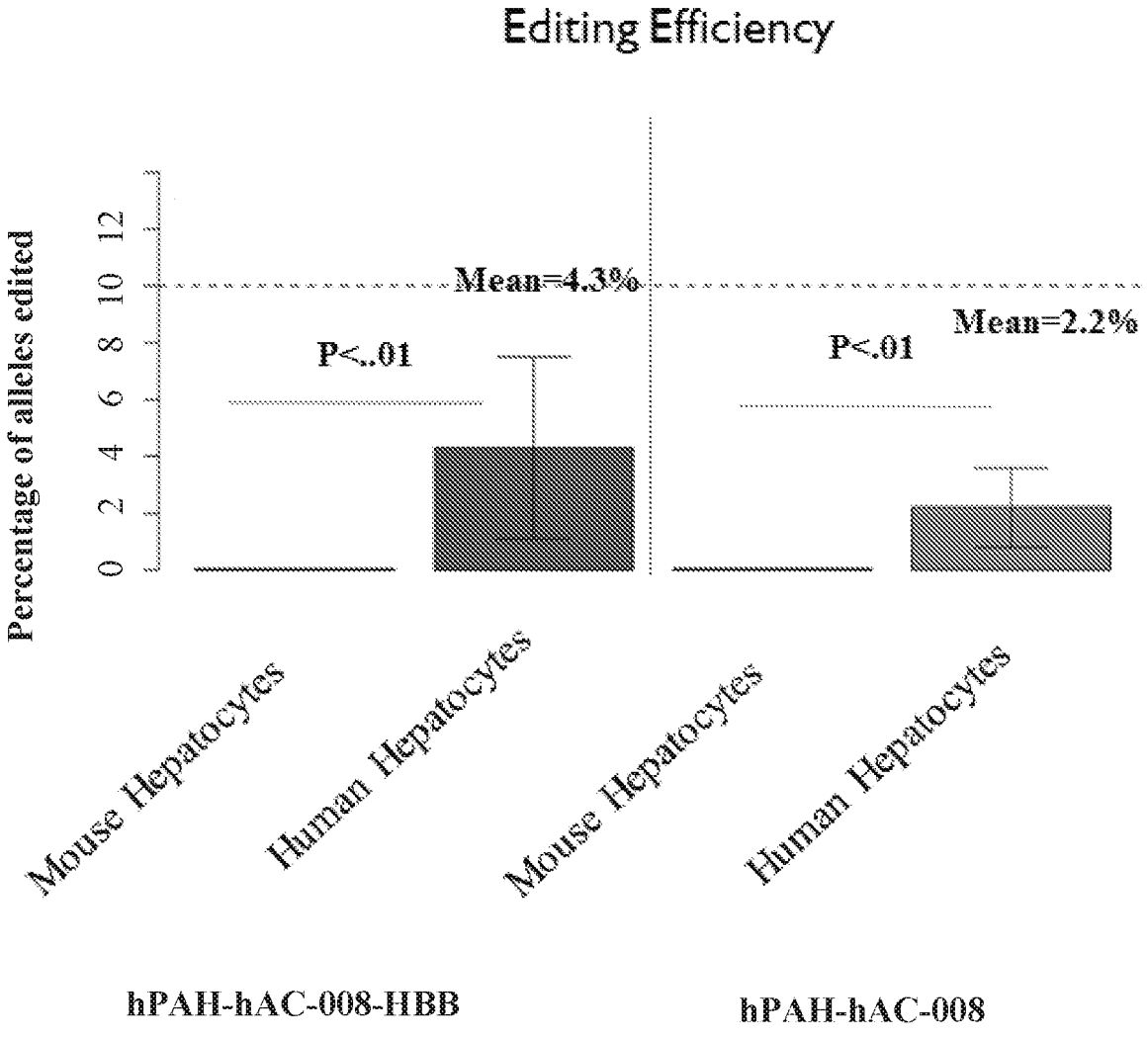

Editing Efficiency
Percentage of alleles edited
0 2 4 6 8 10 12
Mean=4.3%
Mean=2.2%
P<.01
P<.01
Mouse Hepatocytes
Human Hepatocytes
Mouse Hepatocytes
Human Hepatocytes
hPAH-hAC-008-HBB
hPAH-hAC-008

Figure 9A

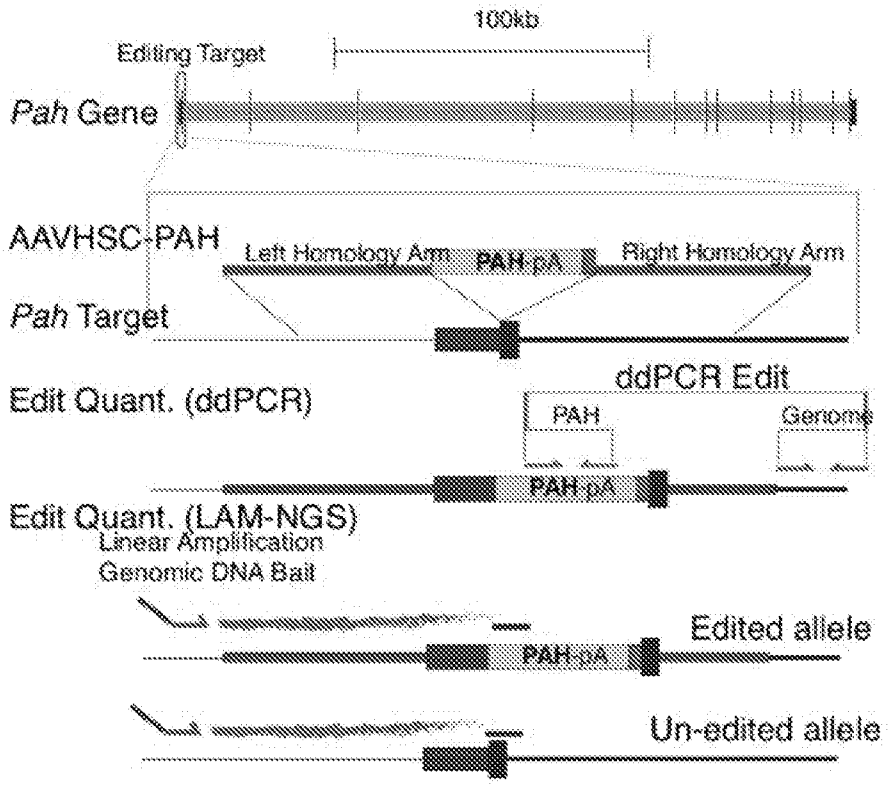


Figure 9B

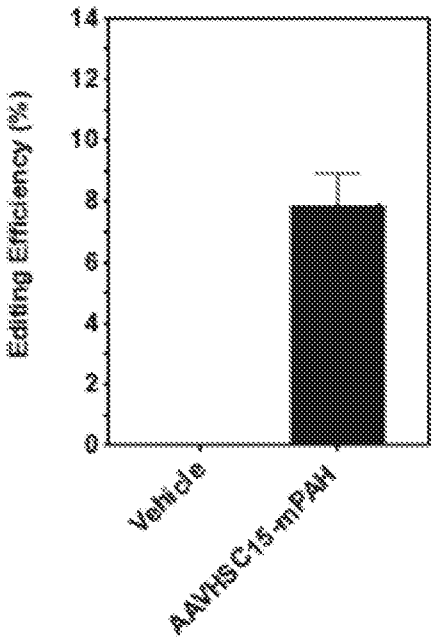


Figure 10A

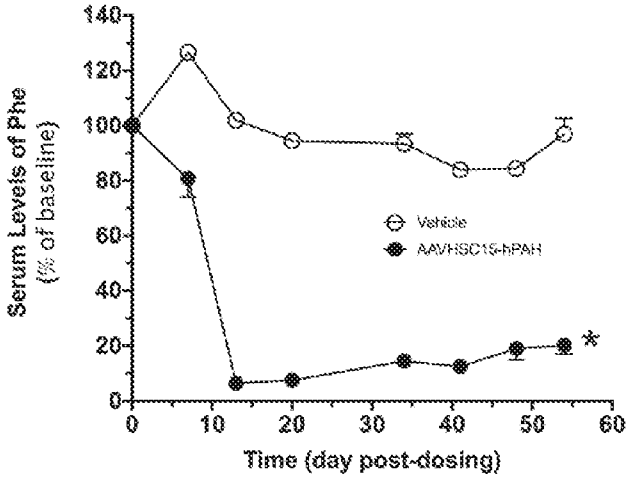


Figure 10B

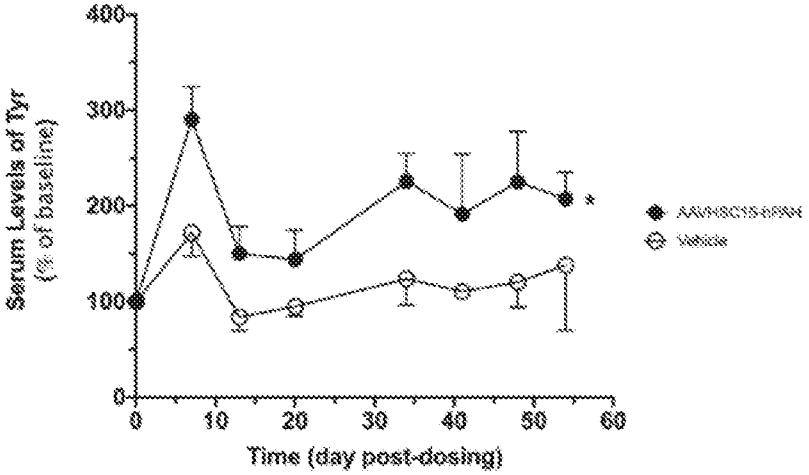


Figure 10C

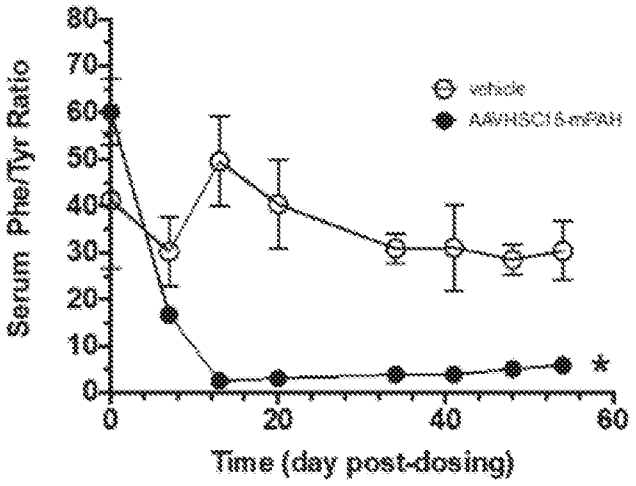


Figure 11A

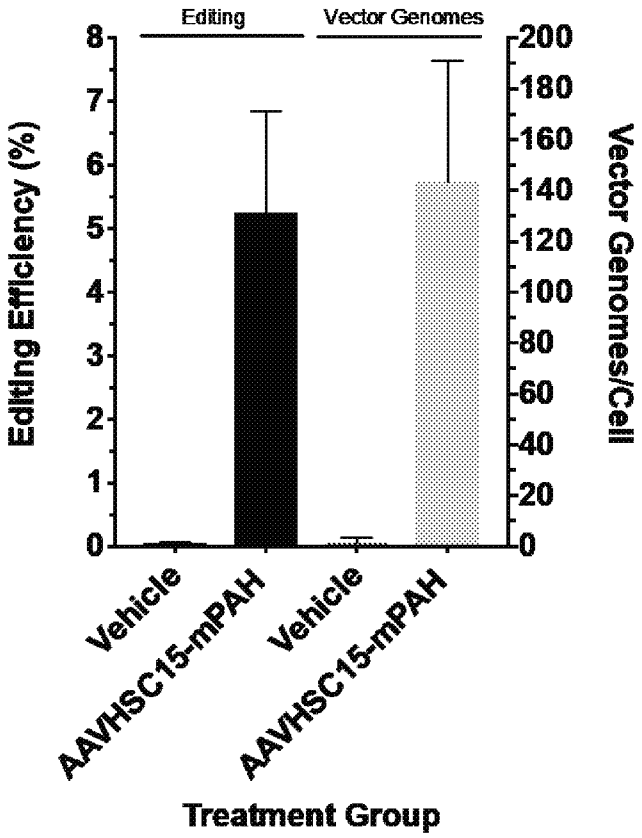


Figure 11B

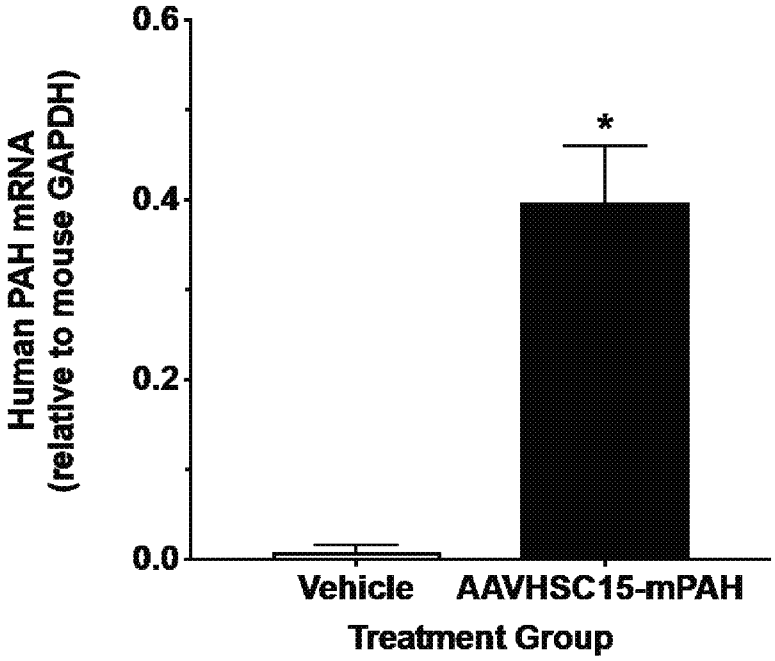


Figure 12A

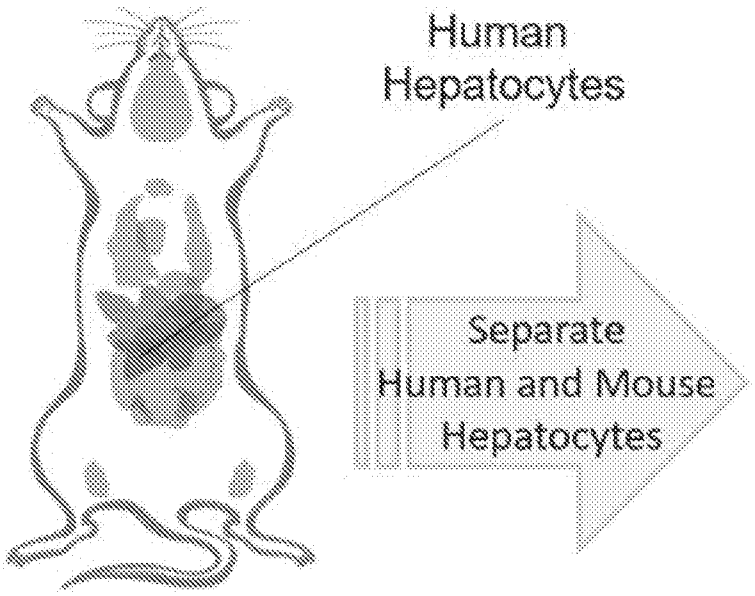


Figure 12B

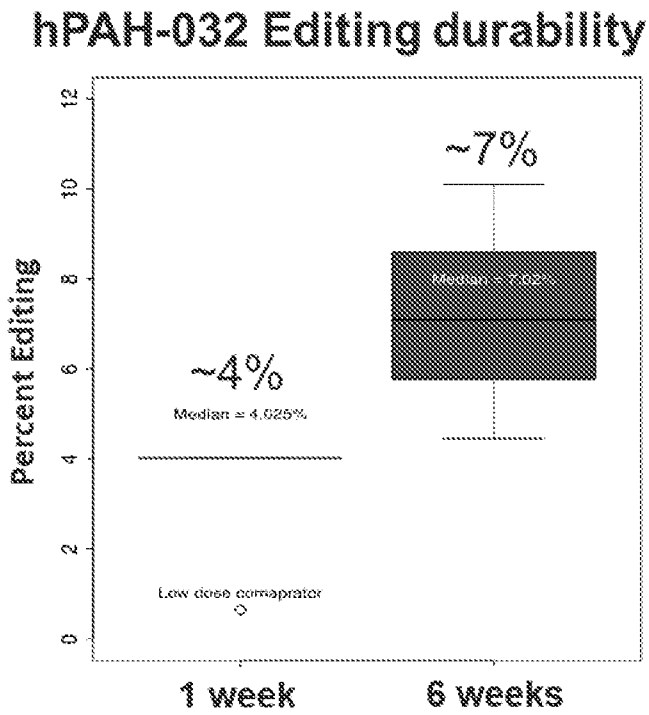


Figure 13

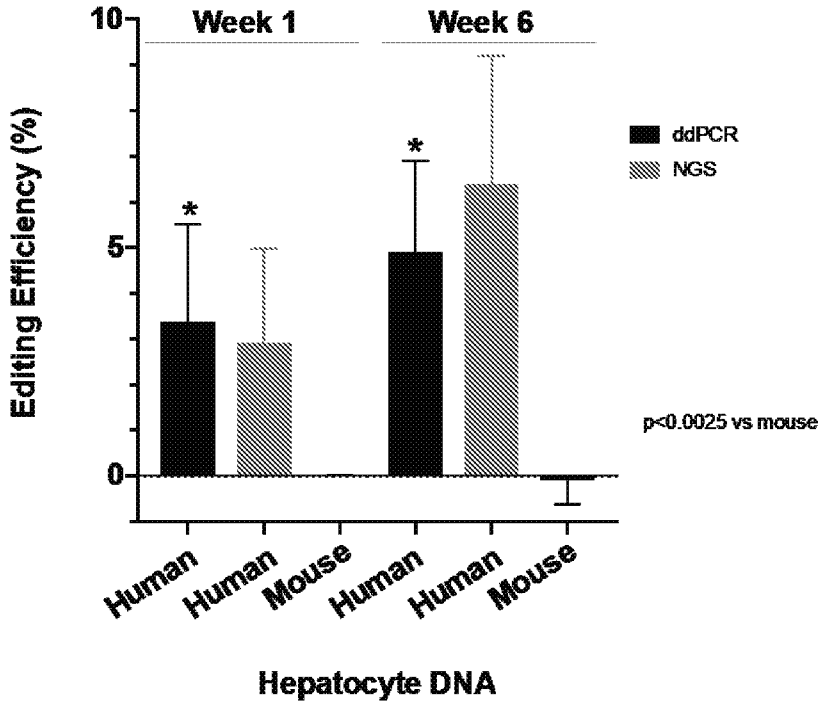


Figure 14

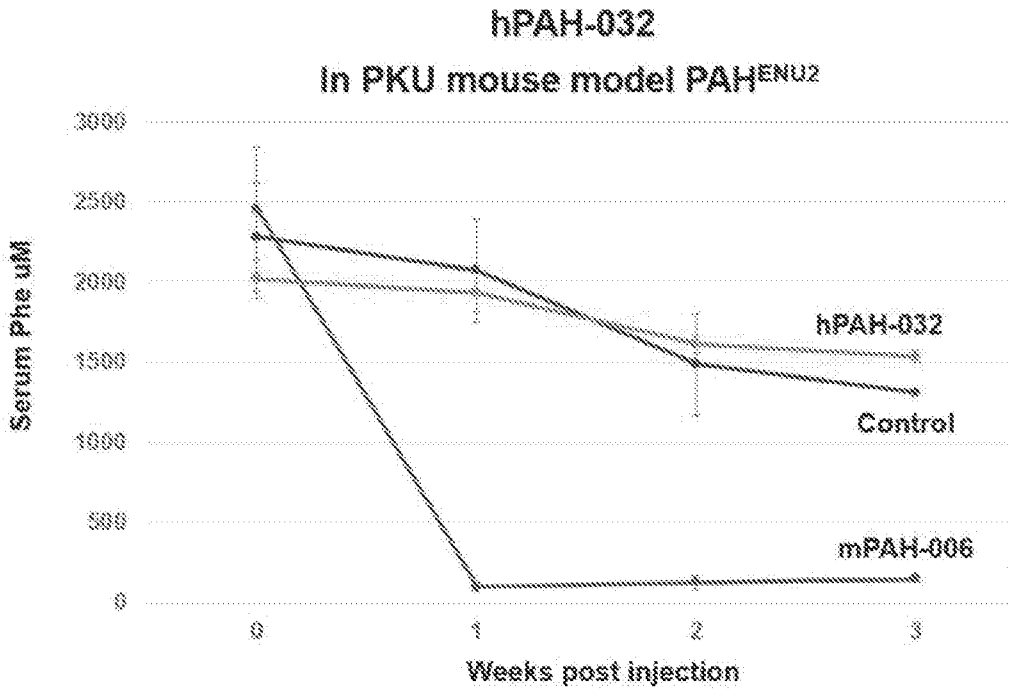


Figure 15A

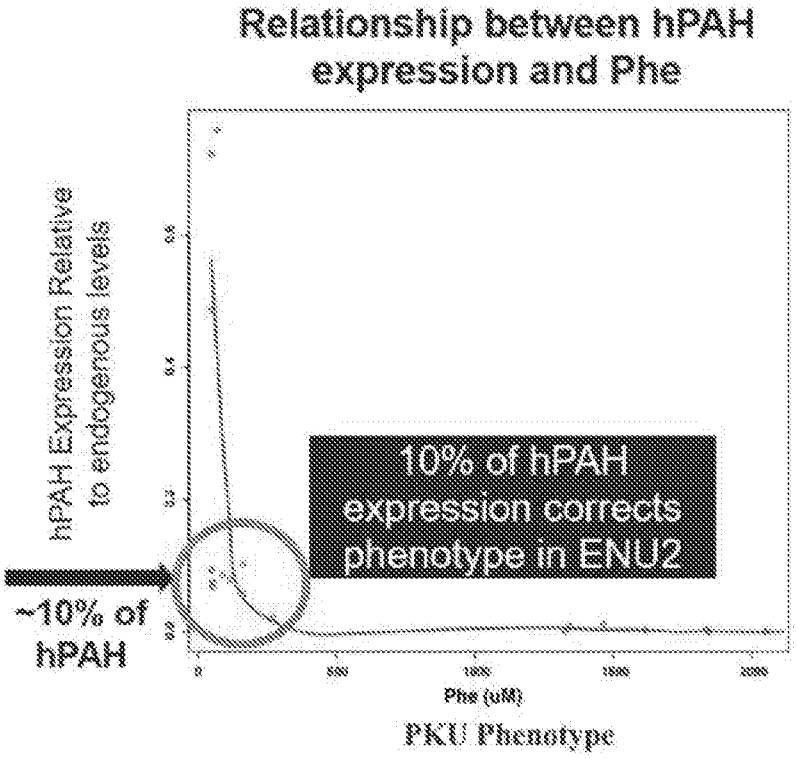


Figure 15B

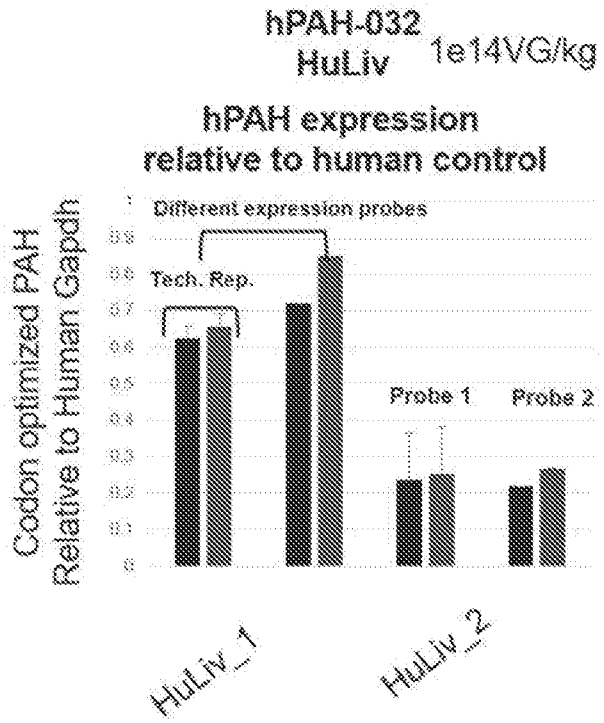


Figure 16

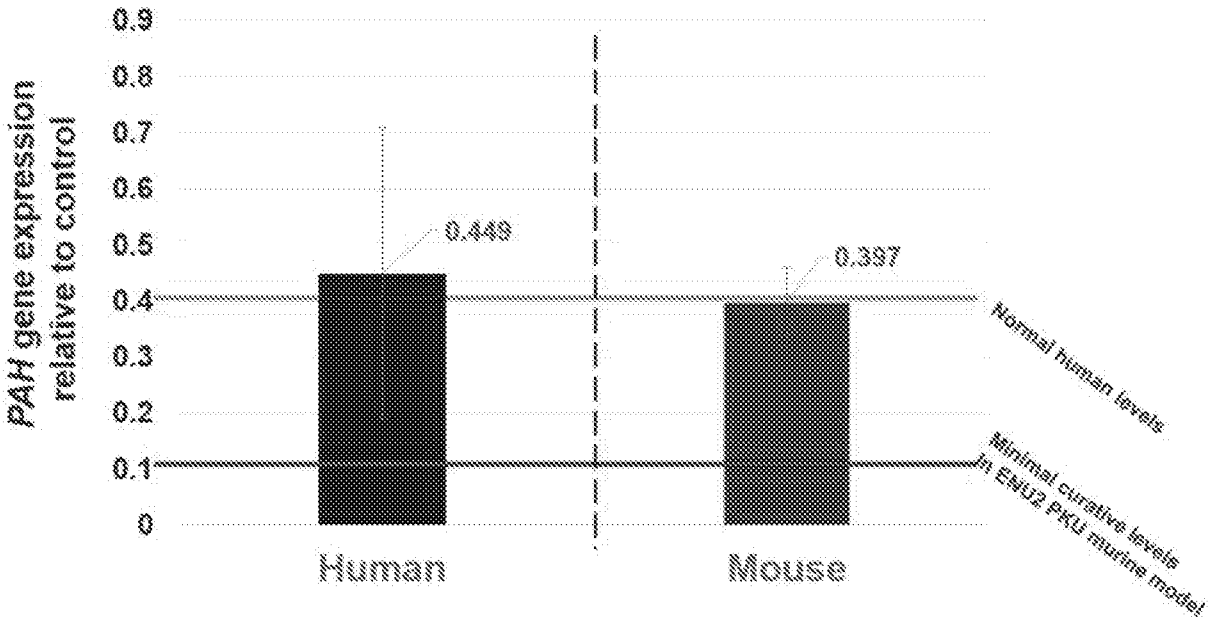

PAH gene expression relative to control
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
0.449
0.397
Human
Mouse
Normal human levels
Minimal curative levels in ENU2 PKU murine model

Figure 17A

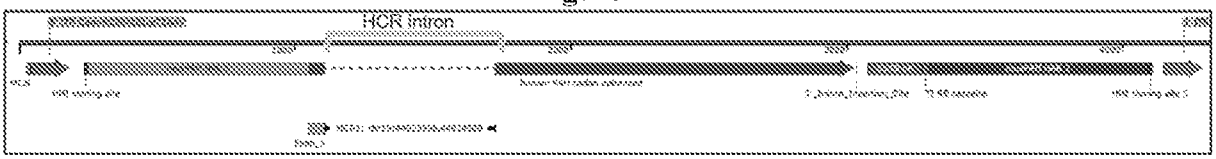


Figure 17B

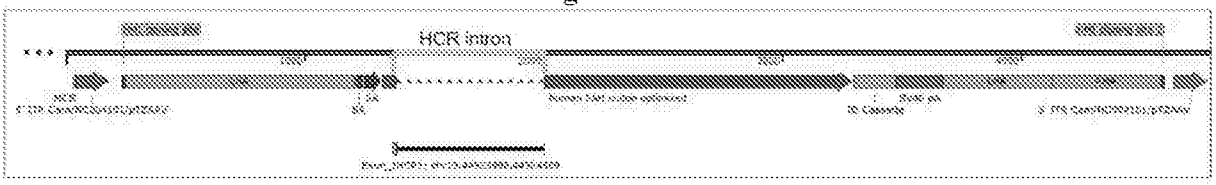


Figure 17C

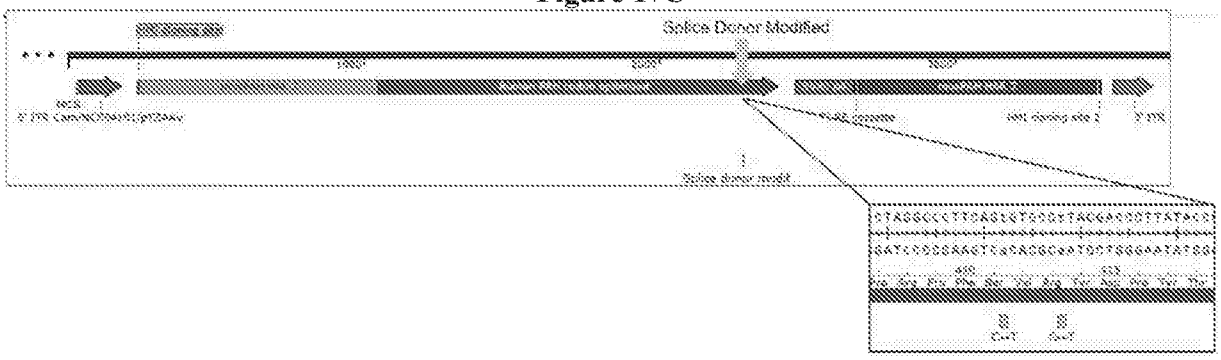


Figure 17D

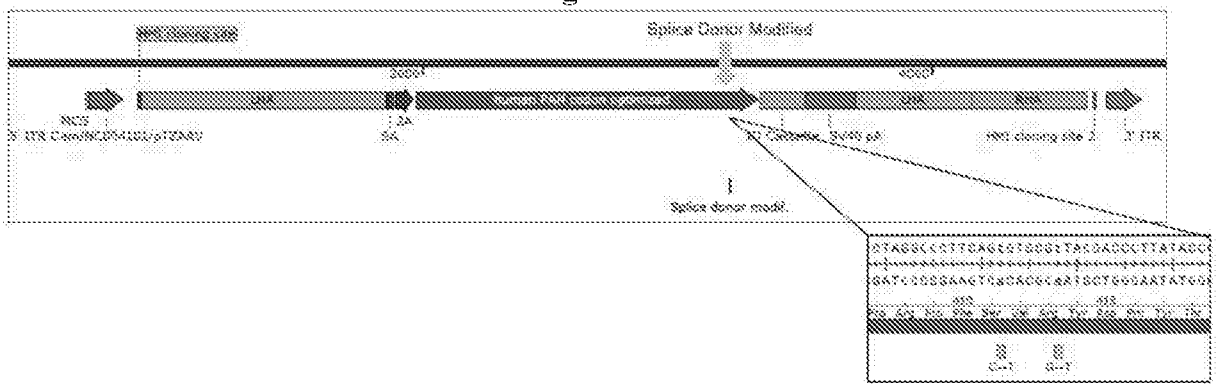


Figure 17E

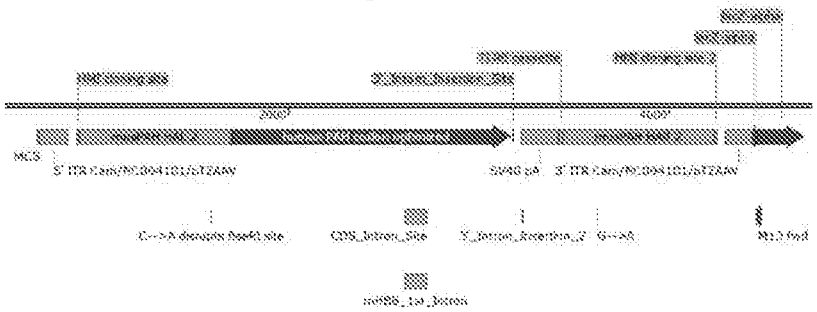
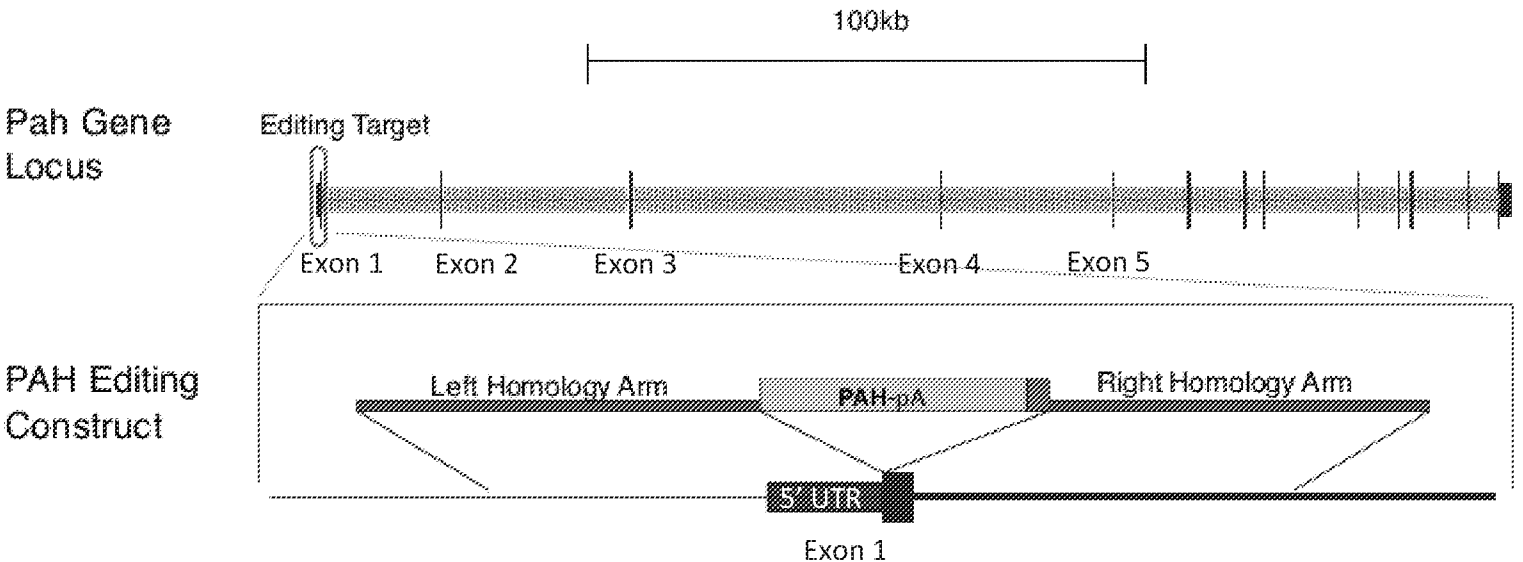


Figure 4A


100kb
Pah Gene Locus
Editing Target
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
PAH Editing Construct
Left Homology Arm
PAH-pA
Right Homology Arm
5' UTR
Exon 1