



(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).

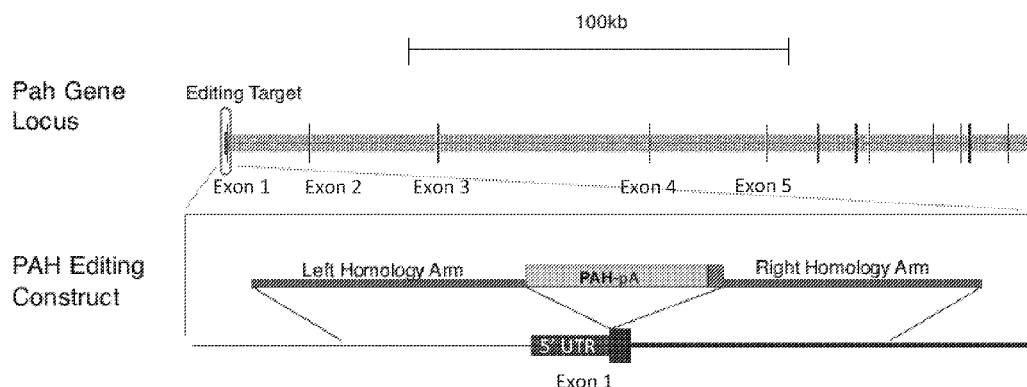
(72) Inventors: **SEYMOUR, Albert, Barnes;** 1 BrickyardLane, Westborough, MA 01581 (US). **AHMED, Seemin,****Seher;** 81 Park Avenue, Apartment 10, Worcester, MA01605 (US). **WRIGHT, Jason, Boke;** 57 Whits EndRoad, Concord, MA 01742 (US). **DOLLIVE, Serena,****Nicole;** 19 Everett St., Unit 1, Waltham, MA 02453 (US).**MCSWIGGEN, James, Anthony;** 15 Beacon Street, UnitA, 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.



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

5 [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
10 (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
15 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
20 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
25 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
30 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.

20

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.

25 [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
- 30 (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.

5 [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:

- 10 (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.

20 [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
 25 (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.

30 [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:

- 5 (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;
 - 10 (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 -
 - 15 432 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
 - 20 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

25 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

30 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.

5 [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
10 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
15 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
20 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
25 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,

30 (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:

5 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:

10 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: 20 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
25 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
30 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.
- 10 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
- 15 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
- 20 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
- 25 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.
- 30 [0043] Figure 1B is a map of the pHMI-hPAH-hIC-007 vector.
- [0044] Figure 1C is a map of the pHMI-hPAH-hIC-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. 5×10^5 HEK 293 cells were transfected with $1 \mu\text{g}$ 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.

5 [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).

10 [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”).

15 [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 2×10^{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.

20 [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 μ L 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.

30 [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 2×10^{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.

10 [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.

15 [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.

20 [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.

25 [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-hIIC-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-hIIC-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-hIIC-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-hIIC-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-hIIC-032-HCR, pKITR-hPAH-mAC-006-SD.3, pHMIA2-hPAH-hIIC-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
5 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
10 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,
15 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
20 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
25 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.
30 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.

5 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
10 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.

15 [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
20 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
25 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
30 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.

5 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
10 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
15 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
20 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.

25 [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%,
30 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
 5 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%,
 10 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
 15 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
 20 203-736 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
 25 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%,
 30 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-
5 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
10 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
15 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
20 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
25 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)
30 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
5 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
10 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%,
15 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
20 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
25 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
30 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).

5 [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.

10 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

15 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

20 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-

25 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

30 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
5 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
10 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
15 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.

20 [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
25 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
30 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.

5 [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.

10 [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
15 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
20 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
25 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
30 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).

5 [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
10 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
15 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
20 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
25 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
30 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
5 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
10 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
15 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
20 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'
25 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
30 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%, or 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.

5 **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
10 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-
15 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
20 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.

25 [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
30 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
 5 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:
 10 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
 15 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
 20 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
 25 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.

30 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 E1, 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
 5 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
 10 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.,
 15 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

20 *b) PAH correction vector pHMI-hPAH-hIC-007*

[00181] PAH correction vector pHMI-hPAH-hIC-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-hIIC-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-hIIC-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-hIIC-032.1

[00184] PAH correction vector pHMIA-hPAH-hIIC-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-hIIC-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-hIIC-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-hIIC-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-hIIC-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-hIIC-032.1

[00185] Eight variants of the pHMIA-hPAH-hIIC-032.1 vector have been designed to improve the expression of the PAH gene locus. These variants, named pHMIA-hPAH-hIIC-032.2 to pHMIA-hPAH-hIIC-032.9, differ from pHMIA-hPAH-hIIC-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-hIIC-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-hIIC-032.2	37	47	56
pHMIA-hPAH-hIIC-032.3	38	48	57
pHMIA-hPAH-hIIC-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-hIIC-032.5	40	50	59
pHMIA-hPAH-hIIC-032.6	41	51	60
pHMIA-hPAH-hIIC-032.7	42	52	61
pHMIA-hPAH-hIIC-032.8	43	53	62
pHMIA-hPAH-hIIC-032.9	44	54	63

[00186] The pHMIA-hPAH-hIIC-032.2 vector was designed to optimize the Kozak sequence for improved ribosome recruitment to the transcript. It differs from pHMIA-hPAH-hIIC-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-hIIC-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-hIIC-032.1 in having the nucleotides A and A at positions -467 and -465, respectively, of the PAH gene.

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

[00189] The pHMIA-hPAH-hIIC-032.5 vector was designed to optimize two cyclic AMP response elements to increase expression. It differs from pHMIA-hPAH-hIIC-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-hIIC-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-hIIC-032.1 in having the nucleotide G at position 194 of the PAH gene.

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

[00192] The pHMIA-hPAH-hIIC-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-hIIC-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-hIIC-032.9 vector was designed to disrupt three G-quadruplexes in the 5' UTR that might suppress expression. It differs from pHMIA-hPAH-hIIC-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-hI-001, a variant of pHMI-hPAH-hIC-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 ccaaateccaccagctcact (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-hI-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-hI-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

I 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 μ L 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	GCAATAGCATCACAAATTTTCAC	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	ACAGCCTATATTTACCATGCTGATCCC	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.

5 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

10 [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
15 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
20 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
25 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 2×10^{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 μ L 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.

5 **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	GCAATAGCATCACAAATTTTAC	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

5 [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
10 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 determined by ddPCR using primer and probe sets to measure the vector,
15 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
20 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)
25 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
30 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).

- 5 [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
- 10 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%.
- 15 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%

- 20 [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.

5 *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
 10 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
 15 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
 20 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
 25 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).

5 **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

10 [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
15 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
20 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
25 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.

30 [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 packaged 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-h11C-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-hIIC-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-hIIC-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-hIIC-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-hIIC-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-hIIC-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-hIIC-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-hIIC-032 vector. Human PAH expression in human hepatocytes was measured relative to human GAPDH in HuLiv mice administered the pHMIK-hPAH-hIIC-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-hIIC-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 pKTR-hPAH-mAC-006-HCR, pKTR-hPAH-hIIC-032-HCR, pKTR-hPAH-mAC-006-SD.3, pHMIA2-hPAH-hIIC-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) *pKlTR-hPAH-mAC-006-HCR*, *pKlTR-hPAH-hIIC-032-HCR*, and *pHMIA2-hPAH-mAC-006-HBB1*

[00239] Vectors *pKlTR-hPAH-mAC-006-HCR* and *pKlTR-hPAH-hIIC-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 (AdV/IgG)	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] *pKlTR-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. *pKlTR-hPAH-hIIC-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.

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

Genetic Element	SEQ ID NO		
	-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-hIIC-032-SD.3* Vectors pKITR-hPAH-mAC-006-SD.3 and pHMIA2-hPAH-hIIC-032-SD.3 were generated by
 10 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.
 15 pHMIA2-hPAH-hIIC-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-hIIC-032-SD.3

Genetic Element	SEQ ID NO	
	-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

5

* * *

[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
5 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.
- 10 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.
- 15 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
20 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.
- 25 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.
- 30 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.
- 5 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.
- 10 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.
- 15 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
20 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
25 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
30 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.

5

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.

10

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.

15

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

20

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.

25

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.

30

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.
- 5 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.
- 10 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.
- 15 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.
- 20 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.
- 25 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
- 30 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.

5

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.

25 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.

30

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.
- 5 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.
- 10 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%
15 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%
20 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.
- 25 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.
- 30 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
5 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
10 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
15 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.

20

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;
- 25 (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;
- 30 (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

5 (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.

10 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.

15 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.

20 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:

5 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:

10 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: 20 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.

25

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.

30

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.

20

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;
- 25 (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;
- 30 (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.

5 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
10 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.

15

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..

20

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.

25

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.

30

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.
- 5 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
10 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.
- 15 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.
20
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
25 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.
- 30 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.

5 100. 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.

10

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
15 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.

20

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
25 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
30 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
5 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%
10 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%
15 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.

20

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
25 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
30 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.

30

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.

15

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;
- 20 (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;
- 25 (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;
- 30 (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:

5 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.

5 136. 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
10 encodes an AAV2 Rep protein.

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

15 139. 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.

20 140. 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.

25 142. 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
30 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).
- 5 146. 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.
- 10 148. 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
15 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.
- 20 151. 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.
- 25 153. 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
30 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

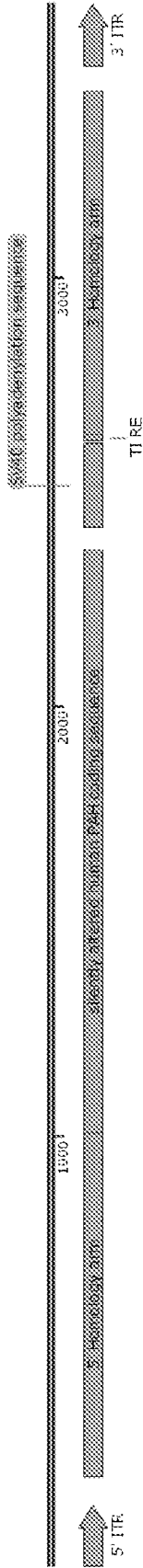


Figure 1B

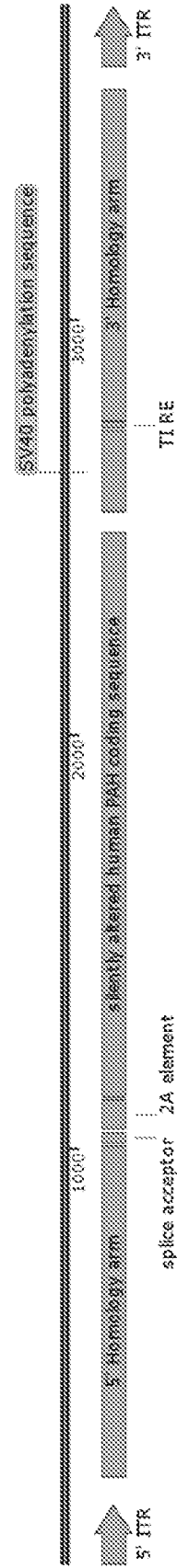


Figure 1C

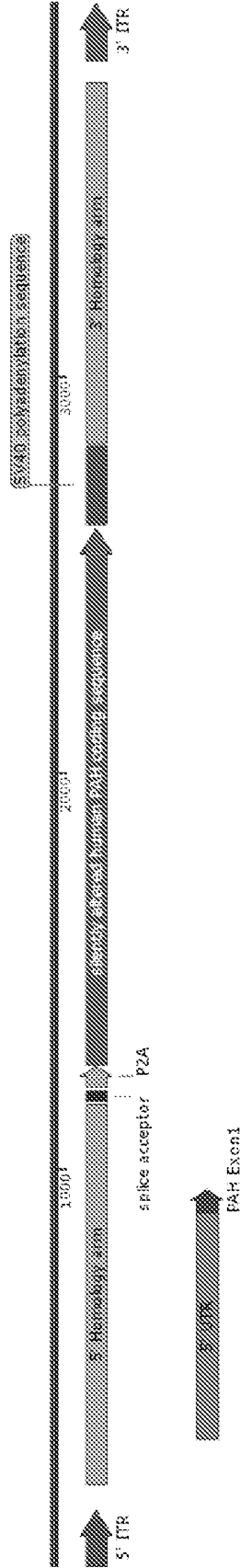
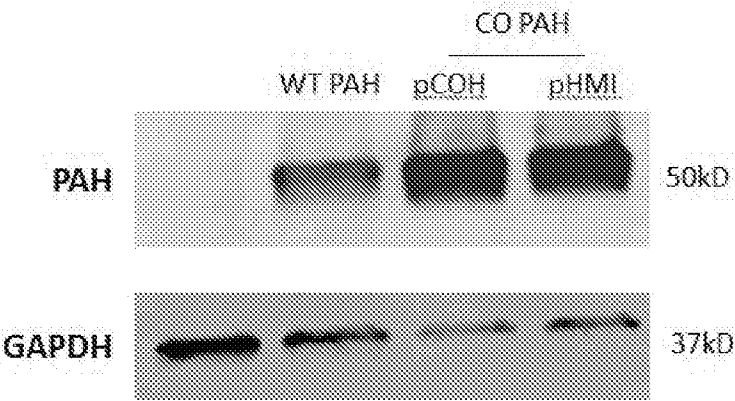


Figure 2



3/26

Figure 3A

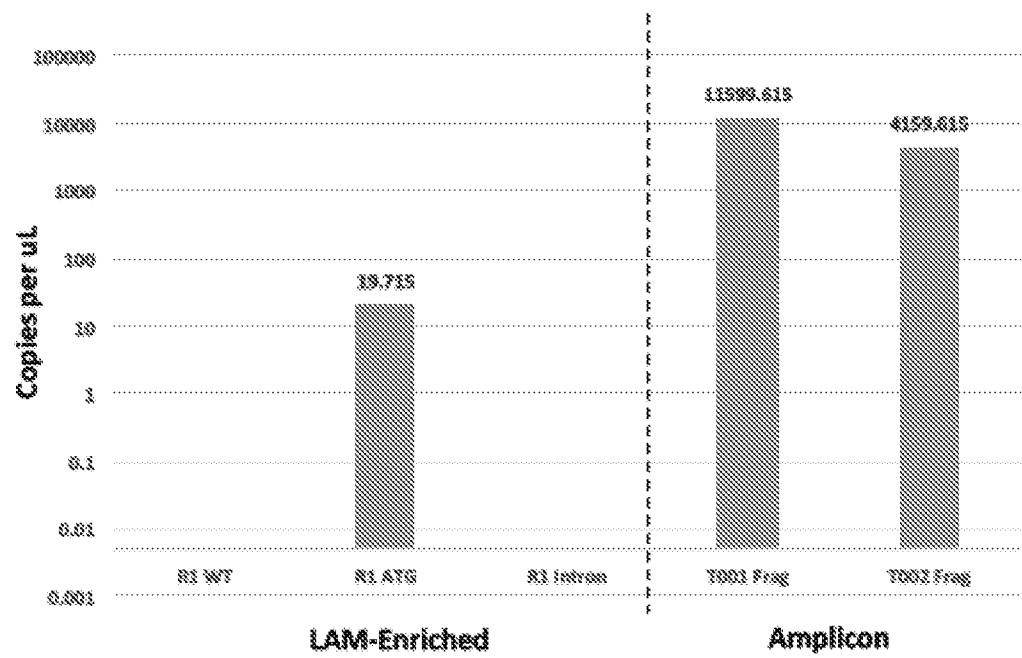


Figure 3B

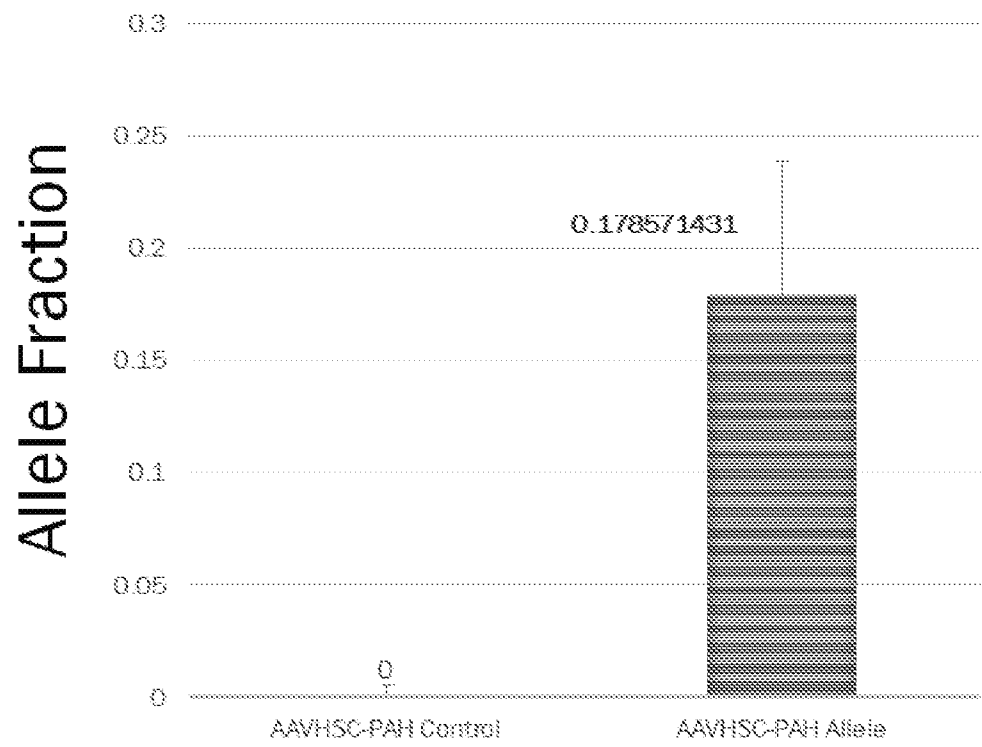
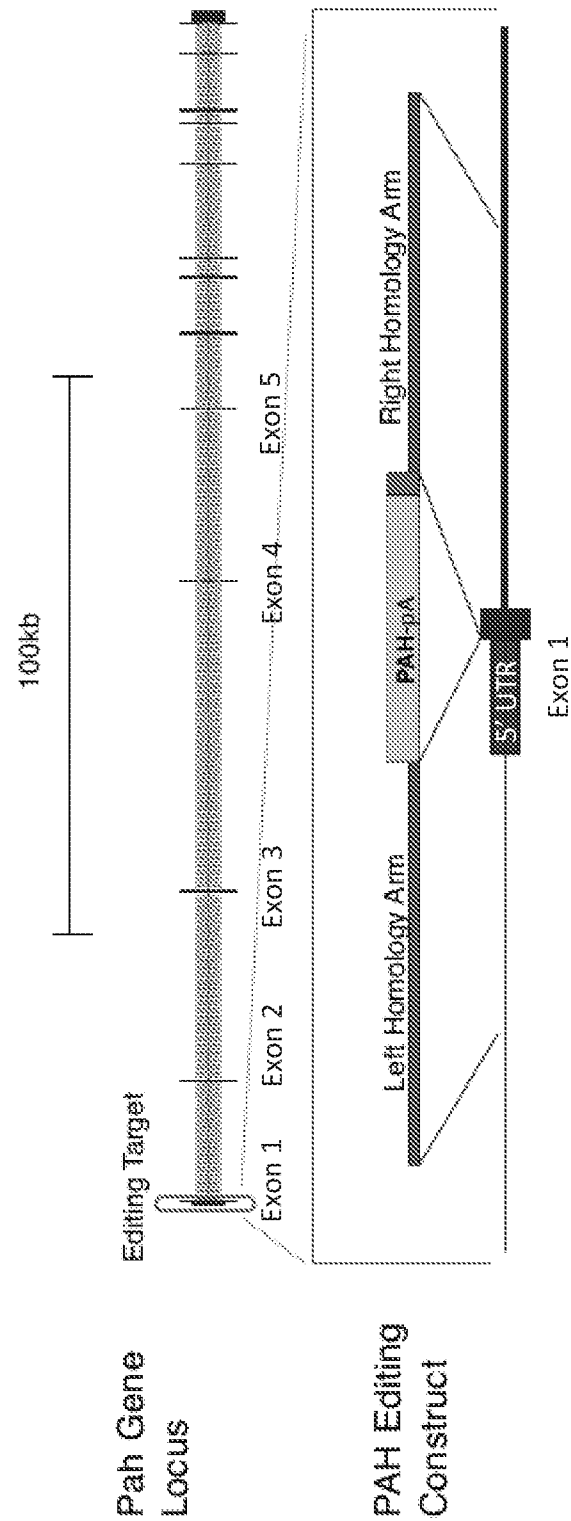


Figure 4A



5/26

Figure 4B

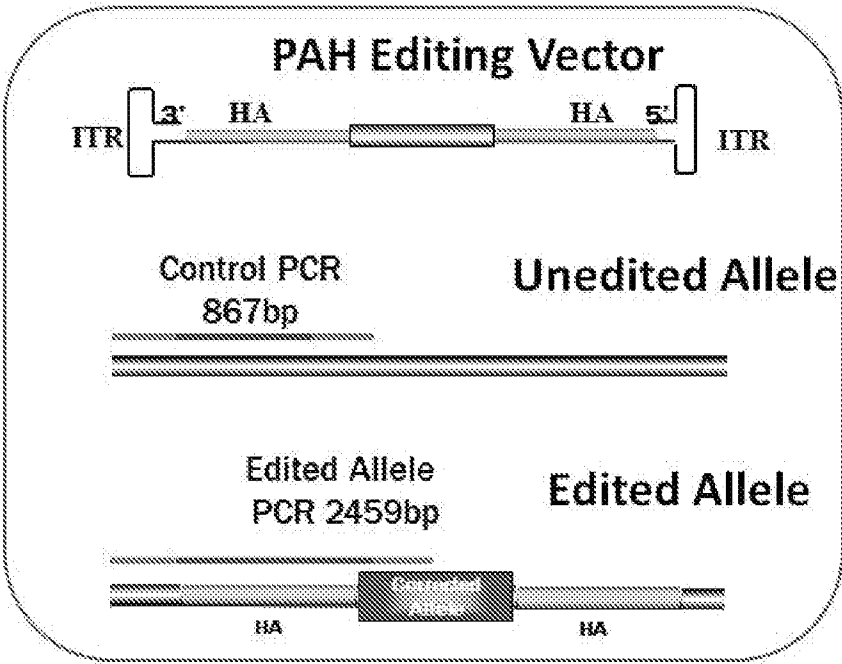
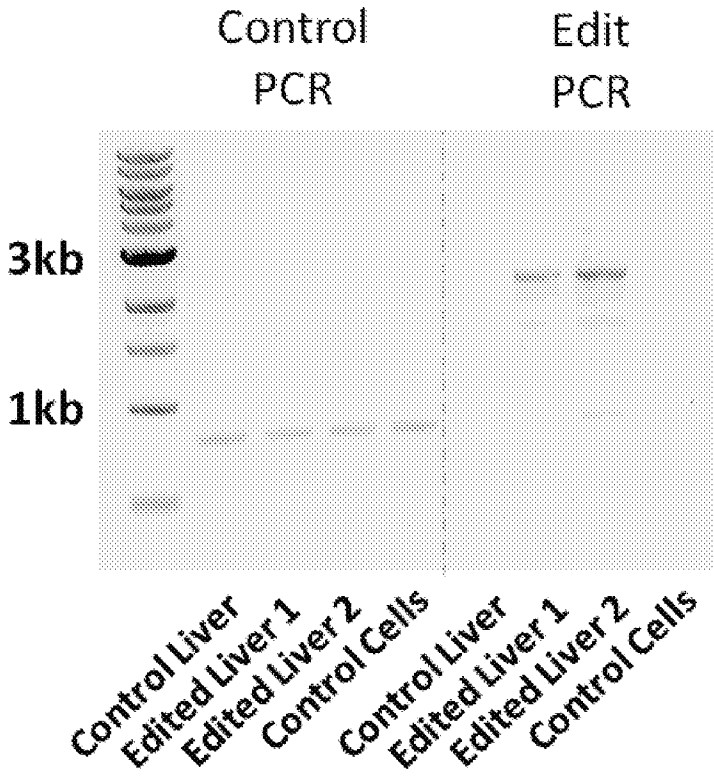


Figure 4C



6/26

Figure 5A

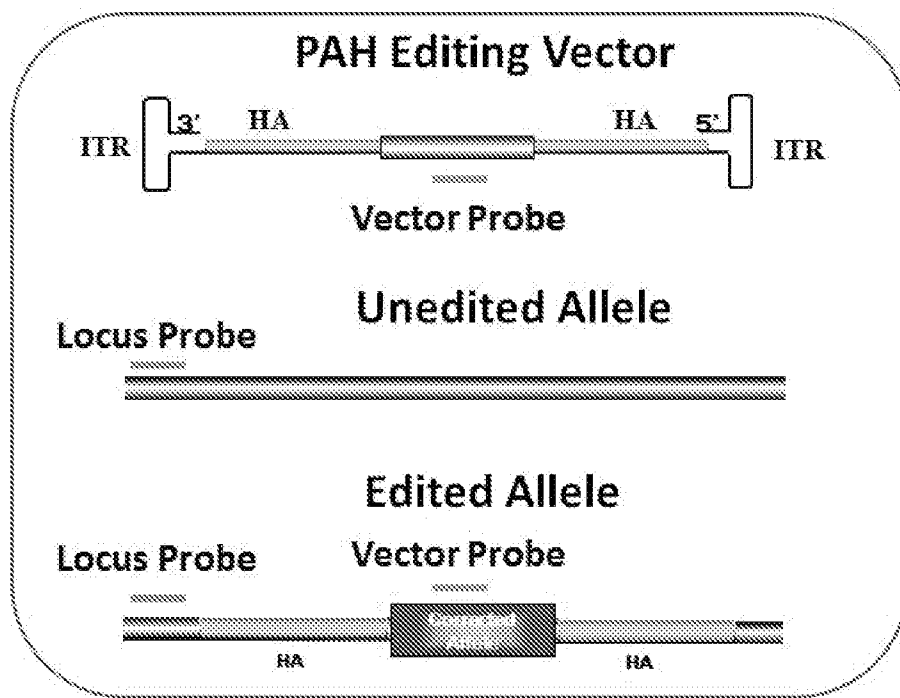
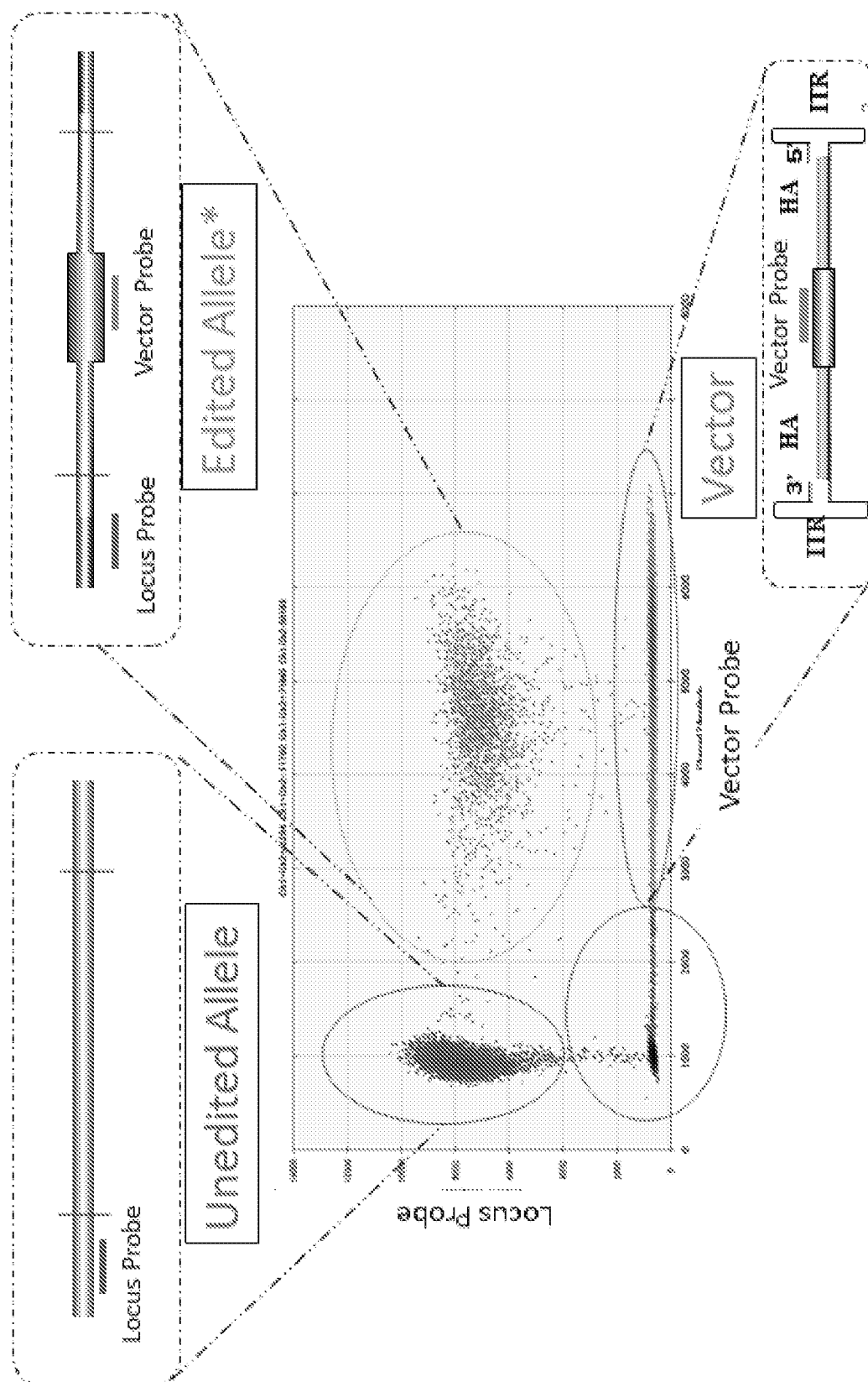


Figure 53



8/26

Figure 5C

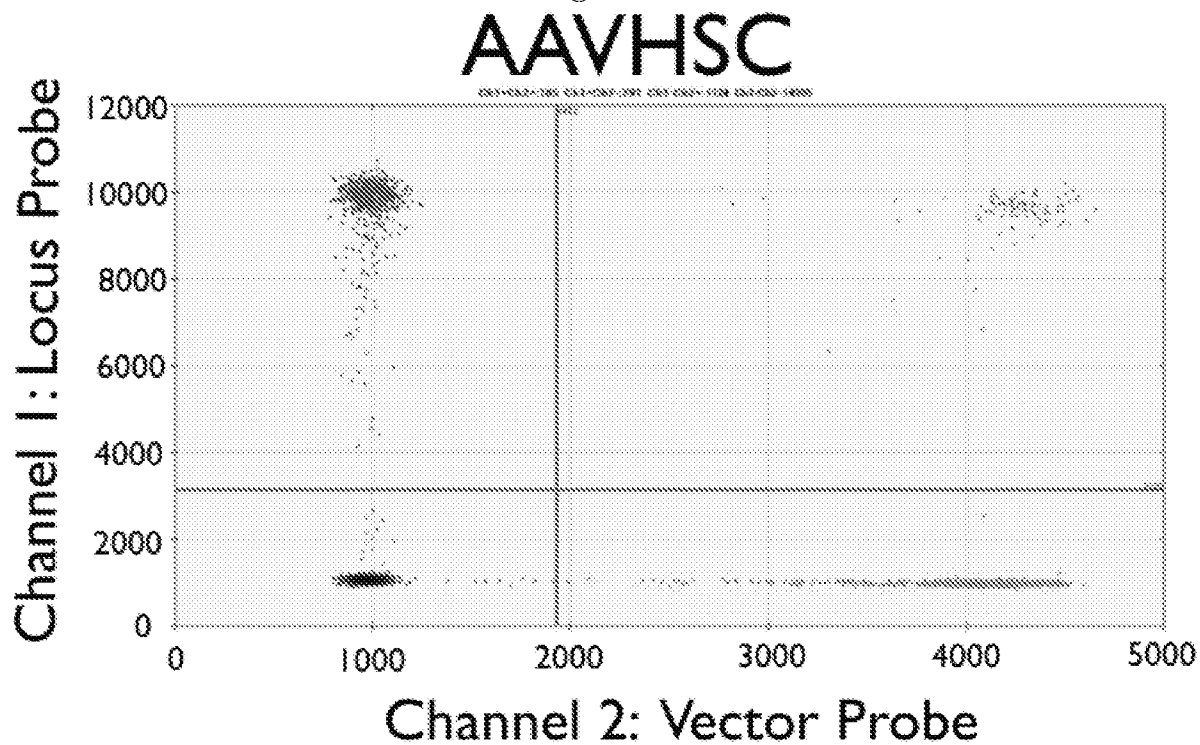
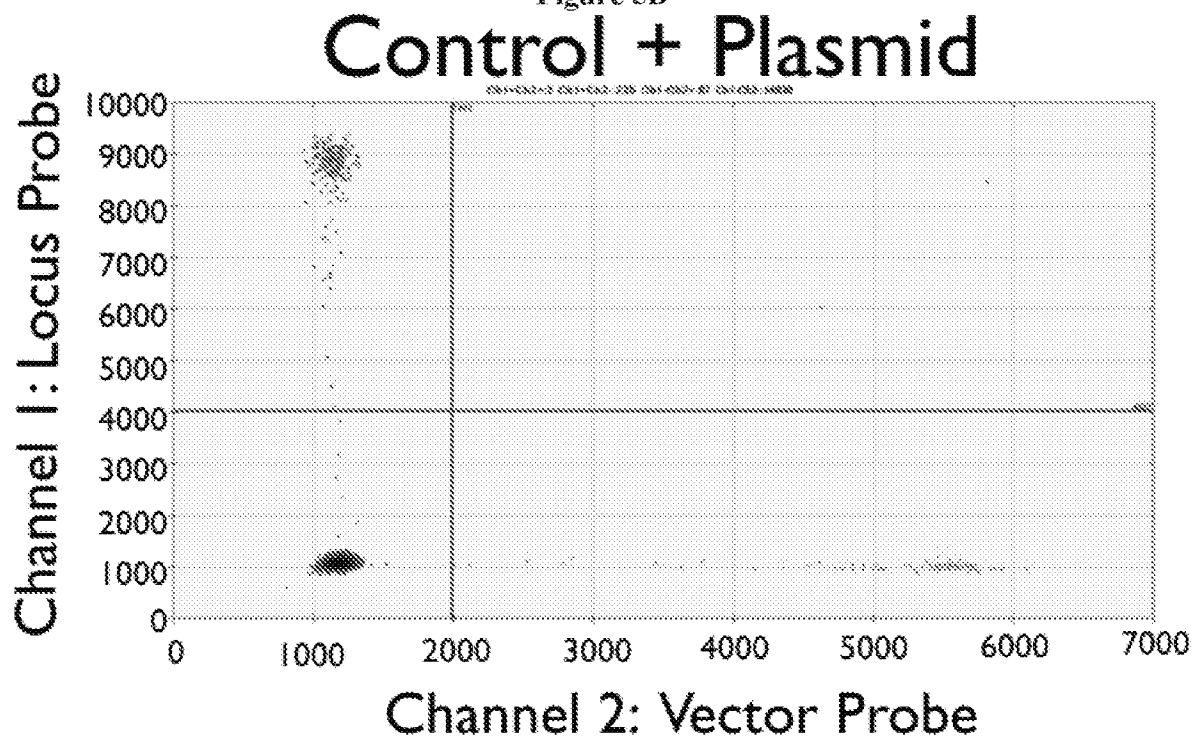
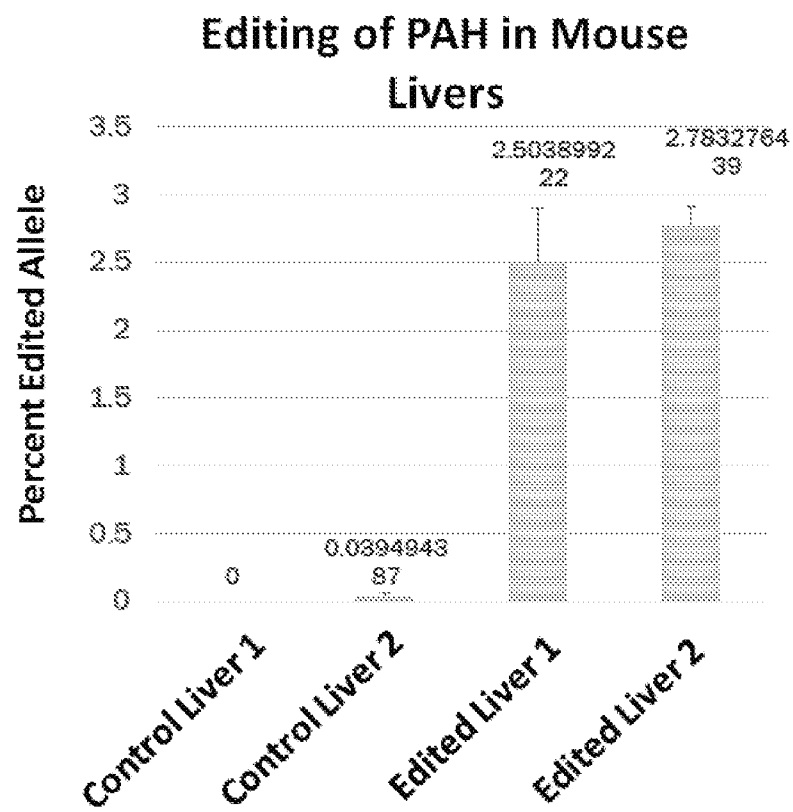


Figure 5D



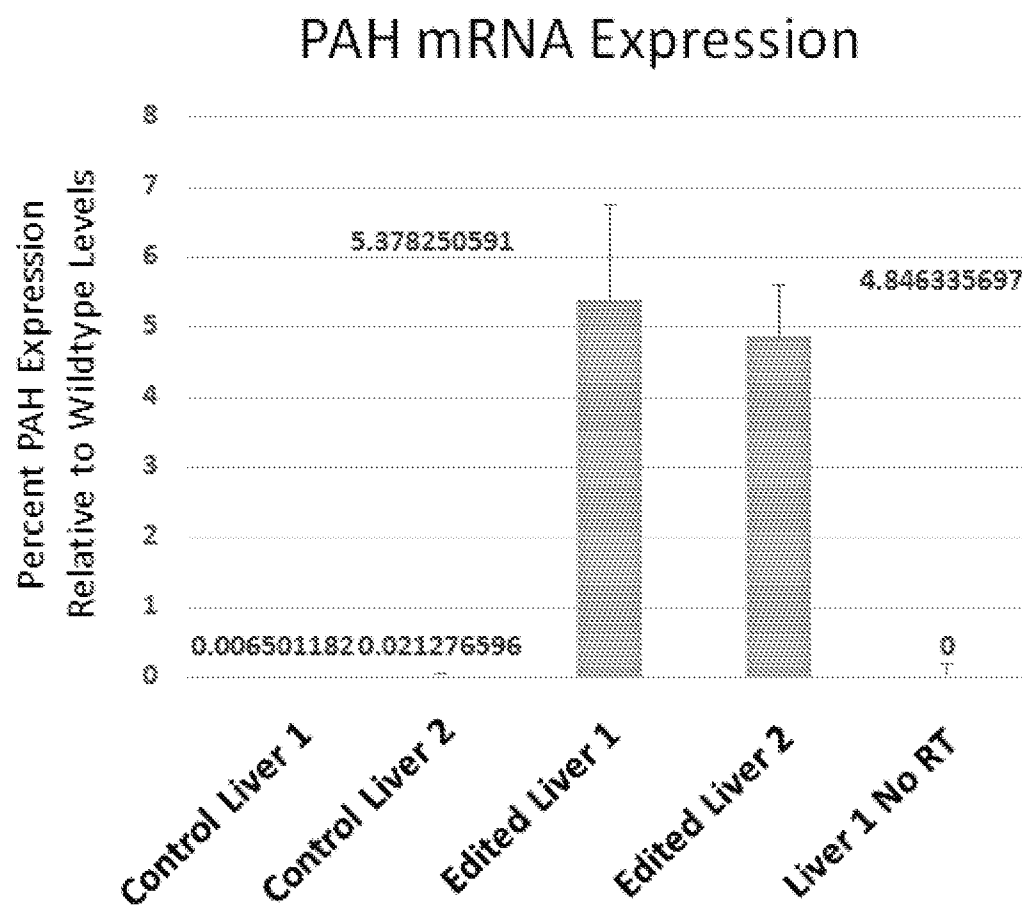
9/26

Figure 5E



10/26

Figure 6



11/26

Figure 7A

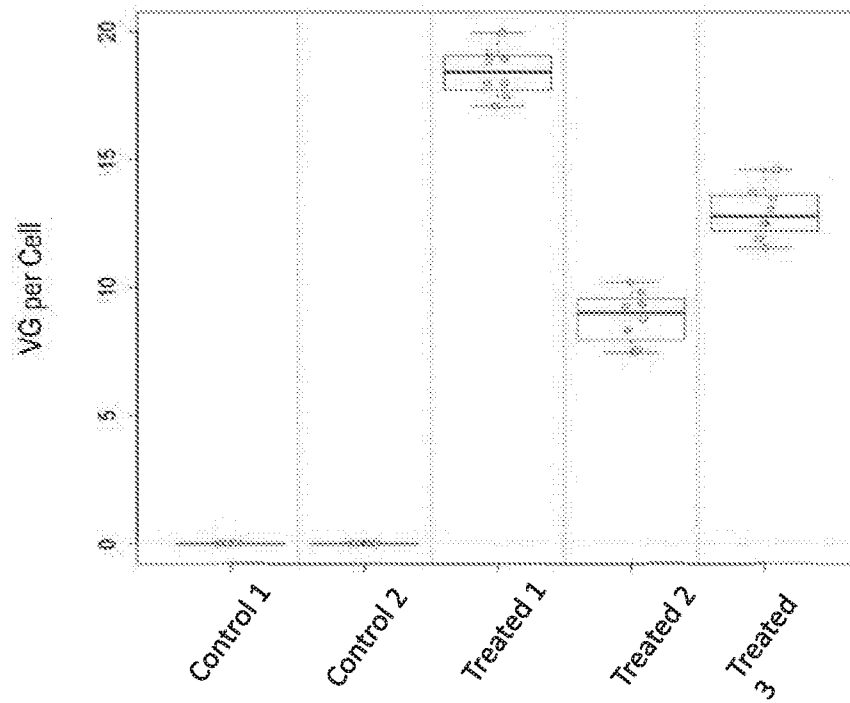
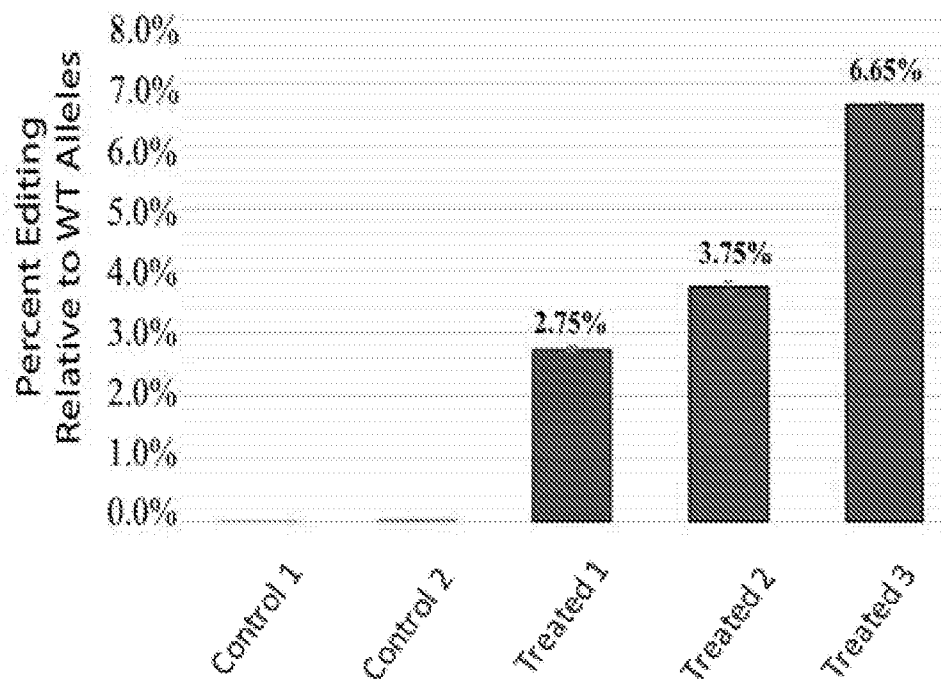


Figure 7B



12/26

Figure 7C

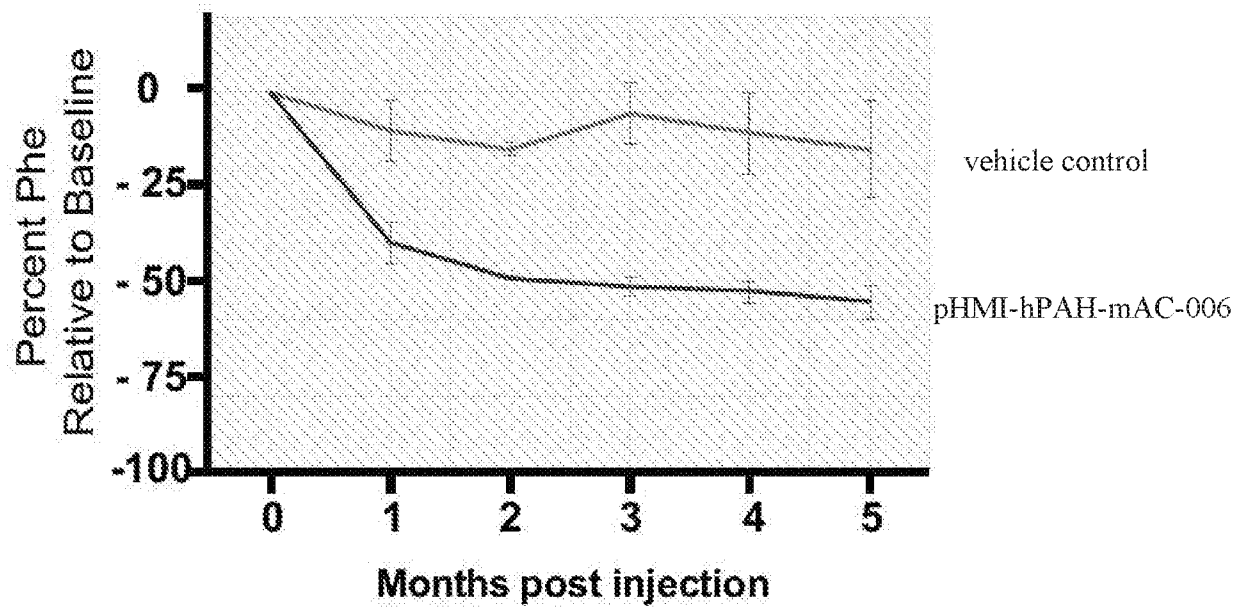
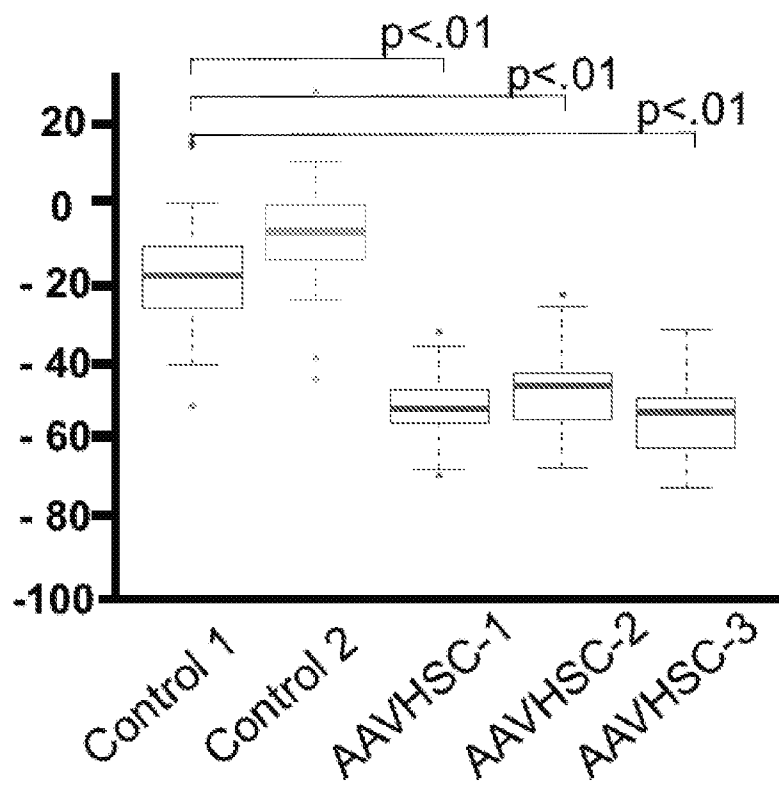
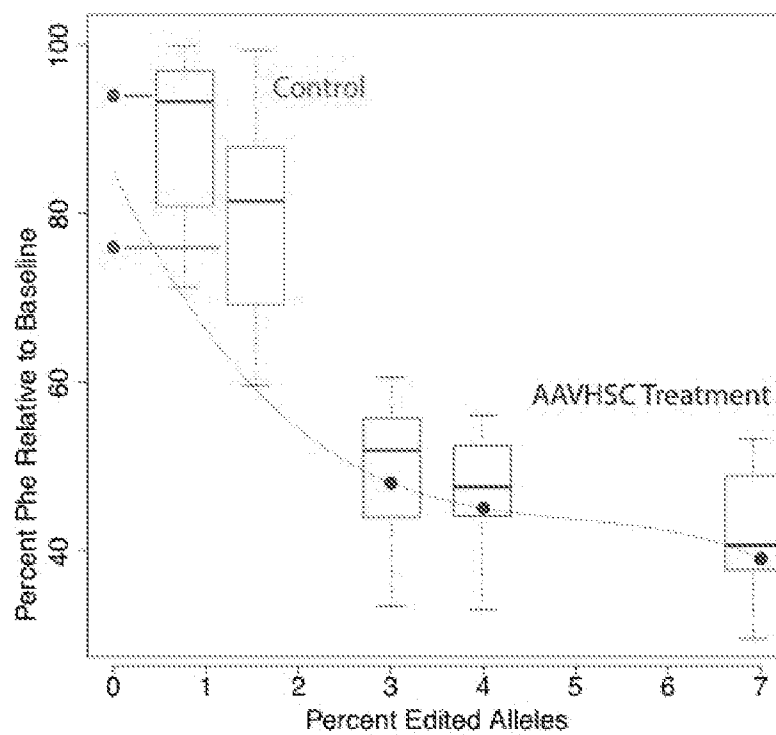


Figure 7D



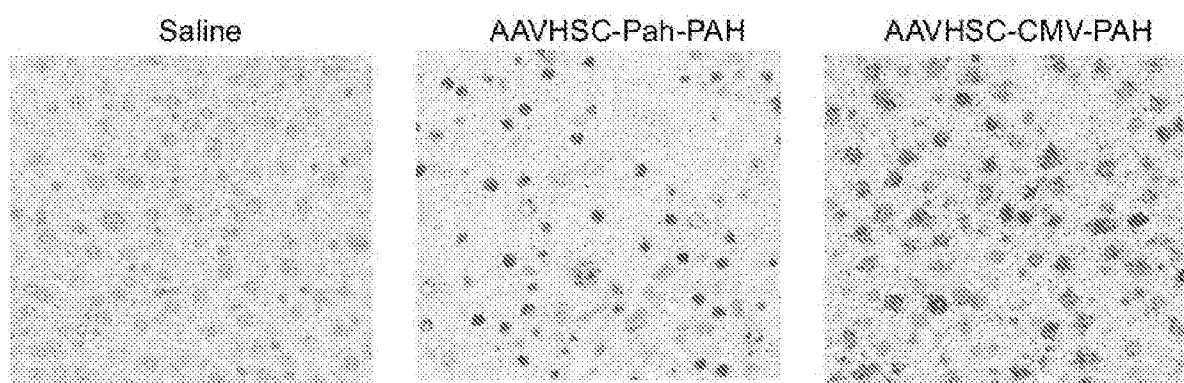
13/26

Figure 7E

Editing to Phenotype Relationship

14/26

Figure 7F



15/26

Figure 8A

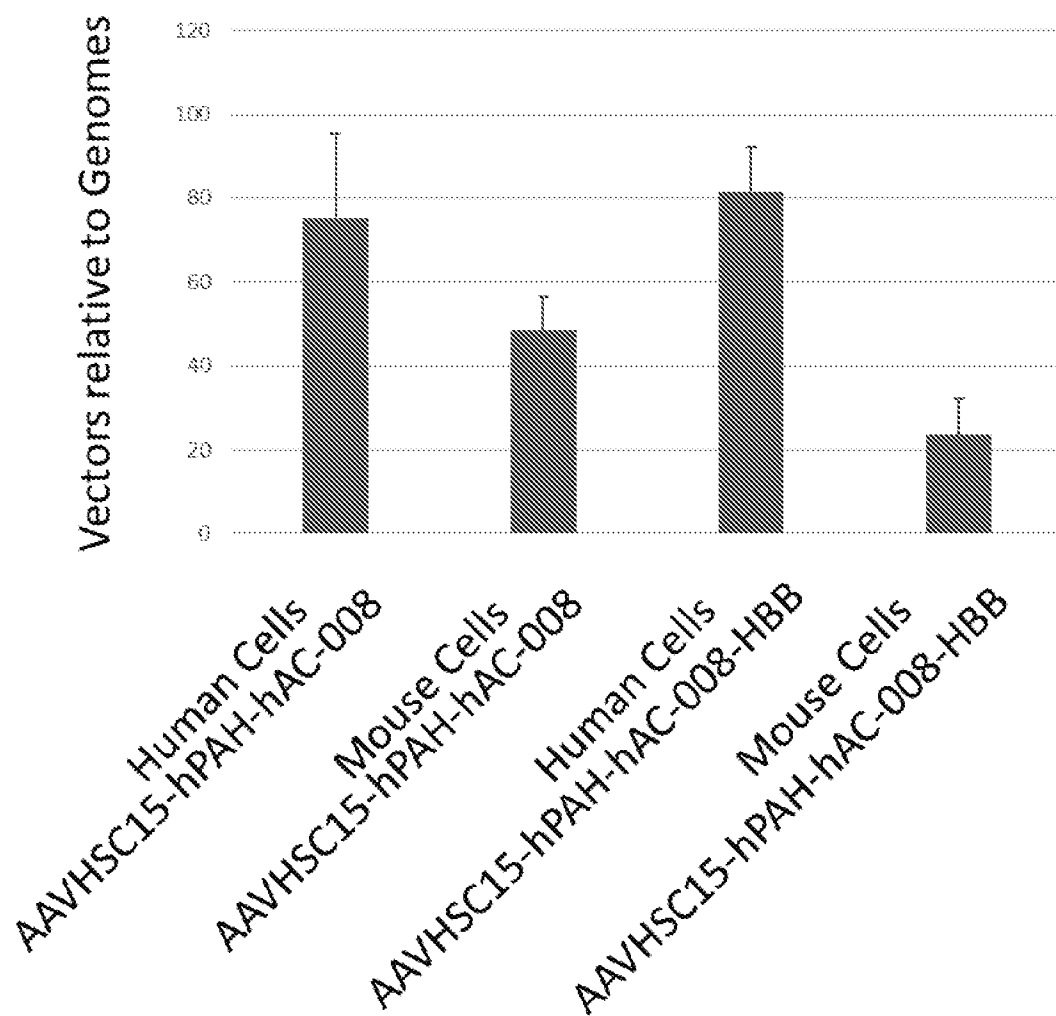
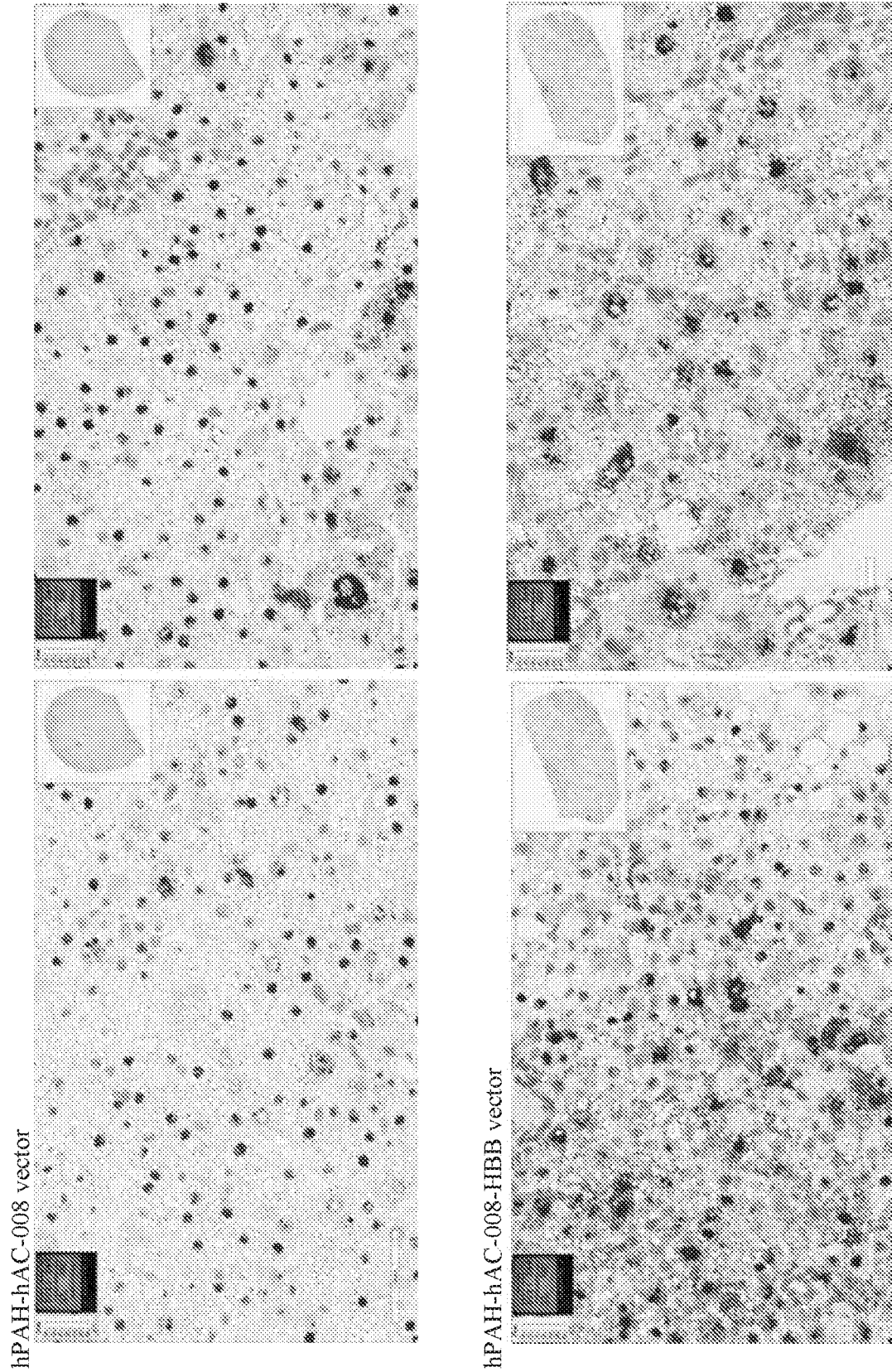
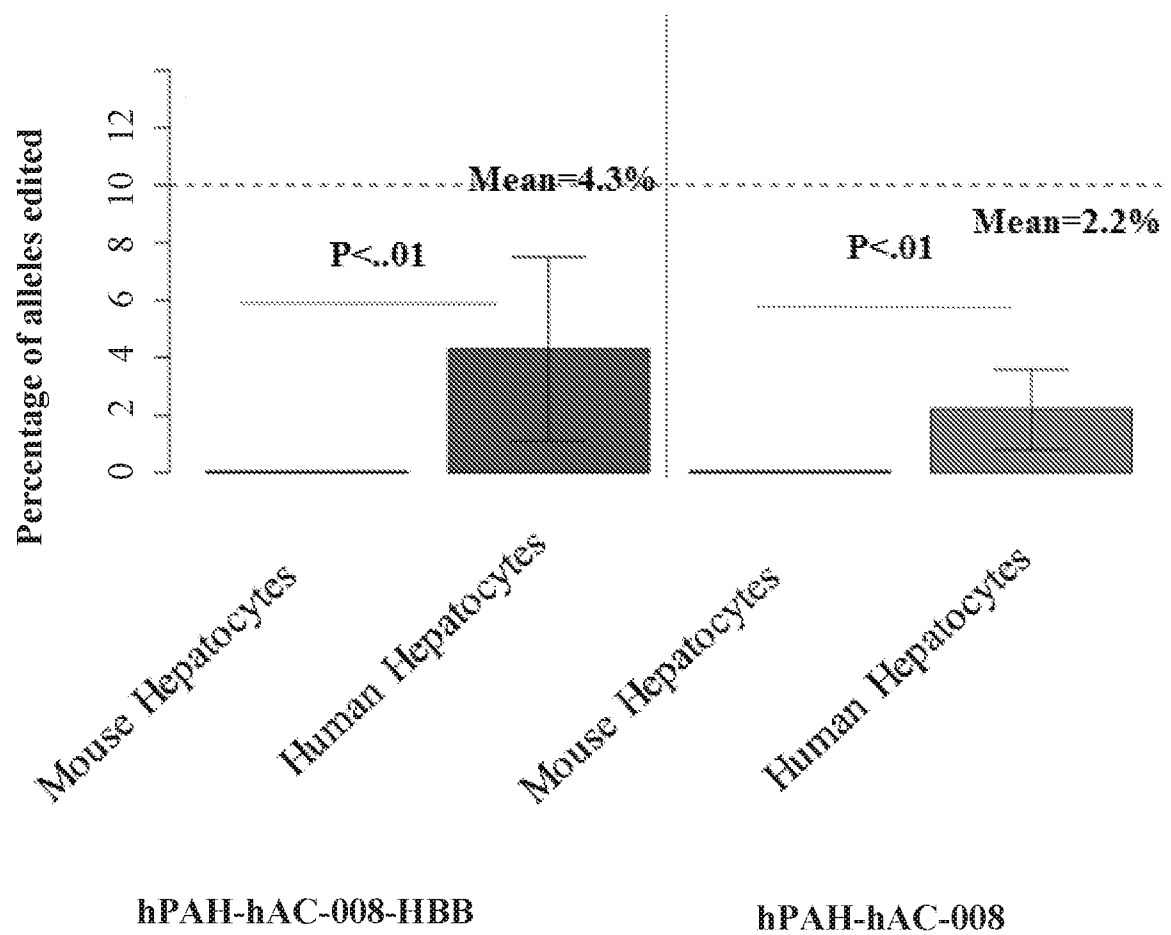
Vector per Genome by Species

Figure 8B



17/26

Figure 8C
Editing Efficiency



18/26

Figure 9A

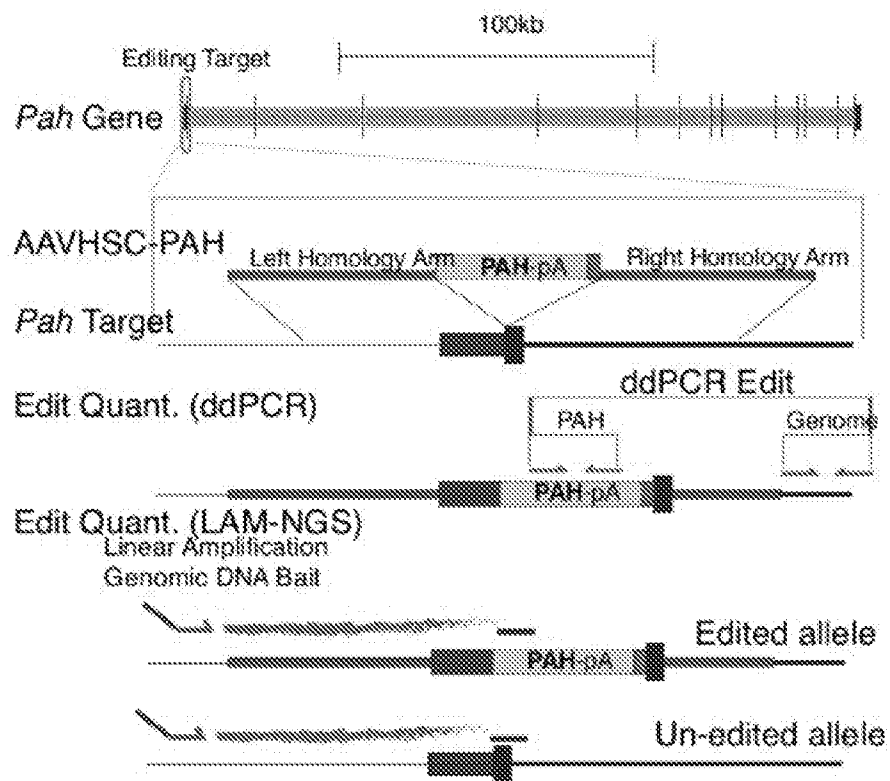
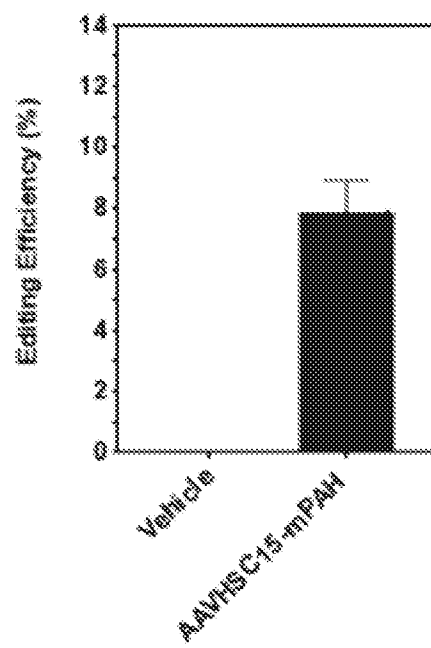


Figure 9B



19/26

Figure 10A

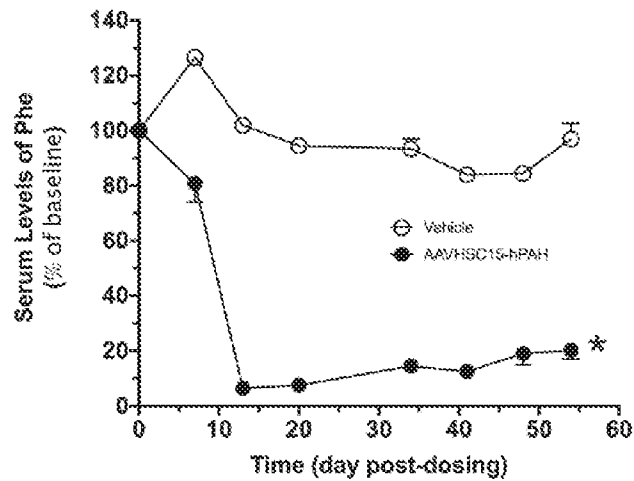


Figure 10B

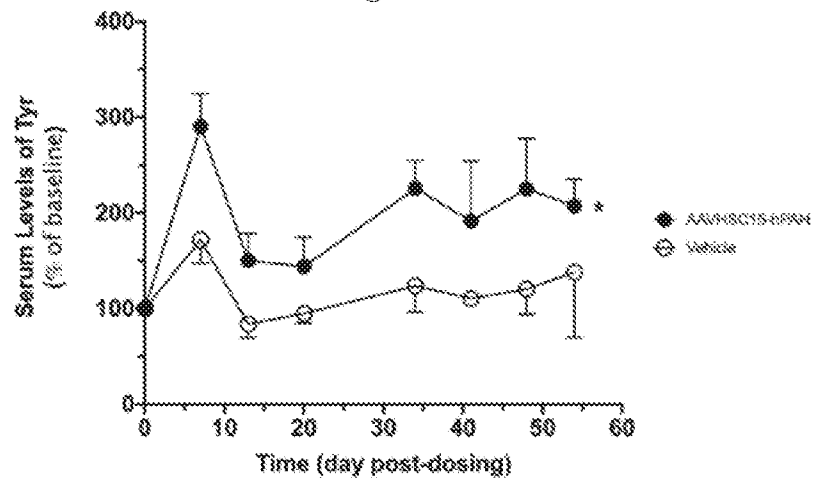
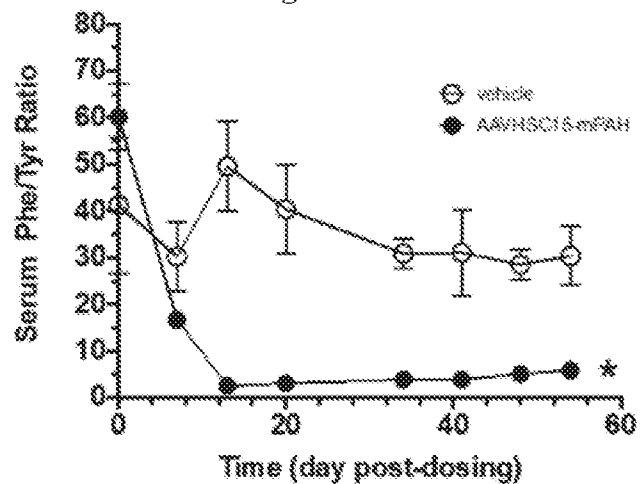
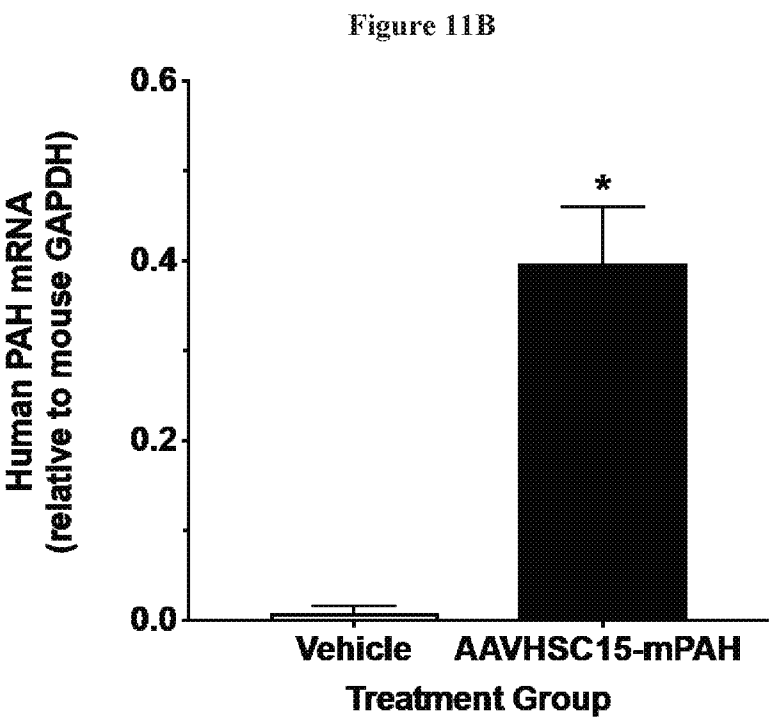
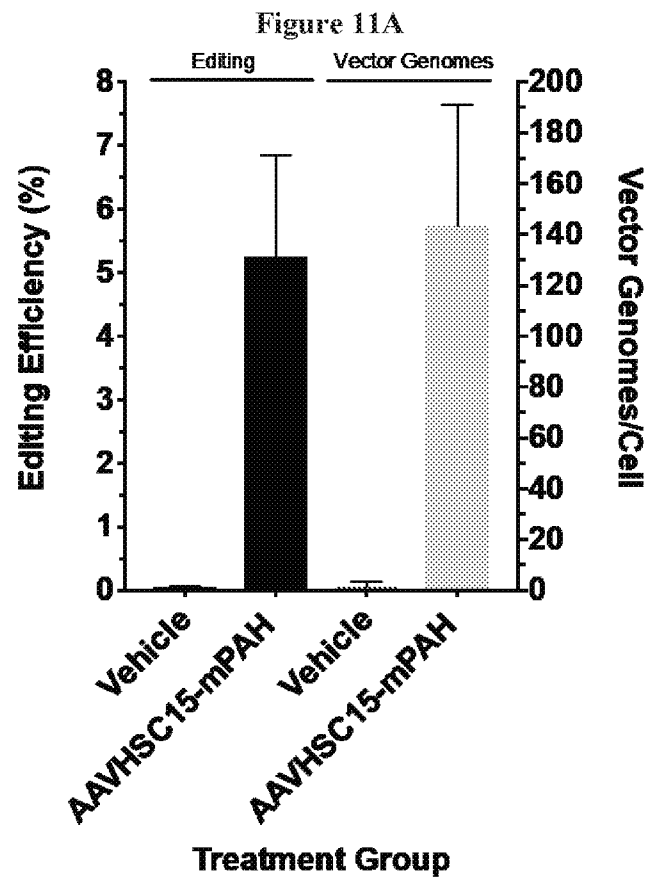


Figure 10C





21/26

Figure 12A

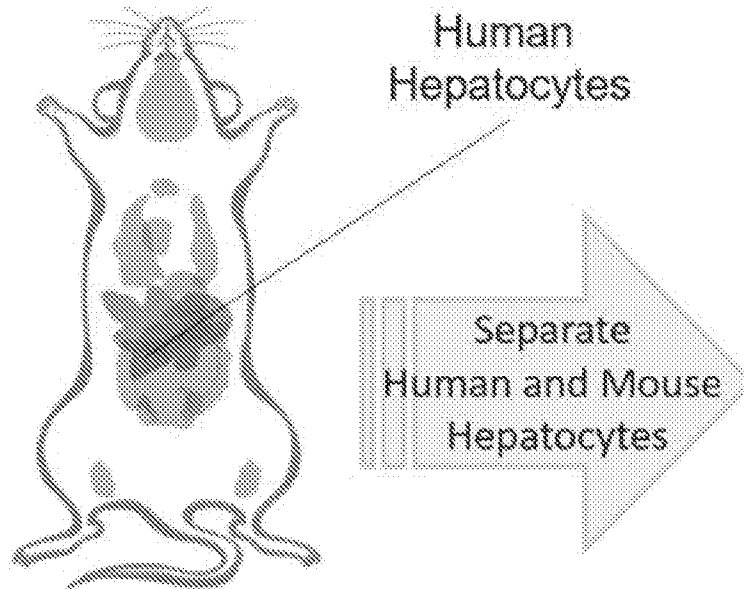
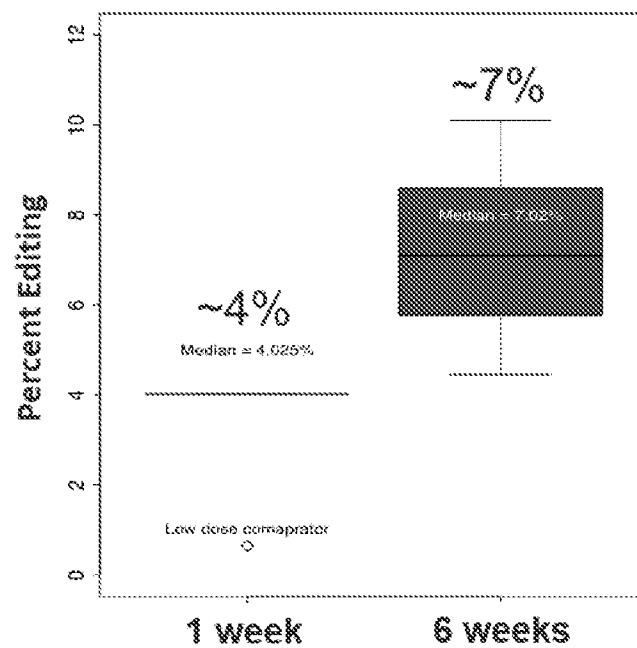


Figure 12B

hPAH-032 Editing durability



22/26

Figure 13

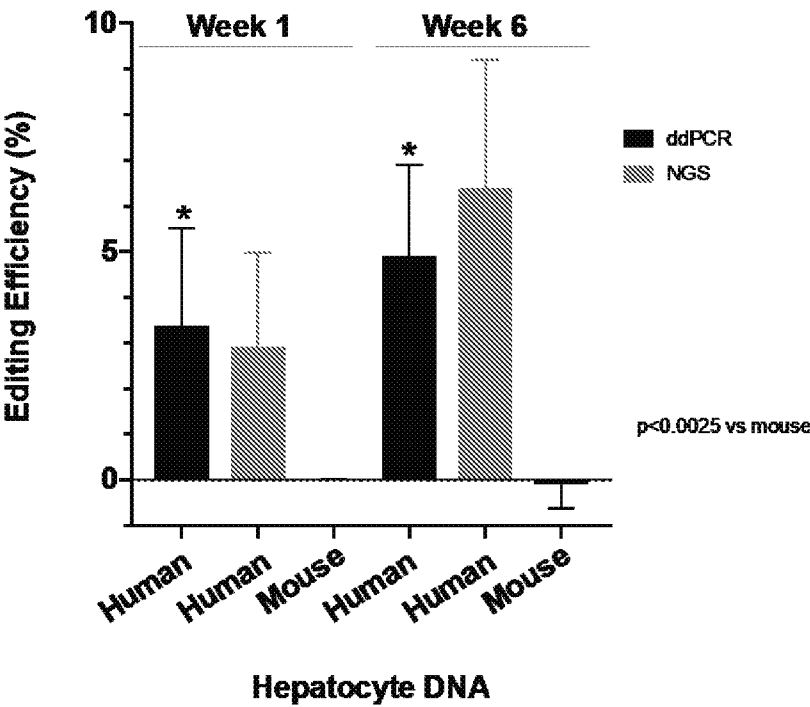
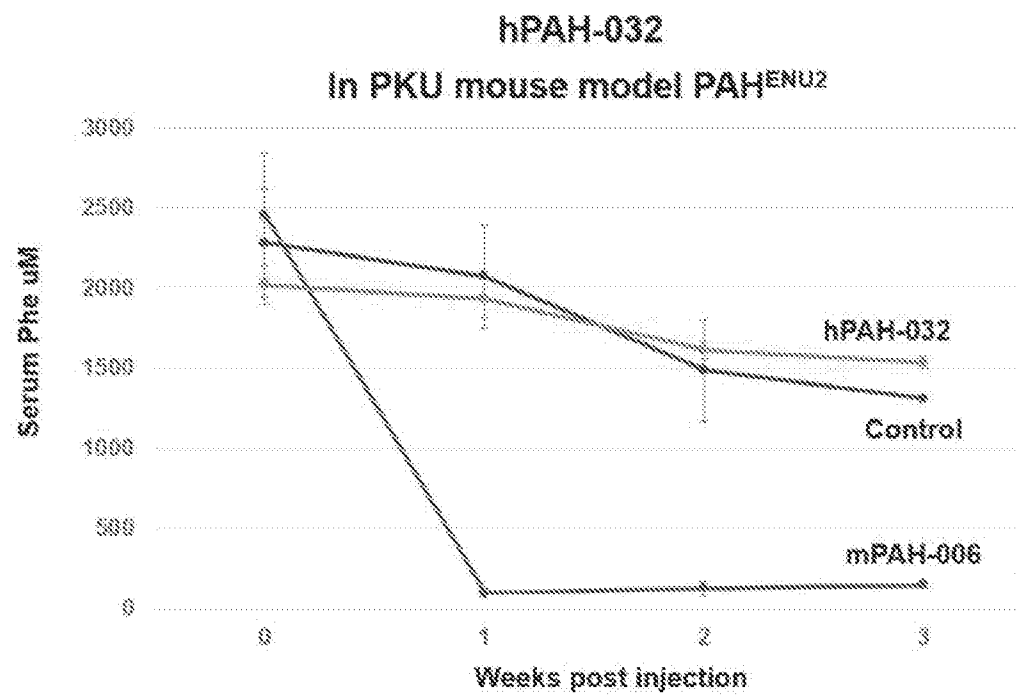


Figure 14



23/26

Figure 15A

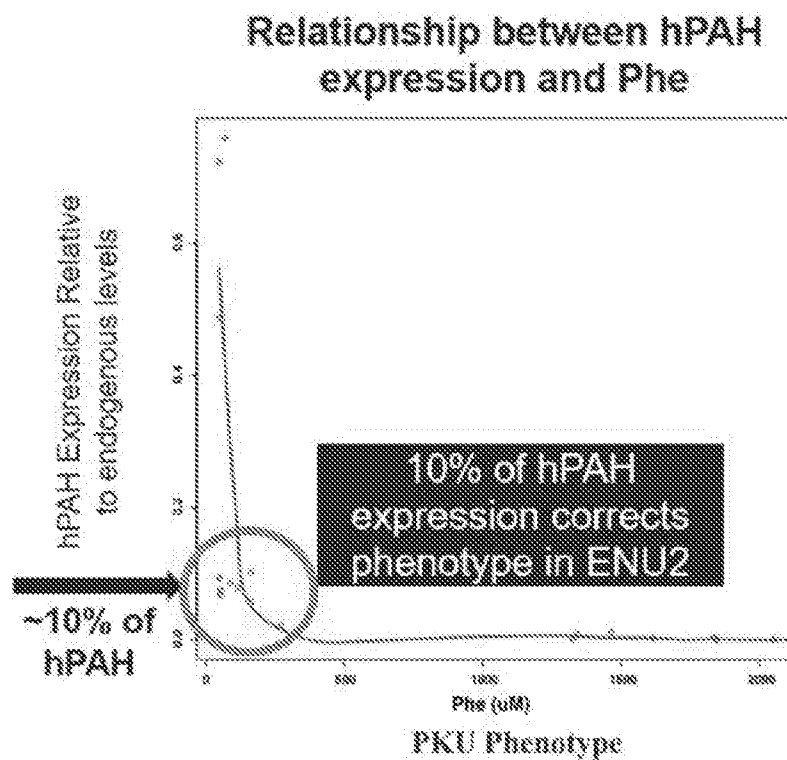


Figure 15B

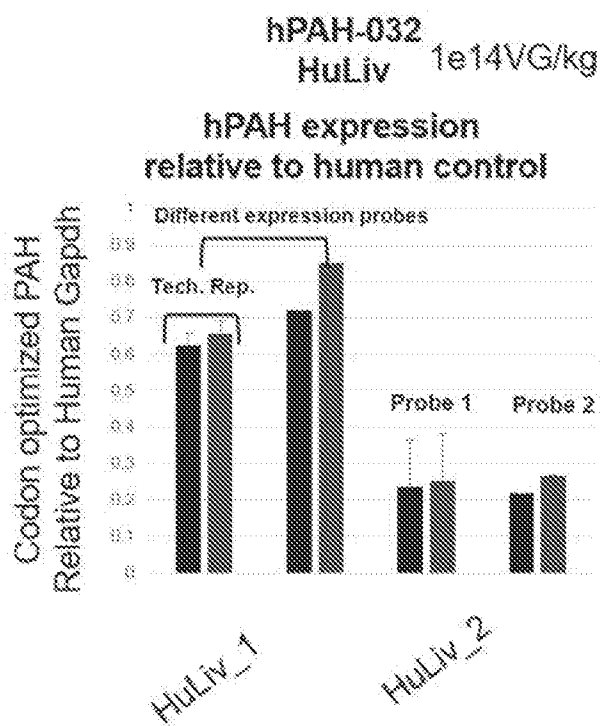


Figure 16

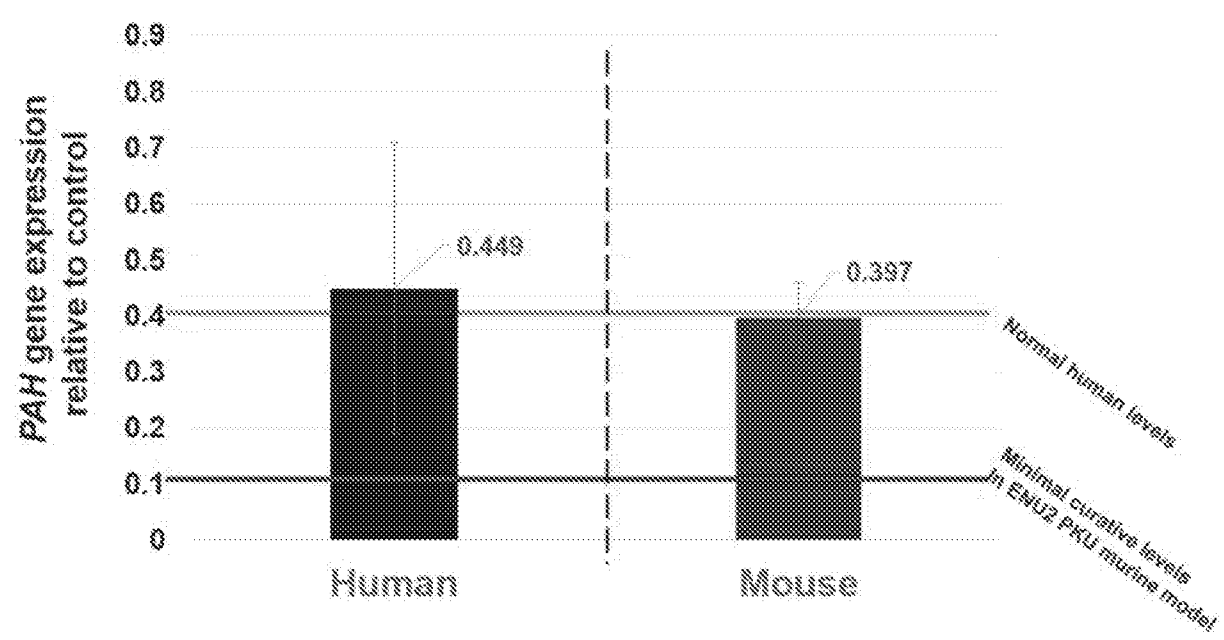


Figure 17A

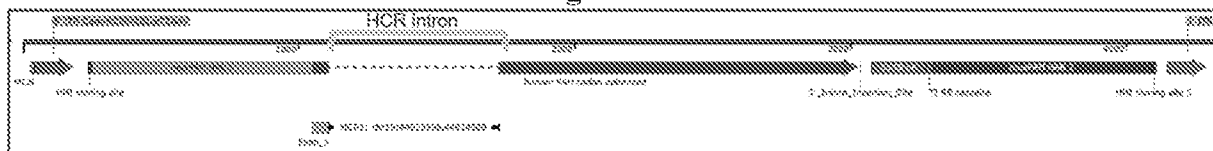


Figure 17B

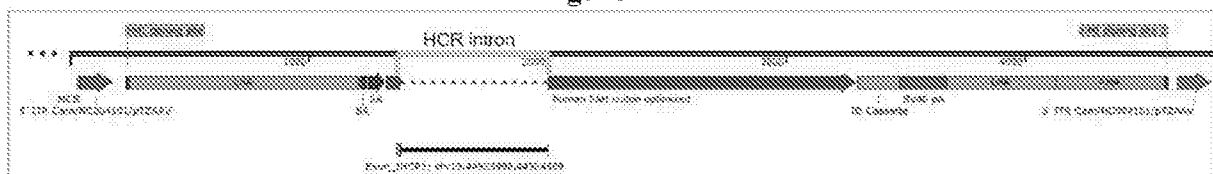


Figure 17C

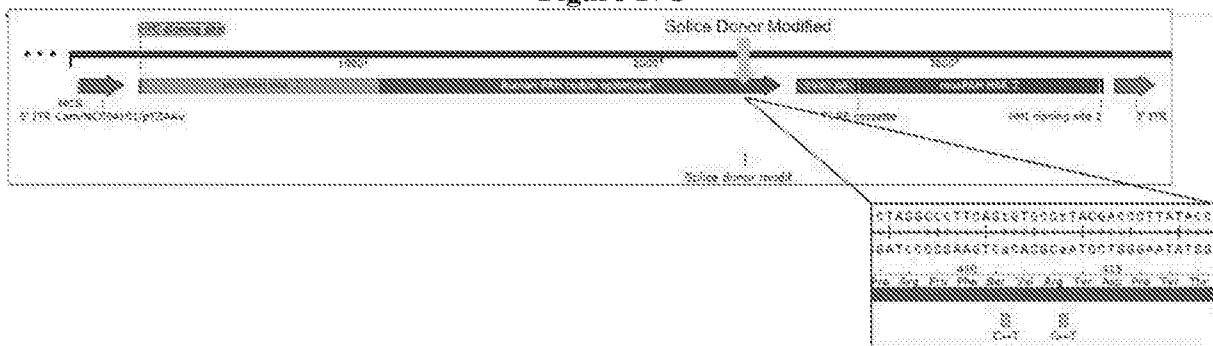


Figure 17D

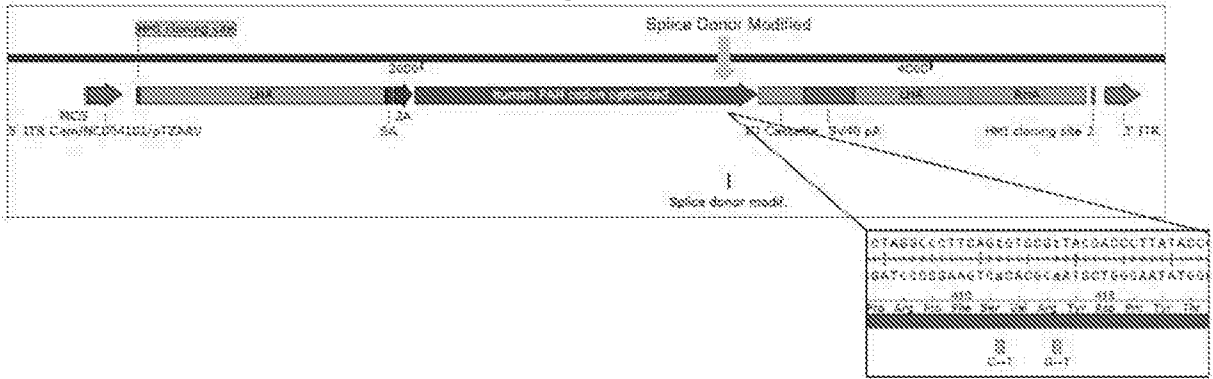


Figure 17E



SEQUENCE LISTING

<110> HOMOLOGY MEDICINES, INC.

<120> ADENO-ASSOCIATED VIRUS COMPOSITIONS FOR RESTORING PAH GENE
FUNCTION AND METHODS OF USE THEREOF

<130> 610062: HMT-024PC

<150> 62/625,149

<151> 2018-02-01

<150> 62/672,377

<151> 2018-05-16

<160> 153

<170> PatentIn version 3.5

<210> 1

<211> 736

<212> PRT

<213> Artificial Sequence

<220>

<223> adeno-associated AAV9

<400> 1

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Val	Asn	Ala	Ala	Asp	Ala	Ala	Ala	Leu	Glu	His	Asp	Lys	Ala	Tyr	Asp	65	70	75	80
Gln	Gln	Leu	Lys	Ala	Gly	Asp	Asn	Pro	Tyr	Leu	Lys	Tyr	Asn	His	Ala	85	90	95	
Asp	Ala	Glu	Phe	Gln	Glu	Arg	Leu	Lys	Glu	Asp	Thr	Ser	Phe	Gly	Gly	100	105	110	
Asn	Leu	Gly	Arg	Ala	Val	Phe	Gln	Ala	Lys	Lys	Arg	Leu	Leu	Glu	Pro	115	120	125	
Leu	Gly	Leu	Val	Glu	Glu	Ala	Ala	Lys	Thr	Ala	Pro	Gly	Lys	Lys	Arg	130	135	140	
Pro	Val	Glu	Gln	Ser	Pro	Gln	Glu	Pro	Asp	Ser	Ser	Ala	Gly	Ile	Gly	145	150	155	160
Lys	Ser	Gly	Ala	Gln	Pro	Ala	Lys	Lys	Arg	Leu	Asn	Phe	Gly	Gln	Thr	165	170	175	
Gly	Asp	Thr	Glu	Ser	Val	Pro	Asp	Pro	Gln	Pro	Ile	Gly	Glu	Pro	Pro	180	185	190	
Ala	Ala	Pro	Ser	Gly	Val	Gly	Ser	Leu	Thr	Met	Ala	Ser	Gly	Gly	Gly	195	200	205	
Ala	Pro	Val	Ala	Asp	Asn	Asn	Glu	Gly	Ala	Asp	Gly	Val	Gly	Ser	Ser	210	215	220	
Ser	Gly	Asn	Trp	His	Cys	Asp	Ser	Gln	Trp	Leu	Gly	Asp	Arg	Val	Ile	225	230	235	240
Thr	Thr	Ser	Thr	Arg	Thr	Trp	Ala	Leu	Pro	Thr	Tyr	Asn	Asn	His	Leu				

245

250

255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
435 440 445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn
500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
545 550 555 560

Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
565 570 575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln
580 585 590

Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln
595 600 605

Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His
610 615 620

Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 2

<211> 736

<212> PRT

<213> Artificial Sequence

<220>

<223> novel AAV isolate

<400> 2

Met Thr Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
100 105 110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
130 135 140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Gln Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
 370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
 385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
 405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
 420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
 435 440 445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
 450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
 465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
 485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn
 500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
 515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
 530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile

545					550					555				560		
Thr	Asn	Glu	Glu	Glu	Ile	Lys	Thr	Thr	Asn	Pro	Val	Ala	Thr	Glu	Ser	
				565					570					575		
Tyr	Gly	Gln	Val	Ala	Thr	Asn	His	Gln	Ser	Ala	Gln	Ala	Gln	Ala	Gln	
			580					585					590			
Thr	Gly	Trp	Val	Gln	Asn	Gln	Gly	Ile	Leu	Pro	Gly	Met	Val	Trp	Gln	
		595					600					605				
Asp	Arg	Asp	Val	Tyr	Leu	Gln	Gly	Pro	Ile	Trp	Ala	Lys	Ile	Pro	His	
	610					615					620					
Thr	Asp	Gly	Asn	Phe	His	Pro	Ser	Pro	Leu	Met	Gly	Gly	Phe	Gly	Met	
625					630					635					640	
Lys	His	Pro	Pro	Pro	Gln	Ile	Leu	Ile	Lys	Asn	Thr	Pro	Val	Pro	Ala	
				645					650					655		
Asp	Pro	Pro	Thr	Ala	Phe	Asn	Lys	Asp	Lys	Leu	Asn	Ser	Phe	Ile	Thr	
			660					665					670			
Gln	Tyr	Ser	Thr	Gly	Gln	Val	Ser	Val	Glu	Ile	Glu	Trp	Glu	Leu	Gln	
		675					680					685				
Lys	Glu	Asn	Ser	Lys	Arg	Trp	Asn	Pro	Glu	Ile	Gln	Tyr	Thr	Ser	Asn	
	690					695					700					
Tyr	Tyr	Lys	Ser	Asn	Asn	Val	Glu	Phe	Ala	Val	Asn	Thr	Glu	Gly	Val	
705					710					715					720	
Tyr	Ser	Glu	Pro	Arg	Pro	Ile	Gly	Thr	Arg	Tyr	Leu	Thr	Arg	Asn	Leu	
				725					730					735		

<210> 3
<211> 736
<212> PRT
<213> Artificial Sequence

<220>
<223> novel AAV isolate

<400> 3

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
100 105 110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg

130

135

140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
 145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
 165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
 180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
 195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
 210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
 225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
 245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
 260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
 275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
 290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
 305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
435 440 445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn
500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
545 550 555 560

Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
565 570 575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln
580 585 590

Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln
595 600 605

Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His
610 615 620

Thr Gly Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Gly Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 4

<211> 736

<212> PRT

<213> Artificial Sequence

<220>

<223> novel AAV isolate

<400> 4

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Ile Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
100 105 110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
130 135 140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser

435

440

445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn
500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
545 550 555 560

Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
565 570 575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln
580 585 590

Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln
595 600 605

Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His
610 615 620

Thr Tyr Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 5

<211> 736

<212> PRT

<213> Artificial Sequence

<220>

<223> novel AAV isolate

<400> 5

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro

20

25

30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
 35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
 50 55 60

Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
 65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
 85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
 100 105 110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
 115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
 130 135 140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Asp
 145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
 165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
 180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
 195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
435 440 445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn
500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
545 550 555 560

Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
565 570 575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln
580 585 590

Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln
595 600 605

Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His
610 615 620

Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 6

<211> 736
<212> PRT
<213> Artificial Sequence

<220>
<223> novel AAV isolate

<400> 6

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
100 105 110

Asn Leu Gly Arg Ala Val Leu Gln Ala Lys Lys Arg Leu Leu Glu Pro
115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
130 135 140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn

325

330

335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
 340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
 355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
 370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
 385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
 405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
 420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
 435 440 445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
 450 455 460

Val Ala Gly Ser Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
 465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
 485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn
 500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
545 550 555 560

Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
565 570 575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln
580 585 590

Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln
595 600 605

Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His
610 615 620

Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 7

<211> 736

<212> PRT

<213> Artificial Sequence

<220>

<223> novel AAV isolate

<400> 7

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
100 105 110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
130 135 140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
435 440 445

Lys	Thr	Ile	Asn	Gly	Ser	Gly	Gln	Asn	Gln	Gln	Thr	Leu	Lys	Phe	Ser	450	455	460	
Val	Ala	Gly	Pro	Ser	Asn	Met	Ala	Val	Gln	Gly	Arg	Asn	Tyr	Ile	Pro	465	470	475	480
Gly	Pro	Ser	Tyr	Arg	Gln	Gln	Arg	Val	Ser	Thr	Thr	Val	Thr	Gln	Asn	485	490	495	
Asn	Asn	Ser	Glu	Phe	Ala	Trp	Pro	Gly	Ala	Ser	Ser	Trp	Ala	Leu	Asn	500	505	510	
Gly	Arg	Asn	Ser	Leu	Met	Asn	Pro	Gly	Pro	Ala	Met	Ala	Ser	His	Lys	515	520	525	
Glu	Gly	Glu	Asp	Arg	Phe	Phe	Pro	Leu	Ser	Gly	Ser	Leu	Ile	Phe	Gly	530	535	540	
Lys	Gln	Gly	Thr	Gly	Arg	Asp	Asn	Val	Asp	Ala	Asp	Lys	Val	Met	Ile	545	550	555	560
Thr	Asn	Glu	Glu	Glu	Ile	Lys	Thr	Thr	Asn	Pro	Val	Ala	Thr	Glu	Ser	565	570	575	
Tyr	Gly	Gln	Val	Ala	Thr	Asn	His	Gln	Ser	Ala	Gln	Ala	Arg	Ala	Gln	580	585	590	
Thr	Gly	Trp	Val	Gln	Asn	Gln	Gly	Ile	Leu	Pro	Gly	Met	Val	Trp	Gln	595	600	605	
Asp	Arg	Asp	Val	Tyr	Leu	Gln	Gly	Pro	Ile	Trp	Ala	Lys	Ile	Pro	His	610	615	620	
Thr	Asp	Gly	Asn	Phe	His	Pro	Ser	Pro	Leu	Met	Gly	Gly	Phe	Gly	Met				

625

630

635

640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 8

<211> 736

<212> PRT

<213> Artificial Sequence

<220>

<223> novel AAV isolate

<400> 8

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
 35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
 50 55 60

Val Asn Ala Val Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
 65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
 85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
 100 105 110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
 115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
 130 135 140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
 145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
 165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
 180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
 195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser

210

215

220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
 225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
 245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
 260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
 275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
 290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
 305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
 325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
 340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
 355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
 370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
 385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
435 440 445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn
500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
545 550 555 560

Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
565 570 575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln
580 585 590

Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln
595 600 605

Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His
610 615 620

Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 9

<211> 736

<212> PRT

<213> Artificial Sequence

<220>

<223> novel AAV isolate

<400> 9

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
100 105 110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
130 135 140

Pro Val Glu Gln Ser Pro Arg Glu Pro Asp Ser Ser Ala Gly Ile Gly
145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
 340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
 355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
 370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
 385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
 405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
 420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
 435 440 445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
 450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
 465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
 485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn
 500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys

515

520

525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
 530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
 545 550 555 560

Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
 565 570 575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln
 580 585 590

Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln
 595 600 605

Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His
 610 615 620

Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
 625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
 645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
 660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
 675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
 690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 10

<211> 736

<212> PRT

<213> Artificial Sequence

<220>

<223> novel AAV isolate

<400> 10

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly

100

105

110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
 115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
 130 135 140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
 145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
 165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
 180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Cys Gly Gly
 195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
 210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
 225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
 245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
 260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
 275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
435 440 445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn
500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
545 550 555 560

Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
565 570 575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln
580 585 590

Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln
595 600 605

Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His
610 615 620

Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
 645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
 660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
 675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
 690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
 705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
 725 730 735

<210> 11
 <211> 736
 <212> PRT
 <213> Artificial Sequence

<220>
 <223> novel AAV isolate

<400> 11

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
 1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
 20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
 35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Arg Ala Tyr Asp
65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
100 105 110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
130 135 140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu

405

410

415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
435 440 445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Gly Ala Ser Ser Trp Ala Leu Asn
500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
545 550 555 560

Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
565 570 575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln
580 585 590

Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln
595 600 605

Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His
610 615 620

Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
675 680 685

Lys Lys Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 12

<211> 736

<212> PRT

<213> Artificial Sequence

<220>

<223> novel AAV isolate

<400> 12

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
100 105 110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
130 135 140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro His Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
435 440 445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Asn
450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Arg Ala Ser Ser Trp Ala Leu Asn
500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
515 520 525

Glu	Gly	Glu	Asp	Arg	Phe	Phe	Pro	Leu	Ser	Gly	Ser	Leu	Ile	Phe	Gly
530						535					540				
Lys	Gln	Gly	Thr	Gly	Arg	Asp	Asn	Val	Asp	Ala	Asp	Lys	Val	Met	Ile
545					550					555					560
Thr	Asn	Glu	Glu	Glu	Ile	Lys	Thr	Thr	Asn	Pro	Val	Ala	Thr	Glu	Ser
				565					570					575	
Tyr	Gly	Gln	Val	Ala	Thr	Asn	His	Gln	Ser	Ala	Gln	Ala	Gln	Ala	Gln
			580					585					590		
Thr	Gly	Trp	Val	Gln	Asn	Gln	Gly	Ile	Leu	Pro	Gly	Met	Val	Trp	Gln
		595					600					605			
Asp	Arg	Asp	Val	Tyr	Leu	Gln	Gly	Pro	Ile	Trp	Ala	Lys	Ile	Pro	His
610						615					620				
Thr	Asp	Gly	Asn	Phe	His	Pro	Ser	Pro	Leu	Met	Gly	Gly	Phe	Gly	Met
625					630					635					640
Lys	His	Pro	Pro	Pro	Gln	Ile	Leu	Ile	Lys	Asn	Thr	Pro	Val	Pro	Ala
				645					650					655	
Asp	Pro	Pro	Thr	Ala	Phe	Asn	Lys	Asp	Lys	Leu	Asn	Ser	Phe	Ile	Thr
			660					665					670		
Gln	Tyr	Ser	Thr	Gly	Gln	Val	Ser	Met	Glu	Ile	Glu	Trp	Glu	Leu	Gln
		675					680					685			
Lys	Glu	Asn	Ser	Lys	Arg	Trp	Asn	Pro	Glu	Ile	Gln	Tyr	Thr	Ser	Asn
690						695					700				
Tyr	Tyr	Lys	Ser	Asn	Asn	Val	Glu	Phe	Ala	Val	Asn	Thr	Glu	Gly	Val

705

710

715

720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 13

<211> 736

<212> PRT

<213> Artificial Sequence

<220>

<223> novel AAV isolate

<400> 13

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
100 105 110

Asn	Leu	Gly	Arg	Ala	Val	Phe	Gln	Ala	Lys	Lys	Arg	Leu	Leu	Glu	Pro	115	120	125	
Leu	Gly	Leu	Val	Glu	Glu	Ala	Ala	Lys	Thr	Ala	Pro	Gly	Lys	Lys	Arg	130	135	140	
Pro	Val	Glu	Gln	Ser	Pro	Gln	Glu	Pro	Asp	Ser	Ser	Ala	Gly	Ile	Gly	145	150	155	160
Lys	Ser	Gly	Ala	Gln	Pro	Ala	Lys	Lys	Arg	Leu	Asn	Phe	Gly	Gln	Thr	165	170	175	
Gly	Asp	Thr	Glu	Ser	Val	Pro	Asp	Pro	Gln	Pro	Ile	Gly	Glu	Pro	Pro	180	185	190	
Ala	Ala	Pro	Ser	Gly	Val	Gly	Ser	Leu	Thr	Met	Ala	Ser	Gly	Gly	Gly	195	200	205	
Ala	Pro	Val	Ala	Asp	Asn	Asn	Glu	Gly	Ala	Asp	Gly	Val	Gly	Ser	Ser	210	215	220	
Ser	Gly	Asn	Trp	His	Cys	Asp	Ser	Gln	Trp	Leu	Gly	Asp	Arg	Val	Ile	225	230	235	240
Thr	Thr	Ser	Thr	Arg	Thr	Trp	Ala	Leu	Pro	Thr	Tyr	Asn	Asn	His	Leu	245	250	255	
Tyr	Lys	Gln	Ile	Ser	Asn	Ser	Thr	Ser	Gly	Gly	Ser	Ser	Asn	Asp	Asn	260	265	270	
Ala	Tyr	Phe	Gly	Tyr	Ser	Thr	Pro	Trp	Gly	Tyr	Phe	Asp	Phe	Asn	Arg	275	280	285	
Phe	His	Cys	His	Phe	Ser	Pro	Arg	Asp	Trp	Gln	Arg	Leu	Ile	Asn	Asn				

290

295

300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
 305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
 325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
 340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
 355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
 370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
 385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
 405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
 420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
 435 440 445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
 450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
 465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Arg Ala Ser Ser Trp Ala Leu Asn
500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
545 550 555 560

Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
565 570 575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln
580 585 590

Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln
595 600 605

Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His
610 615 620

Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 14
<211> 26
<212> DNA
<213> Artificial Sequence

<220>
<223> splice acceptor

<400> 14
c t g a c c t c t t c t c t t c c t c c c a c a g g
26

<210> 15
<211> 736
<212> PRT
<213> Artificial Sequence

<220>
<223> novel AAV isolate

<400> 15

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser

1	5	10	15
Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro	20	25	30
Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro	35	40	45
Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro	50	55	60
Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp	65	70	75
Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala	85	90	95
Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly	100	105	110
Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro	115	120	125
Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg	130	135	140
Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly	145	150	155
Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr	165	170	175
Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro	180	185	190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
420 425 430

Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser
435 440 445

Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser
450 455 460

Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro
465 470 475 480

Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn
485 490 495

Asn Asn Ser Glu Phe Ala Trp Pro Arg Ala Ser Ser Trp Ala Leu Asn
500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
530 535 540

Lys	Gln	Gly	Thr	Gly	Arg	Asp	Asn	Val	Asp	Ala	Asp	Lys	Val	Met	Ile
545					550					555					560
Thr	Asn	Glu	Glu	Glu	Ile	Lys	Thr	Thr	Asn	Pro	Val	Ala	Thr	Glu	Ser
				565					570					575	
Tyr	Gly	Gln	Val	Ala	Thr	Asn	His	Gln	Ser	Ala	Gln	Ala	Gln	Ala	Gln
			580					585					590		
Thr	Gly	Trp	Val	Gln	Asn	Gln	Gly	Ile	Leu	Pro	Gly	Met	Val	Trp	Gln
		595					600					605			
Asp	Arg	Asp	Val	Tyr	Leu	Gln	Gly	Pro	Ile	Trp	Ala	Lys	Ile	Pro	His
	610					615					620				
Thr	Asp	Gly	Asn	Phe	His	Pro	Ser	Pro	Leu	Met	Gly	Gly	Phe	Gly	Met
625					630					635					640
Lys	His	Pro	Pro	Pro	Gln	Ile	Leu	Ile	Lys	Asn	Thr	Pro	Val	Pro	Ala
				645					650					655	
Asp	Pro	Pro	Thr	Ala	Phe	Asn	Lys	Asp	Lys	Leu	Asn	Ser	Phe	Ile	Thr
			660					665					670		
Gln	Tyr	Ser	Thr	Gly	Gln	Val	Ser	Val	Glu	Ile	Glu	Trp	Glu	Arg	Gln
		675					680					685			
Lys	Glu	Asn	Ser	Lys	Arg	Trp	Asn	Pro	Glu	Ile	Gln	Tyr	Thr	Ser	Asn
	690					695					700				
Tyr	Tyr	Lys	Ser	Asn	Asn	Val	Glu	Phe	Ala	Val	Asn	Thr	Glu	Gly	Val
705					710					715					720
Tyr	Ser	Glu	Pro	Arg	Pro	Ile	Gly	Thr	Arg	Tyr	Leu	Thr	Arg	Asn	Leu

725

730

735

<210> 16
<211> 736
<212> PRT
<213> Artificial Sequence

<220>
<223> novel AAV isolate

<400> 16

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
100 105 110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
130 135 140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile

305		310		315		320
Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn						
		325		330		335
Asn Leu Thr Ser Thr Val Gln Val Phe Ala Asp Ser Asp Tyr Gln Leu						
		340		345		350
Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro						
		355		360		365
Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp						
		370		375		380
Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe						
385		390		395		400
Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu						
		405		410		415
Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu						
		420		425		430
Asp Arg Leu Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser						
		435		440		445
Lys Thr Ile Asn Gly Ser Gly Gln Asn Gln Gln Thr Leu Lys Phe Ser						
		450		455		460
Val Ala Gly Pro Ser Asn Met Ala Val Gln Gly Arg Asn Tyr Ile Pro						
465		470		475		480
Gly Pro Ser Tyr Arg Gln Gln Arg Val Ser Thr Thr Val Thr Gln Asn						
		485		490		495

Asn Asn Ser Glu Phe Ala Trp Pro Arg Ala Ser Ser Trp Ala Leu Asn
500 505 510

Gly Arg Asn Ser Leu Met Asn Pro Gly Pro Ala Met Ala Ser His Lys
515 520 525

Glu Gly Glu Asp Arg Phe Phe Pro Leu Ser Gly Ser Leu Ile Phe Gly
530 535 540

Lys Gln Gly Thr Gly Arg Asp Asn Val Asp Ala Asp Lys Val Met Ile
545 550 555 560

Thr Asn Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Ser
565 570 575

Tyr Gly Gln Val Ala Thr Asn His Gln Ser Ala Gln Ala Gln Ala Gln
580 585 590

Thr Gly Trp Val Gln Asn Gln Gly Ile Leu Pro Gly Met Val Trp Gln
595 600 605

Asp Arg Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His
610 615 620

Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
690 695 700

Tyr Tyr Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 17
<211> 736
<212> PRT
<213> Artificial Sequence

<220>
<223> novel AAV isolate

<400> 17

Met Ala Ala Asp Gly Tyr Leu Pro Asp Trp Leu Glu Asp Asn Leu Ser
1 5 10 15

Glu Gly Ile Arg Glu Trp Trp Ala Leu Lys Pro Gly Ala Pro Gln Pro
20 25 30

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro
35 40 45

Gly Tyr Lys Tyr Leu Gly Pro Gly Asn Gly Leu Asp Lys Gly Glu Pro
50 55 60

Val Asn Ala Ala Asp Ala Ala Ala Leu Glu His Asp Lys Ala Tyr Asp
65 70 75 80

Gln Gln Leu Lys Ala Gly Asp Asn Pro Tyr Leu Lys Tyr Asn His Ala
85 90 95

Asp Ala Glu Phe Gln Glu Arg Leu Lys Glu Asp Thr Ser Phe Gly Gly
100 105 110

Asn Leu Gly Arg Ala Val Phe Gln Ala Lys Lys Arg Leu Leu Glu Pro
115 120 125

Leu Gly Leu Val Glu Glu Ala Ala Lys Thr Ala Pro Gly Lys Lys Arg
130 135 140

Pro Val Glu Gln Ser Pro Gln Glu Pro Asp Ser Ser Ala Gly Ile Gly
145 150 155 160

Lys Ser Gly Ala Gln Pro Ala Lys Lys Arg Leu Asn Phe Gly Gln Thr
165 170 175

Gly Asp Thr Glu Ser Val Pro Asp Pro Gln Pro Ile Gly Glu Pro Pro
180 185 190

Ala Ala Pro Ser Gly Val Gly Ser Leu Thr Met Ala Ser Gly Gly Gly
195 200 205

Ala Pro Val Ala Asp Asn Asn Glu Gly Ala Asp Gly Val Gly Ser Ser
210 215 220

Ser Gly Asn Trp His Cys Asp Ser Gln Trp Leu Gly Asp Arg Val Ile
225 230 235 240

Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro Thr Tyr Asn Asn His Leu
245 250 255

Tyr Lys Gln Ile Ser Asn Ser Thr Ser Gly Gly Ser Ser Asn Asp Asn
260 265 270

Ala Tyr Phe Gly Tyr Ser Thr Pro Trp Gly Tyr Phe Asp Phe Asn Arg
275 280 285

Phe His Cys His Phe Ser Pro Arg Asp Trp Gln Arg Leu Ile Asn Asn
290 295 300

Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn Phe Lys Leu Phe Asn Ile
305 310 315 320

Gln Val Lys Glu Val Thr Asp Asn Asn Gly Val Lys Thr Ile Ala Asn
325 330 335

Asn Leu Thr Ser Thr Val Gln Val Phe Thr Asp Ser Asp Tyr Gln Leu
340 345 350

Pro Tyr Val Leu Gly Ser Ala His Glu Gly Cys Leu Pro Pro Phe Pro
355 360 365

Ala Asp Val Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asp
370 375 380

Gly Ser Gln Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe
385 390 395 400

Pro Ser Gln Met Leu Arg Thr Gly Asn Asn Phe Gln Phe Ser Tyr Glu
405 410 415

Phe Glu Asn Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu
420 425 430

Asp	Arg	Leu	Met	Asn	Pro	Leu	Ile	Asp	Gln	Tyr	Leu	Tyr	Tyr	Leu	Ser
		435					440					445			
Lys	Thr	Ile	Asn	Gly	Ser	Gly	Gln	Asn	Gln	Gln	Thr	Leu	Lys	Phe	Ser
		450				455					460				
Val	Ala	Gly	Pro	Ser	Asn	Met	Ala	Val	Gln	Gly	Arg	Asn	Tyr	Ile	Pro
465					470					475					480
Gly	Pro	Ser	Tyr	Arg	Gln	Gln	Arg	Val	Ser	Thr	Thr	Val	Thr	Gln	Asn
				485					490					495	
Asn	Asn	Ser	Glu	Ile	Ala	Trp	Pro	Arg	Ala	Ser	Ser	Trp	Ala	Leu	Asn
			500					505					510		
Gly	Arg	Asn	Ser	Leu	Met	Asn	Pro	Gly	Pro	Ala	Met	Ala	Ser	His	Lys
		515					520					525			
Glu	Gly	Glu	Asp	Arg	Phe	Phe	Pro	Leu	Ser	Gly	Ser	Leu	Ile	Phe	Gly
	530					535					540				
Lys	Gln	Gly	Thr	Gly	Arg	Asp	Asn	Val	Asp	Ala	Asp	Lys	Val	Met	Ile
545					550					555					560
Thr	Asn	Glu	Glu	Glu	Ile	Lys	Thr	Thr	Asn	Pro	Val	Ala	Thr	Glu	Ser
				565					570					575	
Tyr	Gly	Gln	Val	Ala	Thr	Asn	His	Gln	Ser	Ala	Gln	Ala	Gln	Ala	Gln
			580					585					590		
Thr	Gly	Trp	Val	Gln	Asn	Gln	Gly	Ile	Leu	Pro	Gly	Met	Val	Trp	Gln
		595					600					605			
Asp	Arg	Asp	Val	Tyr	Leu	Gln	Gly	Pro	Ile	Trp	Ala	Lys	Ile	Pro	His

610

615

620

Thr Asp Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Met
625 630 635 640

Lys His Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala
645 650 655

Asp Pro Pro Thr Ala Phe Asn Lys Asp Lys Leu Asn Ser Phe Ile Thr
660 665 670

Gln Tyr Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln
675 680 685

Lys Glu Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn
690 695 700

Tyr Cys Lys Ser Asn Asn Val Glu Phe Ala Val Asn Thr Glu Gly Val
705 710 715 720

Tyr Ser Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
725 730 735

<210> 18

<211> 145

<212> DNA

<213> Artificial Sequence

<220>

<223> AAV2 5' ITR

<400> 18

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccggg'gcgacc aaaggtcgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

g c c a a c t c c a t c a c t a g g g g t t c c t
145

<210> 19
<211> 145
<212> DNA
<213> Artificial Sequence

<220>
<223> AAV2 3' ITR

<400> 19
aggaaccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctcactgagg
60

ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc
120

g a g c g c g c a g a g a g g g a g t g g c c a a
145

<210> 20
<211> 167
<212> DNA
<213> Artificial Sequence

<220>
<223> AAV5 5' ITR

<400> 20
ctctccccc tgtcgcgttc gctcgctcgc tggctcgttt ggggggggtgg cagctcaaag
60

agctgccaga cgacggccct ctggccgctc ccccccaaa cgagccagcg agcgagcgaa
120

cgcgacaggg gggagagtgc cacactctca agcaaggggg ttttgta
167

<210> 21
<211> 167

<212> DNA

<213> Artificial Sequence

<220>

<223> AAV5 3' ITR

<400> 21

tacaaaacct ccttgcttga gagtgtggca ctctcccccc tgtcgcggttc gctcgctcgc
60

tggtcgtttt ggggggggtgg cagctcaaag agctgccaga cgacggccct ctggccgtcg
120

cccccccaaa cgagccagcg agcgagcgaa cgcgacaggg gggagag
167

<210> 22

<211> 621

<212> PRT

<213> Artificial Sequence

<220>

<223> AAV2 Rep

<400> 22

Met Pro Gly Phe Tyr Glu Ile Val Ile Lys Val Pro Ser Asp Leu Asp
1 5 10 15

Glu His Leu Pro Gly Ile Ser Asp Ser Phe Val Asn Trp Val Ala Glu
20 25 30

Lys Glu Trp Glu Leu Pro Pro Asp Ser Asp Met Asp Leu Asn Leu Ile
35 40 45

Glu Gln Ala Pro Leu Thr Val Ala Glu Lys Leu Gln Arg Asp Phe Leu
50 55 60

Thr Glu Trp Arg Arg Val Ser Lys Ala Pro Glu Ala Leu Phe Phe Val
65 70 75 80

Gln Phe Glu Lys Gly Glu Ser Tyr Phe His Met His Val Leu Val Glu
85 90 95

Thr Thr Gly Val Lys Ser Met Val Leu Gly Arg Phe Leu Ser Gln Ile
100 105 110

Arg Glu Lys Leu Ile Gln Arg Ile Tyr Arg Gly Ile Glu Pro Thr Leu
115 120 125

Pro Asn Trp Phe Ala Val Thr Lys Thr Arg Asn Gly Ala Gly Gly Gly
130 135 140

Asn Lys Val Val Asp Glu Cys Tyr Ile Pro Asn Tyr Leu Leu Pro Lys
145 150 155 160

Thr Gln Pro Glu Leu Gln Trp Ala Trp Thr Asn Met Glu Gln Tyr Leu
165 170 175

Ser Ala Cys Leu Asn Leu Thr Glu Arg Lys Arg Leu Val Ala Gln His
180 185 190

Leu Thr His Val Ser Gln Thr Gln Glu Gln Asn Lys Glu Asn Gln Asn
195 200 205

Pro Asn Ser Asp Ala Pro Val Ile Arg Ser Lys Thr Ser Ala Arg Tyr
210 215 220

Met Glu Leu Val Gly Trp Leu Val Asp Lys Gly Ile Thr Ser Glu Lys
225 230 235 240

Gln Trp Ile Gln Glu Asp Gln Ala Ser Tyr Ile Ser Phe Asn Ala Ala
245 250 255

Ser Asn Ser Arg Ser Gln Ile Lys Ala Ala Leu Asp Asn Ala Gly Lys
260 265 270

Ile Met Ser Leu Thr Lys Thr Ala Pro Asp Tyr Leu Val Gly Gln Gln
275 280 285

Pro Val Glu Asp Ile Ser Ser Asn Arg Ile Tyr Lys Ile Leu Glu Leu
290 295 300

Asn Gly Tyr Asp Pro Gln Tyr Ala Ala Ser Val Phe Leu Gly Trp Ala
305 310 315 320

Thr Lys Lys Phe Gly Lys Arg Asn Thr Ile Trp Leu Phe Gly Pro Ala
325 330 335

Thr Thr Gly Lys Thr Asn Ile Ala Glu Ala Ile Ala His Thr Val Pro
340 345 350

Phe Tyr Gly Cys Val Asn Trp Thr Asn Glu Asn Phe Pro Phe Asn Asp
355 360 365

Cys Val Asp Lys Met Val Ile Trp Trp Glu Glu Gly Lys Met Thr Ala
370 375 380

Lys Val Val Glu Ser Ala Lys Ala Ile Leu Gly Gly Ser Lys Val Arg
385 390 395 400

Val Asp Gln Lys Cys Lys Ser Ser Ala Gln Ile Asp Pro Thr Pro Val
405 410 415

Ile Val Thr Ser Asn Thr Asn Met Cys Ala Val Ile Asp Gly Asn Ser
420 425 430

Thr	Thr	Phe	Glu	His	Gln	Gln	Pro	Leu	Gln	Asp	Arg	Met	Phe	Lys	Phe
		435					440					445			
Glu	Leu	Thr	Arg	Arg	Leu	Asp	His	Asp	Phe	Gly	Lys	Val	Thr	Lys	Gln
	450					455					460				
Glu	Val	Lys	Asp	Phe	Phe	Arg	Trp	Ala	Lys	Asp	His	Val	Val	Glu	Val
465					470					475					480
Glu	His	Glu	Phe	Tyr	Val	Lys	Lys	Gly	Gly	Ala	Lys	Lys	Arg	Pro	Ala
				485					490					495	
Pro	Ser	Asp	Ala	Asp	Ile	Ser	Glu	Pro	Lys	Arg	Val	Arg	Glu	Ser	Val
			500					505					510		
Ala	Gln	Pro	Ser	Thr	Ser	Asp	Ala	Glu	Ala	Ser	Ile	Asn	Tyr	Ala	Asp
		515					520					525			
Arg	Tyr	Gln	Asn	Lys	Cys	Ser	Arg	His	Val	Gly	Met	Asn	Leu	Met	Leu
	530					535					540				
Phe	Pro	Cys	Arg	Gln	Cys	Glu	Arg	Met	Asn	Gln	Asn	Ser	Asn	Ile	Cys
545					550					555					560
Phe	Thr	His	Gly	Gln	Lys	Asp	Cys	Leu	Glu	Cys	Phe	Pro	Val	Ser	Glu
				565					570					575	
Ser	Gln	Pro	Val	Ser	Val	Val	Lys	Lys	Ala	Tyr	Gln	Lys	Leu	Cys	Tyr
			580					585					590		
Ile	His	His	Ile	Met	Gly	Lys	Val	Pro	Asp	Ala	Cys	Thr	Ala	Cys	Asp
		595					600					605			
Leu	Val	Asn	Val	Asp	Leu	Asp	Asp	Cys	Ile	Phe	Glu	Gln			

610

615

620

<210> 23
 <211> 452
 <212> PRT
 <213> Homo sapiens

<400> 23

Met Ser Thr Ala Val Leu Glu Asn Pro Gly Leu Gly Arg Lys Leu Ser
 1 5 10 15

Asp Phe Gly Gln Glu Thr Ser Tyr Ile Glu Asp Asn Cys Asn Gln Asn
 20 25 30

Gly Ala Ile Ser Leu Ile Phe Ser Leu Lys Glu Glu Val Gly Ala Leu
 35 40 45

Ala Lys Val Leu Arg Leu Phe Glu Glu Asn Asp Val Asn Leu Thr His
 50 55 60

Ile Glu Ser Arg Pro Ser Arg Leu Lys Lys Asp Glu Tyr Glu Phe Phe
 65 70 75 80

Thr His Leu Asp Lys Arg Ser Leu Pro Ala Leu Thr Asn Ile Ile Lys
 85 90 95

Ile Leu Arg His Asp Ile Gly Ala Thr Val His Glu Leu Ser Arg Asp
 100 105 110

Lys Lys Lys Asp Thr Val Pro Trp Phe Pro Arg Thr Ile Gln Glu Leu
 115 120 125

Asp Arg Phe Ala Asn Gln Ile Leu Ser Tyr Gly Ala Glu Leu Asp Ala
 130 135 140

Asp His Pro Gly Phe Lys Asp Pro Val Tyr Arg Ala Arg Arg Lys Gln
145 150 155 160

Phe Ala Asp Ile Ala Tyr Asn Tyr Arg His Gly Gln Pro Ile Pro Arg
165 170 175

Val Glu Tyr Met Glu Glu Glu Lys Lys Thr Trp Gly Thr Val Phe Lys
180 185 190

Thr Leu Lys Ser Leu Tyr Lys Thr His Ala Cys Tyr Glu Tyr Asn His
195 200 205

Ile Phe Pro Leu Leu Glu Lys Tyr Cys Gly Phe His Glu Asp Asn Ile
210 215 220

Pro Gln Leu Glu Asp Val Ser Gln Phe Leu Gln Thr Cys Thr Gly Phe
225 230 235 240

Arg Leu Arg Pro Val Ala Gly Leu Leu Ser Ser Arg Asp Phe Leu Gly
245 250 255

Gly Leu Ala Phe Arg Val Phe His Cys Thr Gln Tyr Ile Arg His Gly
260 265 270

Ser Lys Pro Met Tyr Thr Pro Glu Pro Asp Ile Cys His Glu Leu Leu
275 280 285

Gly His Val Pro Leu Phe Ser Asp Arg Ser Phe Ala Gln Phe Ser Gln
290 295 300

Glu Ile Gly Leu Ala Ser Leu Gly Ala Pro Asp Glu Tyr Ile Glu Lys
305 310 315 320

Leu Ala Thr Ile Tyr Trp Phe Thr Val Glu Phe Gly Leu Cys Lys Gln
325 330 335

Gly Asp Ser Ile Lys Ala Tyr Gly Ala Gly Leu Leu Ser Ser Phe Gly
340 345 350

Glu Leu Gln Tyr Cys Leu Ser Glu Lys Pro Lys Leu Leu Pro Leu Glu
355 360 365

Leu Glu Lys Thr Ala Ile Gln Asn Tyr Thr Val Thr Glu Phe Gln Pro
370 375 380

Leu Tyr Tyr Val Ala Glu Ser Phe Asn Asp Ala Lys Glu Lys Val Arg
385 390 395 400

Asn Phe Ala Ala Thr Ile Pro Arg Pro Phe Ser Val Arg Tyr Asp Pro
405 410 415

Tyr Thr Gln Arg Ile Glu Val Leu Asp Asn Thr Gln Gln Leu Lys Ile
420 425 430

Leu Ala Asp Ser Ile Asn Ser Glu Ile Gly Ile Leu Cys Ser Ala Leu
435 440 445

Gln Lys Ile Lys
450

<210> 24

<211> 1359

<212> DNA

<213> Homo sapiens

<400> 24

atgtccactg cggtcctgga aaaccaggc ttgggcagga aactctctga ctttggacag
60

gaaacaagct atattgaaga caactgcaat caaaatggtg ccatatcact gatcttctca
120

ctcaaagaag aagttggtgc attggccaaa gtattgcgct tatttgagga gaatgatgta
180

aacctgaccc acattgaatc tagaccttct cgtttaaaga aagatgagta tgaatttttc
240

acccatttgg ataaacgtag cctgcctgct ctgacaaaca tcatcaagat cttgaggcat
300

gacattggtg ccactgtcca tgagctttca cgagataaga agaaagacac agtgccttgg
360

ttcccaagaa ccattcaaga gctggacaga ttgccaatc agattctcag ctatggagcg
420

gaactggatg ctgaccaccc tggtttttaa gatcctgtgt accgtgcaag acggaagcag
480

tttgctgaca ttgcctacaa ctaccgccat gggcagccca tccctcgagt ggaatacatg
540

gaggaagaaa agaaaacatg gggcacagtg ttcaagactc tgaagtcctt gtataaaacc
600

catgcttgct atgagtacaa tcacatTTTT ccacttcttg aaaagtactg tggcttccat
660

gaagataaca ttccccagct ggaagacggt tctcaattcc tgcagacttg cactggtttc
720

cgcctccgac ctgtggctgg cctgctttcc tctcgggatt tcttgggtgg cctggccttc
780

cgagtcttcc actgcacaca gtacatcaga catggatcca agcccatgta taccctcgaa
840

cctgacatct gccatgagct gttgggacat gtgcccttgt tttcagatcg cagctttgcc
900

cagttttccc aggaaattgg ccttgcctct ctgggtgcac ctgatgaata cattgaaaag
960

ctcgccacaa tttactgggt tactgtggag tttgggctct gcaaacaagg agactccata
1020

aaggcatatg gtgctgggct cctgtcatcc tttggtgaat tacagtactg cttatcagag
1080

aagccaaagc ttctccccct ggagctggag aagacagcca tccaaaatta cactgtcacg
1140

gagttccagc ccctgtatta cgtggcagag agttttaatg atgccaagga gaaagtaagg
1200

aactttgctg ccacaatacc tcggcccttc tcagttcgct acgaccata cacccaaagg
1260

attgaggtct tggacaatac ccagcagctt aagattttgg ctgattccat taacagtgaa
1320

attggaatcc tttgcagtgc cctccagaaa ataaagtaa
1359

<210> 25

<211> 1359

<212> DNA

<213> Artificial Sequence

<220>

<223> Silently altered PAH coding sequence

<400> 25

atgtccaccg ctgtgctgga gaaccctggg ctgggggagga aactgtcaga cttcgggcag
60

gagacttcac acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc
120

ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg
180

aatctgaccc acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt
240

accacactgg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac
300

gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg
360

ttccctcgga caatccagga gctggataga tttgccaacc agatcctgtc ttacggagca
420

gagctggacg cagatcaccc tggcttcaag gacccagtgt atcgggcccg gagaaagcag
480

tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg
540

gaggaggaga agaagacctg gggcacagtg ttcaagacct tgaagagcct gtacaagaca
600

cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac
660

gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt
720

aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc
780

agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag
840

cccgacatct gtcacgagct gctggggccac gtgcccctgt ttagcgatag atccttcgcc
900

cagttttccc aggagatcgg actggcatct ctgggagcac ctgacgagta catcgagaag
960

ctggccacca tctattgggt cacagtggag tttggcctgt gcaagcaggg cgatagcatc
1020

aaggcctacg gagcaggact gctgtctagc ttcggcgagc tgcagtattg tctgtccgag
1080

aagccaaagc tgctgcccct ggagctggag aagaccgcca tccagaacta caccgtgaca
1140

gagttccagc ccctgtacta tgtggccgag tcttttaacg atgccaagga gaagggtgaga
1200

aatttcgccg ccacaatccc taggcccttc agcgtgcggt acgaccctta tacccagagg
1260

atcgagggtgc tggataatac acagcagctg aagatcctgg ctgactcaat caatagcgaa
1320

atcggaatcc tgtgctccgc cctgcagaaa atcaaatga
1359

<210> 26

<211> 106

<212> DNA

<213> Artificial Sequence

<220>

<223> truncated AAV2 5'ITR

<400> 26

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgctcg ggcgaccttt
60

ggtcgccccgg cctcagtgag cgagcgagcg cgcagagagg gagtgg
106

<210> 27

<211> 143

<212> DNA

<213> Artificial Sequence

<220>

<223> modified AAV2 3'ITR

<400> 27

aggaaccctt agtgatggag ttggccactc cctctctgcg cgctcgctcg ctactgagg
60

ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc
120

g a g c g c g c a g a g a g g g a g t g g c c
143

<210> 28
<211> 130
<212> DNA
<213> Homo sapiens

<400> 28
gttggtatca aggttacaag acaggtttaa ggagaccaat agaaactggg catgtggaga
60

cagagaagac tcttggggtt ctgataggca ctgactctct ctgcctattg gtctattttc
120

c c a c c c t t a g
130

<210> 29
<211> 116
<212> DNA
<213> Mus musculus

<400> 29
gttggtatcc aggttacaag gcagctcaca agaagaagtt gggtgcttgg agacagaggt
60

ctgctttcca gcagacacta actttcagtg tcccctgtct atgtttccct ttttag
116

<210> 30
<211> 92
<212> DNA
<213> Artificial Sequence

<220>
<223> Minute virus in mice

<400> 30
aagaggtaag ggtttaaggg atggttgggt ggtgggggtat taatgtttaa ttacctggag
60

c a c c t g c c t g a a a t c a c t t t t t t t c a g g t t g g
92

<210> 31
<211> 198
<212> DNA
<213> Simian virus 40

<400> 31
gatccagaca tgataagata cattgatgag tttggacaaa ccacaactag aatgcagtga
60

aaaaaatgct ttattttgtga aattttgtgat gctattgctt tattttgtaac cattataagc
120

tgcaataaac aagttaacaa caacaattgc attcatttta tgtttcaggt tcagggggag
180

g t g t g g g a g g t t t t t t a a
198

<210> 32
<211> 122
<212> DNA
<213> Simian virus 40

<400> 32
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca
60

aataaagcat ttttttcact gcattctagt tgtggtttgt ccaaactcat caatgtatct
120

t a
122

<210> 33
<211> 133
<212> DNA
<213> Simian virus 40

<400> 33
tgctttatth gtgaaatttg tgatgctatt gctttatttg taaccattat aagctgcaat
60

aaacaagtta acaacaacaa ttgcattcat tttatgtttc aggttcaggg ggaggtgtgg
120

g a g g t t t t t t a a a
133

<210> 34
<211> 135
<212> DNA
<213> Simian virus 40

<400> 34
gatccagaca tgataagata cattgatgag tttggacaaa ccacaactag aatgcagtga
60

aaaaaatgct ttatttgtga aatttgtgat gctattgctt tatttgtaac cattataagc
120

t g c a a t a a a c a a g t t
135

<210> 35
<211> 1654
<212> DNA
<213> Artificial Sequence

<220>
<223> pHMIA-hPAH-hI1C-032.1 editing element

<400> 35
ctgacctctt ctcttcctcc cacagggcag cggagctact aacttcagcc tgctgaagca
60

ggctggagac gtggaggaga accctggacc tatgtccacc gctgtgctgg agaaccctgg
120

gctggggagg aaactgtcag acttcgggca ggagacttca tacattgagg ataactgtaa
180

ccagaatggc gccatctctc tgatcttcag cctgaaggag gaagtgggcg ccctggcaaa
240

ggtgctgcgc ctgtttgagg agaacgacgt gaatctgacc cacatcgagt cccggccttc
300

tagactgaag aaggacgagt acgagttctt tacccacctg gataagcggg ccctgccagc
360

cctgacaaac atcatcaaga tcctgaggca cgacatcgga gcaaccgtgc acgagctgtc
420

tcgggacaag aagaaggata ccgtgccctg gttccctcgg acaatccagg agctggatag
480

atttgccaac cagatcctgt cttacggagc agagctggac gcagatcacc ctggcttcaa
540

ggacccagtg tatcggggccc ggagaaagca gtttgccgat atcgccctaca attataggca
600

cggacagcca atccctcgcg tggagtatat ggaggaggag aagaagacct ggggcacagt
660

gttcaagacc ctgaagagcc tgtacaagac acacgcctgc tacgagtata accacatctt
720

ccccctgctg gagaagtatt gtggctttca cgaggacaat atccctcagc tggaggacgt
780

gagccagttc ctgcagacct gcacaggcctt taggctgagg ccagtggcag gactgctgag
840

ctcccgggac ttcctgggag gactggcctt cagagtgttt cactgcaccc agtacatcag
900

gcacggctcc aagccaatgt atacaccaga gcccgcacac tgtcacgagc tgctgggcca
960

cgtgcccctg tttagcgata gacccctcgc ccagttttcc caggagatcg gactggcatc
1020

tctgggagca cctgacgagt acatcgagaa gctggccacc atctattggg tcacagtgga
1080

gtttggcctg tgcaagcagg gcgatagcat caaggcctac ggagcaggac tgctgtctag
1140

cttcggcgag ctgcagtatt gtctgtccga gaagccaaag ctgctgcccc tggagctgga
1200

gaagaccgcc atccagaact acaccgtgac agagttccag cccctgtact atgtggccga
1260

gtcttttaac gatgccaaagg agaagggtgag aaattttcgcc gccacaatcc ctaggccctt
1320

cagcgtgcgg tacgaccctt ataccagag gatcgagggtg ctggataata cacagcagct
1380

gaagatcctg gctgactcaa tcaatagcga aatcggaatc ctgtgctccg ccctgcagaa
1440

aatcaaatga ggtaccgatc cagacatgat aagatacatt gatgagtttg gacaaaccac
1500

aactagaatg cagtgaaaaa aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt
1560

tgtaaccatt ataagctgca ataaacaagt taacaacaac aattgcattc attttatgtt
1620

t c a g g t t c a g g g g g a g g t g t g g g a g g t t t t t t a a
1654

<210> 36

<211> 960

<212> DNA

<213> Homo sapiens

<400> 36

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaatagaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acctggagca
480

tccggcaggc tggcctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctcctcccta gtgcgagggtt
600

aaaaccttca gcccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

<210> 37

<211> 960

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.2 vector 5' homology arm

<400> 37

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagaggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggtttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctccctc cggcccgctc tgggcagggtg acctggagca
480

tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgagggtt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccaccatgg cggcggcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

<210> 38
<211> 960
<212> DNA
<213> Artificial Sequence

<220>
<223> 32.3 vector 5' homology arm

<400> 38
gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagaggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctga gagtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggt cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggt acctggagca
480

tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgagggt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

<210> 39

<211> 960

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.4 vector 5' homology arm

<400> 39

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagaggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acctggagca
480

tccggcaggc tgccctggcc tcctgagtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgagggt
600

aaaaccttca gccccacgtg ctgttttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

<210> 40

<211> 960

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.5 vector 5' homology arm

<400> 40

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaagaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcataataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acgtcaagca
480

tccggcaggc tgccctggcc tcctgagtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgaggtt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

<210> 41

<211> 960

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.6 vector 5' homology arm

<400> 41

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acctggagca
480

tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgaggtt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccaactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaatg tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

<210> 42

<211> 960

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.7 vector 5' homology arm

<400> 42

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggtt aacagaagggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttccaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acctggagca
480

tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgaggtt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacaggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcgggtggct ttcacaggaa
960

<210> 43

<211> 968

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.8 vector 5' homology arm

<400> 43

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttccaccag acggtcgcag gcttagtcca attgcagaga acacgctgtt
300

cttcgccccca ggcttctgag agtcccggaa gtgcctaaac ctgtctaatac gacgggggctt
360

gggtggcccg tcgctccctg gcttcttccc tttaccagg gcgggcagcg aagtggtgcc
420

tcctgcgctcc cccacaccct ccctcagccc ctccccctccg gcccgtcctg ggcagggtgac
480

ctggagcatc cggcaggctg ccctggcctc ctgcgtcagg acaacgcccc cgagggggcgt
540

tactgtgcgg agatgcacca cgcaagagac accctttgta actctcttct cctccctagt
600

gcgagggttaa aaccttcagc cccacgtgct gtttgcaaac ctgcctgtac ctgaggccct
660

aaaaagccag agacctcact cccggggagc cagcatgtcc actgcggtcc tggaaaaccc
720

aggcttgggc aggaaactct ctgactttgg acagggtgagc cacggcagcc tgagctgctc
780

agttagggga atttgggcct ccagagaaaag agatccgaag actgctggtg cttcctgggt
840

tcataagctc agtaagaagt ctgaattcgt tggaagctga tgagaatatc caggaagtca
900

acagacaaat gtcctcaaca attgtttcta agtaggagaa catctgtcct cggtggcttt
960

c a c a g g a a
968

<210> 44

<211> 960

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.9 vector 5' homology arm

<400> 44

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcggttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagaggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctga gagtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggctttctg agagtcccgg aagtgcctaa acctgtctaa tcgacagagc ttgggtggcc
360

cgtcgctccc tggctttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc taggcagggtg acctgaagca
480

tccagcaggc tggcctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cagcaagag acaccctttg taactctctt ctccctcccta gtgcgagggtt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

<210> 45

<211> 911
<212> DNA
<213> Homo sapiens

<400> 45

ctgggatggg atgtggaatc cttctagatt tcttttgtaa tatttataaa gtgctctcag
60

caaggtatca aaatggcaaa attgtgagta actatcctcc tttcattttg ggaagaagat
120

gaggcatgaa gagaattcag acagaaactt actcagacca ggggaggcag aaactaagca
180

gagaggaaaa tgaccaagag ttagccctgg gcatggaatg tgaaagaacc ctaaactga
240

cttgaaata atgccaagg tatattccat tctccgggat ttgttggcat tttcttgagg
300

tgaagaattg cagaatacat tctttaatgt gacctacata tttacccatg ggaggaagtc
360

tgctcctgga ctcttgagat tcagtcataa agcccaggcc agggaaataa tgtaagtctg
420

caggcccctg tcatcagtag gattagggag aagagttctc agtagaaaac agggaggctg
480

gagagaaaag aatggttaat gttaacgtta atataactag aaagactgca gaacttagga
540

ctgattttta tttgaatcct taaaaaaaaa aatttcttat gaaaatagta catggctctt
600

aggagacaga acttattgta cagaggaaca gcgtgagagt cagagtgatc ccagaacagg
660

tcctggctcc atcctgcaca tagttttggg gctgctggca atacgggtccc cacaactgtg
720

ggaagggggt aggggcaggg atctcatcag gaaagcatag gggtttaaag ttctttatag
780

agcacttaga agattgagaa tccacaaatt atattaataa caaacaaggt agtgtcgtgt
840

tatataagtaa atgtgaattt gcagacacat ttagggaaaa gttataatta aaaaaatagg
900

c t g t a t a t a t a
911

<210> 46

<211> 3537

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.1 vector correction genome

<400> 46

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acctggagca
480

tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgagggtt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacaggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

aagcttctga cctcttctct tcctcccaca gggcagcgga gctactaact tcagcctgct
1020

gaagcaggct ggagacgtgg aggagaaccc tggacctatg tccaccgctg tgctggagaa
1080

ccctgggctg gggaggaaac tgtcagactt cgggcaggag acttcataca ttgaggataa
1140

ctgtaaccag aatggcgcca tctctctgat cttcagcctg aaggaggaag tgggcgcctt
1200

ggcaaagggtg ctgcgctgtt ttgaggagaa cgacgtgaat ctgaccacaa tcgagtcccg
1260

gccttctaga ctgaagaagg acgagtacga gttctttacc cacctggata agcggtcctt
1320

gccagccctg acaaacatca tcaagatcct gaggcacgac atcggagcaa ccgtgcacga
1380

gctgtctcgg gacaagaaga aggataccgt gccctggttc cctcggacaa tccaggagct
1440

ggatagattt gcccaaccaga tcctgtctta cggagcagag ctggacgcag atcaccctgg
1500

cttcaaggac 1560	ccagtgtatc	gggcccggag	aaagcagttt	gccgatatcg	cctacaatta
taggcacgga 1620	cagccaatcc	ctcgcgtgga	gtatatggag	gaggagaaga	agacctgggg
cacagtgttc 1680	aagaccctga	agagcctgta	caagacacac	gcctgctacg	agtataacca
catcttcccc 1740	ctgctggaga	agtattgtgg	ctttcacgag	gacaatatcc	ctcagctgga
ggacgtgagc 1800	cagttcctgc	agacctgcac	aggctttagg	ctgaggccag	tggcaggact
gctgagctcc 1860	cgggacttcc	tgggaggact	ggccttcaga	gtgtttcact	gcacccagta
catcaggcac 1920	ggctccaagc	caatgtatac	accagagccc	gacatctgtc	acgagctgct
gggccacgtg 1980	cccctgttta	gcgatagatc	cttcgcccag	ttttcccagg	agatcggact
ggcatctctg 2040	ggagcacctg	acgagtacat	cgagaagctg	gccaccatct	attgggtcac
agtggagttt 2100	ggcctgtgca	agcagggcga	tagcatcaag	gcctacggag	caggactgct
gtctagcttc 2160	ggcgagctgc	agtattgtct	gtccgagaag	ccaaagctgc	tgcccctgga
gctggagaag 2220	accgccatcc	agaactacac	cgtgacagag	ttccagcccc	tgtactatgt
ggccgagtct 2280	tttaacgatg	ccaaggagaa	ggtgagaaat	ttcgccgcca	caatccctag
gcccttcagc 2340	gtgcggtacg	acccttatac	ccagaggatc	gaggtgctgg	ataatacaca
gcagctgaag 2400	atcctggctg	actcaatcaa	tagcgaaatc	ggaatcctgt	gctccgccct

gcagaaaatc aaatgaggta ccgatccaga catgataaga tacattgatg agtttggaca
2460

aaccacaact agaatgcagt gaaaaaaatg ctttatttgt gaaatttgtg atgctattgc
2520

tttatttgta accattataa gctgcaataa acaagttaac aacaacaatt gcattcattt
2580

tatgtttcag gttcaggggg aggtgtggga ggtttttttaa ggatccctgg gatgggatgt
2640

ggaatccttc tagatttctt ttgtaatat tataaagtgc tctcagcaag gtatcaaaat
2700

ggcaaaattg tgagtaacta tcctcctttc attttgggaa gaagatgagg catgaagaga
2760

attcagacag aaacttactc agaccagggg aggcagaaac taagcagaga ggaaaatgac
2820

caagagttag ccctgggcat ggaatgtgaa agaaccctaa acgtgacttg gaaataatgc
2880

ccaaggtata ttccattctc cgggatttgt tggcattttc ttgaggtgaa gaattgcaga
2940

atacattctt taatgtgacc tacatatatta cccatgggag gaagtctgct cctggactct
3000

tgagattcag tcataaagcc caggccaggg aaataatgta agtctgcagg cccctgtcat
3060

cagtaggatt agggagaaga gttctcagta gaaaacaggg aggctggaga gaaaagaatg
3120

gttaatgtta acgttaatat aactagaaag actgcagaac ttaggactga tttttatttg
3180

aatccttaaa aaaaaaaatt tcttatgaaa atagtacatg gctcttagga gacagaactt
3240

attgtacaga ggaacagcgt gagagtcaga gtgatcccag aacaggtcct ggctccatcc
3300

tgcacatagt tttggtgctg ctggcaatac ggtccccaca actgtgggaa ggggtaggg
3360

gcagggatct catcaggaaa gcataggggt ttaaagttct ttatagagca cttagaagat
3420

tgagaatcca caaattatat taataacaaa caaagtagtg tcgtgttata tagtaaattgt
3480

gaatttgcag acacatttag ggaaaagtta taattaaaaa aataggctgt atatata
3537

<210> 47

<211> 3537

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.2 vector correction genome

<400> 47

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggt cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggt acctggagca
480

tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgagggt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccaccatgg cggcggcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

aagcttctga cctcttctct tcctcccaca gggcagcgga gctactaact tcagcctgct
1020

gaagcaggct ggagacgtgg aggagaaccc tggacctatg tccaccgctg tgctggagaa
1080

ccctgggctg gggaggaaac tgtcagactt cgggcaggag acttcataca ttgaggataa
1140

ctgtaaccag aatggcgcca tctctctgat cttcagcctg aaggaggaag tgggcgccct
1200

ggcaaagggt ctgcgctgt ttgaggagaa cgacgtgaat ctgaccaca tcgagtcccg
1260

gccttctaga ctgaagaagg acgagtacga gttctttacc cacctggata agcggtcct
1320

gccagccctg acaaacatca tcaagatcct gaggcacgac atcgagagcaa ccgtgcacga
1380

gctgtctcgg gacaagaaga aggataccgt gccctggttc cctcggacaa tccaggagct
1440

ggatagattt gcccaaccaga tcctgtctta cggagcagag ctggacgcag atcacccctgg
1500

cttcaaggac ccagtgtatc gggcccggag aaagcagttt gccgatatcg cctacaatta
1560

taggcacgga cagccaatcc ctcgcgtgga gtatatggag gaggagaaga agacctgggg
1620

cacagtgttc aagaccctga agagcctgta caagacacac gcctgctacg agtataacca
1680

catcttcccc ctgctggaga agtattgtgg ctttcacgag gacaatatcc ctcagctgga
1740

ggacgtgagc cagttcctgc agacctgcac aggctttagg ctgaggccag tggcaggact
1800

gctgagctcc cgggacttcc tgggaggact ggccttcaga gtgtttcact gcacccagta
1860

catcaggcac ggctccaagc caatgtatac accagagccc gacatctgtc acgagctgct
1920

gggccacgtg cccctgttta gcgatagatc cttcgcccag ttttcccagg agatcggact
1980

ggcatctctg ggagcacctg acgagtacat cgagaagctg gccaccatct attgggtcac
2040

agtggagttt ggcctgtgca agcagggcga tagcatcaag gcctacggag caggactgct
2100

gtctagcttc ggcgagctgc agtattgtct gtccgagaag ccaaagctgc tgcccctgga
2160

gctggagaag accgccatcc agaactacac cgtgacagag ttccagcccc tgtactatgt
2220

ggccgagtct tttaacgatg ccaaggagaa ggtgagaaat ttcgccgcca caatccctag
2280

gcccttcagc gtgcggtacg acccttatac ccagaggatc gaggtgctgg ataatacaca
2340

gcagctgaag atcctggctg actcaatcaa tagcgaaatc ggaatcctgt gctccgccct
2400

gcagaaaatc aaatgaggta ccgatccaga catgataaga tacattgatg agtttggaca
2460

aaccacaact agaatgcagt gaaaaaaatg ctttatttgt gaaatttgtg atgctattgc
2520

tttatttgta accattataa gctgcaataa acaagttaac aacaacaatt gcattcattt
2580

tatgtttcag gttcaggggg aggtgtggga ggttttttta ggatccctgg gatgggatgt
2640

ggaatccttc tagatttctt ttgtaatat tataaagtgc tctcagcaag gtatcaaaat
2700

ggcaaaattg tgagtaacta tcctcctttc attttgggaa gaagatgagg catgaagaga
2760

attcagacag aaacttactc agaccagggg aggcagaaac taagcagaga ggaaaatgac
2820

caagagttag ccctgggcat ggaatgtgaa agaaccctaa acgtgacttg gaaataatgc
2880

ccaaggtata ttccattctc cgggatttgt tggcattttc ttgaggtgaa gaattgcaga
2940

atacattctt taatgtgacc tacatatatta cccatgggag gaagtctgct cctggactct
3000

tgagattcag tcataaagcc caggccaggg aaataatgta agtctgcagg cccctgtcat
3060

cagtaggatt agggagaaga gttctcagta gaaaacaggg aggctggaga gaaaagaatg
3120

gttaatgtta acgttaatat aactagaaag actgcagaac ttaggactga tttttatttg
3180

aatccttaaa aaaaaaaatt tcttatgaaa atagtacatg gctcttagga gacagaactt
3240

attgtacaga ggaacagcgt gagagtcaga gtgatcccag aacaggctct ggctccatcc
3300

tgcacatagt tttggtgctg ctggcaatac ggtccccaca actgtgggaa ggggttaggg
3360

gcagggatct catcaggaaa gcataggggt ttaaagttct ttatagagca cttagaagat
3420

tgagaatcca caaattatat taataacaaa caaagtagtg tcgtgttata tagtaaattg
3480

gaatttgcag acacatttag ggaaaagtta taattaaaaa aataggctgt atatata
3537

<210> 48

<211> 3537

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.3 vector correction genome

<400> 48

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctga gagtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgccata acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggctttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acctggagca
480

tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgagggtt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

aagcttctga cctcttctct tcctcccaca gggcagcgga gctactaact tcagcctgct
1020

gaagcaggct ggagacgtgg aggagaaccc tggacctatg tccaccgctg tgctggagaa
1080

ccctgggctg gggaggaaac tgtcagactt cgggcaggag acttcataca ttgaggataa
1140

ctgtaaccag aatggcgcca tctctctgat cttcagcctg aaggaggaag tgggcgccct
1200

ggcaaagggtg ctgcgcctgt ttgaggagaa cgacgtgaat ctgaccacaca tcgagtcccg
1260

gccttctaga ctgaagaagg acgagtacga gttctttacc cacctggata agcggtcct
1320

gccagccctg acaaacatca tcaagatcct gaggcacgac atcggagcaa ccgtgcacga
1380

gctgtctcgg gacaagaaga aggataccgt gccctggttc cctcggacaa tccaggagct
1440

ggatagattt gcccaaccaga tcctgtctta cggagcagag ctggacgcag atcacctgg
1500

cttcaaggac ccagtgtatc gggcccggag aaagcagttt gccgatatcg cctacaatta
1560

taggcacgga cagccaatcc ctgcgctgga gtatatggag gaggagaaga agacctggg
1620

cacagtgttc aagaccctga agagcctgta caagacacac gcctgctacg agtataacca
1680

catcttcccc ctgctggaga agtattgtgg ctttcacgag gacaatatcc ctcagctgga
1740

ggacgtgagc cagttcctgc agacctgcac aggctttagg ctgaggccag tggcaggact
1800

gctgagctcc cgggacttcc tgggaggact ggccttcaga gtgtttcact gcacccagta
1860

catcaggcac ggctccaagc caatgtatac accagagccc gacatctgtc acgagctgct
1920

gggccacgtg cccctgttta gcgatagatc cttcgcccag ttttcccagg agatcggact
1980

ggcatctctg ggagcacctg acgagtacat cgagaagctg gccaccatct attgggtcac
2040

agtggagttt ggctgtgca agcagggcga tagcatcaag gcctacggag caggactgct
2100

gtctagcttc ggcgagctgc agtattgtct gtccgagaag ccaaagctgc tgcccctgga
2160

gctggagaag accgccatcc agaactacac cgtgacagag ttccagcccc tgtactatgt
2220

ggccgagtct tttaacgatg ccaaggagaa ggtgagaaat ttcgccgcca caatccctag
2280

gcccttcagc gtgcggtacg acccttatac ccagaggatc gaggtgctgg ataatacaca
2340

gcagctgaag atcctggctg actcaatcaa tagcgaaatc ggaatcctgt gctccgccct
2400

gcagaaaatc aaatgaggta ccgatccaga catgataaga tacattgatg agtttggaca
2460

aaccacaact agaatgcagt gaaaaaaaaatg ctttatttgt gaaatttgtg atgctattgc
2520

tttatttgta accattataa gctgcaataa acaagttaac aacaacaatt gcattcattt
2580

tatgtttcag gttcaggggg aggtgtggga ggtttttttaa ggatccctgg gatgggatgt
2640

ggaatccttc tagattttctt ttgtaatatatt tataaagtgc tctcagcaag gtatcaaaat
2700

ggcaaaaattg tgagtaacta tcctcctttc attttgggaa gaagatgagg catgaagaga
2760

attcagacag aaacttactc agaccagggg aggcagaaac taagcagaga ggaaaatgac
2820

caagagttag ccctgggcat ggaatgtgaa agaaccctaa acgtgacttg gaaataatgc
2880

ccaaggtata ttccattctc cgggatttgt tggcattttc ttgaggtgaa gaattgcaga
2940

atacattctt taatgtgacc tacatatatta cccatgggag gaagtctgct cctggactct
3000

tgagattcag tcataaagcc caggccaggg aaataatgta agtctgcagg cccctgtcat
3060

cagtaggatt agggagaaga gttctcagta gaaaacaggg aggctggaga gaaaagaatg
3120

gttaatgtta acgttaatat aactagaaag actgcagaac ttaggactga tttttatttg
3180

aatccttaaa aaaaaaaatt tcttatgaaa atagtacatg gctcttagga gacagaactt
3240

attgtacaga ggaacagcgt gagagtcaga gtgatcccag aacaggctct ggctccatcc
3300

tgcacatagt tttggtgctg ctggcaatac ggtccccaca actgtgggaa ggggttaggg
3360

gcagggatct catcaggaaa gcataggggt tttaaagtct ttatagagca cttagaagat
3420

tgagaatcca caaattatat taataacaaa caaagtagtg tcgtgttata tagtaaagt
3480

gaatttgcag acacatttag ggaaaagtta taattaaaaa aataggctgt atatata
3537

<210> 49

<211> 3537

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.4 vector correction genome

<400> 49

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaagtaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcataataatt gttataaccag acgggtcgcag gcttagtcca attgcagaga actcgtttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggctttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acctggagca
480

tccggcaggc tggcctggcc tcctgagtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgaggtt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

aagcttctga cctctttctct tcctcccaca gggcagcgga gctactaact tcagcctgct
1020

gaagcaggct ggagacgtgg aggagaaccc tggacctatg tccaccgctg tgctggagaa
1080

ccctgggctg gggaggaaac tgtcagactt cgggcaggag acttcataca ttgaggataa
1140

ctgtaaccag aatggcgcca tctctctgat cttcagcctg aaggaggaag tgggcgccct
1200

ggcaaagggtg ctgcgccctgt ttgaggagaa cgacgtgaat ctgaccacaca tcgagtcccg
1260

gccttctaga ctgaagaagg acgagtacga gttctttacc cacctggata agcggtcctt
1320

gccagccctg acaaacatca tcaagatcct gaggcacgac atcggagcaa ccgtgcacga
1380

gctgtctcgg gacaagaaga aggataaccgt gccctggttc cctcggacaa tccaggagct
1440

ggatagattt gcccaaccaga tcctgtctta cggagcagag ctggacgcag atcaccttgg
1500

cttcaaggac ccagtgtatc gggcccggag aaagcagttt gccgatatcg cctacaatta
1560

taggcacgga cagccaatcc ctcgcgtgga gtatatggag gaggagaaga agacctgggg
1620

cacagtgttc aagaccctga agagcctgta caagacacac gcctgctacg agtataacca
1680

catcttcccc ctgctggaga agtattgtgg ctttcacgag gacaatatcc ctcagctgga
1740

ggacgtgagc cagttcctgc agacctgcac aggctttagg ctgaggccag tggcaggact
1800

gctgagctcc cgggacttcc tgggaggact ggccttcaga gtgttttact gcacccagta
1860

catcaggcac ggctccaagc caatgtatac accagagccc gacatctgtc acgagctgct
1920

gggccacgtg cccctgttta gcgatagatc cttcgcccag ttttcccagg agatcggact
1980

ggcatctctg ggagcacctg acgagtacat cgagaagctg gccaccatct attgggtcac
2040

agtggagttt 2100	ggcctgtgca	agcagggcga	tagcatcaag	gcctacggag	caggactgct
gtctagcttc 2160	ggcgagctgc	agtattgtct	gtccgagaag	ccaaagctgc	tgcccctgga
gctggagaag 2220	accgccatcc	agaactacac	cgtgacagag	ttccagcccc	tgtactatgt
ggccgagtct 2280	tttaacgatg	ccaaggagaa	ggtgagaaat	ttcgccgcca	caatccctag
gcccttcagc 2340	gtgcggtacg	acccttatac	ccagaggatc	gaggtgctgg	ataatacaca
gcagctgaag 2400	atcctggctg	actcaatcaa	tagcgaaatc	ggaatcctgt	gctccgccct
gcagaaaatc 2460	aaatgaggta	ccgatccaga	catgataaga	tacattgatg	agtttggaca
aaccacaact 2520	agaatgcagt	gaaaaaaaaatg	ctttattttgt	gaaattttgtg	atgctattgc
tttattttgta 2580	accattataa	gctgcaataa	acaagttaac	aacaacaatt	gcattcattt
tatgtttcag 2640	gttcaggggg	aggtgtggga	ggtttttttaa	ggatccctgg	gatgggatgt
ggaatccttc 2700	tagatttcct	ttgtaatatt	tataaagtgc	tctcagcaag	gtatcaaaat
ggcaaaaattg 2760	tgagtaacta	tcctcctttc	attttgggaa	gaagatgagg	catgaagaga
attcagacag 2820	aaacttactc	agaccagggg	aggcagaaac	taagcagaga	ggaaaatgac
caagagttag 2880	ccctggggcat	ggaatgtgaa	agaaccctaa	acgtgacttg	gaaataatgc
ccaaggtata 2940	ttccattctc	cgggattttgt	tggcattttc	ttgaggtgaa	gaattgcaga

atacattcctt taatgtgacc tacatatatta cccatgggag gaagtctgct cctggactct
3000

tgagattcag tcataaagcc caggccaggg aaataatgta agtctgcagg cccctgtcat
3060

cagtaggatt agggagaaga gttctcagta gaaaacaggg aggctggaga gaaaagaatg
3120

gttaatgtta acgttaatat aactagaaag actgcagaac ttaggactga tttttatttg
3180

aatccttaaa aaaaaaaatt tcttatgaaa atagtacatg gctcttagga gacagaactt
3240

attgtacaga ggaacagcgt gagagtcaga gtgatcccag aacaggctct ggctccatcc
3300

tgcacatagt tttggtgctg ctggcaatac ggtccccaca actgtgggaa ggggttaggg
3360

gcagggatct catcaggaaa gcataggggt tttaaagttct ttatagagca cttagaagat
3420

tgagaatcca caaattatat taataacaaa caaagtagtg tcgtgttata tagtaaattg
3480

gaatttgcag acacatttag ggaaaagtta taattaaaaa aataggctgt atatata
3537

<210> 50

<211> 3537

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.5 vector correction genome

<400> 50

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagaggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggctttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggctttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acgtcaagca
480

tccggcaggc tggcctggcc tcctgagtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgagggtt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

aagcttctga cctcttctct tcctcccaca gggcagcgga gctactaact tcagcctgct
1020

gaagcaggct ggagacgtgg aggagaaccc tggacctatg tccaccgctg tgctggagaa
1080

ccctgggctg gggaggaaac tgtcagactt cgggcaggag acttcataca ttgaggataa
1140

ctgtaaccag aatggcgcca tctctctgat cttcagcctg aaggaggaag tgggcgccct
1200

ggcaaagggtg ctgcgccctgt ttgaggagaa cgacgtgaat ctgaccacaca tcgagtcccg
1260

gccttctaga ctgaagaagg acgagtacga gttctttacc cacctggata agcggtcctt
1320

gccagccctg acaaacatca tcaagatcct gaggcacgac atcggagcaa ccgtgcacga
1380

gctgtctcgg gacaagaaga aggataccgt gccctggttc cctcggacaa tccaggagct
1440

ggatagattt gcccaaccaga tcctgtctta cggagcagag ctggacgcag atcacctgg
1500

cttcaaggac ccagtgtatc gggcccggag aaagcagttt gccgatatcg cctacaatta
1560

taggcacgga cagccaatcc ctcgcgtgga gtatatggag gaggagaaga agacctgggg
1620

cacagtgttc aagaccctga agagcctgta caagacacac gcctgctacg agtataacca
1680

catcttcccc ctgctggaga agtattgtgg ctttcacgag gacaatatcc ctcagctgga
1740

ggacgtgagc cagttcctgc agacctgcac aggctttagg ctgaggccag tggcaggact
1800

gctgagctcc cgggacttcc tgggaggact ggccttcaga gtgttttact gcaccagta
1860

catcaggcac ggctccaagc caatgtatac accagagccc gacatctgtc acgagctgct
1920

gggccacgtg cccctgttta gcgatagatc cttcgcccag ttttcccagg agatcggact
1980

ggcatctctg ggagcacctg acgagtacat cgagaagctg gccaccatct attgggtcac
2040

agtggagttt ggcctgtgca agcagggcga tagcatcaag gcctacggag caggactgct
2100

gtctagcttc ggcgagctgc agtattgtct gtccgagaag ccaaagctgc tgcccctgga
2160

gctggagaag accgccatcc agaactacac cgtgacagag ttccagcccc tgtactatgt
2220

ggccgagtct tttaacgatg ccaaggagaa ggtgagaaat ttcgccgcca caatccctag
2280

gcccttcagc gtgcggtacg acccttatac ccagaggatc gaggtgctgg ataatacaca
2340

gcagctgaag atcctggctg actcaatcaa tagcgaaatc ggaatcctgt gctccgccct
2400

gcagaaaatc aaatgaggta ccgatccaga catgataaga tacattgatg agtttggaca
2460

aaccacaact agaatgcagt gaaaaaaaaatg ctttatttgt gaaatttgtg atgctattgc
2520

tttatttgta accattataa gctgcaataa acaagttaac aacaacaatt gcattcattt
2580

tatgtttcag gttcaggggg aggtgtggga ggtttttttaa ggatccctgg gatgggatgt
2640

ggaatccttc tagatttctt ttgtaatat tataaagtgc tctcagcaag gtatcaaaat
2700

ggcaaaattg tgagtaacta tcctcctttc attttgggaa gaagatgagg catgaagaga
2760

attcagacag aaacttactc agaccagggg aggcagaaac taagcagaga ggaaaatgac
2820

caagagttag ccctgggcat ggaatgtgaa agaaccctaa acgtgacttg gaaataatgc
2880

ccaaggtata ttccattctc cgggatttgt tggcattttc ttgaggtgaa gaattgcaga
2940

atacattctt taatgtgacc tacatatatta cccatgggag gaagtctgct cctggactct
3000

tgagattcag tcataaagcc caggccaggg aaataatgta agtctgcagg cccctgtcat
3060

cagtaggatt agggagaaga gttctcagta gaaaacaggg aggctggaga gaaaagaatg
3120

gttaatgtta acgttaatat aactagaaag actgcagaac ttaggactga tttttatttg
3180

aatccttaaa aaaaaaaatt tcttatgaaa atagtacatg gctcttagga gacagaactt
3240

attgtacaga ggaacagcgt gagagtcaga gtgatcccag aacaggctcct ggctccatcc
3300

tgacacatagt tttggtgctg ctggcaatac ggtccccaca actgtgggaa ggggttaggg
3360

gcagggatct catcaggaaa gcataggggt ttaaagttct ttatagagca cttagaagat
3420

tgagaatcca caaattatat taataacaaa caaagtagtg tcgtggtata tagtaaattg
3480

gaatttgcag acacatttag ggaaaagtta taattaaaaa aataggctgt atatata
3537

<210> 51

<211> 3537

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.6 vector correction genome

<400> 51

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acctggagca
480

tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgaggtt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg ttccataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaatg tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

aagcttctga cctctttctc tcctcccaca gggcagcgga gctactaact tcagcctgct
1020

gaagcaggct ggagacgtgg aggagaaccc tggacctatg tccaccgctg tgctggagaa
1080

ccctgggctg gggaggaaac tgtcagactt cgggcaggag acttcataca ttgaggataa
1140

ctgtaaccag aatggcgcca tctctctgat cttcagcctg aaggaggaag tgggcgccct
1200

ggcaaagggtg ctgcgcctgt ttgaggagaa cgacgtgaat ctgaccacaca tcgagtcccg
1260

gccttctaga ctgaagaagg acgagtacga gttctttacc cacctggata agcggtcctt
1320

gccagccctg acaaacatca tcaagatcct gaggcacgac atcggagcaa ccgtgcacga
1380

gctgtctcgg gacaagaaga aggataccgt gccctggttc cctcggacaa tccaggagct
1440

ggatagattt gcccaaccaga tcctgtctta cggagcagag ctggacgcag atcaccctgg
1500

cttcaaggac ccagtgtatc gggcccggag aaagcagttt gccgatatcg cctacaatta
1560

taggcacgga cagccaatcc ctcgcgtgga gtatatggag gaggagaaga agacctgggg
1620

cacagtgttc aagaccctga agagcctgta caagacacac gcctgctacg agtataacca
1680

catcttcccc ctgctggaga agtattgtgg ctttcacgag gacaatatcc ctcagctgga
1740

ggacgtgagc cagttcctgc agacctgcac aggctttagg ctgaggccag tggcaggact
1800

gctgagctcc 1860	cgggacttcc	tgggaggact	ggccttcaga	gtgtttcact	gcacccagta
catcagggcac 1920	ggctccaagc	caatgtatac	accagagccc	gacatctgtc	acgagctgct
gggccacgtg 1980	cccctgttta	gcgatagatc	cttcgcccag	ttttcccagg	agatcggact
ggcatctctg 2040	ggagcacctg	acgagtacat	cgagaagctg	gccaccatct	attgggtcac
agtggagttt 2100	ggcctgtgca	agcagggcga	tagcatcaag	gcctacggag	caggactgct
gtctagcttc 2160	ggcgagctgc	agtattgtct	gtccgagaag	ccaaagctgc	tgcccctgga
gctggagaag 2220	accgccatcc	agaactacac	cgtgacagag	ttccagcccc	tgtactatgt
ggccgagtct 2280	tttaacgatg	ccaaggagaa	ggtgagaaat	ttcgccgcca	caatccctag
gcccttcagc 2340	gtgcggtacg	acccttatac	ccagaggatc	gaggtgctgg	ataatacaca
gcagctgaag 2400	atcctggctg	actcaatcaa	tagcgaaatc	ggaatcctgt	gctccgccct
gcagaaaatc 2460	aaatgaggta	ccgatccaga	catgataaga	tacattgatg	agtttggaca
aaccacaact 2520	agaatgcagt	gaaaaaaaaatg	ctttattttgt	gaaattttgtg	atgctattgc
tttattttgta 2580	accattataa	gctgcaataa	acaagttaac	aacaacaatt	gcattcattt
tatgtttcag 2640	gttcaggggg	aggtgtggga	ggtttttttaa	ggatccctgg	gatgggatgt
ggaatccttc 2700	tagattttctt	ttgtaatat	tataaagtgc	tctcagcaag	gtatcaaaaat

ggcaaaattg tgagtaacta tcctcctttc attttgggaa gaagatgagg catgaagaga
2760

attcagacag aaacttactc agaccagggg aggcagaaac taagcagaga ggaaaatgac
2820

caagagttag ccctgggcat ggaatgtgaa agaaccctaa acgtgacttg gaaataatgc
2880

ccaaggtata ttccattctc cgggatttgt tggcattttc ttgaggtgaa gaattgcaga
2940

atacattctt taatgtgacc tacatatatta cccatgggag gaagtctgct cctggactct
3000

tgagattcag tcataaagcc caggccaggg aaataatgta agtctgcagg cccctgtcat
3060

cagtaggatt agggagaaga gttctcagta gaaaacaggg aggctggaga gaaaagaatg
3120

gttaatgtta acgttaatat aactagaaag actgcagaac ttaggactga tttttatttg
3180

aatccttaaa aaaaaaaatt tcttatgaaa atagtacatg gctcttagga gacagaactt
3240

attgtacaga ggaacagcgt gagagtcaga gtgatcccag aacaggctct ggctccatcc
3300

tgcacatagt tttggtgctg ctggcaatac ggtccccaca actgtgggaa ggggttaggg
3360

gcagggatct catcaggaaa gcataggggt tttaaagttct ttatagagca cttagaagat
3420

tgagaatcca caaattatat taataacaaa caaagtagtg tcgtgttata tagtaaattg
3480

gaatttgcag acacatttag ggaaaagtta taattaaaaa aataggctgt atatata
3537

<211> 3537

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.7 vector correction genome

<400> 52

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagaggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttccaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acctggagca
480

tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgagggt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccaactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

aagcttctga cctcttctct tcctcccaca gggcagcggg gctactaact tcagcctgct
1020

gaagcaggct ggagacgtgg aggagaaccc tggacctatg tccaccgctg tgctggagaa
1080

ccctgggctg gggaggaaac tgtcagactt cgggcaggag acttcataca ttgaggataa
1140

ctgtaaccag aatggcgcca tctctctgat cttcagcctg aaggaggaag tgggcgccct
1200

ggcaaagggtg ctgcgccctgt ttgaggagaa cgacgtgaat ctgaccacaca tcgagtcccg
1260

gccttctaga ctgaagaagg acgagtacga gttctttacc cacctggata agcggtcctt
1320

gccagccctg acaaacatca tcaagatcct gaggcacgac atcggagcaa ccgtgcacga
1380

gctgtctcgg gacaagaaga aggataaccgt gccctgggtc cctcggacaa tccaggagct
1440

ggatagattt gcccaaccaga tcctgtctta cggagcagag ctggacgcag atcaccctgg
1500

cttcaaggac ccagtgtatc gggcccggag aaagcagttt gccgatatcg cctacaatta
1560

taggcacgga cagccaatcc ctcgcgtgga gtatatggag gaggagaaga agacctgggg
1620

cacagtgttc aagaccctga agagcctgta caagacacac gcctgctacg agtataacca
1680

catcttcccc ctgctggaga agtattgtgg ctttcacgag gacaatatcc ctcagctgga
1740

ggacgtgagc cagttcctgc agacctgcac aggctttagg ctgaggccag tggcaggact
1800

gctgagctcc cgggacttcc tgggaggact ggccttcaga gtgtttcact gcacccagta
1860

catcaggcac ggctccaagc caatgtatac accagagccc gacatctgtc acgagctgct
1920

gggccacgtg cccctgttta gcgatagatc cttcgcccag ttttcccagg agatcggact
1980

ggcatctctg ggagcacctg acgagtacat cgagaagctg gccaccatct attggttcac
2040

agtggagttt ggcctgtgca agcagggcga tagcatcaag gcctacggag caggactgct
2100

gtctagcttc ggcgagctgc agtattgtct gtccgagaag ccaaagctgc tgcccctgga
2160

gctggagaag accgccatcc agaactacac cgtgacagag ttccagcccc tgtactatgt
2220

ggccgagtct tttaacgatg ccaaggagaa ggtgagaaat ttcgccgcca caatccctag
2280

gcccttcagc gtgcggtacg acccttatac ccagaggatc gaggtgctgg ataatacaca
2340

gcagctgaag atcctggctg actcaatcaa tagcgaaatc ggaatcctgt gctccgccct
2400

gcagaaaatc aaatgaggta ccgatccaga catgataaga tacattgatg agtttggaca
2460

aaccacaact agaatgcagt gaaaaaaatg ctttatttgt gaaatttgtg atgctattgc
2520

tttatttgta accattataa gctgcaataa acaagttaac aacaacaatt gcattcattt
2580

tatgtttcag gttcaggggg aggtgtggga ggtttttttaa ggatccctgg gatgggatgt
2640

ggaatccttc tagatttctt ttgtaatat tataaagtgc tctcagcaag gtatcaaaat
2700

ggcaaaattg tgagtaacta tcctcctttc attttgggaa gaagatgagg catgaagaga
2760

attcagacag aaacttactc agaccagggg aggcagaaac taagcagaga ggaaaatgac
2820

caagagttag ccctgggcat ggaatgtgaa agaaccctaa acgtgacttg gaaataatgc
2880

ccaaggtata ttccattctc cgggatttgt tggcattttc ttgaggtgaa gaattgcaga
2940

atacattctt taatgtgacc tacatatatta cccatgggag gaagtctgct cctggactct
3000

tgagattcag tcataaagcc caggccaggg aaataatgta agtctgcagg cccctgtcat
3060

cagtaggatt agggagaaga gttctcagta gaaaacaggg aggctggaga gaaaagaatg
3120

gttaatgtta acgttaatat aactagaaag actgcagaac ttaggactga tttttatttg
3180

aatccttaaa aaaaaaaatt tcttatgaaa atagtacatg gctcttagga gacagaactt
3240

attgtacaga ggaacagcgt gagagtcaga gtgatcccag aacaggctct ggctccatcc
3300

tgcacatagt tttggtgctg ctggcaatac ggtccccaca actgtgggaa ggggttaggg
3360

gcagggatct catcaggaaa gcataggggt tttaaagttct ttatagagca cttagaagat
3420

tgagaatcca caaattatat taataacaaa caaagtagtg tcgtgttata tagtaaattgt
3480

gaatttgcag acacatttag ggaaaagtta taattaaaaa aataggctgt atatata
3537

<210> 53

<211> 3545

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.8 vector correction genome

<400> 53

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttccaccag acggtcgcag gcttagtcca attgcagaga acacgctggt
300

cttcgccccca ggcttctgag agtcccggaa gtgcctaaac ctgtctaate gacgggggctt
360

gggtggccccg tcgctccctg gcttcttccc tttaccagg gcgggcagcg aagtggtgcc
420

tcctgcgtcc cccacaccct ccctcagccc ctcccctccg gcccgtcctg ggcagggtgac
480

ctggagcatc cggcaggctg ccctggcctc ctgcgtcagg acaacgcccc cgagggggcgt
540

tactgtgcgg agatgcacca cgcaagagac accctttgta actctcttct cctccctagt
600

gcgagggttaa aaccttcagc cccacgtgct gtttgcaaac ctgcctgtac ctgaggccct
660

aaaaagccag agacctcact cccgggggagc cagcatgtcc actgcggtcc tggaaaaccc
720

aggcttgggc aggaaactct ctgacttttg acaggtgagc cacggcagcc tgagctgctc
780

agttagggga atttgggcct ccagagaaaag agatccgaag actgctggtg cttcctggtt
840

tcataagctc agtaagaagt ctgaattcgt tggaagctga tgagaatatc caggaagtca
900

acagacaaat gtcctcaaca attgtttcta agtaggagaa catctgtcct cggtggcttt
960

cacaggaaaa gcttctgacc tcttctcttc ctcccacagg gcagcggagc tactaacttc
1020

agcctgctga agcaggctgg agacgtggag gagaaccctg gacctatgtc caccgctgtg
1080

ctggagaacc ctgggctggg gaggaaactg tcagacttcg ggcaggagac ttcatacatt
1140

gaggataact gtaaccagaa tggcgccatc tctctgatct tcagcctgaa ggaggaagtg
1200

ggcgccctgg caaagggtgct gcgcctgttt gaggagaacg acgtgaatct gacccacatc
1260

gagtcccggc cttctagact gaagaaggac gagtacgagt tctttaccca cctggataag
1320

cggtccctgc cagccctgac aaacatcatc aagatcctga ggcacgacat cggagcaacc
1380

gtgcacgagc tgtctcggga caagaagaag gataccgtgc cctggttccc tcggacaatc
1440

caggagctgg atagatttgc caaccagatc ctgtcttacg gagcagagct ggacgcagat
1500

caccctggct tcaaggaccc agtgtatcgg gcccgagaaa agcagtttgc cgatatcgcc
1560

tacaattata ggcacggaca gccaatccct cgcgtggagt atatggagga ggagaagaag
1620

acctggggca cagtgttcaa gaccctgaag agcctgtaca agacacacgc ctgctacgag
1680

tataaccaca tcttccccct gctggagaag tattgtggct ttcacgagga caatatccct
1740

cagctggagg acgtgagcca gttcctgcag acctgcacag gctttaggct gaggccagtg
1800

gcaggactgc tgagctcccg ggacttcctg ggaggactgg ccttcagagt gtttcactgc
1860

accagttaca tcaggcacgg ctccaagcca atgtatacac cagagcccga catctgtcac
1920

gagctgctgg gccacgtgcc cctgttttagc gatagatcct tcgcccagtt ttcccaggag
1980

atcggactgg catctctggg agcacctgac gagtacatcg agaagctggc caccatctat
2040

tggttcacag tggagtttgg cctgtgcaag cagggcgata gcatcaaggc ctacggagca
2100

ggactgctgt ctagcttcgg cgagctgcag tattgtctgt ccgagaagcc aaagctgctg
2160

cccctggagc tggagaagac cgccatccag aactacaccg tgacagagtt ccagcccctg
2220

tactatgtgg ccgagtcctt taacgatgcc aaggagaagg tgagaaattt cgccgccaca
2280

atccctaggc ccttcagcgt gcggtacgac ccttataccc agaggatcga ggtgctggat
2340

aatacacagc agctgaagat cctggctgac tcaatcaata gcgaaatcgg aatcctgtgc
2400

tccgccctgc agaaaatcaa atgaggtacc gatccagaca tgataagata cattgatgag
2460

tttggacaaa ccacaactag aatgcagtga aaaaaatgct ttattttgtga aattttgtgat
2520

gctattgctt tattttgtaac cattataagc tgcaataaac aagttaacaa caacaattgc
2580

attcatttta tgtttcaggt tcagggggag gtgtgggagg ttttttaagg atccctggga
2640

tgggatgtgg aatccttcta gatttctttt gtaatattta taaagtgctc tcagcaaggt
2700

atcaaaatgg caaaattgtg agtaactatc ctcctttcat tttgggaaga agatgaggca
2760

tgaagagaat tcagacagaa acttactcag accaggggag gcagaaacta agcagagagg
2820

aaaatgacca agagttagcc ctgggcatgg aatgtgaaag aaccctaaac gtgacttgga
2880

aataatgccc aaggtatatt ccattctccg ggattttgttgc gatttttctt gaggtgaaga
2940

attgcagaat acattcttta atgtgaccta catattttacc catgggagga agtctgctcc
3000

tggactcttg agattcagtc ataaagccca ggccagggaataaatgtaag tctgcaggcc
3060

cctgtcatca gtaggattag ggagaagagt tctcagtaga aaacagggag gctggagaga
3120

aaagaatggg taatgttaac gttaatatataa ctagaaagac tgcagaactt aggactgatt
3180

tttatttgaa tccttaaaaa aaaaaatttc ttatgaaaat agtacatggc tcttaggaga
3240

cagaacttat tgtacagagg aacagcgtga gagtcagagt gatcccagaa caggtcctgg
3300

ctccatcctg cacatagttt tgggtgctgct ggcaatacgg tccccacaac tgtgggaagg
3360

ggttaggggc agggatctca tcaggaaagc atagggggtt aaagttcttt atagagcact
3420

tagaagattg agaatccaca aattatatta ataacaaaca aagtagtgtc gtgttatata
3480

gtaaatgtga atttgcagac acatttaggg aaaagttata attaaaaaaa taggctgtat
3540

a t a t a
3545

<210> 54

<211> 3537

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.9 vector correction genome

<400> 54

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctga gagtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacagagc ttgggtggcc
360

cgtcgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc taggcagggtg acctgaagca
480

tccagcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgagggt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

aagcttctga cctcttctct tcctcccaca gggcagcgga gctactaact tcagcctgct
1020

gaagcaggct ggagacgtgg aggagaaccc tggacctatg tccaccgctg tgctggagaa
1080

ccctgggctg gggaggaaac tgtcagactt cgggcaggag acttcataca ttgaggataa
1140

ctgtaaccag aatggcgcca tctctctgat cttcagcctg aaggaggaag tgggcgccct
1200

ggcaaagggtg ctgcgctgtt ttgaggagaa cgacgtgaat ctgaccacaca tcgagtcccg
1260

gccttctaga ctgaagaagg acgagtacga gttctttacc cacctggata agcggtcct
1320

gccagccctg acaaacatca tcaagatcct gaggcacgac atcgagagcaa ccgtgcacga
1380

gctgtctcgg gacaagaaga aggataccgt gccctggttc cctcggacaa tccaggagct
1440

ggatagattt gcccaaccaga tcctgtctta cggagcagag ctggacgcag atcacccctgg
1500

cttcaaggac ccagtgtatc gggcccggag aaagcagttt gccgatatcg cctacaatta
1560

taggcacgga cagccaatcc ctcgcgtgga gtatatggag gaggagaaga agacctgggg
1620

cacagtgttc aagaccctga agagcctgta caagacacac gcctgctacg agtataacca
1680

catcttcccc ctgctggaga agtattgtgg ctttcacgag gacaatatcc ctcagctgga
1740

ggacgtgagc cagttcctgc agacctgcac aggctttagg ctgaggccag tggcaggact
1800

gctgagctcc cgggacttcc tgggaggact ggccttcaga gtgtttcact gcacccagta
1860

catcaggcac ggctccaagc caatgtatac accagagccc gacatctgtc acgagctgct
1920

gggccacgtg cccctgttta gcgatagatc cttcgcccag ttttcccagg agatcggact
1980

ggcatctctg ggagcacctg acgagtacat cgagaagctg gccaccatct attgggtcac
2040

agtggagttt ggcctgtgca agcagggcga tagcatcaag gcctacggag caggactgct
2100

gtctagcttc ggcgagctgc agtattgtct gtccgagaag ccaaagctgc tgcccctgga
2160

gctggagaag accgccatcc agaactacac cgtgacagag ttccagcccc tgtactatgt
2220

ggccgagtct tttaacgatg ccaaggagaa ggtgagaaat ttcgccgcca caatccctag
2280

gcccttcagc gtgcggtacg acccttatac ccagaggatc gaggtgctgg ataatacaca
2340

gcagctgaag atcctggctg actcaatcaa tagcgaaatc ggaatcctgt gctccgccct
2400

gcagaaaatc aaatgaggta ccgatccaga catgataaga tacattgatg agtttggaca
2460

aaccacaact agaatgcagt gaaaaaaatg ctttatttgt gaaatttgtg atgctattgc
2520

tttatttgta accattataa gctgcaataa acaagttaac aacaacaatt gcattcattt
2580

tatgtttcag gttcaggggg aggtgtggga ggttttttta ggatccctgg gatgggatgt
2640

ggaatccttc tagatttctt ttgtaatat tataaagtgc tctcagcaag gtatcaaaat
2700

ggcaaaattg tgagtaacta tcctcctttc attttgggaa gaagatgagg catgaagaga
2760

attcagacag aaacttactc agaccagggg aggcagaaac taagcagaga ggaaaatgac
2820

caagagttag ccctgggcat ggaatgtgaa agaaccctaa acgtgacttg gaaataatgc
2880

ccaaggtata ttccattctc cgggatttgt tggcattttc ttgaggtgaa gaattgcaga
2940

atacattctt taatgtgacc tacatatatta cccatgggag gaagtctgct cctggactct
3000

tgagattcag tcataaagcc caggccaggg aaataatgta agtctgcagg cccctgtcat
3060

cagtaggatt agggagaaga gttctcagta gaaaacaggg aggctggaga gaaaagaatg
3120

gttaatgtta acgttaatat aactagaaag actgcagaac ttaggactga tttttatttg
3180

aatccttaaa aaaaaaaatt tcttatgaaa atagtacatg gctcttagga gacagaactt
3240

attgtacaga ggaacagcgt gagagtcaga gtgatcccag aacaggctct ggctccatcc
3300

tgcacatagt tttggtgctg ctggcaatac ggtccccaca actgtgggaa ggggttaggg
3360

gcagggatct catcaggaaa gcataggggt ttaaagttct ttatagagca cttagaagat
3420

tgagaatcca caaattatat taataacaaa caaagtagtg tcgtgttata tagtaaattg
3480

gaatttgcag acacatttag ggaaaagtta taattaaaaa aataggctgt atatata
3537

<210> 55

<211> 3943

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.1 vector correction genome (+ ITRs)

<400> 55

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgagc agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgct tcaggagcag ttgtgcgaat agctggagaa
240

caccaggctg gatttaaacc cagatcgctc ttacatttgc tctttacctg ctgtgctcag
300

cgttcacgtg ccctctagct gtagttttct gaagtcagcg cacagcaagg cagtgtgctt
360

agagggttaac agaaggggaaa acaacaacaa caaaaatcta aatgagaatc ctgactgttt
420

cagctggggg taaggggggc ggattattca tataattgtt ataccagacg gtcgcaggct
480

tagtccaatt gcagagaact cgcttcccag gcttctgaga gtcccggaag tgcctaaacc
540

tgtctaatacg acgggggcttg ggtggcccgt cgctccctgg cttcttccct ttacccaggg
600

cgggcagcga agtggtgcct cctgcgtccc ccacaccctc cctcagcccc tcccctccgg
660

cccgtcctgg gcaggtgacc tggagcatcc ggcaggctgc cctggcctcc tgcgtcagga
720

caacgcccac gaggggcggt actgtgcgga gatgcaccac gcaagagaca ccctttgtaa
780

ctctcttctc ctccctagtgc cgagggttaa accttcagcc ccacgtgctg tttgcaaacc
840

tgctgttacc tgaggcccta aaaagccaga gacctcactc ccggggagcc agcatgtcca
900

ctgcggtcct ggaaaaccca ggcttgggca ggaaactctc tgactttgga cagggtgagcc
960

acggcagcct gagctgctca gttaggggaa tttgggcctc cagagaaaga gatccgaaga
1020

ctgctggtgc ttcctgggtt cataagctca gtaagaagtc tgaattcggt ggaagctgat
1080

gagaatatcc aggaagtcaa cagacaaatg tcctcaacaa ttgtttctaa gtaggagaac
1140

atctgtcctc ggtggctttc acaggaaaag cttctgacct cttctcttcc tcccacaggg
1200

cagcggagct actaacttca gcctgctgaa gcaggctgga gacgtggagg agaaccctgg
1260

acctatgtcc 1320	accgctgtgc	tggagaaccc	tgggctgggg	aggaaactgt	cagacttcgg
gcaggagact 1380	tcatacattg	aggataactg	taaccagaat	ggcgccatct	ctctgatctt
cagcctgaag 1440	gaggaagtgg	gcgccctggc	aaaggtgctg	cgcttgtttg	aggagaacga
cgtgaatctg 1500	acccacatcg	agtcccggcc	ttctagactg	aagaaggacg	agtacgagtt
ctttaccac 1560	ctggataagc	ggtccctgcc	agccctgaca	aacatcatca	agatcctgag
gcacgacatc 1620	ggagcaaccg	tgcacgagct	gtctcgggac	aagaagaagg	ataccgtgcc
ctggttcct 1680	cggacaatcc	aggagctgga	tagatttgcc	aaccagatcc	tgtcttacgg
agcagagctg 1740	gacgcagatc	accctggctt	caaggacca	gtgtatcggg	cccggagaaa
gcagtttgcc 1800	gatatcgctt	acaattatag	gcacggacag	ccaatccctc	gcgtggagta
tatggaggag 1860	gagaagaaga	cctggggcac	agtgttcaag	accctgaaga	gcctgtacaa
gacacacgcc 1920	tgctacgagt	ataaccacat	cttccccctg	ctggagaagt	attgtggctt
tcacgaggac 1980	aatatccctc	agctggagga	cgtgagccag	ttcctgcaga	cctgcacagg
ctttaggctg 2040	aggccagtgg	caggactgct	gagctcccgg	gacttcctgg	gaggactggc
cttcagagtg 2100	tttcactgca	cccagtacat	caggcacggc	tccaagccaa	tgtatacacc
agagcccgc 2160	atctgtcacg	agctgctggg	ccacgtgccc	ctgttttagcg	atagatcctt

cgcccagttt 2220	tcccaggaga	tcggactggc	atctctggga	gcacctgacg	agtacatcga
gaagctggcc 2280	accatctatt	ggttcacagt	ggagtttggc	ctgtgcaagc	agggcgatag
catcaaggcc 2340	tacggagcag	gactgctgtc	tagcttcggc	gagctgcagt	attgtctgtc
cgagaagcca 2400	aagctgctgc	ccctggagct	ggagaagacc	gccatccaga	actacaccgt
gacagagttc 2460	cagcccctgt	actatgtggc	cgagtctttt	aacgatgcca	aggagaaggt
gagaaatttc 2520	gccgccacaa	tccctaggcc	cttcagcgtg	cggtagcacc	cttataccca
gaggatcgag 2580	gtgctggata	atacacagca	gctgaagatc	ctggctgact	caatcaatag
cgaaatcgga 2640	atcctgtgct	ccgccctgca	gaaaatcaaa	tgaggtagcg	atccagacat
gataagatac 2700	attgatgagt	ttggacaaac	cacaactaga	atgcagtgaa	aaaaatgctt
tatttgtgaa 2760	atttgtgatg	ctattgcttt	atttghtaacc	attataagct	gcaataaaca
agttaacaac 2820	aacaattgca	ttcattttat	gtttcagggt	cagggggagg	tgtgggaggt
tttttaagga 2880	tccctgggat	gggatgtgga	atccttctag	atttcttttg	taatatttat
aaagtgctct 2940	cagcaaggta	tcaaaatggc	aaaattgtga	gtaactatcc	tcctttcatt
ttgggaagaa 3000	gatgaggcat	gaagagaatt	cagacagaaa	cttactcaga	ccaggggagg
cagaaactaa 3060	gcagagagga	aaatgaccaa	gagttagccc	tgggcatgga	atgtgaaaga

accctaaacg tgacttggaataaatgccca aggtatatattc cattctccgg gatttggttg
3120

cattttcttg aggtgaagaa ttgcagaata cattctttaa tgtgacctac atatttacct
3180

atgggaggaa gtctgctcct ggactcttga gattcagtca taaagcccag gccagggaaa
3240

taatgtaagt ctgcaggccc ctgtcatcag taggattagg gagaagagtt ctcagtagaa
3300

aacagggagg ctggagagaa aagaatggtt aatgttaacg ttaatataac tagaaagact
3360

gcagaactta ggactgattt ttatttgaat ccttaaaaaa aaaaatttct tatgaaaata
3420

gtacatggct cttaggagac agaacttatt gtacagagga acagcgtgag agtcagagtg
3480

atcccagaac aggtcctggc tccatcctgc acatagtttt ggtgctgctg gcaatacgg
3540

ccccacaact gtgggaaggg gttaggggca gggatctcat caggaaagca taggggttta
3600

aagtcttcta tagagcactt agaagattga gaatccacaa attatattaa taacaaacaa
3660

agtagtgtcg tggtatatag taaatgtgaa ttgacagaca catttaggga aaagttataa
3720

ttaaaaaaat aggctgtata tatacctgca ggtctagata cgtagataag tagcatggcg
3780

ggttaatcat taactacaag gaaccctag tgatggagtt ggccactccc tctctgcgcg
3840

ctcgctcgct cactgaggcc gggcgaccaa aggtcgcccg acgcccgggc ttgcccggg
3900

cggcctcagt gagcgagcga gcgcgcagag agggagtggc caa
3943

<210> 56
<211> 3943
<212> DNA
<213> Artificial Sequence

<220>
<223> 32.2 vector correction genome (+ ITRs)

<400> 56
ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgagc agagggagtg
120

gccaactcca tcactagggg ttccctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgct tcaggagcag ttgtgcgaat agctggagaa
240

caccaggctg gatttaaacc cagatcgctc ttacatttgc tctttacctg ctgtgctcag
300

cgttcacgtg ccctctagct gtagttttct gaagtcagcg cacagcaagg cagtgtgctt
360

agaggttaac agaagggaaa acaacaacaa caaaaatcta aatgagaatc ctgactgttt
420

cagctggggg taaggggggc ggattattca tataattggt ataccagacg gtcgcaggct
480

tagtccaatt gcagagaact cgcttcccag gcttctgaga gtcccggaag tgcctaaacc
540

tgtctaatacg acgggggcttg ggtggcccgt cgctccctgg cttcttccct ttaccaggg
600

cgggcagcga agtggtgcct cctgcgtccc ccacaccctc cctcagcccc tcccctccgg
660

cccgtcctgg gcaggtgacc tggagcatcc ggcaggctgc cctggcctcc tgcgtcagga
720

caacgcccac gaggggcggtt actgtgcgga gatgcaccac gcaagagaca ccctttgtaa
780

ctctcttctc ctccctagtg cgaggttaaa accttcagcc ccacgtgctg ttgcaaacc
840

tgctgtacc tgaggcccta aaaagccaga gacctcactc ccggggagcc accatggcgg
900

cggcggtcct ggaaaaccca ggcttgggca ggaaactctc tgactttgga cagggtgagcc
960

acggcagcct gagctgctca gttaggggaa tttgggcctc cagagaaaga gatccgaaga
1020

ctgctggtgc ttcctgggtt cataagctca gtaagaagtc tgaattcggtt ggaagctgat
1080

gagaatatcc aggaagtcaa cagacaaatg tcctcaacaa ttgtttctaa gtaggagaac
1140

atctgtcctc ggtggctttc acaggaaaag cttctgacct cttctcttcc tcccacaggg
1200

cagcggagct actaacttca gcctgctgaa gcaggctgga gacgtggagg agaaccctgg
1260

acctatgtcc accgctgtgc tggagaaccc tgggctgggg aggaaactgt cagacttcgg
1320

gcaggagact tcatacattg aggataactg taaccagaat ggcgccatct ctctgatctt
1380

cagcctgaag gaggaagtgg gcgccctggc aaagggtgctg cgcttgtttg aggagaacga
1440

cgtgaatctg acccacatcg agtcccggcc ttctagactg aagaaggacg agtacgagtt
1500

ctttaccac ctggataagc ggtccctgcc agccctgaca aacatcatca agatcctgag
1560

gcacgacatc ggagcaaccg tgcacgagct gtctcgggac aagaagaagg ataccgtgcc
1620

ctggttcct 1680	cggacaatcc	aggagctgga	tagatttgcc	aaccagatcc	tgtcttacgg
agcagagctg 1740	gacgcagatc	accctggctt	caaggaccca	gtgtatcggg	cccggagaaa
gcagtttgcc 1800	gatatcgctt	acaattatag	gcacggacag	ccaatccctc	gcgtggagta
tatggaggag 1860	gagaagaaga	cctggggcac	agtgttcaag	accctgaaga	gcctgtacaa
gacacacgcc 1920	tgctacgagt	ataaccacat	cttccccctg	ctggagaagt	attgtggctt
tcacgaggac 1980	aatatccctc	agctggagga	cgtgagccag	ttcctgcaga	cctgcacagg
ctttaggctg 2040	aggccagtgg	caggactgct	gagctcccgg	gacttcctgg	gaggactggc
cttcagagtg 2100	tttcaactgca	cccagtacat	caggcacggc	tccaagccaa	tgtatacacc
agagccccgac 2160	atctgtcacg	agctgctggg	ccacgtgccc	ctgttttagcg	atagatcctt
cgcccagttt 2220	tcccaggaga	tcggactggc	atctctggga	gcacctgacg	agtacatcga
gaagctggcc 2280	accatctatt	ggttcacagt	ggagtttggc	ctgtgcaagc	agggcgatag
catcaaggcc 2340	tacggagcag	gactgctgtc	tagcttcggc	gagctgcagt	attgtctgtc
cgagaagcca 2400	aagctgctgc	ccctggagct	ggagaagacc	gccatccaga	actacaccgt
gacagagttc 2460	cagcccctgt	actatgtggc	cgagtctttt	aacgatgcca	aggagaaggt
gagaaatttc 2520	gccgccacaa	tccctaggcc	cttcagcgtg	cggtacgacc	cttataccca

gaggatcgag gtgctggata atacacagca gctgaagatc ctggctgact caatcaatag
2580

cgaaatcgga atcctgtgct ccgccctgca gaaaatcaaa tgagggtaccg atccagacat
2640

gataagatac attgatgagt ttggacaaac cacaactaga atgcagtgaa aaaaatgctt
2700

tatttgtgaa atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca
2760

agttaacaac aacaattgca ttcattttat gtttcagggt cagggggagg tgtgggaggt
2820

ttttaagga tccctgggat gggatgtgga atccttctag atttcttttg taatatttat
2880

aaagtgctct cagcaaggta tcaaaatggc aaaattgtga gtaactatcc tcctttcatt
2940

ttgggaagaa gatgaggcat gaagagaatt cagacagaaa cttactcaga ccaggggagg
3000

cagaaactaa gcagagagga aaatgaccaa gagttagccc tgggcatgga atgtgaaaga
3060

accctaaacg tgacttggaataaatgccca aggtatatct cattctccgg gatttggttg
3120

cattttcttg aggtgaagaa ttgcagaata cattctttaa tgtgacctac atatttacct
3180

atgggaggaa gtctgctcct ggactcttga gattcagtca taaagcccag gccagggaaa
3240

taatgtaagt ctgcaggccc ctgtcatcag taggattagg gagaagagtt ctcagtagaa
3300

aacaggggagg ctggagagaa aagaatgggt aatgttaacg ttaatatata tagaaagact
3360

gcagaactta ggactgattt ttatttgaat ccttaaaaaa aaaaatttct tatgaaaata
3420

gtacatggct cttaggagac agaacttatt gtacagagga acagcgtgag agtcagagtg
3480

atcccagaac aggtcctggc tccatcctgc acatagtttt ggtgctgctg gcaatacggg
3540

ccccacaact gtgggaaggg gttaggggca gggatctcat caggaaagca taggggttta
3600

aagttcttta tagagcactt agaagattga gaatccacaa attatatataa taacaaacaa
3660

agtagtgctg tgttatatag taaatgtgaa tttgcagaca catttaggga aaagttataa
3720

ttaaaaaaat aggctgtata tataacctgca ggtctagata cgtagataag tagcatggcg
3780

ggttaatcat taactacaag gaacccttag tgatggagtt ggccactccc tctctgcgcg
3840

ctcgctcgct cactgaggcc gggcgaccaa aggtcgcccg acgcccgggc ttgcccggg
3900

cggcctcagt gagcgagcga gcgcgcagag agggagtggc caa
3943

<210> 57

<211> 3943

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.3 vector correction genome (+ ITRs)

<400> 57

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccggggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag 240	gtcctgcata	tgcggccgct	tcaggagcag	ttgtgcgaat	agctggagaa
caccaggctg 300	gatttaaacc	cagatcgctc	ttacatttgc	tctttacctg	ctgtgctcag
cgttcacgtg 360	ccctctagct	gtagttttct	gaagtcagcg	cacagcaagg	cagtgtgctt
agaggttaac 420	agaagggaaa	acaacaacaa	caaaaatcta	aatgagaatc	ctgactgttt
cagctgagag 480	taaggggggc	ggattattca	tataattggt	ataccagacg	gtcgcaggct
tagtccaatt 540	gcagagaact	cgcttcccag	gcttctgaga	gtcccggaag	tgcctaaacc
tgtctaatacg 600	acggggcctg	ggtggcccgt	cgctccctgg	cttcttccct	ttaccaggg
cgggcagcga 660	agtgggtgcct	cctgcgtccc	ccacaccctc	cctcagcccc	tcccctccgg
cccgtcctgg 720	gcaggtgacc	tggagcatcc	ggcaggctgc	cctggcctcc	tgcgtcagga
caacgcccac 780	gaggggcggt	actgtgcgga	gatgcaccac	gcaagagaca	ccctttgtaa
ctctcttctc 840	ctccctagtg	cgagggttaa	accttcagcc	ccacgtgctg	tttgcaaacc
tgcctgtacc 900	tgaggcccta	aaaagccaga	gacctcactc	ccggggagcc	agcatgtcca
ctgcggtcct 960	ggaaaaccca	ggcttgggca	ggaaactctc	tgactttgga	caggtgagcc
acggcagcct 1020	gagctgctca	gttaggggaa	tttgggcctc	cagagaaaga	gatccgaaga
ctgctgggtgc 1080	ttcctgggtt	cataagctca	gtaagaagtc	tgaattcggt	ggaagctgat

gagaatatcc aggaagtcaa cagacaaatg tcctcaacaa ttgtttctaa gtaggagAAC
1140

atctgtcctc ggtggctttc acaggaaaag cttctgacct cttctcttcc tcccacaggg
1200

cagcggagct actaacttca gcctgctgaa gcaggctgga gacgtggagg agaaccctgg
1260

acctatgtcc accgctgtgc tggagaaccc tgggctgggg aggaaactgt cagacttcgg
1320

gcaggagact tcatacattg aggataactg taaccagaat ggcgccatct ctctgatctt
1380

cagcctgaag gaggaagtgg gcgccctggc aaaggtgctg cgcctgtttg aggagaacga
1440

cgtgaatctg acccacatcg agtcccggcc ttctagactg aagaaggacg agtacgagtt
1500

ctttaccac ctggataagc ggtccctgcc agccctgaca aacatcatca agatcctgag
1560

gcacgacatc ggagcaaccg tgcacgagct gtctcgggac aagaagaagg ataccgtgcc
1620

ctggttccct cggacaatcc aggagctgga tagatttgcc aaccagatcc tgtcttacgg
1680

agcagagctg gacgcagatc accctggctt caaggacca gtgtatcggg cccggagaaa
1740

gcagtttgcc gatatcgctt acaattatag gcacggacag ccaatccctc gcgtggagta
1800

tatggaggag gagaagaaga cctggggcac agtgttcaag accctgaaga gcctgtacaa
1860

gacacacgcc tgctacgagt ataaccacat cttccccctg ctggagaagt attgtggctt
1920

tcacgaggac aatatccctc agctggagga cgtgagccag ttcctgcaga cctgcacagg
1980

ctttaggctg aggccagtgg caggactgct gagctcccgg gacttcctgg gaggactggc
2040

cttcagagtg tttcactgca cccagtacat caggcacggc tccaagccaa tgtatacacc
2100

agagcccgac atctgtcacg agctgctggg ccacgtgccc ctgttttagcg atagatcctt
2160

cgcccagttt tcccaggaga tcggactggc atctctggga gcacctgacg agtacatcga
2220

gaagctggcc accatctatt gggtcacagt ggagtttggc ctgtgcaagc agggcgatag
2280

catcaaggcc tacggagcag gactgctgtc tagcttcggc gagctgcagt attgtctgtc
2340

cgagaagcca aagctgctgc ccctggagct ggagaagacc gccatccaga actacaccgt
2400

gacagagttc cagcccctgt actatgtggc cgagtctttt aacgatgcca aggagaaggt
2460

gagaaatttc gccgccacaa tccctaggcc cttcagcgtg cgggtacgacc cttataccca
2520

gaggatcgag gtgctggata atacacagca gctgaagatc ctggctgact caatcaatag
2580

cgaaatcgga atcctgtgct ccgccctgca gaaaatcaaa tgagggtaccg atccagacat
2640

gataagatac attgatgagt ttggacaaac cacaactaga atgcagtgaa aaaaatgctt
2700

tatttgtgaa atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca
2760

agttaacaac aacaattgca ttcattttat gtttcagggt caggggggagg tgtggggaggt
2820

tttttaagga tccctgggat gggatgtgga atccttctag atttcttttg taatatttat
2880

aaagtgtctt cagcaaggta tcaaaatggc aaaatttgtga gtaactatcc tccttttcatt
2940

ttgggaagaa gatgagggcat gaagagaatt cagacagaaa cttactcaga ccaggggagg
3000

cagaaactaa gcagagagga aaatgaccaa gagttagccc tgggcatgga atgtgaaaga
3060

accctaaacg tgacttggaa ataatgccca aggtatatcc cattctccgg gatttggttg
3120

cattttcttg aggtgaagaa ttgcagaata cattctttaa tgtgacctac atatttacc
3180

atgggaggaa gtctgtctct ggactcttga gattcagtca taaagcccag gccagggaaa
3240

taatgtaagt ctgcaggccc ctgtcatcag taggattagg gagaagagtt ctcagtagaa
3300

aacagggagg ctggagagaa aagaatgggt aatgttaacg ttaatataac tagaaagact
3360

gcagaactta ggactgattt ttatttgaat ccttaaaaaa aaaaatttct tatgaaaata
3420

gtacatggct cttaggagac agaacttatt gtacagagga acagcgtgag agtcagagtg
3480

atcccagaac aggtcctggc tccatcctgc acatagtttt ggtgctgctg gcaatacgg
3540

ccccacaact gtgggaaggg gttaggggca gggatctcat caggaaagca taggggttta
3600

aagtctctta tagagcactt agaagattga gaatccacaa attatattaa taacaaacaa
3660

agtagtgtcg tgttatatag taaatgtgaa tttgcagaca catttaggga aaagttataa
3720

ttaaaaaaat aggctgtata tatacctgca ggtctagata cgtagataag tagcatggcg
3780

ggttaatcat taactacaag gaaccctag tgatggagtt ggccactccc tctctgcgcg
3840

ctcgcctcgt cactgaggcc gggcgaccaa aggtcgcccg acgcccgggc ttgcccggg
3900

cggcctcagt gagcgagcga ggcgcgagag agggagtggc caa
3943

<210> 58

<211> 3943

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.4 vector correction genome (+ ITRs)

<400> 58

ttggccactc cctctctgcg cgctcgcctc ctcactgagg ccggggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttccctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgct tcaggagcag ttgtgcgaat agctggagaa
240

caccaggctg gatttaaacc cagatcgctc ttacatttgc tctttacctg ctgtgctcag
300

cgttcacgtg ccctctagct gtagttttct gaagtcagcg cacagcaagg cagtgtgctt
360

agaggttaac agaaggggaaa acaacaacaa caaaaatcta aatgagaatc ctgactgttt
420

cagctggggg taaggggggc ggattattca tataattgtt ataccagacg gtcgcaggct
480

tagtccaatt gcagagaact cgcttcccag gcttctgaga gtcccgggaag tgcctaaacc
540

tgtctaatacg acgggggcttg ggtggcccgt cgctccctgg cttcttccct ttacccaggg
600

cgggcagcga agtggtgcct cctgcgtccc ccacaccctc cctcagcccc tcccctccgg
660

cccgctcctgg gcaggtgacc tggagcatcc ggcaggctgc cctggcctcc tgagtcagga
720

caacgcccac gaggggcggtt actgtgcgga gatgcaccac gcaagagaca ccctttgtaa
780

ctctcttctc ctccctagtg cgagggttaa accttcagcc ccacgtgctg tttgcaaacc
840

tgctgtacc tgaggcccta aaaagccaga gacctactc ccggggagcc agcatgtcca
900

ctgcggtcct ggaaaaccca ggcttgggca ggaaactctc tgactttgga caggtgagcc
960

acggcagcct gagctgctca gttaggggaa tttgggcctc cagagaaaga gatccgaaga
1020

ctgctggtgc ttcttggttt cataagctca gtaagaagtc tgaattcggt ggaagctgat
1080

gagaatatcc aggaagtcaa cagacaaatg tcctcaacaa ttgtttctaa gtaggagAAC
1140

atctgtcctc ggtggctttc acaggaaaag cttctgacct cttctcttcc tcccacaggg
1200

cagcggagct actaacttca gcctgctgaa gcaggctgga gacgtggagg agaaccctgg
1260

acctatgtcc accgctgtgc tggagaaccc tgggctgggg aggaaactgt cagacttcgg
1320

gcaggagact tcatacattg aggataactg taaccagaat ggcgccatct ctctgatctt
1380

cagcctgaag gaggaagtgg gcgccctggc aaagggtgctg cgcctgtttg aggagaacga
1440

cgtgaatctg acccacatcg agtcccggcc ttctagactg aagaaggacg agtacgagtt
1500

ctttaccac ctggataagc ggtccctgcc agccctgaca aacatcatca agatcctgag
1560

gcacgacatc ggagcaaccg tgcacgagct gtctcgggac aagaagaagg ataccgtgcc
1620

ctggttccct cggacaatcc aggagctgga tagatttgcc aaccagatcc tgtcttacgg
1680

agcagagctg gacgcagatc accctggctt caaggacca gtgtatcggg cccggagaaa
1740

gcagtttgcc gatatcgctt acaattatag gcacggacag ccaatccctc gcgtggagta
1800

tatggaggag gagaagaaga cctggggcac agtgttcaag accctgaaga gcctgtacaa
1860

gacacacgcc tgctacgagt ataaccacat cttccccctg ctggagaagt attgtggctt
1920

tcacgaggac aatatccctc agctggagga cgtgagccag ttcctgcaga cctgcacagg
1980

ctttaggctg aggccagtgg caggactgct gagctcccg gacttcctgg gaggactggc
2040

cttcagagtg tttcactgca cccagtacat caggcacggc tccaagccaa tgtatacacc
2100

agagcccgac atctgtcacg agctgctggg ccacgtgccc ctgttttagcg atagatcctt
2160

cgcccagttt tcccaggaga tcggactggc atctctggga gcacctgacg agtacatcga
2220

gaagctggcc accatctatt ggttcacagt ggagtttggc ctgtgcaagc agggcgatag
2280

catcaaggcc tacggagcag gactgctgtc tagcttcggc gagctgcagt attgtctgtc
2340

cgagaagcca aagctgctgc ccctggagct ggagaagacc gccatccaga actacaccgt
2400

gacagagttc cagcccctgt actatgtggc cgagtctttt aacgatgcca aggagaaggt
2460

gagaaatttc gccgccacaa tccctaggcc cttcagcgtg cggtacgacc cttataccca
2520

gaggatcgag gtgctggata atacacagca gctgaagatc ctggctgact caatcaatag
2580

cgaaatcgga atcctgtgct ccgccctgca gaaaatcaaa tgaggtaccg atccagacat
2640

gataagatac attgatgagt ttggacaaac cacaactaga atgcagtga aaaaatgctt
2700

tatttgtgaa atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca
2760

agttaacaac aacaattgca ttcattttat gtttcaggtt cagggggagg tgtggggaggt
2820

tttttaagga tccctgggat gggatgtgga atccttctag atttcttttg taatatttat
2880

aaagtgctct cagcaaggta tcaaaatggc aaaattgtga gtaactatcc tcctttcatt
2940

ttgggaagaa gatgaggcat gaagagaatt cagacagaaa cttactcaga ccaggggagg
3000

cagaaactaa gcagagagga aaatgaccaa gagttagccc tgggcatgga atgtgaaaga
3060

accctaaacg tgacttggaataaatgccca aggtatatct cattctccgg gatttggttg
3120

cattttcttg aggtgaagaa ttgcagaata cattctttaa tgtgacctac atatttacct
3180

atgggaggaa gtctgctcct ggactcttga gattcagtca taaagcccag gccagggaaa
3240

taatgtaagt ctgcaggccc ctgtcatcag taggattagg gagaagagtt ctcagtagaa
3300

aacagggagg ctggagagaa aagaatgggt aatgttaacg ttaatatatac tagaaagact
3360

gcagaactta ggactgattt ttatttgaat ccttaaaaaa aaaaatttct tatgaaaata
3420

gtacatggct cttaggagac agaacttatt gtacagagga acagcgtgag agtcagagtg
3480

atcccagaac aggtcctggc tccatcctgc acatagtttt ggtgctgctg gcaatacggg
3540

ccccacaact gtgggaaggg gttaggggca gggatctcat caggaaagca taggggttta
3600

aagttcttta tagagcactt agaagattga gaatccacaa attatattaa taacaaacaa
3660

agtagtgctg tgttatatag taaatgtgaa tttgcagaca catttaggga aaagttataa
3720

ttaaaaaaat aggctgtata tatacctgca ggtctagata cgtagataag tagcatggcg
3780

ggttaatcat taactacaag gaacccttag tgatggagtt ggccactccc tctctgcgcg
3840

ctcgctcgct cactgaggcc gggcgaccaa aggtcgcccg acgcccgggc tttgcccggg
3900

cggcctcagt gagcgagcga gcgcgcagag agggagtggc caa
3943

<210> 59

<211> 3943

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.5 vector correction genome (+ ITRs)

<400> 59

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggctcgcc
60

cgacgcccgg gctttgcccc ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgct tcaggagcag ttgtgcgaat agctggagaa
240

caccaggctg gatttaaacc cagatcgctc ttacatttgc tctttacctg ctgtgctcag
300

cgttcacgtg ccctctagct gtagttttct gaagtcagcg cacagcaagg cagtgtgctt
360

agaggttaac agaagggaaa acaacaacaa caaaaatcta aatgagaatc ctgactgttt
420

cagctggggg taaggggggc ggattattca tataattgtt ataccagacg gtcgcaggct
480

tagtccaatt gcagagaact cgcttcccag gcttctgaga gtcccggaag tgcctaaacc
540

tgtctaatacg acggggcctt ggtggcccgt cgctccctgg cttcttccct ttaccaggg
600

cgggcagcga agtggtgcct cctgcgtccc ccacaccctc cctcagcccc tcccctccgg
660

cccgctcctg gcaggtgacg tcaagcatcc ggcaggctgc cctggcctcc tgagtcagga
720

caacgcccac gaggggcggt actgtgcgga gatgcaccac gcaagagaca ccctttgtaa
780

ctctcttctc ctccctagtg cgagggttaa accttcagcc ccacgtgctg ttgcaaacc
840

tgctgtacc tgaggcccta aaaagccaga gacctcactc ccggggagcc agcatgtcca
900

ctgcggtcct ggaaaaccca ggcttgggca ggaaactctc tgacttttga caggtgagcc
960

acggcagcct gagctgctca gttaggggaa tttgggcctc cagagaaaga gatccgaaga
1020

ctgctggtgc ttcctgggtt cataagctca gtaagaagtc tgaattcggt ggaagctgat
1080

gagaatatcc aggaagtcaa cagacaaatg tcctcaacaa ttgtttctaa gtaggagAAC
1140

atctgtcctc ggtggctttc acaggaaaag cttctgacct cttctcttcc tcccacaggg
1200

cagcggagct actaacttca gcctgctgaa gcaggctgga gacgtggagg agaaccctgg
1260

acctatgtcc accgctgtgc tggagaaccc tgggctgggg aggaaactgt cagacttcgg
1320

gcaggagact tcatacattg aggataactg taaccagaat ggcgccatct ctctgatctt
1380

cagcctgaag gaggaagtgg gcgccctggc aaagggtgctg cgccctgtttg aggagAAC
1440

cgtgaatctg acccacatcg agtcccggcc ttctagactg aagaaggacg agtacgagtt
1500

ctttaccac ctggataagc ggtccctgcc agccctgaca aacatcatca agatcctgag
1560

gcacgacatc ggagcaaccg tgcacgagct gtctcgggac aagaagaagg ataccgtgcc
1620

ctggttcctt cggacaatcc aggagctgga tagatttgcc aaccagatcc tgtcttacgg
1680

agcagagctg gacgcagatc accctggctt caaggacca gtgtatcggg cccggagaaa
1740

gcagtttgcc gatatcgctt acaattatag gcacggacag ccaatccctc gcgtggagta
1800

tatggaggag gagaagaaga cctggggcac agtgttcaag accctgaaga gcctgtacaa
1860

gacacacgcc tgctacgagt ataaccacat cttccccctg ctggagaagt attgtggctt
1920

tcacgaggac aatatccctc agctggagga cgtgagccag ttcctgcaga cctgcacagg
1980

ctttaggctg aggccagtgg caggactgct gagctcccgg gacttcctgg gaggactggc
2040

cttcagagtg tttcactgca cccagtacat caggcacggc tccaagccaa tgtatacacc
2100

agagcccgac atctgtcacg agctgctggg ccacgtgccc ctgttttagcg atagatcctt
2160

cgcccagttt tcccaggaga tcggactggc atctctggga gcacctgacg agtacatcga
2220

gaagctggcc accatctatt gggtcacagt ggagtttggc ctgtgcaagc agggcgatag
2280

catcaaggcc tacggagcag gactgctgtc tagcttcggc gagctgcagt attgtctgtc
2340

cgagaagcca aagctgctgc ccctggagct ggagaagacc gccatccaga actacaccgt
2400

gacagagttc cagcccctgt actatgtggc cgagtctttt aacgatgcca aggagaaggt
2460

gagaaatttc gccgccacaa tccctaggcc cttcagcgtg cggtacgacc cttataccca
2520

gaggatcgag gtgctggata atacacagca gctgaagatc ctggctgact caatcaatag
2580

cgaaatcgga atcctgtgct ccgccctgca gaaaatcaaa tgagggtaccg atccagacat
2640

gataagatac attgatgagt ttggacaaac cacaactaga atgcagtgaa aaaaatgctt
2700

tattttgtgaa atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca
2760

agttaacaac aacaattgca ttcattttat gtttcaggtt cagggggagg tgtgggaggt
2820

tttttaagga tccctgggat gggatgtgga atccttctag atttcttttg taatatttat
2880

aaagtgctct cagcaaggta tcaaaatggc aaaattgtga gtaactatcc tcctttcatt
2940

ttgggaagaa gatgaggcat gaagagaatt cagacagaaa cttactcaga ccaggggagg
3000

cagaaactaa gcagagagga aaatgaccaa gagttagccc tgggcatgga atgtgaaaga
3060

accctaaacg tgacttggaataaatgccca aggtatatct cattctccgg gatttggttg
3120

cattttcttg aggtgaagaa ttgcagaata cattctttaa tgtgacctac atatttacct
3180

atgggaggaa gtctgctcct ggactcttga gattcagtca taaagcccag gccagggaaa
3240

taatgtaagt ctgcaggccc ctgtcatcag taggattagg gagaagagtt ctcagtagaa
3300

aacagggagg ctggagagaa aagaatggtt aatgttaacg ttaatataac tagaaagact
3360

gcagaactta ggactgattt ttatttgaat ccttaaaaaa aaaaatttct tatgaaaata
3420

gtacatggct cttaggagac agaacttatt gtacagagga acagcgtgag agtcagagtg
3480

atcccagaac aggtcctggc tccatcctgc acatagtttt ggtgctgctg gcaatacggg
3540

ccccacaact gtgggaaggg gttaggggca gggatctcat caggaaagca taggggttta
3600

aagttcttta tagagcactt agaagattga gaatccacaa attatattaa taacaaacaa
3660

agtagtgtcg tgttatatag taaatgtgaa tttgcagaca catttaggga aaagttataa
3720

ttaaaaaaat aggctgtata tatacctgca ggtctagata cgtagataag tagcatggcg
3780

ggttaatcat taactacaag gaaccctag tgatggagtt ggccactccc tctctgcgcg
3840

ctcgctcgct cactgaggcc gggcgaccaa aggtcgcccg acgcccgggc ttgcccggg
3900

cggcctcagt gagcgagcga gcgcgcagag agggagtggc caa
3943

<210> 60

<211> 3943

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.6 vector correction genome (+ ITRs)

<400> 60

ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgct tcaggagcag ttgtgcgaat agctggagaa
240

caccaggctg gatttaaacc cagatcgctc ttacatttgc tctttacctg ctgtgctcag
300

cgttcacgtg ccctctagct gtagttttct gaagtcagcg cacagcaagg cagtgtgctt
360

agagggttaac agaaggggaaa acaacaacaa caaaaatcta aatgagaatc ctgactgttt
420

cagctggggg taaggggggc ggattattca tataattgtt ataccagacg gtcgcaggct
480

tagtccaatt gcagagaact cgcttcccag gcttctgaga gtcccggaag tgcctaaacc
540

tgtctaatacg acgggggcttg ggtggcccgt cgctccctgg cttcttccct ttacccaggg
600

cgggcagcga agtggtgcct cctgcgtccc ccacaccctc cctcagcccc tcccctccgg
660

cccgtcctgg gcaggtgacc tggagcatcc ggcaggctgc cctggcctcc tgcgtcagga
720

caacgcccac gaggggcggt actgtgcgga gatgcaccac gcaagagaca ccctttgtaa
780

ctctcttctc ctccctagtgc cgagggttaa accttcagcc ccacgtgctg tttgcaaacc
840

tgctgttacc tgaggcccta aaaagccaga gacctcactc ccggggagcc agcatgtcca
900

ctgcggtcct ggaaaaccca ggcttgggca ggaaactctc tgactttgga caggtagacc
960

acggcagcct gagctgctca gttaggggaa tttgggcctc cagagaaaga gatccgaaga
1020

ctgctggtgc ttcctgggtt cataagctca gtaagaagtc tgaattcggt ggaagctgat
1080

gagaatgtcc aggaagtcaa cagacaaatg tcctcaacaa ttgtttctaa gtaggagaac
1140

atctgtcctc ggtggctttc acaggaaaag cttctgacct cttctcttcc tcccacaggg
1200

cagcggagct actaacttca gcctgctgaa gcaggctgga gacgtggagg agaaccctgg
1260

acctatgtcc 1320	accgctgtgc	tggagaaccc	tgggctgggg	aggaaactgt	cagacttcgg
gcaggagact 1380	tcatacattg	aggataactg	taaccagaat	ggcgccatct	ctctgatctt
cagcctgaag 1440	gaggaagtgg	gcgccctggc	aaaggtgctg	cgcttgtttg	aggagaacga
cgtgaatctg 1500	acccacatcg	agtcccggcc	ttctagactg	aagaaggacg	agtacgagtt
ctttaccac 1560	ctggataagc	ggtccctgcc	agccctgaca	aacatcatca	agatcctgag
gcacgacatc 1620	ggagcaaccg	tgcacgagct	gtctcgggac	aagaagaagg	ataccgtgcc
ctggttcct 1680	cggacaatcc	aggagctgga	tagatttgcc	aaccagatcc	tgtcttacgg
agcagagctg 1740	gacgcagatc	accctggctt	caaggaccca	gtgtatcggg	cccggagaaa
gcagtttgcc 1800	gatatcgctt	acaattatag	gcacggacag	ccaatccctc	gcgtggagta
tatggaggag 1860	gagaagaaga	cctggggcac	agtgttcaag	accctgaaga	gcctgtacaa
gacacacgcc 1920	tgctacgagt	ataaccacat	cttccccctg	ctggagaagt	attgtggctt
tcacgaggac 1980	aatatccctc	agctggagga	cgtgagccag	ttcctgcaga	cctgcacagg
ctttaggctg 2040	aggccagtgg	caggactgct	gagctcccgg	gacttcctgg	gaggactggc
cttcagagtg 2100	tttcactgca	cccagtacat	caggcacggc	tccaagccaa	tgtatacacc
agagcccgc 2160	atctgtcacg	agctgctggg	ccacgtgccc	ctgttttagcg	atagatcctt

cgcccagttt tcccaggaga tcggactggc atctctggga gcacctgacg agtacatcga
2220

gaagctggcc accatctatt ggttcacagt ggagtttggc ctgtgcaagc agggcgatag
2280

catcaaggcc tacggagcag gactgctgtc tagcttcggc gagctgcagt attgtctgtc
2340

cgagaagcca aagctgctgc ccctggagct ggagaagacc gccatccaga actacaccgt
2400

gacagagttc cagcccctgt actatgtggc cgagtctttt aacgatgcca aggagaaggt
2460

gagaaatttc gccgccacaa tccctaggcc cttcagcgtg cggtagcacc cttataccca
2520

gaggatcgag gtgctggata atacacagca gctgaagatc ctggctgact caatcaatag
2580

cgaaatcgga atcctgtgct ccgccctgca gaaaatcaaa tgaggtagcg atccagacat
2640

gataagatac attgatgagt ttggacaaac cacaactaga atgcagtga aaaaatgctt
2700

tatttgtgaa atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca
2760

agttaacaac aacaattgca ttcattttat gtttcagggt cagggggagg tgtgggaggt
2820

tttttaagga tccctgggat gggatgtgga atccttctag atttcttttg taatatttat
2880

aaagtgctct cagcaaggta tcaaaatggc aaaattgtga gtaactatcc tcctttcatt
2940

ttgggaagaa gatgaggcat gaagagaatt cagacagaaa cttactcaga ccaggggagg
3000

cagaaactaa gcagagagga aaatgaccaa gagttagccc tgggcatgga atgtgaaaga
3060

accctaaacg tgacttggaa ataatgccca aggtatatattc cattctccgg gatttgttgg
3120

cattttcttg aggtgaagaa ttgcagaata cattctttaa tgtgacctac atatttacct
3180

atgggaggaa gtctgctcct ggactcttga gattcagtca taaagcccag gccagggaaa
3240

taatgtaagt ctgcaggccc ctgtcatcag taggattagg gagaagagtt ctcagtagaa
3300

aacagggagg ctggagagaa aagaatggtt aatgttaacg ttaatataac tagaaagact
3360

gcagaactta ggactgattt ttatttgaat ccttaaaaaa aaaaatttct tatgaaaata
3420

gtacatggct cttaggagac agaacttatt gtacagagga acagcgtgag agtcagagtg
3480

atcccagaac aggtcctggc tccatcctgc acatagtttt ggtgctgctg gcaatacggc
3540

ccccacaact gtgggaaggg gttaggggca gggatctcat caggaaagca taggggttta
3600

aagtcttcta tagagcactt agaagattga gaatccacaa attatattaa taacaaacaa
3660

agtagtgtcg tggtatatag taaatgtgaa tttgcagaca catttaggga aaagttataa
3720

ttaaaaaaat aggctgtata tatacctgca ggtctagata cgtagataag tagcatggcg
3780

ggttaatcat taactacaag gaaccctag tgatggagtt ggccactccc tctctgcgcg
3840

ctcgctcgct cactgaggcc gggcgaccaa aggtcgcccg acgcccgggc ttgcccggg
3900

cggcctcagt gagcgagcga gcgcgcagag agggagtggc caa
3943

<210> 61
<211> 3943
<212> DNA
<213> Artificial Sequence

<220>
<223> 32.7 vector correction genome (+ ITRs)

<400> 61
ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggctcgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgagc agagggagtg
120

gccaactcca tcactagggg ttccctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgct tcaggagcag ttgtgcgaat agctggagaa
240

caccaggctg gatttaaacc cagatcgctc ttacatttgc tctttacctg ctgtgctcag
300

cgttcacgtg ccctctagct gtagttttct gaagtcagcg cacagcaagg cagtgtgctt
360

agaggttaac agaagggaaa acaacaacaa caaaaatcta aatgagaatc ctgactgttt
420

cagctggggg taaggggggc ggattattca tataattggt ccaccagacg gtcgcaggct
480

tagtccaatt gcagagaact cgcttcccag gcttctgaga gtcccgggaag tgcctaaacc
540

tgtctaatacg acgggggcttg ggtggcccgt cgctccctgg cttcttccct ttaccaggg
600

cgggcagcga agtggtgcct cctgcgtccc ccacaccctc cctcagcccc tcccctccgg
660

cccgtcctgg gcaggtgacc tggagcatcc ggcaggctgc cctggcctcc tgcgtcagga
720

caacgcccac 780	gaggggcggt	actgtgcgga	gatgcaccac	gcaagagaca	ccctttgtaa
ctctcttctc 840	ctccctagt	cgagggttaa	accttcagcc	ccacgtgctg	tttgcaaacc
tgctgtacc 900	tgaggcccta	aaaagccaga	gacctcactc	ccggggagcc	agcatgtcca
ctgcggtcct 960	ggaaaaccca	ggcttgggca	ggaaactctc	tgactttgga	caggtgagcc
acggcagcct 1020	gagctgctca	gttaggggaa	tttgggcctc	cagagaaaga	gatccgaaga
ctgctggtgc 1080	ttcctgggtt	cataagctca	gtaagaagtc	tgaattcggt	ggaagctgat
gagaatatcc 1140	aggaagtcaa	cagacaaatg	tcctcaacaa	ttgtttctaa	gtaggagAAC
atctgtcctc 1200	ggtggctttc	acaggaaaag	cttctgacct	cttctcttcc	tcccacaggg
cagcggagct 1260	actaacttca	gcctgctgaa	gcaggctgga	gacgtggagg	agaaccctgg
acctatgtcc 1320	accgctgtgc	tggagaaccc	tgggctgggg	aggaaactgt	cagacttcgg
gcaggagact 1380	tcatacattg	aggataactg	taaccagaat	ggcgccatct	ctctgatctt
cagcctgaag 1440	gaggaagtgg	gcgccctggc	aaaggtgctg	cgcctgtttg	aggagaacga
cgtgaatctg 1500	acccacatcg	agtcccggcc	ttctagactg	aagaaggacg	agtacgagtt
ctttaccac 1560	ctggataagc	ggtccctgcc	agccctgaca	aacatcatca	agatcctgag
gcacgacatc 1620	ggagcaaccg	tgcacgagct	gtctcgggac	aagaagaagg	ataccgtgcc

ctggttcct 1680	cggacaatcc	aggagctgga	tagatttgcc	aaccagatcc	tgtcttacgg
agcagagctg 1740	gacgcagatc	accctggctt	caaggaccca	gtgtatcggg	cccggagaaa
gcagtttgcc 1800	gatatcgctt	acaattatag	gcacggacag	ccaatccctc	gcgtggagta
tatggaggag 1860	gagaagaaga	cctggggcac	agtgttcaag	accctgaaga	gcctgtacaa
gacacacgcc 1920	tgctacgagt	ataaccacat	cttccccctg	ctggagaagt	attgtggctt
tcacgaggac 1980	aatatccctc	agctggagga	cgtgagccag	ttcctgcaga	cctgcacagg
ctttaggctg 2040	aggccagtgg	caggactgct	gagctcccgg	gacttcctgg	gaggactggc
cttcagagtg 2100	tttcaactgca	cccagtacat	caggcacggc	tccaagccaa	tgtatacacc
agagcccgac 2160	atctgtcacg	agctgctggg	ccacgtgccc	ctgttttagcg	atagatcctt
cgcccagttt 2220	tcccaggaga	tcggactggc	atctctggga	gcacctgacg	agtacatcga
gaagctggcc 2280	accatctatt	ggttcacagt	ggagtttggc	ctgtgcaagc	agggcgatag
catcaaggcc 2340	tacggagcag	gactgctgtc	tagcttcggc	gagctgcagt	attgtctgtc
cgagaagcca 2400	aagctgctgc	ccctggagct	ggagaagacc	gccatccaga	actacaccgt
gacagagttc 2460	cagcccctgt	actatgtggc	cgagtctttt	aacgatgcca	aggagaaggt
gagaaatttc 2520	gccgccacaa	tccctaggcc	cttcagcgtg	cggtacgacc	cttataccca

gaggatcgag gtgctggata atacacagca gctgaagatc ctggctgact caatcaatag
2580

cgaaatcgga atcctgtgct ccgccctgca gaaaatcaaa tgaggtaccg atccagacat
2640

gataagatac attgatgagt ttggacaaac cacaactaga atgcagtgaa aaaaatgctt
2700

tatttgtgaa atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca
2760

agttaacaac aacaattgca ttcattttat gtttcagggt cagggggagg tgtgggaggt
2820

ttttaagga tccctgggat gggatgtgga atccttctag atttcttttg taatatttat
2880

aaagtgctct cagcaaggta tcaaaatggc aaaattgtga gtaactatcc tcctttcatt
2940

ttgggaagaa gatgaggcat gaagagaatt cagacagaaa cttactcaga ccaggggagg
3000

cagaaactaa gcagagagga aaatgaccaa gagttagccc tgggcatgga atgtgaaaga
3060

accctaaacg tgacttggaataaatgccca aggtatatct cattctccgg gatttggttg
3120

cattttcttg aggtgaagaa ttgcagaata cattctttaa tgtgacctac atatttacct
3180

atgggaggaa gtctgctcct ggactcttga gattcagtca taaagcccag gccagggaaa
3240

taatgtaagt ctgcaggccc ctgtcatcag taggattagg gagaagagtt ctcagtagaa
3300

aacaggggagg ctggagagaa aagaatgggt aatgttaacg ttaatatatac tagaaagact
3360

gcagaactta ggactgattt ttatttgaat ccttaaaaaa aaaaatttct tatgaaaata
3420

gtacatggct cttaggagac agaacttatt gtacagagga acagcgtgag agtcagagtg
3480

atcccagaac aggtcctggc tccatcctgc acatagtttt ggtgctgctg gcaatacggg
3540

ccccacaact gtgggaaggg gttaggggca gggatctcat caggaaagca taggggttta
3600

aagttcttta tagagcactt agaagattga gaatccacaa attatatataa taacaaacaa
3660

agtagtgctg tgttatatag taaatgtgaa tttgcagaca catttaggga aaagttataa
3720

ttaaaaaaat aggctgtata tatacctgca ggtctagata cgtagataag tagcatggcg
3780

ggttaatcat taactacaag gaacccttag tgatggagtt ggccactccc tctctgcgcg
3840

ctcgctcgct cactgaggcc gggcgaccaa aggtcgcccg acgcccgggc ttgcccggg
3900

cggcctcagt gagcgagcga gcgcgcagag agggagtggc caa
3943

<210> 62

<211> 3951

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.8 vector correction genome (+ ITRs)

<400> 62

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag 240	gtcctgcata	tgcggccgct	tcaggagcag	ttgtgcgaat	agctggagaa
caccaggctg 300	gatttaaacc	cagatcgctc	ttacatttgc	tctttacctg	ctgtgctcag
cgttcacgtg 360	ccctctagct	gtagttttct	gaagtcagcg	cacagcaagg	cagtgtgctt
agaggttaac 420	agaagggaaa	acaacaacaa	caaaaatcta	aatgagaatc	ctgactgttt
cagctggggg 480	taaggggggc	ggattattca	tataattggt	ccaccagacg	gtcgcaggct
tagtccaatt 540	gcagagaaca	cgctgtttct	cgccccaggc	ttctgagagt	cccggaagtg
cctaaacctg 600	tctaatacgac	ggggcttggg	tggcccgtcg	ctccctggct	tcttcccttt
acccagggcg 660	ggcagcgaag	tggtgccctc	tgcgtcccc	acaccctccc	tcagcccctc
ccctccggcc 720	cgtcctgggc	aggtgacctg	gagcatccgg	caggctgccc	tggcctcctg
cgtcaggaca 780	acgcccacga	ggggcgttac	tgtgcggaga	tgcaccacgc	aagagacacc
ctttgtaact 840	ctcttctcct	ccctagtgcg	aggttaaaac	cttcagcccc	acgtgctggt
tgcaaacctg 900	cctgtacctg	aggccctaaa	aagccagaga	cctcactccc	ggggagccag
catgtccact 960	gcggtcctgg	aaaacccagg	cttgggcagg	aaactctctg	actttggaca
ggtgagccac 1020	ggcagcctga	gctgctcagt	taggggaatt	tgggcctcca	gagaaagaga
tccgaagact 1080	gctggtgctt	cctggtttca	taagctcagt	aagaagtctg	aattcgttgg

aagctgatga gaatatccag gaagtcaaca gacaaatgtc ctcaacaatt gtttctaagt
1140

aggagaacat ctgtcctcgg tggctttcac aggaaaagct tctgacctct tctcttcctc
1200

ccacagggca gcggagctac taacttcagc ctgctgaagc aggctggaga cgtggaggag
1260

aaccctggac ctatgtccac cgctgtgctg gagaaccctg ggctggggag gaaactgtca
1320

gacttcgggc aggagacttc atacattgag gataactgta accagaatgg cgccatctct
1380

ctgatcttca gcctgaagga ggaagtgggc gccctggcaa aggtgctgcg cctgtttgag
1440

gagaacgacg tgaatctgac ccacatcgag tcccggcctt ctagactgaa gaaggacgag
1500

tacgagttct ttaccacact ggataagcgg tccctgccag ccctgacaaa catcatcaag
1560

atcctgaggc acgacatcgg agcaaccgtg cacgagctgt ctcgggacaa gaagaaggat
1620

accgtgccct ggttccctcg gacaatccag gagctggata gatttgccaa ccagatcctg
1680

tcttacggag cagagctgga cgcagatcac cctggcttca aggaccaggt gtatcggggc
1740

cggagaaagc agtttgccga tatcgcctac aattataggc acggacagcc aatccctcgc
1800

gtggagtata tggaggagga gaagaagacc tggggcacag tgttcaagac cctgaagagc
1860

ctgtacaaga cacacgcctg ctacgagtat aaccacatct tccccctgct ggagaagtat
1920

tgtggctttc acgaggacaa tatccctcag ctggaggacg tgagccagtt cctgcagacc
1980

tgcacaggct ttaggctgag gccagtggca ggactgctga gctcccggga cttcctggga
2040

ggactggcct tcagagtgtt tcaactgcacc cagtacatca ggcacggctc caagccaatg
2100

tatacaccag agcccgacat ctgtcacgag ctgctgggcc acgtgcccct gtttagcgat
2160

agatccttcg cccagttttc ccaggagatc ggactggcat ctctgggagc acctgacgag
2220

tacatcgaga agctggccac catctattgg ttcacagtgg agtttggcct gtgcaagcag
2280

ggcgatagca tcaaggccta cggagcagga ctgctgtcta gcttcggcga gctgcagtat
2340

tgtctgtccg agaagccaaa gctgctgccc ctggagctgg agaagaccgc catccagaac
2400

tacaccgtga cagagttcca gcccctgtac tatgtggccg agtcttttaa cgatgccaaag
2460

gagaagggtga gaaatttcgc cgccacaatc cctaggccct tcagcgtgcg gtacgaccct
2520

tatacccaga ggatcgaggt gctggataat acacagcagc tgaagatcct ggctgactca
2580

atcaatagcg aaatcggaat cctgtgctcc gccctgcaga aaatcaaattg aggtaccgat
2640

ccagacatga taagatacat tgatgagttt ggacaaacca caactagaat gcagtgaaaa
2700

aaatgcttta tttgtgaaat ttgtgatgct attgctttat ttgtaaccat tataagctgc
2760

aataaacaag ttaacaacaa caattgcatt cattttatgt ttcaggttca gggggagggtg
2820

tgggagggtt ttttaaggatc cctgggatgg gatgtggaat ccttctagat ttcttttgta
2880

atatttataa 2940	agtgctctca	gcaaggtatc	aaaatggcaa	aattgtgagt	aactatcctc
ctttcatttt 3000	gggaagaaga	tgaggcatga	agagaattca	gacagaaact	tactcagacc
aggggaggca 3060	gaaactaagc	agagaggaaa	atgaccaaga	gtagccctg	ggcatggaat
gtgaaagaac 3120	cctaaacgtg	acttggaat	aatgccaag	gtatattcca	ttctccggga
tttgttggca 3180	ttttcttgag	gtgaagaatt	gcagaataca	ttctttaatg	tgacctacat
atttaccat 3240	gggaggaagt	ctgctcctgg	actcttgaga	ttcagtcata	aagcccaggc
cagggaata 3300	atgtaagtct	gcaggcccct	gtcatcagta	ggattaggga	gaagagttct
cagtagaaaa 3360	cagggaggct	ggagagaaaa	gaatgggtaa	tgtaaactgt	aatataacta
gaaagactgc 3420	agaacttagg	actgattttt	atttgaatcc	ttaaaaaaaa	aaattttcta
tgaaaatagt 3480	acatggctct	taggagacag	aacttattgt	acagaggaac	agcgtgagag
tcagagtgat 3540	cccagaacag	gtcctggctc	catcctgcac	atagttttgg	tgctgctggc
aatacgggtcc 3600	ccacaactgt	gggaaggggt	taggggcagg	gatctcatca	ggaaagcata
ggggtttaaa 3660	gttctttata	gagcacttag	aagattgaga	atccacaaat	tatattaata
acaacaaaag 3720	tagtgctcgtg	ttatatagta	aatgtgaatt	tgacagacaca	tttagggaaa
agttataatt 3780	aaaaaaatag	gctgtatata	tacctgcagg	tctagatacg	tagataagta

gcatggcggg ttaatcatta actacaagga acccctagtg atggagttgg ccactccctc
3840

tctgcgcgct cgctcgctca ctgaggccgg gcgaccaaag gtcgcccgcg gcccgggctt
3900

tgcccgggcg gcctcagtga gcgagcgagc gcgcagagag ggagtggcca a
3951

<210> 63

<211> 3943

<212> DNA

<213> Artificial Sequence

<220>

<223> 32.9 vector correction genome (+ ITRs)

<400> 63

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccggggcgacc aaaggctcgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcgggccgct tcaggagcag ttgtgcgaat agctggagaa
240

caccaggctg gatttaaacc cagatcgctc ttacatttgc tctttacctg ctgtgctcag
300

cgttcacgtg ccctctagct gtagttttct gaagtcagcg cacagcaagg cagtgtgctt
360

agaggttaac agaaggggaaa acaacaacaa caaaaatcta aatgagaatc ctgactgttt
420

cagctgagag taaggggggc ggattattca tataattgtt ataccagacg gtcgcaggct
480

tagtccaatt gcagagaact cgcttcccag gcttctgaga gtcccggaag tgcctaaacc
540

tgtctaatacg acagagcttg ggtggcccgt cgctccctgg cttcttccct ttacccaggg
600

cgggcagcga agtggtgcct cctgcgtccc ccacaccctc cctcagcccc tcccctccgg
660

cccgctcctag gcaggtgacc tgaagcatcc agcaggctgc cctggcctcc tgcgtcagga
720

caacgcccac gaggggcggt actgtgcgga gatgcaccac gcaagagaca ccctttgtaa
780

ctctcttctc ctccctagtg cgagggttaa accttcagcc ccacgtgctg tttgcaaacc
840

tgctgtacc tgaggcccta aaaagccaga gacctactc ccggggagcc agcatgtcca
900

ctgcggtcct ggaaaaccca ggcttgggca ggaaactctc tgactttgga caggtgagcc
960

acggcagcct gagctgctca gttaggggaa tttgggcctc cagagaaaga gatccgaaga
1020

ctgctggtgc ttcttggttt cataagctca gtaagaagtc tgaattcggt ggaagctgat
1080

gagaatatcc aggaagtcaa cagacaaatg tcctcaacaa ttgtttctaa gtaggagAAC
1140

atctgtcctc ggtggctttc acaggaaaag cttctgacct cttctcttcc tcccacaggg
1200

cagcggagct actaacttca gcctgctgaa gcaggctgga gacgtggagg agaaccctgg
1260

acctatgtcc accgctgtgc tggagaaccc tgggctgggg aggaaactgt cagacttcgg
1320

gcaggagact tcatacattg aggataactg taaccagaat ggcgccatct ctctgatctt
1380

cagcctgaag gaggaagtgg gcgccctggc aaagggtgctg cgcctgtttg aggagaacga
1440

cgtgaatctg acccacatcg agtcccggcc ttctagactg aagaaggacg agtacgagtt
1500

ctttacccac ctggataagc ggtccctgcc agccctgaca aacatcatca agatcctgag
1560

gcacgacatc ggagcaaccg tgcacgagct gtctcgggac aagaagaagg ataccgtgcc
1620

ctggttccct cggacaatcc aggagctgga tagatttgcc aaccagatcc tgtcttacgg
1680

agcagagctg gacgcagatc accctggctt caaggaccca gtgtatcggg cccggagaaa
1740

gcagtttgcc gatatcgctt acaattatag gcacggacag ccaatccctc gcgtggagta
1800

tatggaggag gagaagaaga cctggggcac agtgttcaag accctgaaga gcctgtacaa
1860

gacacacgcc tgctacgagt ataaccacat cttccccctg ctggagaagt attgtggctt
1920

tcacgaggac aatatccctc agctggagga cgtgagccag ttccctgcaga cctgcacagg
1980

ctttaggctg aggccagtgg caggactgct gagctcccgg gacttcctgg gaggactggc
2040

cttcagagtg tttcactgca cccagtacat caggcacggc tccaagccaa tgtatacacc
2100

agagcccgac atctgtcacg agctgctggg ccacgtgccc ctgttttagcg atagatcctt
2160

cgcccagttt tcccaggaga tcggactggc atctctggga gcacctgacg agtacatcga
2220

gaagctggcc accatctatt ggttcacagt ggagtttggc ctgtgcaagc agggcgatag
2280

catcaaggcc tacggagcag gactgctgtc tagcttcggc gagctgcagt attgtctgtc
2340

cgagaagcca aagctgctgc ccctggagct ggagaagacc gccatccaga actacaccgt
2400

gacagagttc cagcccctgt actatgtggc cgagtctttt aacgatgcca aggagaaggt
2460

gagaaatttc gccgccacaa tccctaggcc cttcagcgtg cggtacgacc cttataccca
2520

gaggatcgag gtgctggata atacacagca gctgaagatc ctggctgact caatcaatag
2580

cgaaatcgga atcctgtgct ccgccctgca gaaaatcaaa tgaggtaccg atccagacat
2640

gataagatac attgatgagt ttggacaaac cacaactaga atgcagtga aaaaatgctt
2700

tatttgtgaa atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca
2760

agttaacaac aacaattgca ttcattttat gtttcaggtt cagggggagg tgtggggaggt
2820

tttttaagga tccctgggat gggatgtgga atccttctag atttcttttg taatatttat
2880

aaagtgctct cagcaaggta tcaaaatggc aaaattgtga gtaactatcc tcctttcatt
2940

ttgggaagaa gatgaggcat gaagagaatt cagacagaaa cttactcaga ccaggggagg
3000

cagaaactaa gcagagagga aaatgaccaa gagttagccc tgggcatgga atgtgaaaga
3060

accctaaacg tgacttggaataaatgccca aggtatatct cattctccgg gatttggttg
3120

cattttcttg aggtgaagaa ttgcagaata cattctttaa tgtgacctac atatttacct
3180

atgggaggaa gtctgctcct ggactcttga gattcagtca taaagcccag gccagggaaa
3240

taatgtaagt ctgcaggccc ctgtcatcag taggattagg gagaagagtt ctcagtagaa
3300

aacagggagg ctggagagaa aagaatgggt aatgttaacg ttaatatataac tagaaagact
3360

gcagaactta ggactgattt ttatttgaat ccttaaaaaa aaaaatttct tatgaaaata
3420

gtacatggct cttaggagac agaacttatt gtacagagga acagcgtgag agtcagagtg
3480

atcccagaac aggtcctggc tccatcctgc acatagtttt ggtgctgctg gcaatacggg
3540

ccccacaact gtgggaaggg gttaggggca gggatctcat caggaaagca taggggttta
3600

aagttcttta tagagcactt agaagattga gaatccacaa attatatataa taacaaacaa
3660

agtagtgctg tgttatatag taaatgtgaa tttgcagaca catttaggga aaagttataa
3720

ttaaaaaaat aggctgtata tatacctgca ggtctagata cgtagataag tagcatggcg
3780

ggttaatcat taactacaag gaacccttag tgatggagtt ggccactccc tctctgcgcg
3840

ctcgctcgct cactgaggcc gggcgaccaa aggtcgcccg acgcccgggc tttgcccggg
3900

cggcctcagt gagcgagcga gcgcgcagag agggagtggc caa
3943

<210> 64

<211> 912

<212> DNA

<213> Homo sapiens

<400> 64

atgtccactg cggtcctgga aaaccaggc ttgggcagga aactctctga ctttggacag
60

gaaacaagct atattgaaga caactgcaat caaaatgggtg ccatatcact gatctttctca
120

ctcaaagaag aagttgggtgc attggccaaa gtattgcgct tatttgagga gaatgatgta
180

aacctgaccc acattgaatc tagaccttct cgtttaaaga aagatgagta tgaatttttc
240

accattttgg ataaacgtag cctgcctgct ctgacaaaca tcatcaagat cttgaggcat
300

gacattgggtg ccactgtcca tgagctttca cgagataaga agaaagacac agtgccctgg
360

ttcccaagaa ccattcaaga gctggacaga tttgccaatc agattctcag ctatggagcg
420

gaactggatg ctgaccaccc tggtttttaa gatcctgtgt accgtgcaag acggaagcag
480

tttgctgaca ttgcctacaa ctaccgccat gggcagccca tccctcgagt ggaatacatg
540

gaggaagaaa agaaaacatg gggcacagtg ttcaagactc tgaagtcctt gtataaaacc
600

catgcttgct atgagtacaa tcacattttt ccacttcttg aaaagtactg tggcttccat
660

gaagataaca ttccccagct ggaagacggt tctcagttcc tgcagacttg cactggtttc
720

cgctccgac ctgtggctgg cctgctttcc tctcgggatt tcttgggtgg cctggccttc
780

cgagtcttcc actgcacaca gtacatcaga catggatcca agcccatgta taccctcgaa
840

cctgacatct gccatgagct gttgggacat gtgcccttgt tttcagatcg cagctttgcc
900

c a g t t t t c c c **a g**
912

<210> 65
<211> 912
<212> DNA
<213> Artificial Sequence

<220>
<223> from start codon to end of exon 8 of silently altered PAH
coding
sequence

<400> 65
atgtccaccg ctgtgctgga gaaccctggg ctgggggagga aactgtcaga cttcgggcag
60
gagacttcat acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc
120
ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg
180
aatctgaccc acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt
240
accacactgg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac
300
gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg
360
ttccctcgga caatccagga gctggataga ttgccaacc agatcctgtc ttacggagca
420
gagctggacg cagatcacc tggcttcaag gaccagtggt atcgggcccg gagaaagcag
480
tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg
540
gaggaggaga agaagacctg gggcacagtg ttcaagaccc tgaagagcct gtacaagaca
600
cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac
660

gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt
720

aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc
780

agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag
840

cccgacatct gtcacgagct gctggggccac gtgcccctgt ttagcgatag atccttcgcc
900

c a g t t t t c c c a g
912

<210> 66

<211> 447

<212> DNA

<213> Homo sapiens

<400> 66

gaaattggcc ttgcctctct ggggtgcacct gatgaatata ttgaaaagct cgccacaatt
60

tactggttta ctgtggagtt tgggctctgc aaacaaggag actccataaa ggcatatggt
120

gctgggctcc tgtcatcctt tgggtgaatta cagtactgct tatcagagaa gccaaagctt
180

ctccccctgg agctggagaa gacagccatc caaaattaca ctgtcacgga gttccagccc
240

ctctattacg tggcagagag ttttaatgat gccaaggaga aagtaaggaa ctttgctgcc
300

acaatacctc ggcccttctc agttcgctac gaccataca cccaaaggat tgaggtcttg
360

gacaataccc agcagcttaa gatTTTggct gattccatta acagtgaaat tggaatcctt
420

t g c a g t g c c c t c c a g a a a t a a a g t a a
447

<210> 67
<211> 447
<212> DNA
<213> Artificial Sequence

<220>
<223> from start of exon 9 to stop codon of silently altered PAH
coding
sequence

<400> 67
gagatcggac tggcatctct gggagcacct gacgagtaca tcgagaagct ggccaccatc
60

tattggttca cagtggagtt tggcctgtgc aagcagggcg atagcatcaa ggcctacgga
120

gcaggactgc tgtctagctt cggcgagctg cagtattgtc tgtccgagaa gccaaagctg
180

ctgcccctgg agctggagaa gaccgccatc cagaactaca ccgtgacaga gttccagccc
240

ctgtactatg tggccgagtc ttttaacgat gccaaaggaga aggtgagaaa tttcgccgcc
300

acaatcccta ggcccttcag cgtgcggtac gacccttata cccagaggat cgagggtgctg
360

gataatacac agcagctgaa gacccctggct gactcaatca atagcgaaat cggaatcctg
420

t g c t c c g c c c t g c a g a a a a t c a a a t g a
447

<210> 68
<211> 15
<212> DNA
<213> Artificial Sequence

<220>
<223> modification to glucocorticoid binding site

<400> 68

a c g c t g t t c t

t c g c c

15

<210> 69

<211> 800

<212> DNA

<213> Homo sapiens

<400> 69

tatcttccat ttactgagtg tttatgtgga agaactgtac taaatttttaa tgcatttctt
60

tattcctatt cttaaaacct tccagcaagg tggctctacc accctctttt ccgagcttca
120

ggagcagttg tgcgaatagc tggagaacac caggctggat ttaaaccacag atcgctctta
180

catttgctct ttacctgctg tgctcagcgt tcacgtgccc tctagctgta gttttctgaa
240

gtcagcgcac agcaaggcag tgtgcttaga ggtaacaga agggaaaaca acaacaacaa
300

aaatctaaat gagaatcctg actgtttcag ctgggggtaa ggggggcgga ttattcatat
360

aattgttata ccagacggtc gcaggcttag tccaattgca gagaactcgc ttcccaggct
420

tctgagagtc ccggaagtgc ctaaacctgt ctaatcgacg gggcttgggt ggcccgtcgc
480

tccctggctt cttcccttta cccagggcgg gcagcgaagt ggtgcctcct gcgtcccca
540

caccctccct cagcccctcc cctccggccc gtcctgggca ggtgacctgg agcatccggc
600

aggctgccct ggcctcctgc gtcaggacaa cgcccacgag gggcggttact gtgcggagat
660

gcaccacgca agagacaccc tttgtaactc tcttctcctc cctagtgcga ggttaaaacc
720

ttcagcccca cgtgctgttt gcaaacctgc ctgtacctga ggccctaaaa agccagagac
780

c t c a c t c c c g g g g a g c c a g c
800

<210> 70
<211> 800
<212> DNA
<213> Homo sapiens

<400> 70
tccactgcgg tcctggaaaa cccaggcttg ggcaggaaac tctctgactt tggacagggtg
60

agccacggca gcctgagctg ctcagttagg ggaatttggg cctccagaga aagagatccg
120

aagactgctg gtgcttcctg gtttcataag ctcagtaaga agtctgaatt cgttggaagc
180

tgatgagaat atccaggaag tcaacagaca aatgtcctca acaattgttt ctaagtagga
240

gaacatctgt cctcggtggc tttcacagga atgaatgacc attgcttttag ggggttgggg
300

atctggcctc cagaactgcc accaattagc tgtgtgtctt tggacaagtt actgtccctc
360

tctgttgtct gtttactctt ctgtacactg aaggggctgg tccctaataga tctgggatgg
420

gatgtggaat ccttctagat ttcttttgta atatttataa agtgctctca gcaaggatc
480

aaaatggcaa aattgtgagt aactatcctc ctttcatttt gggaagaaga tgaggcatga
540

agagaattca gacagaaact tactcagacc aggggaggca gaaactaagc agagaggaaa
600

atgaccaaga gttagccctg ggcattggaat gtgaaagaac cctaaacgtg acttggaat
660

aatgcccaag gtatatcca ttctccggga ttgttgga tttcttgag gtgaagaatt
720

gcagaatata ttctttaatg tgacctacat atttaccat gggaggaagt ctgctcctgg
780

a c t c t t g a g a t t c a g t c a t a
800

<210> 71

<211> 16

<212> DNA

<213> Artificial Sequence

<220>

<223> targeted integration restriction cassette

<400> 71

t a c g t a c g a t c g t c g a
16

<210> 72

<211> 800

<212> DNA

<213> Homo sapiens

<400> 72

aatggttcca aaattttcta tggttaagaa tcacctggga tggttttgaa atggcagatt
60

ctaagacaac ttgattcaac aggtttaggt aaagcccagg gaactgcatt ataagaagga
120

atcacctgta attttgaggt caagatccaa ggaacactca ttgagaaaca ctgatttaca
180

aagtgcattg agagaaatgg agcaagtga gggggatcag catggtgaaa tataggctgt
240

taggagtgct attgactaac tgtctggtga ctggaccaga gtaaattcttt tactttgcaa
300

gaaacaggac taaattccca tattatgtcc atagcaaagg gaattatgta gaaaaattga
360

taattaggag cctgagttct tgaccagcct ccactaccta tgtggcctca ggtgagttat
420

tttctccctt tggctctaag ttttcccat ctgtaatgta agggagttta actagatgag
480

cactaaggac aaatcaattt ctgtgagtca attattatga aataccatgt gggcatcaaa
540

tgccaagtgg aaagcataga taaagaagtg attgtgcacc tgggctgagg ggaacaaaca
600

tttcctaaga gaattgagac ccaaaagagc ctttaaggaa ggtgagatct tggaaaggga
660

aatttggtga atactctaata gaggagctaa aaaggcaaga aagaaagcag cttggctgga
720

aaggagggtc ctgtaggtgg gcctccagag attcggtacc acagaaactg ccaaacatca
780

g c a a g a a g c c a t g g g g a t g g
800

<210> 73
<211> 800
<212> DNA
<213> Homo sapiens

<400> 73
agcgtttgag ggattctaaa tagaaggaca agagtaaaaa tgtcaggctg gatcgatgca
60

ggccactaag aaatggattc aggtgatggc agtgggaaga aaggacctga tgcccagagg
120

catttctgga gaagatgaga tcagacttgt gattggctga acacacactg tagtggggtg
180

gggttttaggg ggtgactcaa cttcaagccc aggtacattc aagtctgaat tggcctagtc
240

aaaagtggca tctgtggatg tgtatcagaa atatcttact tttcttggaa gccaacagga
300

gaaaagagtg ctaccaagtg aactagagac aggaatatct tttgtcattt caaggaaact
360

ggaaagaaga aggctcagta ttcttttagta ggaagaagac ttaagtcaga gactcatctg
420

tacctctctg gcagggttta aaagggggaa gaggaataga ggctgcaaga gattgtgatt
480

catggacagt atgcagagat caaatgacct gggttcagat cctggctcca ctgctaactg
540

tgtaactata ggcaagttcc ttaacctctc taagccttaa tcttgtcatc aataaaaggg
600

ggcacttggt gcctaataaa acctacctct taggttggtg ccaaattaca tgagataatc
660

caaatcaagt gcttattata ataccagaa attataggct ctaaataaat gtttatatag
720

gctctaaata aatgaagttt tttagaaaga taacatcatg atcaaaatgg gatatttaac
780

a g t t t a g t c t t c c a t t t c a t
800

<210> 74

<211> 72

<212> DNA

<213> Artificial Sequence

<220>

<223> 2A element

<400> 74

agatctggca gcggagaggg cagaggaagt cttctaacat gcggtgacgt ggaggagaat
60

c c c g g c c c t a
72

g g

<210> 75

<211> 8

<212> PRT

<213> Artificial Sequence

<220>

<223> Ribosomal skipping peptide consensus sequence

<220>

<221> VARIANT

<222> (1)..(1)

<223> D or G

<220>

<221> VARIANT

<222> (2)..(2)

<223> V or I

<220>

<221> VARIANT

<222> (4)..(4)

<223> any amino acid residue

<400> 75

Xaa Xaa Glu Xaa Asn Pro Gly Pro
1 5

<210> 76

<211> 18

<212> PRT

<213> Artificial Sequence

<220>

<223> T2A peptide

<400> 76

Glu Gly Arg Gly Ser Leu Leu Thr Cys Gly Asp Val Glu Glu Asn Pro
1 5 10 15

Gly Pro

<210> 77
<211> 19
<212> PRT
<213> Artificial Sequence

<220>
<223> P2A peptide

<400> 77

Ala Thr Asn Phe Ser Leu Leu Lys Gln Ala Gly Asp Val Glu Glu Asn
1 5 10 15

Pro Gly Pro

<210> 78
<211> 54
<212> DNA
<213> Artificial Sequence

<220>
<223> T2A element

<400> 78
gagggcagag gaagtcttct aacatgcggt gacgtggagg agaatcccgg ccct
54

<210> 79
<211> 57
<212> DNA
<213> Artificial Sequence

<220>

<223> P2A element

<400> 79

gctactaact tcagcctgct gaagcaggct ggagacgtgg aggagaaccc tggacct
57

<210> 80

<211> 1489

<212> DNA

<213> Artificial Sequence

<220>

<223> PAH intron-inserted silently altered coding sequence (human
HBB
first intron)

<400> 80

atgtccaccg ctgtgctgga gaaccctggg ctgggggagga aactgtcaga cttcgggcag
60

gagacttcat acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc
120

ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg
180

aatctgacct acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt
240

accacactgg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac
300

gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg
360

ttccctcgga caatccagga gctggataga ttgccaacc agatcctgtc ttacggagca
420

gagctggacg cagatcaccc tggcttcaag gaccacagtgt atcgggcccg gagaaagcag
480

tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg
540

gaggaggaga agaagacctg gggcacagtg ttcaagaccc tgaagagcct gtacaagaca
600

cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac
660

gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt
720

aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc
780

agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag
840

cccgacatct gtcacgagct gctggggccac gtgcccctgt ttagcgatag atccttcgcc
900

cagttttccc aggttggtat caaggttaca agacaggttt aaggagacca atagaaactg
960

ggcatgtgga gacagagaag actcttggggt ttctgatagg cactgactct ctctgcctat
1020

tggctctatth tcccaccctt aggagatcgg actggcatct ctgggagcac ctgacgagta
1080

catcgagaag ctggccacca tctattgggt cacagtggag tttggcctgt gcaagcaggg
1140

cgatagcatc aaggcctacg gagcaggact gctgtctagc ttcggcgagc tgcagtattg
1200

tctgtccgag aagccaaagc tgctgcccct ggagctggag aagaccgcca tccagaacta
1260

caccgtgaca gagttccagc ccctgtacta tgtggccgag tcttttaacg atgccaagga
1320

gaaggtgaga aatttcgccg ccacaatccc taggcccttc agcgtgcggt acgaccctta
1380

taccagagg atcgaggtgc tggataatac acagcagctg aagatcctgg ctgactcaat
1440

caatagcgaa atcggaatcc tgtgctccgc cctgcagaaa atcaaatga
1489

<210> 81

<211> 1475

<212> DNA

<213> Artificial Sequence

<220>

<223> PAH intron-inserted silently altered coding sequence (mouse
HBB
first intron)

<400> 81

atgtccaccg ctgtgctgga gaaccctggg ctggggagga aactgtcaga cttcgggcag
60

gagacttcat acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc
120

ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg
180

aatctgaccc acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt
240

acccacctgg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac
300

gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg
360

ttccctcgga caatccagga gctggataga ttgccaacc agatcctgtc ttacggagca
420

gagctggacg cagatcaccc tggcttcaag gaccagtggt atcgggcccg gagaaagcag
480

tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg
540

gaggaggaga agaagacctg gggcacagtg ttcaagaccc tgaagagcct gtacaagaca
600

cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac
660

gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt
720

aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc
780

agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag
840

cccgacatct gtcacgagct gctgggccac gtgcccctgt ttagcgatag atccttcgcc
900

cagttttccc aggttggtat ccaggttaca aggcagctca caagaagaag ttgggtgctt
960

ggagacagag gtctgctttc cagcagacac taactttcag tgtcccctgt ctatgtttcc
1020

cttttttagga gatcggactg gcatctctgg gagcacctga cgagtacatc gagaagctgg
1080

ccaccatcta ttgggttcaca gtggagtttg gcctgtgcaa gcagggcgat agcatcaagg
1140

cctacggagc aggactgctg tctagcttcg gcgagctgca gtattgtctg tccgagaagc
1200

caaagctgct gcccctggag ctggagaaga ccgccatcca gaactacacc gtgacagagt
1260

tccagcccct gtactatgtg gccgagtctt ttaacgatgc caaggagaag gtgagaaatt
1320

tcgccgccac aatccctagg cccttcagcg tgcggtacga cccttatacc cagaggatcg
1380

aggtgctgga taatacacag cagctgaaga tcctggctga ctcaatcaat agcgaaatcg
1440

g a a t c c t g t g c t c c g c c c t g c a g a a a a t c a a a t g a
1475

<210> 82
<211> 1451
<212> DNA
<213> Artificial Sequence

<220>

<223> PAH intron-inserted silently altered coding sequence (MVM intron)

<400> 82

atgtccaccg ctgtgctgga gaaccctggg ctggggagga aactgtcaga cttcgggcag
60

gagacttcat acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc
120

ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg
180

aatctgaccc acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt
240

accacactgg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac
300

gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg
360

ttccctcgga caatccagga gctggataga ttgccaacc agatcctgtc ttacggagca
420

gagctggacg cagatcaccc tggcttcaag gaccagtggt atcgggcccg gagaaagcag
480

tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg
540

gaggaggaga agaagacctg gggcacagtg ttcaagaccc tgaagagcct gtacaagaca
600

cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac
660

gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt
720

aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc
780

agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag
840

cccgacatct gtcacgagct gctggggccac gtgcccctgt ttagcgatag atccttcgcc
900

cagttttccc agaagaggta agggtttaag ggatgggttg ttggtgggggt attaattgtt
960

aattacctgg agcacctgcc tgaaatcact ttttttcagg ttgggagatc ggactggcat
1020

ctctgggagc acctgacgag tacatcgaga agctggccac catctattgg ttcacagtgg
1080

agtttggcct gtgcaagcag ggcgatagca tcaaggccta cggagcagga ctgctgtcta
1140

gcttcggcga gctgcagtat tgtctgtccg agaagccaaa gctgctgccc ctggagctgg
1200

agaagaccgc catccagaac tacaccgtga cagagttcca gcccctgtac tatgtggccg
1260

agtcttttaa cgatgccaag gagaagggtga gaaatttcgc cgccacaatc cctaggccct
1320

tcagcgtgcg gtacgaccct tatacccaga ggatcgaggt gctggataat acacagcagc
1380

tgaagatcct ggctgactca atcaatagcg aaatcggaat cctgtgctcc gccctgcaga
1440

a a a t c a a a t g a
1451

<210> 83

<211> 1607

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-hAC-008 editing element

<400> 83

atgtccaccg ctgtgctgga gaaccctggg ctggggagga aactgtcaga cttcgggcag
60

gagacttcat acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc
120

ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg
180

aatctgaccc acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt
240

acccacctgg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac
300

gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg
360

ttccctcgga caatccagga gctggataga tttgccaacc agatcctgtc ttacggagca
420

gagctggacg cagatcaccc tggcttcaag gaccagtggt atcgggcccg gagaaagcag
480

tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg
540

gaggaggaga agaagacctg gggcacagtg ttcaagaccc tgaagagcct gtacaagaca
600

cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac
660

gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt
720

aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc
780

agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag
840

cccgacatct gtcacgagct gctgggccac gtgcccctgt ttagcgatag atccttcgcc
900

cagttttccc aggagatcgg actggcatct ctgggagcac ctgacgagta catcgagaag
960

ctggccacca tctattgggtt cacagtggag tttggcctgt gcaagcaggg cgatagcatc
1020

aaggcctacg gagcaggact gctgtctagc ttcggcgagc tgcagtattg tctgtccgag
1080

aagccaaagc tgctgcccct ggagctggag aagaccgcca tccagaacta caccgtgaca
1140

gagttccagc ccctgtacta tgtggccgag tcttttaacg atgccaagga gaaggtgaga
1200

aatttcgccg ccacaatccc taggcccttc agcgtgcggt acgaccctta taccagagg
1260

atcgagggtgc tggataatac acagcagctg aagatcctgg ctgactcaat caatagcgaa
1320

atcggaatcc tgtgctccgc cctgcagaaa atcaaattgag aattcaaggc ctctcgagcc
1380

tctagaacta tagtgagtcg tattacgtag atccagacat gataagatac attgatgagt
1440

ttggacaaac cacaactaga atgcagtgaa aaaaatgctt tattttgtgaa atttgtgatg
1500

ctattgcttt atttgtaacc attataagct gcaataaaca agttaacaac aacaattgca
1560

ttcattttat gtttcagggtt cagggggagg tgtgggagggt ttttttaa
1607

<210> 84

<211> 1719

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-h1C-007 editing element

<400> 84

ctgacctctt ctcttcctcc cacagggcgg taccagatct ggcagcggag agggcagagg
60

aagtcttcta acatgcggtg acgtggagga gaatcccggc cctaggggta ccatgtccac
120

cgctgtgctg gagaaccctg ggctggggag gaaactgtca gacttcgggc aggagacttc
180

atacattgag gataactgta accagaatgg cgccatctct ctgatcttca gcctgaagga
240

ggaagtgggc gccctggcaa aggtgctgcg cctgtttgag gagaacgacg tgaatctgac
300

ccacatcgag tcccggcctt ctagactgaa gaaggacgag tacgagttct ttaccacact
360

ggataagcgg tccctgccag ccctgacaaa catcatcaag atcctgaggc acgacatcgg
420

agcaaccgtg cacgagctgt ctcgggacaa gaagaaggat accgtgccct ggttcctcgc
480

gacaatccag gagctggata gatttgccaa ccagatcctg tcttacggag cagagctgga
540

cgcagatcac cctggcttca aggaccaggt gtatcggggc cggagaaagc agtttgccga
600

tatcgcttac aattataggc acggacagcc aatccctcgc gtggagtata tggaggagga
660

gaagaagacc tggggcacag tgttcaagac cctgaagagc ctgtacaaga cacacgcctg
720

ctacgagtat aaccacatct tccccctgct ggagaagtat tgtggctttc acgaggacaa
780

tatccctcag ctggaggacg tgagccagtt cctgcagacc tgcacaggct ttaggctgag
840

gccagtggca ggactgctga gctcccggga cttcctggga ggactggcct tcagagtgtt
900

tcaactgcacc cagtacatca ggcacggctc caagccaatg tatacaccag agcccgacat
960

ctgtcacgag ctgctgggccc acgtgcccct gtttagcgat agatccttcg cccagttttc
1020

ccaggagatc ggactggcat ctctgggagc acctgacgag tacatcgaga agctggccac
1080

catctatttg ttcacagtgg agtttggcct gtgcaagcag ggcgatagca tcaaggccta
1140

cggagcagga ctgctgtcta gcttcggcga gctgcagtat tgtctgtccg agaagccaaa
1200

gctgctgccc ctggagctgg agaagaccgc catccagaac tacaccgtga cagagttcca
1260

gcccctgtac tatgtggccg agtccttttaa cgatgccaaag gagaagggtga gaaatttcgc
1320

cgccacaatc cctaggccct tcagcgtgcg gtacgaccct tatacccaga ggatcgagggt
1380

gctggataat acacagcagc tgaagatcct ggctgactca atcaatagcg aaatcggaat
1440

cctgtgctcc gccctgcaga aaatcaaattg agaattcaag gcctctcgag cctctagaac
1500

tatagtgagt cgtattacgt agatccagac atgataagat acattgatga gtttggacaa
1560

accacaacta gaatgcagtg aaaaaaatgc tttattttgtg aaattttgtga tgctattgct
1620

ttattttgtaa ccattataag ctgcaataaa caagttaaca acaacaattg cattcatttt
1680

atgttttcagg ttcagggggga ggtgtgggag gtttttttaa
1719

<210> 85
<211> 3227
<212> DNA
<213> Artificial Sequence

<220>
<223> pHMI-hPAH-hAC-008 correction genome from 5' homology arm to
3'
homology arm

<400> 85
tatcttccat ttactgagtg tttatgtgga agaactgtac taaatttttaa tgcattttctt
60

tattcctatt cttaaaacct tccagcaagg tggctctacc accctctttt ccgagcttca
120

ggagcagttg tgcgaatagc tggagaacac caggctggat ttaaaccag atcgctctta
180

catttgctct ttacctgctg tgctcagcgt tcacgtgccc tctagctgta gttttctgaa
240

gtcagcgcac agcaaggcag tgtgcttaga ggtaaacaga agggaaaaca acaacaacaa
300

aaatctaaat gagaatcctg actgtttcag ctgggggtaa ggggggcgga ttattcatat
360

aattgttata ccagacggtc gcaggcttag tccaattgca gagaactcgc ttcccaggct
420

tctgagagtc ccggaagtgc ctaaacctgt ctaatcgacg gggcttgggt ggcccgtcgc
480

tccctggctt cttcccttta cccagggcgg gcagcgaagt ggtgcctcct gcgtcccca
540

caccctccct cagcccctcc cctccggccc gtcctgggca ggtgacctgg agcatccggc
600

aggctgccct ggcctcctgc gtcaggacaa cgcccacgag gggcggttact gtgcggagat
660

gcaccacgca agagacaccc tttgtaactc tcttctcctc cctagtgcga ggttaaaacc
720

ttcagcccca cgtgctgttt gcaaacctgc ctgtacctga ggcctaataa agccagagac
780

ctcactcccg gggagccagc atgtccaccg ctgtgctgga gaacctggg ctggggagga
840

aactgtcaga cttcgggcag gagacttcat acattgagga taactgtaac cagaatggcg
900

ccatctctct gatcttcagc ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc
960

tgtttgagga gaacgacgtg aatctgacct acatcgagtc ccggccttct agactgaaga
1020

aggacgagta cgagttcttt acccacctgg ataagcggtc cctgccagcc ctgacaaaca
1080

tcatcaagat cctgaggcac gacatcggag caaccgtgca cgagctgtct cgggacaaga
1140

agaaggatac cgtgccctgg ttccctcgga caatccagga gctggataga ttgccaacc
1200

agatcctgtc ttacggagca gagctggacg cagatcaccg tggcttcaag gaccagtggt
1260

atcgggcccc gagaaagcag ttgcccagata tcgcctacaa ttataggcac ggacagccaa
1320

tccctcgcgt ggagtatatg gaggaggaga agaagacctg gggcacagtg ttcaagacct
1380

tgaagagcct gtacaagaca cacgcctgct acgagtataa ccacatcttc cccctgctgg
1440

agaagtattg tggctttcac gaggacaata tccctcagct ggaggacgtg agccagttcc
1500

tgcagacctg cacaggcttt aggctgaggg cagtggcagg actgctgagc tcccgggact
1560

tcctgggagg actggccttc agagtgtttc actgcacca gtacatcagg cacggctcca
1620

agccaatgta tacaccagag cccgacatct gtcacgagct gctggggccac gtgccccctgt
1680

ttagcgatag atccttcgcc cagttttccc aggagatcgg actggcatct ctgggagcac
1740

ctgacgagta catcgagaag ctggccacca tctattgggtt cacagtggag tttggcctgt
1800

gcaagcaggg cgatagcatc aaggcctacg gagcaggact gctgtctagc ttcggcgagc
1860

tgcagtattg tctgtccgag aagccaaagc tgctgcccct ggagctggag aagaccgcca
1920

tccagaacta caccgtgaca gagttccagc ccctgtacta tgtggccgag tcttttaacg
1980

atgccaagga gaaggtgaga aatttcgccg ccacaatccc taggcccttc agcgtgcggt
2040

acgaccctta taccagaggg atcgaggtgc tggataatac acagcagctg aagatcctgg
2100

ctgactcaat caatagcgaa atcggaatcc tgtgctccgc cctgcagaaa atcaaatgag
2160

aattcaaggc ctctcgagcc tctagaacta tagtgagtcg tattacgtag atccagacat
2220

gataagatac attgatgagt ttggacaaac cacaactaga atgcagtgaa aaaaatgctt
2280

tatttgtgaa atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca
2340

agttaacaac aacaattgca ttcattttat gtttcagggtt caggggggagg tgtgggaggt
2400

tttttaagct ttacgtacga tcgtcgatcc actgcgggtcc tggaaaaccc aggcttgggc
2460

aggaaactct ctgacttttg acaggtgagc cacggcagcc tgagctgctc agttagggga
2520

atttgggcct ccagagaaaag agatccgaag actgctgggtg cttcctgggtt tcataagctc
2580

agtaagaagt ctgaattcgt tggaagctga tgagaatatc caggaagtca acagacaaat
2640

gtcctcaaca attgtttcta agtaggagaa catctgtcct cgggtggcttt cacaggaatg
2700

aatgaccatt gcttttagggg gttggggatc tggcctccag aactgccacc aattagctgt
2760

gtgtcttttg acaagttact gtccctctct gttgtctgtt tactcttctg tacactgaag
2820

gggctgggtcc ctaatgatct gggatgggat gtggaatcct tctagatttc ttttgtaata
2880

tttataaagt gctctcagca aggtatcaaa atggcaaaat tgtgagtaac tatcctcctt
2940

tcattttggg aagaagatga ggcatagaaga gaattcagac agaaacttac tcagaccagg
3000

ggaggcagaa actaagcaga gaggaaaatg accaagagtt agccctgggc atggaatgtg
3060

aaagaaccct aaacgtgact tggaataaat gccaaggta tattccattc tccgggattt
3120

gttggcattt tcttgagggtg aagaattgca gaatacattc tttaatgtga cctacatatt
3180

tacccatggg aggaagtctg ctcctggact cttgagattc agtcata
3227

<210> 86

<211> 3345

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-h1C-007 correction genome from 5' homology arm to
3'
homology arm

<400> 86

aatggttcca aaattttcta tggttaagaa tcacctggga tggttttgaa atggcagatt
60

ctaagacaac ttgattcaac aggttttaggt aaagcccagg gaactgcatt ataagaagga
120

atcacctgta attttggagt caagatccaa ggaacactca ttgagaaaca ctgatttaca
180

aagtgcattg agagaaatgg agcaagtga gggggatcag catggtgaaa tataggctgt
240

taggagtgct attgactaac tgtctgggtga ctggaccaga gtaaattcttt tactttgcaa
300

gaaacaggac taaattccca tattatgtcc atagcaaagg gaattatgta gaaaaattga
360

taattaggag cctgagttct tgaccagcct ccactaccta tgtggcctca ggtgagttat
420

tttctccctt tggctctaag ttttcccat ctgtaatgta agggagttta actagatgag
480

cactaaggac aaatcaattt ctgtgagtca attattatga aataccatgt gggcatcaaa
540

tgccaagtgg aaagcataga taaagaagtg attgtgcacc tgggctgagg ggaacaaaca
600

tttcctaaga gaattgagac ccaaaagagc ctttaaggaa ggtgagatct tggaaaggga
660

aatttgggtga atactctaata gaggagctaa aaaggcaaga aagaaagcag cttggctgga
720

aaggagggttc ctgtaggtgg gcctccagag attcggtacc acagaaactg ccaaaccatca
780

gcaagaagcc atgggggatgg aagcttctga cctctttctct tcctcccaca gggcgggtacc
840

agatctggca gcggagaggg cagaggaagt cttctaacat gcggtgacgt ggaggagaaat
900

cccggcccta ggggtaccat gtccaccgct gtgctggaga accctgggct ggggaggaaa
960

ctgtcagact tcgggcagga gacttcatac attgaggata actgtaacca gaatggcgcc
1020

atctctctga tcttcagcct gaaggaggaa gtggggcgccc tggcaaaggt gctgcgcctg
1080

tttgaggaga acgacgtgaa tctgaccac atcgagtccc ggccttctag actgaagaag
1140

gacgagtacg agttctttac ccacctggat aagcgggtccc tgccagccct gacaaacatc
1200

atcaagatcc tgaggcacga catcggagca accgtgcacg agctgtctcg ggacaagaag
1260

aaggataccg tgccctggtt ccctcggaca atccaggagc tggatagatt tgccaaccag
1320

atcctgtctt acggagcaga gctggacgca gatcaccctg gcttcaagga cccagtgtat
1380

cgggcccgga gaaagcagtt tgccgatatc gcctacaatt ataggcacgg acagccaatc
1440

cctcgcgtgg agtatatgga ggaggagaag aagacctggg gcacagtgtt caagaccctg
1500

aagagcctgt acaagacaca cgcctgctac gagtataacc acatcttccc cctgctggag
1560

aagtattgtg gctttcacga ggacaatatc cctcagctgg aggacgtgag ccagttcctg
1620

cagacctgca caggcttttag gctgaggcca gtggcaggac tgctgagctc ccgggacttc
1680

ctgggaggac tggccttcag agtgtttcac tgcacccagt acatcaggca cggctccaag
1740

ccaatgtata caccagagcc cgacatctgt cacgagctgc tgggccacgt gcccctgttt
1800

agcgatagat ccttcgcccc gttttcccag gagatcggac tggcatctct gggagcacct
1860

gacgagtaca tcgagaagct ggccaccatc tattgggttca cagtggagtt tggcctgtgc
1920

aagcagggcg atagcatcaa ggcctacgga gcaggactgc tgtctagctt cggcgagctg
1980

cagtattgtc tgtccgagaa gccaaagctg ctgcccctgg agctggagaa gaccgccatc
2040

cagaactaca ccgtgacaga gttccagccc ctgtactatg tggccgagtc ttttaacgat
2100

gccaaaggaga aggtgagaaa tttcgccgcc acaatcccta ggcccttcag cgtgcggtac
2160

gacccttata cccagaggat cgagggtgctg gataatacac agcagctgaa gatcctggct
2220

gactcaatca atagcgaaat cggaatcctg tgctccgccc tgcagaaaaat caaatgagaa
2280

ttcaaggcct ctcgagcctc tagaactata gtgagtcgta ttacgtagat ccagacatga
2340

taagatacat tgatgagttt ggacaaacca caactagaat gcagtgaaaa aaatgcttta
2400

tttgtgaaat ttgtgatgct attgctttat ttgtaaccat tataagctgc aataaacaag
2460

ttaacaacaa caattgcatt cattttatgt ttcagggttca gggggagggtg tgggagggtt
2520

tttaagcttt acgtacgatc gtcgaagcgt ttgagggatt ctaaatagaa ggacaagagt
2580

aaaaatgtca ggctggatcg atgcaggcca ctaagaaatg gattcaggtg atggcagtgg
2640

gaagaaagga cctgatgccc agaggcattt ctggagaaga tgagatcaga cttgtgattg
2700

gctgaacaca cactgtagtg ggggtggggtt taggggggtga ctcaacttca agcccaggta
2760

cattcaagtc tgaattgccc tagtcaaaag tggcatctgt ggatgtgtat cagaaatatc
2820

ttacttttct tggaagccaa caggagaaaa gagtgctacc aagtgaacta gagacaggaa
2880

tatcttttgt catttcaagg aaactggaaa gaagaaggct cagtattctt tagtaggaag
2940

aagacttaag tcagagactc atctgtacct ctctggcagg gtttaaaagg gggaagagga
3000

atagaggctg caagagattg tgattcatgg acagtatgca gagatcaaat gacctggggt
3060

cagatcctgg ctccactgct aactgtgtaa ctataggcaa gttccttaac ctctctaagc
3120

cttaatcttg tcatcaataa aagggggcac ttggtgccta ataaaaccta cctcttaggt
3180

tgttgccaaa ttacatgaga taatccaaat caagtgccta ttataatacc cagaaattat
3240

aggctctaaa taaatgttta tataggctct aaataaatga agtttttttag aaagataaca
3300

tcatgatcaa aatgggatat ttaacagttt agtcttccat ttcatt
3345

<210> 87

<211> 3637

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-hAC-008 correction genome from 5' ITR to 3' ITR

<400> 87

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccggg'gcacc aaagg’tcgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgct atcttccatt tactgagtgt ttatgtggaa
240

gaactgtact aaattttaat gcatttcttt attcctattc ttaaaacctt ccagcaaggt
300

ggctctacca ccctcttttc cgagcttcag gagcagttgt gcgaatagct ggagaacacc
360

aggctggatt taaaccacaga tcgctcttac atttgctctt tacctgctgt gctcagcggt
420

cacgtgccct ctagctgtag ttttctgaag tcagcgcaca gcaaggcagt gtgcttagag
480

gttaacagaa gggaaaacaa caacaacaaa aatctaaatg agaatcctga ctgtttcagc
540

tgggggtaag gggggcggat tattcatata attgttatac cagacggtcg caggcttagt
600

ccaattgcag agaactcgct tcccaggctt ctgagagtcc cggaagtgcc taaacctgtc
660

taatcgacgg ggcttgggtg gcccgtcgct ccctggcttc ttccctttac ccagggcggg
720

cagcgaagtg gtgcctcctg cgtccccac accctccctc agccccctcc ctccggcccc
780

tcctgggcag gtgacctgga gcatccggca ggctgccctg gcctcctgcg tcaggacaac
840

gcccacgagg ggcgttactg tgcggagatg caccacgcaa gagacaccct ttgtaactct
900

cttctcctcc ctagtgcgag gttaaaacct tcagccccac gtgctgtttg caaacctgcc
960

tgtacctgag gccctaataa gccagagacc tcactcccgg ggagccagca tgtccaccgc
1020

tgtgctggag aaccctgggc tggggaggaa actgtcagac ttcgggcagg agacttcata
1080

cattgaggat aactgtaacc agaatggcgc catctctctg atcttcagcc tgaaggagga
1140

agtgggcgcc ctggcaaagg tgctgcgcct gtttgaggag aacgacgtga atctgaccca
1200

catcgagtcc cggccttcta gactgaagaa ggacgagtag gagttcttta cccacctgga
1260

taagcgggtcc ctgccagccc tgacaaacat catcaagatc ctgaggcacg acatcggagc
1320

aaccgtgcac gagctgtctc gggacaagaa gaaggatacc gtgccctggg tccctcggac
1380

aatccaggag ctggatagat ttgccaacca gatcctgtct tacggagcag agctggacgc
1440

agatcacctt ggcttcaagg acccagtgtg tcggggcccgg agaaagcagt ttgccgatat
1500

cgcctacaat tataggcacg gacagccaat ccctcgcgtg gagtatatgg aggaggagaa
1560

gaagacctgg ggcacagtgt tcaagaccct gaagagcctg tacaagacac acgcctgcta
1620

cgagtataac cacatcttcc ccctgctgga gaagtattgt ggctttcacg aggacaatat
1680

ccctcagctg gaggacgtga gccagttcct gcagacctgc acaggcttta ggctgaggcc
1740

agtggcagga ctgctgagct cccgggactt cctgggagga ctggccttca gagtgtttca
1800

ctgcacccag tacatcaggc acggctccaa gccaatgtat acaccagagc ccgacatctg
1860

tcacgagctg ctggggccacg tgcccctggt tagcgataga tccttcgccc agttttccca
1920

ggagatcgga ctggcatctc tgggagcacc tgacgagtac atcgagaagc tggccaccat
1980

ctattgggttc acagtggagt ttggcctgtg caagcagggc gatagcatca aggcctacgg
2040

agcaggactg ctgtctagct tcggcgagct gcagtattgt ctgtccgaga agccaaagct
2100

gctgcccctg gagctggaga agaccgccat ccagaactac accgtgacag agttccagcc
2160

cctgtactat gtggccgagt cttttaacga tgccaaggag aaggtgagaa atttcgccgc
2220

cacaatccct aggcccttca gcgtgcggtg cgacccttat acccagagga tcgagggtgct
2280

ggataataca cagcagctga agatcctggc tgactcaatc aatagcgaaa tcggaatcct
2340

gtgctccgcc ctgcagaaaa tcaaatgaga attcaaggcc tctcgagcct ctagaactat
2400

agtgagtcgt attacgtaga tccagacatg ataagataca ttgatgagtt tggacaaacc
2460

acaactagaa tgcagtgaaa aaaatgcttt atttggtgaaa tttgtgatgc tattgcttta
2520

tttgtaacca ttataagctg caataaaca gttaacaaca acaattgcat tcattttatg
2580

tttcagggttc aggggggaggt gtgggaggtt ttttaagctt tacgtacgat cgtcgatcca
2640

ctgcggtcct ggaaaaccca ggcttgggca ggaaactctc tgacttttga caggtgagcc
2700

acggcagcct gagctgctca gttaggggaa tttgggcctc cagagaaaga gatccgaaga
2760

ctgctggtgc ttcctgggtt cataagctca gtaagaagtc tgaattcggt ggaagctgat
2820

gagaatatcc aggaagtcaa cagacaaatg tcctcaacaa ttgttttctaa gtaggagAAC
2880

atctgtcctc ggtggctttc acaggaatga atgaccattg ctttaggggg ttggggatct
2940

ggcctccaga actgccacca attagctgtg tgtcttttga caagttactg tccctctctg
3000

ttgtctgttt actcttctgt acactgaagg ggctgggtccc taatgatctg ggatgggatg
3060

tggaatcctt ctagatttct tttgtaatat ttataaagtg ctctcagcaa ggtatcaaaa
3120

tggcaaaatt gtgagtaact atcctccttt catttttggga agaagatgag gcatgaagag
3180

aattcagaca gaaacttact cagaccaggg gaggcagaaa ctaagcagag aggaaaatga
3240

ccaagagtta gccctgggca tggaatgtga aagaacccta aacgtgactt ggaaataatg
3300

cccaagggtat attccattct ccgggatttg ttggcatttt cttgagggtga agaattgcag
3360

aatacattct ttaatgtgac ctacatattt acccatggga ggaagtctgc tcctggactc
3420

ttgagattca gtcataaacc tgcagggtcta gatacgtaga taagtagcat ggcgggttaa
3480

tcattaacta caaggaaccc ctagtgatgg agttggccac tccctctctg cgcgctcgct
3540

cgctcactga ggccggggcga ccaaaggctc cccgacgcc gccgtttgcc cgggcggcct
3600

c a g t g a g c g a g c g a g c g c g c a g a g a g g g a g t g g c c a a
3637

<210> 88

<211> 3755

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-h1C-007 correction genome from 5' ITR to 3' ITR

<400> 88

ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgca atggttccaa aattttctat ggттаагаат
240

cacctgggat ggttttgaaa tggcagattc taagacaact tgattcaaca ggtttaggta
300

aagcccaggg aactgcatta taagaaggaa tcacctgtaa ttttggagtc aagatccaag
360

gaacactcat tgagaaacac tgatttaca agtgcattga gagaaatgga gcaagtgaag
420

ggggatcagc atggtgaaat ataggctgtt aggagtgcta ttgactaact gtctggtgac
480

tggaccagag taaatctttt actttgcaag aaacaggact aaattcccat attatgtcca
540

tagcaaaggg aattatgtag aaaaattgat aattaggagc ctgagttctt gaccagcctc
600

cactacctat	gtggcctcag	gtgagttatt	ttctcccttt	ggctctaagt	tttccccatc
660					
tgtaatgtaa	gggagtttaa	ctagatgagc	actaaggaca	aatcaatttc	tgtgagtcaa
720					
ttattatgaa	ataccatgtg	ggcatcaaat	gccaagtgga	aagcatagat	aaagaagtga
780					
ttgtgcacct	gggctgaggg	gaacaaacat	ttcctaagag	aattgagacc	caaaagagcc
840					
tttaaggaag	gtgagatctt	ggaaagggaa	atttggtgaa	tactctaattg	aggagctaaa
900					
aaggcaagaa	agaaagcagc	ttggctggaa	aggaggttcc	tgtagggtggg	cctccagaga
960					
ttcgggtacca	cagaaactgc	caaacatcag	caagaagcca	tgggggatgga	agctttctgac
1020					
ctctttctctt	cctcccacag	ggcgggtacca	gatctggcag	cggagaggggc	agaggaagtc
1080					
ttctaacatg	cggtgacgtg	gaggagaatc	ccggccctag	gggtaccatg	tccaccgctg
1140					
tgctggagaa	ccctgggctg	gggaggaaac	tgtcagactt	cgggcaggag	acttcataca
1200					
ttgaggataa	ctgtaaccag	aatggcgcca	tctctctgat	cttcagcctg	aaggaggaag
1260					
tgggcgccct	ggcaaagggtg	ctgcgcctgt	ttgaggagaa	cgacgtgaat	ctgaccacaca
1320					
tcgagtcccg	gccttctaga	ctgaagaagg	acgagtacga	gttcttttacc	cacctggata
1380					
agcggtcctt	gccagccctg	acaaacatca	tcaagatcct	gaggcacgac	atcggagcaa
1440					
ccgtgcacga	gctgtctcgg	gacaagaaga	aggataccgt	gccctggttc	cctcggacaa
1500					

tccaggagct ggatagattt gcccaaccaga tcctgtctta cggagcagag ctggacgcag
1560

atcaccttgg cttcaaggac ccagtgtatc gggcccggag aaagcagttt gccgatatcg
1620

cctacaatta taggcacgga cagccaatcc ctcgcgtgga gtatatggag gaggagaaga
1680

agacctgggg cacagtgttc aagaccctga agagcctgta caagacacac gcctgctacg
1740

agtataacca catcttcccc ctgctggaga agtattgtgg ctttcacgag gacaatatcc
1800

ctcagctgga ggacgtgagc cagttcctgc agacctgcac aggctttagg ctgaggccag
1860

tggcaggact gctgagctcc cgggacttcc tgggaggact ggccttcaga gtgtttcact
1920

gcacccagta catcaggcac ggctccaagc caatgtatac accagagccc gacatctgtc
1980

acgagctgct gggccacgtg cccctgttta gcgatagatc cttcgcccag ttttcccagg
2040

agatcggact ggcattctctg ggagcacctg acgagtacat cgagaagctg gccaccatct
2100

attggttcac agtggagttt ggcctgtgca agcagggcga tagcatcaag gcctacggag
2160

caggactgct gtctagcttc ggcgagctgc agtattgtct gtccgagaag ccaaagctgc
2220

tgcccctgga gctggagaag accgccatcc agaactacac cgtgacagag ttccagcccc
2280

tgtactatgt ggccgagtct tttaacgatg ccaaggagaa ggtgagaaat ttcgccgcca
2340

caatccctag gcccttcagc gtgcggtacg acccttatac ccagaggatc gaggtgctgg
2400

ataatacacaca 2460	gcagctgaag	atcctggctg	actcaatcaa	tagcgaaatc	ggaatcctgt
gctccgccct 2520	gcagaaaatc	aatgagaat	tcaaggcctc	tcgagcctct	agaactatag
tgagtcgtat 2580	tacgtagatc	cagacatgat	aagatacatt	gatgagtttg	gacaaaccac
aactagaatg 2640	cagtgaaaaa	aatgctttat	ttgtgaaatt	tgtgatgcta	ttgctttatt
tgtaaccatt 2700	ataagctgca	ataaacaagt	taacaacaac	aattgcattc	attttatgtt
tcaggttcag 2760	ggggaggtgt	gggaggtttt	ttaagcttta	cgtacgatcg	tcgaagcggt
tgagggattc 2820	taaatagaag	gacaagagta	aaaatgtcag	gctggatcga	tgcaggccac
taagaaatgg 2880	attcaggtga	tggcagtggg	aagaaaggac	ctgatgccca	gaggcatttc
tggagaagat 2940	gagatcagac	ttgtgattgg	ctgaacacac	actgtagtgg	ggtgggggttt
aggggggtgac 3000	tcaacttcaa	gcccaggtac	attcaagtct	gaattgccct	agtcaaaaagt
ggcatctgtg 3060	gatgtgtatc	agaaatatct	tacttttctt	ggaagccaac	aggagaaaag
agtgctacca 3120	agtgaactag	agacaggaat	atctttttgtc	atttcaagga	aactggaaag
aagaaggctc 3180	agtatttcttt	agtaggaaga	agacttaagt	cagagactca	tctgtacctc
tctggcaggg 3240	tttaaaaggg	ggaagaggaa	tagaggctgc	aagagattgt	gattcatgga
cagtatgcag 3300	agatcaaattg	acctgggttc	agatcctggc	tccactgcta	actgtgtaac

<210> 91
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> PAH_Genomic Set 1, primer F

<400> 91
g c t c c a t c c t g c a c a t a g t t
20

<210> 92
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> PAH_Genomic Set 1, primer R

<400> 92
c c t a t g c t t t c c t g a t g a g a t c c
23

<210> 93
<211> 23
<212> DNA
<213> Artificial Sequence

<220>
<223> PAH_Genomic Set 1, probe

<400> 93
t t g g t g c t g c t g g c a a t a c g g t c
23

<210> 94
<211> 22
<212> DNA
<213> Artificial Sequence

<220>

<223> SV40_FAM Set 1, primer F

<400> 94

g c a a t a g c a t c a c a a a t t t c a c
22

<210> 95

<211> 25

<212> DNA

<213> Artificial Sequence

<220>

<223> SV40_FAM Set 1, primer R

<400> 95

g a t c c a g a c a t g a t a a g a t a c a t t g
25

<210> 96

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> SV40_FAM Set 1, probe

<400> 96

t c a c t g c a t t c t a g t t g t g g t t t g t c c a
28

<210> 97

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> PAH_HA Set 1, primer F

<400> 97

t c c a g t c a c c a g a c a g t t a g t
21

<210> 98
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> PAH_HA Set 1, primer R

<400> 98
g g a g a g a a a t g g a g c a a g t g a a
22

<210> 99
<211> 28
<212> DNA
<213> Artificial Sequence

<220>
<223> PAH_HA Set 1, probe

<400> 99
a c a g c c t a t a t t t c a c c a t g c t g a t c c c
28

<210> 100
<211> 800
<212> DNA
<213> Mus musculus

<400> 100
agcattagct tccatttatg cagtgtaaat ggtgagaaca gccccgactg aatacccaga
60

gcacatctc gtctgtgtca ttcatgcaca taacatatct cagcgagggtg gcccttctgt
120

cctcttttgca gagaccacgc caccatacta gtacctagag aactggctgg atttcagccc
180

cgatacctcc gggcttttgc tcatgttcgc ctcatagggt catctgggtg gttgcctaag
240

gaaaagtatg tcatggagac taacttgctt ggcattgaat aaaagggtgag ttgagagtgg
300

agcgtgttta aattgcaatc ctgcctctat ttctgtgctt gcaggggaaca gtcatcctta
360

attgctatcc tccatcatca tcatgattat ttctggtttt tctctggttg cggagaatcc
420

atactccagg tattccaatg tctcagcatt gccaggcctg tctgagcgtc aggatgtagg
480

tagtctgggc tctctgcctt ctattcttgt ccaggatact ctgccaaaag aatcatgttg
540

tggctgccac ccctcccaca aagcctcccg cttgggtcag tccaggactg gagttgggta
600

tggactgttc atgtctatcc actgctacgt cagggaaca cccactgaga gtgaccttgt
660

agactgcagt gggagacacc cttcaaaacc tctcctctcc tgtcctgaga gccaggttaa
720

aaccatcagc cccgcatcct gagtgcaaac ttttcctaac cctgctgcta agctagacac
780

c t c a c t t a c t g a g a g c c a g c
800

<210> 101

<211> 800

<212> DNA

<213> Mus musculus

<400> 101

gcagctgttg tcctggagaa cggagtcctg agcagaaaac tctcagactt tgggcaggta
60

agcctgttgg gcttccactg ctaggagaga attggttccc cacatgtgaa agcagtctgg
120

gaaatgctgg tatttccagt ctcctaaggc tactaagaaa tatgacttta tttagaggcg
180

aggaaaatgc ccaggaagtc aactgatgag actagtctta acaagttgag gatacagaaa
240

gttggggatc tgagctgcta ccaacatctg tgtgtctttg ggtggctcat tggatatctc
300

tgcctattgg ctttatcttc tgtacactga aaggaaatgg ctggtcctta gtcacctggg
360

gtgggagtcc ctatctctcc agggatactt attcaatcct ttcttctggg tatcaaaatg
420

acaagcttgt aagaaactgt cctctttcgg ctttcaggag gtgatgtcgc atgaagagaa
480

tttggggggg gggacttact cagaaccaag gagggagaaa ttaaacagag agggaaatga
540

acaggagtta gcccggagcc tgaagcacct tggggattat gctgggggtg gagggaatcc
600

attgtcctcc ctagggaggg cttgcagaac atgttctttt ctgtgatatt tgtactttcc
660

ccagattgca aatcatgggt tgtacactga gattcagtct ctggaggtaa tatgcctttt
720

ctagcttttc cttggacagg actaaggggt tgaggggtgc ctggagtcag agaaatttgt
780

g t t a a a g a a g g t t g a t a t g a
800

<210> 102

<211> 22

<212> DNA

<213> Artificial Sequence

<220>

<223> mPAH_ATG_gDNA_FAM Set 1, primer F

<400> 102

c a g c a t c a g a a g c a g a a c a t t t
22

<210> 103
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> mPAH_ATG_gDNA_FAM Set 1, primer R

<400> 103
a a a g c a c a t c a g c a g t t t c a a
21

<210> 104
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> mPAH_ATG_gDNA_FAM Set 1, probe

<400> 104
a g a t g a a a g c a a c t g a a c a t c g a c t a c g a
29

<210> 105
<211> 22
<212> DNA
<213> Artificial Sequence

<220>
<223> SV40_FAM Set 1, primer F

<400> 105
g c a a t a g c a t c a c a a a t t t c a c
22

<210> 106
<211> 25
<212> DNA
<213> Artificial Sequence

<220>

<223> SV40_FAM Set 1, primer R

<400> 106

g a t c c a g a c a t g a t a a g a t a c a t t g
25

<210> 107

<211> 28

<212> DNA

<213> Artificial Sequence

<220>

<223> SV40_FAM Set 1, probe

<400> 107

t c a c t g c a t t c t a g t t g t g g t t t g t c c a
28

<210> 108

<211> 21

<212> DNA

<213> Artificial Sequence

<220>

<223> mPah_1C_LHA_FAM Set 3, primer F

<400> 108

g c a a g c t c c a g a t c a c c a a t a
21

<210> 109

<211> 23

<212> DNA

<213> Artificial Sequence

<220>

<223> mPah_1C_LHA_FAM Set 3, primer R

<400> 109

c t g a g c a a t g c a t t c a g c a a t a a
23

<210> 110
<211> 24
<212> DNA
<213> Artificial Sequence

<220>
<223> mPah_1C_LHA_FAM Set 3, probe

<400> 110
c c c t g a a c a t c c c t t g a c a g a g c a
24

<210> 111
<211> 800
<212> DNA
<213> Mus musculus

<400> 111
agcattagct tccatttatg cagtgtaaat ggtgagaaca gccccgactg aatacccaga
60

gcatcatctc gtctgtgtca ttcatgcaca taacatatct cagcgagggtg gcccttctgt
120

cctctttgca gagaccacgc caccatacta gtacctagag aactggctgg atttcagccc
180

cgatacctcc gggcttttgc tcatgttcgc ctcataggggt catctggggtg gttgcctaag
240

gaaaagtatg tcatggagac taacttgctt ggcattgaat aaaagggtgag ttgagagtgg
300

agcgtgttta aattgcaatc ctgcctctat ttctgtgctt gcaggggaaca gtcaccta
360

attgctatcc tccatcatca tcatgattat ttctggtttt tctctggttg cggagaatcc
420

atactccagg tattccaatg tctcagcatt gccaggcctg tctgagcgtc aggatgtagg
480

tagtctgggc tctctgcctt ctattcttgt ccaggatact ctgccaaaag aatcatgttg
540

tggctgccac ccctcccaca aagcctcccg cttgggtcag tccaggactg gagttgggta
600

tggactgttc atgtctatcc actgctacgt cagggcaaca cccactgaga gtgaccttgt
660

agactgcagt gggagacacc cttcaaaacc tctcctctcc tgtcctgaga gccaggttaa
720

aaccatcagc cccgcatacct gagtgcaaac ttttcctaac cctgctgcta agctagacac
780

c t c a c t t a c t g a g a g c c a g c
800

<210> 112
<211> 800
<212> DNA
<213> Mus musculus

<400> 112
gcagctgttg tcctggagaa cggagtcctg agcagaaaac tctcagactt tgggcaggta
60

agcctgttgg gcttccactg ctaggagaga attggttccc cacatgtgaa agcagtctgg
120

gaaatgctgg tatttccagt ctccctaaggc tactaagaaa tatgacttta tttagaggcg
180

aggaaaatgc ccaggaagtc aactgatgag actagtctta acaagttgag gatacagaaa
240

gttggggatc tgagctgcta ccaacatctg tgtgtctttg ggtggctcat tggatatcctc
300

tgcctattgg ctttatcttc tgtacactga aaggaaatgg ctggtcctta gtcacctggg
360

gtgggagtcc ctatctctcc agggatactt attcaatcct ttcttctggg tatcaaaatg
420

acaagcttgt aagaaactgt cctcttttcgg ctttcaggag gtgatgtcgc atgaagagaa
480

tttggggggg gggacttact cagaaccaag gagggagaaa ttaaacagag agggaaatga
540

acaggagtta gcccggagcc tgaagcacct tggggattat gctgggggtg gagggaatcc
600

attgtcctcc ctagggaggg cttgcagaac atgttctttt ctgtgatatt tgtactttcc
660

ccagattgca aatcatgggt tgtacactga gattcagtct ctggaggtaa tatgcctttt
720

ctagcttttc cttggacagg actaaggggt tgaggggtgc ctggagtcag agaaatttgt
780

g t t a a a g a a g g t t g a t a t g a
800

<210> 113

<211> 3227

<212> DNA

<213> Artificial Sequence

<220>

<223> 006 vector correction genome

<400> 113

agcattagct tccattttatg cagtgtaaat ggtgagaaca gccccgactg aatacccgaga
60

gcatcatctc gtctgtgtca ttcatgcaca taacatatct cagcgagggtg gcccttctgt
120

cctcttttgca gagaccacgc caccatacta gtacctagag aactggctgg atttcagccc
180

cgatacctcc gggctttttgc tcatgttcgc ctcatagggt catctgggtg gttgcctaag
240

gaaaagtatg tcatggagac taacttgctt ggcattgaat aaaagggtgag ttgagagtgg
300

agcgtgttta aattgcaatc ctgcctctat ttctgtgctt gcaggggaaca gtcatcctta
360

attgctatcc tccatcatca tcatgattat ttctggtttt tctctggttg cggagaatcc
420

atactccagg tattccaatg tctcagcatt gccaggcctg tctgagcgtc aggatgtagg
480

tagtctgggc tctctgcctt ctattcttgt ccaggatact ctgccaaaag aatcatgttg
540

tggctgccac ccctcccaca aagcctcccg cttgggtcag tccaggactg gagttgggta
600

tggactgttc atgtctatcc actgctacgt cagggaaca cccactgaga gtgaccttgt
660

agactgcagt gggagacacc cttcaaaacc tctcctctcc tgtcctgaga gccaggttaa
720

aaccatcagc cccgcatcct gagtgcaaac ttttcctaac cctgctgcta agctagacac
780

ctcacttact gagagccagc atgtccaccg ctgtgctgga gaaccctggg ctggggagga
840

aactgtcaga cttcgggcag gagacttcat acattgagga taactgtaac cagaatggcg
900

ccatctctct gatcttcagc ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc
960

tgtttgagga gaacgacgtg aatctgaccc acatcgagtc ccggccttct agactgaaga
1020

aggacgagta cgagttcttt acccacctgg ataagcggtc cctgccagcc ctgacaaaca
1080

tcatcaagat cctgaggcac gacatcggag caaccgtgca cgagctgtct cgggacaaga
1140

agaaggatac cgtgccctgg ttccctcgga caatccagga gctggataga tttgccaacc
1200

agatcctgtc 1260	ttacggagca	gagctggacg	cagatcaccc	tggcttcaag	gacccagtgt
atcgggcccg 1320	gagaaagcag	tttgccgata	tgcctacaa	ttataggcac	ggacagccaa
tccctcgcgt 1380	ggagtatatg	gaggaggaga	agaagacctg	gggcacagtg	ttcaagaccc
tgaagagcct 1440	gtacaagaca	cacgcctgct	acgagtataa	ccacatcttc	cccctgctgg
agaagtattg 1500	tggctttcac	gaggacaata	tccctcagct	ggaggacgtg	agccagttcc
tgcagacctg 1560	cacaggcttt	aggctgaggc	cagtggcagg	actgctgagc	tcccgggact
tcctgggagg 1620	actggccttc	agagtgtttc	actgcaccca	gtacatcagg	cacggctcca
agccaatgta 1680	tacaccagag	cccgacatct	gtcacgagct	gctgggccac	gtgcccctgt
ttagcgatag 1740	atccttcgcc	cagttttccc	aggagatcgg	actggcatct	ctgggagcac
ctgacgagta 1800	catcgagaag	ctggccacca	tctattggtt	cacagtggag	tttggcctgt
gcaagcaggg 1860	cgatagcatc	aaggcctacg	gagcaggact	gctgtctagc	ttcggcgagc
tgcagtattg 1920	tctgtccgag	aagccaaagc	tgctgcccct	ggagctggag	aagaccgcca
tccagaacta 1980	caccgtgaca	gagttccagc	ccctgtacta	tgtggccgag	tcttttaacg
atgccaagga 2040	gaaggtgaga	aatttcgccg	ccacaatccc	taggcccttc	agcgtgcggt
acgaccctta 2100	tacccagagg	atcgaggtgc	tggataatac	acagcagctg	aagatcctgg

ctgactcaat 2160	caatagcgaa	atcggaatcc	tgtgctccgc	cctgcagaaa	atcaaatgag
aattcaaggc 2220	ctctcgagcc	tctagaacta	tagtgagtcg	tattacgtag	atccagacat
gataagatac 2280	attgatgagt	ttggacaaac	cacaactaga	atgcagtgaa	aaaaatgctt
tattttgtgaa 2340	atttgtgatg	ctattgcttt	atttgtaacc	attataagct	gcaataaaca
agttaacaac 2400	aacaattgca	ttcattttat	gtttcagggt	cagggggagg	tgtgggaggt
tttttaagct 2460	ttacgtacga	tcgtcgagca	gctgttgtcc	tggagaacgg	agtcctgagc
agaaaactct 2520	cagactttgg	gcaggtaagc	ctgttgggct	tccactgcta	ggagagaatt
ggttccccac 2580	atgtgaaagc	agtctgggaa	atgctggtat	ttccagtctc	ctaaggctac
taagaaatat 2640	gactttattt	agaggcgagg	aaaatgccca	ggaagtcaac	tgatgagact
agtcttaaca 2700	agttgaggat	acagaaagtt	ggggatctga	gctgctacca	acatctgtgt
gtctttgggt 2760	ggctcattgg	tatcctctgc	ctattggctt	tatcttctgt	acactgaaag
gaaatggctg 2820	gtccttagtc	acctgggggtg	ggagtcacct	tctctccagg	gatacttatt
caatcctttc 2880	ttctgggtat	caaaatgaca	agcttgtaag	aaactgtcct	ctttcggctt
tcaggagggtg 2940	atgtcgcacg	aagagaattt	gggggggggg	acttactcag	aaccaaggag
ggagaaatta 3000	aacagagagg	gaaatgaaca	ggagtttagcc	cggagcctga	agcaccttgg

ggattatgct gggggtggag ggaatccatt gtcctcccta gggagggctt gcagaacatg
3060

ttcttttctg tgatatattgt actttcccca gattgcaaat catggtttgt acactgagat
3120

tcagtctctg gaggtaatat gccttttcta gcttttcctt ggacaggact aaggggttga
3180

gggttgccctg gagtcagaga aatttgtgtt aaagaaggtt gatatga
3227

<210> 114

<211> 3635

<212> DNA

<213> Artificial Sequence

<220>

<223> 006 vector correction genome (+ ITRs)

<400> 114

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttccctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgca gcattagctt ccatttatgc agtgtaaagt
240

gtgagaacag ccccgactga ataccagag catcatctcg tctgtgtcat tcatgcacat
300

aacatatctc agcgagggtg cccttctgtc ctctttgcag agaccagcc accatactag
360

tacctagaga actggctgga tttcagcccc gataacctcg ggcttttgct catgttcgcc
420

tcatagggtc atctgggtgg ttgcctaagg aaaagtatgt catggagact aacttgcttg
480

gcattgaata aaaggtgagt tgagagtgga gcgtgtttta attgcaatcc tgcctctatt
540

tctgtgcttg cagggaaacag tcatccttaa ttgctatcct ccatcatcat catgattatt
600

tctggttttt ctctggttgc ggagaatcca tactccaggt attccaatgt ctcagcattg
660

ccaggcctgt ctgagcgtca ggatgtaggt agtctgggct ctctgccttc tattcttgtc
720

caggatactc tgccaaaaga atcatgttgt ggctgccacc cctcccacaa agcctccgc
780

ttgggtcagt ccaggactgg agttgggtat ggactgttca tgtctatcca ctgctacgtc
840

agggcaacac ccactgagag tgaccttgta gactgcagtg ggagacaccc ttcaaaacct
900

ctcctctcct gtcctgagag ccaggttaaa accatcagcc ccgcatcctg agtgcaaact
960

tttcctaacc ctgctgctaa gctagacacc tcacttactg agagccagca tgtccaccgc
1020

tgtgctggag aaccctgggc tggggaggaa actgtcagac ttcgggcagg agacttcata
1080

cattgaggat aactgtaacc agaatggcgc catctctctg atcttcagcc tgaaggagga
1140

agtgggcgcc ctggcaaagg tgctgcgcct gtttgaggag aacgacgtga atctgaccca
1200

catcgagtcc cggccttcta gactgaagaa ggacgagtac gagttcttta cccacctgga
1260

taagcgggtc ctgccagccc tgacaaacat catcaagatc ctgaggcacg acatcggagc
1320

aaccgtgcac gagctgtctc gggacaagaa gaaggatacc gtgccctggg tccctcggac
1380

aatccaggag ctggatagat ttgccaacca gatcctgtct tacggagcag agctggacgc
1440

agatcaccct ggcttcaagg acccagtgtg tcggggcccgg agaaagcagt ttgccgatat
1500

cgcctacaat tataggcacg gacagccaat ccctcgcgtg gagtatatgg aggaggagaa
1560

gaagacctgg ggcacagtgt tcaagaccct gaagagcctg tacaagacac acgcctgcta
1620

cgagtataac cacatcttcc ccctgctgga gaagtattgt ggctttcacg aggacaatat
1680

ccctcagctg gaggacgtga gccagttcct gcagacctgc acaggcttta ggctgaggcc
1740

agtggcagga ctgctgagct cccgggactt cctgggagga ctggccttca gagtgtttca
1800

ctgcacccag tacatcaggc acggctccaa gccaatgtat acaccagagc ccgacatctg
1860

tcacgagctg ctggggccacg tgcccctgtt tagcgataga tccttcgccc agttttccca
1920

ggagatcgga ctggcatctc tgggagcacc tgacgagtac atcgagaagc tggccaccat
1980

ctattgggtc acagtggagt ttggcctgtg caagcagggc gatagcatca aggcctacgg
2040

agcaggactg ctgtctagct tcggcgagct gcagtattgt ctgtccgaga agccaaagct
2100

gctgcccctg gagctggaga agaccgccat ccagaactac accgtgacag agttccagcc
2160

cctgtactat gtggccgagt cttttaacga tgccaaggag aaggtgagaa atttcgccgc
2220

cacaatccct aggcccttca gcgtgcggta cgacccttat acccagagga tcgaggtgct
2280

ggataataca cagcagctga agatcctggc tgactcaatc aatagcgaaa tcggaatcct
2340

gtgctccgcc ctgcagaaaa tcaaatgaga attcaaggcc tctcgagcct ctagaactat
2400

agtgagtcgt attacgtaga tccagacatg ataagataca ttgatgagtt tggacaaacc
2460

acaactagaa tgcagtgaaa aaaatgcttt atttgtgaaa tttgtgatgc tattgcttta
2520

tttgtaacca ttataagctg caataaaca gttaacaaca acaattgcat tcattttatg
2580

tttcaggttc agggggaggt gtgggaggtt ttttaagctt tacgtacgat cgtcgagcag
2640

ctgttgtcct ggagaacgga gtcctgagca gaaaactctc agactttggg caggtaagcc
2700

tgttgggctt ccactgctag gagagaattg gttccccaca tgtgaaagca gtctgggaaa
2760

tgctgggtatt tccagtctcc taaggctact aagaaatatg actttattta gaggcgagga
2820

aaatgcccag gaagtcaact gatgagacta gtcttaacaa gttgaggata cagaaagttg
2880

gggatctgag ctgctaccaa catctgtgtg tctttgggtg gctcattggg atcctctgcc
2940

tattggcttt atcttctgta cactgaaagg aaatggctgg tccttagtca cctgggggtgg
3000

gagtcacctat ctctccaggg atacttattc aatcctttct tctgggtatc aaaatgacaa
3060

gcttgtaaga aactgtcctc tttcggcttt caggaggtga tgtcgcatga agagaatttg
3120

gggggggggga cttactcaga accaaggagg gagaaattaa acagagaggg aaatgaacag
3180

gagttagccc ggagcctgaa gcaccttggg gattatgctg ggggtggagg gaatccattg
3240

tcctccctag ggagggcctt cagaacatgt tcttttctgt gatatttgta ctttcccag
3300

attgcaaadc atggtttgta cactgagatt cagtctctgg aggtaatatg ctttttctag
3360

cttttccttg gacaggacta aggggttgag ggttgccctg agtcagagaa atttggttta
3420

aagaaggttg atatgacctg caggctctaga tacgtagata agtagcatgg cgggttaadc
3480

attaactaca aggaaccctt agtgatggag ttggccactc cctctctgcg cgctcgctcg
3540

ctcactgagg ccgggcgacc aaaggctgcc cgacgcccgg gctttgcccg ggcggcctca
3600

g t g a g c g a g c g a g c g c g c a g a g a g g g a g t g g c c a a
3635

<210> 115

<211> 960

<212> DNA

<213> Homo sapiens

<400> 115

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagagggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaatagaga atcctgactg tttcagctgg gggtaagggg ggcggttat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggcttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acctggagca
480

tccggcaggc tggcctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctcctcccta gtgcgagggtt
600

aaaaccttca gcccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatctga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

<210> 116

<211> 1353

<212> DNA

<213> Artificial Sequence

<220>

<223> Silently altered PAH coding sequence

<400> 116

tccaccgctg tgctggagaa ccctgggctg gggaggaaac tgtcagactt cgggcaggag
60

acttcataca ttgaggataa ctgtaaccag aatggcgcca tctctctgat cttcagcctg
120

aaggaggaag tgggcgccct ggcaaagggtg ctgcgcctgt ttgaggagaa cgacgtgaat
180

ctgacccaca tcgagtcccg gccttctaga ctgaagaagg acgagtacga gttctttacc
240

cacctggata agcggtcctt gccagccctg acaaacatca tcaagatcct gaggcacgac
300

atcggagcaa ccgtgcacga gctgtctcgg gacaagaaga aggataccgt gccctggttc
360

cctcggacaa tccaggagct ggatagattt gcccaaccaga tcctgtctta cggagcagag
420

ctggacgcag atcaccttgg cttcaaggac ccagtgtatc gggcccggag aaagcagttt
480

gccgatatcg cctacaatta taggcacgga cagccaatcc ctcgcgtgga gtatatggag
540

gaggagaaga agacctgggg cacagtgttc aagaccctga agagcctgta caagacacac
600

gcctgctacg agtataacca catcttcccc ctgctggaga agtattgtgg ctttcacgag
660

gacaatatcc ctcagctgga ggacgtgagc cagttcctgc agacctgcac aggctttagg
720

ctgaggccag tggcaggact gctgagctcc cgggacttcc tgggaggact ggccttcaga
780

gtgtttcact gcacccagta catcaggcac ggctccaagc caatgtatac accagagccc
840

gacatctgtc acgagctgct gggccacgtg cccctgttta gcgatagatc cttcgcccag
900

ttttcccagg agatcggact ggcattctctg ggagcacctg acgagtacat cgagaagctg
960

gccaccatct attggttcac agtggagttt ggcctgtgca agcagggcga tagcatcaag
1020

gcctacggag caggactgct gtctagcttc ggcgagctgc agtattgtct gtccgagaag
1080

ccaaagctgc tgcccctgga gctggagaag accgccatcc agaactacac cgtgacagag
1140

ttccagcccc tgtactatgt ggccgagtct tttaacgatg ccaaggagaa ggtgagaaat
1200

ttcgccgcca caatccctag gcccttcagc gtgcggtacg acccttatac ccagaggatc
1260

gaggtgctgg ataatacaca gcagctgaag atcctggctg actcaatcaa tagcgaaatc
1320

g g a a t c c t g t g c t c c g c c c t g c a g a a a a t c a a a
1353

<210> 117

<211> 910

<212> DNA

<213> Homo sapiens

<400> 117

ctgggatggg atgtggaatc cttctagatt tcttttgtaa tatttataaa gtgctctcag
60

caaggtatca aaatggcaaa attgtgagta actatcctcc tttcattttg ggaagaagat
120

gagggcatgaa gagaattcag acagaaactt actcagacca ggggaggcag aaactaagca
180

gagaggaaaa tgaccaagag ttagccctgg gcatggaatg tgaaagaacc ctaaactga
240

cttggaata atgccaagg tatattccat tctccgggat ttgttgatc tttcttgagg
300

tgaagaattg cagaatacat tctttaatgt gacctacata ttaccatg ggaggaagtc
360

tgctcctgga ctcttgagat tcagtcataa agcccaggcc agggaaataa tgtaagtctg
420

caggcccctg tcatcagtag gattagggag aagagttctc agtagaaaac agggaggctg
480

gagagaaaag aatggttaat gttaacgtta atataactag aaagactgca gaacttagga
540

ctgattttta tttgaatcct taaaaaaaaa atttcttatg aaaatagtac atggctctta
600

ggagacagaa cttattgtac agaggaacag cgtgagagtc agagtgatcc cagaacaggt
660

cctggctcca tcctgcacat agttttggtg ctgctggcaa tacgggtcccc acaactgtgg
720

gaaggggtta ggggcaggga tctcatcagg aaagcatagg ggtttaaagt tctttataga
780

gcacttagaa gattgagaat ccacaaatta tattaataac aaacaaagta gtgtcgtggt
840

atatagtaaa tgtgaatttg cagacacatt tagggaaaag ttataattaa aaaaataggc
900

t g t a t a t a t a
910

<210> 118

<211> 3713

<212> DNA

<213> Artificial Sequence

<220>

<223> 032 correction vector genome

<400> 118

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagaggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggctttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggctttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acctggagca
480

tccggcaggc tggcctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgagggtt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatctga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

aagcttctga cctcttctct tcctcccaca gggcgggtacc agatctggca gcggagaggg
1020

cagaggaagt cttctaacaat gcggtgacgt ggaggagaat cccggcccta ggggtacctc
1080

caccgctgtg ctggagaacc ctgggctggg gaggaactg tcagacttcg ggcaggagac
1140

ttcatacatt gaggataact gtaaccagaa tggcgccatc tctctgatct tcagcctgaa
1200

ggaggaagtg ggcgccctgg caaagggtgct gcgcctgttt gaggagaacg acgtgaatct
1260

gaccacatc gagtcccggc cttctagact gaagaaggac gagtacgagt tctttacca
1320

cctggataag cggtccttgc cagccctgac aaacatcatc aagatcctga ggcacgacat
1380

cgaggaacc gtgcacgagc tgtctcggga caagaagaag gataccgtgc cctgggtccc
1440

tcggacaatc caggagctgg atagatttgc caaccagatc ctgtcttacg gagcagagct
1500

ggacgcagat caccctggct tcaaggaccc agtgtatcgg gcccgagaa agcagtttgc
1560

cgatatcgcc tacaattata ggcacggaca gccaatccct cgcgtggagt atatggagga
1620

ggagaagaag acctggggca cagtgttcaa gaccctgaag agcctgtaca agacacacgc
1680

ctgctacgag tataaccaca tcttccccct gctggagaag tattgtggct ttcacgagga
1740

caatatccct cagctggagg acgtgagcca gttcctgcag acctgcacag gctttaggct
1800

gaggccagtg gcaggactgc tgagctcccg ggacttcctg ggaggactgg ccttcagagt
1860

gtttcactgc acccagtaca tcaggcacgg ctccaagcca atgtatacac cagagcccga
1920

catctgtcac 1980	gagctgctgg	gccacgtgcc	cctgttttagc	gatagatcct	tcgcccagtt
ttcccaggag 2040	atcggactgg	catctctctggg	agcacctgac	gagtacatcg	agaagctggc
caccatctat 2100	tggttcacag	tggagttttgg	cctgtgcaag	cagggcgata	gcatcaaggc
ctacggagca 2160	ggactgctgt	ctagcttcgg	cgagctgcag	tattgtctgt	ccgagaagcc
aaagctgctg 2220	cccctggagc	tggagaagac	cgccatccag	aactacaccg	tgacagagtt
ccagcccctg 2280	tactatgtgg	ccgagtcctt	taacgatgcc	aaggagaagg	tgagaaattt
cgccgccaca 2340	atccctaggc	ccttcagcgt	gcggtacgac	ccttataccc	agaggatcga
ggtgctggat 2400	aatacacagc	agctgaagat	cctggctgac	tcaatcaata	gcgaaatcgg
aatcctgtgc 2460	tccgccctgc	agaaaatcaa	aggtaagcct	atccctaacc	ctctcctcgg
tctcgattct 2520	acgtgatctt	gtggaaagga	cgaaacaccg	gggaattcaa	ggcctctcga
gcctctagaa 2580	tccccgagac	gtttcgtctc	gggatcacta	tagtgagtcg	tattacgtac
acagtgcagg 2640	ggaaagaata	gtagagatcc	agacatgata	agatacattg	atgagtttgg
acaaaccaca 2700	actagaatgc	agtgaaaaaa	atgctttatt	tgtgaaattt	gtgatgctat
tgctttattt 2760	gtaaccatta	taagctgcaa	taaacaagtt	aacaacaaca	attgcattca
ttttatgttt 2820	caggttcagg	gggaggtgtg	ggagggtttt	taactgggat	gggatgtgga

atccttctag atttcttttg taatatattat aaagtgctct cagcaaggta tcaaaatggc
2880

aaaattgtga gtaactatcc tccttttcatt ttgggaagaa gatgaggcat gaagagaatt
2940

cagacagaaa cttactcaga ccaggggagg cagaaactaa gcagagagga aaatgaccaa
3000

gagttagccc tgggcatgga atgtgaaaga accctaaacg tgacttggaa ataatgcca
3060

aggtatatcc cattctccgg gatttgttgg cattttcttg aggtgaagaa ttgcagaata
3120

cattctttta tgtgacctac atatttacc atgggaggaa gtctgctcct ggactcttga
3180

gattcagtca taaagcccag gccagggaaa taatgtaagt ctgcaggccc ctgtcatcag
3240

taggattagg gagaagagtt ctcagtagaa aacagggagg ctggagagaa aagaatggtt
3300

aatgttaacg ttaatatatac tagaaagact gcagaactta ggactgattt ttatttgaat
3360

ccttaaaaaa aaaatttctt atgaaaatag tacatggctc ttaggagaca gaacttattg
3420

tacagaggaa cagcgtgaga gtcagagtga tcccagaaca ggtcctggct ccatcctgca
3480

catagttttg gtgctgctgg caatacggtc cccacaactg tgggaagggg ttaggggcag
3540

ggatctcatc aggaaagcat aggggtttta agttctttat agagcactta gaagattgag
3600

aatccacaaa ttatatatta aacaaacaaa gtagtgctgt gttatatagt aaatgtgaat
3660

ttgcagacac atttagggaa aagttataat taaaaaata ggctgtatat ata
3713

<210> 119
<211> 4139
<212> DNA
<213> Artificial Sequence

<220>
<223> 032 vector correction genome (+ ITRs)

<400> 119
ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgagc agagggagtg
120

gccaactcca tcactagggg ttccctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgct tcaggagcag ttgtgcgaat agctggagaa
240

caccaggctg gatttaaacc cagatcgctc ttacatttgc tctttacctg ctgtgctcag
300

cgttcacgtg ccctctagct gtagttttct gaagtcagcg cacagcaagg cagtgtgctt
360

agaggttaac agaagggaaa acaacaacaa caaaaatcta aatgagaatc ctgactgttt
420

cagctggggg taaggggggc ggattattca tataattggt ataccagacg gtcgcaggct
480

tagtccaatt gcagagaact cgcttcccag gcttctgaga gtcccggaag tgcctaaacc
540

tgtctaatacg acgggggcttg ggtggcccgt cgctccctgg cttcttccct ttaccaggg
600

cgggcagcga agtgggtgcct cctgcgtccc ccacaccctc cctcagcccc tcccctccgg
660

cccgtcctgg gcaggtgacc tggagcatcc ggcaggctgc cctggcctcc tgcgtcagga
720

caacgcccac gaggggcggtt actgtgcgga gatgcaccac gcaagagaca ccctttgtaa
780

ctctcttctc ctccctagtg cgagggttaa accttcagcc ccacgtgctg ttgcaaacc
840

tgctgtacc tgaggcccta aaaagccaga gacctcactc ccggggagcc agcatgtcca
900

ctgcggtcct ggaaaaccca ggcttgggca ggaaactctc tgactttgga cagggtgagcc
960

acggcagcct gagctgctca gttaggggaa tttgggcctc cagagaaaga gatctgaaga
1020

ctgctggtgc ttcctgggtt cataagctca gtaagaagtc tgaattcgtt ggaagctgat
1080

gagaatatcc aggaagtcaa cagacaaatg tcctcaacaa ttgtttctaa gtaggagaac
1140

atctgtcctc ggtggctttc acaggaaaag cttctgacct cttctcttcc tcccacaggg
1200

cggtaccaga tctggcagcg gagagggcag aggaagtctt ctaacatgcg gtgacgtgga
1260

ggagaatccc ggccctaggg gtacctccac cgctgtgctg gagaaccctg ggctggggag
1320

gaaactgtca gacttcgggc aggagacttc atacattgag gataactgta accagaatgg
1380

cgccatctct ctgatcttca gcctgaagga ggaagtgggc gccctggcaa aggtgctgcg
1440

cctgtttgag gagaacgacg tgaatctgac ccacatcgag tcccggcctt ctagactgaa
1500

gaaggacgag tacgagttct ttaccacact ggataagcgg tccctgccag ccctgacaaa
1560

catcatcaag atcctgaggc acgacatcgg agcaaccgtg cacgagctgt ctcgggacaa
1620

gaagaaggat accgtgccct ggttccctcg gacaatccag gagctggata gatttgccaa
1680

ccagatcctg tcttacggag cagagctgga cgcagatcac cctggcttca aggacccagt
1740

gtatcgggcc cggagaaagc agtttgccga tatcgacctac aattataggc acggacagcc
1800

aatccctcgc gtggagtata tggaggagga gaagaagacc tggggcacag tgttcaagac
1860

cctgaagagc ctgtacaaga cacacgcctg ctacgagtat aaccacatct tccccctgct
1920

ggagaagtat tgtggctttc acgaggacaa tatccctcag ctggaggacg tgagccagtt
1980

cctgcagacc tgcacaggct ttaggctgag gccagtggca ggactgctga gctcccggga
2040

cttcctggga ggactggcct tcagagtgtt tcaactgcacc cagtacatca ggcacggctc
2100

caagccaatg tatacaccag agcccgacat ctgtcacgag ctgctggggc acgtgcccct
2160

gtttagcgat agatccttcg cccagttttc ccaggagatc ggactggcat ctctgggagc
2220

acctgacgag tacatcgaga agctggccac catctattgg ttcacagtgg agtttggcct
2280

gtgcaagcag ggcgatagca tcaaggccta cggagcagga ctgctgtcta gcttcggcga
2340

gctgcagtat tgtctgtccg agaagccaaa gctgctgccc ctggagctgg agaagaccgc
2400

catccagaac tacaccgtga cagagttcca gcccctgtac tatgtggccg agtcttttaa
2460

cgatgccaag gagaagggtga gaaatttcgc cgccacaatc cctaggccct tcagcgtgcg
2520

gtacgaccct tatacccaga ggatcgaggt gctggataat acacagcagc tgaagatcct
2580

ggctgactca atcaatagcg aaatcggaat cctgtgctcc gccctgcaga aaatcaaagg
2640

taagcctatc cctaaccctc tcctcgggtct cgattctacg tgatcttgtg gaaaggacga
2700

aacaccgggg aattcaaggc ctctcgagcc tctagaatcc ccgagacggt tcgtctcggg
2760

atcactatag tgagtcgtat tacgtacaca gtgcagggga aagaatagta gagatccaga
2820

catgataaga tacattgatg agtttggaca aaccacaact agaatgcagt gaaaaaaaaatg
2880

ctttattttgt gaaattttgtg atgctattgc tttattttgta accattataa gctgcaataa
2940

acaagttaac aacaacaatt gcattcattt tatgttttcag gttcaggggg aggtgtggga
3000

ggtttttttaa ctgggatggg atgtggaatc cttctagatt tcttttgtaa tatttataaa
3060

gtgctctcag caaggatatca aaatggcaaa attgtgagta actatcctcc tttcattttg
3120

ggaagaagat gaggcataaa gagaattcag acagaaactt actcagacca ggggaggcag
3180

aaactaagca gagaggaaaa tgaccaagag ttagccctgg gcatggaatg tgaaagaacc
3240

ctaaacgtga cttggaaata atgccaagg tatattccat tctccgggat ttgttggcat
3300

tttcttgagg tgaagaattg cagaatacat tctttaatgt gacctacata ttacccatg
3360

ggaggaagtc tgctcctgga ctcttgagat tcagtcataa agcccaggcc agggaaataa
3420

tgtaagtctg caggcccctg tcatcagtag gattagggag aagagttctc agtagaaaac
3480

agggaggctg gagagaaaag aatggttaat gttaacgtta atataactag aaagactgca
3540

gaacttagga ctgattttta tttgaatcct taaaaaaaaa atttcttatg aaaatagtac
3600

atggctctta ggagacagaa cttattgtac agaggaacag cgtgagagtc agagtgatcc
3660

cagaacaggt cctggctcca tcctgcacat agttttggtg ctgctggcaa tacgggtccc
3720

acaactgtgg gaaggggtta ggggcaggga tctcatcagg aaagcatagg ggtttaagtc
3780

tctttataga gcacttagaa gattgagaat ccacaaatta tattaataac aaacaaagta
3840

gtgtcgtgtt atatagtaaa tgtgaatttg cagacacatt tagggaaaag ttataattaa
3900

aaaaataggc tgtatatata ctagatacgt agataagtag cctgcaggtc tagatacgta
3960

gataagtagc atggcggggtt aatcattaac tacaaggaaac ccctagtgat ggagttggcc
4020

actccctctc tgcgcgctcg ctcgctcact gaggccgggc gaccaaaggt cgcccgacgc
4080

ccgggctttg cccggggcggc ctcagtgagc gagcgagcgc gcagagaggg agtggccaa
4139

<210> 120

<211> 97

<212> DNA

<213> Artificial Sequence

<220>

<223> Chymeric MVM intron (ChiMVM)

<400> 120
gtaaggggtt aagggatggt tggttggtgg ggtattaatg ttttaattacc tggagcacct
60

gcctgaaatc acctgacctc ttctcttcct cccacag
97

<210> 121

<211> 97

<212> DNA

<213> Artificial Sequence

<220>

<223> SV40 intron

<400> 121
gtaagtttag tctttttgtc ttttatttca ggtcccggat ccggtggtgg tgcaaataca
60

agaactgctc ctcagtggat gttgccttta cttctag
97

<210> 122

<211> 521

<212> DNA

<213> Artificial Sequence

<220>

<223> Adenovirus tripartite leader intron (AdTPL)

<400> 122
gtcctcactc tcttccgcat cgctgtctgc gagggccagc tggtgggctc gcggttgagg
60

acaaactctt cgcggtcttt ccagtactct tggatcggaa acccgtcggc ctccgaacgg
120

tactccgcca ccgagggacc tgagcgagtc cgcatcgacc ggatcggaaa acctctcgag
180

aaaggcgtct aaccagtcac agtcgcaagg taggctgagc accgtggcgg gcggcagcgg
240

gtggcggtcg gggttgtttc tggcggaggt gctgctgatg atgtaattaa agtaggcggt
300

cttgagacgg cggatggtcg aggtgaggtg tggcaggctt gagatccagc tgttggggtg
360

agtactccct ctcaaaagcg ggcattactt ctgcgctaag attgtcagtt tccaaaaacg
420

aggaggattt gatattcacc tggcccgatc tggccataca cttgagtgac aatgacatcc
480

actttgcctt tctctccaca ggtgtccact cccagggtcca g
521

<210> 123

<211> 124

<212> DNA

<213> Homo sapiens

<400> 123

gtaagtatca aggttacaag acaggtttaa ggagaccaat agaaactggg catgtggaga
60

cagagaagac tcttgggttt ctgataggca ctgactctct tcctttgtcc tgttcccatt
120

t c a g
124

<210> 124

<211> 230

<212> DNA

<213> Artificial Sequence

<220>

<223> AdV/Ig chimeric intron (AdVIgG)

<400> 124

gtgagtactc cctctcaaaa gcgggcatga cttctgcgct aagattgtca gtttccaaaa
60

acgaggagga tttgatattc acctggcccg cggtgatgcc tttgaggggtg gccgcgtcca
120

tctggtcaga aaagacaatc tttttgttgt caagcttgag gtgtggcagg cttgagatct
180

ggccatacac ttgagtgaca atgacatcca ctttgccttt ctctccacag
230

<210> 125

<211> 133

<212> DNA

<213> Artificial Sequence

<220>

<223> Beta globin Ig heavy chain intron (beta-globinIg)

<400> 125

gtaagtatca aggttacaag acaggtttaa ggagaccaat agaaactggg cttgtcgaga
60

cagagaagac tcttgcgttt ctgataggca cctattggtc ttactgacat ccactttgcc
120

t t t c t c t c c a c a g
133

<210> 126

<211> 82

<212> DNA

<213> Artificial Sequence

<220>

<223> MVM intron

<400> 126

gtaagggttt aagggatggt tggttggtgg ggtattaatg ttttaattacc tggagcacct
60

g c c t g a a a t c a c t t t t t t t c a g
82

<212> DNA
<213> Homo sapiens

<400> 128
gtgagtctat gggacccttg atgttttctt tccccttctt ttctatgggt aagttcatgt
60
cataggaagg ggagaagtaa caggggtacac atattgacca aatcagggtta attttgcatt
120
tgtaatttta aaaaatgctt tcttctttta atatactttt ttgtttatct tattttctaat
180
actttcccta atctcttttct ttcagggtcaa taatgataca atgtatcatg cctctttgca
240
ccatttctaaa gaataacagt gataatttct gggttaaggc aatagcaata tttctgcata
300
taaataatttc tgcataataa ttgtaactga tgtaagaggt ttcataattgc taatagcagc
360
tacaatccag ctaccattct gctttttattt tatggttggg ataaggctgg attattctga
420
gtccaagcta ggcccttttg ctaatcatgt tcataacctt tatcttctc ccacag
476

<210> 129
<211> 299
<212> DNA
<213> Artificial Sequence

<220>
<223> tFIX intron (FIX intron)

<400> 129
gtaagtttcc ttttttaaaa tacattgagt atgcttgcct tttagatata gaaatatctg
60
atgctgtctt cttcactaaa ttttgattac atgatttgac agcaatattg aagagtctaa
120
cagccagcac gcaggttggc aactactgtg ggaacatcac agattttggc tccatgccct
180

aaagagaaat tggctttcag attatttgga ttaaaaacaa agactttctt aagagatgta
240

aaattttcat gatgttttct tttttgctaa aactaaagaa ttattctttt acatttcag
299

<210> 130

<211> 900

<212> DNA

<213> Artificial Sequence

<220>

<223> ch2BLood intron (BloodEnh)

<400> 130

gtaagtagtc tacactgggg ctaagtaaac atttactgag tgaaagaata aatacgtgca
60

gaaccagcc aacagagagc atccccggag gtggggccat gccctcaggc actcagagga
120

ggcaggcctc tggcagtagc taagcagatt cccgtctcag atgtcagccc tgcccaccct
180

tctgtgcccc cccgcccccc tggttggtgt ttgtgggaca gtttccactg tgttgccctg
240

gaaacgaggc atcctgccac caccactccc cacctccggg ctgccaacac ctaccacgcc
300

cggcttgggg gttttggctg gtttcccttg tgtcctagtc agcttgagaa accctcggat
360

gcactcgcca cttttacagt gtggcctttgc tcatctgggt gggagggtact gcctgggggtg
420

agctcatcac ctctgggttcc tcagtgcaag gcactaaagt ctcttggcac tgagttctta
480

caacttggtc caggggatat ttggtgggtg atggtcacca cccatgtgtc atggtccttg
540

agcatcccc gccacccac tgctccctat tttggcagtt actgtcaccc tctagggatg
600

cagggatgca gtttacctat ctagagtgcg gtgatcaggt ggggaagtac aaatggcaat
660

tgacagctag agagcttgaa accctccac ttggctctgc tcaccctgga ctctgggaga
720

cctcagctgc ccagcagtgg aggctggggc agcagagggg ccagacctgg gagcccagaa
780

gcctggatct gagtctgact cactgctgac ctttgacccc tggaggcttg ggcaagctga
840

ggaacttctc tgggtcttaat cctgtgggtg actgaccatg ccttcttctt tttcctacag
900

<210> 131

<211> 1977

<212> DNA

<213> Artificial Sequence

<220>

<223> 006-HCR vector PAH coding sequence

<400> 131

atgtccaccg ctgtgctgga gaaccctggg ctggggagga aactgtcaga cttcgggcag
60

gtaagtttta tggaatgtga atcataattc aatttttcaa catgcgttag gagggacatt
120

tcaaactctt ttttacccta gactttccta ccatcaccca gagtatccag ccaggagggg
180

aggggctaga gacaccagaa gtttagcagg gaggagggcg tagggattcg gggaatgaag
240

ggatgggatt cagactaggg ccaggaccca gggatggaga gaaagagatg agagtggttt
300

gggggcttgg tgacttagag aacagagctg caggctcaga ggcacacagg agtttctggg
360

ctggcccggc ccccttccaa cccctcatta tggaatccag cagctgtttg tgtgctgcct
420

ctgaagtcca 480	cactgaacaa	acttcagcct	actcatgtcc	catataaggg	aaattatgga
atcagcaaac 540	agcaaacaca	cagccctccc	tgcttgctga	ccttggacct	ggggcagagg
tcagagacct 600	ccttggctct	atgccacctc	caacatccac	tcgacccctt	ggaatttcgg
tgagagaggag 660	cagaggttgt	cctggcgtgg	tttaggtagt	gtgagagggc	tgaccatgcc
ttcttctttt 720	tcctacagga	aacttcatac	attgaggata	actgtaacca	gaatggcgcc
atctctctga 780	tcttcagcct	gaaggaggaa	gtgggcgccc	tggtcaaagg	gctgcgcctg
tttgaggaga 840	acgacgtgaa	tctgaccac	atcgagtccc	ggccttctag	actgaagaag
gacgagtacg 900	agttctttac	ccacctggat	aagcgggtccc	tgccagccct	gacaaacatc
atcaagatcc 960	tgaggcacga	catcggagca	accgtgcacg	agctgtctcg	ggacaagaag
aaggataaccg 1020	tgccctgggt	ccctcggaca	atccaggagc	tggtatagatt	tgccaaccag
atcctgtctt 1080	acggagcaga	gctggacgca	gatcaccctg	gcttcaagga	cccagtgtat
cgggcccggg 1140	gaaagcagtt	tgccgatatc	gcctacaatt	ataggcacga	acagccaatc
cctcgcgtgg 1200	agtatatgga	ggaggagaag	aagacctggg	gcacagtgtt	caagaccctg
aagagcctgt 1260	acaagacaca	cgcttgctac	gagtataacc	acatcttccc	cctgctggag
aagtattgtg 1320	gctttcacga	ggacaatatc	cctcagctgg	aggacgtgag	ccagttcctg

cagacctgca caggcttttag gctgaggcca gtggcaggac tgctgagctc ccgggacttc
1380

ctgggaggac tggccttcag agtgtttcac tgcacccagt acatcaggca cggctccaag
1440

ccaatgtata caccagagcc cgacatctgt cacgagctgc tgggccacgt gcccctgttt
1500

agcgatagat ccttcgccca gttttcccag gagatcggac tggcatctct gggagcacct
1560

gacgagtaca tcgagaagct ggccaccatc tattgggtca cagtggagtt tggcctgtgc
1620

aagcagggcg atagcatcaa ggcctacgga gcaggactgc tgtctagctt cggcgagctg
1680

cagtattgtc tgtccgagaa gccaaagctg ctgcccctgg agctggagaa gaccgccatc
1740

cagaactaca ccgtgacaga gttccagccc ctgtactatg tggccgagtc ttttaacgat
1800

gccaaggaga aggtgagaaa tttcgccgcc acaatcccta ggcccttcag cgtgcggtac
1860

gacccttata cccagaggat cgagggtgctg gataatacac agcagctgaa gatcctggct
1920

gactcaatca atagcgaaat cggaatcctg tgctccgccc tgcagaaaat caaatga
1977

<210> 132

<211> 1971

<212> DNA

<213> Artificial Sequence

<220>

<223> 032-HCR vector PAH coding sequence

<400> 132

tccaccgctg tgctggagaa ccctgggctg gggaggaaac tgtcagactt cgggcaggta
60

agtttttatgg aatgtgaatc ataattcaat ttttcaacat gcgttaggag ggacatttca
120

aactctttttt taccctagac tttcctacca tcaccagag tatccagcca ggaggggagg
180

ggctagagac accagaagtt tagcaggag gagggcgtag ggattcgggg aatgaaggga
240

tgggattcag actagggcca ggaccaggg atggagagaa agagatgaga gtggtttggg
300

ggcttggtga cttagagaac agagctgcag gctcagaggc acacaggagt ttctgggctg
360

gcccggcccc cttccaaccc ctcattatgg aatccagcag ctgtttgtgt gctgcctctg
420

aagtccacac tgaacaaact tcagcctact catgtcccat ataaggga aa ttatggaatc
480

agcaaacagc aaacacacag ccctccctgc ctgctgacct tggacctggg gcagagggtca
540

gagacctcct tggctctatg ccacctcaa catccactcg accccttgga atttcgggtg
600

agaggagcag aggttgtcct ggcgtgggttt aggtagtgtg agagggctga ccatgccttc
660

ttctttttcc tacaggaaac ttcatacatt gaggataact gtaaccagaa tggcgccatc
720

tctctgatct tcagcctgaa ggaggaagtg ggcgccctgg caaagggtgct gcgcctgttt
780

gaggagaacg acgtgaatct gaccacatc gagtcccggc cttctagact gaagaaggac
840

gagtacgagt tctttaccca cctggataag cggtcctcgc cagccctgac aaacatcatc
900

aagatcctga ggcacgacat cggagcaacc gtgcacgagc tgtctcggga caagaagaag
960

gataccgtgc cctgggttccc tcggacaatc caggagctgg atagatttgc caaccagatc
1020

ctgtcttacg gagcagagct ggacgcagat caccctggct tcaaggaccc agtgtatcgg
1080

gcccggagaa agcagtttgc cgatatcgcc tacaattata ggcacggaca gccaatccct
1140

cgcggtggagt atatggagga ggagaagaag acctggggca cagtgttcaa gaccctgaag
1200

agcctgtaca agacacacgc ctgctacgag tataaccaca tcttccccct gctggagaag
1260

tattgtggct ttcacgagga caatatccct cagctggagg acgtgagcca gttcctgcag
1320

acctgcacag gctttaggct gaggccagtg gcaggactgc tgagctcccg ggacttcctg
1380

ggaggactgg ccttcagagt gtttcaactgc acccagtaca tcaggcacgg ctccaagcca
1440

atgtatacac cagagcccga catctgtcac gagctgctgg gccacgtgcc cctgttttagc
1500

gatagatcct tcgcccagtt ttcccaggag atcggactgg catctctggg agcacctgac
1560

gagtacatcg agaagctggc caccatctat tggttcacag tggagtttgg cctgtgcaag
1620

cagggcgata gcatcaaggc ctacggagca ggactgctgt ctagcttcgg cgagctgcag
1680

tattgtctgt ccgagaagcc aaagctgctg cccctggagc tggagaagac cgccatccag
1740

aactacaccg tgacagagtt ccagcccctg tactatgtgg ccgagtcttt taacgatgcc
1800

aaggagaagg tgagaaattt cgccgccaca atccctaggc ccttcagcgt gcggtacgac
1860

ccttataccc agaggatcga ggtgctggat aatacacagc agctgaagat cctggctgac
1920

tcaatcaata gcgaaatcgg aatcctgtgc tccgccctgc agaaaatcaa a
1971

<210> 133

<211> 174

<212> DNA

<213> Artificial Sequence

<220>

<223> ID cassette

<400> 133

ggtaagccta tccctaacc tctcctcggt ctcgattcta cgtgatcttg tggaaaggac
60

gaaacaccgg ggaattcaag gcctctcgag cctctagaat ccccgagacg tttcgtctcg
120

ggatcactat agtgagtcgt attacgtaca cagtgcaggg gaaagaatag taga
174

<210> 134

<211> 800

<212> DNA

<213> Artificial Sequence

<220>

<223> 006-HCR vector correction genome

<400> 134

gcagctgttg tcctggagaa cggagtcctg agcagaaaac tctcagactt tgggcaggta
60

agcctgttgg gcttccactg ctaggagaga attggttccc cacatgtgaa agcagtctgg
120

gaaatgctgg tatttccagt ctccctaaggc tactaagaaa tatgacttta tttagaggcg
180

aggaaaatgc ccaggaagtc aactgatgag actagtctta acaagttgag gatacagaaa
240

gttggggatc tgagctgcta ccaacatctg tgtgtctttg ggtggctcat tggatatctc
300

tgcctattgg ctttatcttc tgtacactga aaggaaatgg ctggtcctta gtcacctggg
360

gtgggagtcc ctatctctcc agggatactt attcaatcct ttcttctggg tatcaaaatg
420

acaagcttgt aagaaactgt cctcttttcgg ctttcaggag gtgatgtcgc atgaagagaa
480

tttggggggg gggacttact cagaaccaag gagggagaaa ttaaacagag agggaaatga
540

acaggagtta gcccggagcc tgaagcacct tggggattat gctgggggtg gagggaatcc
600

attgtcctcc ctagggaggg cttgcagaac atgttctttt ctgtgatatt tgtactttcc
660

ccagattgca aatcatgggt tgtacactga gattcagtct ctggaggtaa tatgcctttt
720

ctagcttttc cttggacagg actaaggggt tgagggttgc ctggagtcag agaaatttgt
780

g t t a a a g a a g
800

g t t g a t a t g a

<210> 135

<211> 4253

<212> DNA

<213> Artificial Sequence

<220>

<223> 006-HCR vector correction genome (+ITRs)

<400> 135

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgca gcattagctt ccatttatgc agtgtaaagt
240

gtgagaacag ccccgactga ataccagag catcatctcg tctgtgtcat tcatgcacat
300

aacatatctc agcgagggtg cccttctgtc ctctttgcag agaccagcc accatactag
360

tacctagaga actggctgga tttcagcccc gataacctcg ggcttttgct catgttcgcc
420

tcatagggtc atctgggtgg ttgcctaagg aaaagtatgt catggagact aacttgcttg
480

gcattgaata aaaggtgagt tgagagtgga gcgtgtttaa attgcaatcc tgcctctatt
540

tctgtgcttg cagggaacag tcatccttaa ttgctatcct ccatcatcat catgattatt
600

tctggttttt ctctggttgc ggagaatcca tactccaggt attccaatgt ctcagcattg
660

ccaggcctgt ctgagcgtca ggatgtaggt agtctgggct ctctgccttc tattcttgtc
720

caggatactc tgccaaaaga atcatgttgt ggctgccacc cctcccacaa agcctcccgc
780

ttgggtcagt ccaggactgg agttgggtat ggactgttca tgtctatcca ctgctacgtc
840

agggcaacac ccactgagag tgaccttgta gactgcagtg ggagacaccc ttcaaacct
900

ctcctctcct gtcctgagag ccaggttaaa accatcagcc ccgcatcctg agtgcaaact
960

tttcctaacc ctgctgctaa gctagacacc tcacttactg agagccagca tgtccaccgc
1020

tgtgctggag aaccctgggc tggggaggaa actgtcagac ttcgggcagg taagttttat
1080

ggaatgtgaa tcataattca atttttcaac atgcgttagg agggacattt caaactcttt
1140

tttaccctag actttcctac catcacccag agtatccagc caggagggga ggggctagag
1200

acaccagaag tttagcaggg aggagggcgt agggattcgg ggaatgaagg gatgggattc
1260

agactagggc caggacccag ggatggagag aaagagatga gagtggtttg ggggcttggt
1320

gacttagaga acagagctgc aggctcagag gcacacagga gtttctgggc tggcccggcc
1380

cccttccaac ccctcattat ggaatccagc agctgtttgt gtgctgcctc tgaagtccac
1440

actgaacaaa cttcagccta ctcatgtccc atataaggga aattatggaa tcagcaaaca
1500

gcaaacacac agccctccct gcctgctgac cttggacctg gggcagaggt cagagacctc
1560

cttggtctta tgccacctcc aacatccact cgacccttg gaatttcggt ggagaggagc
1620

agaggttgtc ctggcgtggg ttaggtagtg tgagagggct gaccatgcct tcttcttttt
1680

cctacaggaa acttcataca ttgaggataa ctgtaaccag aatggcgcca tctctctgat
1740

cttcagcctg aaggaggaag tgggcgccct ggcaaagggt ctgcgccctgt ttgaggagaa
1800

cgacgtgaat ctgaccacaca tcgagtcccg gccttctaga ctgaagaagg acgagtacga
1860

gttcctttacc cacctggata agcggtcctt gccagccctg acaaacatca tcaagatcct
1920

gaggcacgac atcggagcaa ccgtgcacga gctgtctcgg gacaagaaga aggataccgt
1980

gccctgggttc cctcggacaa tccaggagct ggatagattt gcccaaccaga tcctgtctta
2040

cggagcagag ctggacgcag atcaccctgg cttcaaggac ccagtgtatc gggcccggag
2100

aaagcagttt gccgatatcg cctacaatta taggcacgaa cagccaatcc ctgcgctgga
2160

gtatatggag gaggagaaga agacctgggg cacagtgttc aagaccctga agagcctgta
2220

caagacacac gcctgctacg agtataacca catcttcccc ctgctggaga agtattgtgg
2280

ctttcacgag gacaatatcc ctcagctgga ggacgtgagc cagttcctgc agacctgcac
2340

aggctttagg ctgaggccag tggcaggact gctgagctcc cgggacttcc tgggaggact
2400

ggccttcaga gtgtttcact gcaccagta catcaggcac ggctccaagc caatgtatac
2460

accagagccc gacatctgtc acgagctgct gggccacgtg cccctgttta gcgatagatc
2520

cttcgcccag ttttcccagg agatcggact ggcattctctg ggagcacctg acgagtacat
2580

cgagaagctg gccaccatct attggttcac agtggagttt ggctgtgca agcagggcga
2640

tagcatcaag gcctacggag caggactgct gtctagcttc ggcgagctgc agtattgtct
2700

gtccgagaag ccaaagctgc tgcccctgga gctggagaag accgccatcc agaactacac
2760

cgtgacagag ttccagcccc tgtactatgt ggccgagtct tttaacgatg ccaaggagaa
2820

ggtgagaaat ttcgccgcca caatccctag gcccttcagc gtgcggtacg acccttatac
2880

ccagaggatc gaggtgctgg ataatacaca gcagctgaag atcctggctg actcaatcaa
2940

tagcgaaatc ggaatcctgt gctccgccct gcagaaaatc aaatgagaat tcaaggcctc
3000

tcgagcctct agaactatag tgagtcgtat tacgtagatc cagacatgat aagatacatt
3060

gatgagtttg gacaaaccac aactagaatg cagtgaaaaa aatgctttat ttgtgaaatt
3120

tgtgatgcta ttgctttatt tgtaaccatt ataagctgca ataaacaagt taacaacaac
3180

aattgcattc attttatggt tcaggttcag ggggaggtgt gggagggtttt ttaagcttta
3240

cgtacgatcg tcgagcagct gttgtcctgg agaacggagt cctgagcaga aaactctcag
3300

actttgggca ggtaagcctg ttgggcttcc actgctagga gagaattggt tccccacatg
3360

tgaaagcagt ctgggaaatg ctggtatttc cagtctccta aggctactaa gaaatatgac
3420

tttatattaga ggcgaggaaa atgcccagga agtcaactga tgagactagt cttaacaagt
3480

tgaggataca gaaagttggg gatctgagct gctaccaaca tctgtgtgtc tttgggtggc
3540

tcattggtat cctctgccta ttggctttat cttctgtaca ctgaaaggaa atggctggtc
3600

cttagtcacc tgggggtggga gtccctatct ctccagggat acttattcaa tcctttcttc
3660

tggttatcaa aatgacaagc ttgtaagaaa ctgtcctctt tcggctttca ggaggatg
3720

tcgcatgaag agaatttggg gggggggact tactcagaac caaggaggga gaaattaaac
3780

agagagggaa atgaacagga gttagcccgg agcctgaagc accttgggga ttatgctggg
3840

ggtggaggga atccattgtc ctccctaggg agggcttgca gaacatgttc ttttctgtga
3900

tatttgtact ttccccagat tgcaaatcat ggtttgtaca ctgagattca gtctctggag
3960

gtaatatgcc ttttctagct tttccttgga caggactaag gggttgaggg ttgcctggag
4020

tcagagaaat ttgtgttaaa gaaggttgat atgacctgca ggtctagata cgtagataag
4080

tagcatggcg ggttaatcat taactacaag gaaccctag tgatggagtt ggccactccc
4140

tctctgcgcg ctgcctcgct cactgaggcc gggcgaccaa aggtcgcccg acgcccgggc
4200

tttgcccggg cggcctcagt gagcgagcga gcgcgcagag agggagtggc caa
4253

<210> 136

<211> 4331

<212> DNA

<213> Artificial Sequence

<220>

<223> 032-HCR vector correction genome

<400> 136

gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg
60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt
120

tctgaagtca gcgcacagca aggcagtgtg cttagaggtt aacagaaggg aaaacaacaa
180

caacaaaaat ctaaattgaga atcctgactg tttcagctgg gggtaagggg ggcggattat
240

tcatataatt gttataaccag acggtcgcag gcttagtcca attgcagaga actcgcttcc
300

caggctttctg agagtcccgg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc
360

cgtcgctccc tggctttcttc cctttaccca gggcgggcag cgaagtgggtg cctcctgcgt
420

ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggtg acctggagca
480

tccggcaggc tggcctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc
540

ggagatgcac cacgcaagag acaccctttg taactctctt ctccctcccta gtgcgagggtt
600

aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc
660

agagacctca ctcccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg
720

gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg
780

gaatttgggc ctccagagaa agagatctga agactgctgg tgcttcctgg tttcataagc
840

tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa
900

atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa
960

aagcttctga cctcttctct tcctcccaca gggcgggtacc agatctggca gcggagaggg
1020

cagaggaagt cttctaacaat gcggtgacgt ggaggagaat cccggcccta ggggtacctc
1080

caccgctgtg ctggagaacc ctgggctggg gaggaactg tcagacttcg ggcaggtaag
1140

ttttatggaa tgtgaatcat aattcaattt ttcaacatgc gttaggaggg acatttcaaa
1200

ctctttttta ccctagactt tcctaccatc acccagagta tccagccagg aggggagggg
1260

ctagagacac cagaagttta gcagggagga gggcgtaggg attcggggaa tgaagggatg
1320

ggattcagac tagggccagg acccagggat ggagagaaag agatgagagt ggtttggggg
1380

cttggtgact tagagaacag agctgcaggc tcagaggcac acaggagttt ctgggctggc
1440

ccggccccct tccaaccctt cattatggaa tccagcagct gtttggtgtgc tgcctctgaa
1500

gtccacactg aacaaacttc agcctactca tgtcccatat aagggaatt atggaatcag
1560

caaacagcaa acacacagcc ctccctgcct gctgaccttg gacctggggc agaggtcaga
1620

gacctccttg gctctatgcc acctccaaca tccactcgac cccttggaat ttcggtggag
1680

aggagcagag gttgtcctgg cgtgggttag gtagtgtgag agggctgacc atgccttctt
1740

ctttttccta caggaaactt catacattga ggataactgt aaccagaatg gcgccatctc
1800

tctgatcttc agcctgaagg aggaagtggg cgccctggca aagggtgctgc gcctgtttga
1860

ggagaacgac gtgaatctga cccacatcga gtcccggcct tctagactga agaaggacga
1920

gtacgagttc 1980	tttaccacc	tggataagcg	gtccctgcca	gccctgacaa	acatcatcaa
gatcctgagg 2040	cacgacatcg	gagcaaccgt	gcacgagctg	tctcgggaca	agaagaagga
taccgtgccc 2100	tggttccctc	ggacaatcca	ggagctggat	agatttgcca	accagatcct
gtcttacgga 2160	gcagagctgg	acgcagatca	ccctggcttc	aaggacccag	tgtatcgggc
ccggagaaaag 2220	cagtttgccg	atatcgccta	caattatagg	cacggacagc	caatccctcg
cgtggagtat 2280	atggaggagg	agaagaagac	ctggggcaca	gtgttcaaga	ccctgaagag
cctgtacaag 2340	acacacgcct	gctacgagta	taaccacatc	ttccccctgc	tggagaagta
ttgtggcttt 2400	cacgaggaca	atatccctca	gctggaggac	gtgagccagt	tcctgcagac
ctgcacaggc 2460	tttaggctga	ggccagtggc	aggactgctg	agctcccggg	acttcctggg
aggactggcc 2520	ttcagagtgt	ttcactgcac	ccagtacatc	aggcacggct	ccaagccaat
gtatacacca 2580	gagcccgaca	tctgtcacga	gctgctgggc	cacgtgcccc	tgtttagcga
tagatccttc 2640	gcccagtttt	cccaggagat	cggactggca	tctctgggag	cacctgacga
gtacatcgag 2700	aagctggcca	ccatctattg	gttcacagtg	gagtttggcc	tgtgcaagca
gggcgatagc 2760	atcaaggcct	acggagcagg	actgctgtct	agcttcggcg	agctgcagta
ttgtctgtcc 2820	gagaagccaa	agctgctgcc	cctggagctg	gagaagaccg	ccatccagaa

ctacaccgtg acagagttcc agcccctgta ctatgtggcc gagtctttta acgatgccaa
2880

ggagaagggtg agaaattttcg ccgccacaat ccctaggccc ttcagcgtgc ggtacgaccc
2940

ttatacccag aggatcgagg tgctggataa tacacagcag ctgaagatcc tggctgactc
3000

aatcaatagc gaaatcggaa tcctgtgctc cgccctgcag aaaatcaaag gtaagcctat
3060

ccctaaccct ctcctcggtc tcgattctac gtgatcttgt ggaaaggacg aaacaccggg
3120

gaattcaagg cctctcgagc ctctagaatc cccgagacgt ttcgtctcgg gatcactata
3180

gtgagtcgta ttacgtacac agtgcagggg aaagaatagt agagatccag acatgataag
3240

atacattgat gagtttggac aaaccacaac tagaatgcag tgaaaaaaat gctttatttg
3300

tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata aacaagttaa
3360

caacaacaat tgcattcatt ttatgtttca gggttcagggg gaggtgtggg aggtttttta
3420

actgggatgg gatgtggaat ccttctagat ttcttttgta atatttataa agtgctctca
3480

gcaaggatc aaaatggcaa aattgtgagt aactatcctc ctttcatttt gggaagaaga
3540

tgaggcatga agagaattca gacagaaact tactcagacc aggggaggca gaaactaagc
3600

agagaggaaa atgaccaaga gttagccctg ggcatggaat gtgaaagaac cctaaacgtg
3660

acttggaat aatgcccaag gtatatcca ttctccggga tttgttggca ttttcttgag
3720

gtgaagaatt gcagaataca ttctttaatg tgacctacat atttaccat gggaggaagt
3780

ctgctcctgg actcttgaga ttcagtcata aagcccaggc cagggaaata atgtaagtct
3840

gcaggcccct gtcatcagta ggattaggga gaagagttct cagtagaaaa cagggaggct
3900

ggagagaaaa gaatgggttaa tggttaacgtt aatataacta gaaagactgc agaacttagg
3960

actgattttt atttgaatcc ttaaaaaaaaa aatttcttat gaaaatagta catggctctt
4020

aggagacaga acttattgta cagaggaaca gcgtgagagt cagagtgatc ccagaacagg
4080

tcctggctcc atcctgcaca tagttttggg gctgctggca atacgggtccc cacaactgtg
4140

ggaaggggtt aggggcaggg atctcatcag gaaagcatag gggtttaaag ttctttatag
4200

agcacttaga agattgagaa tccacaaatt atattaataa caaacaaggt agtgtcgtgt
4260

tatatagtaa atgtgaattt gcagacacat ttagggaaaa gttataatta aaaaaatagg
4320

c t g t a t a t a t a
4331

<210> 137

<211> 4737

<212> DNA

<213> Artificial Sequence

<220>

<223> 032-HCR vector correction genome (+ ITRs)

<400> 137

ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccc ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgct tcaggagcag ttgtgcgaat agctggagaa
240

caccaggctg gatttaaacc cagatcgctc ttacatttgc tctttacctg ctgtgctcag
300

cgttcacgtg ccctctagct gtagttttct gaagtcagcg cacagcaagg cagtgtgctt
360

agaggttaac agaagggaaa acaacaacaa caaaaatcta aatgagaatc ctgactgttt
420

cagctggggg taaggggggc ggattattca tataattgtt ataccagacg gtcgcaggct
480

tagtccaatt gcagagaact cgcttcccag gcttctgaga gtcccggaag tgcctaaacc
540

tgtctaatacg acgggggcttg ggtggcccgt cgctccctgg cttcttccct ttaccaggg
600

cgggcagcga agtggtgcct cctgcgtccc ccacaccctc cctcagcccc tcccctccgg
660

cccgtcctgg gcaggtgacc tggagcatcc ggcaggctgc cctggcctcc tgcgtcagga
720

caacgcccac gaggggcgtt actgtgcgga gatgcaccac gcaagagaca ccctttgtaa
780

ctctcttctc ctccctagtg cgaggttaaa accttcagcc ccacgtgctg ttgcaaacc
840

tgctgtacc tgaggcccta aaaagccaga gacctcactc ccggggagcc agcatgtcca
900

ctgcggtcct ggaaaaccca ggcttgggca ggaaactctc tgactttgga caggtagacc
960

acggcagcct 1020	gagctgctca	gttaggggaa	tttgggcctc	cagagaaaga	gatctgaaga
ctgctggtgc 1080	ttcctggttt	cataagctca	gtaagaagtc	tgaattcggt	ggaagctgat
gagaatatcc 1140	aggaagtcaa	cagacaaatg	tcctcaacaa	ttgtttctaa	gtaggagAAC
atctgtcctc 1200	ggtggctttc	acaggaaaag	cttctgacct	cttctcttcc	tcccacaggg
cggTaccaga 1260	tctggcagcg	gagagggcag	aggaagtctt	ctaacatgcg	gtgacgtgga
ggagaatccc 1320	ggccctaggg	gtacctccac	cgctgtgctg	gagaaccctg	ggctggggag
gaaactgtca 1380	gacttcgggc	aggtaagttt	tatggaatgt	gaatcataat	tcaatttttc
aacatgcggt 1440	aggagggaca	tttcaaactc	ttttttaccc	tagactttcc	taccatcacc
cagagtatcc 1500	agccaggagg	ggagggggcta	gagacaccag	aagtttagca	gggaggaggg
cgtagggatt 1560	cggggaatga	agggatggga	ttcagactag	ggccaggacc	cagggatgga
gagaaagaga 1620	tgagagtggT	ttgggggctt	ggtgacttag	agaacagagc	tgcaggctca
gaggcacaca 1680	ggagtttctg	ggctggcccg	gcccccttcc	aaccctcat	tatggaatcc
agcagctggt 1740	tgtgtgctgc	ctctgaagtc	cacactgaac	aaacttcagc	ctactcatgt
cccatataag 1800	ggaaattatg	gaatcagcaa	acagcaaaca	cacagccctc	cctgcctgct
gaccttggac 1860	ctgggggcaga	ggtcagagac	ctccttggct	ctatgccacc	tccaacatcc

actcgacccc ttggaatttc ggtggagagg agcagaggtt gtcctggcgt ggtttaggta
1920

gtgtgagagg gctgaccatg ccttcttctt tttcctacag gaaacttcat acattgagga
1980

taactgtaac cagaatggcg ccatctctct gatcttcagc ctgaaggagg aagtgggcgc
2040

cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg aatctgaccc acatcgagtc
2100

ccggccttct agactgaaga aggacgagta cgagtctctt acccacctgg ataagcggtc
2160

cctgccagcc ctgacaaaca tcatcaagat cctgaggcac gacatcggag caaccgtgca
2220

cgagctgtct cgggacaaga agaaggatac cgtgccctgg ttcctcggga caatccagga
2280

gctggataga ttgccaacc agatcctgtc ttacggagca gagctggacg cagatcacc
2340

tggttcaag gaccagtgat atcgggcccg gagaaagcag tttgccgata tcgcctacaa
2400

ttataggcac ggacagccaa tccctcgcgt ggagtatatg gaggaggaga agaagacctg
2460

gggcacagtg ttcaagaccc tgaagagcct gtacaagaca cacgcctgct acgagtataa
2520

ccacatcttc cccctgctgg agaagtattg tggttttcac gaggacaata tccctcagct
2580

ggaggacgtg agccagttcc tgcagacctg cacaggcttt aggctgaggc cagtggcagg
2640

actgctgagc tcccgggact tcctgggagg actggccttc agagtgtttc actgcacca
2700

gtacatcagg cacggctcca agccaatgta tacaccagag cccgacatct gtcacgagct
2760

gctggggccac gtgccccctgt ttagcgatag atccttcgcc cagttttccc aggagatcgg
2820

actggcatct ctgggagcac ctgacgagta catcgagaag ctggccacca tctattgggtt
2880

cacagtggag tttggcctgt gcaagcaggg cgatagcatc aaggcctacg gagcaggact
2940

gctgtctagc ttcggcgagc tgcagtattg tctgtccgag aagccaaagc tgctgcccct
3000

ggagctggag aagaccgcca tccagaacta caccgtgaca gagttccagc ccctgtacta
3060

tgtggccgag tcttttaacg atgccaagga gaagggtgaga aatttcgccg ccacaatccc
3120

taggcccttc agcgtgcggt acgaccctta taccagagg atcgaggtgc tggataatac
3180

acagcagctg aagatcctgg ctgactcaat caatagcgaa atcggaatcc tgtgctccgc
3240

cctgcagaaa atcaaaggta agcctatccc taaccctctc ctcggtctcg attctacgtg
3300

atcttgtgga aaggacgaaa caccggggaa ttcaaggcct ctcgagcctc tagaatcccc
3360

gagacgtttc gtctcgggat cactatagtg agtcgtatta cgtacacagt gcaggggaaa
3420

gaatagtaga gatccagaca tgataagata cattgatgag tttggacaaa ccacaactag
3480

aatgcagtga aaaaaatgct ttattttgtga aatttgtgat gctattgctt tatttgtaac
3540

cattataagc tgcaataaac aagttaacaa caacaattgc attcatttta tgtttcaggt
3600

tcaggggggag gtgtgggagg ttttttaact gggatgggat gtggaatcct tctagatttc
3660

ttttgtaata tttataaagt gctctcagca aggtatcaaa atggcaaaat tgtgagtaac
3720

tatcctcctt tcattttggg aagaagatga ggcatgaaga gaattcagac agaaacttac
3780

tcagaccagg ggaggcagaa actaagcaga gaggaaaatg accaagagtt agccctgggc
3840

atggaatgtg aaagaaccct aaacgtgact tggaaataat gccaaggta tattccattc
3900

tccgggattt gttggcattt tcttgagggtg aagaattgca gaatacattc tttaatgtga
3960

cctacatatt taccatggg aggaagtctg ctcttgact cttgagattc agtcataaag
4020

cccaggccag ggaaataatg taagtctgca ggcccctgtc atcagtagga ttagggagaa
4080

gagttctcag tagaaaacag ggaggctgga gagaaaagaa tggttaatgt taacgttaat
4140

ataactagaa agactgcaga acttaggact gatttttatt tgaatcctta aaaaaaaaaat
4200

ttcttatgaa aatagtacat ggctcttagg agacagaact tattgtacag aggaacagcg
4260

tgagagtcag agtgatccca gaacagggtcc tggctccatc ctgcacatag ttttggtgct
4320

gctggcaata cgggtcccccac aactgtggga aggggttagg ggcagggatc tcatcaggaa
4380

agcatagggg tttaaagttc tttatagagc acttagaaga ttgagaatcc acaaattata
4440

ttaataacaa acaaagtagt gtcgtgttat atagtaaattg tgaatttgca gacacattta
4500

gggaaaagtt ataattaaaa aaataggctg tatatatacc tgcagggtcta gatacgtaga
4560

taagtagcat ggcgggttaa tcattaacta caaggaaccc ctagtgatgg agttggccac
4620

tccctctctg cgcgctcgct cgctcactga ggccgggcga ccaaaggctg cccgacgccc
4680

gggctttgcc cgggcggcct cagtgagcga gcgagcgcgc agagagggag tggccaa
4737

<210> 138

<211> 1359

<212> DNA

<213> Artificial Sequence

<220>

<223> 006-SD.3 vector PAH coding sequence

<400> 138

atgtccaccg ctgtgctgga gaaccctggg ctggggagga aactgtcaga cttcgggcag
60

gagacttcat acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc
120

ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg
180

aatctgaccc acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt
240

accacactgg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac
300

gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg
360

ttccctcgga caatccagga gctggataga tttgccaaacc agatcctgtc ttacggagca
420

gagctggacg cagatcaccc tggcttcaag gaccagtggt atcgggcccg gagaaagcag
480

tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg
540

gaggaggaga agaagacctg gggcacagtg ttcaagaccc tgaagagcct gtacaagaca
600

cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac
660

gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt
720

aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc
780

agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag
840

cccgacatct gtcacgagct gctgggccac gtgcccctgt ttagcgatag atccttcgcc
900

cagttttccc aggagatcgg actggcatct ctgggagcac ctgacgagta catcgagaag
960

ctggccacca tctattgggt cacagtggag tttggcctgt gcaagcaggg cgatagcatc
1020

aaggcctacg gagcaggact gctgtctagc ttcggcgagc tgcagtattg tctgtccgag
1080

aagccaaagc tgctgcccct ggagctggag aagaccgcca tccagaacta caccgtgaca
1140

gagttccagc ccctgtacta tgtggccgag tcttttaacg atgccaagga gaaggtgaga
1200

aatttcgccg ccacaatccc taggcccttc agtgtgcggt acgaccctta taccagagg
1260

atcgaggtgc tggataatac acagcagctg aagatcctgg ctgactcaat caatagcgaa
1320

atcggaatcc tgtgctccgc cctgcagaaa atcaaatga
1359

<210> 139

<211> 1353

<212> DNA

<213> Artificial Sequence

<220>

<223> 032-SD.3 vector PAH coding sequence

<400> 139

tccaccgctg tgctggagaa ccctgggctg gggaggaaac tgtcagactt cgggcaggag
60

acttcataca ttgaggataa ctgtaaccag aatggcgcca tctctctgat cttcagcctg
120

aaggaggaag tgggcgccct ggcaaagggtg ctgcgccctgt ttgaggagaa cgacgtgaat
180

ctgaccacaca tcgagtcccg gccttctaga ctgaagaagg acgagtacga gttctttacc
240

cacctggata agcggtcacct gccagccctg acaaacatca tcaagatcct gaggcacgac
300

atcggagcaa ccgtgcacga gctgtctcgg gacaagaaga aggataccgt gccctgggttc
360

cctcggacaa tccaggagct ggatagattt gcccaaccaga tcctgtctta cggagcagag
420

ctggacgcag atcaccttgg cttcaaggac ccagtgtatc gggcccggag aaagcagttt
480

gccgatatcg cctacaatta taggcacgga cagccaatcc ctcgcgtgga gtatatggag
540

gaggagaaga agacctgggg cacagtgttc aagaccctga agagcctgta caagacacac
600

gcctgctacg agtataacca catcttcccc ctgctggaga agtattgtgg ctttcacgag
660

gacaatatcc ctcagctgga ggacgtgagc cagttcctgc agacctgcac aggctttagg
720

ctgaggccag tggcaggact gctgagctcc cgggacttcc tgggaggact ggccttcaga
780

gtgtttcact gcacccagta catcaggcac ggctccaagc caatgtatac accagagccc
840

gacatctgtc acgagctgct gggccacgtg cccctgttta gcgatagatc cttcgcccag
900

ttttcccagg agatcggact ggcattctctg ggagcacctg acgagtacat cgagaagctg
960

gccaccatct attggttcac agtggagttt ggcctgtgca agcagggcga tagcatcaag
1020

gcctacggag caggactgct gtctagcttc ggcgagctgc agtattgtct gtccgagaag
1080

ccaaagctgc tgcccctgga gctggagaag accgccatcc agaactacac cgtgacagag
1140

ttccagcccc tgtactatgt ggccgagtct tttaacgatg ccaaggagaa ggtgagaaat
1200

ttcgccgcca caatccctag gcccttcagt gtgcgttacg acccttatac ccagaggatc
1260

gaggtgctgg ataatacaca gcagctgaag atcctggctg actcaatcaa tagcgaaatc
1320

g g a a t c c t g t g c t c c g c c c t g c a g a a a a t c a a a
1353

<210> 140

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> 37 bp additional 3' ITR sequence from wtAAV2

<400> 140

g t a g a t a a g t a g c a t g g c g g g t t a a t c a t t a a c t a c a
37

<210> 141
<211> 180
<212> DNA
<213> Artificial Sequence

<220>
<223> 3' ITR with additional 37 bp sequence

<400> 141
gtagataagt agcatggcgg gttaatcatt aactacaagg aacccttagt gatggagttg
60
gccactccct ctctgcgcgc tcgctcgctc actgaggccg ggcgaccaa ggtcgcccga
120
cgcccgggct ttgcccgggc ggcctcagtg agcgagcgag cgcgacagaga gggagtggcc
180

<210> 142
<211> 800
<212> DNA
<213> Mus musculus

<400> 142
agcattagct tccatttatg cagtgtaaat ggtgagaaca gcccgcactg aataccaga
60
gcatcatctc gtctgtgtca ttcatgcaca taacatatct cagcgagggtg gcccttctgt
120
cctcttttgca gagaccagc caccatacta gtacctagag aactggctgg atttcagccc
180
cgatacctcc gggctttttgc tcatgttcgc ctcatagggt catctgggtg gttgcctaag
240
gaaaagtatg tcatggagac taacttgctt ggcattgaat aaaagggtgag ttgagagtgg
300
agcgtgttta aattgcaatc ctgcctctat ttctgtgctt gcagggaaca gtcaccta
360
attgctatcc tccatcatca tcatgattat ttctggtttt tctctggttg cggagaatcc
420

atactccagg tattccaatg tctcagcatt gccaggcctg tctgagcgtc aggatgtagg
480

tagtctgggc tctctgcctt ctattcttgt ccaggatact ctgccaaaag aatcatgttg
540

tggctgccac ccctcccaca aagcctcccg cttgggtcag tccaggactg gagttgggta
600

tggactgttc atgtctatcc actgctacgt cagggcaaca cccactgaga gtgaccttgt
660

agactgcagt gggagacacc cttcaaaacc tctcatctcc tgtcctgaga gccaggttaa
720

aaccatcagc cccgcacacct gagtgcaaac ttttcctaac cctgctgcta agctagacac
780

c t c a c t t a c t g a g a g c c a g c
800

<210> 143

<211> 1475

<212> DNA

<213> Artificial Sequence

<220>

<223> 006-HBB1 vector PAH coding sequence

<400> 143

atgtccaccg ctgtgctgga gaaccctggg ctgggggagga aactgtcaga cttcgggcag
60

gagacttcat acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc
120

ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg
180

aatctgaccc acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt
240

accacactgg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac
300

gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg
360

ttccctcgga caatccagga gctggataga tttgccaacc agatcctgtc ttacggagca
420

gagctggacg cagatcaccc tggcttcaag gacccagtgt atcgggcccg gagaaagcag
480

tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg
540

gaggaggaga agaagacctg gggcacagtg ttcaagacct tgaagagcct gtacaagaca
600

cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac
660

gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt
720

aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc
780

agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag
840

cccgacatct gtcacgagct gctggggccac gtgcccctgt ttagcgatag atccttcgcc
900

cagttttccc aggttggtat ccaggttaca aggcagctca caagaagaag ttgggtgctt
960

ggagacagag gtctgctttc cagcagacac taactttcag tgtcccctgt ctatgtttcc
1020

cttttttagga gatcggactg gcatctcttg gagcacctga cgagtacatc gagaagctgg
1080

ccaccatcta ttgggttcaca gtggagtttg gcctgtgcaa gcagggcgat agcatcaagg
1140

cctacggagc aggactgctg tctagcttcg gcgagctgca gtattgtctg tccgagaagc
1200

caaagctgct gcccctggag ctggagaaga ccgccatcca gaactacacc gtgacagagt
1260

tccagcccct gtactatgtg gccgagtctt ttaacgatgc caaggagaag gtgagaaatt
1320

tcgccgccac aatccctagg cccttcagcg tgcggtacga cccttatacc cagaggatcg
1380

agggtgctgga taatacacag cagctgaaga tcctggctga ctcaatcaat agcgaaatcg
1440

g a a t c c t g t g c t c c g c c c t g c a g a a a a t c a a a t g a
1475

<210> 144

<211> 800

<212> DNA

<213> Mus musculus

<400> 144

gcagctgttg tcctggagaa cggagtcctg agcagaaaac tctcagactt tgggcaggta
60

agcctgttgg gcttccactg ctaggagaga attggttccc cacatgtgaa agcagtctgg
120

gaaatgctgg tatttccagt ctccctaaggc tactaagaaa tatgacttta tttagaggcg
180

agaaaaatgc ccaggaagtc aactgatgag actagtctta acaagttgag gatacagaaa
240

gttggggatc tgagctgcta ccaacatctg tgtgtctttg ggtggctcat tggatcctc
300

tgcctattgg ctttatcttc tgtacactga aaggaaatgg ctggtcctta gtcacctggg
360

gtgggagtcc ctatctctcc agggatactt attcaatcct ttcttctggg tatcaaaatg
420

acaagcttgt aagaaactgt cctctttcgg ctttcaggag gtgatgtcgc atgaagagaa
480

tttggggggg gggacttact cagaaccaag gagggagaaa ttaaacagag agggaaatga
540

acaggagtta gcccggagcc tgaagcacct tggggattat gctgggggtg gagggaatcc
600

attgtcctcc ctagggaggg cttgcagaac atgttctttt ctgtgatatt tgtactttcc
660

ccagattgca aatcatgggt tgtacactga gattcagtct ctggaggtaa tatgcctttt
720

ctagcttttc cttggacagg actaaggggt tgaggggtgc ctggagtcag agaaatttgt
780

g t t a a a g a a g g t t g a t a t g a
800

<210> 145

<211> 3343

<212> DNA

<213> Artificial Sequence

<220>

<223> 006-HBB1 vector correction genome

<400> 145

agcattagct tccatttatg cagtgtaaat ggtgagaaca gccccgactg aataccaga
60

gcatcatctc gtctgtgtca ttcatgcaca taacatatct cagcgagggtg gcccttctgt
120

cctctttgca gagaccagc caccatacta gtacctagag aactggctgg atttcagccc
180

cgatacctcc gggcttttgc tcatgttcgc ctcatagggt catctgggtg gttgcctaag
240

gaaaagtatg tcatggagac taacttgctt ggcattgaat aaaagggtgag ttgagagtgg
300

agcgtgttta aattgcaatc ctgcctctat ttctgtgctt gcaggggaaca gtcatcctta
360

attgctatcc tccatcatca tcatgattat ttctgggttt tctctgggtg cggagaatcc
420

atactccagg tattccaatg tctcagcatt gccaggcctg tctgagcgtc aggatgtagg
480

tagtctgggc tctctgcctt ctattcttgt ccaggatact ctgccaaaag aatcatgttg
540

tggctgccac ccctcccaca aagcctcccg cttgggtcag tccaggactg gagttgggta
600

tggactgttc atgtctatcc actgctacgt cagggaaca cccactgaga gtgaccttgt
660

agactgcagt gggagacacc cttcaaaacc tctcatctcc tgtcctgaga gccaggttaa
720

aaccatcagc cccgcctcct gagtgcaaac ttttcctaac cctgctgcta agctagacac
780

ctcacttact gagagccagc atgtccaccg ctgtgctgga gaaccctggg ctggggagga
840

aactgtcaga cttcggggcag gagacttcat acattgagga taactgtaac cagaatggcg
900

ccatctctct gatcttcagc ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc
960

tgtttgagga gaacgacgtg aatctgaccc acatcgagtc ccggccttct agactgaaga
1020

aggacgagta cgagttcttt acccacctgg ataagcggtc cctgccagcc ctgacaaaca
1080

tcatcaagat cctgaggcac gacatcggag caaccgtgca cgagctgtct cgggacaaga
1140

agaaggatac cgtgccctgg ttccctcgga caatccagga gctggataga tttgccaacc
1200

agatcctgtc ttacggagca gagctggacg cagatcacc cggcttcaag gaccagtggt
1260

atcgggcccg gagaaagcag tttgccgata tcgcctacaa ttataggcac ggacagccaa
1320

tccctcgcgt ggagtatatg gaggaggaga agaagacctg gggcacagtg ttcaagaccc
1380

tgaagagcct gtacaagaca cacgcctgct acgagtataa ccacatcttc cccctgctgg
1440

agaagtattg tggctttcac gaggacaata tccctcagct ggaggacgtg agccagttcc
1500

tgcagacctg cacaggcttt aggctgaggc cagtggcagg actgctgagc tcccgggact
1560

tcctgggagg actggccttc agagtgtttc actgcacca gtacatcagg cacggctcca
1620

agccaatgta tacaccagag cccgacatct gtcacgagct gctgggccac gtgcccctgt
1680

ttagcgatag atccttcgcc cagttttccc aggttggtat ccaggttaca aggcagctca
1740

caagaagaag ttgggtgctt ggagacagag gtctgctttc cagcagacac taactttcag
1800

tgtcccctgt ctatgtttcc ctttttagga gatcggactg gcatctctgg gagcacctga
1860

cgagtacatc gagaagctgg ccaccatcta ttggttcaca gtggagtttg gcctgtgcaa
1920

gcagggcgat agcatcaagg cctacggagc aggactgctg tctagcttcg gcgagctgca
1980

gtattgtctg tccgagaagc caaagctgct gcccctggag ctggagaaga ccgccatcca
2040

gaactacacc gtgacagagt tccagcccct gtactatgtg gccgagtctt ttaacgatgc
2100

caaggagaag gtgagaaatt tcgccgccac aatccctagg cccttcagcg tgcggtacga
2160

cccttataacc cagaggatcg aggtgctgga taatacacag cagctgaaga tcctggctga
2220

ctcaatcaat agcgaaatcg gaatcctgtg ctccgccctg cagaaaatca aatgagaatt
2280

caaggcctct cgagcctcta gaactatagt gagtcgtatt acgtagatcc agacatgata
2340

agatacattg atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt
2400

tgtgaaattt gtgatgctat tgctttattht gtaaccatta taagctgcaa taaacaagtt
2460

aacaacaaca attgcattca ttttatgttt caggttcagg gggagggtgtg ggagggttttt
2520

taagctttac gtacgatcgt cgagcagctg ttgtcctgga gaacggagtc ctgagcagaa
2580

aactctcaga ctttgggcag gtaagcctgt tgggcttcca ctgctaggag agaattggtt
2640

ccccacatgt gaaagcagtc tgggaaatgc tggatatttc agtctcctaa ggctactaag
2700

aaatatgact ttatttagag gcgagaaaaa tgcccaggaa gtcaactgat gagactagtc
2760

ttaacaagtt gaggatacag aaagttgggg atctgagctg ctaccaacat ctgtgtgtct
2820

ttgggtggct cattgggtatc ctctgcctat tggctttatc ttctgtacac tgaaaggaaa
2880

tggctgggtcc ttagtcacct ggggtgggag tccctatctc tccaggata cttattcaat
2940

cctttcttct gggtatcaaa atgacaagct tgtaagaaac tgtcctcttt cggctttcag
3000

gagggtgatgt cgcatagaaga gaatttgggg ggggggactt actcagaacc aaggaggagg
3060

aaattaaaca gagagggaaa tgaacaggag ttagcccgga gcctgaagca ccttggggat
3120

tatgctgggg gtggagggaa tccattgtcc tccctagggg gggcttgcag aacatgttct
3180

tttctgtgat atttgtactt tccccagatt gcaaactcatg gtttgtacac tgagattcag
3240

tctctggagg taatatgcct tttctagctt ttccttggac aggactaagg ggttgagggg
3300

tgccctggagt cagagaaatt tgtgttaaag aagggttgata tga
3343

<210> 146

<211> 3751

<212> DNA

<213> Artificial Sequence

<220>

<223> 006-HBB1 vector correction genome (+ ITRs)

<400> 146

ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggcgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgagc agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcata tgcggccgca gcattagctt ccatttatgc agtgtaaagt
240

gtgagaacag ccccgactga ataccagag catcatctcg tctgtgtcat tcatgcacat
300

aacatatctc agcgagggtg cccttctgtc ctctttgcag agaccagcc accatactag
360

tacctagaga actggctgga tttcagcccc gataacctcg ggcttttgct catgttcgcc
420

tcataagggtc atctgggttg ttgcctaagg aaaagtatgt catggagact aacttgcttg
480

gcattgaata aaaggtgagt tgagagtgga gcgtgtttta attgcaatcc tgcctctatt
540

tctgtgcttg caggaacag tcatccttaa ttgctatcct ccatcatcat catgattatt
600

tctggttttt ctctggttgc ggagaatcca tactccaggt attccaatgt ctcagcattg
660

ccaggcctgt ctgagcgtca ggatgtaggt agtctgggct ctctgccttc tattcttgct
720

caggatactc tgccaaaaga atcatgttgt ggctgccacc cctcccacaa agcctcccgc
780

ttgggtcagt ccaggactgg agttgggtat ggactgttca tgtctatcca ctgctacgtc
840

agggcaacac ccactgagag tgaccttgta gactgcagtg ggagacaccc ttcaaacct
900

ctcatctcct gtcctgagag ccagggttaa accatcagcc ccgcatcctg agtgcaaact
960

tttcctaacc ctgctgctaa gctagacacc tcacttactg agagccagca tgtccaccgc
1020

tgtgctggag aaccctgggc tggggaggaa actgtcagac ttcgggcagg agacttcata
1080

cattgaggat aactgtaacc agaatggcgc catctctctg atcttcagcc tgaaggagga
1140

agtgggcgcc ctggcaaagg tgctgcgcct gtttgaggag aacgacgtga atctgaccca
1200

catcgagtcc cggccttcta gactgaagaa ggacgagtac gagttcttta cccacctgga
1260

taagcgggtc ctgccagccc tgacaaacat catcaagatc ctgaggcacg acatcggagc
1320

aaccgtgcac gagctgtctc gggacaagaa gaaggatacc gtgccctggt tccctcggac
1380

aatccaggag ctggatagat ttgccaacca gatcctgtct tacggagcag agctggacgc
1440

agatcaccct ggcttcaagg acccagtgtat tcggggcccgg agaaagcagt ttgccgatat
1500

cgcctacaat tataggcacg gacagccaat ccctcgcgtg gagtatatgg aggaggagaa
1560

gaagacctgg ggcacagtgt tcaagaccct gaagagcctg tacaagacac acgcctgcta
1620

cgagtataac cacatcttcc ccctgctgga gaagtattgt ggctttcacg aggacaatat
1680

ccctcagctg gaggacgtga gccagttcct gcagacctgc acaggcttta ggctgaggcc
1740

agtggcagga ctgctgagct cccgggactt cctgggagga ctggccttca gagtgtttca
1800

ctgcacccag tacatcaggc acggctccaa gccaatgtat acaccagagc ccgacatctg
1860

tcacgagctg ctggggccacg tgcccctggt tagcgataga tccttcgccc agttttccca
1920

ggttggtatc caggttacaa ggcagctcac aagaagaagt tgggtgcttg gagacagagg
1980

tctgctttcc agcagacact aactttcagt gtcccctgtc tatgtttccc tttttaggag
2040

atcggactgg catctctggg agcacctgac gagtacatcg agaagctggc caccatctat
2100

tggttcacag tggagtttgg cctgtgcaag cagggcgata gcatcaaggc ctacggagca
2160

ggactgctgt ctagcttcgg cgagctgcag tattgtctgt ccgagaagcc aaagctgctg
2220

cccctggagc tggagaagac cgccatccag aactacaccg tgacagagtt ccagcccctg
2280

tactatgtgg ccgagtcctt taacgatgcc aaggagaagg tgagaaattt cgccgccaca
2340

atccctaggc ccttcagcgt gcggtacgac ccttataccc agaggatcga ggtgctggat
2400

aatacacagc agctgaagat cctggctgac tcaatcaata gcgaaatcgg aatcctgtgc
2460

tccgccctgc agaaaatcaa atgagaattc aaggcctctc gagcctctag aactatagtg
2520

agtcgtatta cgtagatcca gacatgataa gatacattga tgagtttgga caaaccacaa
2580

ctagaatgca gtgaaaaaaaa tgctttatct gtgaaatttg tgatgctatt gctttatttg
2640

taaccattat aagctgcaat aaacaagtta acaacaacaa ttgcattcat tttatgtttc
2700

aggttcaggg ggaggtgtgg gaggtttttt aagctttacg tacgatcgtc gagcagctgt
2760

tgtcctggag aacggagtcc tgagcagaaa actctcagac tttgggcagg taagcctggt
2820

gggcttccac tgctaggaga gaattgggtc cccacatgtg aaagcagtct gggaaatgct
2880

ggtatttcca gtctcctaag gctactaaga aatatgactt tatttagagg cgagaaaaat
2940

gccaggaag tcaactgatg agactagtct taacaagttg aggatacaga aagttgggga
3000

tctgagctgc taccaacatc tgtgtgtctt tgggtggctc attgggtatcc tctgcctatt
3060

ggctttatct tctgtacact gaaaggaaat ggctgggtcct tagtcacctg ggggtgggagt
3120

ccctatctct ccagggatac ttattcaatc ctttcttctg ggtatcaaaa tgacaagctt
3180

gtaagaaact gtcctctttc ggctttcagg aggtgatgtc gcatgaagag aatttggggg
3240

gggggactta ctcagaacca aggagggaga aattaaacag agaggggaaat gaacaggagt
3300

tagcccggag cctgaagcac cttggggatt atgctggggg tggagggaat ccattgtcct
3360

ccctagggag ggcttgcaga acatgttctt ttctgtgata tttgtacttt ccccagattg
3420

caaatcatgg tttgtacact gagattcagt ctctggaggt aatatgcctt ttctagcttt
3480

tccttggaca ggactaaggg gttgagggtt gcctggagtc agagaaattt gtgttaaaga
3540

aggttgatat gacctgcagg tctagatacg tagataagta gcatggcggg ttaatcatta
3600

actacaagga acccctagtg atggagttgg ccactccctc tctgcgcgct cgctcgctca
3660

ctgaggccgg gcgaccaaag gtcgcccgcac gcccgggctt tgcccggggcg gcctcagtga
3720

g c g a g c g a g c g c g c a g a g a g g g a g t g g c c a a
3751

<210> 147

<211> 1607

<212> DNA

<213> Artificial Sequence

<220>

<223> 006 vector editing element

<400> 147

atgtccaccg ctgtgctgga gaaccctggg ctgggggagga aactgtcaga cttcgggcag
60

gagacttcat acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc
120

ctgaaggagg aagtgggcg cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg
180

aatctgaccc acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt
240

accacactgg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac
300

gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg
360

ttccctcgga caatccagga gctggataga ttgccaacc agatcctgtc ttacggagca
420

gagctggacg cagatcaccc tggcttcaag gaccagtggt atcgggcccg gagaaagcag
480

tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg
540

gaggaggaga agaagacctg gggcacagtg ttcaagaccc tgaagagcct gtacaagaca
600

cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac
660

gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt
720

aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc
780

agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag
840

cccgacatct gtcacgagct gctggggccac gtgcccctgt ttagcgatag atccttcgcc
900

cagttttccc aggagatcgg actggcatct ctgggagcac ctgacgagta catcgagaag
960

ctggccacca tctattgggtt cacagtggag tttggcctgt gcaagcaggg cgatagcatc
1020

aaggcctacg gagcaggact gctgtctagc ttcggcgagc tgcagtattg tctgtccgag
1080

aagccaaagc tgctgcccct ggagctggag aagaccgcca tccagaacta caccgtgaca
1140

gagttccagc ccctgtacta tgtggccgag tctttttaacg atgccaagga gaagggtgaga
1200

aatttcgccg ccacaatccc taggcccttc agcgtgcggt acgaccctta taccagagg
1260

atcgaggtgc tggataatac acagcagctg aagatcctgg ctgactcaat caatagcgaa
1320

atcggaatcc tgtgctccgc cctgcagaaa atcaaattgag aattcaaggc ctctcgagcc
1380

tctagaacta tagtgagtcg tattacgtag atccagacat gataagatac attgatgagt
1440

ttggacaaac cacaactaga atgcagtgaa aaaaatgctt tatttgtgaa atttgtgatg
1500

ctattgcttt atttgtaacc attataagct gcaataaaca agttaacaac aacaattgca
1560

ttcattttat gtttcagggtt cagggggagg tgtgggaggt tttttaa
1607

<210> 148

<211> 1837

<212> DNA

<213> Artificial Sequence

<220>

<223> 032 vector editing element

<400> 148

ctgacctctt ctcttcctcc cacagggcgg taccagatct ggcagcggag agggcagagg
60

aagtcttcta acatgcggtg acgtggagga gaatcccggc cctaggggta cctccaccgc
120

tgtgctggag aaccctgggc tggggaggaa actgtcagac ttcgggcagg agacttcata
180

cattgaggat aactgtaacc agaatggcgc catctctctg atcttcagcc tgaaggagga
240

agtgggcgcc ctggcaaagg tgctgcgctt gtttgaggag aacgacgtga atctgaccga
300

catcgagtcc cggccttcta gactgaagaa ggacgagtag gagttcttta cccacctgga
360

taagcgggtc ctgccagccc tgacaaacat catcaagatc ctgaggcacg acatcggagc
420

aaccgtgcac gagctgtctc gggacaagaa gaaggatacc gtgccctggg tccctcggac
480

aatccaggag ctggatagat ttgccaacca gatcctgtct tacggagcag agctggacgc
540

agatcacctt ggcttcaagg acccagtgtg tcggggcccgg agaaagcagt ttgccgatat
600

cgcctacaat tataggcacg gacagccaat ccctcgcgtg gagtatatgg aggaggagaa
660

gaagacctgg ggcacagtgt tcaagaccct gaagagcctg tacaagacac acgcctgcta
720

cgagtataac cacatcttcc ccctgctgga gaagtattgt ggctttcacg aggacaatat
780

ccctcagctg gaggacgtga gccagttcct gcagacctgc acaggcttta ggctgaggcc
840

agtggcagga ctgctgagct cccgggactt cctgggagga ctggccttca gagtgtttca
900

ctgcacccag tacatcaggc acggctccaa gccaatgtat acaccagagc ccgacatctg
960

tcacgagctg ctgggccacg tgcccctggt tagcgataga tccttcgccc agttttccca
1020

ggagatcgga ctggcatctc tgggagcacc tgacgagtac atcgagaagc tggccaccat
1080

ctattggttc acagtggagt ttggcctgtg caagcagggc gatagcatca aggcctacgg
1140

agcaggactg ctgtctagct tcggcgagct gcagtattgt ctgtccgaga agccaaagct
1200

gctgcccctg gagctggaga agaccgccat ccagaactac accgtgacag agttccagcc
1260

cctgtactat gtggccgagt cttttaacga tgccaaggag aaggtgagaa atttcgccgc
1320

cacaatccct aggcccttca gcgtgcggta cgacccttat acccagagga tcgaggtgct
1380

ggataataca cagcagctga agatcctggc tgactcaatc aatagcgaaa tcggaatcct
1440

gtgctccgcc ctgcagaaaa tcaaaggtaa gcctatccct aaccctctcc tcggtctcga
1500

ttctacgtga tcttgtggaa aggacgaaac accggggaat tcaaggcctc tcgagcctct
1560

agaatccccg agacgtttcg tctcgggac actatagtga gtcgtattac gtacacagtg
1620

caggggaaag aatagtagag atccagacat gataagatac attgatgagt ttggacaaac
1680

cacaactaga atgcagtgaa aaaaatgctt tatttgtgaa atttgtgatg ctattgcttt
1740

atttghtaacc attataagct gcaataaaca agttaacaac aacaattgca ttcattttat
1800

g t t t c a g g t t c a g g g g g a g g t g t g g g a g g t t t t t a a
1837

<210> 149
<211> 2225
<212> DNA
<213> Artificial Sequence

<220>
<223> 006-HCR vector editing element

<400> 149
atgtccaccg ctgtgctgga gaaccctggg ctgggggagga aactgtcaga cttcgggcag
60

gtaagtttta tggaatgtga atcataattc aatttttcaa catgcgttag gagggacatt
120

tcaaactctt ttttacccta gactttccta ccatcaccca gagtatccag ccaggagggg
180

aggggctaga gacaccagaa gtttagcagg gaggagggcg tagggattcg gggaatgaag
240

ggatgggatt cagactaggg ccaggaccca gggatggaga gaaagagatg agagtggttt
300

gggggcttgg tgacttagag aacagagctg caggctcaga ggcacacagg agtttctggg
360

ctggcccggc ccccttccaa cccctcatta tggaatccag cagctgtttg tgtgctgcct
420

ctgaagtcca cactgaacaa acttcagcct actcatgtcc catataaggg aaattatgga
480

atcagcaaac agcaaacaca cagccctccc tgccctgctga ccttggacct ggggcagagg
540

tcagagacct ccttggctct atgccacctc caacatccac tcgacccctt ggaatttcgg
600

tgagagaggag cagaggttgt cctggcggtg tttaggtagt gtgagagggc tgaccatgcc
660

ttctttcttt tcctacagga aacttcatac attgaggata actgtaacca gaatggcgcc
720

atctctctga	tcttcagcct	gaaggaggaa	gtgggcgccc	tggcaaaggt	gctgcgccctg
780					
tttgaggaga	acgacgtgaa	tctgaccac	atcgagtccc	ggccttctag	actgaagaag
840					
gacgagtacg	agttctttac	ccacctggat	aagcgggtccc	tgccagccct	gacaaacatc
900					
atcaagatcc	tgaggcacga	catcggagca	accgtgcacg	agctgtctcg	ggacaagaag
960					
aaggataccg	tgccctgggt	ccctcggaca	atccaggagc	tggatagatt	tgccaaccag
1020					
atcctgtctt	acggagcaga	gctggacgca	gatcaccctg	gcttcaagga	cccagtgtat
1080					
cgggcccgga	gaaagcagtt	tgccgatatc	gcctacaatt	ataggcacga	acagccaatc
1140					
cctcgcgtgg	agtatatgga	ggaggagaag	aagacctggg	gcacagtgtt	caagacctg
1200					
aagagcctgt	acaagacaca	cgctgctac	gagtataacc	acatcttccc	cctgctggag
1260					
aagtattgtg	gctttcacga	ggacaatatc	cctcagctgg	aggacgtgag	ccagttcctg
1320					
cagacctgca	caggctttag	gctgaggcca	gtggcaggac	tgctgagctc	ccgggacttc
1380					
ctgggaggac	tggccttcag	agtgtttcac	tgcacccagt	acatcaggca	cggctccaag
1440					
ccaatgtata	caccagagcc	cgacatctgt	cacgagctgc	tgggccacgt	gcccctgttt
1500					
agcgatagat	ccttcgccc	gttttcccag	gagatcggac	tggcatctct	gggagcacct
1560					
gacgagtaca	tcgagaagct	ggccaccatc	tattgggtca	cagtggagtt	tggcctgtgc
1620					

aagcagggcg atagcatcaa ggcctacgga gcaggactgc tgtctagctt cggcgagctg
1680

cagtattgtc tgtccgagaa gccaaagctg ctgcccctgg agctggagaa gaccgccatc
1740

cagaactaca ccgtgacaga gttccagccc ctgtactatg tggccgagtc ttttaacgat
1800

gccaaaggaga aggtgagaaa tttcgccgcc acaatcccta ggcccttcag cgtgcggtac
1860

gacccttata cccagaggat cgagggtgctg gataatacac agcagctgaa gatcctggct
1920

gactcaatca atagcgaaat cggaatcctg tgctccgccc tgcagaaaat caaatgagaa
1980

ttcaaggcct ctcgagcctc tagaactata gtgagtcgta ttacgtagat ccagacatga
2040

taagatacat tgatgagttt ggacaaacca caactagaat gcagtgaaaa aaatgcttta
2100

tttgtgaaat ttgtgatgct attgctttat ttgtaaccat tataagctgc aataaacaag
2160

ttaacaacaa caattgcatt cattttatgt ttcagggtca gggggagggtg tgggagggtt
2220

t t t a a
2225

<210> 150

<211> 2455

<212> DNA

<213> Artificial Sequence

<220>

<223> 032-HCR vector editing element

<400> 150

ctgacctctt ctcttcctcc cacagggcgg taccagatct ggcagcggag agggcagagg
60

aagtcttcta acatgcggtg acgtggagga gaatcccggc cctaggggta cctccaccgc
120

tgtgctggag aaccctgggc tggggaggaa actgtcagac ttcgggcagg taagttttat
180

ggaatgtgaa tcataattca atttttcaac atgcgttagg agggacattt caaactcttt
240

tttaccctag actttcctac catcacccag agtatccagc caggagggga ggggctagag
300

acaccagaag tttagcaggg aggagggcgt agggattcgg ggaatgaagg gatgggattc
360

agactagggc caggacccag ggatggagag aaagagatga gagtggtttg ggggcttggt
420

gacttagaga acagagctgc aggctcagag gcacacagga gtttctgggc tggcccggcc
480

cccttccaac ccctcattat ggaatccagc agctgtttgt gtgctgcctc tgaagtccac
540

actgaacaaa cttcagccta ctcatgtccc atataaggga aattatggaa tcagcaaaca
600

gcaaacacac agccctccct gcctgctgac cttggacctg gggcagaggt cagagacctc
660

cttggtctta tgccacctcc aacatccact cgacccttg gaatttcggt ggagaggagc
720

agaggttgtc ctggcgtggt ttaggtagtg tgagagggct gaccatgcct tcttcttttt
780

cctacaggaa acttcataca ttgaggataa ctgtaaccag aatggcgcca tctctctgat
840

cttcagcctg aaggaggaag tgggcgccct ggcaaagggt ctgcgcctgt ttgaggagaa
900

cgacgtgaat ctgaccacaca tcgagtcccg gccttctaga ctgaagaagg acgagtacga
960

gttcctttacc cacctggata agcggtcctt gccagccctg acaaacatca tcaagatcct
1020

gaggcacgac atcggagcaa ccgtgcacga gctgtctcgg gacaagaaga aggataccgt
1080

gccctgggttc cctcggacaa tccaggagct ggatagattt gccaaccaga tcctgtctta
1140

cggagcagag ctggacgcag atcacccctgg cttcaaggac ccagtgtatc gggcccggag
1200

aaagcagttt gccgatatcg cctacaatta taggcacgga cagccaatcc ctgcgctgga
1260

gtatatggag gaggagaaga agacctgggg cacagtgttc aagaccctga agagcctgta
1320

caagacacac gcctgctacg agtataacca catcttcccc ctgctggaga agtattgtgg
1380

ctttcacgag gacaatatcc ctcagctgga ggacgtgagc cagttcctgc agacctgcac
1440

aggctttagg ctgaggccag tggcaggact gctgagctcc cgggacttcc tgggaggact
1500

ggccttcaga gtgtttcact gcaccagta catcaggcac ggctccaagc caatgtatac
1560

accagagccc gacatctgtc acgagctgct gggccacgtg cccctgttta gcgatagatc
1620

cttcgcccag ttttcccagg agatcggact ggcattctctg ggagcacctg acgagtacat
1680

cgagaagctg gccaccatct attggttcac agtggagttt ggcctgtgca agcagggcga
1740

tagcatcaag gcctacggag caggactgct gtctagcttc ggcgagctgc agtattgtct
1800

gtccgagaag ccaaagctgc tgcccctgga gctggagaag accgccatcc agaactacac
1860

cgtagacagag ttccagcccc tgtactatgt ggccgagtct tttaacgatg ccaaggagaa
1920

ggtgagaaat ttcgccgcca caatccctag gcccttcagc gtgcggtacg acccttatac
1980

ccagaggatc gaggtgctgg ataatacaca gcagctgaag atcctggctg actcaatcaa
2040

tagcgaaatc ggaatcctgt gctccgccct gcagaaaatc aaaggtaagc ctatccctaa
2100

ccctctcctc ggtctcgatt ctacgtgatc ttgtggaaag gacgaaacac cggggaattc
2160

aaggcctctc gagcctctag aatccccgag acgtttcgtc tcgggatcac tatagtgagt
2220

cgtattacgt acacagtgca ggggaaagaa tagtagagat ccagacatga taagatacat
2280

tgatgagttt ggacaaacca caactagaat gcagtgaaaa aaatgcttta tttgtgaaat
2340

ttgtgatgct attgctttat ttgtaaccat tataagctgc aataaacaag ttaacaacaa
2400

caattgcatt catTTTTatgt ttcaggttca gggggagggtg tgggagggtt tttaa
2455

<210> 151

<211> 1723

<212> DNA

<213> Artificial Sequence

<220>

<223> 006-HBB1 vector editing element

<400> 151

atgtccaccg ctgtgctgga gaaccctggg ctggggagga aactgtcaga cttcgggcag
60

gagacttcac acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc
120

ctgaaggagg aagtgggcg cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg
180

aatctgaccc acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt
240

acccacctgg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac
300

gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg
360

ttccctcgga caatccagga gctggataga ttgccaacc agatcctgtc ttacggagca
420

gagctggacg cagatcaccc tggcttcaag gaccagtggt atcgggcccg gagaaagcag
480

tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg
540

gaggaggaga agaagacctg gggcacagtg ttcaagaccc tgaagagcct gtacaagaca
600

cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac
660

gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt
720

aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc
780

agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag
840

cccgacatct gtcacgagct gctggggccac gtgcccctgt ttagcgatag atccttcgcc
900

cagttttccc aggttggtat ccaggttaca aggcagctca caagaagaag ttgggtgctt
960

ggagacagag gtctgctttc cagcagacac taactttcag tgtcccctgt ctatgtttcc
1020

cttttttagga gatcggactg gcatctcttg gagcacctga cgagtacatc gagaagctgg
1080

ccaccatcta ttgggttcaca gtggagtttg gcctgtgcaa gcagggcgat agcatcaagg
1140

cctacggagc aggactgctg tctagcttcg gcgagctgca gtattgtctg tccgagaagc
1200

caaagctgct gcccctggag ctggagaaga ccgccatcca gaactacacc gtgacagagt
1260

tccagcccct gtactatgtg gccgagtctt ttaacgatgc caaggagaag gtgagaaatt
1320

tcgccgccac aatccctagg cccttcagcg tgcggtacga cccttatacc cagaggatcg
1380

aggtgctgga taatacacag cagctgaaga tcctggctga ctcaatcaat agcgaaatcg
1440

gaatcctgtg ctccgccctg cagaaaatca aatgagaatt caaggcctct cgagcctcta
1500

gaactatagt gagtcgtatt acgtagatcc agacatgata agatacattg atgagtttgg
1560

acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt gtgatgctat
1620

tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca attgcattca
1680

ttttatgttt caggttcagg gggagggtgtg ggagggttttt taa
1723

<210> 152

<211> 1607

<212> DNA

<213> Artificial Sequence

<220>

<223> 006-SD.3 vector editing element

<400> 152

atgtccaccg ctgtgctgga gaaccctggg ctgggggagga aactgtcaga cttcgggcag
60

gagacttcat acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc
120

ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg
180

aatctgaccc acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt
240

acccacctgg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac
300

gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg
360

ttccctcgga caatccagga gctggataga tttgccaacc agatcctgtc ttacggagca
420

gagctggacg cagatcaccc tggcttcaag gaccagtggt atcggggccc gagaaagcag
480

tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg
540

gaggaggaga agaagacctg gggcacagtg ttcaagaccc tgaagagcct gtacaagaca
600

cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac
660

gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt
720

aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc
780

agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag
840

cccgacatct gtcacgagct gctggggccac gtgcccctgt ttagcgatag atccttcgcc
900

cagttttccc aggagatcgg actggcatct ctgggagcac ctgacgagta catcgagaag
960

ctggccacca tctattgggtt cacagtggag tttggcctgt gcaagcaggg cgatagcatc
1020

aaggcctacg gagcaggact gctgtctagc ttcggcgagc tgcagtattg tctgtccgag
1080

aagccaaagc tgctgcccct ggagctggag aagaccgcca tccagaacta caccgtgaca
1140

gagttccagc ccctgtacta tgtggccgag tcttttaacg atgccaagga gaaggtgaga
1200

aatttcgccg ccacaatccc taggcccttc agtgtgctgt acgaccctta taccagagg
1260

atcgaggtgc tggataatac acagcagctg aagatcctgg ctgactcaat caatagcgaa
1320

atcggaatcc tgtgctccgc cctgcagaaa atcaaattgag aattcaaggc ctctcgagcc
1380

tctagaacta tagtgagtcg tattacgtag atccagacat gataagatac attgatgagt
1440

ttggacaaac cacaactaga atgcagtgaa aaaaatgctt tatttgatgaa atttgatgag
1500

ctattgcttt atttgtaacc attataagct gcaataaaca agttaacaac aacaattgca
1560

ttcattttat gtttcagggtt cagggggagg tgtgggaggt ttttttaa
1607

<210> 153

<211> 1837

<212> DNA

<213> Artificial Sequence

<220>

<223> 032-SD.3 vector editing element

<400> 153

ctgacctctt ctcttcctcc cacagggcgg taccagatct ggcagcggag agggcagagg
60

aagtctttcta acatgcggtg acgtggagga gaatcccggc cctaggggta cctccaccgc
120

tgtgctggag aaccctgggc tggggaggaa actgtcagac ttcgggcagg agacttcata
180

cattgaggat aactgtaacc agaatggcgc catctctctg atcttcagcc tgaaggagga
240

agtgggcgcc ctggcaaagg tgctgcgctt gtttgaggag aacgacgtga atctgaccca
300

catcgagtcc cggccttcta gactgaagaa ggacgagtac gagttcttta cccacctgga
360

taagcgggtc ctgccagccc tgacaaacat catcaagatc ctgaggcacg acatcggagc
420

aaccgtgcac gagctgtctc gggacaagaa gaaggatacc gtgccctggg tccctcggac
480

aatccaggag ctggatagat ttgccaacca gatcctgtct tacggagcag agctggacgc
540

agatcacctt ggcttcaagg acccagtgtg tcggggcccgg agaaagcagt ttgccgatat
600

cgcctacaat tataggcacg gacagccaat ccctcgcgtg gagtatatgg aggaggagaa
660

gaagacctgg ggcacagtgt tcaagaccct gaagagcctg tacaagacac acgcctgcta
720

cgagtataac cacatcttcc ccctgctgga gaagtattgt ggctttcacg aggacaatat
780

ccctcagctg gaggacgtga gccagttcct gcagacctgc acaggcttta ggctgaggcc
840

agtggcagga ctgctgagct cccgggactt cctgggagga ctggccttca gagtgtttca
900

ctgcacccag tacatcaggc acggctccaa gccaatgtat acaccagagc ccgacatctg
960

tcacgagctg ctggggccacg tgcccctgtt tagcgataga tccttcgccc agttttccca
1020

ggagatcgga ctggcatctc tgggagcacc tgacgagtac atcgagaagc tggccaccat
1080

ctattggttc acagtggagt ttggcctgtg caagcagggc gatagcatca aggcctacgg
1140

agcaggactg ctgtctagct tcggcgagct gcagtattgt ctgtccgaga agccaaagct
1200

gctgcccctg gagctggaga agaccgccat ccagaactac accgtgacag agttccagcc
1260

cctgtactat gtggccgagt cttttaacga tgccaaggag aaggtgagaa atttcgccgc
1320

cacaatccct aggcccttca gtgtgcgtta cgacccttat acccagagga tcgaggtgct
1380

ggataataca cagcagctga agatcctggc tgactcaatc aatagcgaaa tcggaatcct
1440

gtgctccgcc ctgcagaaaa tcaaaggtaa gcctatccct aaccctctcc tcggtctcga
1500

ttctacgtga tcttgtggaa aggacgaaac accggggaat tcaaggcctc tcgagcctct
1560

agaatccccg agacgtttcg tctcgggatc actatagtga gtcgtattac gtacacagtg
1620

caggggaaag aatagtagag atccagacat gataagatac attgatgagt ttggacaaac
1680

cacaactaga atgcagtgaa aaaaatgctt tatttgtgaa atttgtgatg ctattgcttt
1740

atttgtaacc attataagct gcaataaaca agttaacaac aacaattgca ttcattttat
1800

gtttcaggtt
1837

cagggggagg

tgtgggaggt

tttttaa