



(51) International Patent Classification:

A61K 38/43 (2006.01) C12N 9/02 (2006.01)

A61K 35/761 (2015.01) C12N 5/10 (2006.01)

A61K 48/00 (2006.01) A61P 3/00 (2006.01)

C12N 15/864 (2006.01)

(21) International Application Number:

PCT/US2019/016351

(22) International Filing Date:

01 February 2019 (01.02.2019)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/625,150 01 February 2018 (01.02.2018) US

(71) Applicant: **HOMOLOGY MEDICINES, INC.** [US/US];
45 Wiggins Avenue, Bedford, Massachusetts 01730 (US).(72) Inventors: **SEYMOUR, Albert Barnes**; 1 Brickyard Lane, Westborough, Massachusetts 01581 (US). **AHMED, Seemin Seher**; 81 Park Avenue, Apartment 10, Worcester, Massachusetts 01605 (US). **WRIGHT, Jason Boke**; 57 Whits End Road, Concord, MA 01742 (US). **DOLLIVE, Serena Nicole**; 19 Everett St., Unit 1, Waltham, Massachusetts 02453 (US). **RUBIN, Hillard**; 10 Gale Street, Apt. 2, Northborough, Massachusetts 01532 (US).(74) Agent: **WILKINS, Andrew T.** et al.; Lathrop Gage LLP, 28 State Street, Suite 700, Boston, Massachusetts 02109 (US).

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

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

Declarations under Rule 4.17:

- as to the identity of the inventor (Rule 4.17(i))
- 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))

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

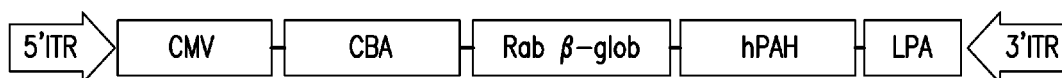


FIG.8A

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



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

RELATED APPLICATIONS

5 [0001] This application claims priority to U.S. Provisional Patent Application Serial No. 62/625,150, filed February 1, 2018, the entire disclosure of which is 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-025PC_SeqList_ST25.txt; Size: 34,214 bytes; and Date of Creation: January 27, 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 medical
20 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 intellectual disability of the disorder. However, even on a low-phenylalanine 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 once children reach 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, raising 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). Non-integrating vectors, on the other hand, often suffer insufficient expression level or inadequate duration of expression *in vivo*.

[0007] Accordingly, there is a need in the art for improved gene therapy compositions and methods that can efficiently and safely restore PAH gene function in PKU patients.

SUMMARY

[0008] 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.

[0009] Accordingly, in one aspect, the instant disclosure provides a method for expressing a PAH polypeptide in a cell, the method comprising transducing the cell with a replication-defective adeno-associated virus (AAV) comprising:

- (a) an AAV capsid comprising an AAV Clade F capsid protein; and
- (b) a transfer genome comprising a transcriptional regulatory element operably linked to a PAH coding sequence.

[0010] 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.

[0011] In another aspect, the instant disclosure provides a method for treating a subject having a disease or disorder associated with a PAH gene mutation, the method comprising administering to the subject an effective amount of a replication-defective AAV comprising:

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

(b) a transfer genome comprising a transcriptional regulatory element operably linked to a PAH coding sequence.

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

5 [0013] In another aspect, the instant disclosure provides a replication-defective adeno-associated virus (AAV) comprising:

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

(b) a transfer genome comprising a transcriptional regulatory element operably linked to a PAH coding sequence.

10 [0014] The following embodiments apply to each of the foregoing aspects.

[0015] 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 nucleotide 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
15 comprises the nucleotide sequence set forth in SEQ ID NO: 25.

[0016] In certain embodiments, the transcriptional regulatory element is capable of mediating transcription in 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 transcriptional regulatory element is capable of mediating
20 transcription in a hepatocyte or renal cell. In certain embodiments, the transcriptional regulatory element comprises one of more of the elements selected from the group consisting of a CAG promoter, a human EF-1 α promoter, a human hepatic control region 1 (HCR1), a human α 1-antitrypsin (hAAT) promoter, a hepatic specific regulatory module of the hAAT promoter, an SV40 intron, and a minute virus of mouse (MVM) intron. In certain
25 embodiments, the transcriptional regulatory element comprises a nucleotide sequence at least 90% identical to a sequence selected from the group consisting of SEQ ID NOs: 28-30 and 32-41. In certain embodiments, the transcriptional regulatory element comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 28-30 and 32-41. In certain
30 embodiments, the transcriptional regulatory element comprises from 5' to 3' the nucleotide sequences set forth in SEQ ID NOs: 29, 30, and 31. In certain embodiments, the transcriptional regulatory element comprises the nucleotide sequences set forth in SEQ ID NO: 32.

[0017] In certain embodiments, the transfer genome further comprises an intron operably linked to the PAH coding sequence. In certain embodiments, the intron comprises a nucleotide sequence at least 90% identical to the sequence set forth in SEQ ID NO: 31 or 35.

In certain embodiments, the intron comprises the nucleotide sequence set forth in SEQ ID NO: 31 or 35. In certain embodiments, the transfer genome comprises from 5' to 3': a non-coding exon, the intron, and the PAH coding sequence.

5 [0018] In certain embodiments, the transfer genome further comprises a polyadenylation sequence 3' to the PAH coding sequence. In certain embodiments, the polyadenylation sequence is an exogenous polyadenylation sequence. In certain embodiments, the exogenous polyadenylation sequence is an SV40 polyadenylation sequence. In certain embodiments, the SV40 polyadenylation sequence comprises a sequence selected from the group consisting of SEQ ID NOs: 42, 43, and 45.

10 [0019] In certain embodiments, the transfer genome comprises a sequence selected from the group consisting of SEQ ID NOs: 46-50, 61, 64, 67, 74, 76, 78, 80, 82, 84, 86, and 89.

[0020] In certain embodiments, the transfer genome further comprises a 5' inverted terminal repeat (5' ITR) nucleotide sequence 5' of the genome, and a 3' inverted terminal repeat (3' ITR) nucleotide sequence 3' of the genome. 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
15 20 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.

[0021] In certain embodiments, the transfer genome comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 51-55, 62, 65, 68, 75, 77, 79, 81, 83, 85, 87, and 90. In certain embodiments, the transfer genome consists of a nucleotide sequence selected from the group consisting of SEQ ID NOs: 51-55, 62, 65, 68, 75, 77, 79, 81, 83, 85, 87, and 90. In certain embodiments, the transfer genome consists of the nucleotide sequence set forth in SEQ ID NO: 52.

[0022] In certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 95% sequence identity with the amino acid sequence of amino acids 203-736 of SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q;

the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A;
 the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N;
 the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S;
 the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I;
 5 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
 10 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.

[0023] In certain embodiments, (a) the amino acid in the capsid protein corresponding
 15 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
 20 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;

25 (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,
 30 and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is
 C.

[0024] 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.

[0025] 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.

[0026] In certain embodiments, (a) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G;

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

(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.

[0027] 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.

[0028] In certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 95% sequence identity with the amino acid sequence of amino acids 1-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 2 is T; the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 68 of SEQ ID NO: 2 is V; the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 2 is L; the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 2 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is

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

[0029] 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.

[0030] 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.

[0031] In another aspect, the instant disclosure provides a pharmaceutical composition comprising an AAV disclosed herein.

[0032] In another aspect, the instant disclosure provides a packaging system for recombinant preparation of an AAV, wherein the packaging system comprises:

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

10 [0033] 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 transfer 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
15 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.

[0034] 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
20 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.

[0035] In certain embodiments, the helper virus is selected from the group consisting
25 of adenovirus, herpes virus, vaccinia virus, and cytomegalovirus (CMV). In certain embodiments, the helper virus is adenovirus. In certain embodiments, the adenovirus genome comprises one or more adenovirus RNA genes selected from the group consisting of E1, E2, E4 and VA. In certain embodiments, the helper virus is herpes simplex virus (HSV). In certain embodiments, the HSV genome comprises one or more of HSV genes selected from the group
30 consisting of UL5/8/52, ICPO, ICP4, ICP22 and UL30/UL42.

[0036] In certain embodiments, the first vector and the third vector are contained within a first transfecting plasmid. In certain embodiments, the nucleotides of the second vector and the third vector are contained within a second transfecting plasmid. In certain embodiments, the nucleotides of the first vector and the third vector are cloned into a recombinant helper

virus. In certain embodiments, the nucleotides of the second vector and the third vector are cloned into a recombinant helper virus.

[0037] In another aspect, the instant disclosure provides a method for recombinant preparation of an AAV, the method comprising introducing a packaging system as described herein into a cell under conditions operative for enclosing the transfer genome or the transfer genome in the capsid to form the AAV.

BRIEF DESCRIPTION OF THE DRAWINGS

[0038] Figure 1A, 1B, 1C, 1D, and 1E are vector maps of the pHMI-hPAH-TC-004, pHMI-hPAH-TC-025, pHMI-hPAH-TC-010, pHMI-hPAH-TC-011, and pHMI-hPAH-TC-012 vectors, respectively.

[0039] 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.

[0040] Figures 3A and 3B are graphs showing the Phe level in the serum of two *pah*^{-/-} mice ("Mouse H1" and "Mouse H5") each administered with 5×10^{13} vector genomes of the rAAV-CBA-mPAH vector packaged in an AAVHSC capsid per kg of body weight intravenously via the tail vein. Serum samples were collected in a time course. The Phe levels were measured with a BioAssay Systems ELISA kit EPHE-100 (Figure 3A) or mass spectrometry (Figure 3B).

[0041] Figure 4 is a graph and a table showing the numbers of vector genomes per 10^6 cells detected in major organs. The rAAV-CBA-mPAH vector packaged in an AAVHSC capsid was injected to *pah*^{-/-} mice intravenously via the tail vein at a dose of 5×10^{13} vector genomes per kg of body weight. Organs of the mice were collected 4 weeks after the administration. The numbers of vector genomes per 10^6 cells were measured by the following method: (1) the weight/volume concentration of the vector genome in a sample was measured by Taqman PCR using a standard curve generated with serial dilutions of the vector plasmid; (2) the mass of a single vector genome was calculated based on the sequence of the vector; (3) the number/volume concentration of the vector genome in the sample was calculated; (4) the weight/volume concentration of genomic DNA in the same sample was

measured by Taqman PCR of the apolipoprotein B gene using a standard curve generated with serial dilutions of calculated amounts of genomic DNA isolated from mouse tissues; (5) the number/volume concentration of cell genome in the sample was calculated based on copies of ApoB; and (6) the number of vector genomes per 10^6 cells was calculated by dividing the number/volume concentration of the vector genome by the number/volume concentration of the cell genome and multiplying the result by 10^6 .

[0042] **Figures 5A and 5B** are graphs showing the aspartate aminotransferase (AST) and alanine aminotransferase (ALT) levels in the serum of *pah*^{-/-} mice administered with the rAAV-CBA-mPAH vector. The rAAV-CBA-mPAH vector packaged in an AAVHSC capsid was injected to *pah*^{-/-} mice intravenously via the tail vein at a dose of 5×10^{13} vector genomes per kg of body weight. Serum samples were collected 4 weeks after the administration. The levels of AST (Figure 5A) and ALT (Figure 5B) were measured by ELISA using the Sigma MAK055 and Sigma MAK052 kits, respectively.

[0043] **Figures 6A-6H** are graphs showing the levels of phenylalanine (Figures 6A, 6C, 6E, and 6G) or tyrosine (Figures 6B, 6D, 6F, and 6H) in the serum of male (Figures 6A, 6B, 6E, and 6F) or female (Figures 6C, 6D, 6G, and 6H) mice administered with the indicated doses of the pHMI-hPAH-TC-004, pHMI-hPAH-TC-025, pHMI-hPAH-TC-010, pHMI-hPAH-TC-011, or pHMI-hPAH-TC-012 vector.

[0044] **Figures 6I-6J** are graphs showing the long-term efficacy on levels of phenylalanine in the serum of male (Figure 6I) or female (Figure 6J) mice administered the indicated doses of the pHMI-hPAH-TC-025 vector.

[0045] **Figures 7A-7D** are graphs showing the levels of phenylalanine (Figures 7A and 7C) or tyrosine (Figures 7B and 7D) in the serum of male (Figures 7A and 7B) or female (Figures 7C and 7D) mice administered with the indicated doses of the pHMI-hPAH-TC-025 vector.

[0046] **Figure 8A, 8B, 8C** are vector maps of pHMI-hPAH-TC-009, pHMI-hPAH-TC-013 and pHMI-hPAH-TC-017 vectors, respectively.

[0047] **Figure 9A-9B** depict the quantification of Western blots of human PAH expression, from the indicated AAV vectors, in Huh7 cells (Figure 9A) and HEK293 cells (Figure 9B).

[0048] **Figure 10A-10C** are graphs showing serum phenylalanine levels in mice that have been administered the indicated AAV vectors. **Figure 10A** is a graph showing serum Phe levels over time of male *Pah*^{-/-} PAH^{enu2} mice. **Figure 10B** is a graph showing serum Phe levels over time of female *Pah*^{-/-} PAH^{enu2} mice. **Figure 10C** is a graph showing the average

baseline serum Phe levels of the male and female mice in the study (55 mice per group; **** indicates $p < 0.05$).

[0049] Figure 11A, 11B, 11C, 11D, 11E, and 11F are vector maps of pHMI-hPAH-TC-018, pHMI-hPAH-TC-019, pHMI-hPAH-TC-020, pHMI-hPAH-TC-021, pHMI-hPAH-TC-022, and pHMI-hPAH-TC-023 vectors, respectively.

[0050] Figure 12 depicts the quantification of Western blots of human PAH expression from HEK293 cells transfected with the indicated AAV vectors under the control of a CBA promoter.

[0051] Figure 13 is a graph showing serum phenylalanine levels over time of male *Pah*^{-/-} PAH^{enu2} mice administered the indicated AAV vectors.

DETAILED DESCRIPTION

[0052] 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

[0053] As used herein, the term “replication-defective adeno-associated virus” refers to an AAV comprising a genome lacking Rep and Cap genes.

[0054] As used herein, the term “PAH gene” refers to the phenylalanine hydroxylase 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.

[0055] As used herein, the term “transfer genome” refers to a recombinant AAV genome comprising a coding sequence operably linked to an exogenous transcriptional regulatory element that mediates expression of the coding sequence when the transfer genome is introduced into a cell. In certain embodiments, the transfer genome does not integrate in the chromosomal DNA of the cell. The skilled artisan will appreciate that the portion of a transfer genome comprising the transcriptional regulatory element operably linked to a PAH coding sequence can be in the sense or antisense orientation relative to direction of transcription of the PAH coding sequence.

[0056] As used herein, the term “Clade F capsid protein” refers to an AAV VP1, VP2, or VP3 capsid protein that has 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.

[0057] As used herein, the identity between two nucleotide sequences or between two amino acid sequences is determined by the number of identical nucleotides or amino acids in alignment divided by the full length of the longer nucleotide or amino acid sequence.

[0058] 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 mutation of a PAH gene. In certain embodiments, the disease or disorder associated with a PAH gene mutation is phenylketonuria (PKU).

[0059] 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, alternative translation initiation, and variation within the population. A coding sequence may either be wild-type or codon-altered. An exemplary wild-type PAH coding sequence is set forth in SEQ ID NO: 24.

[0060] As used herein, the term “silently altered” refers to alteration of a coding sequence or a stuffer-inserted coding sequence of a gene (*e.g.*, by nucleotide substitution) without changing the amino acid sequence of the polypeptide encoded by the coding sequence or stuffer-inserted coding sequence. Such silent alteration is advantageous in that it may increase the translation efficiency of a coding sequence.

[0061] 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. An exemplary nucleotide 1 of the human PAH gene is nucleotide 5,473 of the NCBI Reference Sequence: NG_008690.1, and an exemplary nucleotide 3 of the human PAH gene is nucleotide 5,475 of the NCBI Reference Sequence: NG_008690.1. The nucleotide adjacently 5' to the start codon is nucleotide -1.

[0062] 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. An exemplary exon 1 of the human PAH

gene is nucleotides 5001-5532 of the NCBI Reference Sequence: NG_008690.1. An exemplary intron 1 of the human PAH gene is nucleotides 5533-9704 of the NCBI Reference Sequence: NG_008690.1.

[0063] As used herein, the term “transcriptional regulatory element” or “TRE” refers to a cis-acting nucleotide sequence, for example, a DNA sequence, that regulates (e.g., controls, increases, or reduces) transcription of an operably linked nucleotide sequence by an RNA polymerase to form an RNA molecule. A TRE relies on one or more trans-acting molecules, such as transcription factors, to regulate transcription. Thus, one TRE may regulate transcription in different ways when it is in contact with different trans-acting molecules, for example, when it is in different types of cells. A TRE may comprise one or more promoter elements and/or enhancer elements. A skilled artisan would appreciate that the promoter and enhancer elements in a gene may be close in location, and the term “promoter” may refer to a sequence comprising a promoter element and an enhancer element. Thus, the term “promoter” does not exclude an enhancer element in the sequence. The promoter and enhancer elements do not need to be derived from the same gene or species, and the sequence of each promoter or enhancer element may be either identical or substantially identical to the corresponding endogenous sequence in the genome.

[0064] As used herein, the term “operably linked” is used to describe the connection between a TRE and a coding sequence to be transcribed. Typically, gene expression is placed under the control of a TRE comprising one or more promoter and/or enhancer elements. The coding sequence is “operably linked” to the TRE if the transcription of the coding sequence is controlled or influenced by the TRE. The promoter and enhancer elements of the TRE may be in any orientation and/or distance from the coding sequence, as long as the desired transcriptional activity is obtained. In certain embodiments, the TRE is upstream from the coding sequence.

[0065] 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).

[0066] As used herein, “exogenous polyadenylation sequence” refers to a polyadenylation sequence not identical or substantially identical to the endogenous polyadenylation sequence of a PAH gene (e.g., human PAH gene). In certain embodiments, an exogenous polyadenylation sequence is a polyadenylation sequence of a non-PAH gene in

the same species (e.g., human). In certain embodiments, an exogenous polyadenylation sequence is a polyadenylation sequence of a different species (e.g., a virus).

[0067] As used herein, the term “effective amount” in the context of the administration of an AAV to a subject refers to the amount of the AAV that achieves a desired prophylactic or therapeutic effect.

II. Adeno-Associated Virus Compositions

[0068] In one aspect, provided herein are novel replication-defective AAV compositions useful for expressing PAH polypeptide in cells with reduced or otherwise defective PAH gene function. In certain embodiments, the AAV disclosed herein comprise: an AAV capsid comprising an AAV Clade F capsid protein; and a transfer genome comprising a transcriptional regulatory element operably linked to a PAH coding sequence, allowing for extrachromosomal expression of PAH.

[0069] 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.

[0070] 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.

[0071] For example, in certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the amino acid sequence of amino acids 1-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17. In certain embodiments, the AAV Clade F capsid protein comprises an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%,

90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the amino acid sequence of amino acids 1-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17, wherein: the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 2 is T; the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 68 of SEQ ID NO: 2 is V; the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 2 is L; the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 2 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 2 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 2 is T, and the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 2 is I, and the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is Y. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 2 is L, and the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S. In certain embodiments, the amino acid in the capsid protein corresponding

to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C.

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

[0072] In certain embodiments, the AAV capsid comprises two or more of: (a) a Clade F capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17. In certain embodiments, the AAV capsid comprises: (a) a Clade F capsid protein having an amino acid sequence consisting of amino acids 203-736 of SEQ ID NO: 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17; (b) a Clade F capsid protein having an amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17; and (c) a Clade F capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

[0073] In certain embodiments, the AAV capsid comprises one or more of: (a) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 203-736 of SEQ ID NO: 8; (b) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%

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

5 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

10 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

15 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.

[0074] 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%,

20 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

25 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

30 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.

[0075] In certain embodiments, the AAV capsid comprises one or more of: (a) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 203-736 of SEQ ID NO: 13; (b) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 138-736 of SEQ ID NO: 13; and (c) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 1-736 of SEQ ID NO: 13. In certain embodiments, the AAV capsid comprises one or more of: (a) a Clade F capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 13; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 13; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 13. In certain embodiments, the AAV capsid comprises two or more of: (a) a Clade F capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 13; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 13; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 13. In certain embodiments, the AAV capsid comprises: (a) a Clade F capsid protein having an amino acid sequence consisting of amino acids 203-736 of SEQ ID NO: 13; (b) a Clade F capsid protein having an amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 13; and (c) a Clade F capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 13.

[0076] In certain embodiments, the AAV capsid comprises one or more of: (a) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 203-736 of SEQ ID NO: 16; (b) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%,

85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 138-736 of SEQ ID NO: 16; and (c) a Clade F capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity with the sequence of amino acids 1-736 of SEQ ID NO: 16. In certain embodiments, the AAV capsid comprises one or more of: (a) a Clade F capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 16; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 16; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 16. In certain embodiments, the AAV capsid comprises two or more of: (a) a Clade F capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 16; (b) a Clade F capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 16; and (c) a Clade F capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 16. In certain embodiments, the AAV capsid comprises: (a) a Clade F capsid protein having an amino acid sequence consisting of amino acids 203-736 of SEQ ID NO: 16; (b) a Clade F capsid protein having an amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 16; and (c) a Clade F capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 16.

[0077] Transfer genomes useful in the AAV compositions disclosed herein generally comprise a transcriptional regulatory element (TRE) operably linked to a PAH coding sequence. In certain embodiments, the transfer genome comprises a 5' inverted terminal repeat (5' ITR) nucleotide sequence 5' of the TRE and PAH coding sequence, and a 3' inverted terminal repeat (3' ITR) nucleotide sequence 3' of the TRE and PAH coding sequence.

[0078] In certain embodiments, the PAH coding sequence comprises all or substantially all of a coding sequence of a PAH gene. In certain embodiments, the transfer genome comprises a nucleotide sequence encoding SEQ ID NO: 23 and can optionally further comprise an exogenous polyadenylation sequence 3' to the PAH coding sequence. In certain embodiments, the nucleotide sequence encoding SEQ ID NO: 23 is wild-type (e.g., having the sequence set forth in SEQ ID NO: 24). In certain embodiments, the nucleotide sequence encoding SEQ ID NO: 23 is codon-altered (e.g., having the sequence set forth in SEQ ID NO: 25). In certain embodiments, the nucleotide sequence encoding SEQ ID NO: 23 is codon-altered (e.g., having the sequence set forth in SEQ ID NO: 69, 70, 71, 72, or 73).

[0079] In certain embodiments, the PAH coding sequence encodes a polypeptide comprising all or substantially all of the amino acids sequence of a PAH protein. In certain

embodiments, the PAH coding sequence encodes the amino acid sequence of a wild-type PAH protein (e.g., human PAH protein). In certain embodiments, the PAH coding sequence encodes the amino acid sequence of a mutant PAH protein (e.g., human PAH protein), wherein the mutant PAH polypeptide is a functional equivalent of the wild-type PAH polypeptide, i.e., can function as a wild-type PAH polypeptide. In certain embodiments, the functionally equivalent PAH polypeptide further comprises at least one characteristic not found in the wild-type PAH polypeptide, e.g., the ability to stabilize PAH protein (e.g., dimer or tetramer), or the ability to resist protein degradation.

[0080] The transfer genome can be used to express PAH in any mammalian cells (e.g., human cells). Thus, the TRE can be active in any mammalian cells (e.g., human cells). In certain embodiments, the TRE is active in a broad range of human cells. Such TREs may comprise constitutive promoter and/or enhancer elements including cytomegalovirus (CMV) promoter/enhancer (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 58), SV40 promoter, chicken beta actin (CBA) promoter (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 59), human elongation factor 1 alpha (EF1 α) promoter (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 40), minute virus of mouse (MVM) intron which comprises transcription factor binding sites (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 35), human phosphoglycerate kinase (PGK1) promoter, human ubiquitin C (Ubc) promoter, human beta actin promoter, human neuron-specific enolase (ENO2) promoter, human beta-glucuronidase (GUSB) promoter, a rabbit beta-globin element (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 60), and/or human Methyl-CpG Binding Protein 2 (MeCP2) promoter. Any of these TREs can be combined in any order to drive efficient transcription. For example, a transfer genome may comprise a CMV enhancer, a CBA promoter, and the splice acceptor from exon 3 of the rabbit beta-globin gene, collectively called a CAG promoter (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 28). For example, a transfer genome may comprise a hybrid of CMV enhancer and CBA promoter followed by a splice donor and splice acceptor, collectively called a CASI promoter region (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 63).

[0081] Alternatively, the TRE may be a tissue-specific TRE, i.e., it is active in specific tissue(s) and/or organ(s). A tissue-specific TRE comprises one or more tissue-specific promoter and/or enhancer elements, and optionally one or more constitutive promoter and/or enhancer elements. A skilled artisan would appreciate that tissue-specific promoter and/or enhancer elements can be isolated from genes specifically expressed in the tissue by methods well known in the art. In certain embodiments, the TRE is liver-specific (e.g., hepatocyte-specific). Exemplary liver-specific TREs may comprise one or more elements selected from the group consisting of human albumin promoter, human transthyretin (TTR) promoter (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 34), human APOE/C-I hepatic control region (HCR) 1 or 2 (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 29 or 37), human APOH promoter, and human SERPINA1 (hAAT) promoter (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 30 or 38) or a hepatic specific regulatory module thereof (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 33). In certain embodiments, an hAAT promoter region comprises a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 66. More liver-specific promoter elements are disclosed in WO 2009/130208 and Kramer et al. (Molecular Therapy (2003) 7, 375–385), which are incorporated by reference herein in their entirety.

[0082] In certain embodiments, the TRE is kidney-specific (e.g., renal epithelial cell-specific). Exemplary kidney-specific TREs may comprise one or more elements selected from the group consisting of human nephrin promoter, human parathyroid hormone receptor promoter, human uromodulin promoter, and human SLC12A1 promoter. In certain embodiments, the TRE is brain-specific (e.g., neuron-specific, glial cell-specific, astrocyte-specific, oligodendrocyte-specific, microglia-specific and/or central nervous system-specific). Exemplary brain-specific TREs may comprise one or more elements selected from the group consisting of human glial fibrillary acidic protein (GFAP) promoter and human synapsin I (SYN1) promoter. More brain-specific promoter elements are disclosed in WO 2016/100575A1, which is incorporated by reference herein in its entirety.

[0083] In certain embodiments, the transfer genome comprises two or more TREs, optionally comprising at least one of the TREs disclosed above. A skilled person in the art would appreciate that any of these TREs can be combined in any order, and combinations of a

constitutive TRE and a tissue-specific TRE can drive efficient and tissue-specific transcription. For example, in certain embodiments, the transfer genome comprises a human HCR1 (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 or 37) and a human EF-1 α promoter (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 40), optionally wherein the human HCR1 is 5' to the human EF-1 α promoter. In certain embodiments, the transfer genome comprises a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the sequence set forth in SEQ ID NO: 41. In certain embodiments, the transfer genome comprises a human HCR1 (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 29 or 37) and a human EF-1 α promoter (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 40), optionally wherein the human HCR1 is 5' to the human EF-1 α promoter. In certain embodiments, the transfer genome comprises the sequence set forth in SEQ ID NO: 41.

[0084] Similarly, combinations of two or more tissue-specific TREs can drive efficient and tissue-specific transcription. For example, in certain embodiments, the transfer genome comprises a human HCR1 (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) and a hAAT promoter (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), optionally wherein the human HCR1 is 5' to the hAAT promoter. In certain embodiments, the transfer genome comprises a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the sequence set forth in SEQ ID NO: 32. In certain embodiments, the transfer genome comprises a human HCR1 (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 29) and a hAAT promoter (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 30), optionally wherein the human HCR1 is 5' to the hAAT promoter. In certain embodiments, the transfer genome comprises the nucleotide sequence set forth in SEQ ID NO: 32.

[0085] In certain embodiments, the transfer genome comprises a hepatic specific regulatory module of hAAT promoter (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 33) and a human TTR promoter (e.g., comprising a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 34), optionally wherein the hepatic specific regulatory module is 5' to the human TTR promoter. In certain embodiments, the transfer genome comprises a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the sequence set forth in SEQ ID NO: 36. In certain

embodiments, the transfer genome comprises a hepatic specific regulatory module of hAAT promoter (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 33) and a human TTR promoter (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 34), optionally wherein the hepatic specific regulatory module is 5' to the human TTR promoter. In certain embodiments, the transfer genome comprises the nucleotide sequence set forth in SEQ ID NO: 36.

[0086] In certain embodiment, the transfer genome comprises a human HCR1 (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 or 37) and a hAAT promoter (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 or 38), optionally wherein the human HCR1 is 5' to the hAAT promoter. In certain embodiments, the transfer genome comprises a nucleotide sequence at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the sequence set forth in SEQ ID NO: 39. In certain embodiment, the transfer genome comprises a human HCR1 (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 29 or 37) and a hAAT promoter (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 30 or 38), optionally wherein the human HCR1 is 5' to the hAAT promoter. In certain embodiments, the transfer genome comprises the nucleotide sequence set forth in SEQ ID NO: 39.

[0087] In certain embodiments, the transfer vector further comprises an intron 5' to or inserted in the PAH coding sequence. Such introns can increase transgene expression, for example, by reducing transcriptional silencing and enhancing mRNA export from the nucleus to the cytoplasm. In certain embodiments, the transfer genome comprises from 5' to 3': a non-coding exon, an intron, and the PAH coding sequence. In certain embodiments, an intron sequence is inserted in the PAH coding sequence, optionally wherein the intron is inserted at an internucleotide bond that links two native exons. In certain embodiments, the intron is inserted at an internucleotide bond that links native exon 1 and exon 2.

[0088] The intron can comprise a native intron sequence of the PAH gene, an intron sequence from a different species or a different gene from the same species, and/or a synthetic intron sequence. A skilled worker will appreciate that synthetic intron sequences can be designed to mediate RNA splicing by introducing any consensus splicing motifs known in the art (e.g., in Sibley et al., (2016) Nature Reviews Genetics, 17, 407-21, which is incorporated by reference herein in its entirety). Exemplary intron sequences are provided in Lu et al. (2013) Molecular Therapy 21(5): 954-63, and Lu et al. (2017) Hum. Gene Ther. 28(1): 125-34, which are incorporated by reference herein in their entirety. In certain embodiments, the transfer

genome comprises an SV40 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: 31) or 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: 35). In certain embodiments, the transfer genome comprises an SV40 intron (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 31) or a minute virus of mouse (MVM) intron (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 35).

[0089] In certain embodiments, the transfer genome disclosed herein further comprises a transcription terminator (e.g., a polyadenylation sequence). In certain embodiments, the transcription terminator is 3' to the PAH coding sequence. The transcription terminator may be any sequence that effectively terminates transcription, and a skilled artisan would appreciate that such sequences can be isolated from any genes that are expressed in the cell in which transcription of the PAH coding sequence is desired. In certain embodiments, the transcription terminator comprises a polyadenylation sequence. In certain embodiments, the polyadenylation sequence is identical or substantially identical to the endogenous polyadenylation sequence of the human PAH gene. In certain embodiments, the polyadenylation sequence is an exogenous polyadenylation sequence. In certain embodiments, the polyadenylation sequence is an SV40 polyadenylation sequence (e.g., comprising the nucleotide sequence set forth in SEQ ID NO: 42, 43, or 45, or a nucleotide sequence complementary thereto). In certain embodiments, the polyadenylation sequence comprises the sequence set forth in SEQ ID NO: 43.

[0090] In certain embodiments, the transfer genome comprises from 5' to 3': a TRE, optionally a non-coding exon and an intron, a PAH coding sequence, and a polyadenylation sequence. In certain embodiments, the TRE has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any one of SEQ ID NOs: 28-30 and 32-41; the intron has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 31 or 35; the PAH coding sequence has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 25; and/or the polyadenylation sequence has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any one of SEQ ID NOs: 42, 43, and 45. In certain embodiments, the TRE comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 28-30 and 32-41; the intron comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 31 and 35; the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25; and/or the polyadenylation sequence comprises a nucleotide sequence selected from the group

consisting of SEQ ID NOs: 42, 43, and 45. In certain embodiments, the TRE comprises the sequence set forth in SEQ ID NO: 28; the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25; and/or the polyadenylation sequence comprises the sequence set forth in SEQ ID NO: 42. In certain embodiments, the TRE comprises from 5' to 3' the sequence set forth in SEQ ID NO: 29, and the sequence set forth in SEQ ID NO: 30 (e.g., the TRE comprises the sequence set forth in SEQ ID NO: 32); the intron comprises the sequence set forth in SEQ ID NO: 31; the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25; and/or the polyadenylation sequence comprises the sequence set forth in SEQ ID NO: 43. In certain embodiments, the TRE comprises from 5' to 3' the sequence set forth in SEQ ID NO: 33, and the sequence set forth in SEQ ID NO: 34 (e.g., the TRE comprises the sequence set forth in SEQ ID NO: 36); the intron comprises the sequence set forth in SEQ ID NO: 35; the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25; and/or the polyadenylation sequence comprises the sequence set forth in SEQ ID NO: 45. In certain embodiments, the TRE comprises from 5' to 3' the sequence set forth in SEQ ID NO: 37, and the sequence set forth in SEQ ID NO: 38 (e.g., the TRE comprises the sequence set forth in SEQ ID NO: 39); the intron comprises the sequence set forth in SEQ ID NO: 35; the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25; and/or the polyadenylation sequence comprises the sequence set forth in SEQ ID NO: 45. In certain embodiments, the TRE comprises from 5' to 3' the sequence set forth in SEQ ID NO: 37, and the sequence set forth in SEQ ID NO: 40 (e.g., the TRE comprises the sequence set forth in SEQ ID NO: 41); and/or the polyadenylation sequence comprises the sequence set forth in SEQ ID NO: 45.

[0091] In certain embodiments, the transfer genome comprises a sequence at least 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 46, 47, 48, 49, 50, 61, 64, 67, 74, 76, 78, 80, 82, 84, 86, or 89. In certain embodiments, the transfer genome comprises the nucleotide sequence set forth in SEQ ID NO: 46, 47, 48, 49, 50, 61, 64, 67, 74, 76, 78, 80, 82, 84, 86, or 89. In certain embodiments, the transfer genome consists of the nucleotide sequence set forth in SEQ ID NO: 46, 47, 48, 49, 50, 61, 64, 67, 74, 76, 78, 80, 82, 84, 86, or 89. In certain embodiments, the transfer genome comprises the nucleotide sequence set forth in SEQ ID NO: 47. In certain embodiments, the transfer genome consists of the nucleotide sequence set forth in SEQ ID NO: 47.

[0092] In certain embodiments, the transfer genomes disclosed herein further comprise a 5' inverted terminal repeat (5' ITR) nucleotide sequence 5' of the TRE, and a 3' inverted terminal repeat (3' ITR) nucleotide sequence 3' of the PAH coding sequence. ITR sequences

from any AAV serotype or variant thereof can be used in the transfer 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 transfer genomes disclosed herein are set forth in SEQ ID NOs: 18-21, 26, and 27 herein.

5 [0093] In certain embodiments, the 5' ITR or 3' ITR is from AAV2. In certain embodiments, both the 5' ITR and the 3' ITR are from AAV2. In certain embodiments, the 5' ITR nucleotide sequence has at least 90% (e.g., at least 95%, 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 90% (e.g., at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%) sequence identity to SEQ ID NO: 19. In certain embodiments, the 5' ITR nucleotide sequence has at least 90% (e.g., at least 95%, 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 90% (e.g., at least 95%, 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 transfer genome comprises a nucleotide sequence set forth in any one of SEQ ID NOs: 46-50, 61, 64, 67, 74, 76, 78, 80, 82, 84, 86, and 89, 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.

[0094] In certain embodiments, the 5' ITR or 3' ITR are from AAV5. In certain embodiments, both the 5' ITR and 3' ITR are from AAV5. In certain embodiments, the 5' ITR nucleotide sequence has at least 90% (e.g., at least 95%, 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 90% (e.g., at least 95%, 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 5' ITR nucleotide sequence has at least 90% (e.g., at least 95%, at least 96%, at least 97%, at least 98%, at least 99%, or 100%) sequence identity to SEQ ID NO: 20, and the 3' ITR nucleotide sequence has at least 90% (e.g., at least 95%, 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 transfer genome comprises a nucleotide sequence set forth in any one of SEQ ID NOs: 46-50, 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.

30 [0095] 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).

[0096] 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,
5 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,
10 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
15 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
20 nucleotide sequence set forth in SEQ ID NO: 19.

[0097] 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*
25 *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: 56. In certain embodiments, the 3' ITR comprises a nucleotide sequence at least 95%, 96%, 97%, 98%, or 99% identical to SEQ ID NO: 57. In certain embodiments, the 3' ITR
30 comprises the nucleotide sequence set forth in SEQ ID NO: 57. 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: 57. In certain embodiments, the nucleotide sequence of the 3' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 57.

[0098] In certain embodiments, the transfer genome comprises from 5' to 3': a 5' ITR; an internal element comprising from 5' to 3': a TRE, optionally a non-coding exon and an intron, a PAH coding sequence, and a polyadenylation sequence, as disclosed herein; a non-resolvable ITR; a nucleotide sequence complementary to the internal element; and a 3' ITR.

- 5 Such transfer genome can form a self-complementary, double-stranded DNA genome of the AAV after infection and before replication.

[0099] In certain embodiments, the transfer genome comprises from 5' to 3': a 5' ITR, a TRE, optionally a non-coding exon and an intron, a PAH coding sequence, a polyadenylation sequence, and a 3' ITR. In certain embodiments, the 5' ITR has at least 90%,
 10 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID: 18, 20, or 26; the TRE has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any one of SEQ ID NOs: 28-30 and 32-41; the intron has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID NO: 31 or 35; the PAH coding sequence has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%,
 15 or 99% sequence identity to SEQ ID NO: 25; the polyadenylation sequence has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to any one of SEQ ID NOs: 42, 43, and 45; and/or the 3' ITR has at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% sequence identity to SEQ ID: 19, 21, or 27. In certain
 20 embodiments, the 5' ITR comprises or consists of a nucleotide sequence selected from the group consisting of SEQ ID NOs: 18, 20, and 26; the TRE comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 28-30 and 32-41; the intron comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 31 and 35; the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25; the polyadenylation sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID
 25 NOs: 42, 43, and 45; and/or the 3' ITR comprises or consists of a nucleotide sequence selected from the group consisting of SEQ ID NOs: 19, 21, and 27. In certain embodiments, the 5' ITR comprises or consists of the sequence set forth in SEQ ID NO: 18; the TRE comprises the sequence set forth in SEQ ID NO: 28; the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25; the polyadenylation sequence comprises the sequence set forth in SEQ ID NO: 42; and/or the 3' ITR comprises or consists of the sequence set forth
 30 in SEQ ID NO: 19. In certain embodiments, the 5' ITR comprises or consists of the sequence set forth in SEQ ID NO: 26; the TRE comprises from 5' to 3' the sequence set forth in SEQ ID NO: 29, and the sequence set forth in SEQ ID NO: 30 (e.g., the TRE comprises the sequence set forth in SEQ ID NO: 32); the intron comprises the sequence set forth in SEQ ID

NO: 31; the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25; the polyadenylation sequence comprises the sequence set forth in SEQ ID NO: 43 and/or the 3' ITR comprises or consists of the sequence set forth in SEQ ID NO: 27. In certain embodiments, the 5' ITR comprises or consists of the sequence set forth in SEQ ID NO: 18; the TRE comprises from 5' to 3' the sequence set forth in SEQ ID NO: 33, and the sequence set forth in SEQ ID NO: 34 (e.g., the TRE comprises the sequence set forth in SEQ ID NO: 36); the intron comprises the sequence set forth in SEQ ID NO: 35; the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25; the polyadenylation sequence comprises the sequence set forth in SEQ ID NO: 45; and/or the 3' ITR comprises or consists of the sequence set forth in SEQ ID NO: 19. In certain embodiments, the 5' ITR comprises or consists of the sequence set forth in SEQ ID NO: 18; the TRE comprises from 5' to 3' the sequence set forth in SEQ ID NO: 37, and the sequence set forth in SEQ ID NO: 38 (e.g., the TRE comprises the sequence set forth in SEQ ID NO: 39); the intron comprises the sequence set forth in SEQ ID NO: 35; the PAH coding sequence comprises the sequence set forth in SEQ ID NO: 25; the polyadenylation sequence comprises the sequence set forth in SEQ ID NO: 45; and/or the 3' ITR comprises or consists of the sequence set forth in SEQ ID NO: 19. In certain embodiments, the 5' ITR comprises or consists of the sequence set forth in SEQ ID NO: 18; the TRE comprises from 5' to 3' the sequence set forth in SEQ ID NO: 37, and the sequence set forth in SEQ ID NO: 40 (e.g., the TRE comprises the sequence set forth in SEQ ID NO: 41); the polyadenylation sequence comprises the sequence set forth in SEQ ID NO: 45; and/or the 3' ITR comprises or consists of the sequence set forth in SEQ ID NO: 19.

[00100] In certain embodiments, the transfer genome comprises a sequence at least 80% (e.g., at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%) identical to the sequence set forth in SEQ ID NO: 51, 52, 53, 54, 55, 62, 65, 68, 75, 77, 79, 81, 83, 85, 87, or 90. In certain embodiments, the transfer genome comprises the sequence set forth in SEQ ID NO: 51, 52, 53, 54, 55, 62, 65, 68, 75, 77, 79, 81, 83, 85, 87, or 90. In certain embodiments, the transfer genome comprises the sequence set forth in SEQ ID NO: 52. In certain embodiments, the transfer genome consists of the sequence set forth in SEQ ID NO: 51, 52, 53, 54, 55, 62, 65, 68, 75, 77, 79, 81, 83, 85, 87, or 90. In certain embodiments, the transfer genome consists of the sequence set forth in SEQ ID NO: 52.

[00101] 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' following genetic elements: a 5' ITR element (e.g., the 5' ITR of SEQ ID NOs: 26), a human HCR1 (e.g., the HCR1 of SEQ ID

NOS: 29), an hAAT promoter (e.g., the hAAT promoter of SEQ ID NOS: 30), an SV40 intron
 (e.g., the SV40 intron of SEQ ID NOS: 31), a silently altered human PAH coding sequence
 (e.g., the PAH coding sequence of SEQ ID NOS: 25), an SV40 polyadenylation sequence
 (e.g., the SV40 polyadenylation sequence of SEQ ID NOS: 43), and a 3' ITR element (e.g.,
 5 the 3' ITR of SEQ ID NOS: 27); (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' following genetic elements: a 5' ITR element (e.g., the 5' ITR of SEQ ID NOS: 26), a
 human HCR1 (e.g., the HCR1 of SEQ ID NOS: 29), an hAAT promoter (e.g., the hAAT
 promoter of SEQ ID NOS: 30), an SV40 intron (e.g., the SV40 intron of SEQ ID NOS: 31), a
 10 silently altered human PAH coding sequence (e.g., the PAH coding sequence of SEQ ID
 NOS: 25), an SV40 polyadenylation sequence (e.g., the SV40 polyadenylation sequence of
 SEQ ID NOS: 43), and a 3' ITR element (e.g., the 3' ITR of SEQ ID NOS: 27; 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' following genetic elements: a 5' ITR element (e.g., the 5' ITR of
 15 SEQ ID NOS: 26), a human HCR1 (e.g., the HCR1 of SEQ ID NOS: 29), an hAAT promoter
 (e.g., the hAAT promoter of SEQ ID NOS: 30), an SV40 intron (e.g., the SV40 intron of SEQ
 ID NOS: 31), a silently altered human PAH coding sequence (e.g., the PAH coding sequence
 of SEQ ID NOS: 25), an SV40 polyadenylation sequence (e.g., the SV40 polyadenylation
 sequence of SEQ ID NOS: 43), and a 3' ITR element (e.g., the 3' ITR of SEQ ID NOS: 27).
 20 **[00102]** 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 the nucleotide sequence set forth in any one of
 SEQ ID NOS: 24, 25, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 61, 62, 64, 65, 67, 68, 74, 75, 76,
 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 89, or 90; (b) an AAV capsid protein comprising the
 25 amino acid sequence of amino acids 138-736 of SEQ ID NO: 16, and a transfer genome
 comprising the nucleotide sequence set forth in any one of SEQ ID NOS: 24, 25, 46, 47, 48,
 49, 50, 51, 52, 53, 54, 55, 61, 62, 64, 65, 67, 68, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85,
 86, 87, 89, or 90; and/or (c) an AAV capsid protein comprising the amino acid sequence of
 SEQ ID NO: 16, and a transfer genome comprising the nucleotide sequence set forth in any
 30 one of SEQ ID NOS: 24, 25, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 61, 62, 64, 65, 67, 68, 74,
 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 89, or 90.

[00103] In another aspect, provided herein is a polynucleotide comprising a nucleic
 acid sequence that is at least 80% (e.g., at least 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%,
 97%, 98%, or 99%) identical to the nucleic acid sequence set forth in SEQ ID NO: 88, 91, or

92. In certain embodiments, the polynucleotide comprises the nucleic acid sequence set forth in SEQ ID NO: 88, 91, or 92. In certain embodiments, the polynucleotide comprises the nucleic acid sequence set forth in SEQ ID NO: 92. In certain embodiments, the polynucleotide consists of the nucleic acid sequence set forth in SEQ ID NO: 88, 91, or 92. In certain embodiments, the polynucleotide consists of the nucleic acid sequence set forth in SEQ ID NO: 92.

[00104] 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.

[00105] In another aspect, the instant disclosure provides a polynucleotide comprising a coding sequence encoding a human PAH protein or a fragment thereof, wherein the coding sequence has been codon-altered to have less than 100% (e.g., less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%, or 50%) identical to a wild-type human PAH gene. In certain embodiments, the polynucleotide comprises the sequence set forth in SEQ ID NO: 25. In certain embodiments, the polynucleotide comprises nucleotides 4 to 1359 of the sequence set forth in SEQ ID NO: 25. The polynucleotide can comprise DNA, RNA, modified DNA, modified RNA, or a combination thereof. In certain embodiments, the polynucleotide is an expression vector.

III. Method of Use

[00106] In another aspect, the instant disclosure provides methods for 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 restoring PAH

expression. Accordingly, in certain embodiments, the methods disclosed herein involve transducing the cell with a replication-defective AAV as disclosed herein.

[00107] 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 are active in Phe metabolism are of particular interest. Accordingly, in certain embodiments, the method is applied to cells in the liver, kidney, brain, pituitary gland, adrenal gland, pancreas, urinary bladder, gallbladder, colon, small intestine, or breast. In certain embodiments, the method is applied to hepatocytes and/or renal cells.

[00108] The methods disclosed herein can be performed *in vitro* for research purposes or can be performed *ex vivo* or *in vivo* for therapeutic purposes.

[00109] In certain embodiments, the cell to be transduced is in a mammalian subject and the AAV is administered to the subject in an amount effective to transduce the cell in the subject. Accordingly, in certain embodiments, the instant disclosure provides a method for treating a subject having a disease or disorder associated with a PAH gene mutation, the method generally comprising administering to the subject an effective amount of a replication-defective AAV as disclosed herein. The subject can be a human subject, a non-human primate subject (e.g., a cynomolgus), or a rodent subject (e.g., a mouse) with a PAH mutation, or a non-human primate subject (e.g., a cynomolgus) or a rodent subject (e.g., a mouse) containing PAH-mutant 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.

[00110] 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' following genetic elements: a 5' ITR element (e.g., the 5' ITR of SEQ ID NOs: 26), a human HCR1 (e.g., the HCR1 of SEQ ID NOs: 29), an hAAT promoter (e.g., the hAAT promoter of SEQ ID NOs: 30), an SV40 intron (e.g., the SV40 intron of SEQ ID NOs: 31), a silently altered human PAH coding sequence (e.g., the PAH coding sequence of SEQ ID NOs: 25), an SV40 polyadenylation sequence (e.g., the SV40 polyadenylation sequence of SEQ ID NOs: 43), and a 3' ITR element (e.g., the 3' ITR of SEQ ID NOs: 27); (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' following genetic elements: a 5' ITR element (e.g., the 5' ITR of SEQ ID NOs: 26), a human HCR1 (e.g., the HCR1 of SEQ ID NOs: 29), an hAAT

promoter (e.g., the hAAT promoter of SEQ ID NOs: 30), an SV40 intron (e.g., the SV40 intron of SEQ ID NOs: 31), a silently altered human PAH coding sequence (e.g., the PAH coding sequence of SEQ ID NOs: 25), an SV40 polyadenylation sequence (e.g., the SV40 polyadenylation sequence of SEQ ID NOs: 43), and a 3' ITR element (e.g., the 3' ITR of SEQ ID NOs: 27; 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' following genetic elements: a 5' ITR element (e.g., the 5' ITR of SEQ ID NOs: 26), a human HCR1 (e.g., the HCR1 of SEQ ID NOs: 29), an hAAT promoter (e.g., the hAAT promoter of SEQ ID NOs: 30), an SV40 intron (e.g., the SV40 intron of SEQ ID NOs: 31), a silently altered human PAH coding sequence (e.g., the PAH coding sequence of SEQ ID NOs: 25), an SV40 polyadenylation sequence (e.g., the SV40 polyadenylation sequence of SEQ ID NOs: 43), and a 3' ITR element (e.g., the 3' ITR of SEQ ID NOs: 27).

[00111] 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 the nucleotide sequence set forth in any one of SEQ ID NOs: 24, 25, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 61, 62, 64, 65, 67, 68, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 89, or 90; (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 the nucleotide sequence set forth in any one of SEQ ID NOs: 24, 25, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 61, 62, 64, 65, 67, 68, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 89, or 90; and/or (c) an AAV capsid protein comprising the amino acid sequence of SEQ ID NO: 16, and a transfer genome comprising the nucleotide sequence set forth in any one of SEQ ID NOs: 24, 25, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 61, 62, 64, 65, 67, 68, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 89, or 90.

[00112] The methods disclosed herein are particularly advantageous in that they are capable of expressing a PAH protein in a cell with high efficiency both *in vivo* and *in vitro*. In certain embodiments, the expression level of the PAH protein is at least 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80%, 90%, or 100% of the expression level of the endogenous PAH protein in a cell of the same type that does not have a mutation in the PAH gene. In certain embodiments, the expression level of the PAH protein is at least 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 3, 4, 5, 6, 7, 8, 9, or 10 fold higher than the expression level of the endogenous PAH protein in a cell of the same type that does not have a mutation in the PAH gene. Any methods of determining the expression level of the PAH protein can be employed including, without limitation, ELISA, Western blotting, immunostaining, and mass spectrometry.

[00113] 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.

[00114] An AAV composition disclosed herein can be administered to a subject by any appropriate route including, without limitation, intravenous, intraperitoneal, subcutaneous, intramuscular, intranasal, topical or intradermal routes. In certain embodiments, the composition is formulated for administration via intravenous injection or subcutaneous injection.

IV. AAV Packaging Systems

[00115] 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 transfer genome for expression of the PAH gene as disclosed herein, wherein the packaging system is operative in a cell for enclosing the transfer genome in the capsid to form the AAV.

[00116] 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 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.).

[00117] 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.

[00118] 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
5 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.

[00119] 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
10 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.

[00120] 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
15 embodiments the second vector and the third vector are contained within a second transfecting plasmid.

[00121] 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
20 embodiments, the second vector and the third vector are contained within a recombinant helper virus.

[00122] 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 transfer genome in the capsid to form the AAV as described herein. Exemplary
25 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
30 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 transfer genome as described herein being delivered in the form of a transfecting plasmid or a recombinant helper virus).

5 V. Examples

[00123] The recombinant AAV vectors disclosed herein mediate highly efficient gene transfer *in vitro* and *in vivo*. The following examples demonstrate the efficient restoration of the expression of the PAH gene, which is mutated in certain human diseases, such as phenylketonuria, using an AAV-based vector as disclosed herein. These examples are
10 offered by way of illustration, and not by way of limitation.

Example 1: Human PAH transfer vector

[00124] This example provides human PAH transfer vectors pHMI-hPAH-TC-004, pHMI-hPAH-TC-025, pHMI-hPAH-TC-010, pHMI-hPAH-TC-011, and pHMI-hPAH-TC-012 for expression of human PAH in a human or mouse cell.

15 a) pHMI-hPAH-TC-004

[00125] PAH transfer vector pHMI-hPAH-TC-004, as shown in Figure 1A, comprises 5' to 3' the following genetic elements: a 5' ITR element, a CAG promoter, a silently altered human PAH coding sequence, an SV40 polyadenylation sequence, and a 3' ITR element. The sequences of these elements are set forth in Table 1. This vector is capable of expressing a
20 human PAH protein in a cell (e.g., a human cell or a mouse cell) to which the vector is transduced.

Table 1: Genetic elements in human PAH transfer vector pHMI-hPAH-TC-004

Genetic Element	SEQ ID NO
5' ITR element	18
CAG promoter	28
codon-altered human PAH coding sequence	25
SV40 polyadenylation sequence	42
3' ITR element	19
Transfer genome (from promoter to polyadenylation sequence)	46
Transfer genome (from 5' ITR to 3' ITR)	51

b) pHMI-hPAH-TC-025

[00126] PAH transfer vector pHMI-hPAH-TC-025, as shown in Figure 1B, comprises 5' to 3' the following genetic elements: a truncated 5' ITR element, a human hepatic control region 1 (HCR1), a human α 1-antitrypsin (hAAT) promoter, an SV40 intron, a silently altered human PAH coding sequence, an SV40 polyadenylation sequence, and a modified 3' ITR element. The sequences of these elements are set forth in Table 2. The truncated 5' ITR allows the vector to form a double-stranded AAV genome after transduction into cells. This vector is capable of expressing a human PAH protein in a human hepatocyte to which the vector is transduced.

Table 2: Genetic elements in human PAH transfer vector pHMI-hPAH-TC-025

Genetic Element	SEQ ID NO
truncated 5' ITR element	26
human HCR1	29
human α 1-antitrypsin (hAAT) promoter	30
SV40 intron	31
transcriptional regulatory region comprising the human HCR1 and hAAT promoter	32
codon-altered human PAH coding sequence	25
SV40 polyadenylation sequence	43
modified 3' ITR element	27
Transfer genome (from HCR1 to polyadenylation sequence)	47
Transfer genome (from 5' ITR to 3' ITR)	52
Full sequence of transfer vector	92

c) pHMI-hPAH-TC-010

[00127] PAH transfer vector pHMI-hPAH-TC-010, as shown in Figure 1C, comprises 5' to 3' the following genetic elements: a 5' ITR element, a hepatic specific regulatory module of hAAT promoter, a human TTR promoter, a modified minute virus of mouse (MVM) intron, a silently altered human PAH coding sequence, an SV40 polyadenylation sequence, and a 3' ITR element. The sequences of these elements are set forth in Table 3. This vector is capable of expressing a human PAH protein in a cell (e.g., a human cell or a mouse cell) to which the vector is transduced, particularly at a high level in a hepatocyte.

Table 3: Genetic elements in human PAH transfer vector pHMI-hPAH-TC-010

Genetic Element	SEQ ID NO
5' ITR element	18
hepatic specific regulatory module of hAAT promoter	33
human TTR promoter	34
modified minute virus of mouse (MVM) intron	35
transcriptional regulatory region comprising the hepatic specific regulatory module (HSRM) and human TTR promoter	36
codon-altered human PAH coding sequence	25
SV40 polyadenylation sequence	45
3' ITR element	19
Transfer genome (from HSRM to polyadenylation sequence)	48
Transfer genome(from 5' ITR to 3' ITR)	53

d) pHMI-hPAH-TC-011

[00128] PAH transfer vector pHMI-hPAH-TC-011, as shown in Figure 1D, comprises

- 5 5' to 3' the following genetic elements: a 5' ITR element, a human HCR1, a human α 1-antitrypsin (hAAT) promoter, an modified MVM intron, a silently altered human PAH coding sequence, an SV40 polyadenylation sequence, and a 3' ITR element. The sequences of these elements are set forth in Table 4. This vector is capable of expressing a human PAH protein in a human hepatocyte to which the vector is transduced.

Table 4: Genetic elements in human PAH transfer vector pHMI-hPAH-TC-011

Genetic Element	SEQ ID NO
truncated 5' ITR element	18
human HCR1	37
human α 1-antitrypsin (hAAT) promoter	38
modified minute virus of mouse (MVM) intron	35
transcriptional regulatory region comprising the human HCR1 and hAAT promoter	39
codon-altered human PAH coding sequence	25
SV40 polyadenylation sequence	45

Genetic Element	SEQ ID NO
modified 3' ITR element	19
Transfer genome (from HCR1 to polyadenylation sequence)	49
Transfer genome (from 5' ITR to 3' ITR)	54

e) pHMI-hPAH-TC-012

[00129] PAH transfer vector pHMI-hPAH-TC-012, as shown in Figure 1E, comprises 5' to 3' the following genetic elements: a 5' ITR element, a human hepatic control region 1 (HCR1), a modified human EF-1 α promoter, a silently altered human PAH coding sequence, an SV40 polyadenylation sequence, and a 3' ITR element. The sequences of these elements are set forth in Table 5. This vector is capable of expressing a human PAH protein in a human hepatocyte to which the vector is transduced.

Table 5: Genetic elements in human PAH transfer vector pHMI-hPAH-TC-012

Genetic Element	SEQ ID NO
truncated 5' ITR element	18
human hepatic control region 1 (HCR1)	37
modified human EF-1 α promoter	40
transcriptional regulatory region comprising the human HCR1 and modified human EF-1 α promoter	41
codon-altered human PAH coding sequence	25
SV40 polyadenylation sequence	45
modified 3' ITR element	19
Transfer genome (from HCR1 to polyadenylation sequence)	50
Transfer genome (from 5' ITR to 3' ITR)	55

[00130] The silent alteration significantly improves 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 CAG 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 CAG promoter operably linked to a codon-altered human PAH coding sequence as set forth in SEQ ID NO: 25. The pCOH-CO-PAH and pHMI-CO-PAH vectors are 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 several fold higher than VG-GT-PAH (“WT-hPAH”).

[00131] The vectors disclosed herein can be packaged in an AAV clade F capsid, such as an AAVHSC5, AAVHSC7, AAVHSC15 or AAVHSC17 capsid. The packaged viral particles can be administered to a wild-type animal, a PAH deficient animal, or a reconstituted animal having human hepatocytes obtained from a patient with phenylketonuria caused by a PAH mutation. The gene transfer efficiency can be measured by collecting liver samples and quantifying the percentage of PAH-positive cells (e.g., cells that have a unique nucleotide sequence from the vector, cells that express a wild-type PAH protein, or cells with a higher PAH activity than in cells from a control animal not receiving the PAH expression vector). The restoration of phenylalanine metabolism, which indicates the efficacy of the PAH expression vectors, can be assessed by measuring the Phe level in the blood and by observing the coat color of the mouse. Safety of the viral particle administration can be evaluated by measuring the aspartate aminotransferase (AST) and alanine aminotransferase (ALT) levels in serum.

Example 2: Mouse PAH gene transfer in a mouse model

[00132] This example provides a mouse PAH transfer vector rAAV-CBA-mPAH that is similar to the human PAH transfer vector pHMI-hPAH-TC-004 described in Example 1 except that a wild-type mouse PAH coding sequence is substituted for the codon-altered human PAH coding sequence. This vector is capable of expressing a mouse PAH protein in a cell (e.g., a human cell or a mouse cell) to which the vector is transduced.

[00133] Briefly, *Pah*^{-/-} (*PAH*^{enu2}) mice were housed in clear polycarbonate cages with contact bedding in an isolator. Picolab Mouse Diet 5058 was provided to the animals *ad libitum*. Spring or tap water acidified with 1N HCl to a targeted pH of 2.5 - 3.0 was provided *ad libitum*. Vectors packaged in AAVHSC15 capsid were prepared in PBS (with Ca and Mg), supplemented with 35 mM NaCl, 1% sucrose, and 0.05% Pluronic F-68. The formulation was injected intravenously via the tail vein.

[00134] Blood samples were collected every week after the administration of the PAH transfer vector (0 week: prior to administration) by facial vein puncture or tail snip. The samples were allowed to clot at room temperature for at least 30 minutes, centrifuged at ambient temperature at minimum 1000 x g for 10 minutes and the serum samples were extracted. Serum samples were stored at -70 °C. Serum phenylalanine and tyrosine levels were measured by tandem mass spectrometry.

[00135] For collection of tissue samples, the animals underwent cardiac perfusion with saline. Liver (caudate lobe), kidney (left), brain, heart, and muscle (quadriceps) tissues were snap frozen in liquid nitrogen and stored at -70 °C. The snap frozen tissues were ground into powder in liquid nitrogen in a mortar and pestle and divided into aliquots to test for PAH expression for vector genome biodistribution by qPCR.

[00136] The safety of the rAAV-CBA-mPAH vector was assessed by measuring the aspartate aminotransferase (AST) and alanine aminotransferase (ALT) levels in the treated animals. Serum samples were collected pre-dose and one week after administration of the viral particles. The levels of AST and ALT were measured by the Sigma MAK055 and Sigma MAK052 ELISA kits.

[00137] The *pah*^{-/-} mice manifested phenylketonuria and had lighter coat color than wild-type mice. As shown in Figures 3A and 3B, the administration of the rAAV-CBA-mPAH vector lead to significant reduction of Phe levels in the serum within one week, and the Phe levels remained low for four weeks. The coat color also changed from brown to black within one week.

[00138] Expression of mPAH was also observed in tissue samples. As shown in Figure 4, DNA of the rAAV-CBA-mPAH vector was detectable in many organs, wherein the numbers of viral genomes per 10⁶ cells was the highest in liver, heart, and kidney.

[00139] With respect to the safety of the AAV administration, the AST and ALT levels remained low after administration (Figures 5A and 5B), suggesting that the rAAV-CBA-mPAH vector was not toxic to the liver.

Example 3: Human PAH gene transfer in a mouse model

[00140] This example demonstrates that the PAH transfer vectors described in Example 1 effectively reversed the phenotype caused by PAH gene deficiency in a mouse model. The mouse model, AAV packaging and formulation, and methods for examining gene transfer efficiency were identical to the ones described in Example 2.

[00141] To examine the efficacy of the five PAH transfer vectors in reversing the phenotypes, a single dose of 2.6×10^{13} vector genomes per kg of body weight for male mice, or a dose of 6×10^{13} vector genomes per kg of body weight for female mice. The *pah*^{-/-} mice manifested increased level of phenylalanine (Phe) and reduced level of tyrosine (Tyr) in the serum. As shown in Figures 6A-6H, the administration of any one of the five vectors led to significant reduction of Phe levels and increase of Tyr levels within one week. The efficacy lasted for at least 12 weeks in male mice, and at least 6 weeks in female mice. Other than

pHMI-hPAH-TC-004, all the vectors maintained complete reduction of serum Phe levels during the time examined.

[00142] To examine the long-term efficacy of pHMI-hPAH-TC-025 in reversing the phenotype caused by PAH gene deficiency, a single dose of 2.6×10^{13} vector genomes per kg of body weight was administered to male mice, or a single dose of 6×10^{13} vector genomes per kg of body weight was administered to female mice. As shown in Figures 6I and 6J, the administration of the pHMI-hPAH-TC-025 vector led to significant reduction of Phe levels within one week. This reduction persisted for at least 48 weeks in male mice, and at least 46 weeks in female mice. Additionally, within two weeks post administration of the AAV, the coat color of the mice administered with pHMI-hPAH-TC-004 changed from brown to black. An increase of PAH mRNA was observed by ddPCR in the liver samples of these mice collected 4 weeks post injection relative to the mice not administered with AAV vectors. An increase of the PAH enzymatic activity was also detected in liver samples by mass spectrometry.

[00143] The efficacy of different doses of the pHMI-hPAH-TC-025 vector was further assessed. A single dose of 2.6×10^{11} , 2.6×10^{12} , or 2.6×10^{13} vector genomes per kg of body weight was administered to male mice and female mice, and the serum levels of Phe and Tyr were measured. As shown in Figures 7A-7D, the dose of 2.6×10^{13} vector genomes per kg of body weight reduced the Phe levels and increased the Tyr levels more significantly than the two lower doses, and maintained complete reduction of serum Phe levels during the time examined in both male and female subjects.

Example 3: Additional human PAH transfer vectors

[00144] This example provides human PAH transfer vectors pHMI-hPAH-TC-009, pHMI-hPAH-TC-013, and pHMI-hPAH-TC-017 for expression of human PAH in a human or mouse cell. Vector maps are shown in Figure 8A, 8B, and 8C, respectively.

a) pHMI-hPAH-TC-009

[00145] PAH transfer vector pHMI-hPAH-TC-009, as shown in Figure 8A, comprises 5' to 3' the following genetic elements: a 5' ITR element, a CMV enhancer, a CBA promoter, a rabbit β -globin element, a human PAH coding sequence, a polyadenylation sequence, and a 3' ITR element. The sequences of these elements are set forth in Table 6. This vector is capable of expressing a human PAH protein in a cell (e.g., a human cell or a mouse cell) to which the vector is transduced.

Table 6: Genetic elements in human PAH transfer vector pHMI-hPAH-TC-009

Genetic Element	SEQ ID NO
5' ITR element	18
CMV enhancer	58
CBA promoter	59
Rabbit β -globin element	60
codon-altered human PAH coding sequence	25
Polyadenylation sequence	45
3' ITR element	19
Transfer genome (from CMV to polyadenylation sequence)	61
Transfer genome (from 5' ITR to 3' ITR)	62

b) pHMI-hPAH-TC-013

[00146] PAH transfer vector pHMI-hPAH-TC-013, as shown in **Figure 8B**, comprises 5' to 3' the following genetic elements: a 5' ITR element, a CASI promoter region (comprising a CMV enhancer, a CASI promoter, and a ubiquitin C enhancer element (hUBC exon)), a human PAH coding sequence, a polyadenylation sequence, and a 3' ITR element. The sequences of these elements are set forth in Table 7. This vector is capable of expressing a human PAH protein in a cell (e.g., a human cell or a mouse cell) to which the vector is transduced.

Table 7: Genetic elements in human PAH transfer vector pHMI-hPAH-TC-013

Genetic Element	SEQ ID NO
5' ITR element	18
CASI promoter region	63
Human PAH coding sequence	25
Polyadenylation sequence	45
3' ITR element	19
Transfer genome (from promoter region to polyadenylation sequence)	64
Transfer genome (from 5' ITR to 3' ITR)	65

f) pHMI-hPAH-TC-017

[00147] PAH transfer vector pHMI-hPAH-TC-017, as shown in **Figure 8C**, comprises 5' to 3' the following genetic elements: a 5' ITR element, an hAAT promoter region (comprising an ABMP enhancer (an enhancer region adjacent to a gene on chromosome 9 that expresses highly in liver, 5' to the ATG), a TTR enhancer, an hAAT promoter, and an MVM intron), a human PAH coding sequence, a polyadenylation sequence, and a 3' ITR element. The sequences of these elements are set forth in Table 8. This vector is capable of expressing a human PAH protein in a cell (e.g., a human cell or a mouse cell) to which the vector is transduced.

Table 8: Genetic elements in human PAH transfer vector pHMI-hPAH-TC-017

Genetic Element	SEQ ID NO
5' ITR element	18
hAAT promoter region	66
Human PAH coding sequence	25
Polyadenylation sequence	45
3' ITR element	19
Transfer genome (from promoter region to polyadenylation sequence)	67
Transfer genome (from 5' ITR to 3' ITR)	68

[00148] The vectors described in this example were tested for expression in two different cell lines. 5×10^5 HEK293 cells (kidney; non-liver) and 5×10^5 Huh7 cells (liver) were transfected with 1 μ g each of the following vectors: pHMI-hPAH-TC-004 (PAH-004); pHMI-hPAH-TC-009 (PAH-009); pHMI-hPAH-TC-010 (PAH-010); pHMI-hPAH-TC-011 (PAH-011); pHMI-hPAH-TC-012 (PAH-012); pHMI-hPAH-TC-013 (PAH-013); pHMI-hPAH-TC-025 (LP1); pHMI-hPAH-TC-017 (PAH-017). 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 used as a loading control. PAH expression levels of all vectors were normalized to pHMI-hPAH-TC-004 expression level; data was collected from multiple independent transfections and plotted in **Figure 9**. **Figure 9A** shows the normalized PAH expression level of the indicated vectors in Huh7 cells. **Figure 9B** shows the normalized PAH expression level of the indicated vectors in HEK293 cells.

[00149] **Figure 10A and 10B** are graphs showing the serum phenylalanine levels over time of male and female homozygous *Pah*^{-/-} PAH^{enu2} mice respectively. Male and female mice were dosed at 2e13 vg/kg and 6e13 vg/kg respectively with pHMI-hPAH-TC-010 (hPAH-TC-010), pHMI-hPAH-TC-025 (hPAH-TC-025), pHMI-hPAH-TC-004 (hPAH-TC-004), pHMI-hPAH-TC-011 (hPAH-TC-011), or pHMI-hPAH-TC-012 (hPAH-TC-012) vectors packaged in AAVHSC15 capsid. Serum samples were collected weekly then biweekly after the administration. Serum phenylalanine concentrations were assessed by LC-MS/MS. **Figure 10C** is a graph showing the average baseline serum phenylalanine level for the male and female homozygous *Pah*^{-/-} PAH^{enu2} mice in the study. The data represents a total of 55 mice per group.

[00150] As shown in **Figure 10**, the administration of certain vectors led to significant reduction of Phe levels within one week of administration, and this reduction persisted for at least 45 weeks. **Figure 10** demonstrates that some of the PAH transfer vectors effectively reversed the phenotype caused by PAH gene deficiency in a mouse model. The mouse model, AAV packaging and formulation, and methods for examining gene transfer efficiency were identical to those described in Example 2 herein. The sizes of the AAV vectors were as follows: pHMI-hPAH-TC-010 (hPAH-TC-010): 2391 bp; pHMI-hPAH-TC-025 (hPAH-TC-025): 2351 bp; pHMI-hPAH-TC-004 (hPAH-TC-004): 3781 bp; pHMI-hPAH-TC-011 (hPAH-TC-011): 3158 bp; and pHMI-hPAH-TC-012 (hPAH-TC-012): 3799 bp.

Example 4: Additional human PAH transfer vectors

[00151] This example examines the effect of PAH gene CpG content on PAH protein expression, using the PAH transfer vectors pHMI-hPAH-TC-018, pHMI-hPAH-TC-019, pHMI-hPAH-TC-020, pHMI-hPAH-TC-021, pHMI-hPAH-TC-022, and pHMI-hPAH-TC-023. Vector maps are shown in **Figure 11A, 11B, 11C, 11D, 11E, and 11F**, respectively. These PAH transfer vectors comprise the sequences and elements set forth in Table 9.

Table 9: Genetic elements in PAH transfer vectors

	pHMI-hPAH-TC-XXX Vector					
	SEQ ID NO					
Genetic Element	-018	-019	-020	-021	-022	-023
5' ITR element	26	26	26	26	26	26
HCR1	29	29	29	29	29	29
hAAT promoter	30	30	30	30	30	30
SV40 intron	31	31	31	31	31	31

	pHMI-hPAH-TC-XXX Vector					
	SEQ ID NO					
Genetic Element	-018	-019	-020	-021	-022	-023
PAH coding sequence	69	70	71	72	73	24
Late SV40 polyadenylation sequence	45	45	45	45	45	45
3' ITR element	27	27	27	27	27	27
Transfer genome (from HCR1 to polyadenylation sequence)	74	76	78	80	82	84
Transfer genome (from 5' ITR to 3' ITR)	75	77	79	81	83	85

[00152] The vectors described in this example were tested for expression in HEK293 cells but under the control of a CBA promoter. 5×10^5 HEK293 cells were transfected with 1 μ g each of the following vectors: pHMI-hPAH-TC-004 (PAH-TC-004); pHMI-hPAH-TC-009 (PAH-TC-009); pHMI-hPAH-TC-018 (PAH-TC-018); pHMI-hPAH-TC-019 (PAH-TC-019); pHMI-hPAH-TC-020 (PAH-TC-020); pHMI-hPAH-TC-021 (PAH-TC-021); pHMI-hPAH-TC-022 (PAH-TC-022); pHMI-hPAH-TC-023 (PAH-TC-023). 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 used as a loading control. PAH expression levels of all vectors were normalized to pHMI-hPAH-TC-004 expression level; data was plotted in **Figure 12**.

[00153] **Figure 13** is a graph showing the serum phenylalanine levels over time of male homozygous *Pah*^{-/-} PAH^{enu2} mice. Male mice have been dosed at 2×10^{13} vg/kg with pHMI-hPAH-TC-018 (hPAH-TC-018); pHMI-hPAH-TC-019 (hPAH-TC-019); pHMI-hPAH-TC-020 (hPAH-TC-020); pHMI-hPAH-TC-021 (hPAH-TC-021); pHMI-hPAH-TC-022 (hPAH-TC-022); pHMI-hPAH-TC-023 (hPAH-TC-023); and pHMI-hPAH-TC-025 (hPAH-TC-025) vectors packaged in AAVHSC15 capsid. Serum samples were collected weekly after the administration. Serum phenylalanine concentration was assessed by LC-MS/MS.

[00154] As shown in **Figure 13**, the administration of certain vectors led to significant reduction of Phe levels within one week of administration, and this reduction persisted for at least 25 weeks. The mouse model, AAV packaging and formulation, and methods for examining gene transfer efficiency were identical to those previously described in Example 2 herein. The CpG content of the vectors were as follows: pHMI-hPAH-TC-018 (hPAH-TC-018): 2; pHMI-hPAH-TC-019 (hPAH-TC-019): 7; pHMI-hPAH-TC-020 (hPAH-TC-020): 22; pHMI-hPAH-TC-021 (hPAH-TC-021): 10; pHMI-hPAH-TC-022 (hPAH-TC-022): 7; pHMI-hPAH-TC-023 (hPAH-TC-023): 23; and pHMI-hPAH-TC-025 (hPAH-TC-025): 60.

10 **Example 5: Alternative ITR human PAH transfer vectors**

[00155] This example provides human PAH transfer vectors pHMI-01004 and pHMI-01008 for expression of human PAH in a human or mouse cell. Vector maps are shown in **Figure 14A** and **14B**, respectively. These PAH transfer vectors comprise the sequences and elements set forth in Table 10.

15 **Table 10: Genetic elements in PAH transfer vectors pHMI-01004 and pHMI-01008**

	pHMI-01004	pHMI-01008
Genetic Element	SEQ ID NO	SEQ ID NO
5' ITR element	26	26
HCR1	29	29
hAAT promoter region	30	30
SV40 intron	31	31
Human PAH coding sequence	25	25
Polyadenylation sequence	43	43
3' ITR element	27	57
Transfer genome (from HCR1 to polyadenylation sequence)	86	89
Transfer genome (from 5' ITR to 3' ITR)	87	90
Full sequence of transfer vector	88	91

* * *

[00156] 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.

[00157] 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 expressing a PAH polypeptide in a cell, the method comprising transducing the cell with a replication-defective adeno-associated virus (AAV) comprising:
 - 5 (a) an AAV capsid comprising an AAV Clade F capsid protein; and
 - (b) a transfer genome comprising a transcriptional regulatory element operably linked to a PAH coding sequence.
2. The method of claim 1, wherein the cell is a hepatocyte, a renal cell, or a cell in the
10 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.
15
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 adeno-associated virus (AAV) comprising:
 - (a) an AAV capsid comprising an AAV Clade F capsid protein; and
 - 20 (b) a transfer genome comprising a transcriptional regulatory element operably linked to a PAH coding sequence.
5. The method of claim 4, wherein the disease or disorder is phenylketonuria.
- 25 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 PAH coding sequence encodes an amino acid sequence set forth in SEQ ID NO: 23.
- 30 8. The method of claim 7, wherein the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 24.
9. The method of claim 7, wherein the PAH coding sequence is silently altered.

10. The method of claim 9, wherein the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 25.
11. The method of any one of the preceding claims, wherein the transcriptional regulatory
5 element is capable of mediating transcription in a hepatocyte, a renal cell, or a cell in the brain, pituitary gland, adrenal gland, pancreas, urinary bladder, gallbladder, colon, small intestine, or breast.
12. The method of claim 11, wherein the transcriptional regulatory element comprises one
10 of more of the elements selected from the group consisting of a CAG promoter, a human EF-1 α promoter, a human hepatic control region 1 (HCR1), a human α 1-antitrypsin (hAAT) promoter, a hepatic specific regulatory module of hAAT promoter, an SV40 intron, and a minute virus of mouse (MVM) intron.
13. The method of claim 12, wherein the transcriptional regulatory element comprises a
15 nucleotide sequence at least 90% identical to a sequence selected from the group consisting of SEQ ID NOs: 28-30 and 32-41.
14. The method of claim 13, wherein the transcriptional regulatory element comprises a
20 nucleotide sequence selected from the group consisting of SEQ ID NOs: 28-30 and 32-41.
15. The method of claim 14, wherein the transcriptional regulatory element comprises from 5' to 3' the nucleotide sequences set forth in SEQ ID NOs: 29, 30, and 31.
16. The method of claim 15, wherein the transcriptional regulatory element comprises the
25 nucleotide sequences set forth in SEQ ID NO: 32.
17. The method of any one of the preceding claims, wherein the transfer genome further comprises an intron operably linked to the PAH coding sequence.
30
18. The method of claim 17, wherein the intron comprises a nucleotide sequence at least 90% identical to the sequence set forth in SEQ ID NO: 31 or 35.

19. The method of claim 18, wherein the intron comprises the nucleotide sequence set forth in SEQ ID NO: 31 or 35.
20. The method of any one of claims 17-19, wherein the transfer genome comprises from
5 5' to 3': a non-coding exon, the intron, and the PAH coding sequence.
21. The method of any one of the preceding claims, wherein the transfer genome further comprises a polyadenylation sequence 3' to the PAH coding sequence.
- 10 22. The method of claim 21, wherein the polyadenylation sequence is an exogenous polyadenylation sequence.
23. The method of claim 22, wherein the exogenous polyadenylation sequence is an SV40 polyadenylation sequence.
- 15 24. The method of claim 23, wherein the SV40 polyadenylation sequence comprises a sequence selected from the group consisting of SEQ ID NOs: 42, 43, and 45.
25. The method of any one of the preceding claims, wherein the transfer genome comprises
20 a sequence selected from the group consisting of SEQ ID NOs: 46-50, 61, 64, 67, 74, 76, 78, 80, 82, 84, 86, and 89.
26. The method of any one of the preceding claims, wherein the transfer genome further comprises a 5' inverted terminal repeat (5' ITR) nucleotide sequence 5' of the genome, and a 3'
25 inverted terminal repeat (3' ITR) nucleotide sequence 3' of the genome.
27. The method of claim 26, 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.
- 30 28. The method of claim 26, wherein the 5' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO: 20, and the 3' ITR nucleotide sequence has at least 95% sequence identity to SEQ ID NO: 21.

29. The method of claim 26, 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.

5 30. The method of any one of the preceding claims, wherein the transfer genome comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 51-55, 62, 65, 68, 75, 77, 79, 81, 83, 85, 87, and 90.

31. The method of any one of the preceding claims, wherein the transfer genome consists
10 of a nucleotide sequence selected from the group consisting of SEQ ID NOs: 51-55, 62, 65, 68, 75, 77, 79, 81, 83, 85, 87, and 90.

32. The method of any one of the preceding claims, wherein the transfer genome consists of the nucleotide sequence set forth in SEQ ID NO: 52.

15

33. 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.

20

34. The method of claim 33, 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
25 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
30 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.

35. The method of claim 34, wherein:

- 5 (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;
- 10 (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;
- 15 (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.
- 20

36. The method of claim 34, 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.

25

37. 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.

30

38. The method of claim 37, 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.

39. The method of claim 38, 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.

40. The method of claim 38, wherein the capsid protein comprises the amino acid sequence
5 of amino acids 138-736 of SEQ ID NO: 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17.

41. 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,
10 15, 16, or 17.

42. The method of claim 41, 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.

43. The method of claim 42, 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

5 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.

44. The method of claim 42, 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.
45. A replication-defective adeno-associated virus (AAV) comprising:
- 5 (a) an AAV capsid comprising an AAV Clade F capsid protein; and
(b) a transfer genome comprising a transcriptional regulatory element operably linked to a PAH coding sequence.
46. The AAV of claim 45, wherein the PAH coding sequence encodes an amino acid
10 sequence set forth in SEQ ID NO: 23.
47. The AAV of claim 46, wherein the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 24.
- 15 48. The AAV of claim 46, wherein the PAH coding sequence is codon-altered and comprises the sequence set forth in SEQ ID NOs: 25 and 69-73.
49. The AAV of claim 46, wherein the PAH coding sequence is codon-altered and comprises the sequence set forth in SEQ ID NO: 25.
- 20 50. The AAV of any one of claims 45-49, wherein the transcriptional regulatory element is capable of mediating transcription in a hepatocyte, a renal cell, or a cell in the brain, pituitary gland, adrenal gland, pancreas, urinary bladder, gallbladder, colon, small intestine, or breast.
- 25 51. The AAV of claim 50, wherein the transcriptional regulatory element comprises one of more of the elements selected from the group consisting of a CAG promoter, a human EF-1 α promoter, a human HCR1, a human α 1-antitrypsin (hAAT) promoter, a hepatic specific regulatory module of hAAT promoter, an SV40 intron, and a minute virus of mouse (MVM) intron.
- 30 52. The AAV of claim 51, wherein the transcriptional regulatory element comprises a nucleotide sequence at least 90% identical to a sequence selected from the group consisting of SEQ ID NOs: 28-30 and 32-41.

53. The AAV of claim 52, wherein the transcriptional regulatory element comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 28-30 and 32-41.
54. The AAV of claim 53, wherein the transcriptional regulatory element comprises from
5 5' to 3' the nucleotide sequences set forth in SEQ ID NOs: 29, 30, and 31.
55. The AAV of claim 54, wherein the transcriptional regulatory element comprises the nucleotide sequences set forth in SEQ ID NO: 32.
- 10 56. The AAV of any one of claims 45-55, wherein the transfer genome further comprises an intron operably linked to the PAH coding sequence.
57. The AAV of claim 56, wherein the intron comprises a nucleotide sequence at least 90% identical to the sequence set forth in SEQ ID NO: 31 or 35.
- 15 58. The AAV of claim 57, wherein the intron comprises the nucleotide sequence set forth in SEQ ID NO: 31 or 35.
59. The AAV of any one of claims 56-58, wherein the transfer genome comprises from 5'
20 to 3': a non-coding exon, the intron, and the PAH coding sequence.
60. The AAV of any one of claims 45-59, wherein the transfer genome further comprises a polyadenylation sequence 3' to the PAH coding sequence.
- 25 61. The AAV of claim 60, wherein the polyadenylation sequence is an exogenous polyadenylation sequence.
62. The AAV of claim 61, wherein the exogenous polyadenylation sequence is an SV40 polyadenylation sequence.
- 30 63. The AAV of claim 62, wherein the SV40 polyadenylation sequence comprises a sequence selected from the group consisting of SEQ ID NOs: 42, 43, and 45.

64. The AAV of any one of claims 45-63, wherein the transfer genome comprises a sequence selected from the group consisting of SEQ ID NOs: 46-50, 61, 64, 67, 74, 76, 78, 80, 82, 84, 86, and 89.

5 65. The AAV of any one of claims 45-64, wherein the transfer genome further comprises a 5' inverted terminal repeat (5' ITR) nucleotide sequence 5' of the genome, and a 3' inverted terminal repeat (3' ITR) nucleotide sequence 3' of the genome.

66. The AAV of claim 65, 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.

67. The AAV of claim 65, 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.

68. The AAV of claim 65, 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

69. The AAV of any one of claims 45-68, wherein the transfer genome comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 51-55, 62, 65, 68, 75, 77, 79, 81, 83, 85, 87, and 90.

25 70. The AAV of any one of claims 45-69, wherein the transfer genome consists of a nucleotide sequence selected from the group consisting of SEQ ID NOs: 51-55, 62, 65, 68, 75, 77, 79, 81, 83, 85, 87, and 90.

71. The AAV of any one of claims 45-70, wherein the transfer genome consists of the
30 nucleotide sequence set forth in SEQ ID NO: 52.

72. The AAV of any one of claims 44-71, 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.

73. The AAV of claim 72, 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.

74. The AAV of claim 73, 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.

5

75. The AAV of claim 73, 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.

76. The AAV of any one of claims 45-75, 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.

77. The AAV of claim 76, 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.

78. The AAV of claim 77, 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.

79. The AAV of claim 77, 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.

80. The AAV of any one of claims 45-79, 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.

81. The AAV of claim 80, 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.

82. The AAV of claim 81, wherein:
- (a) the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 2 is T, and the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 2 is Q;
 - (b) the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 2 is I, and the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is Y;
 - (c) the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 2 is K;
 - (d) the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 2 is L, and the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 2 is S;
 - (e) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 2 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 2 is G;

- (f) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 2 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 2 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 2 is M;
- (g) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 2 is R;
- (h) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 2 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R; or
- (i) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 2 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 2 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 2 is C.

83. The AAV of claim 81, 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.
84. A pharmaceutical composition comprising an AAV of any one of claims 45-83.
85. 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 72-83; and
 - (c) a transfer genome as set forth in any one of claims 45-71, wherein the packaging system is operative in a cell for enclosing the transfer genome in the capsid to form the AAV.
86. The packaging system of claim 85, wherein the packaging system comprises a first vector comprising the Rep nucleotide sequence and the Cap nucleotide sequence, and a second vector comprising the transfer genome.

87. The packaging system of claims 85 or 86, wherein the Rep nucleotide sequence encodes an AAV2 Rep protein.
88. The packaging system of claim 87, wherein the AAV2 Rep protein is 78/68 or Rep
5 68/52.
89. The packaging system of claim 87 or 88, 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
10 70% across the length of the amino acid sequence encoding the AAV2 Rep protein.
90. The packaging system of any one of claims 85-89, further comprising a third vector, wherein the third vector is a helper virus vector.
- 15 91. The packaging system of claim 90, wherein the helper virus vector is an independent third vector.
92. The packaging system of claim 90, wherein the helper virus vector is integral with the first vector.
20
93. The packaging system of claim 90, wherein the helper virus vector is integral with the second vector.
94. The packaging system of any one of claims 90-93, wherein the third vector comprises
25 genes encoding helper virus proteins.
95. The packaging system of any one of claims 90-94, wherein the helper virus is selected from the group consisting of adenovirus, herpes virus, vaccinia virus, and cytomegalovirus (CMV).
30
96. The packaging system of claim 95, wherein the helper virus is adenovirus.
97. The packaging system of claim 96, wherein the adenovirus genome comprises one or more adenovirus RNA genes selected from the group consisting of E1, E2, E4 and VA.

98. The packaging system of claim 95, wherein the helper virus is herpes simplex virus (HSV).
- 5 99. The packaging system of claim 98, wherein the HSV genome comprises one or more of HSV genes selected from the group consisting of UL5/8/52, ICPO, ICP4, ICP22 and UL30/UL42.
100. The packaging system of any one of claims 90-99, wherein the first vector and the third
10 vector are contained within a first transfecting plasmid.
101. The packaging system of any one of claims 90-99, wherein the nucleotides of the second vector and the third vector are contained within a second transfecting plasmid.
- 15 102. The packaging system of any one of claims 90-99, wherein the nucleotides of the first vector and the third vector are cloned into a recombinant helper virus.
103. The packaging system of any one of claims 90-99, wherein the nucleotides of the second vector and the third vector are cloned into a recombinant helper virus.
20
104. A method for recombinant preparation of a replication-defective AAV, the method comprising introducing the packaging system of any one of claims 85-103 into a cell under conditions operative for enclosing the transfer genome in the capsid to form the AAV.

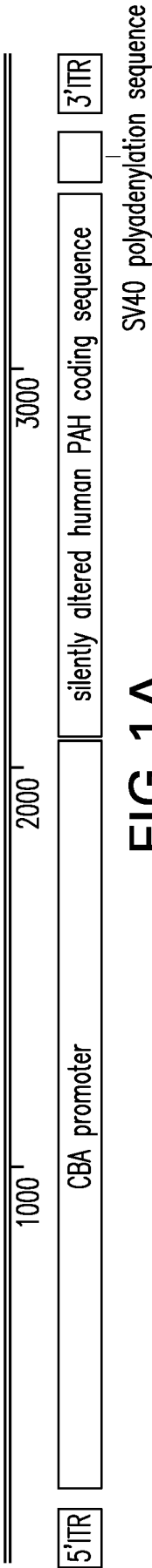


FIG. 1A

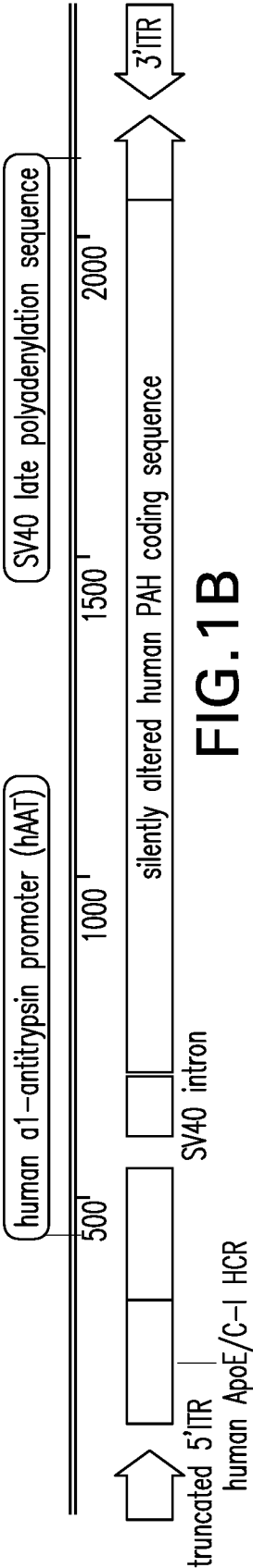


FIG. 1B

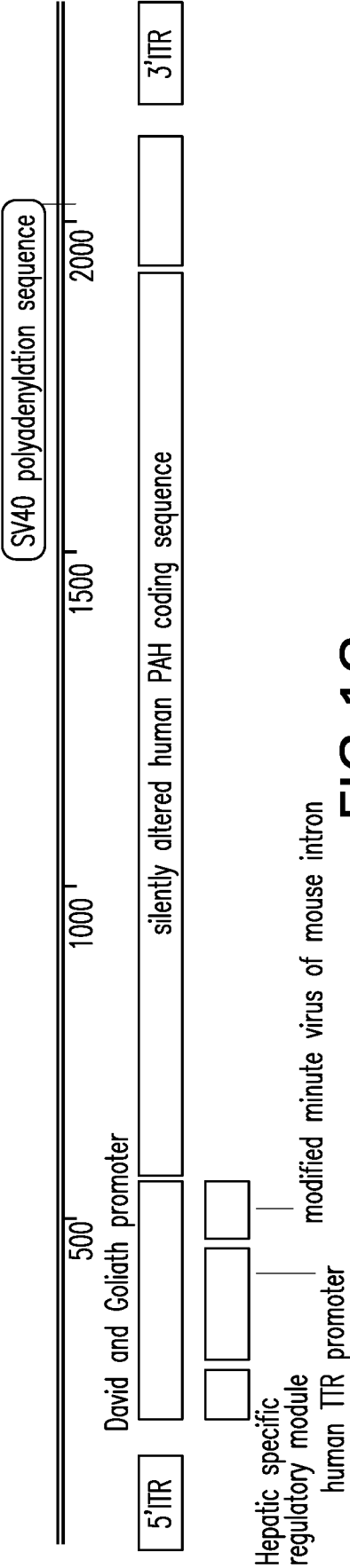


FIG. 1C

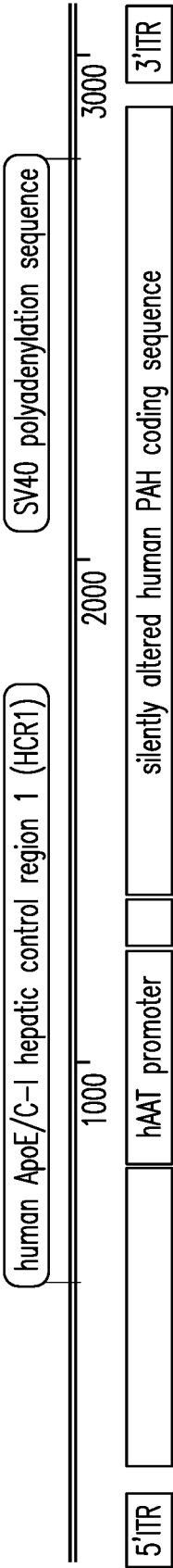


FIG. 1D

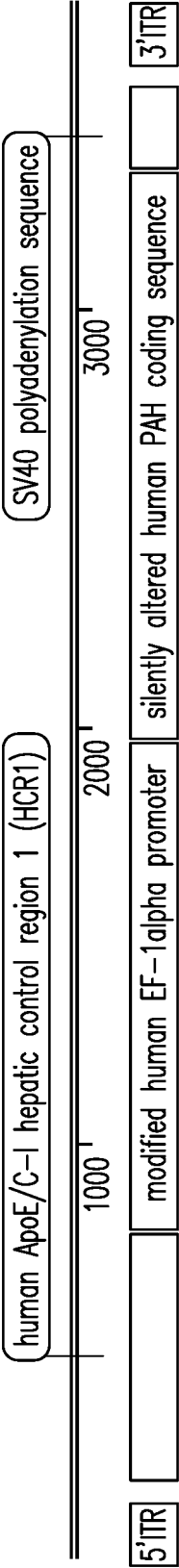


FIG. 1E

3/23

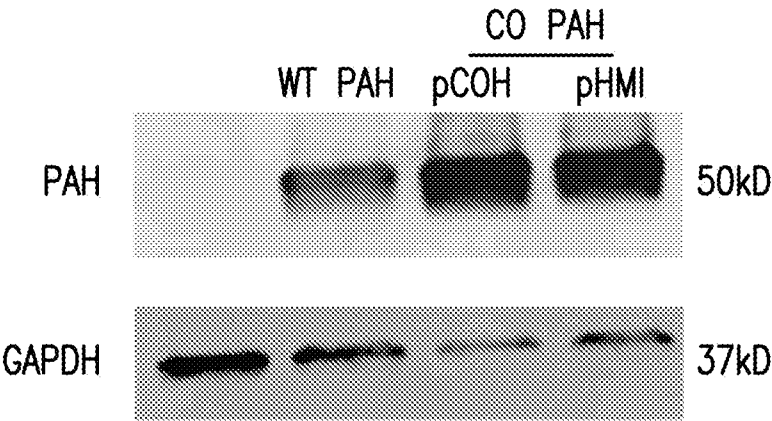


FIG.2

4/23

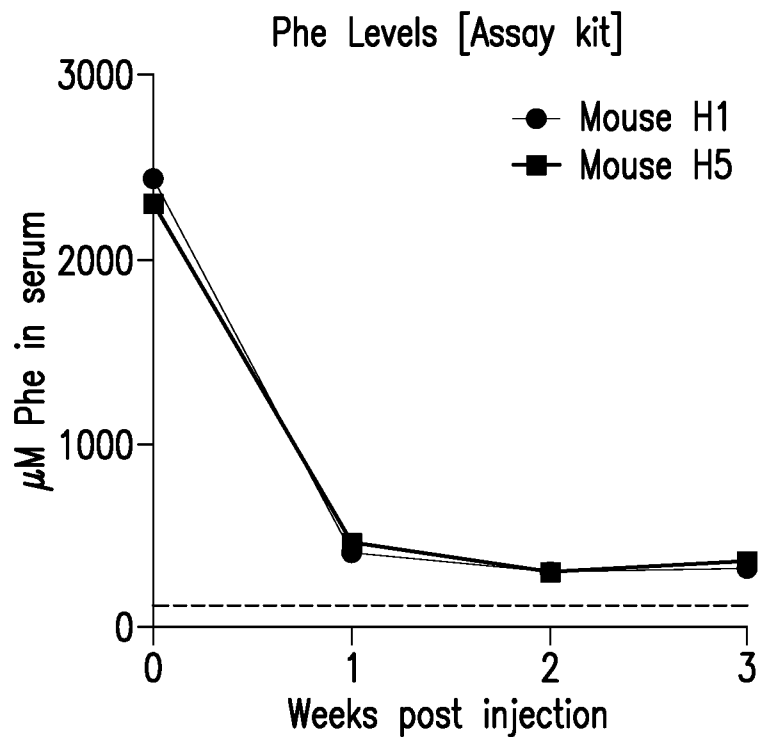


FIG.3A

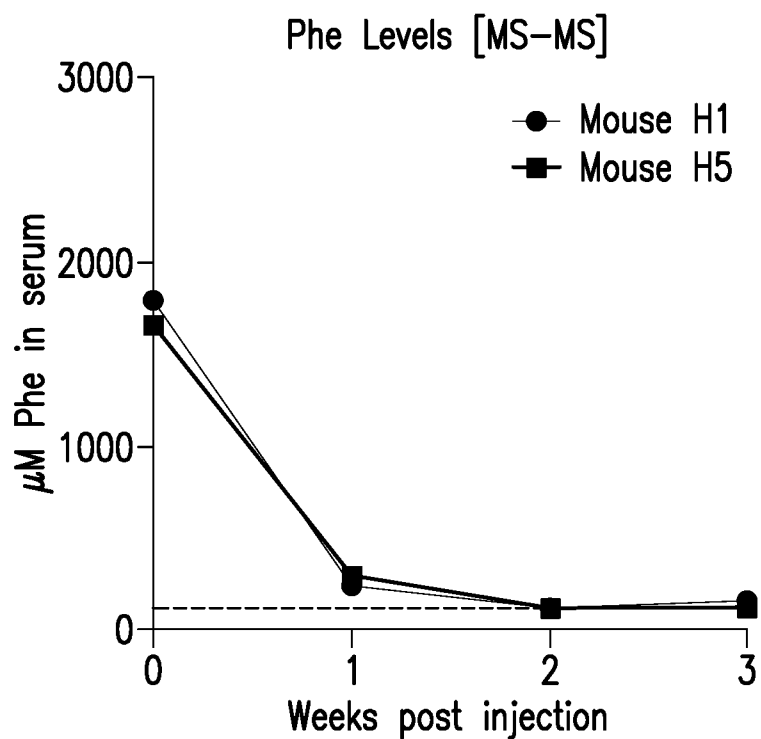


FIG.3B

5/23

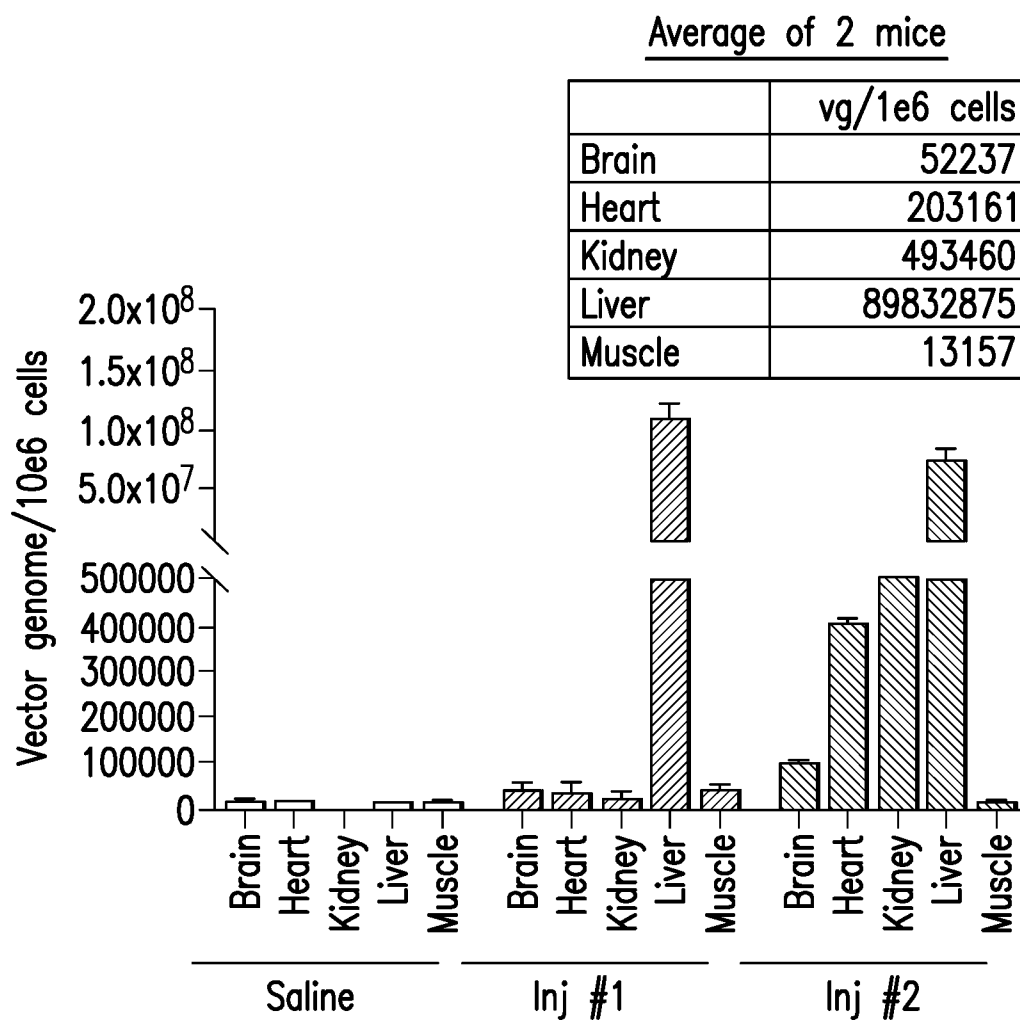


FIG.4

6/23

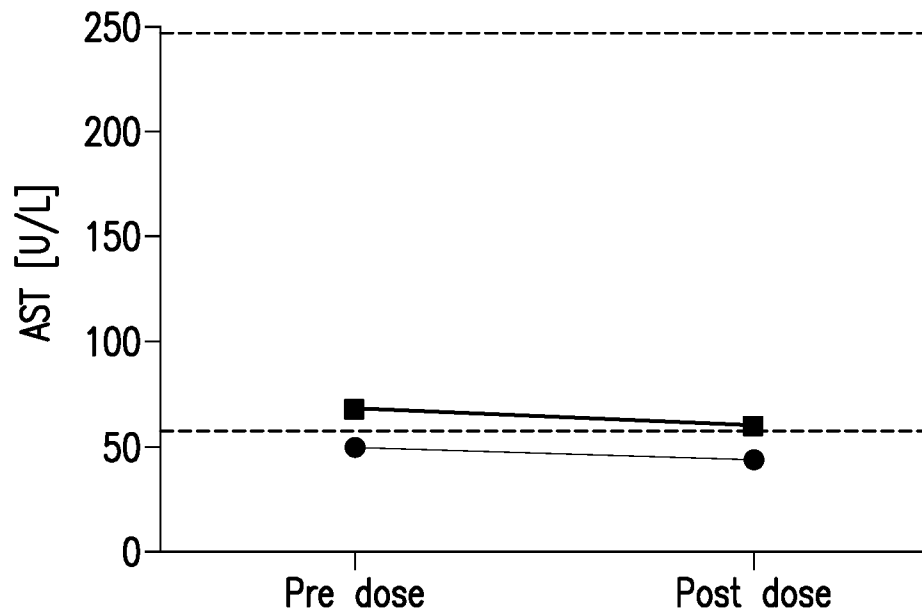


FIG.5A

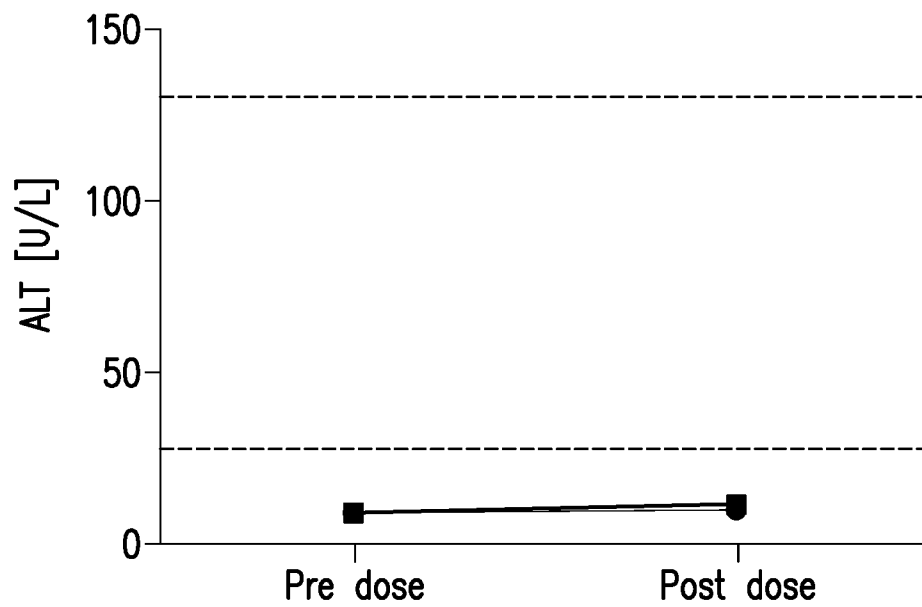


FIG.5B

7/23

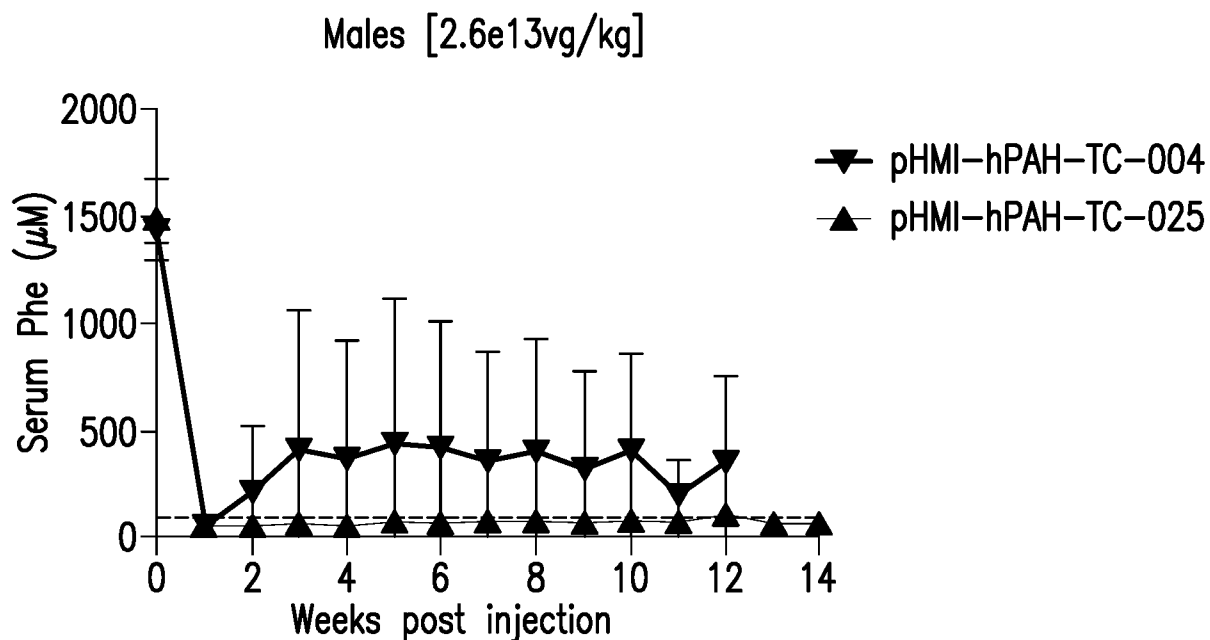


FIG.6A

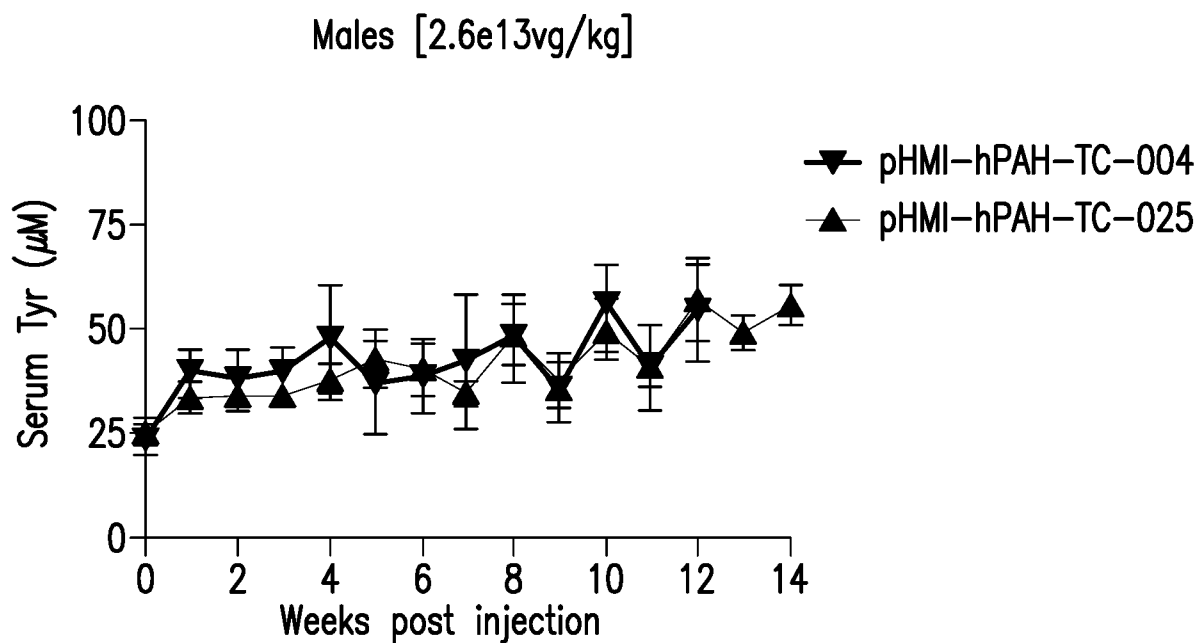


FIG.6B

8/23

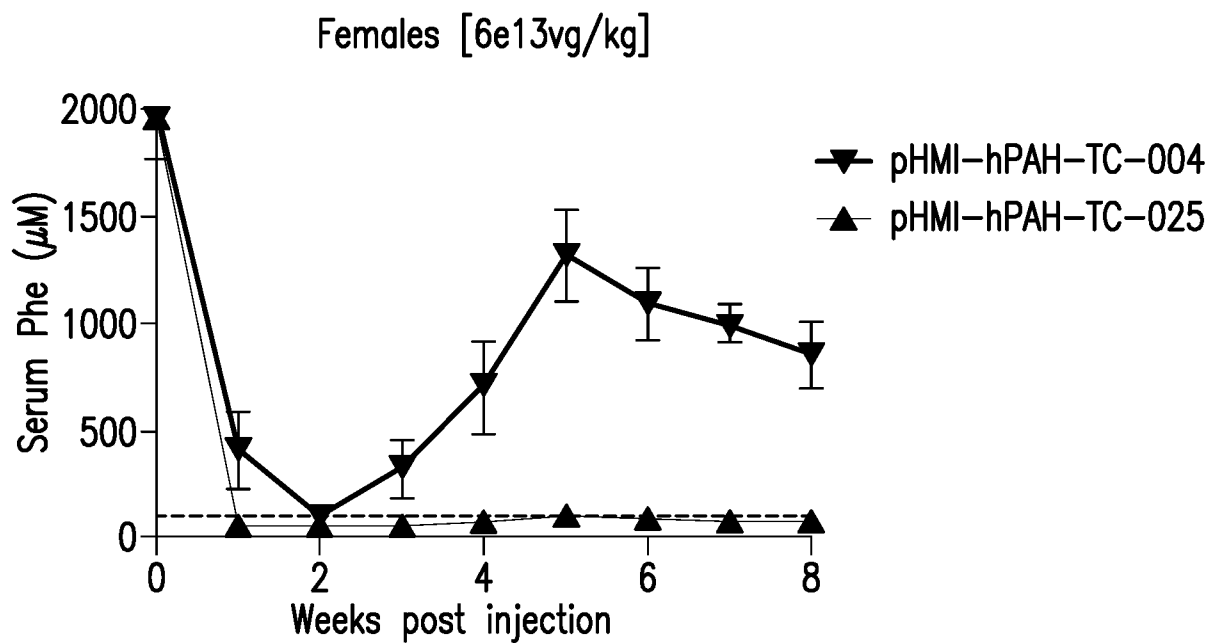


FIG.6C

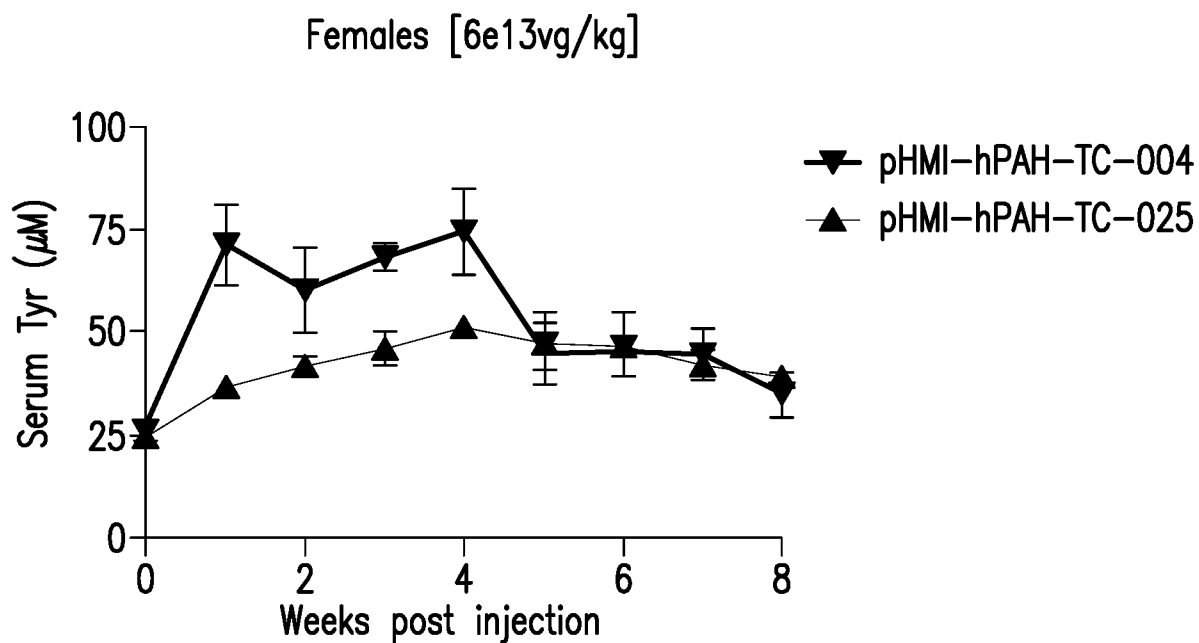


FIG.6D

9/23

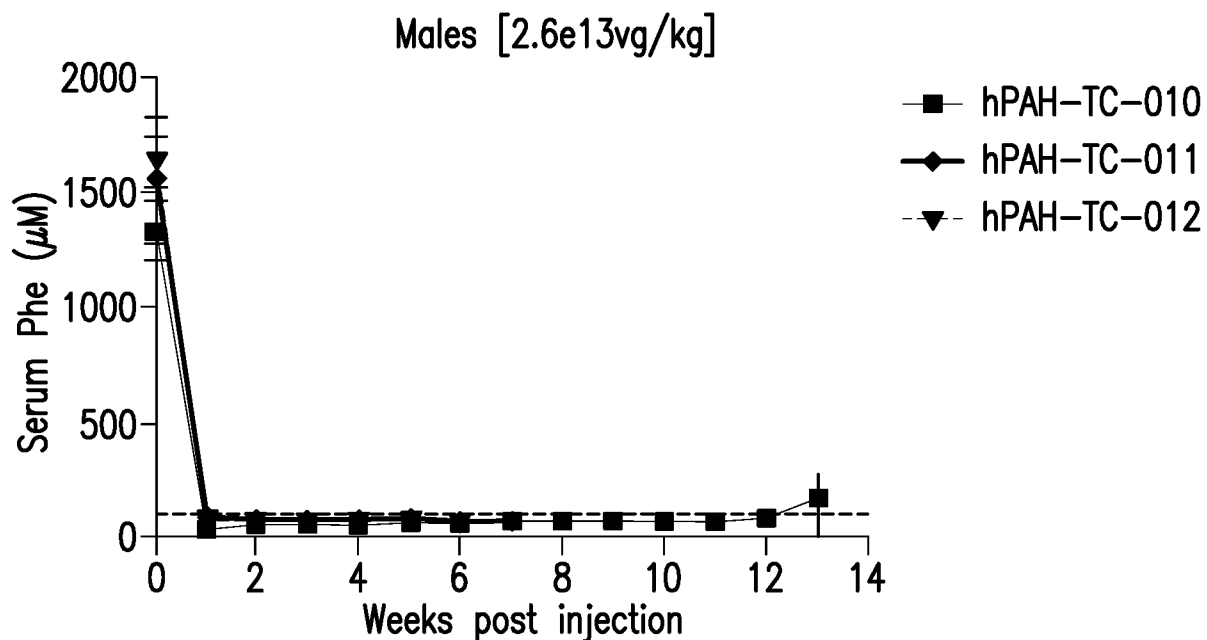


FIG.6E

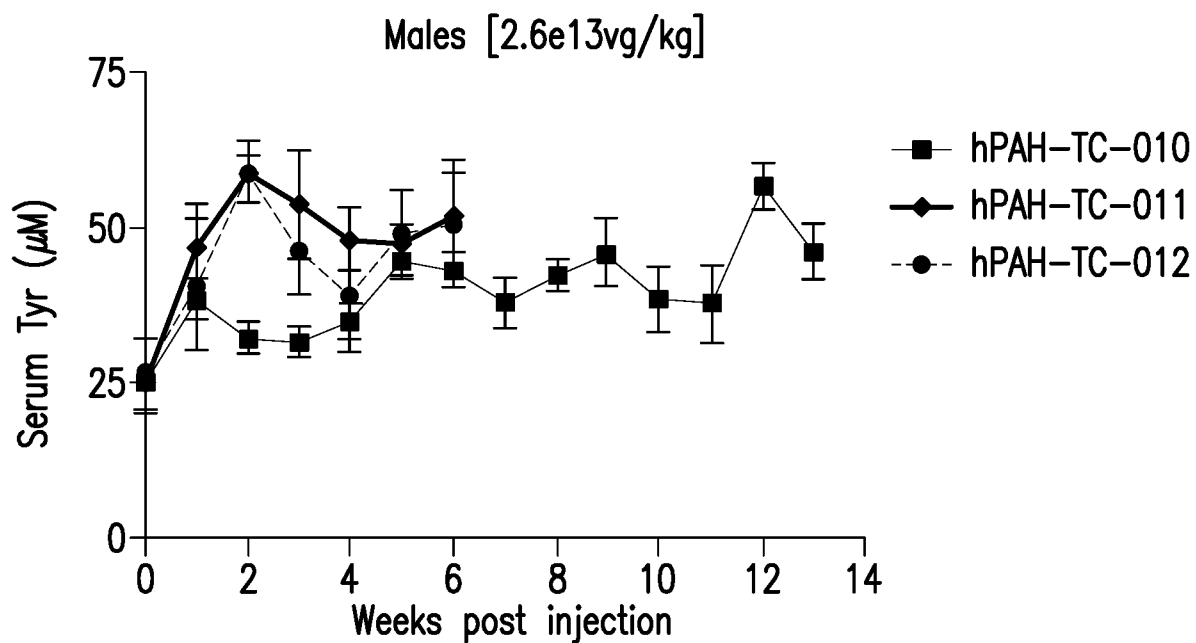


FIG.6F

10/23

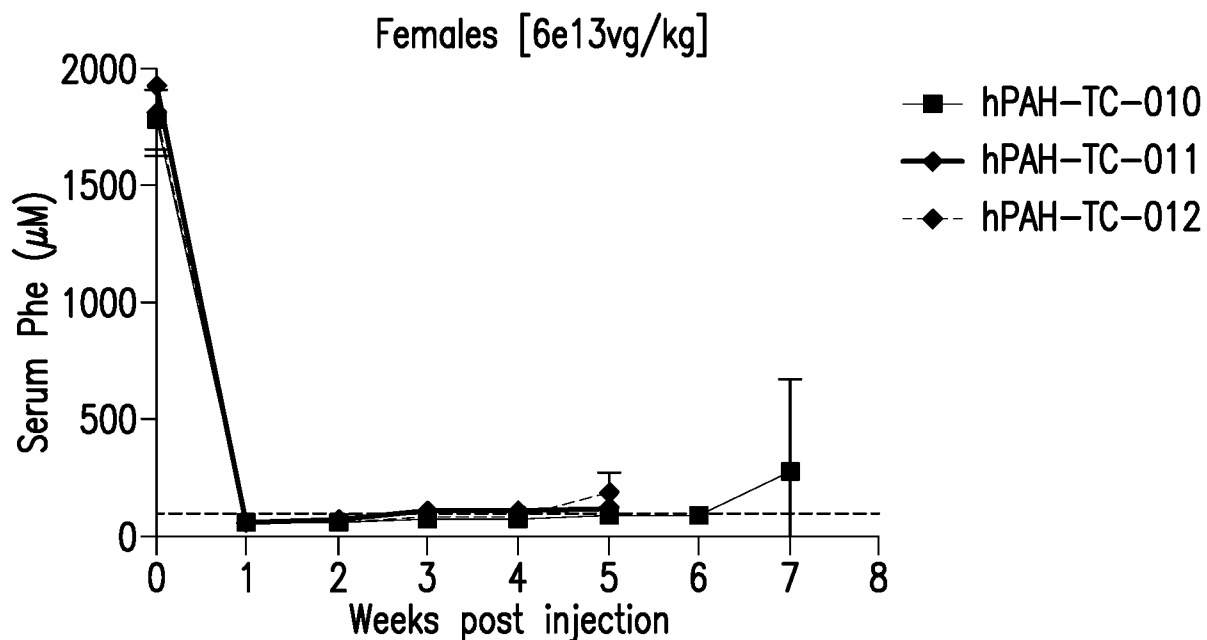


FIG.6G

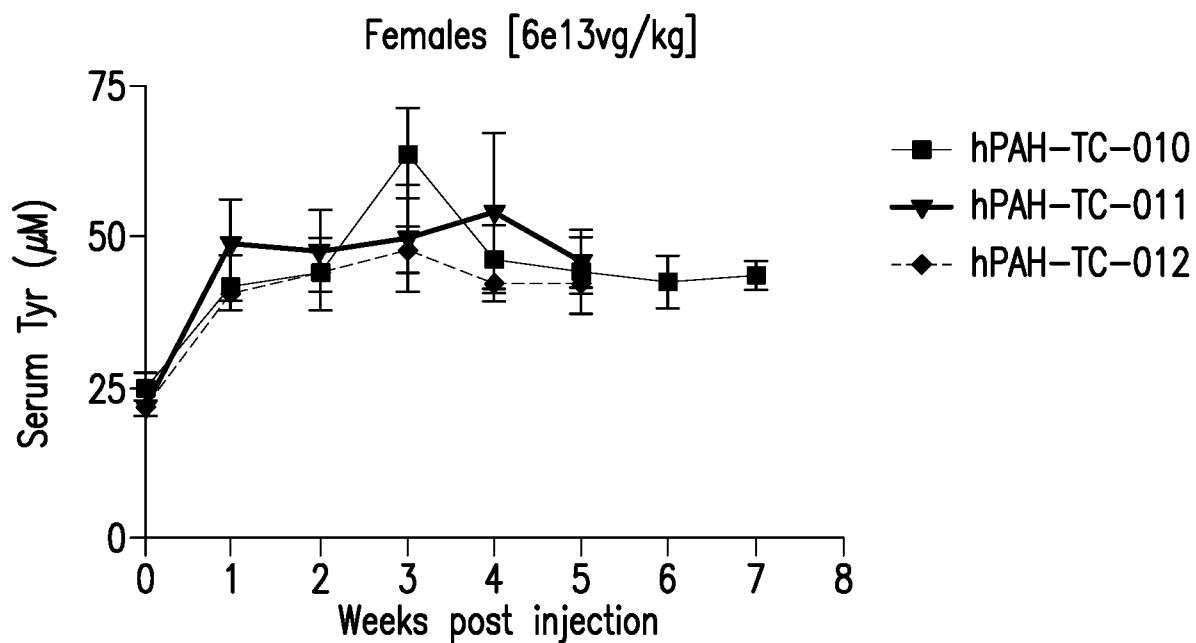


FIG.6H

11/23

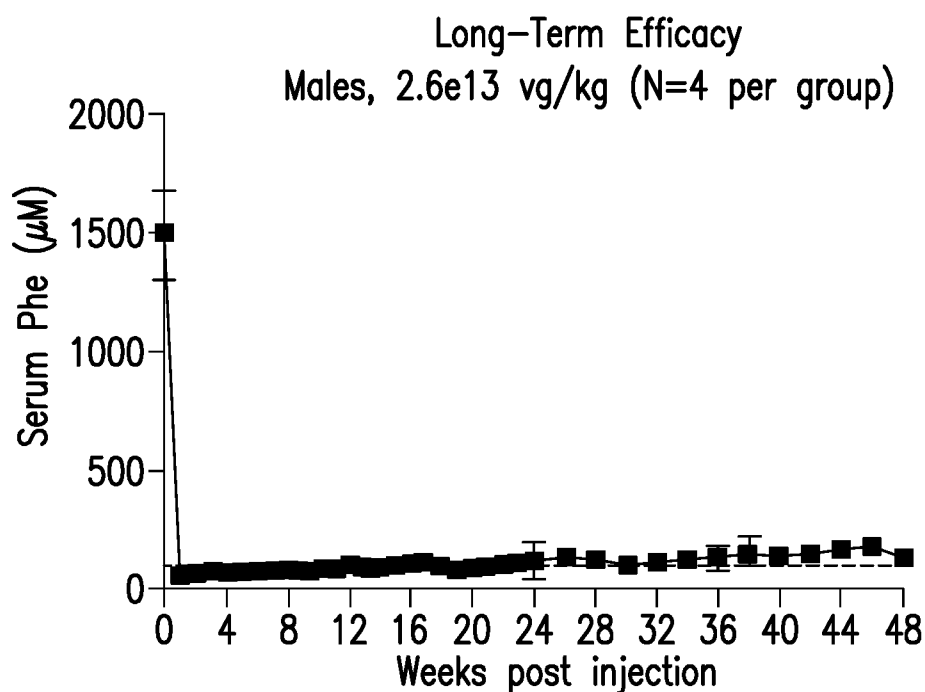


FIG.6I

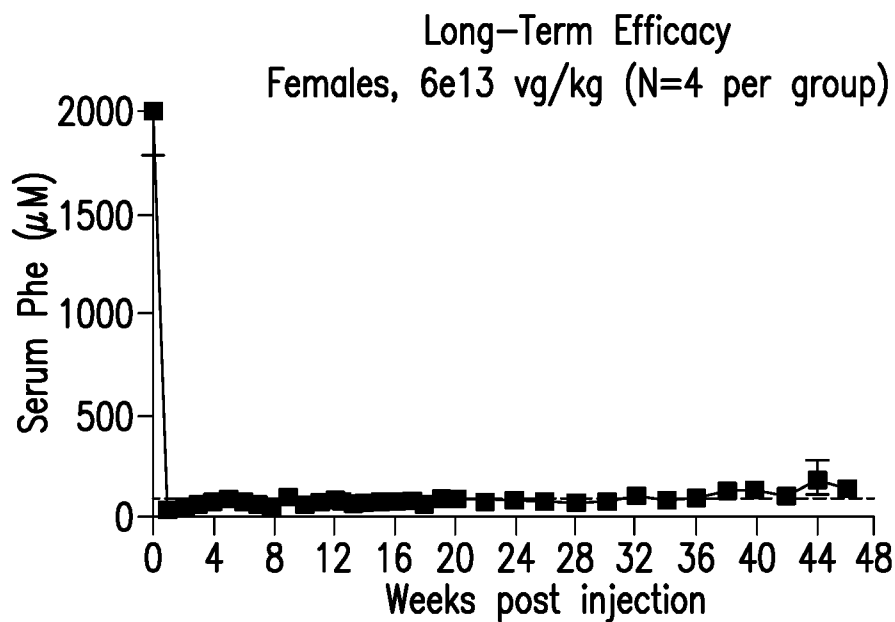
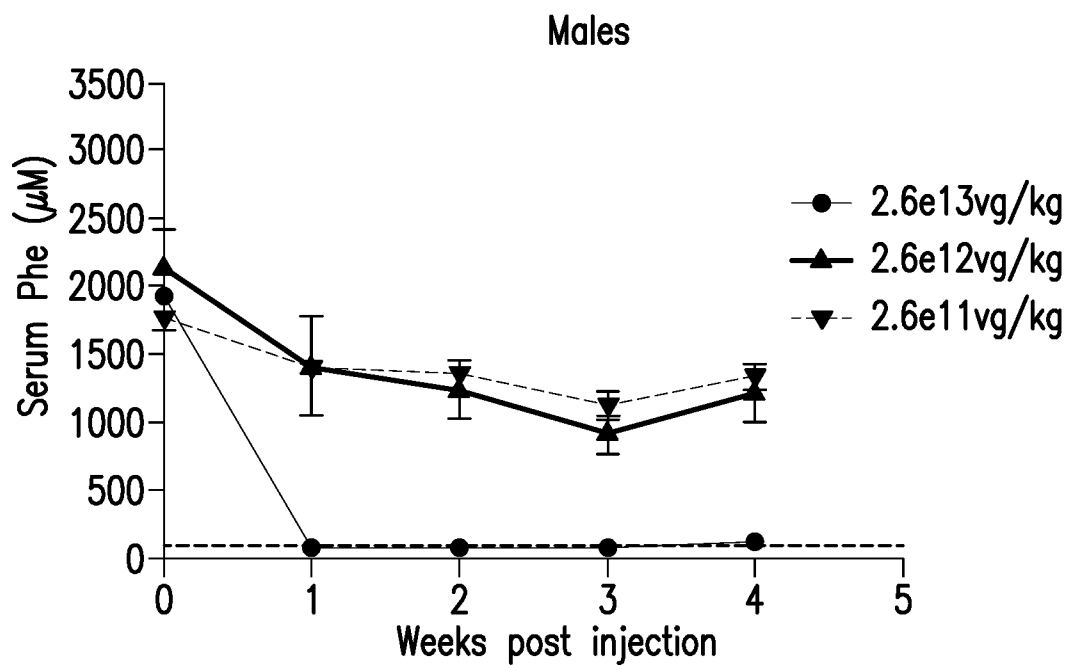
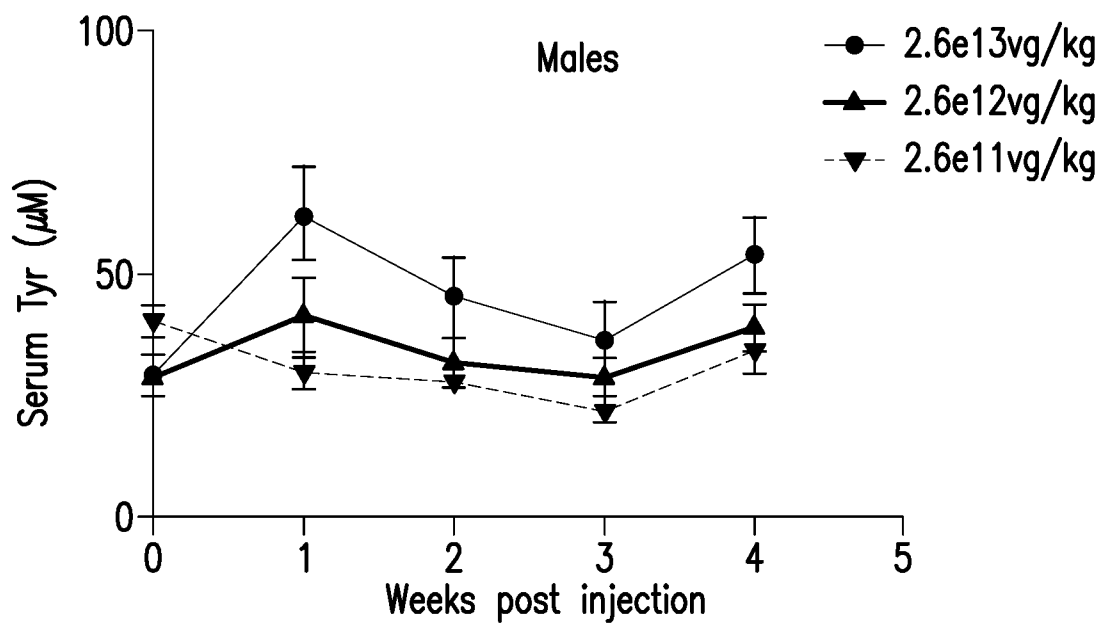


FIG.6J

12/23

**FIG.7A****FIG.7B**

13/23

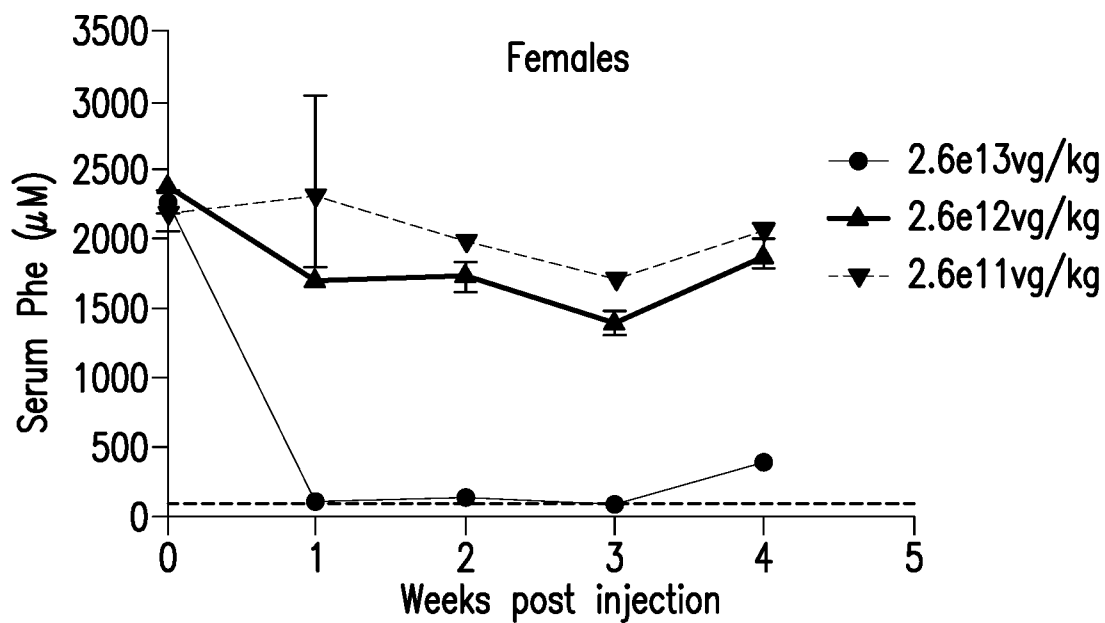


FIG.7C

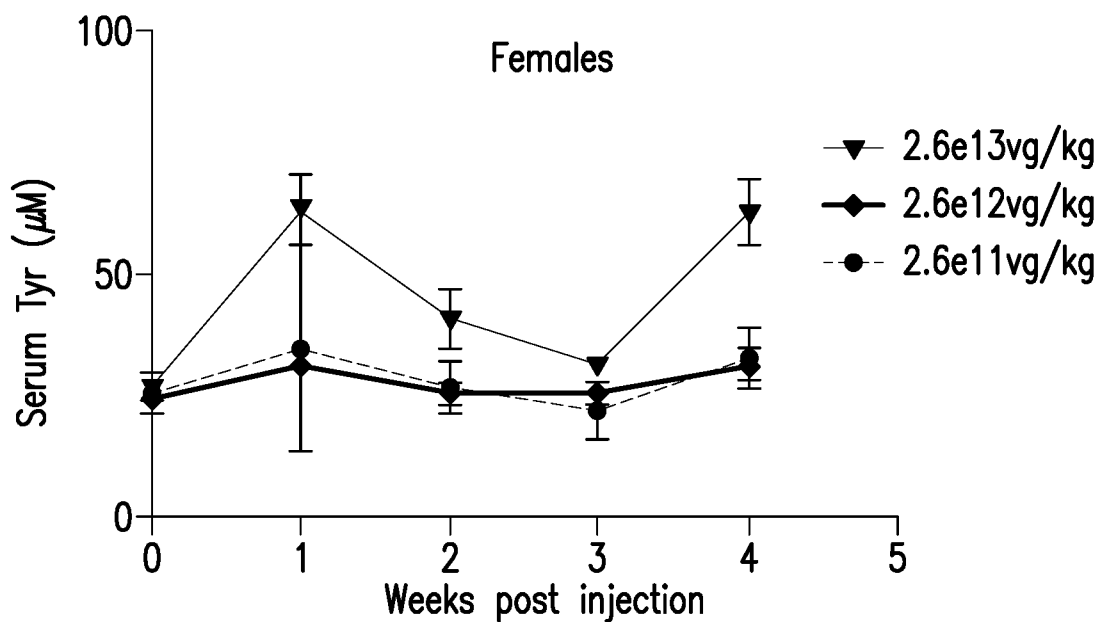


FIG.7D



FIG. 8A



FIG. 8B

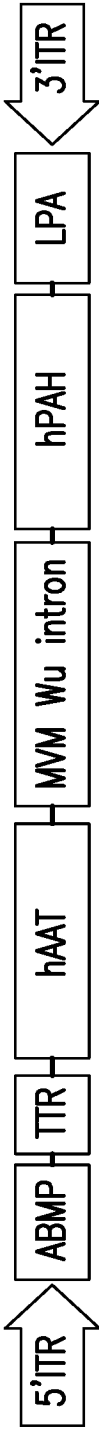


FIG. 8C

15/23

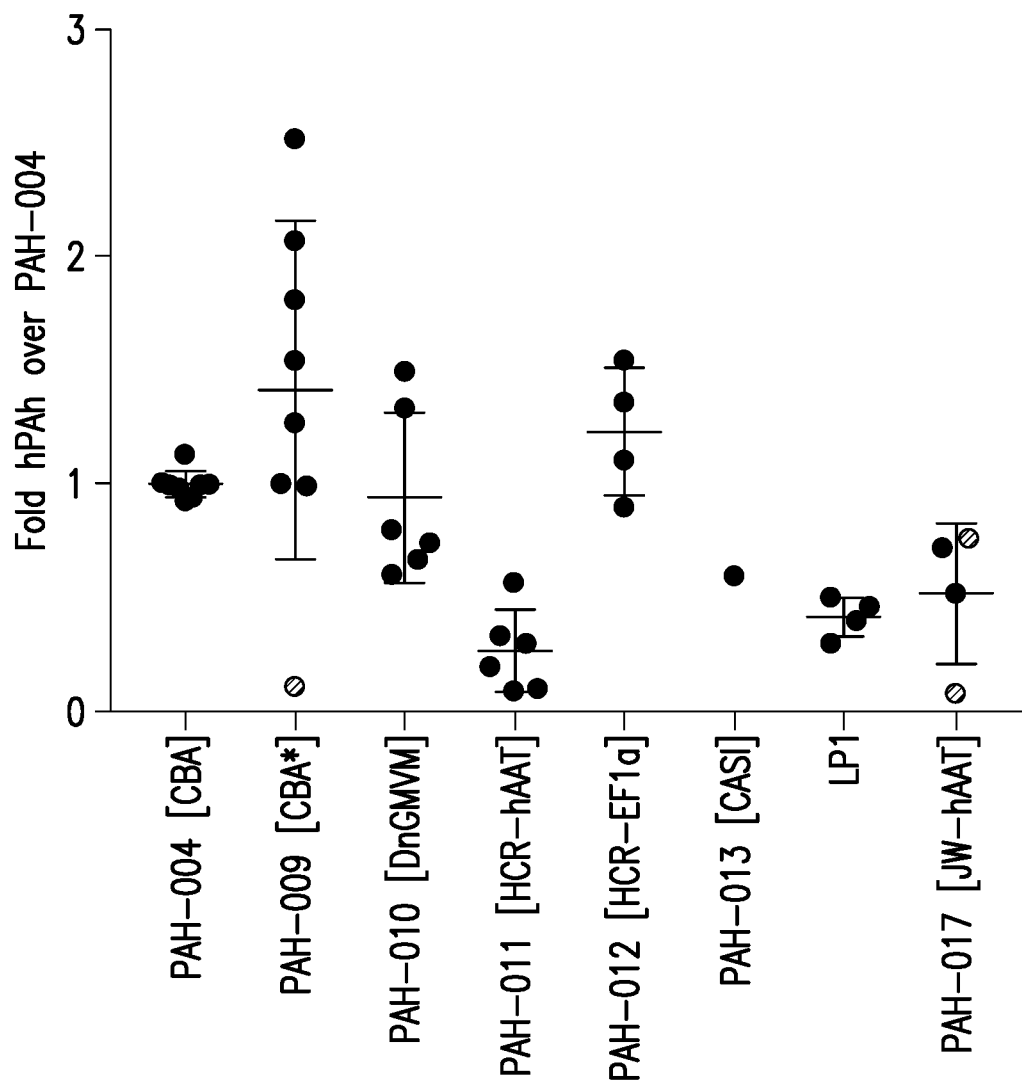


FIG.9A

16/23

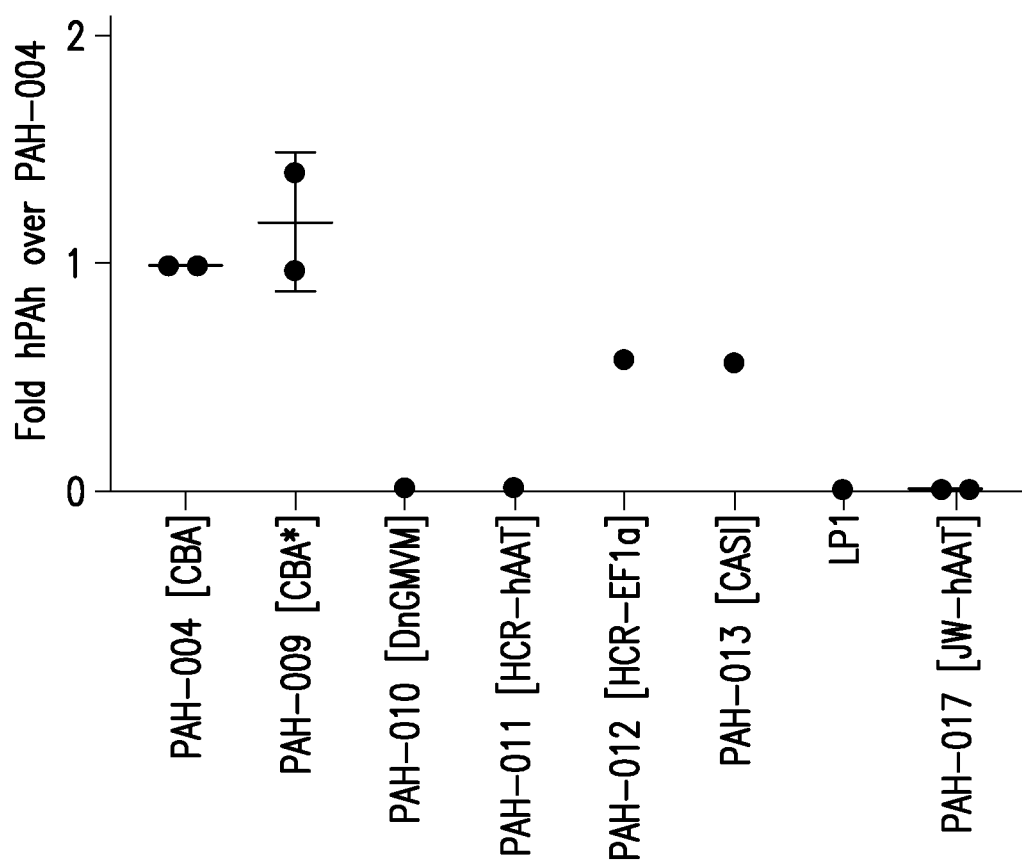


FIG.9B

17/23

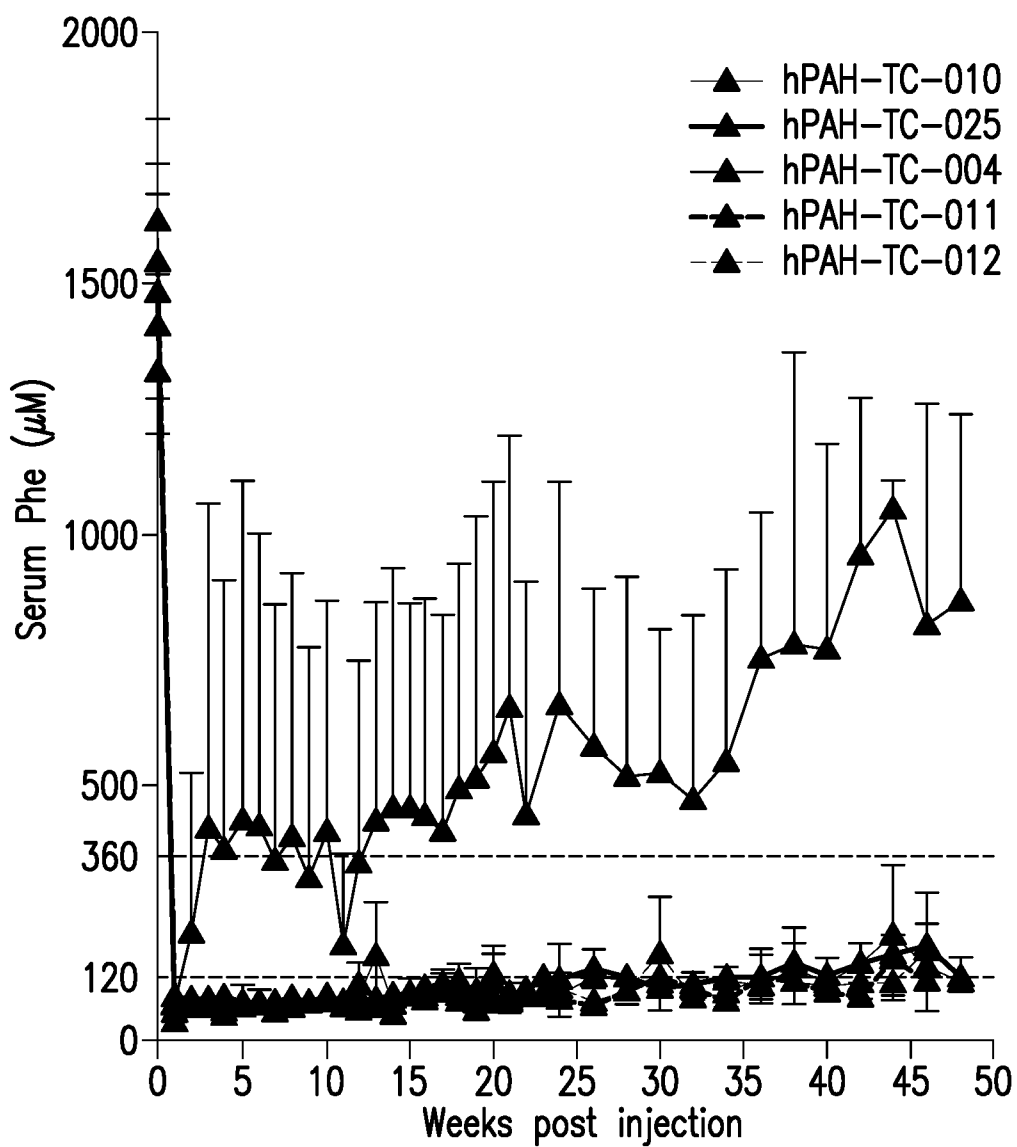


FIG.10A

18/23

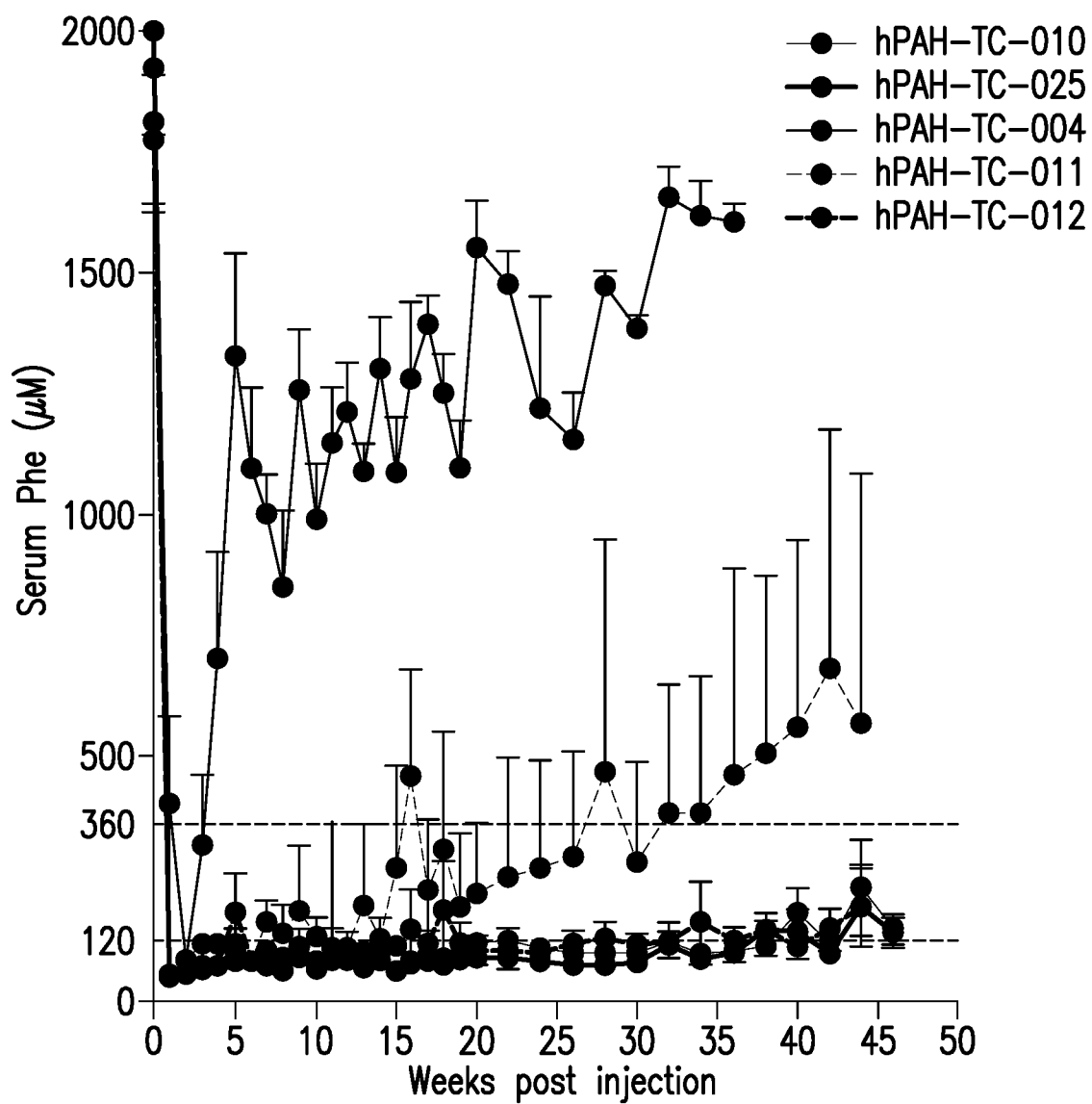


FIG.10B

19/23

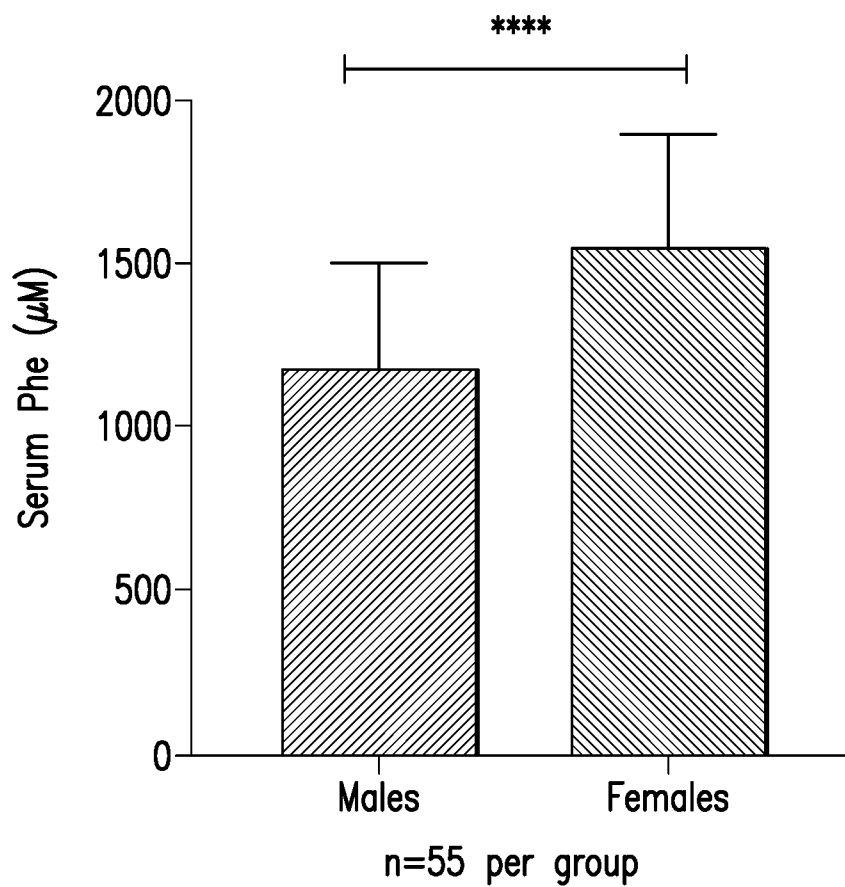


FIG.10C



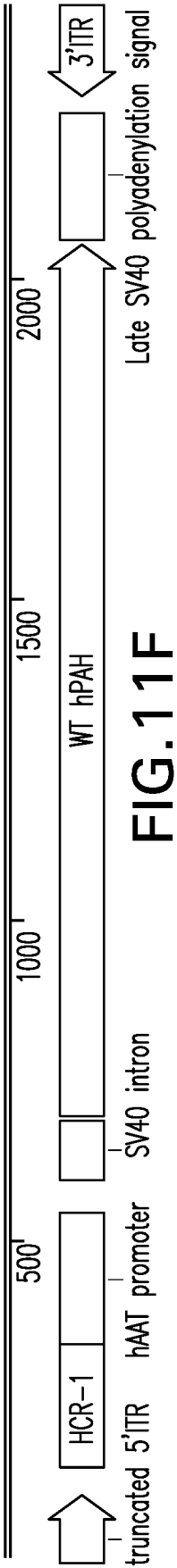
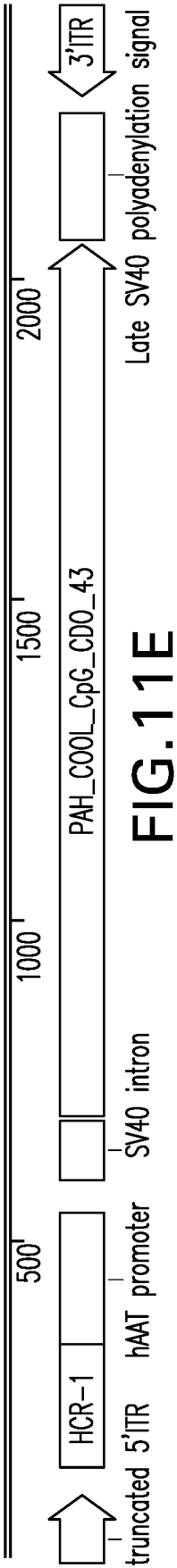
FIG. 11A



FIG. 11B



FIG. 11C



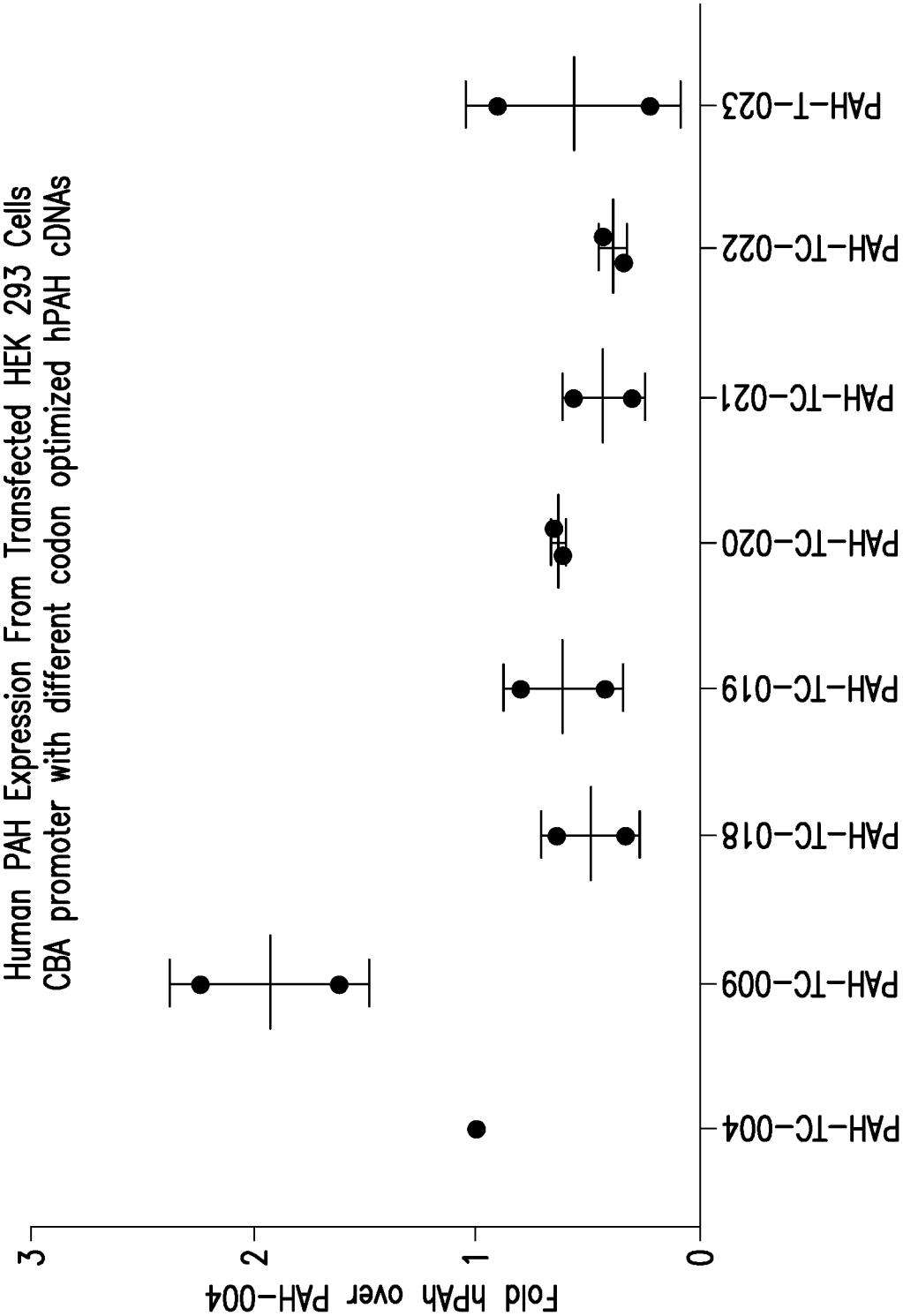


FIG.12

23/23

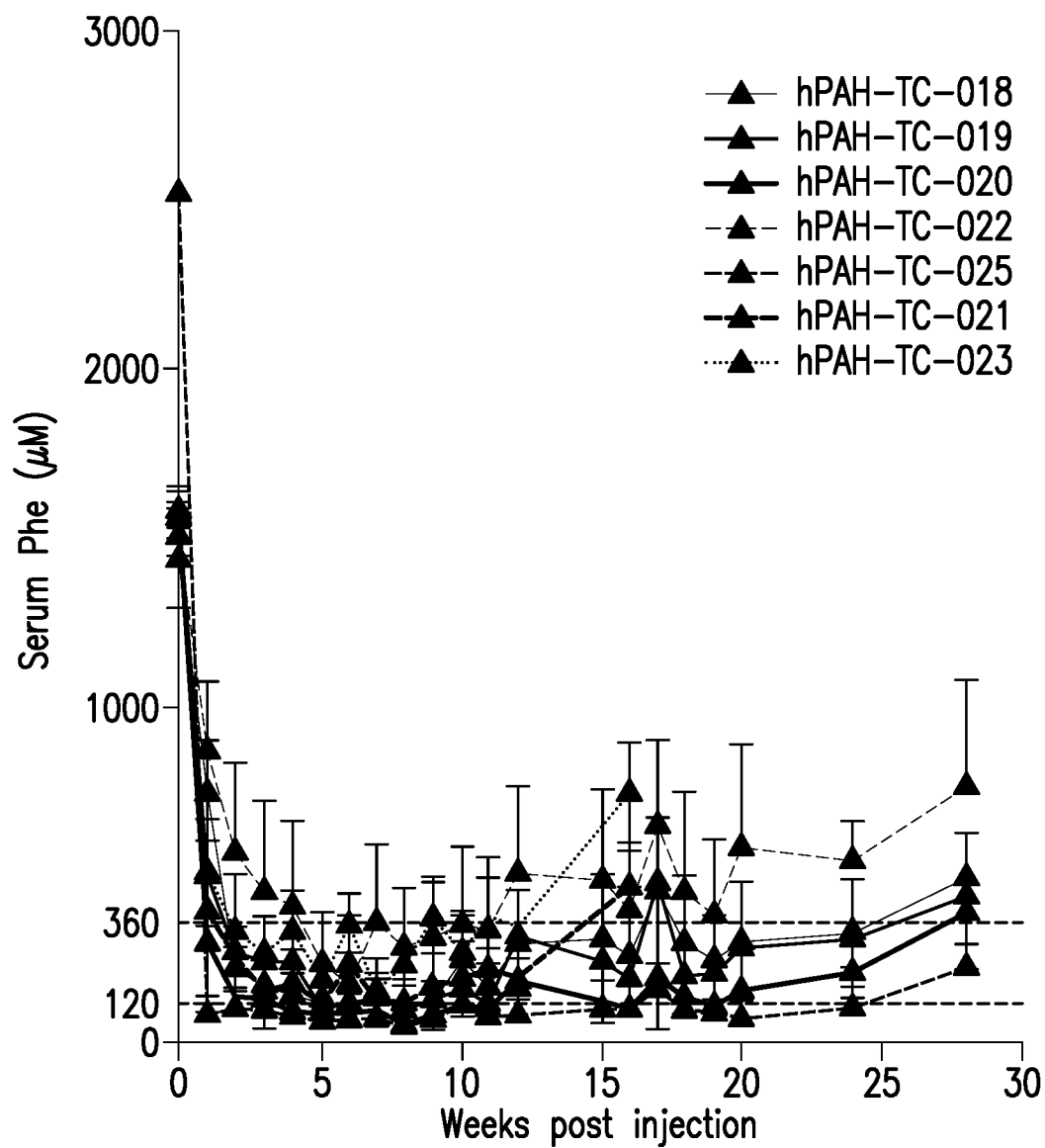


FIG.13

SEQUENCE LISTING

<110> HOMOLOGY MEDICINES, INC.

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

<130> 610061: HMT-025PC

<150> US 62/625,150

<151> 2018-02-01

<160> 92

<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				

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 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 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 ccgggcgacc aaaggtcgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgagc agagggagtg
120

g c c a a c t c c a
145

t c a c t a g g g g

t t c c t

<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 tgctcgcttc 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 tgtcgcgttc gctcgcctgc
60

tggctcgttt 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

accattttgg ataaacgtag cctgcctgct ctgacaaaca tcatcaagat cttgaggcat
300

gacattggtg ccactgtcca tgagctttca cgagataaga agaaagacac agtgccctgg
360

ttccaagaa 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 ggaagacgtt tctcaattcc tgcagacttg cactggtttc
720

cgctccgac ctgtggctgg cctgctttcc tctcgggatt tcttgggtgg cctggccttc
780

cgagtcttcc actgcacaca gtacatcaga catggatcca agcccatgta taccctcgaa
840

cctgacatct gccatgagct gttgggacat gtgcccttgt tttcagatcg cagctttgcc
900

cagttttccc aggaaattgg ccttgccctc ctgggtgcac ctgatgaata cattgaaaag
960

ctcgccacaa tttactgggt tactgtggag tttgggctct gcaaacaagg agactccata
1020

aaggcatatg gtgctgggct cctgtcatcc tttgggtgaat 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 aagatTTTtg 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

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 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 actgcacca gtacatcagg cacggctcca agccaatgta tacaccagag
840

cccgacatct gtcacgagct gctggggccac gtgcccctgt ttagcgatag atccttcgcc
900

cagttttccc aggagatcgg actggcatct ctgggagcac ctgacgagta catcgagaag
960

ctggccacca tctattgggt cacagtggag ttggcctgt 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 taccagagg
1260

atcgaggtgc 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 ccgggcgctc ggcgaccttt
60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtgg
106

<210> 27

<211> 143

<212> DNA

<213> Artificial Sequence

<220>

<223> modified AAV2 3'ITR

<400> 27

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 a g t g g c c
143

<210> 28

<211> 1873

<212> DNA

<213> Artificial Sequence

<220>

<223> CAG promoter

<400> 28

gatcttcaat attggccatt agccatatta ttcattgggtt atatagcata aatcaatatt
60

ggctattggc cattgcatac gttgtatcta tatcataata tgtacattta tattggctca
120

tgtccaatat gaccgccatg ttggcattga ttattgacta gttattaata gtaatcaatt
180

acgggggtcat tagttcatag cccatatatg gagttccgcg ttacataact tacggtaaatt
240

ggcccgccctg gctgaccgcc caacgacccc cgcccatatga cgtcaataat gacgtatgtt
300

cccatagtaa cgccaatagg gactttccat tgacgtcaat ggggtggagta ttacggtaa
360

actgcccact tggcagtaca tcaagtgtat catatgccaa gtccgcccccc tattgacgtc
420

aatgacggta aatggcccgc ctggcattat gcccagtaca tgaccttacg ggactttcct
480

acttggcagt acatctacgt attagtcatc gctattacca tggtcgaggt gagccccacg
540

ttctgcttca ctctcccat ctccccccc tccccacccc caattttgta tttatttatt
600

ttttaattat tttgtgcagc gatgggggcg gggggggggg gggggcgcgcg gccaggcggg
660

gcggggcggg gcgaggggcg gggcggggcg aggcggagag gtgcggcggc agccaatcag
720

agcggcgcgcg tccgaaagtt tccttttatg gcgaggcggc ggcggcgggcg gccctataaa
780

aagcgaagcg cgcgggcgggc gggagtcgct gcgacgctgc cttegccccg tgccccgctc
840

cgccgccgcc tcgcgccgcc cgccccggct ctgactgacc gcgttactcc cacagggtgag
900

cgggcgggac ggccctttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt
960

ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg
1020

ggagcggctc ggggggtgcg tgcgtgtgtg tgtgcgtggg gagcgccgcg tgcggcccg
1080

gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgcagtgt
1140

gcgcgagggg agcgcgggccg ggggcggtgc cccgcggtgc ggggggggct gcgaggggaa
1200

caaaggctgc gtgcgggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcggcggt
1260

cgggctgtaa cccccccctg cccccccctc cccgagttgc tgagcacggc ccggcttcgg
1320

gtgcgggggct ccgtacgggg cgtggcgcgg ggctcgccgt gccgggaggg gggtaggggc
1380

aggtgggggt gccgggaggg gcggggccgc ctcgggaggg ggagggctcg ggggaggggc
1440

gcggcggccc ccggagcgcc ggcggctgtc gaggcgcggc gagccgcagc cattgccttt
1500

tatggtaatc gtgcgagagg gcgcagggac ttcctttgtc ccaaattctgt gcggagccga
1560

aatctgggag gcgccgccgc accccctcta gcgggagcgg ggcgaagcgg tgcggcgccg
1620

gcaggaagga aatgggaggg gagggccttc gtgcgtcgcc gcgccgccgt ccccttctcc
1680

ctctccagcc tcggggctgt ccgcgggggg acggctgcct tcggggggga cggggcaggg
1740

cgggggttcgg cttctggcgt gtgaccggcg gctctagagc ctctgctaac catgttcatg
1800

ccttcttctt tttcctacag ctcctgggca acgtgctggt tattgtgctg tctcatcatt
1860

t t g g c a a a g a a t t
1873

<210> 29

<211> 192

<212> DNA

<213> Homo sapiens

<400> 29

ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc
60

tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc
120

cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt
180

a g t g t g a g a g g g
192

<210> 30

<211> 205

<212> DNA

<213> Homo sapiens

<400> 30

aatgactcct ttcggttaagt gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc
60

agcgtaggcg ggcgactcag atcccagcca gtggacttag cccctgtttg ctcctccgat
120

aactggggtg accttggtta atattcacca gcagcctccc ccgttgcccc tctggatcca
180

c t g c t t a a a t a c g g a c g a g g a c a g g
205

<210> 31
<211> 93
<212> DNA
<213> Simian virus 40

<400> 31
ctctaaggta aatataaaat ttttaagtgt ataatgtggt aaactactga ttctaattgt
60

t t c t c t c t t t t a g a t t c c a a c c t t t g g a a c t g a
93

<210> 32
<211> 398
<212> DNA
<213> Artificial Sequence

<220>
<223> pHMI-hPAH-TC-025 transcriptional regulatory region

<400> 32
ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc
60

tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc
120

cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt
180

agtgtgagag gggaatgact cctttcggtta agtgcagtgg aagctgtaca ctgcccaggc
240

aaagcgtccg ggcagcgtag gcgggcgact cagatcccag ccagtggact tagcccctgt
300

ttgctcctcc gataactggg gtgaccttgg ttaatatcca ccagcagcct cccccgttgc
360

ccctctggat ccactgctta aatacggacg aggacagg
398

<210>	33
<211>	72
<212>	DNA
<213>	Homo sapiens

```
<400> 33
gggggaggct gctggtgaat attaaccaag gtcaccccag ttatcggagg agcaaacagg
60
```

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

<210>	34
<211>	170
<212>	DNA
<213>	Homo sapiens

```
<400> 34
cgatgctcta atctctctag acaagggtca tatttgatg ggttacttat tctctctttg
60
```

ttgactaagt caataatcag aatcagcagg tttgcagtca gattggcagg gataagcagc
120

ctagctcagg agaagtgagt ataaaagccc caggctggga gcagccatca
170

<210>	35
<211>	92
<212>	DNA
<213>	Artificial Sequence

<220>
<223> modified MVM intron

```
<400> 35
aagaggtaag ggtttaaggg atggttggtt ggtgggggat taatgtttaa ttacctggag
60
```


ggatgggatt cagactaggg ccaggaccca gggatggaga gaaagagatg agagtgggtt
 240

gggggcttgg tgacttagag aacagagctg caggctcaga ggcacacagg agtttctggg
 300

ctcaccctgc ccccttccaa cccctcagtt cccatcctcc agcagctgtt tgtgtgctgc
 360

ctctgaagtc cacactgaac aaacttcagc ctactcatgt ccctaaaatg ggcaaacatt
 420

gcaagcagca aacagcaaac acacagccct ccctgcctgc tgaccttgga gctggggcag
 480

aggtcagaga cctctctggg cccatgccac ctccaacatc cactcgaccc cttggaattt
 540

cggtggagag gagcagaggt tgtcctggcg tggtttaggt agtgtgagag gg
 592

<210> 38
 <211> 423
 <212> DNA
 <213> Homo sapiens

<400> 38
 gctctaaccc actctgatct cccagggcgg cagtaagtct tcagcatcag gcattttggg
 60

gtgactcagt aaatggtaga tcttgctacc agtggaacag ccactaagga ttctgcagtg
 120

agagcagagg gccagctaag tgggtactctc ccagagactg tctgactcac gccaccccct
 180

ccaccttgga cacaggacgc tgtgggtttct gagccaggta caatgactcc tttcggttaag
 240

tgcagtggaa gctgtacact gcccaggcaa agcgtccggg cagcgtaggc gggcgactca
 300

gatcccagcc agtggactta gcccctgttt gctcctccga taactgggggt gaccttggtt
 360

aatattcacc agcagcctcc cccgttgccc ctctggatcc actgcttaaa tacggacgag
420

g a c
423

<210> 39

<211> 1021

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-011 transcriptional regulatory region

<400> 39

gtaaatttta tggaatgtga atcataattc aatTTTTcaa catgcgttag gagggacatt
60

tcaaactctt ttttacccta gactttccta ccatcaccca gagtatccag ccaggagggg
120

aggggctaga gacaccagaa gtttagcagg gagggaggcg tagggattcg gggaatgaag
180

ggatgggatt cagactaggg ccaggaccca gggatggaga gaaagagatg agagtggttt
240

gggggcttgg tgacttagag aacagagctg caggctcaga ggcacacagg agtttctggg
300

ctcacctgc ccccttccaa cccctcagtt cccatcctcc agcagctgtt tgtgtgctgc
360

ctctgaagtc cacactgaac aaacttcagc ctactcatgt ccctaaaatg ggcaaacatt
420

gcaagcagca aacagcaaac acacagccct ccctgcctgc tgaccttgga gctggggcag
480

aggtcagaga cctctctggg cccatgccac ctccaacatc cactcgaccc cttggaattt
540

cggtggagag gagcagaggt tgtcctggcg tggtttaggt agtgtgagag ggcttaaggc
600

tctaaccac tctgatctcc cagggcggca gtaagtcttc agcatcaggc attttggggt
660

gactcagtaa atggtagatc ttgctaccag tggaacagcc actaaggatt ctgcagtgag
720

agcagagggc cagctaagtg gtactctccc agagactgtc tgactcacgc caccctcc
780

accttggaca caggacgctg tggtttctga gccagggtaca atgactcctt tcggtaagtg
840

cagtggaagc tgtacactgc ccaggcaaag cgtccgggca gcgtaggcgg gcgactcaga
900

tcccagccag tggacttagc ccctgtttgc tcctccgata actgggggtga ccttggttaa
960

tattcaccag cagcctcccc cgttgcccct ctggatccac tgcttaaata cggacgagga
1020

c
1021

<210> 40

<211> 1168

<212> DNA

<213> Artificial Sequence

<220>

<223> modified human EF-1alpha promoter

<400> 40

cgtgaggctc cggtgcccgt cagtgggcag agcgcacatc gcccacagtc cccgagaagt
60

tgggggggagg ggtcggcaat tgaaccggtg cctagagaag gtggcgcggg gtaaactggg
120

aaagtgatgt cgtgtactgg ctccgccttt ttcccgaggg tgggggagaa ccgtatataa
180

gtgcagtagt cgccgtgaac gttctttttc gcaacggggt tgccgccaga acacaggtaa
240

gtgccgtgtg tggttcccgc gggcctggcc tctttacggg ttatggccct tgcgtgcctt
300

gaattacttc cacctggctc cagtacgtga ttcttgatcc cgagctggag ccaggggcgg
360

gccttgcgct ttaggagccc cttcgccctcg tgcttgagtt gaggcctggc ctgggcgctg
420

gggccgccgc gtgcgaatct ggtggcacct tcgcgccctgt ctcgctgctt tcgataagtc
480

tctagccatt taaaattttt gatgacctgc tgcgacgctt tttttctggc aagatagtct
540

tgtaaatacg ggccaggatc tgcacactgg tatttcggtt tttggggccg cgggcggcga
600

cggggcccgt gcgtcccagc gcacatgttc ggcgaggcgg ggcctgcgag cgcggccacc
660

gagaatcgga cgggggtagt ctcaagctgg ccggcctgct ctggtgcctg gcctcgcgcc
720

gccgtgtatc gccccgccct gggcggcaag gctggcccgg tcggcaccag ttgcgtgagc
780

ggaaagatgg ccgcttcccg gccctgctcc aggggggctca aaatggagga cgcggcgctc
840

gggagagcgg gcgggtgagt caccacaca aaggaaaggg gcctttccgt cctcagccgt
900

cgcttcatgt gactccacgg agtaccgggc gccgtccagg cacctcgatt agttctggag
960

cttttggagt acgtcgtctt taggttgggg ggaggggttt tatgcgatgg agtttcccca
1020

cactgagtgg gtggagactg aagttaggcc agcttggcac ttgatgtaat tctccttgga
1080

at ttgccctt tttgagtttg gatcttggtt cattctcaag cctcagacag tggttcaaag
1140

t t t t t t t c t t
1168

c c a t t t c a g g

t g t c g t g a

<210> 41

<211> 1766

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-012 transcriptional regulatory region

<400> 41

gtaaatttta tggaatgtga atcataattc aatttttcaa catgcttag gagggacatt
60

tcaaactctt ttttacccta gactttccta ccatcaccca gagtatccag ccaggagggg
120

aggggctaga gacaccagaa gtttagcagg gaggagggcg tagggattcg gggaatgaag
180

ggatgggatt cagactaggg ccaggaccca gggatggaga gaaagagatg agagtggttt
240

gggggcttgg tgacttagag aacagagctg caggctcaga ggcacacagg agtttctggg
300

ctcacctgc ccccttccaa cccctcagtt cccatcctcc agcagctggt tgtgtgctgc
360

ctctgaagtc cacactgaac aaacttcagc ctactcatgt ccctaaaatg ggcaaacatt
420

gcaagcagca aacagcaaac acacagccct ccctgcctgc tgaccttgga gctggggcag
480

aggtcagaga cctctctggg cccatgccac ctccaacatc cactcgaccc cttggaattt
540

cggtggagag gagcagaggt tgtcctggcg tggtttaggt agtgtgagag ggcttaagcg
600

tgaggctccg gtgcccgtca gtgggcagag cgcacatcgc ccacagtccc cgagaagttg
660

gggggagggg tcggcaattg aaccggtgcc tagagaaggt ggcgcggggg aaactgggaa
720

agtgatgtcg tgtactggct ccgccttttt cccgaggggtg ggggagaacc gtatataagt
780

gcagtagtcg ccgtgaacgt tcttttttcgc aacgggtttg ccgccagAAC acaggtaagt
840

gccgtgtgtg gttcccgcgg gcctggcctc ttacgggtt atggcccttg cgtgccttga
900

attacttcca cctggctcca gtacgtgatt cttgatcccg agctggagcc aggggcgggc
960

cttgcgcttt aggagcccct tcgcctcgtg cttgagttga ggcctggcct gggcgctggg
1020

gccgccgcgt gcgaatctgg tggcaccttc gcgcctgtct cgctgctttc gataagtctc
1080

tagccattta aaatttttga tgacctgctg cgacgctttt tttctggcaa gatagtcttg
1140

taaatgcggg ccaggatctg cacactggta tttcggtttt tggggccgcg ggcggcgacg
1200

gggcccgtgc gtcccagcgc acatgttcgg cgaggcgggg cctgcgagcg cggccaccga
1260

gaatcggacg ggggtagtct caagctggcc ggcctgctct ggtgcctggc ctcgcgccgc
1320

cgtgtatcgc cccgccctgg gcggcaaggc tggcccggtc ggcaccagtt gcgtgagcgg
1380

aaagatggcc gcttcccggc cctgctccag ggggctcaaa atggaggacg cggcgctcgg
1440

gagagcgggc gggtgagtca cccacacaaa ggaaaggggc ctttccgtcc tcagccgtcg
1500

cttcatgtga ctccacggag taccgggcgc cgtccaggca cctcgattag ttctggagct
1560

tttggagtac gtcgtcttta ggttgggggg aggggtttta tgcgatggag tttccccaca
1620

ctgagtgggt ggagactgaa gttaggccag cttggcactt gatgtaattc tccttggaat
1680

ttgccctttt tgagtttgga tcttggttca ttctcaagcc tcagacagtg gttcaaagtt
1740

t t t t t c t t c c a t t t c a g g t g t c g t g a
1766

<210> 42
<211> 122
<212> DNA
<213> Simian virus 40

<400> 42
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca
60

aataaagcat ttttttcact gcattctagt tgtggtttgt ccaaactcat caatgtatct
120

t a
122

<210> 43
<211> 133
<212> DNA
<213> Simian virus 40

<400> 43
tgctttatatt 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> 44

<400> 44

000

<210> 45

<211> 198

<212> DNA

<213> Simian virus 40

<400> 45

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
198

t t t t t t a a

<210> 46

<211> 3392

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-004 transfer genome

<400> 46

gatcttcaat attggccatt agccatatta ttcattgggtt atatagcata aatcaatatt
60

ggctattggc cattgcatac gttgtatcta tatcataata tgtacattta tattggctca
120

tgtccaatat gaccgccatg ttggcattga ttattgacta gttattaata gtaatcaatt
180

acgggggtcat tagttcatag cccatatatg gagttccgcg ttacataact tacggtaaatt
240

ggcccgccctg gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatggt
300

cccatagtaa cgccaatagg gactttccat tgacgtcaat ggggtggagta ttacggtaa
360

actgcccact tggcagtaca tcaagtgtat catatgccaa gtccgcccc tattgacgtc
420

aatgacggta aatggcccg cttggcattat gcccagtaca tgaccttacg ggactttcct
480

acttggcagt acatctacgt attagtcac gctattacca tggtcgaggt gagccccacg
540

ttctgcttca ctctcccat ccccccccc tccccacccc caattttgta ttattttatt
600

ttttaattat tttgtgcagc gatgggggcg gggggggggg gggggcgcg gccaggcggg
660

gcggggcggg gcgaggggcg gggcggggcg aggcggagag gtgcggcggc agccaatcag
720

agcggcgcg tccgaaagtt tccttttatg gcgagggcggc ggcggcgggc gccctataaa
780

aagcgaagcg cgcggcgggc gggagtcgct gcgacgctgc cttcgcccc tgccccgctc
840

cgcgcgcgcc tcgcgcgcgcc cgcggcggt ctgactgacc gcgttactcc cacagggtgag
900

cgggcgggac ggcccttctc ctccgggctg taattagcg ttggtttaat gacggcttgt
960

ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg
1020

ggagcggctc ggggggtgcg tgcgtgtgtg tgtgcgtggg gagcgccgcg tgcggcccg
1080

gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgcagtgt
1140

gcgcgagggg agcgcggccg ggggcggtgc cccgcggtgc ggggggggct gcgaggggaa
1200

caaaggctgc gtgcgggggtg tgtgcgtggg ggggtgagca ggggggtgtgg gcgcggcggt
1260

cgggctgtaa cccccccctg cccccccctc cccgagttgc tgagcacggc ccggcttcgg
1320

gtgcgggggct ccgtacgggg cgtggcgcgg ggctcgccgt gccgggaggg gggtaggggc
1380

aggtgggggt gccgggaggg gcggggccgc ctcgggaggg ggagggtcgc ggggaggggc
1440

gcggcgggcc ccggagcgcc ggcggctgtc gaggcgcggc gagccgcagc cattgccttt
1500

tatggtaatc gtgcgagagg gcgcagggac ttcctttgtc ccaaattctgt gcggagccga
1560

aatctgggag gcgccgccgc accccctcta gcgggagcgg ggcaagcgg tcgggagccg
1620

gcaggaagga aatgggaggg gagggccttc gtgcgtcgcc gcgccgccgt ccccttctcc
1680

ctctccagcc tcgggggtgt ccgcgggggg acggctgcct tcggggggga cggggcaggg
1740

cgggggttcgg cttctggcgt gtgaccggcg gctctagagc ctctgctaac catgttcatt
1800

ccttcttctt tttcctacag ctccctggga acgtgctggt tattgtgctg tctcatcatt
1860

ttggcaaaga attccgccac catgtccacc gctgtgctgg agaaccctgg gctggggagg
1920

aaactgtcag acttcgggca ggagacttca tacattgagg ataactgtaa ccagaatggc
1980

gccatctctc tgatcttcag cctgaaggag gaagtgggag ccctggcaaa ggtgctgcgc
2040

ctgtttgagg 2100	agaacgacgt	gaatctgacc	cacatcgagt	cccggccttc	tagactgaag
aaggacgagt 2160	acgagttctt	taccacacctg	gataagcggg	ccctgccagc	cctgacaaac
atcatcaaga 2220	tcctgaggca	cgacatcgga	gcaaccgtgc	acgagctgtc	tcgggacaag
aagaaggata 2280	ccgtgccctg	gttccctcgg	acaatccagg	agctggatag	atttgccaac
cagatcctgt 2340	cttacggagc	agagctggac	gcagatcacc	ctggcttcaa	ggacccagtg
tatcgggccc 2400	ggagaaagca	gtttgccgat	atcgcctaca	attataggca	cggacagcca
atccctcgcg 2460	tggagtatat	ggaggaggag	aagaagacct	ggggcacagt	gttcaagacc
ctgaagagcc 2520	tgtacaagac	acacgcctgc	tacgagtata	accacatctt	ccccctgctg
gagaagtatt 2580	gtggctttca	cgaggacaat	atccctcagc	tggaggacgt	gagccagttc
ctgcagacct 2640	gcacaggctt	taggctgagg	ccagtggcag	gactgctgag	ctcccgggac
ttcctgggag 2700	gactggcctt	cagagtgttt	cactgcaccc	agtacatcag	gcacggctcc
aagccaatgt 2760	atacaccaga	gcccgacatc	tgtcacgagc	tgctgggcca	cgtgcccttg
tttagcgata 2820	gatccttcgc	ccagttttcc	caggagatcg	gactggcatc	tctgggagca
cctgacgagt 2880	acatcgagaa	gctggccacc	atctattggg	tcacagtgga	gtttggcctg
tgcaagcagg 2940	gcgatagcat	caaggcctac	ggagcaggac	tgctgtctag	cttcggcgag

ctgcagtatt gtctgtccga gaagccaaag ctgctgcccc tggagctgga gaagaccgcc
3000

atccagaact acaccgtgac agagttccag cccctgtact atgtggccga gtcttttaac
3060

gatgccaagg agaaggtgag aaatttcgcc gccacaatcc ctaggccctt cagcgtgcgg
3120

tacgaccctt ataccagag gatcgagggtg ctggataata cacagcagct gaagatcctg
3180

gctgactcaa tcaatagcga aatcggaatc ctgtgctccg ccctgcagaa aatcaaatga
3240

atcgattcta gagtcgagcc gcggactagt aacttgttta ttgcagctta taatggttac
3300

aaataaagca atagcatcac aaatttcaca aataaagcat ttttttcact gcattctagt
3360

t g t g g t t t g t c c a a a c t c a t c a a t g t a t c t t a
3392

<210> 47

<211> 2042

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-025 transfer genome

<400> 47

ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc
60

tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc
120

cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt
180

agtgtgagag gggaatgact cctttcggta agtgcagtgg aagctgtaca ctgcccaggc
240

aaagcgtccg ggcagcgtag gcgggcgact cagatcccag ccagtggact tagccccgtg
300

ttgctcctcc gataactggg gtgaccttgg ttaatatcca ccagcagcct cccccgttgc
360

ccctctggat ccactgctta aatacggacg aggacagggc cctgtctcct cagcttcagg
420

caccaccact gacctgggac agtgaatcct ctaaggtaaa tataaaattt ttaagtgtat
480

aatgtgttaa actactgatt ctaattgttt ctctctttta gattccaacc tttggaactg
540

accgccacca tgtccaccgc tgtgctggag aaccctgggc tggggaggaa actgtcagac
600

ttcgggcagg agacttcata cattgaggat aactgtaacc agaatggcgc catctctctg
660

atcttcagcc tgaaggagga agtgggcgcc ctggcaaagg tgctgcgctt gtttgaggag
720

aacgacgtga atctgaccca catcgagtcc cggccttcta gactgaagaa ggacgagtac
780

gagttcttta cccacctgga taagcgggtcc ctgccagccc tgacaaacat catcaagatc
840

ctgaggcacg acatcggagc aaccgtgcac gagctgtctc gggacaagaa gaaggatacc
900

gtgccctggg tccctcggac aatccaggag ctggatagat ttgccaacca gatcctgtct
960

tacggagcag agctggacgc agatcacctt ggcttcaagg acccagtgtg tcgggcccgg
1020

agaaagcagt ttgccgatat cgcctacaat tataggcacg gacagccaat ccctcgcgtg
1080

gagtatatgg aggaggagaa gaagacctgg ggcacagtgt tcaagaccct gaagagcctg
1140

tacaagacac acgcctgcta cgagtataac cacatcttcc ccctgctgga gaagtattgt
1200

ggctttcacg aggacaatat ccctcagctg gaggacgtga gccagttcct gcagacctgc
1260

acaggcttta ggctgaggcc agtggcagga ctgctgagct cccgggactt cctgggagga
1320

ctggccttca gagtgtttca ctgcacccag tacatcaggc acgggtccaa gccaatgtat
1380

acaccagagc ccgacatctg tcacgagctg ctggggccacg tgcccctgtt tagcgataga
1440

tccttcgccc agttttccca ggagatcgga ctggcatctc tgggagcacc tgacgagtac
1500

atcgagaagc tggccaccat ctattgggtc acagtggagt ttggcctgtg caagcagggc
1560

gatagcatca aggcctacgg agcaggactg ctgtctagct tcggcgagct gcagtattgt
1620

ctgtccgaga agccaaagct gctgcccctg gagctggaga agaccgccat ccagaactac
1680

accgtgacag agttccagcc cctgtactat gtggccgagt cttttaacga tgccaaggag
1740

aaggtgagaa atttcgccgc cacaatccct aggcccttca gcgtgcggta cgacccttat
1800

accagagga tcgaggtgct ggataataca cagcagctga agatcctggc tgactcaatc
1860

aatagcgaaa tcggaatcct gtgctccgcc ctgcagaaaa tcaaatgaat gctttatttg
1920

tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata aacaagttaa
1980

caacaacaat tgcattcatt ttatgtttca ggttcagggg gaggtgtggg aggtttttta
2040

a
2042

a

<210> 48
<211> 1927
<212> DNA
<213> Artificial Sequence

<220>
<223> pHMI-hPAH-TC-010 transfer genome

<400> 48
gggggaggct gctggtgaat attaaccaag gtcaccccag ttatcggagg agcaaacagg
60

ggctaagtcc acctcgagcc atggcgatgc tctaattctct ctagacaagg ttcataatttg
120

tatggggttac ttattctctc tttgttgact aagtcaataa tcagaatcag caggtttgca
180

gtcagattgg cagggataag cagcctagct caggagaagt gagtataaaa gccccaggct
240

gggagcagcc atcagctagc gccggcaaga ggtaagggtt taagggatgg ttggttggtg
300

gggtattaat gtttaattac ctggagcacc tgcctgaaat cacttttttt caggttgggc
360

caccatgtcc accgctgtgc tggagaaccc tgggctgggg aggaaactgt cagacttcgg
420

gcaggagact tcatacattg aggataactg taaccagaat ggcgccatct ctctgatctt
480

cagcctgaag gaggaagtgg gcgccctggc aaagggtgctg cgcctgtttg aggagaacga
540

cgtgaatctg acccacatcg agtcccggcc ttctagactg aagaaggacg agtacgagtt
600

ctttaccac ctggataagc ggtccctgcc agccctgaca aacatcatca agatcctgag
660

gcacgacatc ggagcaaccg tgcacgagct gtctcgggac aagaagaagg ataccgtgcc
720

ctggttccct cggacaatcc aggagctgga tagatttgcc aaccagatcc tgtcttacgg
780

agcagagctg gacgcagatc accctggcct caaggaccca gtgtatcggg cccggagaaa
840

gcagtttgcc gatatcgcct acaattatag gcacggacag ccaatccctc gcgtggagta
900

tatggaggag gagaagaaga cctggggcac agtgttcaag accctgaaga gcctgtacaa
960

gacacacgcc tgctacgagt ataaccacat cttccccctg ctggagaagt attgtggcctt
1020

tcacgaggac aatatccctc agctggagga cgtgagccag ttcctgcaga cctgcacagg
1080

ctttaggctg aggccagtgg caggactgct gagctcccgg gacttcctgg gaggactggc
1140

cttcagagtg tttcactgca cccagtacat caggcacggc tccaagccaa tgtatacacc
1200

agagcccgac atctgtcacg agctgctggg ccacgtgccc ctgttttagcg atagatcctt
1260

cgcccagttt tcccaggaga tcggactggc atctctggga gcacctgacg agtacatcga
1320

gaagctggcc accatctatt ggttcacagt ggagtttgcc ctgtgcaagc agggcgatag
1380

catcaaggcc tacggagcag gactgctgtc tagcttcggc gagctgcagt attgtctgtc
1440

cgagaagcca aagctgctgc ccctggagct ggagaagacc gccatccaga actacaccgt
1500

gacagagttc cagcccctgt actatgtggc cgagtctttt aacgatgcca aggagaaggt
1560

gagaaatttc gccgccacaa tccctaggcc cttcagcgtg cggtacgacc cttataccca
1620

gaggatcgag gtgctggata atacacagca gctgaagatc ctggctgact caatcaatag
1680

cgaaatcgga atcctgtgct ccgccctgca gaaaatcaaa tgaatcgtag atccagacat
1740

gataagatac attgatgagt ttggacaaac cacaactaga atgcagtgaa aaaaatgctt
1800

tattttgtgaa atttgtgatg ctattgcttt atttgtaacc attataagct gcaataaaca
1860

agttaacaac aacaattgca ttcattttat gtttcagggt cagggggagg tgtgggaggt
1920

t t t t t a a
1927

<210> 49

<211> 2694

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-011

<400> 49

gtaaatttta tggaatgtga atcataattc aatttttcaa catgcgttag gagggacatt
60

tcaaactctt ttttacccta gactttccta ccatcaccca gagtatccag ccaggagggg
120

aggggctaga gacaccagaa gtttagcagg gaggagggcg tagggattcg gggaatgaag
180

ggatgggatt cagactaggg ccaggaccca gggatggaga gaaagagatg agagtggttt
240

gggggcttgg tgacttagag aacagagctg caggctcaga ggcacacagg agtttctggg
300

ctcacctgc ccccttccaa cccctcagtt cccatcctcc agcagctggt tgtgtgctgc
360

ctctgaagtc cacactgaac aaacttcagc ctactcatgt ccctaaaatg ggcaaacatt
420

gcaagcagca aacagcaaac acacagccct ccctgcctgc tgaccttgga gctggggcag
480

aggtcagaga cctctctggg cccatgccac ctccaacatc cactcgaccc cttggaattt
540

cggtggagag gagcagaggt tgtcctggcg tggtttaggt agtgtgagag ggcttaaggc
600

tctaaccac tctgatctcc cagggcggca gtaagtcttc agcatcaggc attttggggt
660

gactcagtaa atggtagatc ttgctaccag tggaacagcc actaaggatt ctgcagtgag
720

agcagagggc cagctaagtg gtactctccc agagactgtc tgactcacgc caccctcc
780

accttggaac caggacgctg tggtttctga gccagggtaca atgactcctt tcggtaagtg
840

cagtggaagc tgtacactgc ccaggcaaag cgtccgggca gcgtaggcgg gcgactcaga
900

tcccagccag tggacttagc ccctgtttgc tcctccgata actgggggtga ccttggttaa
960

tattcaccag cagcctcccc cgttgcccct ctggatccac tgcttaaata cggacgagga
1020

cgctagcgcc ggcaagaggt aagggttttaa gggatggttg gttggtgggg tattaatgtt
1080

taattacctg gagcacctgc ctgaaatcac tttttttcag gttgggccac catgtccacc
1140

gctgtgctgg agaaccctgg gctggggagg aaactgtcag acttcgggca ggagacttca
1200

tacattgagg ataactgtaa ccagaatggc gccatctctc tgatcttcag cctgaaggag
1260

gaagtgggcg ccctggcaaa ggtgctgcgc ctgtttgagg agaacgacgt gaatctgacc
1320

cacatcgagt cccggccttc tagactgaag aaggacgagt acgagttctt taccacactg
1380

gataagcggc ccctgccagc cctgacaaac atcatcaaga tcctgaggca cgacatcgga
1440

gcaaccgtgc acgagctgtc tcgggacaag aagaaggata ccgtgccctg gttccctcgg
1500

acaatccagg agctggatag atttgccaac cagatcctgt cttacggagc agagctggac
1560

gcagatcacc ctggcttcaa ggacccagtg tatcggggcc ggagaaagca gtttgccgat
1620

atgcctaca attataggca cggacagcca atccctcgcg tggagtatat ggaggaggag
1680

aagaagacct ggggcacagt gttcaagacc ctgaagagcc tgtacaagac acacgcctgc
1740

tacgagtata accacatctt cccctgctg gagaagtatt gtggctttca cgaggacaat
1800

atccctcagc tggaggacgt gagccagttc ctgcagacct gcacaggctt taggctgagg
1860

ccagtggcag gactgctgag ctcccgggac ttcctgggag gactggcctt cagagtgttt
1920

cactgcaccc agtacatcag gcacggctcc aagccaatgt atacaccaga gcccgacatc
1980

tgtcacgagc tgctgggcca cgtgccctg tttagcgata gatccttcgc ccagtgtttc
2040

caggagatcg gactggcatc tctgggagca cctgacgagt acatcgagaa gctggccacc
2100

atctattggt tcacagtgga gtttggcctg tgcaagcagg gcgatagcat caaggcctac
2160

ggagcaggac tgctgtctag cttcggcgag ctgcagtatt gtctgtccga gaagccaaag
2220

ctgctgcccc tggagctgga gaagaccgcc atccagaact acaccgtgac agagttccag
2280

cccctgtact atgtggccga gtctttttaac gatgccaagg agaagggtgag aaatttcgcc
2340

gccacaatcc ctaggccctt cagcgtgcgg tacgaccctt atacccagag gatcgagggtg
2400

ctggataata cacagcagct gaagatcctg gctgactcaa tcaatagcga aatcggaatc
2460

ctgtgctccg ccctgcagaa aatcaaata gaatcgtatgc cagacatgat aagatacatt
2520

gatgagtttg gacaaaccac aactagaatg cagtgaacaaa aatgctttat ttgtgaaatt
2580

tgtgatgcta ttgctttatt tgtaaccatt ataagctgca ataaacaagt taacaacaac
2640

aattgcattc attttatggt tcagggttcag ggggaggtgt gggaggtttt ttaa
2694

<210> 50

<211> 3335

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-012

<400> 50

gtaaatttta tggaatgtga atcataattc aatttttcaa catgcgtag gagggacatt
60

tcaaactctt ttttacccta gactttccta ccatcaccca gagtatccag ccaggagggg
120

aggggctaga gacaccagaa gtttagcagg gaggagggcg tagggattcg gggaatgaag
180

ggatgggatt cagactaggg ccaggaccca gggatggaga gaaagagatg agagtggttt
240

gggggcttgg tgacttagag aacagagctg caggctcaga ggcacacagg agtttctggg
300

ctcacctgc ccccttccaa cccctcagtt cccatcctcc agcagctgtt tgtgtgctgc
360

ctctgaagtc cacactgaac aaacttcagc ctactcatgt ccctaaaatg ggcaaacatt
420

gcaagcagca aacagcaaac acacagccct ccctgcctgc tgaccttga gctggggcag
480

aggtcagaga cctctctggg cccatgccac ctccaacatc cactcgaccc cttggaattt
540

cggtggagag gagcagaggt tgtcctggcg tggtttaggt agtgtgagag ggcttaagcg
600

tgaggctccg gtgcccgtca gtgggcagag cgcacatcgc ccacagtccc cgagaagttg
660

gggggagggg tcggcaattg aaccggtgcc tagagaaggt ggcgcgggggt aaactgggaa
720

agtgatgtcg tgtactggct ccgccttttt cccgaggggtg ggggagaacc gtatataagt
780

gcagtagtcg ccgtgaacgt tctttttcgc aacggggttg ccgccagaac acaggtaagt
840

gccgtgtgtg gttcccgcgg gcctggcctc ttacgggtt atggcccttg cgtgccttga
900

attacttcca cctggctcca gtacgtgatt cttgatcccg agctggagcc aggggcgggc
960

cttgcgcttt aggagcccct tcgcctcgtg cttgagttga ggcctggcct gggcgctggg
1020

gccgccgcgt gcgaatctgg tggcaccttc gcgcctgtct cgctgctttc gataagtctc
1080

tagccattta aaatTTTTga tgacctgctg cgacgctttt tttctggcaa gatagtcttg
1140

taaatgcggg ccaggatctg cacactggta tttcggtttt tggggccgcg ggcggcgacg
1200

gggcccgctg gtcccagcgc acatgttcgg cgaggcgggg cctgcgagcg cggccaccga
1260

gaatcggacg ggggtagtct caagctggcc ggcctgctct ggtgcctggc ctgcgcgccg
1320

cgtgtatcgc cccgccctgg gcggcaaggc tggcccggtc ggcaccagtt gcgtgagcgg
1380

aaagatggcc gcttcccggc cctgctccag ggggctcaaa atggaggacg cggcgctcgg
1440

gagagcgggc gggtgagtca cccacacaaa ggaaaggggc ctttccgtcc tcagccgtcg
1500

cttcatgtga ctccacggag taccgggcgc cgtccaggca cctcgattag ttctggagct
1560

tttgaggtac gtcgtcttta gggtgggggg aggggtttta tgcgatggag tttccccaca
1620

ctgagtgggt ggagactgaa gttaggccag cttggcactt gatgtaattc tccttggaat
1680

ttgccctttt tgagtttgga tcttggttca ttctcaagcc tcagacagtg gttcaaagtt
1740

tttttcttcc atttcagggtg tcgtgagcca ccatgtccac cgctgtgctg gagaaccctg
1800

ggctggggag gaaactgtca gacttcgggc aggagacttc atacattgag gataactgta
1860

accagaatgg cgccatctct ctgatcttca gcctgaagga ggaagtgggc gccctggcaa
1920

aggtgctgcg cctgttttgag gagaacgacg tgaatctgac ccacatcgag tcccggcctt
1980

ctagactgaa gaaggacgag tacgagttct ttaccacact ggataagcgg tccctgccag
2040

ccctgacaaa catcatcaag atcctgaggc acgacatcgg agcaaccgtg cacgagctgt
2100

ctcgggacaa gaagaaggat accgtgccct ggttccctcg gacaatccag gagctggata
2160

gatttgccaa ccagatcctg tcttacggag cagagctgga cgcagatcac cctggcttca
2220

aggacccagt gtatcggggc cggagaaagc agtttgccga tatcgacctac aattataggc
2280

acggacagcc aatccctcgc gtggagtata tggaggagga gaagaagacc tggggcacag
2340

tgttcaagac cctgaagagc ctgtacaaga cacacgcctg ctacgagtat aaccacatct
2400

tccccctgct ggagaagtat tgtggctttc acgaggacaa tatccctcag ctggaggacg
2460

tgagccagtt cctgcagacc tgcacaggct ttaggctgag gccagtggca ggactgctga
2520

gctcccggga cttcctggga ggactggcct tcagagtgtt tctctgcacc cagtacatca
2580

ggcacggctc caagccaatg tatacaccag agcccgacat ctgtcacgag ctgctgggcc
2640

acgtgcccct gtttagcgat agatccttcg cccagttttc ccaggagatc ggactggcat
2700

ctctgggagc acctgacgag tacatcgaga agctggccac catctattgg ttcacagtgg
2760

agtttggcct gtgcaagcag ggcgatagca tcaaggccta cggagcagga ctgctgtcta
2820

gcttcggcga gctgcagtat tgtctgtccg agaagccaaa gctgctgccc ctggagctgg
2880

agaagaccgc catccagaac tacaccgtga cagagttcca gcccctgtac tatgtggccg
2940

agtcttttaa cgatgccaag gagaagggtga gaaatttcgc cgccacaatc cctaggccct
3000

tcagcgtgcg gtacgaccct tatacccaga ggatcgaggt gctggataat acacagcagc
3060

tgaagatcct ggctgactca atcaatagcg aaatcggaat cctgtgctcc gccctgcaga
3120

aaatcaaatg aatcgtagat ccagacatga taagatacat tgatgagttt ggacaaacca
3180

caactagaat gcagtgaaaa aaatgcttta tttgtgaaat ttgtgatgct attgctttat
3240

ttgtaaccat tataagctgc aataaacaag ttaacaacaa caattgcatt cattttatgt
3300

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 t a a
3335

<210> 51

<211> 3783

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-004 transfer genome (from 5' ITR to 3' ITR)

<400> 51

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccc ggcggcctca gtgagcgagc gagcgcgagc agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcaga tcttcaatat tggccattag ccatattatt cattggttat
240

atagcataaa tcaatattgg ctattggcca ttgcatacgt tgtatctata tcataatatg
300

tacattttata ttggctcatg tccaatatga ccgccatggt ggcatgatt attgactagt
360

tattaatagt aatcaattac ggggtcatta gttcatagcc catatatgga gttccgcgtt
420

acataactta cggtaaattg cccgcctggc tgaccgcca acgacccccg cccattgacg
480

tcaataatga cgtatgttcc catagtaacg ccaataggga ctttccattg acgtcaatgg
540

gtggagtatt tacggtaaac tgcccacttg gcagtacatc aagtgtatca tatgccaagt
600

ccgcccccta ttgacgtcaa tgacggtaaa tggcccgctt ggcatatgc ccagtacatg
660

accttacggg actttcctac ttggcagtac atctacgtat tagtcatcgc tattaccatg
720

gtcgaggtga gccccacgtt ctgcttcact ctccccatct cccccccctc cccacccccca
780

attttgtatt tattttatfff ttaattatfff tgtgcagcga tgggggfcggg gggggggggg
840

gggcgcgcgc caggcggggc ggggcggggc gaggggfcggg gcggggfcgag gcggagaggt
900

gcggcggcag ccaatcagag cggcgcgcgc cgaaagtffc cttttatggc gaggcggcgg
960

cggcggcggc cctataaaaa gcgaagcgcg cggcgggfcgg gagtcgctgc gacgctgcct
1020

tcgccccgtg ccccgctccg ccgccgcctc gcgccgcccg ccccggtctt gactgaccgc
1080

gttactccca caggtgagcg ggcgggacgg cccttctcct ccgggctgta attagcgctt
1140

ggtttaatga cggcttggtt cttttctgtg gctgcgtgaa agccttgagg ggctccggga
1200

gggccctttg tgcggggggg agcggctcgg ggggtgcgtg cgtgtgtgtg tgcgtgggga
1260

gcgccgcgtg cggcccgcgc tgcccggcgg ctgtgagcgc tgcgggcgcg gcgcggggct
1320

ttgtgcgtc cgcagtgtgc gcgaggggag cgcggccggg ggcggtgccc cgcggtgcgg
1380

ggggggctgc gaggggaaca aaggctgcgt gcggggtgtg tgcgtggggg ggtgagcagg
1440

gggtgtgggc gcggcggtcg ggctgtaacc cccccctgca cccccctccc cgagttgctg
1500

agcacggccc ggcttcgggt gcggggctcc gtacggggcg tggcgcgggg ctcgccgtgc
1560

cgggcggggg gtggcggcag gtgggggtgc cgggcggggc ggggccgcct cgggccgggg
1620

agggtcggg ggaggggcgc ggcggccccc ggagcgccgg cggctgtcga ggcgcggcga
1680

gccgcagcca ttgcctttta tggtaatcgt gcgagagggc gcagggactt cctttgtccc
1740

aaatctgtgc ggagccgaaa tctgggaggc gccgccgcac cccctctagc gggcgcgggg
1800

cgaagcgggt cggcgccggc aggaaggaaa tgggcgggga gggccttcgt gcgtcgccgc
1860

gccgccgtcc ctttctccct ctccagcctc ggggctgtcc gcggggggac ggctgccttc
1920

gggggggacg gggcagggcg gggttcggct tctggcgtgt gaccggcggc tctagagcct
1980

ctgctaacca tgttcatgcc ttcttctttt tcctacagct cctgggcaac gtgctggtta
2040

ttgtgctgtc tcatcatttt ggcaaagaat tccgccacca tgtccaccgc tgtgctggag
2100

aaccctgggc tggggaggaa actgtcagac ttcgggcagg agacttcata cattgaggat
2160

aactgtaacc agaatggcgc catctctctg atcttcagcc tgaaggagga agtgggcgcc
2220

ctggcaaagg tgctgcgctt gtttgaggag aacgacgtga atctgacca catcgagtcc
2280

cggccttcta gactgaagaa ggacgagtac gagttcttta cccacctgga taagcgggcc
2340

ctgccagccc tgacaaacat catcaagatc ctgaggcacg acatcggagc aaccgtgcac
2400

gagctgtctc gggacaagaa gaaggatacc gtgccctggg tccctcggac aatccaggag
2460

ctggatagat ttgccaacca gatcctgtct tacggagcag agctggacgc agatcacctt
2520

ggcttcaagg acccagtgtg tcggggcccgg agaaagcagt ttgccgatat cgcctacaat
2580

tataggcacg gacagccaat ccctcgcgtg gagtatatgg aggaggagaa gaagacctgg
2640

ggcacagtgt tcaagaccct gaagagcctg tacaagacac acgcctgcta cgagtataac
2700

cacatcttcc ccctgctgga gaagtattgt ggctttcacg aggacaatat ccctcagctg
2760

gaggacgtga gccagttcct gcagacctgc acaggcttta ggctgaggcc agtggcagga
2820

ctgctgagct cccgggactt cctgggagga ctggccttca gagtgtttca ctgcacccag
2880

tacatcaggc acggctccaa gccaatgtat acaccagagc ccgacatctg tcacgagctg
2940

ctgggccacg tgcccctggt tagcgataga tccttcgccc agttttccca ggagatcgga
3000

ctggcatctc tgggagcacc tgacgagtac atcgagaagc tggccaccat ctattggttc
3060

acagtggagt ttggcctgtg caagcagggc gatagcatca aggcctacgg agcaggactg
3120

ctgtctagct tcggcgagct gcagtattgt ctgtccgaga agccaaagct gctgcccctg
3180

gagctggaga agaccgccat ccagaactac accgtgacag agttccagcc cctgtactat
3240

gtggccgagt cttttaacga tgccaaggag aaggtgagaa atttcgccgc cacaatccct
3300

aggcccttca gcgtgcggta cgacccttat acccagagga tcgaggtgct ggataataca
3360

cagcagctga agatcctggc tgactcaatc aatagcgaaa tcggaatcct gtgctccgcc
3420

ctgcagaaaa tcaaatgaat cgattctaga gtcgagccgc ggactagtaa cttgtttatt
3480

gcagcttata atggttacia ataaagcaat agcatcacia atttcacaaa taaagcattt
3540

ttttcactgc attctagttg tggtttgtcc aaactcatca atgtatctta ggtctagata
3600

cgtagataag tagcatggcg ggttaatcat taactacaag gaacccttag tgatggagtt
3660

ggccactccc tctctgcgcg ctcgctcgct cactgaggcc gggcgaccaa aggtcgcccg
3720

acgcccgggc ttgcccggg cggcctcagt gagcgagcga gcgcgcagag agggagtggc
3780

c
3783

a

a

<210> 52
<211> 2356
<212> DNA
<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-025 transfer genome (from 5' ITR to 3' ITR)

<400> 52

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgctcg ggcgaccttt
60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggaatt cacgcgtgga
120

tctgaattca attcacgcgt ggtacctccc taaaatgggc aaacattgca agcagcaaac
180

agcaaacaca cagccctccc tgcctgctga ccttggagct ggggcagagg tcagagacct
240

ctctgggccc atgccacctc caacatccac tcgacctt ggaatttcgg tggagaggag
300

cagaggttgt cctggcgtgg tttaggtagt gtgagagggg aatgactcct ttcggtaagt
360

gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc agcgtaggcg ggcgactcag
420

atcccagcca gtggacttag cccctgtttg ctcctccgat aactgggggtg accttggtta
480

atattcacca gcagcctccc ccgttgcccc tctggatcca ctgcttaaat acggacgagg
540

acagggccct gtctcctcag cttcaggcac caccactgac ctgggacagt gaatcctcta
600

aggtaaatat aaaattttta agtgtataat gtgttaaact actgattcta attgtttctc
660

tcttttagat tccaaccttt ggaactgacc gccaccatgt ccaccgctgt gctggagaac
720

cctgggctgg ggaggaaact gtcagacttc gggcaggaga cttcatacat tgaggataac
780

tgtaaccaga atggcgccat ctctctgata ttcagcctga aggaggaagt gggcgccctg
840

gcaaagggtgc tgcgcctggt tgaggagaac gacgtgaatc tgaccacat cgagtcccgg
900

ccttctagac tgaagaagga cgagtacgag ttctttaccc acctggataa gcggtccctg
960

ccagccctga caaacatcat caagatcctg aggcacgaca tcggagcaac cgtgcacgag
1020

ctgtctcggg acaagaagaa ggataccgtg ccctgggttc ctcggacaat ccaggagctg
1080

gatagatttg ccaaccagat cctgtcttac ggagcagagc tggacgcaga tcaccctggc
1140

ttcaaggacc cagtgtatcg ggcccggaga aagcagtttg ccgatatcgc ctacaattat
1200

aggcacggac agccaatccc tcgctggag tatatggagg aggagaagaa gacctggggc
1260

acagtgttca agaccctgaa gagcctgtac aagacacacg cctgctacga gtataaccac
1320

atcttcccc tgctggagaa gtattgtggc tttcacgagg acaatatccc tcagctggag
1380

gacgtgagcc agttcctgca gacctgcaca ggctttaggc tgaggccagt ggcaggactg
1440

ctgagctccc gggacttcct gggaggactg gccttcagag tgtttcactg caccagtac
1500

atcaggcacg gctccaagcc aatgtataca ccagagcccg acatctgtca cgagctgctg
1560

ggccacgtgc ccctgttttag cgatagatcc ttcgcccagt tttcccagga gatcggactg
1620

gcatctcttg gagcacctga cgagtacatc gagaagctgg ccaccatcta ttggttcaca
1680

gtggagtttg gcctgtgcaa gcagggcgat agcatcaagg cctacggagc aggactgctg
1740

tctagcttcg gcgagctgca gtattgtctg tccgagaagc caaagctgct gcccctggag
1800

ctggagaaga ccgccatcca gaactacacc gtgacagagt tccagcccct gtactatgtg
1860

gccgagtctt ttaacgatgc caaggagaag gtgagaaatt tcgccgccac aatccctagg
1920

cccttcagcg tgcggtacga cccttatacc cagaggatcg aggtgctgga taatacacag
1980

cagctgaaga tcctggctga ctcaatcaat agcgaaatcg gaatcctgtg ctccgccctg
2040

cagaaaatca aatgaatgct ttattttgtga aattttgtgat gctattgctt tattttgtaac
2100

cattataagc tgcaataaac aagttaacaa caacaattgc attcatttta tgtttcaggt
2160

tcaggggggag gtgtgggagg ttttttaaag catgctgggg agagatcgat ctgaggaacc
2220

cctagtgatg gagttggcca ctccctctct gcgcgctcgc tcgctcactg aggccgggcg
2280

accaaaggtc gcccgacgcc cgggctttgc ccgggcggcc tcagtgagcg agcgagcgcg
2340

c a g a g a g g g a g t g g c c
2356

<211> 2323

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-010 transfer genome (from 5' ITR to 3' ITR)

<400> 53

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggcgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcagg gggaggctgc tgggtgaatat taaccaaggt caccgccagtt
240

atcggaggag caaacagggg ctaagtccac ctcgagccat ggcgatgctc taatctctct
300

agacaagggt catatttgta tgggttactt attctctctt tgttgactaa gtcaataatc
360

agaatcagca ggtttgcagt cagattggca gggataagca gcctagctca ggagaagtga
420

gtataaaagc cccaggctgg gagcagccat cagctagcgc cggcaagagg taagggttta
480

agggatgggt ggttgggtgg gtattaatgt ttaattacct ggagcacctg cctgaaatca
540

ctttttttca ggttggggcca ccatgtccac cgctgtgctg gagaaccctg ggctggggag
600

gaaactgtca gacttcgggc aggagacttc atacattgag gataactgta accagaatgg
660

cgccatctct ctgatcttca gcctgaagga ggaagtgggc gccctggcaa aggtgctgcg
720

cctgtttgag gagaacgacg tgaatctgac ccacatcgag tcccggcctt ctagactgaa
780

gaaggacgag tacgagttct ttaccacact ggataagcgg tccctgccag ccctgacaaa
840

catcatcaag atcctgaggc acgacatcgg agcaaccgtg cacgagctgt ctcgggacaa
900

gaagaaggat accgtgccct ggttccctcg gacaatccag gagctggata gatttgccaa
960

ccagatcctg tcttacggag cagagctgga cgcagatcac cctggcttca aggaccagct
1020

gtatcgggcc cggagaaagc agtttgccga tatcgcctac aattataggc acggacagcc
1080

aatccctcgc gtggagtata tggaggagga gaagaagacc tggggcacag tgttcaagac
1140

cctgaagagc ctgtacaaga cacacgcctg ctacgagtat aaccacatct tccccctgct
1200

ggagaagtat tgtggctttc acgaggacaa tatccctcag ctggaggacg tgagccagtt
1260

cctgcagacc tgcacaggct ttaggctgag gccagtggca ggactgctga gctcccggga
1320

cttcctggga ggactggcct tcagagtgtt tcaactgcacc cagtacatca ggcacggctc
1380

caagccaatg tatacaccag agcccgacat ctgtcacgag ctgctgggcc acgtgcccct
1440

gtttagcgat agatccttcg cccagttttc ccaggagatc ggactggcat ctctgggagc
1500

acctgacgag tacatcgaga agctggccac catctattgg ttcacagtgg agtttggcct
1560

gtgcaagcag ggcgatagca tcaaggccta cggagcagga ctgctgtcta gcttcggcga
1620

gctgcagtat tgtctgtccg agaagccaaa gctgctgccc ctggagctgg agaagaccgc
1680

catccagaac tacaccgtga cagagttcca gcccctgtac tatgtggccg agtcttttaa
1740

cgatgccaaag gagaagggtga gaaatttcgc cgccacaatc cctaggccct tcagcgtgcg
1800

gtacgaccct tatacccaga ggatcgaggt gctggataat acacagcagc tgaagatcct
1860

ggctgactca atcaatagcg aaatcggaat cctgtgctcc gccctgcaga aaatcaaattg
1920

aatcgtagat ccagacatga taagatacat tgatgagttt ggacaaacca caactagaat
1980

gcagtgaaaa aaatgcttta tttgtgaaat ttgtgatgct attgctttat ttgtaaccat
2040

tataagctgc aataaacaag ttaacaacaa caattgcatt cattttatgt ttcaggttca
2100

gggggaggtg tgggaggttt ttttaagcttg tttaaacgta cgtagataag tagcatggcg
2160

ggttaatcat taactacaag gaacccttag tgatggagtt ggccactccc tctctgcgcg
2220

ctcgctcgct cactgaggcc gggcgaccaa aggtcgcccg acgcccgggc tttgcccggg
2280

cggcctcagt gagcgagcga gcgcgagag agggagtggc caa
2323

<210> 54

<211> 3090

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-011 transfer genome (from 5' ITR to 3' ITR)

<400> 54

ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcagt aaattttatg gaatgtgaat cataattcaa tttttcaaca
240

tgcgttagga gggacatttc aaactctttt ttaccctaga ctttcctacc atcaccaga
300

gtatccagcc aggaggggag gggctagaga caccagaagt ttagcaggga ggagggcgta
360

gggattcggg gaatgaagg atgggattca gactagggcc aggaccagc gatggagaga
420

aagagatgag agtggttttg gggcttggtg acttagagaa cagagctgca ggctcagagg
480

cacacaggag tttctgggct caccctgcc cctccaacc cctcagttcc catcctccag
540

cagctgtttg tgtgctgcct ctgaagtcca cactgaacaa acttcagcct actcatgtcc
600

ctaaaatggg caaacattgc aagcagcaaa cagcaaacac acagccctcc ctgcctgctg
660

accttggagc tggggcagag gtcagagacc tctctgggcc catgccacct ccaacatcca
720

ctcgaccct tggaatttcg gtggagagga gcagaggttg tcctggcgtg gtttaggtag
780

tgtgagaggg ctttaaggctc taaccactc tgatctccca ggcggcagt aagtcttcag
840

catcaggcat tttgggggtga ctcagtaa atgtagatctt gctaccagtg gaacagccac
900

taaggattct gcagtgagag cagagggcca gctaagtggg actctcccag agactgtctg
960

actcacgcca cccctccac cttggacaca ggacgctgtg gtttctgagc caggtacaat
1020

gactcctttc ggtaagtgca gtggaagctg tacactgccc aggcaaagcg tccgggcagc
1080

gtaggcgggc gactcagatc ccagccagtg gacttagccc ctgtttgctc ctccgataac
1140

tggggtgacc ttggttaata ttcaccagca gcctcccccg ttgcccctct ggatccactg
1200

cttaaatacg gacgaggacg ctagcgccgg caagaggtaa gggtttaagg gatggttggt
1260

tgggtgggta ttaatgttta attacctgga gcacctgcct gaaatcactt tttttcaggt
1320

tgggccacca tgtccaccgc tgtgctggag aaccctgggc tggggaggaa actgtcagac
1380

ttcgggcagg agacttcata cattgaggat aactgtaacc agaatggcgc catctctctg
1440

atcttcagcc tgaaggagga agtgggcgcc ctggcaaagg tgctgcgcct gtttgaggag
1500

aacgacgtga atctgacca catcgagtcc cggccttcta gactgaagaa ggacgagtac
1560

gagttcttta cccacctgga taagcgggtcc ctgccagccc tgacaaacat catcaagatc
1620

ctgaggcacg acatcggagc aaccgtgcac gagctgtctc gggacaagaa gaaggatacc
1680

gtgccctggt tccctcggac aatccaggag ctggatagat ttgccaacca gatcctgtct
1740

tacggagcag agctggacgc agatcaccct ggcttcaagg acccagtgta tcgggcccgg
1800

agaaagcagt ttgccgatat cgcctacaat tataggcacg gacagccaat ccctcgcgtg
1860

gagtatatgg aggaggagaa gaagacctgg ggcacagtgt tcaagaccct gaagagcctg
1920

tacaagacac acgcctgcta cgagtataac cacatcttcc ccctgctgga gaagtattgt
1980

ggctttcacg aggacaatat ccctcagctg gaggacgtga gccagttcct gcagacctgc
2040

acaggcttta ggctgaggcc agtggcagga ctgctgagct cccgggactt cctgggagga
2100

ctggccttca gagtgtttca ctgcacccag tacatcaggc acggctccaa gccaatgtat
2160

acaccagagc ccgacatctg tcacgagctg ctgggccacg tgcccctgtt tagcgataga
2220

tccttcgccc agttttccca ggagatcgga ctggcatctc tgggagcacc tgacgagtac
2280

atcgagaagc tggccaccat ctattgggtc acagtggagt ttggcctgtg caagcagggc
2340

gatagcatca aggcctacgg agcaggactg ctgtctagct tcggcgagct gcagtattgt
2400

ctgtccgaga agccaaagct gctgcccctg gagctggaga agaccgccat ccagaactac
2460

accgtgacag agttccagcc cctgtactat gtggccgagt cttttaacga tgccaaggag
2520

aaggtgagaa atttcgccgc cacaatccct aggcccttca gcgtgcggta cgacccttat
2580

accagagga tcgaggtgct ggataataca cagcagctga agatcctggc tgactcaatc
2640

aatagcgaat tcggaatcct gtgctccgcc ctgcagaaaa tcaaatgaat cgtagatcca
2700

gacatgataa gatacattga tgagtttgga caaaccacaa ctagaatgca gtgaaaaaaaa
2760

tgctttatatt gtgaaatttg tgatgctatt gctttatttg taaccattat aagctgcaat
2820

aaacaagtta acaacaacaa ttgcattcat tttatgtttc aggttcaggg ggaggtgtgg
2880

gagggtttttt aagcttggtt aaacgtacgt agataagtag catggcgggg taatcattaa
2940

ctacaaggaa cccctagtga tggagttggc cactccctct ctgcgcgctc gctcgctcac
3000

tgaggccggg cgaccaaagg tcgcccgcgc cccgggcttt gcccgggcgg cctcagtgag
3060

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
3090

<210> 55

<211> 3731

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-012 transfer genome (from 5' ITR to 3' ITR)

<400> 55

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgagc agagggagtg
120

gccaactcca tcactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcagt aaattttatg gaatgtgaat cataattcaa tttttcaaca
240

tgcgtagga gggacatttc aaactctttt ttaccctaga ctttcctacc atcaccaga
300

gtatccagcc aggaggggag gggctagaga caccagaagt ttagcaggga ggagggcgta
360

gggattcggg gaatgaaggg atgggattca gactagggcc aggacccagg gatggagaga
420

aagagatgag agtggttttg gggcttggtg acttagagaa cagagctgca ggctcagagg
480

cacacaggag tttctgggct caccctgccc ccttccaacc cctcagttcc catcctccag
540

cagctgtttg tgtgctgcct ctgaagtcca cactgaacaa acttcagcct actcatgtcc
600

ctaaaatggg caaacattgc aagcagcaaa cagcaaacac acagccctcc ctgcctgctg
660

accttggagc tggggcagag gtcagagacc tctctgggcc catgccacct ccaacatcca
720

ctcgaccctt tggaatttcg gtggagagga gcagaggttg tcctggcgtg gtttaggtag
780

tgtgagaggg cttaagcgtg aggctccggt gcccgtcagt gggcagagcg cacatcgccc
840

acagtccccg agaagttggg gggaggggtc ggcaattgaa ccggtgccta gagaagggtg
900

cgcggggtaa actgggaaag tgatgtcgtg tactggctcc gcctttttcc cgagggtggg
960

ggagaaccgt atataagtgc agtagtcgcc gtgaacgttc tttttcgcaa cgggtttgcc
1020

gccagaacac aggtaagtgc cgtgtgtggt tcccgcgggc ctggcctctt tacgggttat
1080

ggcccttgcg tgccttgaat tacttccacc tggctccagt acgtgattct tgatcccag
1140

ctggagccag gggcgggcct tgcgcttttag gagccccttc gcctcgtgct tgagttgagg
1200

cctggcctgg gcgctggggc cgccgcgtgc gaatctggtg gcaccttcgc gcctgtctcg
1260

ctgcttttcga taagtcttcta gccattttaa atttttgatg acctgctgcg acgctttttt
1320

tctggcaaga tagtcttgta aatgcggggc aggatctgca cactggtatt tcggtttttg
1380

gggccgcggg cggcgacggg gcccgtgctt cccagcgcac atgttcggcg aggcgggggc
1440

tgcgagcgcg gccaccgaga atcggacggg ggtagtctca agctggccgg cctgctctgg
1500

tgccctggcct cgcgccgccg tgtatcgccc cgccctgggc ggcaaggctg gcccggtcgg
1560

caccagttgc gtgagcgga agatggccgc ttcccggccc tgctccaggg ggctcaaaat
1620

ggaggacgcg gcgctcggga gagcggggcg gtgagtcacc cacacaaagg aaaggggcct
1680

ttccgtcctc agccgtcgct tcatgtgact ccacggagta ccgggcgccg tccaggcacc
1740

tcgattagtt ctggagcttt tggagtacgt cgtcttttagg ttgggggggag gggttttatg
1800

cgatggagtt tccccacact gagtgggtgg agactgaagt taggccagct tggcacttga
1860

tgtaattctc cttggaattt gccctttttg agtttggatc ttgggttcatt ctcaagcctc
1920

agacagtggg tcaaagtttt tttcttccat ttcagggtgc gtgagccacc atgtccaccg
1980

ctgtgctgga gaaccctggg ctggggagga aactgtcaga cttcgggcag gagacttcat
2040

acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc ctgaaggagg
2100

aagtgggcgc cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg aatctgaccc
2160

acatcgagtc 2220	ccggccttct	agactgaaga	aggacgagta	cgagttcttt	acccacctgg
ataagcggtc 2280	cctgccagcc	ctgacaaaca	tcatcaagat	cctgaggcac	gacatcggag
caaccgtgca 2340	cgagctgtct	cgggacaaga	agaaggatac	cgtgccctgg	ttccctcgga
caatccagga 2400	gctggataga	tttgccaacc	agatcctgtc	ttacggagca	gagctggacg
cagatcaccc 2460	tggcttcaag	gacccagtgt	atcggggccg	gagaaagcag	tttgccgata
tgcctacaa 2520	ttataggcac	ggacagccaa	tccctcgcgt	ggagtatatg	gaggaggaga
agaagacctg 2580	gggcacagtg	ttcaagaccc	tgaagagcct	gtacaagaca	cacgcctgct
acgagtataa 2640	ccacatcttc	cccctgctgg	agaagtattg	tggctttcac	gaggacaata
tccctcagct 2700	ggaggacgtg	agccagttec	tgcagacctg	cacaggcttt	aggctgaggc
cagtggcagg 2760	actgctgagc	tcccgggact	tcctgggagg	actggccttc	agagtgtttc
actgcaccca 2820	gtacatcagg	cacggctcca	agccaatgta	tacaccagag	cccgacatct
gtcacgagct 2880	gctggggccac	gtgcccctgt	ttagcgatag	atccttcgcc	cagttttccc
aggagatcgg 2940	actggcatct	ctgggagcac	ctgacgagta	catcgagaag	ctggccacca
tctattgggt 3000	cacagtggag	tttggcctgt	gcaagcaggg	cgatagcatc	aaggcctacg
gagcaggact 3060	gctgtctagc	ttcggcgagc	tgcagtattg	tctgtccgag	aagccaaagc

tgctgcccct ggagctggag aagaccgcca tccagaacta caccgtgaca gagttccagc
3120

ccctgtacta tgtggccgag tctttttaacg atgccaagga gaagggtgaga aatttcgccg
3180

ccacaatccc taggcccttc agcgtgcggt acgaccctta taccagagg atcgaggtgc
3240

tgataatac acagcagctg aagatcctgg ctgactcaat caatagcgaa atcggaatcc
3300

tgtgctccgc cctgcagaaa atcaaatgaa tcgtagatcc agacatgata agatacattg
3360

atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt
3420

gtgatgctat tgctttatatt gtaaccatta taagctgcaa taaacaagtt aacaacaaca
3480

attgcattca ttttatgttt caggttcagg gggagggtgtg ggagggttttt taagcttggt
3540

taaacgtacg tagataagta gcatggcggg ttaatcatta actacaagga acccctagtg
3600

atggagttgg ccaactccctc tctgcgcgct cgctcgctca ctgaggccgg gcgaccaaag
3660

gtcgcccgac gcccgggctt tgcccgggcg gcctcagtga gcgagcgagc gcgcagagag
3720

g g a g t g g c c a a
3731

<210> 56

<211> 37

<212> DNA

<213> Artificial Sequence

<220>

<223> 37 bp additional 3' ITR sequence from wtAAV2

<400> 56
gtagataagt agcatggcgg gttaatcatt aactaca
37

<210> 57
<211> 180
<212> DNA
<213> Artificial Sequence

<220>
<223> 3'ITR with additional 37 bp sequence

<400> 57
gtagataagt agcatggcgg gttaatcatt aactacaagg aaccctagt gatggagttg
60

gccactccct ctctgcgcgc tcgctcgctc actgaggccg ggcgaccaa ggtcgccga
120

cgcccgggct ttgcccgggc ggcctcagt agcgagcgag cgcgagaga gggagtggcc
180

<210> 58
<211> 380
<212> DNA
<213> Artificial Sequence

<220>
<223> CMV enhancer

<400> 58
ggcattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc
60

catatatgga gttccgcgtt acataactta cggtaaattgg cccgcctggc tgaccgcca
120

acgacccccg cccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga
180

ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc
240

aagtgtatca tatgccaagt ccgcccccta ttgacgtcaa tgacggtaaa tggcccgctt
300

ggcattatgc ccagtacatg accttacggg actttcctac ttggcagtac atctacgtat
360

t a g t c a t c g c t a t t a c c a t g
380

<210> 59

<211> 1246

<212> DNA

<213> Artificial Sequence

<220>

<223> CBA promoter

<400> 59

tcgaggtgag ccccacgttc tgcttcactc tccccatctc cccccctcc ccacccccaa
60

ttttgtattt atttatTTTT taattatTTT gtgcagcgat gggggcgggg gggggggggg
120

ggcgcgcgcc aggcggggcg gggcggggag aggggagggg cggggcgagg cggagaggtg
180

cggcggcagc caatcagagc ggcgcgctcc gaaagtttcc ttttatggcg aggcggcggc
240

ggcgggcgcc ctataaaaag cgaagcgcgc ggcggggcgag agtcgctgag cgctgccttc
300

gccccgtgcc ccgctccgcc gccgcctcgc gccgccccgc ccggctctga ctgaccgcgt
360

tactcccaca ggtgagcgag cgggacggcc cttctcctcc gggctgtaat tagcgcttgg
420

tttaatgacg gcttgtttct tttctgtggc tgcgtgaaag ccttgagggg ctccgggagg
480

gccctttgtg cgggggggag ggctcggggg gtgcgtgcgt gtgtgtgtgc gtggggagcg
540

ccgcgtgcgg ctccgcgctg cccggcggct gtgagcgtg cgggcgcggc gcggggcttt
600

gtgcgtccg cagtgtgcgc gaggggagcg cggccggggg cggtgccccg cggtgcgggg
660

ggggctgcga ggggaacaaa ggctgcgtgc ggggtgtgtg cgtggggggg tgagcagggg
720

gtgtgggcgc gtcggtcggg ctgcaacccc ccctgcaccc ccctccccga gttgctgagc
780

acggccccggc ttcgggtgcg gggctccgta cggggcgtgg cgcggggctc gccgtgccgg
840

gcgggggggtg gcggcaggtg ggggtgccgg gcggggcggg gccgcctcgg gccggggagg
900

gctcggggga ggggcgcggc ggcccccgga gcgccggcgg ctgtcgaggc gcggcgagcc
960

gcagccattg ctttttatgg taatcgtgcg agagggcgca gggacttcct ttgtcccaaa
1020

tctgtgcgga gccgaaatct gggaggcgcc gccgcacccc ctctagcggg cgcggggcgga
1080

agcgggtgcgg cggcggcagg aaggaaatgg gcggggaggg ccttcgtgcg tcgccgcgcc
1140

gccgtcccct tctccctctc cagcctcggg gctgtccgcg gggggacggc tgccttcggg
1200

ggggacgggg cagggcgggg ttcggcttct ggcgtgtgac cggcgg
1246

<210> 60

<211> 95

<212> DNA

<213> Artificial Sequence

<220>

<223> rabbit beta-globin element

<400> 60
cctctgctaa ccatgttcat gccttcttct ttttcctaca gctcctgggc aacgtgctgg
60

t t a t t g t g c t g t c t c a t c a t t t t g g c a a a g a a t t c
95

<210> 61

<211> 3301

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-009 transfer genome

<400> 61

tggcattgat tattgactag ttattaatag taatcaatta cgggggtcatt agttcatagc
60

ccatatatgg agttccgcgt tacataactt acggtaaagt gcccgcttg ctgaccgccc
120

aacgaccccc gcccattgac gtcaataatg acgtatgttc ccatagtaac gccaataggg
180

actttccatt gacgtcaatg ggtggagtat ttacggtaaa ctgcccactt ggcagtacat
240

caagtgtatc atatgccaag tccgccccct attgacgtca atgacggtaa atggcccgcc
300

tggcattatg ccagtacat gaccttacgg gactttccta cttggcagta catctacgta
360

ttagtcatcg ctattaccat ggtcgaggtg agccccacgt tctgcttcac tctccccatc
420

tcccccccct cccaccccc aattttgtat ttattttatt ttttaattatt ttgtgcagcg
480

atgggggfcgg ggggggggggg ggggcgcgcg ccaggcgggg cggggcgggg cgaggggfcgg
540

ggcggggcga ggcggagagg tgcggcggca gccaatcaga gcggcgcgct ccgaaagttt
600

ccttttatgg cgaggcggcg gcggcggcgg ccctataaaa agcgaagcgc gcggcgggcg
660

ggagtcgctg cgcgctgcct tcgccccgtg ccccgctccg ccgccgcctc gcgccgcccg
720

ccccggctct gactgaccgc gttactccca caggtgagcg ggcgggacgg cccttctcct
780

ccgggctgta attagcgctt ggtttaatga cggcttgttt cttttctgtg gctgcgtgaa
840

agccttgagg ggctccggga gggccctttg tgcgggggga gcggctcggg gggtgcgtagc
900

gtgtgtgtgt gcgtggggag cgccgcgtgc ggctccgcgc tgcccggcgg ctgtgagcgc
960

tgcgggcgcg gcgcggggct ttgtgcgctc cgcagtgtgc gcgaggggag cgcggccggg
1020

ggcggtgccc cgcggtgcgg ggggggctgc gaggggaaca aaggctgcgt gcgggggtgtg
1080

tgcgtggggg ggtgagcagg ggggtgtgggc gcgtcggtcg ggctgcaacc cccctgcac
1140

ccccctcccc gagttgctga gcacggcccc gcttcgggtg cggggctccg tacggggcgt
1200

ggcgcggggc tcgccgtgcc gggcgggggg tggcggcagg tgggggtgcc gggcggggcg
1260

gggccgcctc gggccgggga gggctcgggg gaggggcgcg gcggcccccg gagcgccggc
1320

ggctgtcgag gcgcggcgag ccgcagccat tgccttttat ggtaatcgtg cgagagggcg
1380

cagggacttc ctttgtccca aatctgtgcg gagccgaaat ctgggaggcg ccgccgcacc
1440

ccctctagcg ggcgcggggc gaagcgggtgc ggcgccggca ggaaggaaat gggcggggag
1500

ggccttcgtg cgtcgccgcg ccgccgtccc cttctccctc tccagcctcg gggctgtccg
1560

cggggggacg gctgccttcg ggggggacgg ggcagggcgg ggttcggctt ctggcgtgtg
1620

accggcggct ctagagcctc tgctaaccat gttcatgcct tcttcttttt cctacagctc
1680

ctgggcaacg tgctggttat tgtgctgtct catcattttg gcaaagaatt ccgccaccat
1740

gtccaccgct gtgctggaga accctgggct ggggaggaaa ctgtcagact tcgggcagga
1800

gacttcatac attgaggata actgtaacca gaatggcgcc atctctctga tcttcagcct
1860

gaaggaggaa gtgggcgccc tggcaaaggt gctgcgcctg tttgaggaga acgacgtgaa
1920

tctgaccac atcgagtccc ggccttctag actgaagaag gacgagtacg agttctttac
1980

ccacctggat aagcgggtccc tgccagccct gacaaacatc atcaagatcc tgaggcacga
2040

catcggagca accgtgcacg agctgtctcg ggacaagaag aaggataccg tgccctgggt
2100

ccctcggaca atccaggagc tggatagatt tgccaaccag atcctgtctt acggagcaga
2160

gctggacgca gatcaccctg gcttcaagga ccagtgatat cgggcccgga gaaagcagtt
2220

tgccgatatc gcctacaatt ataggcacgg acagccaatc cctcgcgtgg agtatatgga
2280

ggaggagaag aagacctggg gcacagtgtt caagaccctg aagagcctgt acaagacaca
2340

cgcttgctac gagtataacc acatcttccc cctgctggag aagtattgtg gctttcacga
2400

ggacaatatc cctcagctgg aggacgtgag ccagttcctg cagacctgca caggctttag
2460

gctgaggcca gtggcaggac tgctgagctc ccgggacttc ctgggaggac tggccttcag
2520

agtgtttcac tgcacccagt acatcaggca cggctccaag ccaatgtata caccagagcc
2580

cgacatctgt cacgagctgc tgggccacgt gcccctgttt agcgatagat ccttcgcccc
2640

gttttcccag gagatcggac tggcatctct gggagcacct gacgagtaca tcgagaagct
2700

ggccaccatc tattggttca cagtggagtt tggcctgtgc aagcagggcg atagcatcaa
2760

ggcctacgga gcaggactgc tgtctagctt cggcgagctg cagtattgtc tgtccgagaa
2820

gccaaagctg ctgcccctgg agctggagaa gaccgccatc cagaactaca ccgtgacaga
2880

gttccagccc ctgtactatg tggccgagtc ttttaacgat gccaaaggaga aggtgagaaa
2940

tttcgccgcc acaatcccta ggcccttcag cgtgcggtac gacccttata cccagaggat
3000

cgaggtgctg gataatacac agcagctgaa gatcctggct gactcaatca atagcgaaat
3060

cggaatcctg tgctccgccc tgcagaaaat caaatgaatc gtagatccag acatgataag
3120

atacattgat gagtttggac aaaccacaac tagaatgcag tgaaaaaat gctttatttg
3180

tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata aacaagttaa
3240

caacaacaat tgcattcatt ttatgtttca gggtcagggg gaggtgtggg aggtttttta
3300

a
3301

<210> 62
<211> 3701
<212> DNA
<213> Artificial Sequence

<220>
<223> pHMI-hPAH-TC-009 transfer genome

<400> 62
ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggctgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgagc agagggagtg
120

gccaactcca tcactagggg ttccctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcaga tctggcattg attattgact agttattaat agtaatcaat
240

tacgggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa
300

tggcccgccct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt
360

tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta
420

aactgcccac ttggcagtac atcaagtgtg tcatatgcc aagtcgcccc ctattgacgt
480

caatgacggg aaatggcccc cctggcatta tgcccagtac atgaccttac gggactttcc
540

tacttggcag tacatctacg tattagtcac cgctattacc atgggtcgagg tgagccccac
600

gttctgcttc actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat
660

ttttaatta ttttgtgcag cgatgggggc gggggggggg ggggggcgcg cgccaggcgg
720

ggcggggcgg ggcgaggggc ggggcggggc gaggcggaga ggtgcggcgg cagccaatca
780

gagcggcgcg ctccgaaagt ttccttttat ggcgaggcgg cggcggcggc ggccctataa
840

aaagcgaagc gcgcggcggg cgggagtcgc tgcgcgctgc cttcgccccg tgccccgctc
900

cgccgccgcc tcgcgccgcc cgccccggct ctgactgacc gcgttactcc cacaggtag
960

cgggcgggac ggcccttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt
1020

ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg
1080

gagcggctcg gggggtgcgt gcgtgtgtgt gtgcgtgggg agcgccgcgt gcggctccgc
1140

gctgcccggc ggctgtgagc gctgcggggc cggcgcgggg ctttgtgcgc tccgcagtgt
1200

gcgcgagggg agcgcgggccg ggggcgggtgc cccgcgggtgc gggggggggct gcgaggggaa
1260

caaaggctgc gtgcgggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcgtcgg
1320

cgggctgcaa cccccctgc acccccctcc ccgagttgct gagcacggcc cggcttcggg
1380

tgccggggctc cgtacggggc gtggcgcggg gctcgccgtg ccgggcgggg ggtggcggca
1440

ggtgggggtg ccgggcgggg cggggccgcc tcgggcccgg gagggctcgg gggaggggag
1500

cggcggcccc cggagcgccg gcggctgtcg aggcgcggcg agccgcagcc attgcctttt
1560

atggtaatcg tgcgagaggg cgcagggact tcctttgtcc caaatctgtg cggagccgaa
1620

atctgggagg cgccgccgca cccctctag cgggcgcggg gcgaagcggg gcggcgccgg
1680

caggaaggaa atgggcgggg agggccttcg tgcgtcgccg cgccgccgtc cccttctccc
1740

tctccagcct cggggctgtc cgcgggggga cggctgcctt cgggggggac ggggcagggc
1800

ggggttcggc ttctggcgtg tgaccggcgg ctctagagcc tctgctaacc atgttcatgc
1860

cttcttcttt ttctacagc tcctgggcaa cgtgctggtt attgtgctgt ctcatcattt
1920

tggcaaagaa ttccgccacc atgtccaccg ctgtgctgga gaacctggg ctggggagga
1980

aactgtcaga cttcgggcag gagacttcat acattgagga taactgtaac cagaatggcg
2040

ccatctctct gatcttcagc ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc
2100

tgtttgagga gaacgacgtg aatctgaccc acatcgagtc ccggccttct agactgaaga
2160

aggacgagta cgagttcttt acccacctgg ataagcggtc cctgccagcc ctgacaaaca
2220

tcatcaagat cctgaggcac gacatcggag caaccgtgca cgagctgtct cgggacaaga
2280

agaaggatac cgtgccctgg ttccctcgga caatccagga gctggataga tttgccaacc
2340

agatcctgtc ttacggagca gagctggacg cagatcaccg tggcttcaag gaccagtggt
2400

atcgggccccg 2460	gagaaagcag	tttgccgata	tcgcctacaa	ttataggcac	ggacagccaa
tccctcgcgt 2520	ggagtatatg	gaggaggaga	agaagacctg	gggcacagtg	ttcaagaccc
tgaagagcct 2580	gtacaagaca	cacgcctgct	acgagtataa	ccacatcttc	cccctgctgg
agaagtattg 2640	tggctttcac	gaggacaata	tccctcagct	ggaggacgtg	agccagttcc
tgcagacctg 2700	cacaggcttt	aggctgaggc	cagtggcagg	actgctgagc	tcccgggact
tcctgggagg 2760	actggccttc	agagtgtttc	actgcacca	gtacatcagg	cacggctcca
agccaatgta 2820	tacaccagag	cccgacatct	gtcacgagct	gctgggccac	gtgcccctgt
ttagcgatag 2880	atccttcgcc	cagttttccc	aggagatcgg	actggcatct	ctgggagcac
ctgacgagta 2940	catcgagaag	ctggccacca	tctattggtt	cacagtggag	tttggcctgt
gcaagcaggg 3000	cgatagcatc	aaggcctacg	gagcaggact	gctgtctagc	ttcggcgagc
tgcagtattg 3060	tctgtccgag	aagccaaagc	tgctgcccct	ggagctggag	aagaccgcca
tccagaacta 3120	caccgtgaca	gagttccagc	ccctgtacta	tgtggccgag	tcttttaacg
atgccaagga 3180	gaaggtgaga	aatttcgccg	ccacaatccc	taggcccttc	agcgtgcggt
acgaccctta 3240	taccagagag	atcgaggtgc	tggataatac	acagcagctg	aagatcctgg
ctgactcaat 3300	caatagcgaa	atcggaatcc	tgtgctccgc	cctgcagaaa	atcaaatgaa

tcgtagatcc agacatgata agatacattg atgagtttgg acaaaccaca actagaatgc
3360

agtgaaaaaa atgcttttatt tgtgaaattt gtgatgctat tgctttatatt gtaaccatta
3420

taagctgcaa taaacaagtt aacaacaaca attgcattca ttttatgttt caggttcagg
3480

gggaggtgtg ggagggttttt taagcttggt taaacgtacg tagataagta gcatggcggg
3540

ttaatcatta actacaagga acccctagtg atggagttgg ccactccctc tctgcgcgct
3600

cgctcgctca ctgaggccgg gcgaccaaag gtcgcccgcac gcccgggctt tgcccgggcg
3660

gcctcagtga gcgagcgcgc gcgcagagag ggagtggcca a
3701

<210> 63

<211> 1061

<212> DNA

<213> Artificial Sequence

<220>

<223> CASI promoter region

<400> 63

tagggaggtc ctgcacgtta cataacttac ggtaaattggc ccgcctggct gaccgcccac
60

cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc caatagggac
120

tttccattga cgtcaatggg tggagtattt acggtaaact gcccaattgg cagtacatca
180

agtgtatcat atgccaagta cgccccctat tgacgtcaat gacggtaaata ggcccgccctg
240

gcattatgcc cagtacatga ccttatggga ctttcctact tggcagtaca tctacgtatt
300

agtcatcgct attaccatgg tcgagggtgag ccccacgttc tgcttcactc tccccatctc
360

ccccccctcc ccacccccaa ttttgtatatt atttatTTTT taattatTTT gtgcagcgat
420

gggggcgggg ggggggggggg gcgcgcgccca ggcggggcg ggcggggcga ggggcggggc
480

ggggcgaggc ggagagggtgc ggcggcagcc aatcagagcg gcgcgctccg aaagtttcct
540

tttatggcga ggcggcgggcg gcggcgggccc tataaaaagc gaagcgcgcg gcgggcggga
600

gtcgctgcgc gctgccttcg ccccgctgcc cgctccgccg ccgcctcgcg ccgcccgcc
660

cggtcttgac tgaccgcgtt actaaaacag gtaagtccgg cctccgcgcc gggTTTTggc
720

gcctcccgcg ggcgcccccc tcctcacggc gagcgctgcc acgtcagacg aagggcgag
780

cgagcgctcct gatccttcg cccggacgct caggacagcg gcccgctgct cataagactc
840

ggccttagaa cccagtatc agcagaagga catTTtagga cgggacttgg gtgactctag
900

ggcactgggtt ttctttccag agagcggaac aggcgaggaa aagtagtccc ttctcggcga
960

ttctgcggag ggatctccgt ggggcggtga acgccgatga tgcctctact aacctgttc
1020

atgTTTTctt tttttttcta caggtcctgg gtgacgaaca g
1061

<210> 64

<211> 2630

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-013 transfer genome

<400> 64

tagggagggtc ctgcacgtta cataacttac ggtaaattggc ccgcctgggt gaccgccc
60

cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc caatagggac
120

tttccattga cgtcaatggg tggagtattt acggtaaact gccacttgg cagtacatca
180

agtgtatcat atgccaagta cggcccctat tgacgtcaat gacggtaaata ggccgcctg
240

gcattatgcc cagtacatga ccttatggga ctttcctact tggcagtaca tctacgtatt
300

agtcacgct attaccatgg tcgagggtgag cccacggtc tgcttcactc tccccatctc
360

ccccccctcc ccacccccaa ttttgtattt atttattttt taattatttt gtgcagcgat
420

gggggcgggg gggggggggg gcgcgcgcca ggcggggcg ggcggggcga ggggcggggc
480

ggggcgaggg ggagaggtgc ggcggcagcc aatcagagcg gcgcgctccg aaagtttcct
540

tttatggcga ggcggcgggc gcggcgggcc tataaaaagc gaagcgcgcg gcgggcggga
600

gtcgctgcgc gctgccttcg ccccggtgcc cgctccgccc ccgcctcgcg ccgcccggcc
660

cggctctgac tgaccgcgtt actaaaacag gtaagtccgg cctccgcgcc gggttttggc
720

gcctcccgcg ggcgcccccc tcctcacggc gagcgctgcc acgtcagacg aagggcgcag
780

cgagcgtcct gatccttccg cccggacgct caggacagcg gcccgctgct cataagactc
840

ggccttagaa ccccagtatc agcagaagga catttttagga cgggacttgg gtgactctag
900

ggcactgggtt ttctttccag agagcggaac aggcgaggaa aagtagtccc ttctcggcga
960

ttctgcggag ggatctccgt ggggcggtga acgccgatga tgcctctact aacctgttc
1020

atgttttctt tttttttcta caggtcctgg gtgacgaaca ggccaccatg tccaccgctg
1080

tgctggagaa ccctgggctg gggaggaaac tgtcagactt cgggcaggag acttcataca
1140

ttgaggataa ctgtaaccag aatggcgcca tctctctgat cttcagcctg aaggaggaag
1200

tgggcgccct ggcaaagggtg ctgcgccctgt ttgaggagaa cgacgtgaat ctgaccacaca
1260

tcgagtcccg gccttctaga ctgaagaagg acgagtacga gttctttacc cacctggata
1320

agcggtcctt gccagccctg acaaacatca tcaagatcct gaggcacgac atcggagcaa
1380

ccgtgcacga gctgtctcgg gacaagaaga aggataccgt gccctgggttc cctcggacaa
1440

tccaggagct ggatagattt gcccaaccaga tcctgtctta cggagcagag ctggacgcag
1500

atcacctgg cttcaaggac ccagtgtatc gggcccggag aaagcagttt gccgatatcg
1560

cctacaatta taggcacgga cagccaatcc ctcgcgtgga gtatatggag gaggagaaga
1620

agacctgggg cacagtgttc aagaccctga agagcctgta caagacacac gcctgctacg
1680

agtataacca catcttcccc ctgctggaga agtattgtgg ctttcacgag gacaatatcc
1740

ctcagctgga ggacgtgagc cagttcctgc agacctgcac aggctttagg ctgaggccag
1800

tggcaggact gctgagctcc cgggacttcc tgggaggact ggccttcaga gtgtttcact
1860

gcacccagta catcaggcac ggctccaagc caatgtatac accagagccc gacatctgtc
1920

acgagctgct gggccacgtg cccctgttta gcgatagatc cttcgcccag ttttcccagg
1980

agatcggact ggcattctctg ggagcacctg acgagtacat cgagaagctg gccaccatct
2040

attggttcac agtggagttt ggcctgtgca agcagggcga tagcatcaag gcctacggag
2100

caggactgct gtctagcttc ggcgagctgc agtattgtct gtccgagaag ccaaagctgc
2160

tgcccctgga gctggagaag accgccatcc agaactacac cgtgacagag ttccagcccc
2220

tgtactatgt ggccgagtct tttaacgatg ccaaggagaa ggtgagaaat ttcgccgcca
2280

caatccctag gcccttcagc gtgcggtacg acccttatac ccagaggatc gaggtgctgg
2340

ataatacaca gcagctgaag atcctggctg actcaatcaa tagcgaaatc ggaatcctgt
2400

gctccgccct gcagaaaatc aaatgaatcg tagatccaga catgataaga tacattgatg
2460

agtttggaca aaccacaact agaatgcagt gaaaaaaaaatg ctttatttgt gaaatttgtg
2520

atgctattgc tttatttgta accattataa gctgcaataa acaagttaac aacaacaatt
2580

gcattcattt tatgtttcag gttcaggggg aggtgtggga ggtttttttaa
2630

<210> 65
<211> 3011
<212> DNA
<213> Artificial Sequence

<220>
<223> pHMI-hPAH-TC-013 transfer genome

<400> 65
ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggtcgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttccctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcacg ttacataact tacggtaaact ggcccgcctg gctgaccgcc
240

caacgacccc cgcccattga cgtcaataat gacgtatggt cccatagtaa cgccaatagg
300

gactttccat tgacgtcaat ggggtggagta ttacggtaa actgcccact tggcagtaca
360

tcaagtgtat catatgccaa gtacgcccc tattgacgtc aatgacggta aatggcccgc
420

ctggcattat gcccagtaca tgaccttatg ggactttcct acttggcagt acatctacgt
480

attagtcatc gctattacca tggtcgaggt gagccccacg ttctgcttca ctctcccat
540

ctcccccccc tccccacccc caattttgta ttattttatt ttttaattat tttgtgcagc
600

gatggggggcg gggggggggg ggggcgcgcg ccaggcgggg cggggcgggg cgaggggcgg
660

ggcggggcga ggcggagagg tgcggcggca gccaatcaga gcggcgcgct ccgaaagttt
720

ccttttatgg cgaggcggcg gcggcggcgg ccctataaaa agcgaagcgc gcggcgggcg
780

ggagtcgctg cgcgctgcct tcgccccgtg ccccgctccg ccgccgcctc gcgccgcccg
840

ccccggctct gactgaccgc gttactaaaa caggtaagtc cggcctccgc gccgggtttt
900

ggcgcctccc gcgggcgccc ccctcctcac ggcgagcgct gccacgtcag acgaagggcg
960

cagcgagcgt cctgatacct cgcgccggac gctcaggaca gcggcccgtc gctcataaga
1020

ctcggcctta gaaccccagt atcagcagaa ggacatttta ggacgggact tgggtgactc
1080

tagggcactg gttttctttc cagagagcgg aacaggcgag gaaaagtagt cccttctcgg
1140

cgattctgcg gagggatctc cgtggggcgg tgaacgccga tgatgcctct actaaccatg
1200

ttcatgtttt cttttttttt ctacagggtc tgggtgacga acaggccacc atgtccaccg
1260

ctgtgctgga gaaccctggg ctgggggagga aactgtcaga cttcgggcag gagacttcat
1320

acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc ctgaaggagg
1380

aagtgggcgc cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg aatctgaccc
1440

acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt acccacctgg
1500

ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac gacatcggag
1560

caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg ttccctcgga
1620

caatccagga gctggataga tttgcccaacc agatcctgtc ttacggagca gagctggacg
1680

cagatcaccc tggcttcaag gacccagtgt atcggggcccg gagaaagcag tttgccgata
1740

tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg gaggaggaga
1800

agaagacctg gggcacagtg ttcaagaccc tgaagagcct gtacaagaca cacgcctgct
1860

acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac gaggacaata
1920

tccctcagct ggaggacgtg agccagttec tgcagacctg cacaggcttt aggctgaggc
1980

cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc agagtgtttc
2040

actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag cccgacatct
2100

gtcacgagct gctggggccac gtgcccctgt ttagcgatag atccttcgcc cagttttccc
2160

aggagatcgg actggcatct ctgggagcac ctgacgagta catcgagaag ctggccacca
2220

tctattgggt cacagtggag tttggcctgt gcaagcaggg cgatagcatc aaggcctacg
2280

gagcaggact gctgtctagc ttcggcgagc tgcagtattg tctgtccgag aagccaaagc
2340

tgctgcccct ggagctggag aagaccgcca tccagaacta caccgtgaca gagttccagc
2400

ccctgtacta tgtggccgag tcttttaacg atgccaagga gaagggtgaga aatttcgccg
2460

ccacaatccc taggcccttc agcgtgcggt acgaccctta taccagagg atcgagggtgc
2520

tgataataac acagcagctg aagatcctgg ctgactcaat caatagcgaa atcggaatcc
2580

tgtgctccgc cctgcagaaa atcaaataaa tcgtagatcc agacatgata agatacattg
2640

atgagtttgg acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt
2700

gtgatgctat tgctttatatt gtaaccatta taagctgcaa taaacaagtt aacaacaaca
2760

attgcattca ttttatgttt caggttcagg gggagggtgtg ggagggttttt taagcttggt
2820

taaacgtacg tagataagta gcatggcggg ttaatcatta actacaagga acccctagt
2880

atggagttgg ccactccctc tctgcgcgct cgctcgctca ctgaggccgg gcgaccaaag
2940

gtcgcccgac gcccgggctt tgcccgggcg gcctcagtga gcgagcgagc gcgagagag
3000

g g a g t g g c c a a
3011

<210> 66

<211> 913

<212> DNA

<213> Artificial Sequence

<220>

<223> hAAT promoter region

<400> 66

tagggaggtc ctgcacagaa ggggaggagg gggcagcagc tgtctgacca ctgttggtct
60

tgcaacttgt gtccccaggt taatttttaa aaagcagtca aaagtccaag tggcccttgg
120

cagcatttac tctctctgtt tgctctggtt aataatctca ggagcacaaa cattcctgga
180

ggcaggagaa gaaatcaaca tcctggactt atcctctggg cctctcccca ccccaggat
240

tgtaactgaa atgcttcact ggtgctcctt ttgttttaag gcattggatc ttcataagcta
300

ctgatcgtgc ccaagcacac agtatctgca gcaaccactt aggcctccag gaatgtggtg
360

accattgacc ctaattcatt ccccttcatt gatcctatgt aaccatcctc caaaaagagc
420

tttcgcaaac tcaaataaac acaggaaagg aagaccttct tatctttgag agtatatgtt
480

tagccctata gctctaacct actctgatct cccagggcgg cagtaagtct tcagcatcag
540

gcattttggg gtgactcagt aaatggtaga tcttgctacc agtggaacag ccactaagga
600

ttctgcagtg agagcagagg gccagctaag tgggtactctc ccagagactg tctgactcac
660

gccacccct ccaccttgga cacaggacgc tgtggtttct gagccaggta caatgactcc
720

tttcggtaag tgcagtggaa gctgtacact gcccaggcaa agcgtccggg cagcgtaggc
780

gggcgactca gatcccagcc agtggactta gccctgttt gctcctccga taactggggt
840

gaccttggtt aatattcacc agcagcctcc cccgttgccc ctctggatcc actgcttaaa
900

t a c g g a c g a g g a c
913

<210> 67

<211> 2600

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-017 transfer genome

<400> 67

tagggagggtc ctgcacagaa ggggaggagg gggcagcagc tgtctgacca ctgttggtct
60

tgcaacttgt gtccccaggt taatttttaa aaagcagtca aaagtccaag tggcccttgg
120

cagcatttac tctctctggt tgctctgggt aataatctca ggagcacaaa cattcctgga
180

ggcaggagaa gaaatcaaca tcctggactt atcctctggg cctctcccca cccccaggat
240

tgtaactgaa atgcttcact ggtgctcctt ttgttttaag gcattggatc ttcatagcta
300

ctgatcgtgc ccaagcacac agtatctgca gcaaccactt aggcctccag gaatgtggtg
360

accattgacc ctaattcatt ccccttcatt gatcctatgt aaccatcctc caaaaagagc
420

tttcgcaaac tcaaataaac acaggaaagg aagaccttct tatctttgag agtatatggt
480

tagccctata gctctaaccc actctgatct cccagggcgg cagtaagtct tcagcatcag
540

gcattttggg gtgactcagt aaatggtaga tcttgctacc agtggaacag ccactaagga
600

ttctgcagtg agagcagagg gccagctaag tgggtactctc ccagagactg tctgactcac
660

gccaccccct ccaccttgga cacaggacgc tgttggtttct gagccaggta caatgactcc
720

tttcggtaag tgcagtggaa gctgtacact gcccaggcaa agcgtccggg cagcgtaggc
780

gggcgactca gatcccagcc agtggactta gcccctgttt gctcctccga taactggggt
840

gaccttggtt aatattcacc agcagcctcc cccgttgccc ctctggatcc actgcttaaa
900

tacggacgag gacgctagcg ccggcaagag gtaagggttt aagggatggg tggttggtgg
960

ggtattaatg ttttaattacc tggagcacct gcctgaaatc actttttttc aggttggtta
1020

attaaggatc cgccaccatg tccaccgctg tgctggagaa ccctgggctg gggaggaaac
1080

tgtcagactt cgggcaggag acttcataca ttgaggataa ctgtaaccag aatggcgcca
1140

tctctctgat cttcagcctg aaggaggaag tgggcgccct ggcaaagggtg ctgcgcctgt
1200

ttgaggagaa cgacgtgaat ctgaccacaca tcgagtcccg gccttctaga ctgaagaagg
1260

acgagtacga gttctttacc cacctggata agcggtcctt gccagccctg acaaacatca
1320

tcaagatcct gaggcacgac atcggagcaa ccgtgcacga gctgtctcgg gacaagaaga
1380

aggataccgt gccctgggtc cctcggacaa tccaggagct ggatagattt gccaaccaga
1440

tcctgtctta cggagcagag ctggacgcag atcacccctgg cttcaaggac ccagtgtatc
1500

gggcccggag aaagcagttt gccgatatcg cctacaatta taggcacgga cagccaatcc
1560

ctcgcgtgga gtatatggag gaggagaaga agacctgggg cacagtgttc aagaccctga
1620

agagcctgta caagacacac gcctgctacg agtataacca catcttcccc ctgctggaga
1680

agtattgtgg ctttcacgag gacaatatcc ctcagctgga ggacgtgagc cagttcctgc
1740

agacctgcac aggcctttagg ctgaggccag tggcaggact gctgagctcc cgggacttcc
1800

tgggaggact ggccttcaga gtgtttcact gcacccagta catcaggcac ggctccaagc
1860

caatgtatac accagagccc gacatctgtc acgagctgct gggccacgtg cccctgttta
1920

gcgatagatc cttcgcccag ttttcccagg agatcggact ggcatctctg ggagcacctg
1980

acgagtacat cgagaagctg gccaccatct attggttcac agtggagttt ggctgtgca
2040

agcagggcga tagcatcaag gcctacggag caggactgct gtctagcttc ggcgagctgc
2100

agtattgtct gtccgagaag ccaaagctgc tgcccctgga gctggagaag accgccatcc
2160

agaactacac cgtgacagag ttccagcccc tgtactatgt ggccgagtct tttaacgatg
2220

ccaaggagaa ggtgagaaat ttcgccgcca caatccctag gcccttcagc gtgcggtacg
2280

acccttatac ccagaggatc gaggtgctgg ataatacaca gcagctgaag atcctggctg
2340

actcaatcaa tagcgaaatc ggaatcctgt gctccgccct gcagaaaatc aaatgaatcg
2400

tagatccaga catgataaga tacattgatg agtttggaca aaccacaact agaatgcagt
2460

gaaaaaaatg ctttatttgt gaaatttgtg atgctattgc tttatttgta accattataa
2520

gctgcaataa acaagttaac aacaacaatt gcattcattt tatgtttcag gttcaggggg
2580

a g g t g t g g g a g g t t t t t t a a
2600

<210> 68
<211> 2981
<212> DNA
<213> Artificial Sequence

<220>
<223> pHMI-hPAH-TC-017 transfer genome

<400> 68
ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggtcgcc
60

cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg
120

gccaactcca tcactagggg ttccctggagg ggtggagtcg tgacgtgaat tacgtcatag
180

ggttagggag gtcctgcaca gaaggggagg agggggcagc agctgtctga ccactgttgg
240

tcttgcaact tgtgtcccca gggttaatttt taaaaagcag tcaaaagtcc aagtggccct
300

tggcagcatt tactctctct gtttgctctg gttaataatc tcaggagcac aaacattcct
360

ggaggcagga gaagaaatca acatcctgga cttatcctct gggcctctcc ccacccccag
420

gattgtaact gaaatgcttc actggtgctc cttttgtttt aaggcattgg atcttcatag
480

ctactgatcg tgcccaagca cacagtatct gcagcaacca cttaggcctc caggaatgtg
540

gtgaccattg accctaattc attccccttc atggatccta tgtaaccatc ctccaaaaag
600

agctttcgca aactcaaata aacacaggaa aggaagacct tcttatcttt gagagtatat
660

gtttagccct atagctctaa cccactctga tctcccaggg cggcagtaag tcttcagcat
720

caggcatttt ggggtgactc agtaaattggt agatcttgct accagtggaa cagccactaa
780

ggattctgca gtgagagcag agggccagct aagtgggtact ctcccagaga ctgtctgact
840

cacgccaccc cctccacctt ggacacagga cgctgtgggtt tctgagccag gtacaatgac
900

tcctttcggg aagtgcagtg gaagctgtac actgcccagg caaagcgtcc gggcagcgta
960

ggcgggcgac tcagatccca gccagtggac ttagcccctg tttgctcctc cgataactgg
1020

ggtgaccttg gttaatatc accagcagcc tccccggtg cccctctgga tccactgctt
1080

aaatacggac gaggacgcta gcgccggcaa gaggtaaggg ttttaagggat gggttggttg
1140

tgggggtatta atgtttaatt acctggagca cctgcctgaa atcacttttt ttcagggttg
1200

ttaattaagg atccgccacc atgtccaccg ctgtgctgga gaaccctggg ctggggagga
1260

aactgtcaga cttcgggcag gagacttcat acattgagga taactgtaac cagaatggcg
1320

ccatctctct gatcttcagc ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc
1380

tgtttgagga gaacgacgtg aatctgacct acatcgagtc ccggccttct agactgaaga
1440

aggacgagta cgagttcttt acccacctgg ataagcggtc cctgccagcc ctgacaaaca
1500

tcatcaagat cctgaggcac gacatcggag caaccgtgca cgagctgtct cgggacaaga
1560

agaaggatac cgtgccctgg ttccctcgga caatccagga gctggataga tttgcccaacc
1620

agatcctgtc ttacggagca gagctggacg cagatcaccc tggcttcaag gacccagtgt
1680

atcgggcccg gagaaagcag tttgccgata tcgcctacaa ttataggcac ggacagccaa
1740

tccctcgcgt ggagtatatg gaggaggaga agaagacctg gggcacagtg ttcaagaccc
1800

tgaagagcct gtacaagaca cacgcctgct acgagtataa ccacatcttc cccctgctgg
1860

agaagtattg tggctttcac gaggacaata tccctcagct ggaggacgtg agccagttcc
1920

tgcagacctg cacaggcttt aggctgaggc cagtggcagg actgctgagc tcccgggact
1980

tcctgggagg actggccttc agagtgtttc actgcaccca gtacatcagg cacggctcca
2040

agccaatgta tacaccagag cccgacatct gtcacgagct gctgggccac gtgcccctgt
2100

ttagcgatag atccttcgcc cagttttccc aggagatcgg actggcatct ctgggagcac
2160

ctgacgagta catcgagaag ctggccacca tctattggtt cacagtggag tttggcctgt
2220

gcaagcaggg cgatagcatc aaggcctacg gagcaggact gctgtctagc ttcggcgagc
2280

tgcagtattg tctgtccgag aagccaaagc tgctgcccct ggagctggag aagaccgcca
2340

tccagaacta caccgtgaca gagttccagc ccctgtacta tgtggccgag tcttttaacg
2400

atgccaagga gaaggtgaga aatttcgccg ccacaatccc taggcccttc agcgtgcggt
2460

acgaccctta taccagagg atcgaggtgc tggataatac acagcagctg aagatcctgg
2520

ctgactcaat caatagcgaa atcggaatcc tgtgctccgc cctgcagaaa atcaaatgaa
2580

tcgtagatcc agacatgata agatacattg atgagtttgg acaaaccaca actagaatgc
2640

agtgaaaaaa atgcttttatt tgtgaaattt gtgatgctat tgctttatatt gtaaccatta
2700

taagctgcaa taaacaagtt aacaacaaca attgcattca ttttatgttt caggttcagg
2760

gggaggtgtg ggagggttttt taagcttggt taaacgtacg tagataagta gcatggcggg
2820

ttaatcatta actacaagga acccctagtg atggagttgg ccactccctc tctgcgcgct
2880

cgctcgctca ctgaggccgg gcgaccaaag gtcgcccgcg gcccgggctt tgccccggcg
2940

gcctcagtg gcgagcgagc gcgcagagag ggagtggcca a
2981

<210> 69

<211> 1359

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-018 PAH sequence

<400> 69

atgtccactg ctgtgctgga gaaccaggc ctgggcagga agttgagtga ctttgggcag
60

gagacctcct acatagaaga caattgcaat cagaatgggg ccactctctt gatcttcagc
120

cttaaagagg aggtgggtgc tctggcaaaa gtgctcagac tctttgagga gaatgatgtg
180

aatctcaccc acattgaatc caggcccagc agactcaaaa aggatgaata tgaattcttc
240

accacttgg	acaaaaggtc	cttacctgcc	cttacaata	tcatcaaat	tttgagacat
300					
gacatagggg	caactgtaca	tgaactgagt	agagataaaa	aaaaagacac	agtcccctgg
360					
ttccccagga	ccatccagga	attggacagg	tttgcaaacc	agatactgag	ctatggtgct
420					
gagttggatg	ctgatcaccc	aggcttcaag	gaccctgtgt	acagagcaag	gagaaagcag
480					
tttgctgaca	ttgcctacaa	ttacaggcac	ggccaaccca	ttcctagagt	cgagtacatg
540					
gaagaagaga	agaaaacctg	gggcactgtc	ttcaagaccc	tgaagtcact	gtacaagaca
600					
catgcctgct	atgaatacaa	ccacatattt	ccactcctag	agaaatactg	tggattccat
660					
gaggacaata	taccccaatt	ggaggatgtc	tcacagtttc	tgcagacttg	tacaggtttt
720					
aggctgaggc	cagtggctgg	gcttctcagc	agcagggact	tcctgggtgg	actggccttc
780					
agggtgtttc	actgtacaca	atacatcaga	catggtagca	aaccaatgta	tactcctgaa
840					
ccagacatct	gccatgagct	gcttgggcat	gtgcctctgt	tttcagacag	gtcctttgct
900					
cagttctcac	aagagattgg	gctagcttca	ctgggagctc	cagatgagta	tattgaaaaa
960					
ctggcaacaa	tttactgggt	tacagtggag	tttggacttt	gtaagcaggg	agactccatc
1020					
aaggcctatg	gtgcaggatt	gttgtcttcc	tttggggaac	tgcaatattg	tctctctgaa
1080					
aagcctaagt	tgctaccact	ggagcttgag	aagactgcca	ttcagaacta	cacagtgact
1140					

gaattccagc ccctctacta tgttgcagag tctttcaatg atgccaagga gaaggtagg
1200

aactttgctg caacaatccc cagacctttc agtgtgaggt atgacccta cactcagaga
1260

attgaagttc tggataacac ccagcagctg aaaattctgg cagatagtat caactctgag
1320

attggaatcc tgtgttctgc cctgcagaag atcaagtga
1359

<210> 70

<211> 1359

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-019 PAH sequence

<400> 70

atgtccactg ctgtgctgga gaaccaggc ctgggcagga agttgagtga ctttgggcag
60

gagacctcct acatagaaga caattgcaat cagaatgggg ccatctctct gatcttcagc
120

cttaaagagg aggtgggtgc tctggcaaaa gtgctcagac tgtttgagga gaatgatgtg
180

aatctcaccc acattgaatc caggcccagc agactcaaaa aggatgaata tgaattcttc
240

accacttgg acaaaaggtc cttacctgcc cttacaaata tcatcaaaat tttgagacat
300

gacatagggg caactgtaca tgaactgagt agagataaaa aaaaagacac agtcccctgg
360

ttccccagga ccatccagga attggacagg ttgcaaacc agatactgag ctatggtgct
420

gagttggatg ctgatcaccc aggcttcaag gaccctgtgt acagagcaag gagaaagcag
480

tttgctgaca ttgcctacaa ttacaggcac ggccaaccca ttcctagagt cgagtacatg
540

gaagaagaga agaaaacctg gggcactgtc ttcaagaccc tgaagtcact gtacaagaca
600

catgcctgct atgaatacaa ccacatattt ccactcctag agaaatactg tggattccat
660

gaggacaata taccccaatt ggaggatgtc tcacagtttc tgcagacttg tacaggtttt
720

aggctgaggc cagtggctgg gctcctcagc agcagggact tcctgggtgg actggccttt
780

cgagttttcc actgtactca gtatatcaga catggctcca agcctatgta taccccagaa
840

cctgacatct gccatgaact gcttgggcat gtgcctctct tttcagaccg ttcctttgcc
900

cagttttctc aggagattgg actagccagc ctaggtgcac cagatgagta cattgagaag
960

ttagcaacca tttactgggt cacagtggag ttcggtctct gcaagcaagg ggactcaata
1020

aaggcctatg gagcaggcct cctgtcaagt tttggagaac tccaatactg cctatctgag
1080

aagcctaaat tattaccctt ggaactagaa aaaactgcaa tacagaacta cacagtgact
1140

gagtttcagc cactctacta tgtggcagag tcctttaatg atgccaaaga aaagggtccga
1200

aattttgctg caacaattcc caggcccttc tctgttcgct atgatccata cacccaaaga
1260

attgaagtcc tagataacac ccagcagctg aaaatcctgg cagacagtat caactctgaa
1320

attggaatcc tctgttctgc cctgcagaag atcaagtga
1359

<210> 71
<211> 1359
<212> DNA
<213> Artificial Sequence

<220>
<223> pHMI-hPAH-TC-020 PAH sequence

<400> 71
atgtccactg ctgtgctgga gaacccaggc ctgggcagga agttgagtga ctttgggcag
60

gagacctcct acatagaaga caattgcaat cagaatgggg ccatctctct gatcttcagc
120

cttaaagagg aggtgggtgc tctggcaaaa gtgctcagac tctttgagga gaatgatgtg
180

aatctcaccc acattgaatc caggcccagc agactcaaaa aggatgaata tgaattcttc
240

accacttgg acaaaaggtc cttacctgcc cttacaaata tcatcaaaat tttgagacat
300

gacatagggg caactgtaca tgaactgagt agagataaaa aaaaagacac agtcccctgg
360

ttccccagga ccatccagga attggacagg tttgcaaacc agatactgag ctatggtgct
420

gagttggatg ctgatcaccc aggcttcaag gaccctgtgt acagagcaag gagaaagcag
480

tttgctgaca ttgcctacaa ttacaggcac ggccaaccca ttcctagagt cgagtacatg
540

gaagaagaga agaaaacctg gggcactgtc ttcaagaccc tgaagtcact gtacaagaca
600

catgcctgct atgaatacaa ccacatattt cactcctag agaaatactg tggattccat
660

gaggacaata taccccaatt ggaggatgtc tcacagtttc tgcagacttg tacaggtttt
720

aggctgaggc cagtggctgg gcttctcagc agcagggact tcctgggtgg actggccttc
780

agggtgtttc actgtacaca atacatcaga catggtagca aaccaatgta tactcctgaa
840

ccagacatct gccatgagct gcttgggcat gtgcctctgt tcagcgacag aagctttgct
900

cagtttagcc aggagattgg gctggccagc ctgggcgccc ctgatgagta tatcgagaaa
960

ctggccacaa tctactgggtt cacagtggag ttcggcctgt gcaagcaggg cgactcaatc
1020

aaggcctatg gcgccggcct gctgagcagc ttcggcgaac tgcagtactg cctgagcgag
1080

aagcccaagc tgctgccact ggagctggag aaaaccgcca tccagaacta cacagtgaca
1140

gagttccagc ctctgtacta tgtggccgag agcttcaacg atgccaagga gaaggtagg
1200

aattttgccg ccactatccc caggcctttc tccgtgagat atgacccta caccagcga
1260

atcgaggtgc tggacaatac ccagcagctg aagatcctgg ccgattccat caactctgag
1320

atcggcattc tgtgtagcgc cctgcagaag attaagtga
1359

<210> 72

<211> 1359

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-021 PAH sequence

<400> 72

atgtccactg ctgtgctgga gaaccaggc ctgggaagga agctgagtga ctttggccag
60

gagacctcct acatagagga caactgcaat cagaacgggg ccatcagcct gatcttcagc
120

cttaaagagg aggtaggcgc tctggcaaaa gtgctcagac tctttgagga gaatgatgtg
180

aatctcaccc acattgaatc caggcccagc agactcaaaa aggatgaata tgaatttttc
240

accacttgg acaaaaggtc cttacctgcc cttacaaata tcatcaaaat ttgagacat
300

gacatagggg caactgtaca tgaactgagt agagataaaa aaaaagacac agtcccctgg
360

ttccccagga ccatacagga attggacagg ttgcaaacc agatactgag ctatggtgct
420

gaattggatg ctgatcaccc aggcttcaag gaccctgtgt acagagcacg aagaaagcag
480

tttgctgaca ttgcctacaa ttacaggcac ggccaacca ttcctagagt cgagtacatg
540

gaagaagaaa agaaaacctg gggcactgtg ttcaagacct tgaagtcact gtacaagaca
600

catgcctgct atgaatacaa ccacatattt ccactcctag agaaatactg tggattccat
660

gaggacaaca taccccaatt ggaggatgtg tcacagtttc tgcagacttg tacaggtttt
720

aggctgaggc cagtggcagg gcttctcagc agcagggact tcctgggtgg actggccttc
780

agggtgtttc actgtacaca gtacatcaga catggtagca aaccaatgta tactcctgaa
840

ccagacatct gccatgagct gcttgggcat gtgcctctgt tttcagacag gtcctttgct
900

caattctcac aggagattgg actagccagc ctaggtgcac cagatgagta cattgagaag
960

ttagcaacca tttactgggtt cacagtggag ttcggccttt gcaagcaagg ggactcaata
1020

aaggcctatg gagcaggcct cctgtcaagt tttggagaac tacaatactg cctatctgag
1080

aagcctaaat tattaccctt ggaactagag aaaactgcaa tacagaacta cacagtgact
1140

gagtttcagc cactctacta tgtggccgag tccttcaatg atgccaaaga aaagggtccga
1200

aattttgctg caacaattcc caggcctttc tctgttcgct atgataccta cacccaaaga
1260

attgaagtcc tagataacac ccagcagctg aagatcctgg ctgatagcat aaacagcgaa
1320

attggaatcc tctgttctgc cctgcagaag atcaagtga
1359

<210> 73

<211> 1359

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-022 PAH sequence

<400> 73

atgtccactg ctgtgctgga gaaccaggc ctgggcagga agttgagtga ctttgggcag
60

gagacctcct acatagaaga caattgcaat cagaatgggg ccatctctct gatcttcagc
120

cttaaagagg aggtgggtgc tctggcaaaa gtgctcagac tctttgagga gaatgatgtg
180

aatctcaccc acattgaatc caggcccagc agactcaaaa aggatgaata tgaattcttc
240

accacttgg acaaaaggtc cttacctgcc cttacaaata tcatcaaaat tttgagacat
300

gacatagggg caactgtaca tgaactgagt agagataaaa aaaaagacac agtcccctgg
360

ttccccagga ccatccagga attggacagg tttgcaaacc agatactgag ctatgggtgct
420

gagttggatg ctgatcaccc aggcttcaag gaccctgtgt acagagcaag gagaaagcag
480

tttgctgaca ttgcctacaa ttacaggcac ggccaaccca ttcctagagt cgagtacatg
540

gaagaagaga agaaaacctg gggcactgtc ttcaagacct tgaagtcact gtacaagaca
600

catgcctgct atgaatacaa ccacatattt ccactcctag agaaatactg tggattccat
660

gaggacaata taccccaatt ggaggatgtc tcacagtttc tgcagacttg tacaggtttt
720

aggctgaggc cagtggctgg gcttctcagc agcagggact tcctgggtgg actggccttt
780

cgagtttttc actgtactca gtatatcaga catggctcca agcccatgta taccccagaa
840

cctgacatct gccatgaact gcttgggcat gtgcctctgt tttcagaccg ttcctttgcc
900

cagttttctc aggagattgg actagccagc ctaggtgcac cagatgagta cattgagaag
960

ttagcaacca tttactgggt cacagtggag ttcggccttt gcaagcaagg ggactcaata
1020

aaggcctatg gagcaggcct cctgtcaagt tttggagAAC tacaatactg cctatctgag
1080

aagcctaaat tattaccctt ggaactagaa aaaactgcaa tacagaacta cacagtgact
1140

gagtttcagc cactctacta tgtggcagag tcctttaatg atgccaaaga aaagggtccga
1200

aattttgctg caacaattcc caggcctttc tctgttcgct atgatccata cacccaaaga
1260

attgaagtcc tagataacac ccagcagctg aaaatcctgg cagacagcat caactctgaa
1320

attggaatcc tctgtttctgc cctgcagaag atcaagtga
1359

<210> 74

<211> 2111

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-018 transfer genome

<400> 74

ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc
60

tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc
120

cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt
180

agtgtgagag gggaatgact cctttcggta agtgcagtgg aagctgtaca ctgcccaggc
240

aaagcgtccg ggcagcgtag gcgggcgact cagatcccag ccagtggact tagcccctgt
300

ttgctcctcc gataactggg gtgaccttgg ttaatattca ccagcagcct cccccgttgc
360

ccctctggat ccactgctta aatacggacg aggacagggc cctgtctcct cagcttcagg
420

caccaccact gacctgggac agtgaatcct ctaaggtaaa tataaaattt ttaagtgtat
480

aatgtgttaa actactgatt ctaattgttt ctctctttta gattccaacc tttggaactg
540

accgccacca tgtccactgc tgtgctggag aaccaggcc tgggcaggaa gttgagtac
600

tttgggcagg agacctccta catagaagac aattgcaatc agaatggggc catctctctg
660

atcttcagcc ttaaagagga ggtgggtgct ctggcaaaag tgctcagact ctttgaggag
720

aatgatgtga atctcaccca cattgaatcc aggcccagca gactcaaaaa ggatgaatat
780

gaattcttca cccacttgga caaaagggtcc ttacctgccc ttacaaatat catcaaaatt
840

ttgagacatg acataggggc aactgtacat gaactgagta gagataaaaa aaaagacaca
900

gtcccctggg tccccaggac catccaggaa ttggacaggt ttgcaaacca gatactgagc
960

tatgggtgctg agttggatgc tgatcaccca ggcttcaagg accctgtgta cagagcaagg
1020

agaaagcagt ttgctgacat tgcctacaat tacaggcacg gccaacccat tcctagagtc
1080

gagtacatgg aagaagagaa gaaaacctgg ggcactgtct tcaagaccct gaagtcactg
1140

tacaagacac atgcctgcta tgaatacaac cacatatctc cactcctaga gaaatactgt
1200

ggattccatg aggacaatat accccaattg gaggatgtct cacagtttct gcagacttgt
1260

acaggtttta ggctgaggcc agtggctggg cttctcagca gcagggactt cctgggtgga
1320

ctggccttca ggggtgtttca ctgtacacaa tacatcagac atggtagcaa accaatgtat
1380

actcctgaac cagacatctg ccatgagctg cttgggcatg tgcctctgtt ttcagacagg
1440

tcctttgctc agttctcaca agagattggg ctagcttcac tgggagctcc agatgagtat
1500

attgaaaaac tggcaacaat ttactgggtt acagtggagt ttggactttg taagcagggg
1560

gactccatca aggcctatgg tgcaggattg ttgtcttcct ttggggaact gcaatattgt
1620

ctctctgaaa agcctaagtt gctaccactg gagcttgaga agactgccat tcagaactac
1680

acagtgactg aattccagcc cctctactat gttgcagagt ctttcaatga tgccaaggag
1740

aaggttagga actttgctgc aacaatcccc agacctttca gtgtgaggta tgaccctac
1800

actcagagaa ttgaagttct ggataacacc cagcagctga aaattctggc agatagtatc
1860

aactctgaga ttggaatcct gtgttctgcc ctgcagaaga tcaagtgact cgagatccag
1920

acatgataag atacattgat gagtttggac aaaccacaac tagaatgcag tgaaaaaat
1980

gctttatttg tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata
2040

aacaagttaa caacaacaat tgcattcatt ttatgtttca gggttcagggg gaggtgtggg
2100

a g g t t t t t t a a
2111

<210> 75

<211> 2426

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-018 transfer genome

<400> 75

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt
60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggaatt cacgcgtgga
120

tctgaattca attcacgcgt ggtacctccc taaaatgggc aaacattgca agcagcaaac
180

agcaaacaca cagccctccc tgcctgctga ccttgagagct ggggcagagg tcagagacct
240

ctctgggccc atgccacctc caacatccac tcgacctctt ggaatttcgg tggagaggag
300

cagaggttgt cctggcgtgg tttaggtagt gtgagagggg aatgactcct ttcggtaagt
360

gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc agcgtaggcg ggcgactcag
420

atcccagcca gtggacttag cccctgtttg ctctccgat aactgggggtg accttggtta
480

atattcacca gcagcctccc ccgttgcccc tctggatcca ctgcttaaata acggacgagg
540

acagggccct gtctcctcag cttcaggcac caccactgac ctgggacagt gaatcctcta
600

aggtaaatat aaaattttta agtgtataat gtgttaaact actgattcta attgtttctc
660

tcttttagat tccaaccttt ggaactgacc gccaccatgt cactgctgt gctggagaac
720

ccaggcctgg gcaggaagtt gagtgacttt gggcaggaga cctcctacat agaagacaat
780

tgcaatcaga atggggccat ctctctgac ttcagcctta aagaggaggt gggtgctctg
840

gcaaaagtgc tcagactctt tgaggagaat gatgtgaatc tcaccacat tgaatccagg
900

cccagcagac 960	tcaaaaagga	tgaatatgaa	ttcttcaccc	acttggacaa	aaggtcctta
cctgccctta 1020	caaatatcat	caaaattttg	agacatgaca	taggggcaac	tgtacatgaa
ctgagtagag 1080	ataaaaaaaaa	agacacagtc	ccctgggttc	ccaggaccat	ccaggaattg
gacaggtttg 1140	caaaccagat	actgagctat	ggtgctgagt	tggatgctga	tcaccaggc
ttcaaggacc 1200	ctgtgtacag	agcaaggaga	aagcagtttg	ctgacattgc	ctacaattac
aggcacggcc 1260	aaccatttc	tagagtcgag	tacatggaag	aagagaagaa	aacctggggc
actgtcttca 1320	agaccctgaa	gtcactgtac	aagacacatg	cctgctatga	atacaaccac
atatttccac 1380	tcctagagaa	atactgtgga	ttccatgagg	acaatatacc	ccaattggag
gatgtctcac 1440	agtttctgca	gacttgtaca	ggttttaggc	tgaggccagt	ggctgggctt
ctcagcagca 1500	gggacttcct	gggtggactg	gccttcaggg	tgtttcactg	tacacaatac
atcagacatg 1560	gtagcaaacc	aatgtatact	cctgaaccag	acatctgcca	tgagctgctt
gggcatgtgc 1620	ctctgttttc	agacagggtc	tttgctcagt	tctcacaaga	gattgggcta
gcttcactgg 1680	gagctccaga	tgagtatatt	gaaaaactgg	caacaattta	ctggttttaca
gtggagtttg 1740	gactttgtaa	gcagggagac	tccatcaagg	cctatgggtgc	aggattgttg
tcttcctttg 1800	gggaactgca	atattgtctc	tctgaaaagc	ctaagttgct	accactggag

cttgagaaga ctgccattca gaactacaca gtgactgaat tccagcccct ctactatggt
1860

gcagagtctt tcaatgatgc caaggagaag gttaggaact ttgctgcaac aatccccaga
1920

cctttcagtg tgaggatatga cccctacact cagagaattg aagttctgga taacaccag
1980

cagctgaaaa ttctggcaga tagtatcaac tctgagattg gaatcctgtg ttctgccctg
2040

cagaagatca agtgactcga gatccagaca tgataagata cattgatgag tttggacaaa
2100

ccacaactag aatgcagtg aaaaaatgct ttatttgtga aatttgtgat gctattgctt
2160

tatttgtaac cattataagc tgcaataaac aagttaacaa caacaattgc attcatttta
2220

tgtttcaggt tcagggggag gtgtgggagg ttttttaaag catgctgggg agagatcgat
2280

ctgaggaacc cctagtgatg gagttggcca ctccctctct gcgctcgc tcgctcactg
2340

aggccgggcg accaaaggct gcccgacgcc cgggctttgc ccgggcggcc tcagtgagcg
2400

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
2426

<210> 76

<211> 2111

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-019 transfer genome

<400> 76

ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc
60

tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc
120

cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt
180

agtgtgagag gggaatgact cctttcggta agtgcagtgg aagctgtaca ctgcccaggc
240

aaagcgtccg ggcagcgtag gcgggcgact cagatcccag ccagtggact tagcccctgt
300

ttgctcctcc gataactggg gtgaccttgg ttaatatcca ccagcagcct cccccgttgc
360

ccctctggat ccactgctta aatacggacg aggacagggc cctgtctcct cagcttcagg
420

caccaccact gacctgggac agtgaatcct ctaaggtaaa tataaaattt ttaagtgtat
480

aatgtgttaa actactgatt ctaattgttt ctctctttta gattccaacc tttggaactg
540

accgccacca tgtccactgc tgtgctggag aaccagggc tgggcaggaa gttgagtgac
600

tttgggcagg agacctccta catagaagac aattgcaatc agaatggggc catctctctg
660

atcttcagcc ttaaagagga ggtgggtgct ctggcaaaag tgctcagact gtttgaggag
720

aatgatgtga atctcaccca cattgaatcc aggcccagca gactcaaaaa ggatgaatat
780

gaattcttca cccacttgga caaaagggtcc ttacctgccc ttacaaatat catcaaaatt
840

ttgagacatg acataggggc aactgtacat gaactgagta gagataaaaa aaaagacaca
900

gtcccctggg tccccaggac catccaggaa ttggacaggt ttgcaaacca gatactgagc
960

tatggtgctg agttggatgc tgatcaccca ggcttcaagg accctgtgta cagagcaagg
1020

agaaagcagt ttgctgacat tgcctacaat tacaggcacg gccaacccat tcctagagtc
1080

gagtacatgg aagaagagaa gaaaacctgg ggcaactgtct tcaagaccct gaagtcactg
1140

tacaagacac atgcctgcta tgaatacaac cacatatattc cactcctaga gaaatactgt
1200

ggattccatg aggacaatat accccaattg gaggatgtct cacagtttct gcagacttgt
1260

acaggtttta ggctgaggcc agtggctggg ctcctcagca gcagggactt cctgggtgga
1320

ctggcctttc gagttttcca ctgtactcag tataatcagac atggctccaa gcctatgtat
1380

accccagaac ctgacatctg ccatgaactg cttggggcatg tgcctctctt ttcagaccgt
1440

tcctttgccc agtttttctca ggagattgga ctagccagcc taggtgcacc agatgagtac
1500

attgagaagt tagcaaccat ttactgggtc acagtggagt tcggtctctg caagcaaggg
1560

gactcaataa aggcctatgg agcaggcctc ctgtcaagtt ttggagaact ccaatactgc
1620

ctatctgaga agcctaaatt attacccttg gaactagaaa aaactgcaat acagaactac
1680

acagtgactg agtttcagcc actctactat gtggcagagt cctttaatga tgccaaagaa
1740

aagggtccgaa attttgctgc aacaattccc aggcccttct ctgttcgcta tgatccatac
1800

acccaaagaa ttgaagtcct agataacacc cagcagctga aaatcctggc agacagtatc
1860

aactctgaaa ttggaatcct ctgttctgcc ctgcagaaga tcaagtgact cgagatccag
1920

acatgataag atacattgat gagtttggac aaaccacaac tagaatgcag tgaaaaaaat
1980

gctttatttg tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata
2040

aacaagttaa caacaacaat tgcattcatt ttatgtttca gggttcagggg gaggtgtggg
2100

a g g t t t t t t a a
2111

<210> 77

<211> 2426

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-019 transfer genome

<400> 77

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgctc ggcgaccttt
60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggaatt cacgcgtgga
120

tctgaattca attcacgcgt ggtacctccc taaaatgggc aaacattgca agcagcaaac
180

agcaaacaca cagccctccc tgccctgctga ccttgagact ggggcagagg tcagagacct
240

ctctgggccc atgccacctc caacatccac tcgacctt ggaatttcgg tggagaggag
300

cagaggttgt cctggcggtg tttaggtagt gtgagagggg aatgactcct ttcggtaagt
360

gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc agcgtaggcg ggcgactcag
420

atcccagcca 480	gtggacttag	cccctgtttg	ctcctccgat	aactgggggtg	accttggtta
atattcacca 540	gcagcctccc	ccgttgcccc	tctggatcca	ctgcttaaata	acggacgagg
acagggccct 600	gtctcctcag	cttcaggcac	caccactgac	ctgggacagt	gaatcctcta
aggtaaata 660	aaaattttta	agtgtataat	gtgttaaact	actgattcta	attgtttctc
tcttttagat 720	tccaaccttt	ggaactgacc	gccaccatgt	ccactgctgt	gctggagaac
ccaggcctgg 780	gcaggaagtt	gagtgacttt	gggcaggaga	cctcctacat	agaagacaat
tgcaatcaga 840	atggggccat	ctctctgac	ttcagcctta	aagaggaggt	gggtgctctg
gcaaaagtgc 900	tcagactggt	tgaggagaat	gatgtgaatc	tcaccacat	tgaatccagg
cccagcagac 960	tcaaaaagga	tgaatatgaa	ttcttcaccc	acttggaaca	aaggtcctta
cctgccctta 1020	caaatatcat	caaaattttg	agacatgaca	taggggcaac	tgtacatgaa
ctgagtagag 1080	ataaaaaaaaa	agacacagtc	ccctgggttc	ccaggacat	ccaggaattg
gacaggtttg 1140	caaaccagat	actgagctat	ggtgctgagt	tggatgctga	tcaccaggc
ttcaaggacc 1200	ctgtgtacag	agcaaggaga	aagcagtttg	ctgacattgc	ctacaattac
aggcacggcc 1260	aaccatttc	tagagtcgag	tacatggaag	aagagaagaa	aacctggggc
actgtcttca 1320	agaccctgaa	gtcactgtac	aagacacatg	cctgctatga	atacaaccac

atatttccac 1380	tcctagagaa	atactgtgga	ttccatgagg	acaatatacc	ccaattggag
gatgtctcac 1440	agtttctgca	gacttgtaca	ggtttttaggc	tgaggccagt	ggctgggctc
ctcagcagca 1500	gggacttcct	gggtggactg	gcctttcgag	ttttccactg	tactcagtat
atcagacatg 1560	gctccaagcc	tatgtatacc	ccagaacctg	acatctgcca	tgaactgctt
gggcatgtgc 1620	ctctcttttc	agaccgttcc	tttgcccagt	tttctcagga	gattggacta
gccagcctag 1680	gtgcaccaga	tgagtacatt	gagaagttag	caaccattta	ctggttcaca
gtggagttcg 1740	gtctctgcaa	gcaaggggac	tcaataaagg	cctatggagc	aggcctcctg
tcaagttttg 1800	gagaactcca	atactgccta	tctgagaagc	ctaaattatt	acccttggaa
ctagaaaaaa 1860	ctgcaataca	gaactacaca	gtgactgagt	ttcagccact	ctactatgtg
gcagagtcct 1920	ttaatgatgc	caaagaaaag	gtccgaaatt	ttgctgcaac	aattcccagg
cccttctctg 1980	ttcgctatga	tccatacacc	caaagaattg	aagtcctaga	taacaccag
cagctgaaaa 2040	tcctggcaga	cagtatcaac	tctgaaattg	gaatcctctg	ttctgccctg
cagaagatca 2100	agtgactcga	gatccagaca	tgataagata	cattgatgag	tttggacaaa
ccacaactag 2160	aatgcagtga	aaaaaatgct	ttattttgtga	aattttgtgat	gctattgctt
tatttgtaac 2220	cattataagc	tgcaataaac	aagttaacaa	caacaattgc	attcatttta

tgtttcaggt tcagggggag gtgtgggagg ttttttaaag catgctgggg agagatcgat
2280

ctgaggaacc cctagtgatg gagttggcca ctccctctct gcgcgctcgc tcgctcactg
2340

aggccgggcg accaaaggtc gcccgcgcgc cgggcctttgc ccgggcggcc tcagtgcgcg
2400

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
2426

<210> 78

<211> 2111

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-020 transfer genome

<400> 78

ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc
60

tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc
120

cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt
180

agtgtgagag gggaatgact cctttcggta agtgcagtgg aagctgtaca ctgcccaggc
240

aaagcgtccg ggcagcgtag gcgggcgact cagatcccag ccagtggact tagcccctgt
300

ttgctcctcc gataactggg gtgaccttgg ttaatatcca ccagcagcct cccccgttgc
360

ccctctggat ccactgctta aatacggacg aggacagggc cctgtctcct cagcttcagg
420

caccaccact gacctgggac agtgaatcct ctaaggtaaa tataaaattt ttaagtgtat
480

aatgtgttaa actactgatt ctaattgttt ctctctttta gattccaacc tttggaactg
540

accgccacca tgtccactgc tgtgctggag aaccaggcc tgggcaggaa gttgagtac
600

tttgggcagg agacctccta catagaagac aattgcaatc agaatggggc catctctctg
660

atcttcagcc ttaaagagga ggtgggtgct ctggcaaaag tgctcagact ctttgaggag
720

aatgatgtga atctcaccca cattgaatcc aggcccagca gactcaaaaa ggatgaatat
780

gaattcttca cccacttgga caaaagggtcc ttacctgccc ttacaaatat catcaaaatt
840

ttgagacatg acataggggc aactgtacat gaactgagta gagataaaaa aaaagacaca
900

gtcccctggg tccccaggac catccaggaa ttggacaggt ttgcaaacca gatactgagc
960

tatgggtgctg agttggatgc tgatcaccca ggcttcaagg accctgtgta cagagcaagg
1020

agaaagcagt ttgctgacat tgcctacaat tacaggcacg gccaacccat tcctagagtc
1080

gagtacatgg aagaagagaa gaaaacctgg ggcactgtct tcaagaccct gaagtcactg
1140

tacaagacac atgcctgcta tgaatacaac cacatatctc cactcctaga gaaatactgt
1200

ggattccatg aggacaatat accccaattg gaggatgtct cacagtttct gcagacttgt
1260

acaggtttta ggctgaggcc agtggctggg cttctcagca gcagggactt cctgggtgga
1320

ctggccttca ggggtgtttca ctgtacacaa tacatcagac atggtagcaa accaatgtat
1380

actcctgaac cagacatctg ccatgagctg cttgggcatg tgcctctgtt cagcgacaga
1440

agctttgctc agtttagcca ggagattggg ctggccagcc tgggcgcccc tgatgagtat
1500

atcgagaaac tggccacaat ctactgggtc acagtggagt tcggcctgtg caagcagggc
1560

gactcaatca aggcctatgg cgccggcctg ctgagcagct tcggcgaaact gcagtactgc
1620

ctgagcgaga agcccaagct gctgccactg gagctggaga aaaccgccat ccagaactac
1680

acagtgacag agttccagcc tctgtactat gtggccgaga gcttcaacga tgccaaggag
1740

aaggtgagga attttgccgc cactatcccc aggcctttct ccgtgagata tgaccctac
1800

accagcgaa tcgaggtgct ggacaatacc cagcagctga agatcctggc cgattccatc
1860

aactctgaga tcggcattct gtgtagcgcc ctgcagaaga ttaagtgact cgagatccag
1920

acatgataag atacattgat gagtttggac aaaccacaac tagaatgcag tgaaaaaat
1980

gctttatttg tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata
2040

aacaagttaa caacaacaat tgcattcatt ttatgtttca gggttcagggg gaggtgtggg
2100

a g g t t t t t a a
2111

<210> 79

<211> 2426

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-020 transfer genome

<400> 79

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgctcg ggcgaccttt
60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggaatt cacgcgtgga
120

tctgaattca attcacgcgt ggtacctccc taaaatgggc aaacattgca agcagcaaac
180

agcaaacaca cagccctccc tgcctgctga ccttggagct ggggcagagg tcagagacct
240

ctctgggccc atgccacctc caacatccac tcgacctt ggaatttcgg tggagaggag
300

cagaggttgt cctggcgctgg tttaggtagt gtgagagggg aatgactcct ttcggtaagt
360

gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc agcgtaggcg ggcgactcag
420

atcccagcca gtggacttag cccctgtttg ctccctccgat aactgggggtg accttggtta
480

atattcacca gcagcctccc ccgttgcccc tctggatcca ctgcttaaata acggacgagg
540

acagggccct gtctcctcag cttcaggcac caccactgac ctgggacagt gaatcctcta
600

aggtaaatat aaaattttta agtgtataat gtgttaaact actgattcta attgtttctc
660

tcttttagat tccaaccttt ggaactgacc gccaccatgt ccactgctgt gctggagAAC
720

ccaggcctgg gcaggaagtt gagtgacttt gggcaggaga cctcctacat agaagacaat
780

tgcaatcaga atggggccat ctctctgatc ttcagcctta aagaggaggt ggggtgctctg
840

gcaaaagtgc tcagactctt tgaggagaat gatgtgaatc tcacccacat tgaatccagg
900

cccagcagac tcaaaaagga tgaatatgaa ttcttcaccc acttggacaa aaggtcctta
960

cctgccctta caaatatcat caaaattttg agacatgaca taggggcaac tgtacatgaa
1020

ctgagtagag ataaaaaaaa agacacagtc ccctgggttc ccaggaccat ccaggaattg
1080

gacaggtttg caaaccagat actgagctat ggtgctgagt tggatgctga tcacccaggc
1140

ttcaaggacc ctgtgtacag agcaaggaga aagcagtttg ctgacattgc ctacaattac
1200

aggcacggcc aaccatttc tagagtcgag tacatggaag aagagaagaa aacctggggc
1260

actgtcttca agaccctgaa gtcactgtac aagacacatg cctgctatga atacaaccac
1320

atatttccac tcctagagaa atactgtgga ttccatgagg acaatatacc ccaattggag
1380

gatgtctcac agtttctgca gacttgtaca ggttttaggc tgaggccagt ggctgggctt
1440

ctcagcagca gggacttcct ggggtggactg gccttcaggg tgtttcactg tacacaatac
1500

atcagacatg gtagcaaacc aatgtatact cctgaaccag acatctgcca tgagctgctt
1560

gggcatgtgc ctctgttcag cgacagaagc tttgctcagt ttagccagga gattgggctg
1620

gccagcctgg gcgcccctga tgagtatatc gagaaactgg ccacaatcta ctggttcaca
1680

gtggagttcg gcctgtgcaa gcagggcgac tcaatcaagg cctatggcgc cggcctgctg
1740

agcagcttcg gcgaactgca gtactgcctg agcgagaagc ccaagctgct gccactggag
1800

ctggagaaaa ccgccatcca gaactacaca gtgacagagt tccagcctct gtactatgtg
1860

gccgagagct tcaacgatgc caaggagaag gtgaggaatt ttgccgccac tatccccagg
1920

cctttctccg tgagatatga cccctacacc cagcgaatcg aggtgctgga caatacccag
1980

cagctgaaga tcctggccga ttccatcaac tctgagatcg gcattctgtg tagcgccctg
2040

cagaagatta agtgactcga gatccagaca tgataagata cattgatgag tttggacaaa
2100

ccacaactag aatgcagtga aaaaaatgct ttattttgtga aattttgtgat gctattgctt
2160

tattttgtaac cattataagc tgcaataaac aagttaacaa caacaattgc attcatttta
2220

tgtttcaggt tcaggggggag gtgtggggagg ttttttaaag catgctgggg agagatcgat
2280

ctgaggaacc cctagtgatg gagttggcca ctccctctct gcgcgctcgc tcgctcactg
2340

aggccgggcg accaaaggtc gcccgacgcc cgggctttgc ccgggcggcc tcagtgagcg
2400

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
2426

<210> 80

<211> 2111

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-021 transfer genome

<400> 80

ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc
60

tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc
120

cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt
180

agtgtgagag gggaatgact cctttcggta agtgcagtgg aagctgtaca ctgcccaggc
240

aaagcgtccg ggcagcgtag gcgggcgact cagatcccag ccagtggact tagcccctgt
300

ttgctcctcc gataactggg gtgaccttgg ttaatatcca ccagcagcct cccccgttgc
360

ccctctggat ccactgctta aatacggacg aggacagggc cctgtctcct cagcttcagg
420

caccaccact gacctgggac agtgaatcct ctaaggtaaa tataaaattt ttaagtgtat
480

aatgtgttaa actactgatt ctaattgttt ctctctttta gattccaacc tttggaactg
540

accgccacca tgtccactgc tgtgctggag aaccacaggc tgggaaggaa gctgagtgac
600

tttggccagg agacctccta catagaggac aactgcaatc agaacggggc catcagcctg
660

atcttcagcc ttaaagagga ggtaggcgct ctggcaaaag tgctcagact ctttgaggag
720

aatgatgtga atctcaccca cattgaatcc aggcccagca gactcaaaaa ggatgaatat
780

gaatttttca cccacttgga caaaagggtc ttacctgccc ttacaaatat catcaaaatt
840

ttgagacatg acataggggc aactgtacat gaactgagta gagataaaaa aaaagacaca
900

gtccccctggt 960	tccccaggac	catacaggaa	ttggacaggt	ttgcaaacca	gatactgagc
tatgggtgctg 1020	aattggatgc	tgatcaccca	ggcttcaagg	accctgtgta	cagagcacga
agaaagcagt 1080	ttgctgacat	tgcctacaat	tacaggcacg	gccaacccat	tcctagagtc
gagtacatgg 1140	aagaagaaaa	gaaaacctgg	ggcactgtgt	tcaagaccct	gaagtcactg
tacaagacac 1200	atgcctgcta	tgaatacaac	cacatatattc	cactcctaga	gaaatactgt
ggattccatg 1260	aggacaacat	accccaattg	gaggatgtgt	cacagtttct	gcagacttgt
acaggtttta 1320	ggctgaggcc	agtggcaggg	cttctcagca	gcagggactt	cctgggtgga
ctggccttca 1380	gggtgtttca	ctgtacacag	tacatcagac	atggtagcaa	accaatgtat
actcctgaac 1440	cagacatctg	ccatgagctg	cttgggcatg	tgcctctggt	ttcagacagg
tcctttgctc 1500	aattctcaca	ggagattgga	ctagccagcc	taggtgcacc	agatgagtac
attgagaagt 1560	tagcaaccat	ttactgggtc	acagtggagt	tcggcctttg	caagcaaggg
gactcaataa 1620	aggcctatgg	agcaggcctc	ctgtcaagtt	ttggagaact	acaatactgc
ctatctgaga 1680	agcctaaatt	attacccttg	gaactagaga	aaactgcaat	acagaactac
acagtgactg 1740	agtttcagcc	actctactat	gtggccgagt	ccttcaatga	tgccaaagaa
aagggtccgaa 1800	attttgctgc	aacaattccc	aggcctttct	ctgttcgcta	tgatccttac

acccaaagaa ttgaagtcct agataacacc cagcagctga agatcctggc tgatagcata
1860

aacagcgaaa ttggaatcct ctgttctgcc ctgcagaaga tcaagtgact cgagatccag
1920

acatgataag atacattgat gagtttggac aaaccacaac tagaatgcag tgaaaaaaat
1980

gctttatttg tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata
2040

aacaagttaa caacaacaat tgcattcatt ttatgtttca gggttcagggg gaggtgtggg
2100

a g g t t t t t a a
2111

<210> 81

<211> 2426

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-021 transfer genome

<400> 81

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgctc ggcgaccttt
60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggaatt cacgcgtgga
120

tctgaattca attcacgcgt ggtacctccc taaaatgggc aaacattgca agcagcaaac
180

agcaaacaca cagccctccc tgccctgctga ccttggagct ggggcagagg tcagagacct
240

ctctgggccc atgccacctc caacatccac tcgacccctt ggaatttcgg tggagaggag
300

cagaggttgt cctggcgtgg tttaggtagt gtgagagggg aatgactcct ttcggtaagt
360

gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc agcgtaggcg ggcgactcag
420

atcccagcca gtggacttag cccctgtttg ctcctccgat aactgggggtg accttggtta
480

atattcacca gcagcctccc ccgttgcccc tctggatcca ctgcttaaata acggacgagg
540

acagggccct gtctcctcag cttcaggcac caccactgac ctgggacagt gaatcctcta
600

aggtaaataat aaaattttta agtgtataat gtgttaaact actgattcta attgtttctc
660

tcttttagat tccaaccttt ggaactgacc gccaccatgt ccactgctgt gctggagAAC
720

ccaggcctgg gaaggaagct gagtgacttt ggccaggaga cctcctacat agaggacaac
780

tgcaatcaga acgggggcat cagcctgac ttcagcctta aagaggaggt aggcgctctg
840

gcaaaagtgc tcagactctt tgaggagaat gatgtgaatc tcaccacat tgaatccagg
900

cccagcagac tcaaaaagga tgaatatgaa tttttcacc acttggaaca aaggtcctta
960

cctgccctta caaatatcat caaaattttg agacatgaca taggggcaac tgtacatgaa
1020

ctgagtagag ataaaaaaaa agacacagtc ccctgggttc ccaggacat acaggaattg
1080

gacaggtttg caaaccagat actgagctat ggtgctgaat tggatgctga tcaccaggc
1140

ttcaaggacc ctgtgtacag agcacgaaga aagcagtttg ctgacattgc ctacaattac
1200

aggcacggcc aaccattcc tagagtcgag tacatggaag aagaaaagaa aacctggggc
1260

actgtgtttca agaccctgaa gtcactgtac aagacacatg cctgctatga atacaaccac
1320

atattttccac tcctagagaa atactgtgga ttccatgagg acaacatacc ccaattggag
1380

gatgtgtcac agttttctgca gacttgtaca ggtttttaggc tgaggccagt ggcagggcctt
1440

ctcagcagca gggacttcct ggggtggactg gccttcaggg tgtttcactg tacacagtac
1500

atcagacatg gtagcaaacc aatgtatact cctgaaccag acatctgcca tgagctgctt
1560

gggcatgtgc ctctgttttc agacaggtec ttgctcaat tctcacagga gattggacta
1620

gccagcctag gtgcaccaga tgagtacatt gagaagttag caaccattta ctggttcaca
1680

gtggagttcg gcctttgcaa gcaaggggac tcaataaagg cctatggagc aggcctcctg
1740

tcaagttttg gagaactaca atactgccta tctgagaagc ctaaattatt acccttggaa
1800

ctagagaaaa ctgcaataca gaactacaca gtgactgagt ttcagccact ctactatgtg
1860

gccgagtcct tcaatgatgc caaagaaaag gtccgaaatt ttgctgcaac aattcccagg
1920

ccttttctctg ttcgctatga tccttacacc caaagaattg aagtcctaga taacaccag
1980

cagctgaaga tcctggctga tagcataaac agcgaaattg gaatcctctg ttctgccctg
2040

cagaagatca agtgactcga gatccagaca tgataagata cattgatgag ttggacaaa
2100

ccacaactag aatgcagtga aaaaaatgct ttattttgtga aatttgtgat gctattgctt
2160

tatttgtaac cattataagc tgcaataaac aagttaacaa caacaattgc attcatttta
2220

tgtttcaggt tcagggggag gtgtgggagg ttttttaaag catgctgggg agagatcgat
2280

ctgaggaacc cctagtgatg gagttggcca ctccctctct gcgcgctcgc tcgctcactg
2340

aggccgggcg accaaaggtc gcccgcgccc cgggctttgc ccgggcggcc tcagtgcgcg
2400

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
2426

<210> 82

<211> 2111

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-022 transfer genome

<400> 82

ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc
60

tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc
120

cactcgaccc cttggaattt cggtggagag gagcagaggt tgtcctggcg tggtttaggt
180

agtgtgagag gggaatgact cctttcggta agtgcagtgg aagctgtaca ctgcccaggc
240

aaagcgtccg ggcagcgtag gcgggcgact cagatcccag ccagtggact tagcccctgt
300

ttgctcctcc gataactggg gtgaccttgg ttaatatcca ccagcagcct cccccgttgc
360

ccctctggat ccactgctta aatacggacg aggacagggc cctgtctcct cagcttcagg
420

caccaccact 480	gacctgggac	agtgaatcct	ctaaggtaaa	tataaaattt	ttaagtgtat
aatgtgttaa 540	actactgatt	ctaattgttt	ctctctttta	gattccaacc	tttggaactg
accgccacca 600	tgtccactgc	tgtgctggag	aaccaggcc	tgggcaggaa	gttgagtgc
tttgggcagg 660	agacctccta	catagaagac	aattgcaatc	agaatggggc	catctctctg
atcttcagcc 720	ttaaagagga	ggtgggtgct	ctggcaaaag	tgctcagact	ctttgaggag
aatgatgtga 780	atctcaccca	cattgaatcc	aggcccagca	gactcaaaaa	ggatgaatat
gaattcttca 840	cccacttgga	caaaagggtcc	ttacctgccc	ttacaaatat	catcaaaatt
ttgagacatg 900	acataggggc	aactgtacat	gaactgagta	gagataaaaa	aaaagacaca
gtcccctggc 960	tccccaggac	catccaggaa	ttggacaggt	ttgcaaacca	gatactgagc
tatggtgctg 1020	agttggatgc	tgatcaccca	ggcttcaagg	accctgtgta	cagagcaagg
agaaagcagt 1080	ttgctgacat	tgcttacaat	tacaggcacg	gccaacccat	tcctagagtc
gagtacatgg 1140	aagaagagaa	gaaaacctgg	ggcactgtct	tcaagaccct	gaagtcactg
tacaagacac 1200	atgcctgcta	tgaatacaac	cacatatttc	cactcctaga	gaaatactgt
ggattccatg 1260	aggacaatat	acccaattg	gaggatgtct	cacagtttct	gcagacttgt
acaggtttta 1320	ggctgaggcc	agtggctggg	cttctcagca	gcagggactt	cctgggtgga

ctggcctttc gagttttcca ctgtactcag tataatcagac atggctccaa gcccatgtat
1380

accccagaac ctgacatctg ccatgaactg cttgggcatg tgcctctgtt ttcagaccgt
1440

tcctttgccc agtttttctca ggagattgga ctagccagcc taggtgcacc agatgagtac
1500

attgagaagt tagcaaccat ttactgggtc acagtggagt tcggcctttg caagcaaggg
1560

gactcaataa aggcctatgg agcaggcctc ctgtcaagtt ttggagaact acaatactgc
1620

ctatctgaga agcctaaatt attacccttg gaactagaaa aaactgcaat acagaactac
1680

acagtgactg agtttcagcc actctactat gtggcagagt cctttaatga tgccaaagaa
1740

aaggtccgaa attttgctgc aacaattccc aggcctttct ctgttcgcta tgatccatac
1800

acccaaagaa ttgaagtcct agataacacc cagcagctga aaatcctggc agacagcatc
1860

aactctgaaa ttggaatcct ctgttctgcc ctgcagaaga tcaagtgact cgagatccag
1920

acatgataag atacattgat gagtttggac aaaccacaac tagaatgcag tgaaaaaaat
1980

gctttatttg tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata
2040

aacaagttaa caacaacaat tgcattcatt ttatgtttca gggttcagggg gaggtgtggg
2100

a g g t t t t t t a a
2111

<211> 2426

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-022 transfer genome

<400> 83

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt
60

ggtcgccccg cctcagtgag cgagcgagcg cgcagagagg gagtggaatt cacgcgtgga
120

tctgaattca attcacgctt ggtacctccc taaaatgggc aaacattgca agcagcaaac
180

agcaaacaca cagccctccc tgcctgctga ccttggagct ggggcagagg tcagagacct
240

ctctgggccc atgccacctc caacatccac tcgacctcctt ggaatttcgg tggagaggag
300

cagaggttgt cctggcgtgg tttaggtagt gtgagagggg aatgactcct ttcggtaagt
360

gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc agcgtaggcg ggcgactcag
420

atcccagcca gtggacttag cccctgtttg ctccctccgat aactgggggtg accttggtta
480

atattcacca gcagcctccc ccgttgcccc tctggatcca ctgcttaaata acggacgagg
540

acagggccct gtctcctcag cttcaggcac caccactgac ctgggacagt gaatcctcta
600

aggtaaatat aaaattttta agtgtataat gtgttaaact actgattcta attgtttctc
660

tcttttagat tccaaccttt ggaactgacc gccaccatgt ccactgctgt gctggagAAC
720

ccaggcctgg gcaggaagtt gagtgacttt gggcaggaga cctcctacat agaagacaat
780

tgcaatcaga atggggccat ctctctgac ttcagcctta aagaggaggt gggtgctctg
840

gcaaaagtgc tcagactctt tgaggagaat gatgtgaatc tcaccacat tgaatccagg
900

cccagcagac tcaaaaagga tgaatatgaa ttcttcaccc acttggaaca aaggtcctta
960

cctgccctta caaatatcat caaaattttg agacatgaca taggggcaac tgtacatgaa
1020

ctgagtagag ataaaaaaaa agacacagtc ccctgggtcc ccaggaccat ccaggaattg
1080

gacaggtttg caaaccagat actgagctat ggtgctgagt tggatgctga tcaccaggc
1140

ttcaaggacc ctgtgtacag agcaaggaga aagcagtttg ctgacattgc ctacaattac
1200

aggcacggcc aaccattcc tagagtcgag tacatggaag aagagaagaa aacctggggc
1260

actgtcttca agaccctgaa gtcactgtac aagacacatg cctgctatga atacaaccac
1320

atatttccac tcctagagaa atactgtgga ttccatgagg acaatatacc ccaattggag
1380

gatgtctcac agtttctgca gacttgtaca ggttttaggc tgaggccagt ggctgggctt
1440

ctcagcagca gggacttcct gggtggaactg gcctttcgag ttttccactg tactcagtat
1500

atcagacatg gctccaagcc catgtatacc ccagaacctg acatctgcca tgaactgctt
1560

gggcatgtgc ctctgttttc agaccgttcc tttgcccagt tttctcagga gattggacta
1620

gccagcctag gtgcaccaga tgagtacatt gagaagttag caaccattta ctgggttcaca
1680

gtggagttcg gcctttgcaa gcaaggggac tcaataaagg cctatggagc aggcctcctg
1740

tcaagttttg gagaactaca atactgccta tctgagaagc ctaaattatt acccttgga
1800

ctagaaaaaa ctgcaataca gaactacaca gtgactgagt ttcagccact ctactatgtg
1860

gcagagtcct ttaatgatgc caaagaaaag gtccgaaatt ttgctgcaac aattcccagg
1920

cctttctctg ttcgctatga tccatacacc caaagaattg aagtcctaga taacaccag
1980

cagctgaaaa tcctggcaga cagcatcaac tctgaaattg gaatcctctg ttctgccctg
2040

cagaagatca agtgactcga gatccagaca tgataagata cattgatgag tttggacaaa
2100

ccacaactag aatgcagtga aaaaaatgct ttattttgtga aatttgtgat gctattgctt
2160

tatttgtaac cattataagc tgcaataaac aagttaacaa caacaattgc attcatttta
2220

tgtttcaggt tcaggggggag gtgtggggagg ttttttaaag catgctgggg agagatcgat
2280

ctgaggaacc cctagtgatg gagttggcca ctccctctct gcgcgctcgc tcgctcactg
2340

aggccgggcg accaaaggtc gcccgacgcc cgggctttgc ccgggcgggc tcagtgagcg
2400

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
2426

<210> 84

<211> 2111

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-023 transfer genome

<400> 84

ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc
60

tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc
120

cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt
180

agtgtgagag gggaatgact cctttcggta agtgcagtgg aagctgtaca ctgcccaggc
240

aaagcgtccg ggcagcgtag gcgggcgact cagatcccag ccagtggact tagcccctgt
300

ttgctcctcc gataactggg gtgaccttgg ttaatatcca ccagcagcct cccccgttgc
360

ccctctggat ccactgctta aatacggacg aggacagggc cctgtctcct cagcttcagg
420

caccaccact gacctgggac agtgaatcct ctaaggtaaa tataaaattt ttaagtgtat
480

aatgtgttaa actactgatt ctaattgttt ctctctttta gattccaacc tttggaactg
540

accgccacca tgtccactgc ggtcctggaa aaccacaggct tgggcaggaa actctctgac
600

tttggacagg aaacaagcta tattgaagac aactgcaatc aaaatggtgc catatcactg
660

atcttctcac tcaaagaaga agttggtgca ttggccaaag tattgcgctt atttgaggag
720

aatgatgtaa acctgaccca cattgaatct agaccttctc gtttaaagaa agatgagtat
780

gaatttttca cccatttgga taaacgtagc ctgcctgctc tgacaaacat catcaagatc
840

ttgaggcatg acattggtgc cactgtccat gagctttcac gagataagaa gaaagacaca
900

gtgccctggg tcccaagaac cattcaagag ctggacagat ttgccaatca gattctcagc
960

tatggagcgg aactggatgc tgaccaccct ggtttttaag atcctgtgta ccgtgcaaga
1020

cggaagcagt ttgctgacat tgcctacaac taccgccatg ggcagcccat ccctcgagtg
1080

gaatacatgg aggaagaaaa gaaaacatgg ggcacagtgt tcaagactct gaagtccttg
1140

tataaaaccc atgcttgcta tgagtacaat cacatttttc cacttcttga aaagtactgt
1200

ggcttccatg aagataacat tccccagctg gaagacgttt ctcaattcct gcagacttgc
1260

actggtttcc gcctccgacc tgtggctggc ctgctttcct ctcgggattt cttgggtggc
1320

ctggccttcc gagtcttcca ctgcacacag tacatcagac atggatccaa gcccatgtat
1380

acccccgaac ctgacatctg ccatgagctg ttgggacatg tgcccttggt ttcagatcgc
1440

agctttgccc agttttccca ggaaattggc cttgcctctc tgggtgcacc tgatgaatac
1500

attgaaaagc tcgccacaat ttactggttt actgtggagt ttgggctctg caaacaagga
1560

gactccataa aggcataatg tgctgggctc ctgtcatcct ttggtgaatt acagtactgc
1620

ttatcagaga agccaaagct tctccccctg gagctggaga agacagccat ccaaaattac
1680

actgtcacgg agttccagcc cctgtattac gtggcagaga gttttaatga tgccaaggag
1740

aaagtaagga actttgctgc cacaatacct cggcccttct cagttcgcta cgacccatac
1800

acccaaagga ttgaggtctt ggacaatacc cagcagctta agattttggc tgattccatt
1860

aacagtgaaa ttggaatcct ttgcagtgcc ctccagaaaa taaagtaact cgagatccag
1920

acatgataag atacattgat gagtttggac aaaccacaac tagaatgcag tgaaaaaaat
1980

gctttatttg tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata
2040

aacaagttaa caacaacaat tgcattcatt ttatgtttca gggttcagggg gaggtgtggg
2100

a g g t t t t t t a a
2111

<210> 85

<211> 2426

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-023 transfer genome

<400> 85

ctgcgcgctc gctcgcctcac tgaggccgcc cgggcaaagc ccgggcgctc ggcgaccttt
60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggaatt cacgcgtgga
120

tctgaattca attcacgcgt ggtacctccc taaaatgggc aaacattgca agcagcaaac
180

agcaaacaca cagccctccc tgcctgctga ccttggagct ggggcagagg tcagagacct
240

ctctgggccc atgccacctc caacatccac tcgacccctt ggaatttcgg tggagaggag
300

cagaggttgt cctggcgtgg tttaggtagt gtgagagggg aatgactcct ttcggtaagt
360

gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc agcgtaggcg ggcgactcag
420

atcccagcca gtggacttag cccctgtttg ctcctccgat aactgggggtg accttggtta
480

atattcacca gcagcctccc ccgttgcccc tctggatcca ctgcttaaata acggacgagg
540

acagggccct gtctcctcag cttcaggcac caccactgac ctgggacagt gaatcctcta
600

aggtaaataa aaaattttta agtgtataat gtgttaaact actgattcta attgtttctc
660

tcttttagat tccaaccttt ggaactgacc gccaccatgt ccactgcggt cctggaaaac
720

ccaggcttgg gcaggaaact ctctgacttt ggacaggaaa caagctatat tgaagacaac
780

tgcaatcaaa atggtgccat atcactgac ttctcactca aagaagaagt tgggtgcattg
840

gccaaagtat tgcgcttatt tgaggagaat gatgtaaacc tgaccacat tgaatctaga
900

ccttctcggt taaagaaaga tgagtatgaa tttttcacc atttgataa acgtagcctg
960

cctgctctga caaacatcat caagatcttg aggcattgaca ttggtgccac tgtccattgag
1020

ctttcacgag ataagaagaa agacacagt ccctgggtcc caagaaccat tcaagagctg
1080

gacagatttg ccaatcagat tctcagctat ggagcggaac tggatgctga ccaccctggg
1140

tttaaagatc ctgtgtaccg tgcaagacgg aagcagtttg ctgacattgc ctacaactac
1200

cgccatgggc agcccatccc tcgagtggaa tacatggagg aagaaaagaa aacatggggc
1260

acagtgttca agactctgaa gtccttgtat aaaacccatg cttgctatga gtacaatcac
1320

atttttccac ttcttgaaaa gtactgtggc ttccatgaag ataacattcc ccagctggaa
1380

gacgtttctc aattcctgca gacttgcact ggtttccgcc tccgacctgt ggctggcctg
1440

ctttcctctc gggatttctt ggggtggcctg gccttccgag tcttccactg cacacagtac
1500

atcagacatg gatccaagcc catgtatacc cccgaacctg acatctgcca tgagctgttg
1560

ggacatgtgc ccttgttttc agatcgcagc tttgcccagt tttcccagga aattggcctt
1620

gcctctctgg gtgcacctga tgaatacatt gaaaagctcg ccacaattta ctggtttact
1680

gtggagtttg ggctctgcaa acaaggagac tccataaagg catatggtgc tgggctcctg
1740

tcatacctttg gtgaattaca gtactgctta tcagagaagc caaagcttct ccccctggag
1800

ctggagaaga cagccatcca aaattacact gtcacggagt tccagcccct gtattacgtg
1860

gcagagagtt ttaatgatgc caaggagaaa gtaaggaact ttgctgccac aatacctcgg
1920

cccttctcag ttcgctacga cccatacacc caaaggattg aggtcttgga caatacccag
1980

cagcttaaga ttttggctga ttccattaac agtgaaattg gaatcctttg cagtgcctc
2040

cagaaaataa agtaactcga gatccagaca tgataagata cattgatgag tttggacaaa
2100

ccacaactag aatgcagtga aaaaaatgct ttattttgtga aattttgtgat gctattgctt
2160

tattttgtaac cattataagc tgcaataaac aagttaacaa caacaattgc attcatttta
2220

tgttttcaggt tcaggggggag gtgtgggagg ttttttaaag catgctgggg agagatcgat
2280

ctgaggaacc cctagtgatg gagttggcca ctccctctct gcgcgctcgc tcgctcactg
2340

aggccgggcg accaaaggtc gcccgcgccc cgggctttgc ccgggcggcc tcagtgcgcg
2400

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
2426

<210> 86

<211> 2042

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-01004 transfer genome

<400> 86

ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc
60

tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc
120

cactcgaccc cttggaattt cggtggagag gagcagaggt tgtcctggcg tggtttaggt
180

agtgtgagag gggaatgact cttttcggtg agtgcagtgg aagctgtaca ctgcccaggc
240

aaagcgtccg ggcagcgtag gcgggcgact cagatcccag ccagtggact tagcccctgt
300

ttgctcctcc gataactggg gtgaccttgg ttaatatcca ccagcagcct cccccgttgc
360

ccctctggat ccactgctta aatacggacg aggacagggc cctgtctcct cagcttcagg
420

caccaccact gacctgggac agtgaatcct ctaaggtaaa tataaaattt ttaagtgtat
480

aatgtgttaa actactgatt ctaattgttt ctctctttta gattccaacc tttggaactg
540

accgccacca tgtccaccgc tgtgctggag aaccctgggc tggggaggaa actgtcagac
600

ttcgggcagg agacttcata cattgaggat aactgtaacc agaatggcgc catctctctg
660

atcttcagcc tgaaggagga agtgggcgcc ctggcaaagg tgctgcgcct gtttgaggag
720

aacgacgtga atctgaccca catcgagtcc cggccttcta gactgaagaa ggacgagtac
780

gagttcttta cccacctgga taagcgggcc ctgccagccc tgacaaacat catcaagatc
840

ctgaggcacg acatcggagc aaccgtgcac gagctgtctc gggacaagaa gaaggatacc
900

gtgccctggt tccctcggac aatccaggag ctggatagat ttgccaacca gatcctgtct
960

tacggagcag agctggacgc agatcacctt ggcttcaagg acccagtgta tcgggcccgg
1020

agaaagcagt ttgccgatat cgcctacaat tataggcacg gacagccaat ccctcgcgtg
1080

gagtatatgg aggaggagaa gaagacctgg ggcacagtgt tcaagacctt gaagagcctg
1140

tacaagacac acgcctgcta cgagtataac cacatcttcc ccctgctgga gaagtattgt
1200

ggctttcacg aggacaatat ccctcagctg gaggacgtga gccagttcct gcagacctgc
1260

acaggcttta ggctgaggcc agtggcagga ctgctgagct cccgggactt cctgggagga
1320

ctggccttca gagtgtttca ctgcacccag tacatcaggc acggctccaa gccaatgtat
1380

acaccagagc ccgacatctg tcacgagctg ctggggccacg tgcccctgtt tagcgataga
1440

tccttcgccc agttttccca ggagatcgga ctggcatctc tgggagcacc tgacgagtac
1500

atcgagaagc tggccaccat ctattgggtc acagtggagt ttggcctgtg caagcagggc
1560

gatagcatca aggcctacgg agcaggactg ctgtctagct tcggcgagct gcagtattgt
1620

ctgtccgaga agccaaagct gctgcccctg gagctggaga agaccgccat ccagaactac
1680

accgtgacag agttccagcc cctgtactat gtggccgagt cttttaacga tgccaaggag
1740

aaggtgagaa atttcgccgc cacaatccct aggcccttca gcgtgcggta cgacccttat
1800

accagagga tcgaggtgct ggataataca cagcagctga agatcctggc tgactcaatc
1860

aatagcgaaa tcggaatcct gtgctccgcc ctgcagaaaa tcaaatgaat gctttatttg
1920

tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata aacaagttaa
1980

caacaacaat tgcattcatt ttatgtttca gggtcagggg gaggtgtggg aggtttttta
2040

a a
2042

<211> 2356

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-01004 transfer genome

<400> 87

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt
60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggaatt cacgcgtgga
120

tctgaattca attcacgctt ggtacctccc taaaatgggc aaacattgca agcagcaaac
180

agcaaacaca cagccctccc tgcctgctga ccttggagct ggggcagagg tcagagacct
240

ctctgggccc atgccacctc caacatccac tcgacctcctt ggaatttcgg tggagaggag
300

cagaggttgt cctggcgtgg tttaggtagt gtgagagggg aatgactcct ttcggtaagt
360

gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc agcgtaggcg ggcgactcag
420

atcccagcca gtggacttag cccctgtttg ctccctccgat aactgggggtg accttggtta
480

atattcacca gcagcctccc ccgttgcccc tctggatcca ctgcttaaata acggacgagg
540

acagggccct gtctcctcag cttcaggcac caccactgac ctgggacagt gaatcctcta
600

aggtaaatat aaaattttta agtgtataat gtgttaaact actgattcta attgtttctc
660

tcttttagat tccaaccttt ggaactgacc gccaccatgt ccaccgctgt gctggagAAC
720

cctgggctgg ggaggaaact gtcagacttc gggcaggaga cttcatacat tgaggataac
780

tgtaaccaga atggcgccat ctctctgac ttcagcctga aggaggaagt gggcgccctg
840

gcaaaggtgc tgcgcctgtt tgaggagaac gacgtgaatc tgaccacat cgagtcccgg
900

ccttctagac tgaagaagga cgagtacgag ttctttaccc acctggataa gcggtccctg
960

ccagccctga caaacatcat caagatcctg aggcacgaca tcggagcaac cgtgcacgag
1020

ctgtctcggg acaagaagaa ggataccgtg ccctgggttc ctcggacaat ccaggagctg
1080

gatagatttg ccaaccagat cctgtcttac ggagcagagc tggacgcaga tcaccctggc
1140

ttcaaggacc cagtgtatcg ggcccggaga aagcagtttg ccgatatcgc ctacaattat
1200

aggcacggac agccaatccc tcgctggag tatatggagg aggagaagaa gacctggggc
1260

acagtgttca agaccctgaa gagcctgtac aagacacacg cctgctacga gtataaccac
1320

atcttcccc tgctggagaa gtattgtggc tttcacgagg acaatatccc tcagctggag
1380

gacgtgagcc agttcctgca gacctgcaca ggctttaggc tgaggccagt ggcaggactg
1440

ctgagctccc gggacttcct gggaggactg gccttcagag tgtttcactg caccagtac
1500

atcaggcacg gctccaagcc aatgtataca ccagagcccg acatctgtca cgagctgctg
1560

ggccacgtgc ccctgttttag cgatagatcc ttcgcccagt tttcccagga gatcggactg
1620

gcattctctgg gagcacctga cgagtacatc gagaagctgg ccaccatcta ttggttcaca
1680

gtggagtttg gcctgtgcaa gcagggcgat agcatcaagg cctacggagc aggactgctg
1740

tctagcttcg gcgagctgca gtattgtctg tccgagaagc caaagctgct gcccctggag
1800

ctggagaaga ccgccatcca gaactacacc gtgacagagt tccagcccct gtactatgtg
1860

gccgagtctt ttaacgatgc caaggagaag gtgagaaatt tcgccgccac aatccctagg
1920

cccttcagcg tgcggtacga cccttatacc cagaggatcg aggtgctgga taatacacag
1980

cagctgaaga tcctggctga ctcaatcaat agcgaaatcg gaatcctgtg ctccgccctg
2040

cagaaaatca aatgaatgct ttatttgtga aatttgtgat gctattgctt tatttgtaac
2100

cattataagc tgcaataaac aagttaacaa caacaattgc attcatttta tgtttcaggt
2160

tcaggggggag gtgtgggagg ttttttaaag catgctgggg agagatcgat ctgaggaacc
2220

cctagtgatg gagttggcca ctccctctct gcgcgctcgc tcgctcactg aggccgggcg
2280

accaaaggtc gcccgacgcc cgggctttgc ccgggcggcc tcagtgagcg agcgagcgcg
2340

c a g a g a g g g a g t g g c c
2356

<210> 88

<211> 5041

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-01004 full sequence

<400> 88

agaaaaactc atcgagcatc aaatgaaatt gcaattttatt catatcagga ttatcaatac
60

catatTTTTg aaaaagccgt ttctgtaatg aaggagaaaa ctcaccgagg cagttccata
120

ggatggcaag atcctggtat cggctctgca ttccgactcg tccaacatca atacaaccta
180

ttaatttccc ctcgtcaaaa ataaggttat caagtgagaa atcaccatga gtgacgactg
240

aatccggtga gaatggcaaa agtttatgca tttctttcca gacttgttca acaggccagc
300

cattacgctc gtcatcaaaa tcaactcgcat caaccaaacc gttattcatt cgtgattgcg
360

cctgagcgag gcgaaatacg cgatcgctgt taaaaggaca attacaaaca ggaatcgagt
420

gcaaccggcg caggaacact gccagcgcat caacaatatt ttcacctgaa tcaggatatt
480

cttctaatac ctggaacgct gtttttccgg ggatcgcagt ggtgagtaac catgcatcat
540

caggagtacg gataaaatgc ttgatggtcg gaagtggcat aaattccgctc agccagttta
600

gtctgaccat ctcatctgta acatcattgg caacgctacc tttgccatgt ttcagaaaca
660

actctggcgc atcgggcttc ccatacaagc gatagattgt cgcacctgat tgcccgacat
720

tatcgcgagc ccatttatac ccatataaat cagcatccat gttggaattt aatcgcggcc
780

tcgacgtttc ccgttgaata tggctcatat tcttcctttt tcaatattat tgaagcattt
840

atcagggtta ttgtctcatg agcggataca tatttgaatg tatttagaaa aataaaca
900

taggggtcag tgttacaacc aattaaccaa ttctgaacat tatcgcgagc ccattttatac
960

ctgaatatgg ctcataacac cccttgtttg cctggcggca gtagcgcggt ggtcccacct
1020

gaccccatgc cgaactcaga agtgaaacgc cgtagcgccg atggtagtgt ggggactccc
1080

catgcgagag tagggaactg ccaggcatca aataaaacga aaggctcagt cgaaagactg
1140

ggcctttcgc ccaaaccata tgattgacat gctagtttta cgattaccgt tcatcgccct
1200

gcgcgctcgc tcgctcactg aggccgcccg ggcaaagccc gggcgtcggg cgacctttgg
1260

tcgcccggcc tcagtgagcg agcgagcgcg cagagaggga gtggaattca cgcggtggatc
1320

tgaattcaat tcacgcgtgg tacctcccta aaatgggcaa acattgcaag cagcaaacag
1380

caaacacaca gccctccctg cctgctgacc ttggagctgg ggcagaggtc agagacctct
1440

ctggggcccat gccacctcca acatccactc gaccccttgg aatttcggtg gagaggagca
1500

gaggttgtcc tggcgtgggt taggtagtgt gagaggggaa tgactccttt cggttaagtgc
1560

agtggaagct gtacactgcc caggcaaagc gtccgggcag cgtaggcggg cgactcagat
1620

cccagccagt ggacttagcc cctgtttgct cctccgataa ctggggtgac cttggttaat
1680

attcaccagc agcctcccc gttgccctc tggatccact gcttaaatac ggacgaggac
1740

agggccctgt ctctcagct tcaggcacca ccactgacct gggacagtga atcctctaag
1800

gtaaataataa 1860	aatttttaag	tgtataatgt	gttaaactac	tgatttctaataat	tgttttctctc
tttttagattc 1920	caaccttttg	aactgaccgc	caccatgtcc	accgctgtgc	tgagagaaccc
tgggctgggg 1980	aggaaactgt	cagacttcgg	gcaggagact	tcatacattg	aggataactg
taaccagaat 2040	ggcgccatct	ctctgatctt	cagcctgaag	gaggaagtgg	gcgccctggc
aaaggtgctg 2100	cgctgtttg	aggagaacga	cgtgaatctg	accacatcg	agtcccggcc
ttctagactg 2160	aagaaggacg	agtacgagtt	ctttaccac	ctggataagc	ggccctgcc
agccctgaca 2220	aacatcatca	agatcctgag	gcacgacatc	ggagcaaccg	tgacgagct
gtctcgggac 2280	aagaagaagg	ataccgtgcc	ctggttccct	cggacaatcc	aggagctgga
tagatttgcc 2340	aaccagatcc	tgtcttacgg	agcagagctg	gacgcagatc	accctggctt
caaggaccca 2400	gtgtatcggg	cccggagaaa	gcagtttgcc	gatatcgctt	acaattatag
gcacggacag 2460	ccaatccctc	gcgtggagta	tatggaggag	gagaagaaga	cctggggcac
agtgttcaag 2520	accctgaaga	gcctgtacaa	gacacacgcc	tgctacgagt	ataaccacat
cttccccctg 2580	ctggagaagt	attgtggctt	tcacgaggac	aatatccctc	agctggagga
cgtgagccag 2640	ttcctgcaga	cctgcacagg	ctttaggctg	aggccagtgg	caggactgct
gagctcccgg 2700	gacttcctgg	gaggactggc	cttcagagtg	tttactgca	cccagtacat

caggcacggc tccaagccaa tgtatacacc agagcccgac atctgtcacg agctgctggg
2760

ccacgtgccc ctgttttagcg atagatcctt cgcccagttt tcccaggaga tcggactggc
2820

atctctggga gcacctgacg agtacatcga gaagctggcc accatctatt ggttcacagt
2880

ggagtttggc ctgtgcaagc agggcgatag catcaaggcc tacggagcag gactgctgtc
2940

tagcttcggc gagctgcagt attgtctgtc cgagaagcca aagctgctgc ccctggagct
3000

ggagaagacc gccatccaga actacaccgt gacagagttc cagcccctgt actatgtggc
3060

cgagtctttt aacgatgcc aaggagaaggt gagaaatttc gccgccacaa tccctaggcc
3120

cttcagcgtg cggtagcacc cttataccca gaggatcgag gtgctggata atacacagca
3180

gctgaagatc ctggctgact caatcaatag cgaaatcgga atcctgtgct ccgccctgca
3240

gaaaatcaaa tgaatgcttt atttgtgaaa tttgtgatgc tattgcttta tttgtaacca
3300

ttataagctg caataaacia gttaacaaca acaattgcat tcattttatg tttcaggttc
3360

agggggaggt gtgggaggtt ttttaaagca tgctggggag agatcgatct gaggaacccc
3420

tagtgatgga gttggccact ccctctctgc gcgctcgctc gctcactgag gccgggagac
3480

caaaggctcg cccgacgccc ggctttgccc gggcggcctc agtgagcgag cgagcgcgca
3540

gagagggagt ggcccatatg cggtagcaga attcgggtct agacgtcaaa agggcgacac
3600

aaaattttatt ctaaattgcat aataaatact gataacatct tatagtttgt attatatattt
3660

gtattatcgt tgacatgtat aattttgata tcaaaaactg attttccctt tattattttc
3720

gagattttatt ttcttaattc tctttaacaa actagaaata ttgtatatac aaaaaatcat
3780

aaataataga tgaatagttt aattataggt gttcatcaat cgaaaaagca acgtatctta
3840

tttaaagtgc gttgcttttt tctcatttat aaggttaaat aattctcata tatcaagcaa
3900

agtgacaggc gcccttaaatt attctgacaa atgctctttc cctaaactcc ccccataaaa
3960

aaacccgccg aagcggggtt ttacgttatt tgcggattaa cgattactcg ttatcagaac
4020

cgcccagggg gcccgagctt aagactggcc gtcgtttttac aacacagaaa gagttttag
4080

aaacgcaaaa aggccatccg tcaggggcct tctgcttagt ttgatgcctg gcagttccct
4140

actctgcct tccgcttcct cgctcactga ctcgctgcgc tcggtcgttc ggctgcggcg
4200

agcggtatca gctcactcaa aggcggtaat acggttatcc acagaatcag gggataacgc
4260

aggaaagaac atgtgagcaa aaggccagca aaaggccagg aaccgtaaaa aggccgcgtt
4320

gctggcggtt ttccataggc tccgcccccc tgacgagcat cacaaaaatc gacgctcaag
4380

tcagaggtgg cgaaacccga caggactata aagataccag gcgtttcccc ctggaagctc
4440

cctcgtgcgc tctcctgttc cgaccctgcc gcttaccgga tacctgtccg cctttctccc
4500

ttcgggaagc gtggcgcttt ctcatagctc acgctgtagg tatctcagtt cgggtgtaggt
4560

cgttcgctcc aagctgggct gtgtgcacga accccccggt cagcccgacc gctgcgctt
4620

atccggtaac tatcgtcttg agtccaaccc ggtaagacac gacttatcgc cactggcagc
4680

agccactggg aacaggatta gcagagcgag gtatgtaggc ggtgctacag agttcttgaa
4740

gtgggtgggct aactacggct acactagaag aacagtattt ggtatctgcg ctctgctgaa
4800

gccagttacc ttcggaaaaa gagttggtag ctcttgatcc ggcaaacaaa ccaccgctgg
4860

tagcggtggt ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga
4920

agatcctttg atcttttcta cgggggtctga cgctcagtgg aacgacgcgc gcgtaactca
4980

cgttaaggga ttttggtcat gagcttgccg cgccccgtca agtcagcgta atgctctgct
5040

t
5041

<210> 89

<211> 2042

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-01008 transfer genome

<400> 89

ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc
60

tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc
120

cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt
180

agtgtgagag gggaaatgact cctttcggta agtgcagtgg aagctgtaca ctgcccaggc
240

aaagcgtccg ggcagcgtag gcggggcgact cagatcccag ccagtggact tagcccctgt
300

ttgctcctcc gataactggg gtgaccttgg ttaatatcca ccagcagcct cccccgttgc
360

ccctctggat ccactgctta aatacggacg aggacagggc cctgtctcct cagcttcagg
420

caccaccact gacctgggac agtgaatcct ctaaggtaaa tataaaattt ttaagtgtat
480

aatgtgttaa actactgatt ctaattgttt ctctctttta gattccaacc tttggaactg
540

accgccacca tgtccaccgc tgtgctggag aaccctgggc tggggaggaa actgtcagac
600

ttcgggcagg agacttcata cattgaggat aactgtaacc agaatggcgc catctctctg
660

atcttcagcc tgaaggagga agtgggcgcc ctggcaaagg tgctgcgcct gtttgaggag
720

aacgacgtga atctgacca catcgagtcc cggccttcta gactgaagaa ggacgagtac
780

gagttcttta cccacctgga taagcgggtcc ctgccagccc tgacaaacat catcaagatc
840

ctgaggcacg acatcggagc aaccgtgcac gagctgtctc gggacaagaa gaaggatacc
900

gtgccctggg tccctcggac aatccaggag ctggatagat ttgccaacca gatcctgtct
960

tacggagcag agctggacgc agatcaccct ggcttcaagg acccagtgtg tcgggcccgg
1020

agaaagcagt ttgccgatat cgcctacaat tataggcacg gacagccaat ccctcgcgtg
1080

gagtatatgg aggaggagaa gaagacctgg ggcacagtgt tcaagaccct gaagagcctg
1140

tacaagacac acgcctgcta cgagtataac cacatcttcc ccctgctgga gaagtattgt
1200

ggctttcacg aggacaatat ccctcagctg gaggacgtga gccagttcct gcagacctgc
1260

acaggcttta ggctgaggcc agtggcagga ctgctgagct cccgggactt cctgggagga
1320

ctggccttca gagtgtttca ctgcaccag tacatcaggc acggctccaa gccaatgtat
1380

acaccagagc ccgacatctg tcacgagctg ctgggccacg tgcccctgtt tagcgataga
1440

tccttcgccc agttttccca ggagatcgga ctggcatctc tgggagcacc tgacgagtac
1500

atcgagaagc tggccaccat ctattgggtc acagtggagt ttggcctgtg caagcagggc
1560

gatagcatca aggcctacgg agcaggactg ctgtctagct tcggcgagct gcagtattgt
1620

ctgtccgaga agccaaagct gctgcccctg gagctggaga agaccgccat ccagaactac
1680

accgtgacag agttccagcc cctgtactat gtggccgagt cttttaacga tgccaaggag
1740

aaggtgagaa atttcgccgc cacaatccct aggcccttca gcgtgcggta cgacccttat
1800

accagagga tcgaggtgct ggataataca cagcagctga agatcctggc tgactcaatc
1860

aatagcgaaa tcggaatcct gtgctccgcc ctgcagaaaa tcaaatgaat gctttatttg
1920

tgaaatttgt gatgctattg ctttatttgt aaccattata agctgcaata aacaagttaa
1980

caacaacaat tgcattcatt ttatgtttca gggtcagggg gaggtgtggg aggtttttta
2040

a a
2042

<210> 90

<211> 2393

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-01008 transfer genome

<400> 90

ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgctc ggcgaccttt
60

ggtcgcccgg cctcagttag cgagcgagcg cgcagagagg gagtggaatt cacgcgtgga
120

tctgaattca attcacgcgt ggtacctccc taaaatgggc aaacattgca agcagcaaac
180

agcaaacaca cagccctccc tgcctgctga ccttgagact ggggcagagg tcagagacct
240

ctctgggccc atgccacctc caacatccac tcgacctt ggaatttcgg tggagaggag
300

cagaggttgt cctggcgtgg tttaggtagt gtgagagggg aatgactcct ttcggtaagt
360

gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc agcgtaggcg ggcgactcag
420

atcccagcca gtggacttag cccctgtttg ctcctccgat aactgggggtg accttggtta
480

atattcacca gcagcctccc ccggtgcccc tctggatcca ctgcttaaata acggacgagg
540

acagggccct 600	gtctcctcag	cttcaggcac	caccactgac	ctgggacagt	gaatcctcta
aggtaaatat 660	aaaattttta	agtgtataat	gtgttaaact	actgattcta	attgtttctc
tcttttagat 720	tccaaccttt	ggaactgacc	gccaccatgt	ccaccgctgt	gctggagaac
cctgggctgg 780	ggaggaaact	gtcagacttc	gggcaggaga	cttcatacat	tgaggataac
tgtaaccaga 840	atggcgccat	ctctctgata	ttcagcctga	aggaggaagt	gggcgccctg
gcaaagggtgc 900	tgcgctgttt	tgaggagaac	gacgtgaatc	tgaccacat	cgagtcccgg
ccttctagac 960	tgaagaagga	cgagtacgag	ttctttaccc	acctggataa	gcggtccctg
ccagccctga 1020	caaacatcat	caagatcctg	aggcacgaca	tcggagcaac	cgtgcacgag
ctgtctcggg 1080	acaagaagaa	ggataccgtg	ccctgggttc	ctcggacaat	ccaggagctg
gatagatttg 1140	ccaaccagat	cctgtcttac	ggagcagagc	tggacgcaga	tcaccctggc
ttcaaggacc 1200	cagtgtatcg	ggcccggaga	aagcagtttg	ccgatatcgc	ctacaattat
aggcacggac 1260	agccaatccc	tcgcgtggag	tatatggagg	aggagaagaa	gacctggggc
acagtgttca 1320	agaccctgaa	gagcctgtac	aagacacacg	cctgctacga	gtataaccac
atcttcccc 1380	tgctggagaa	gtattgtggc	tttcacgagg	acaatatccc	tcagctggag
gacgtgagcc 1440	agttcctgca	gacctgcaca	ggcttttaggc	tgaggccagt	ggcaggactg

ctgagctccc gggacttcct gggaggactg gccttcagag tgtttcactg caccagtac
1500

atcaggcacg gctccaagcc aatgtataca ccagagcccg acatctgtca cgagctgctg
1560

ggccacgtgc ccctgttttag cgatagatcc ttcgcccagt tttcccagga gatcggactg
1620

gcatctctgg gagcacctga cgagtacatc gagaagctgg ccaccatcta ttggttcaca
1680

gtggagtttg gcctgtgcaa gcagggcgat agcatcaagg cctacggagc aggactgctg
1740

tctagcttcg gcgagctgca gtattgtctg tccgagaagc caaagctgct gcccctggag
1800

ctggagaaga ccgccatcca gaactacacc gtgacagagt tccagcccct gtactatgtg
1860

gccgagtctt ttaacgatgc caaggagaag gtgagaaatt tcgccgccac aatccctagg
1920

cccttcagcg tgcgggtacga cccttatacc cagaggatcg aggtgctgga taatacacag
1980

cagctgaaga tcctggctga ctcaatcaat agcgaaatcg gaatcctgtg ctccgccctg
2040

cagaaaatca aatgaatgct ttatttgtga aatttgtgat gctattgctt tatttgtaac
2100

cattataagc tgcaataaac aagttaacaa caacaattgc attcatttta tgtttcaggt
2160

tcagggggag gtgtgggagg ttttttaaag catgctgggg agagatcgat ctggtagata
2220

agtagcatgg cgggttaatc attaactaca aggaaccctt agtgatggag ttggccactc
2280

cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggctgcc cgacgcccg
2340

gctttgcccg ggcggcctca gtgagcgagc gagcgcgcag agaggggagtg gcc
2393

<210> 91

<211> 5078

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-01008 full sequence

<400> 91

agaaaaactc atcgagcatc aaatgaaatt gcaattttatt catatcagga ttatcaatac
60

catatTTTTTg aaaaagccgt ttctgtaatg aaggagaaaa ctcaccgagg cagttccata
120

ggatggcaag atcctgggtat cggctctgcga ttccgactcg tccaacatca atacaaccta
180

ttaattttccc ctcgtcaaaa ataaggttat caagtgagaa atcaccatga gtgacgactg
240

aatccgggtga gaatggcaaa agttttatgca tttctttcca gacttgttca acaggccagc
300

cattacgctc gtcatcaaaa tcaactcgcat caaccaaacc gttattcatt cgtgattgcg
360

cctgagcgag gcgaaatacg cgatcgctgt taaaaggaca attacaaaca ggaatcgagt
420

gcaaccggcg caggaacact gccagcgcat caacaatatt ttcacctgaa tcaggatatt
480

cttctaatac ctggaacgct gtttttccgg ggatcgcagt ggtgagtaac catgcatcat
540

caggagtacg gataaaatgc ttgatgggtcg gaagtggcat aaattccgtc agccagttta
600

gtctgaccat ctcatctgta acatcattgg caacgctacc tttgccatgt ttcagaaaca
660

actctggcgc atcgggcttc ccatacaagc gatagattgt cgcacctgat tgcccgcacat
720

tatcgcgagc ccattttatac ccatataaat cagcatccat gttggaattt aatcgcggcc
780

tcgacgtttc ccgttgaata tggctcatat tcttcctttt tcaatattat tgaagcattt
840

atcaggggta ttgtctcatg agcggataca tatttgaatg tatttagaaa aataaacaaa
900

taggggtcag tgttacaacc aattaaccaa ttctgaacat tatcgcgagc ccattttatac
960

ctgaatatgg ctcataacac cccttgtttg cctggcggca gtagcgcggt ggtcccacct
1020

gaccccatgc cgaactcaga agtgaaacgc cgtagcgccg atggtagtgt ggggactccc
1080

catgcgagag tagggaactg ccaggcatca aataaaacga aaggctcagt cgaaagactg
1140

ggcctttcgc ccaaaccata tgattgacat gctagtttta cgattaccgt tcatcgccct
1200

gcgcgctcgc tcgctcactg aggccgcccg ggcaaagccc gggcgtcggg cgacctttgg
1260

tcgcccggcc tcagtgagcg agcgagcgcg cagagaggga gtggaattca cgcgtggatc
1320

tgaattcaat tcacgcgtgg tacctcccta aaatgggcaa acattgcaag cagcaaacag
1380

caaacacaca gccctccctg cctgctgacc ttggagctgg ggcagaggtc agagacctct
1440

ctggggcccat gccacctcca acatccactc gaccccttgg aatttcggtg gagaggagca
1500

gaggttgtcc tggcgtgggt taggtagtgt gagaggggaa tgactccttt cggttaagtgc
1560

agtggaagct gtacactgcc caggcaaagc gtccgggcag cgtaggcggg cgactcagat
1620

cccagccagt ggacttagcc cctgtttgct cctccgataa ctgggggtgac cttgggttaat
1680

attcaccagc agcctcccc gttgcccctc tggatccact gcttaaatac ggacgaggac
1740

agggccctgt ctctcagct tcaggcacca ccactgacct gggacagtga atcctctaag
1800

gtaaatataa aatttttaag tgtataatgt gttaaactac tgattctaata tgtttctctc
1860

ttttagattc caacctttgg aactgaccgc caccatgtcc accgctgtgc tggagaacct
1920

tgggctgggg aggaaactgt cagacttcgg gcaggagact tcatacattg aggataactg
1980

taaccagaat ggcgccatct ctctgatctt cagcctgaag gaggaagtgg gcgccctggc
2040

aaagggtgctg cgcctgtttg aggagaacga cgtgaatctg acccacatcg agtcccggcc
2100

ttctagactg aagaaggacg agtacgagtt ctttaccac ctggataagc ggtccctgcc
2160

agccctgaca aacatcatca agatcctgag gcacgacatc ggagcaaccg tgcacgagct
2220

gtctcgggac aagaagaagg ataccgtgcc ctggttccct cggacaatcc aggagctgga
2280

tagatttgcc aaccagatcc tgtcttacgg agcagagctg gacgcagatc accctggctt
2340

caaggaccca gtgtatcggg cccggagaaa gcagtttgcc gatatcgctt acaattatag
2400

gcacggacag ccaatccctc gcgtggagta tatggaggag gagaagaaga cctggggcac
2460

agtgttcaag accctgaaga gcctgtacaa gacacacgcc tgctacgagt ataaccacat
2520

cttccccctg ctggagaagt attgtggctt tcacgaggac aatatccctc agctggagga
2580

cgtgagccag ttcctgcaga cctgcacagg ctttaggctg aggccagtgg caggactgct
2640

gagctcccgg gacttcctgg gaggactggc cttcagagtg tttcactgca cccagtacat
2700

caggcacggc tccaagccaa tgtatacacc agagcccgac atctgtcacg agctgctggg
2760

ccacgtgccc ctgttttagcg atagatcctt cgcccagttt tcccaggaga tcggactggc
2820

atctctggga gcacctgacg agtacatcga gaagctggcc accatctatt ggttcacagt
2880

ggagtttggc ctgtgcaagc agggcgatag catcaaggcc tacggagcag gactgctgtc
2940

tagcttcggc gagctgcagt attgtctgtc cgagaagcca aagctgctgc ccctggagct
3000

ggagaagacc gccatccaga actacaccgt gacagagttc cagcccctgt actatgtggc
3060

cgagtctttt aacgatgcc aaggagaaggt gagaaatttc gccgccacaa tccctaggcc
3120

cttcagcgtg cggtacgacc cttataccca gaggatcgag gtgctggata atacacagca
3180

gctgaagatc ctggctgact caatcaatag cgaaatcgga atcctgtgct ccgccctgca
3240

gaaaatcaaa tgaatgcttt atttgtgaaa tttgtgatgc tattgcttta tttgtaacca
3300

ttataagctg caataaacia gttaacaaca acaattgcat tcattttatg tttcaggttc
3360

agggggaggt gtgggaggtt ttttaaagca tgctggggag agatcgatct ggtagataag
3420

tagcatggcg ggttaatcat taactacaag gaaccctag tgatggagtt ggccactccc
3480

tctctgcgcg ctgcctcgct cactgaggcc gggcgaccaa aggtcgcccg acgcccgggc
3540

tttggccggg cggcctcagt gagcgagcga gcgcgcagag agggagtggc ccatatgcgg
3600

taccagaatt cgggtctaga cgtcaaaagg gcgacacaaa atttattcta aatgcataat
3660

aaatactgat aacatcttat agtttgtatt atattttgta ttatcgttga catgtataat
3720

tttgatatca aaaactgatt ttccctttat tattttcgag atttattttc ttaattctct
3780

ttaacaaact agaaatattg tatatacaaa aaatcataaa taatagatga atagtttaat
3840

tataggtggt catcaatcga aaaagcaacg tatcttattt aaagtgcgtt gcttttttct
3900

catttataag gttaaataat tctcatatat caagcaaagt gacaggcgcc cttaaattatt
3960

ctgacaaatg ctctttccct aaactcccc cataaaaaaa cccgccgaag cgggttttta
4020

cgttatttgc ggattaacga ttactcgta tcagaaccgc ccagggggcc cgagcttaag
4080

actggccgtc gttttacaac acagaaagag tttgtagaaa cgcaaaaagg ccatccgtca
4140

ggggccttct gcttagtttg atgcctggca gttccctact ctgccttcc gcttcctcgc
4200

tcactgactc gctgcgctcg gtcgttcggc tgcggcgagc ggtatcagct cactcaaagg
4260

cggtaatacg gttatccaca gaatcagggg ataacgcagg aaagaacatg tgagcaaaag
4320

gccagcaaaa ggccaggaac cgtaaaaagg ccgcgttgct ggcgtttttc cataggctcc
4380

gccccctga cgagcatcac aaaaatcgac gctcaagtca gaggtggcga aacccgacag
4440

gactataaag ataccaggcg tttccccctg gaagctccct cgtgcgctct cctgttccga
4500

ccctgccgct taccggatac ctgtccgcct ttctcccttc gggaagcgtg gcgctttctc
4560

atagctcacg ctgtaggtat ctcagttcgg tgtaggtcgt tcgctccaag ctgggctgtg
4620

tgcacgaacc ccccgttcag cccgaccgct gcgccttata cggtaaactat cgtcttgagt
4680

ccaacccggt aagacacgac ttatcgccac tggcagcagc cactggtaac aggattagca
4740

gagcgaggta tgtaggcggt gctacagagt tcttgaagtg gtgggctaac tacggctaca
4800

ctagaagaac agtattttggt atctgcgctc tgctgaagcc agttaccttc ggaaaaagag
4860

ttggtagctc ttgatccggc aaacaaacca ccgctggtag cggtgggttt tttgtttgca
4920

agcagcagat tacgcgcaga aaaaaaggat ctcaagaaga tcctttgatc ttttctacgg
4980

ggtctgacgc tcagtggaac gacgcgcgcg taactcacgt taagggattt tggtcatgag
5040

cttgccgctg cccgtcaagt cagcgtaatg ctctgctt
5078

<211> 6020

<212> DNA

<213> Artificial Sequence

<220>

<223> pHMI-hPAH-TC-025 full sequence

<400> 92

aacgccagca acgcggcctt tttacgggtc ctggcctttt gctggccttt tgctcacatg
60

ttctttcctg cgttatcccc tgattctgtg gataaccgta ttaccgcctt tgagtgagct
120

gataccgctc gccgcagccg aacgaccgag cgcagcgcgt cagtgcgcga ggaagcggaa
180

gagcgcccaa tacgcaaacc gcctctcccc gcgcgttggc cgattcatta atgcagcagc
240

tggcgtaata gcgaagaggc ccgcaccgat cgcccttccc aacagttgcg cagcctgaat
300

ggcgaatgga attccagacg attgagcgtc aaaatgtagg tatttccatg agcgtttttc
360

ctgttgcaat ggctggcggg aatattgttc tggatattac cagcaaggcc gatagtttga
420

gttcttctac tcaggcaagt gatgttatta ctaatcaaag aagtattgcg acaacggtta
480

atttgcgtaga tggacagact cttttactcg gtggcctcac tgattataaa aacacttctc
540

aggattctgg cgtaccgttc ctgtctaaaa tccctttaat cggcctcctg tttagctccc
600

gctctgattc taacgaggaa agcacgttat acgtgctcgt caaagcaacc atagtacgcg
660

ccctgtagcg gcgcattaag cgcggcgggt gtggtgggta cgcgcagcgt gaccgctaca
720

cttgccagcg ccctagcgcc cgctcctttc gctttcttcc cttcctttct cgccacgttc
780

gccggctttc cccgtcaagc tctaaatcgg gggctccctt tagggttccg atttagtgct
840

ttacggcacc tcgaccccaa aaaacttgat tagggtgatg gttcacgtag tgggccatcg
900

ccctgataga cggttttttcg ccctttgacg ttggagtcca cgttctttaa tagtggactc
960

ttgttccaaa ctggaacaac actcaaccct atctcgggtct attcttttga tttataaggg
1020

atthttgccga tttcggccta ttggttaaaa aatgagctga ttttaacaaaa atttaacgcg
1080

aattttaaca aaatattaac gcttacaatt taaatatttg cttatacaat cttcctgttt
1140

ttggggcctt tctgattatc aaccggggta catatgattg acatgctagt tttacgatta
1200

ccgttcatcg ccctgcgcgc tcgctcgctc actgaggccg cccgggcaaa gcccgggcgt
1260

cgggcgacct ttggtcgccc ggcctcagtg agcgagcgag cgcgagagaga gggagtggaa
1320

ttcacgcgtg gatctgaatt caattcacgc gtggtacctc cctaaaatgg gcaaacattg
1380

caagcagcaa acagcaaaca cacagccctc cctgcctgct gaccttggag ctggggcaga
1440

ggtcagagac ctctctgggc ccatgccacc tccaacatcc actcgacccc ttggaatttc
1500

ggtggagagg agcagagggt gtcctggcgt ggtttaggta gtgtgagagg ggaatgactc
1560

ctttcggtaa gtgcagtgga agctgtacac tgcccaggca aagcgtccgg gcagcgtagg
1620

cgggcgactc agatcccagc cagtggactt agcccctgtt tgctcctccg ataactgggg
1680

tgaccttggt taatattcac cagcagcctc ccccgttgcc cctctggatc cactgcttaa
1740

atacggacga ggacagggcc ctgtctcctc agcttcaggc accaccactg acctgggaca
1800

gtgaatcctc taaggtaaata ataaaatttt taagtgtata atgtgttaaa ctactgattc
1860

taattgtttc tctcttttag attccaacct ttggaactga ccgccaccat gtccaccgct
1920

gtgctggaga accctgggct ggggaggaaa ctgtcagact tcgggcagga gacttcatac
1980

attgaggata actgtaacca gaatggcgcc atctctctga tcttcagcct gaaggaggaa
2040

gtgggcgccc tggcaaaggt gctgcgcctg tttgaggaga acgacgtgaa tctgaccac
2100

atcgagtcctc ggccttctag actgaagaag gacgagtacg agttctttac ccacctggat
2160

aagcggctccc tgccagccct gacaaacatc atcaagatcc tgaggcacga catcggagca
2220

accgtgcacg agctgtctcg ggacaagaag aaggataccg tgccctgggt ccctcggaca
2280

atccaggagc tggatagatt tgccaaccag atcctgtctt acggagcaga gctggacgca
2340

gatcacctg gcttcaagga cccagtgtat cgggcccga gaaagcagtt tgccgatatc
2400

gcctacaatt ataggcacgg acagccaatc cctcgcgtgg agtatatgga ggaggagaag
2460

aagacctggg gcacagtgtt caagaccctg aagagcctgt acaagacaca cgcttgctac
2520

gagtataacc acatcttccc cctgctggag aagtattgtg gctttcacga ggacaatatc
2580

cctcagctgg aggacgtgag ccagttcctg cagacctgca caggcttttag gctgaggcca
2640

gtggcaggac tgctgagctc ccgggacttc ctgggaggac tggccttcag agtgtttcac
2700

tgcacccagt acatcaggca cggctccaag ccaatgtata caccagagcc cgacatctgt
2760

cacgagctgc tgggccacgt gcccctgttt agcgatagat ccttcgccca gttttcccag
2820

gagatcggac tggcatctct gggagcacct gacgagtaca tcgagaagct ggccaccatc
2880

tattggttca cagtggagtt tggcctgtgc aagcagggcg atagcatcaa ggcctacgga
2940

gcaggactgc tgtctagctt cggcgagctg cagtattgtc tgtccgagaa gccaaagctg
3000

ctgcccctgg agctggagaa gaccgccatc cagaactaca ccgtgacaga gttccagccc
3060

ctgtactatg tggccgagtc ttttaacgat gccaaaggaga aggtgagaaa tttcgccgcc
3120

acaatcccta ggcccttcag cgtgcggtac gacccttata cccagaggat cgagggtgctg
3180

gataatacac agcagctgaa gacacctggct gactcaatca atagcgaaat cggaatcctg
3240

tgctccgccc tgcagaaaat caaatgaatg ctttatttgt gaaatttgtg atgctattgc
3300

tttatttgta accattataa gctgcaataa acaagttaac aacaacaatt gcattcattt
3360

tatgtttcag gttcaggggg aggtgtggga ggtttttttaa agcatgctgg ggagagatcg
3420

atctgaggaa cccctagtga tggagttggc cactccctct ctgcgcgctc gctcgctcac
3480

tgaggccggg cgaccaaagg tcgcccgacg cccgggcttt gcccgggcg cctcagtgag
3540

cgagcgagcg cgcagagagg gagtggcccc ccccccccc cccccggcg attctcttgt
3600

ttgctccaga ctctcaggca atgacctgat agcctttgta gagacctctc aaaaatagct
3660

accctctccg gcatgaattt atcagctaga acggttgaat atcatattga tgggtgatttg
3720

actgtctccg gcctttctca cccgtttgaa tctttaccta cacattactc aggcattgca
3780

tttaaaatat atgagggttc taaaaatttt tatccttgcg ttgaaataaa ggcttctccc
3840

gcaaaagtat tacagggtca taatgttttt ggtacaaccg atttagcttt atgctctgag
3900

gctttattgc ttaattttgc taattctttg ccttgccctgt atgatttatt ggatgttgga
3960

atcgctgat gcggtatttt ctcccttacgc atctgtgcgg tatttcacac cgcatatggg
4020

gcactctcag tacaatctgc tctgatgccg catagttaag ccagccccga caccgcgcaa
4080

caccgctga cgcgccctga cgggcttgtc tgctcccggc atccgcttac agacaagctg
4140

tgaccgtctc cgggagctgc atgtgtcaga ggttttcacc gtcatcaccg aaacgcgcga
4200

gacgaaagg cctcgtgata cgcctatttt tatagggttaa tgtcatgata ataatggttt
4260

cttagacgtc aggtggcact tttcggggaa atgtgcgcgg aaccctatt tgtttatttt
4320

tctaaataca ttcaaatatg tatccgctca tgagacaata accctgataa atgcttcaat
4380

aatattgaaa aaggaagagt atgagtattc aacatttccg tgtcgccctt attccctttt
4440

ttgcggcatt ttgccttcct gtttttgctc acccagaaac gctggtgaaa gtaaaagatg
4500

ctgaagatca gttgggtgca cgagtgggtt acatcgaact ggatctcaac agcggtaaga
4560

tccttgagag ttttcgcccc gaagaacgtt ttccaatgat gagcactttt aaagttctgc
4620

tatgtggcgc ggtattatcc cgtattgacg ccgggcaaga gcaactcggc cgccgcatac
4680

actattctca gaatgacttg gttgagtact caccagtcac agaaaagcat cttacggatg
4740

gcatgacagt aagagaatta tgcagtgctg ccataaccat gagtgataac actgcggcca
4800

acttacttct gacaacgacg ggaggaccga aggagctaac cgcttttttg cacaacatgg
4860

gggatcatgt aactcgcctt gatcgttggg aaccggagct gaatgaagcc ataccaaacg
4920

acgagcgtga caccacgatg cctgtagcaa tggcaacaac gttgcgcaaa ctattaactg
4980

gcgaactact tactctagct tcccggcaac aattaataga ctggatggag gcggataaag
5040

ttgcaggacc acttctgcgc tcggcccttc cggctggctg gtttattgct gataaatctg
5100

gagccggtga gcgtgggtct cgcggtatca ttgcagcact ggggccagat ggtaagccct
5160

cccgtatcgt agttatctac acgacgggga gtcaggcaac tatggatgaa cgaaatagac
5220

agatcgctga gatagggtgcc tcaactgatta agcattggta actgtcagac caagtttact
5280

catatatact ttagattgat ttaaaacttc atttttaatt taaaaggatc taggtgaaga
5340

tcctttttga taatctcatg accaaaatcc cttaacgtga gttttcgttc cactgagcgt
5400

cagaccccggt agaaaagatc aaaggatctt cttgagatcc tttttttctg cgcgtaatct
5460

gctgcttgca aacaaaaaaaa ccaccgctac cagcgggtggg ttgtttgccg gatcaagagc
5520

taccaactct ttttccgaag gtaactggct tcagcagagc gcagatacca aatactgtcc
5580

ttctagtgtg gccgtagtta ggccaccact tcaagaactc tgtagcaccg cctacatacc
5640

tcgctctgct aatcctgtta ccagtggctg ctgccagtgg cgataagtcg tgtcttaccg
5700

ggttggactc aagacgatag ttaccggata aggcgcagcg gtcgggctga acgggggggtt
5760

cgtgcacaca gcccagcttg gagcgaacga cctacaccga actgagatac ctacagcgtg
5820

agctatgaga aagcgccacg cttcccgaag ggagaaaggc ggacaggtat ccggtaagcg
5880

gcaggggtcgg aacaggagag cgcacgaggg agcttccagg gggaaacgcc tggatatcttt
5940

atagtcctgt cggggtttcgc cacctctgac ttgagcgtcg atttttgtga tgctcgtcag
6000

g g g g g c g g a g c c t a t g g a a a
6020