



(51) International Patent Classification:

C12N 15/90 (2006.01) C12N 7/00 (2006.01)
A61K 35/761 (2015.01)

(21) International Application Number:

PCT/US2021/070615

(22) International Filing Date:

27 May 2021 (27.05.2021)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

63/030,341 27 May 2020 (27.05.2020) US
63/117,252 23 November 2020 (23.11.2020) US

(71) Applicant: **HOMOLOGY MEDICINES, INC.** [US/US];
1 Patriots Park, Bedford, Massachusetts 01730 (US).

(72) Inventors: **WRIGHT, Jason Boke**; c/o Homology Medi-
cines, Inc., 1 Patriots Park, Bedford, Massachusetts 01730
(US). **SOOKIASIAN, Danielle Lauren**; c/o Homology
Medicines, Inc., 1 Patriots Park, Bedford, Massachusetts
01730 (US). **ST. MARTIN, Thia Baboval**; c/o Homolo-
gy Medicines, Inc., 1 Patriots Park, Bedford, Massachusetts
01730 (US). **FRANCONE, Omar**; c/o Homology Medi-
cines, Inc., 1 Patriots Park, Bedford, Massachusetts 01730

(US). **SEYMOUR, Albert Barnes**; c/o Homology Medi-
cines, Inc., 1 Patriots Park, Bedford, Massachusetts 01730
(US).

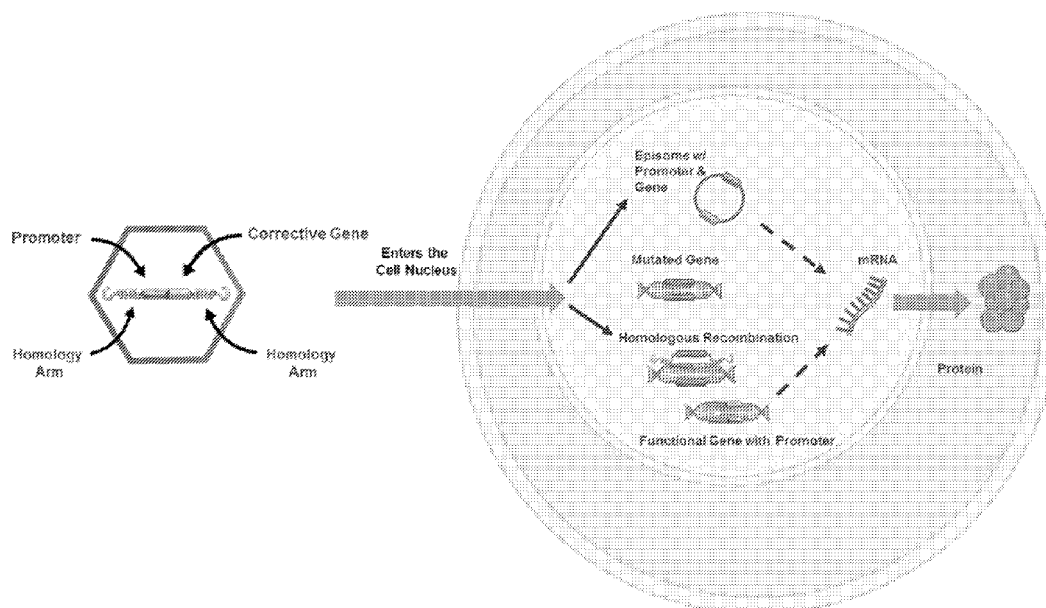
(74) Agent: **WILKINS, Andrew Thomas** et al.; c/o Dechert
LLP, One International Place, 40th Floor, 100 Oliver Street,
Boston, Massachusetts 02110-2605 (US).

(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, IT, 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, ST, SV, SY, TH, TJ, TM, TN,
TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, 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,

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

FIG. 1



(57) Abstract: Provided herein are recombinant adeno-associated virus (rAAV) compositions that can restore phenylalanine hydroxylase (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 nucleic acids, vectors, packaging systems, and methods for making the adeno-associated virus compositions.

TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW,
KM, ML, MR, NE, SN, TD, TG).

Published:

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

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

RELATED APPLICATIONS

- 5 **[0001]** This application claims priority to U.S. Provisional Patent Application Serial Nos. 63/030,341, filed May 27, 2020, and 63/117,252, filed November 23, 2020, the entire disclosures of which are hereby incorporated herein by reference.

SEQUENCE LISTING

- 10 **[0002]** This application contains a sequence listing which has been submitted electronically in ASCII format and is hereby incorporated by reference in its entirety (said ASCII copy, created on May 24, 2021, is named “404217-HMW-040WO-183161_SL.txt” and is 213,352 bytes in size).

BACKGROUND

- 15 **[0003]** Phenylketonuria (PKU) is an autosomal recessive genetic disorder where the majority of cases are caused by mutations in the phenylalanine hydroxylase (PAH) gene. The PAH gene encodes a hepatic enzyme that catalyzes the hydroxylation of L-phenylalanine (Phe) to L-tyrosine (Tyr) upon multimerization. Reduction or loss of PAH activity leads to
20 phenylalanine accumulation and its conversion into phenylpyruvate (also known as phenylketone). This abnormality in phenylalanine metabolism impairs neuronal maturation and the synthesis of myelin, resulting in mental retardation, seizures, and other serious medical problems.

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

- 30 **[0005]** New treatment strategies have recently emerged, including large neutral amino acid (LNAA) supplementation, cofactor tetrahydrobiopterin therapy, enzyme replacement therapy, and genetically modified probiotic therapy. However, these strategies suffer from

shortcomings. The LNAA supplementation is suitable only for adults not adhering to a low Phe diet. The cofactor tetrahydrobiopterin can only be used in some mild forms of PKU. Enzyme replacement by administration of a substitute for PAH, *e.g.*, phenylalanine ammonia-lyase (PAL), can lead to immune responses that reduce the efficacy and/or cause side effects.

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

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

10 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*).

15 [0007] It has also been speculated that nuclease-based gene editing technologies, such as meganucleases, zinc finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs), and clustered, regularly interspaced, short palindromic repeat (CRISPR) technology, may be used to correct defects in the PAH gene in PKU patients. However, each of these technologies raises safety concerns due to the potential for off-target mutation of sites

20 in the human genome similar in sequence to the intended target site.

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

SUMMARY

25 [0009] Provided herein are recombinant adeno-associated virus (rAAV) 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 nucleic acids, vectors, packaging systems, and methods for making the adeno-associated virus compositions. The rAAV compositions provided herein are particularly advantageous in that they can

30 efficiently edit the genome of cells (*e.g.*, liver cells) in a subject to express PAH under the control of a liver-specific promoter, and thereby offer a potential cure for PKU patients.

[0010] Accordingly, in one aspect, the instant disclosure provides a recombinant adeno-associated virus (rAAV) comprising:

- (a) an AAV capsid comprising an AAV capsid protein; and
- (b) an rAAV genome comprising: (i) an editing element for editing a target locus in a PAH gene, comprising at least a portion of a PAH coding sequence operably linked to a transcriptional regulatory element; (ii) a 5' homology arm nucleotide sequence position 5' of the editing element, having homology to a first genomic region 5' to the target locus; and (iii) a 3' homology arm nucleotide sequence positioned 3' of the editing element, having homology to a second genomic region 3' to the target locus.

[0011] In certain embodiments, the editing element comprises a PAH coding sequence. In certain embodiments, the PAH coding sequence encodes an amino acid sequence set forth in SEQ ID NO: 33. In certain embodiments, the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 53. In certain embodiments, the PAH coding sequence is silently altered. In certain embodiments, the PAH coding sequence comprises a nucleotide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the nucleotide sequence set forth in SEQ ID NO: 28, 63, or 83. In certain embodiments, the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 28. In certain embodiments, the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 63. In certain embodiments, the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 83.

[0012] 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 endogenous to the PAH gene. In certain embodiments, the transcriptional regulatory element is exogenous to the PAH gene. In certain embodiments, the transcriptional regulatory element is liver specific, optionally wherein the transcriptional regulatory element comprises one or more elements selected from the group consisting of a human albumin promoter, a human transthyretin (TTR) promoter, a human ApoE/C-I hepatic control region (HCR) 1 or 2, a human ApoH promoter, a human SERPINA1 (hAAT) promoter, and a hepatic specific regulatory module thereof. In certain embodiments, the transcriptional regulatory element comprises a nucleotide sequence at least 90% identical to a sequence selected from the group consisting of SEQ ID NO: 25, 26, 27, or 69. In certain embodiments, the transcriptional regulatory element comprises a nucleotide sequence at least 90% identical to the nucleotide sequence set forth in SEQ ID NO: 27. In

certain embodiments, the transcriptional regulatory element comprises the nucleotide sequence set forth in SEQ ID NO: 27. In certain embodiments, the transcriptional regulatory element consists of the nucleotide sequence set forth in SEQ ID NO: 27.

[0013] In certain embodiments, the editing element further comprises an intron element positioned 5' to the PAH coding sequence and 3' to the transcriptional regulatory element. In certain embodiments, the intron element is an exogenous intron element, optionally wherein the exogenous intron element is an SV40 intron element. In certain embodiments, the SV40 intron element comprises a nucleotide sequence at least 90% identical to the nucleotide sequence set forth in SEQ ID NO: 29. In certain embodiments, the SV40 intron element comprises the nucleotide sequence set forth in SEQ ID NO: 29. In certain embodiments, the SV40 intron element consists of the nucleotide sequence set forth in SEQ ID NO: 29.

[0014] In certain embodiments, the editing element further comprises a polyadenylation sequence 3' to the PAH coding sequence. In certain embodiments, the polyadenylation sequence is an exogenous polyadenylation sequence, optionally wherein the exogenous polyadenylation sequence is an SV40 polyadenylation sequence. In certain embodiments, the SV40 polyadenylation sequence comprises a nucleotide sequence at least 90% identical to the nucleotide sequence set forth in SEQ ID NO: 31. In certain embodiments, the SV40 polyadenylation sequence comprises the nucleotide sequence set forth in SEQ ID NO: 31. In certain embodiments, the SV40 polyadenylation sequence consists of the nucleotide sequence set forth in SEQ ID NO: 31.

[0015] In certain embodiments, the nucleotide 5' to the target locus is in an intron of a PAH gene. In certain embodiments, the nucleotide 5' to the target locus is in intron 1 of a PAH gene. In certain embodiments, the nucleotide 3' to the target locus is in an intron of a PAH gene. In certain embodiments, the nucleotide 3' to the target locus is in intron 1 of a PAH gene.

[0016] In certain embodiments, the PAH gene is a human PAH gene. In certain embodiments, the human PAH gene is wild-type. In certain embodiments, the human PAH gene is a variant PAH gene.

[0017] In certain embodiments, the editing element comprises a nucleotide sequence selected from the group consisting of SEQ ID NO: 25, 26, 27, 28, 29, 31, 50, 51, 52, 69, or 70.

[0018] In certain embodiments, the 5' homology arm nucleotide sequence is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the first genomic region. In certain embodiments, the 3' homology arm nucleotide sequence is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the second genomic region. In certain embodiments, the first genomic region is located in a first editing window,

and the second genomic region is located in a second editing window. In certain embodiments, the first editing window consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 37. In certain embodiments, the second editing window consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 38. In certain embodiments, the first genomic region consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 37. In certain embodiments, the second genomic region consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 38. In certain embodiments, each of the 5' and 3' homology arm nucleotide sequences independently has a length of about 100 to about 2000 nucleotides. In certain embodiments, the 5' homology arm comprises a nucleotide sequence that is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleotide sequence set forth in SEQ ID NO: 23. In certain embodiments, the 5' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 23. In certain embodiments, the nucleotide sequence of the 5' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 23. In certain embodiments, the 3' homology arm comprises a nucleotide sequence that is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleotide sequence set forth in SEQ ID NO: 24. In certain embodiments, the 3' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 24. In certain embodiments, the nucleotide sequence of the 3' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 24.

[0019] In certain embodiments, the rAAV genome comprises a nucleotide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleotide sequence set forth in SEQ ID NO: 43. In certain embodiments, the rAAV genome comprises the nucleotide sequence set forth in SEQ ID NO: 43.

[0020] In certain embodiments, the rAAV genome further comprises a 5' inverted terminal repeat (5' ITR) nucleotide sequence 5' of the 5' homology arm nucleotide sequence, and a 3' inverted terminal repeat (3' ITR) nucleotide sequence 3' of the 3' homology arm nucleotide sequence. In certain embodiments, the 5' ITR nucleotide sequence is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleotide sequence set forth in SEQ ID NO: 14, and the 3' ITR nucleotide sequence is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or 100% identical to the nucleotide sequence set forth in SEQ ID NO: 18.

[0021] In certain embodiments, the rAAV genome comprises the nucleotide sequence set forth in SEQ ID NO: 73 and/or 74. In certain embodiments, the rAAV genome comprises the nucleotide sequence set forth in SEQ ID NO: 75 and/or 76.

[0022] In certain embodiments, the rAAV genome comprises a nucleotide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleotide sequence set forth in SEQ ID NO: 45. In certain embodiments, the rAAV genome comprises the nucleotide sequence set forth in SEQ ID NO: 45. In certain embodiments, the nucleotide sequence of the rAAV genome consists of the nucleotide sequence set forth in SEQ ID NO: 45.

[0023] In certain embodiments, the AAV capsid protein comprises an amino acid sequence that is at least 95% identical to the amino acid sequence of amino acids 203-736 of SEQ ID NO: 1, 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 206 of SEQ ID NO: 16 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 16 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 16 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 16 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G. In certain embodiments:

(a) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G;

(b) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is

R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M;

(c) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO:

5 16 is R;

(d) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; or

(e) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is

10 I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C.

[0024] In certain embodiments, the capsid protein comprises the amino acid sequence of amino acids 203-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

15 **[0025]** In certain embodiments, the AAV capsid protein comprises an amino acid sequence that is at least 95% identical to the amino acid sequence of amino acids 138-736 of SEQ ID NO: 1, 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 151 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 16 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 16 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 16 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 16 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G or

20 Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 16 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is

25

30

C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G. In certain embodiments,

(a) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO:

5 16 is G;

(b) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO:

10 16 is M;

(c) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R;

(d) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; or

(e) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 20 16 is C.

[0026] 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.

[0027] In certain embodiments, the AAV capsid protein comprises an amino acid sequence that is at least 95% identical to the amino acid sequence of amino acids 1-736 of SEQ ID NO: 1, 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: 16 is T; the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 16 is I; the amino acid in the capsid protein corresponding to amino acid 68 of SEQ ID NO: 16 is V; the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 16 is L; the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 16 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 16 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H; the amino

acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 16 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 16 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 16 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G. In certain embodiments:

(a) the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 16 is T, and the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 16 is Q;

(b) the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 16 is I, and the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is Y;

(c) the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 16 is K;

(d) the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 16 is L, and the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 16 is S;

(e) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G;

(f) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M;

(g) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R;

(h) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; or

(i) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C.

[0028] In certain embodiments, the capsid protein comprises the amino acid sequence of amino acids 1-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

[0029] In certain embodiments, the integration efficiency of the editing element into the target locus is at least 1% when the AAV is administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions. In certain embodiments, the allelic frequency of integration of the editing element into the target locus is at least 0.5% when the AAV is administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions.

[0030] In another aspect, the instant disclosure provides a pharmaceutical composition comprising an rAAV disclosed herein.

[0031] In another aspect, the instant disclosure provides a polynucleotide comprising the nucleic acid sequence set forth in SEQ ID NO: 43, 45, 51, or 52.

[0032] In another aspect, the instant disclosure provides a method for treating a subject having phenylketonuria (PKU), the method comprising administering to the subject an effective amount of an rAAV, or pharmaceutical composition disclosed herein. In certain embodiments, the rAAV or pharmaceutical composition is administered intravenously. In certain embodiments, the PKU is associated with a PAH gene mutation. In certain embodiments, the subject is a human subject.

[0033] In another aspect, the instant disclosure provides a packaging system for preparation of an rAAV, wherein the packaging system comprises: (a) a first nucleotide sequence encoding one or more AAV Rep proteins; (b) a second nucleotide sequence encoding a capsid protein disclosed herein; and (c) a third nucleotide sequence comprising an rAAV genome sequence of an rAAV disclosed herein. In certain embodiments, the packaging system

comprises a first vector comprising the first nucleotide sequence and the second nucleotide sequence, and a second vector comprising the third nucleotide sequence. In certain embodiments, the packaging system further comprises a fourth nucleotide sequence comprising one or more helper virus genes. In certain embodiments, the fourth nucleotide sequence is comprised within a third vector. In certain embodiments, the fourth nucleotide sequence comprises one or more genes from a virus selected from the group consisting of adenovirus, herpes virus, vaccinia virus, and cytomegalovirus (CMV). In certain embodiments, the first vector, second vector, and/or the third vector is a plasmid.

[0034] In another aspect, the instant disclosure provides a method for recombinant preparation of an rAAV, the method comprising introducing a packaging system disclosed herein into a cell under conditions whereby the rAAV is produced.

[0035] An rAAV, pharmaceutical composition, or polynucleotide disclosed herein, for use as a medicament. An rAAV, pharmaceutical composition, or polynucleotide disclosed herein, for use in the treatment of PKU. An rAAV, pharmaceutical composition, or polynucleotide disclosed herein, for use in a method of treating a subject having PKU, the method comprising administering to the subject an effective amount of the rAAV, the pharmaceutical composition, or the polynucleotide.

BRIEF DESCRIPTION OF THE DRAWINGS

[0036] FIG. 1 is a schematic showing the gene transfer and gene editing mechanism of an rAAV as disclosed herein.

[0037] FIGs. 2A-2D are schematics showing the plasmid maps of PAH-006m (FIG. 2A), PAH-006m-LP-1 (FIG. 2B), PAH-032h (FIG. 2C), and hPAH-hI1C-032-LP1-SD3 (FIG. 2D).

[0038] FIG. 3 is a plot showing the effect of PAH-006m on blood Phe levels in PAH^{enu2} mice.

[0039] FIG. 4 is a plot showing on-target insertion at the mouse PAH locus following administration of PAH-006m.

[0040] FIG. 5 is a plot showing the effect of PAH-006m on mRNA expression relative to mouse GAPDH.

[0041] FIG. 6 is a plot showing on-target insertion in FRG® mice following administration of PAH-032h as measured by species specific ddPCR editing assays.

[0042] **FIG. 7** is a plot showing editing of human hepatocytes in FRG mice as measured by quantitative next generation sequencing assays.

[0043] **FIGs. 8A-8C** are plots showing the effect of PAH-006m-LP-1 on blood Phe concentration in PAH^{enu2} mice over time. **FIG. 8A** shows the full data set, while **FIG. 8B** shows the data over a certain range in time with a reduced y-axis scale in order to show differences between the dose levels. **FIG. 8C** shows the on-target integration of PAH-006m-LP-1 in PAH^{enu2} mice, as measured by next generation sequencing.

[0044] **FIGs. 9A-9L** are plots showing the efficacy of the gene transfer/gene editing vector PAH-006m-LP-1 and an episomal transgene expression vector. **FIGs. 9A and 9B** are plots showing the effect of the episomal transgene expression vector and PAH-006m-LP-1, respectively, packaged in AAVHSC15 capsid, on the serum Phe levels of PAH^{enu2} mice, up to 12 weeks post-dosing at the indicated doses. **FIGs. 9C and 9D** are plots showing the effect of the episomal transgene expression vector and PAH-006m-LP-1, respectively, packaged in AAVHSC15 capsid, on the serum Tyr levels of PAH^{enu2} mice, up to 12 weeks post-dosing at the indicated doses. **FIGs. 9E and 9F** are plots showing the effect of the episomal transgene expression vector and PAH-006m-LP-1, respectively, packaged in AAVHSC15 capsid, on the level of vector genomes detected per ug of genomic DNA in PAH^{enu2} mice, at the indicated doses. **FIGs. 9G and 9H** are plots showing the effect of the episomal transgene expression vector and PAH-006m-LP-1, respectively, packaged in AAVHSC15 capsid, on the level of mRNA expression per 10 ng of total RNA detected in PAH^{enu2} mice, at the indicated doses. **FIG. 9I** is a plot showing the level of on-target integration presented by the number of viral genomes per allele detected by ddPCR across various doses as indicated, at 12 weeks post-dosing of vector packaged in AAVHSC15 as indicated. **FIG. 9J** is a plot showing the frequency of on-target insertion detected across various doses as indicated, at 12 weeks post-dosing of vector packaged in AAVHSC15 as indicated. **FIGs. 9K and 9L** are plots showing time courses out to 42 weeks (**FIG. 9K**) or 40 weeks (**FIG. 9L**) post-dosing of PAH^{enu2} mice with vector and dosage as indicated, of the results of the effect of a partial hepatectomy (PHx) on serum Phe levels. PHx or sham surgery was performed at around 2 weeks post-dosing.

[0045] **FIGs. 10A-10P** are plots showing dose range differences between the mouse-specific gene transfer/gene editing AAV vector (PAH-006m-LP-1; “mouse design”), and the human-specific gene transfer/gene editing AAV vector (hPAH-hIIC-032-LP1-SD3; “human design”) in 4-week (**FIGs. 10A-10H**) or 10-week-old (**FIGs. 10I-10P**) PAH^{enu2} mice administered the vector packaged in AAVHSC15 capsid as indicated, at the dosage as indicated.

[0046] **FIGs. 11A-11K** are plots showing the effect the age of PAHenu2 mice has on the response to a single dose of the mouse-specific gene transfer/gene editing AAV vector (PAH-006m-LP-1; “mouse design”), or the human-specific gene transfer/gene editing AAV vector (hPAH-hIIC-032-LP1-SD3; “human design”). Serum Phe and Tyr levels (**FIGs. 11A-11D**) and vector genome and mRNA levels (**FIGs. 11E-11K**) were examined over time for the various dosing groups set forth in Table 2.

[0047] **FIGs. 12A-12C** are plots showing the kinetics and durability of integration over time, of PAH^{enu2} mice administered a mouse-specific gene transfer/gene editing AAV vector (PAH-006m-LP-1) packaged in AAVHSC15 capsid at a dose of 1E14 vg/kg. **FIGs. 12A and 12B** are plots showing the effect of PAH-006m-LP-1 packaged in AAVHSC15 capsid on the level of vector genomes detected per ug of genomic DNA, and the level of mRNA expression detected per 10 ng of total RNA in PAH^{enu2} mice, respectively, over time. **FIG. 12C** is a plot showing the frequency of on-target insertion detected at various time points post-dosing of PAH-006m-LP-1 packaged in AAVHSC15 capsid. **FIGs. 12D and 12E** are graphs showing the effect on serum Phe (**FIG. 12D**) and Tyr (**FIG. 12E**) in PAH^{enu2} mice administered PAH-006m-LP-1 packaged in AAVHSC15 capsid at a dose of 1E14 vg/kg, up to 42 weeks post-injection.

[0048] **FIGs. 13A and 13B** are plots showing the effect of hPAH-hIIC-032-LP1-SD3 packaged in AAVHSC15 on the level of vector genomes detected per ug of genomic DNA, and the level of mRNA expression detected per 10 ng of total RNA in human hepatocytes isolated from HuLiv mice, respectively, at the indicated doses and timepoints. **FIG. 13C** is a plot showing the frequency of on-target insertion detected in human hepatocytes isolated from HuLiv mice at various time points post-dosing of hPAH-hIIC-032-LP1-SD3 packaged in AAVHSC15, at the indicated doses.

[0049] **FIGs. 14A-14D** are plots showing the amount of integration at predicted off-target integration sites detected in genomic DNA isolated from HuLiv mice administered the human-specific gene transfer/gene editing vector, hPAH-hIIC-032-LP1-SD3, packaged in AAVHSC15 capsid.

DETAILED DESCRIPTION

[0050] Provided herein are recombinant adeno-associated virus (rAAV) 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 nucleic acids, vectors, packaging systems, and methods for making the adeno-associated virus compositions.

I. Definitions

5 [0051] As used herein, the terms “recombinant adeno-associated virus” or “rAAV” refers to an AAV comprising a genome lacking functional rep and cap genes.

[0052] As used herein, the term “PAH gene” refers to the phenylalanine hydroxylase (PAH) gene, including but not limited to the coding regions, exons, introns, 5' UTR, 3' UTR, and transcriptional regulatory regions of the PAH gene. The human PAH gene is identified by
10 Entrez Gene ID 5053. An exemplary nucleotide sequence of a PAH mRNA is provided as SEQ ID NO: 53. An exemplary amino acid sequence of a PAH polypeptide is provided as SEQ ID NO: 33. In certain embodiments, the PAH gene is a variant PAH gene. Variant PAH genes are known to those of skill in the art and may comprise one or more nucleotide differences as compared to the reference human genome. In certain embodiments, a variant
15 PAH gene is a common variant observed in the general populace. For example, variant rs1522296 comprises a single nucleotide difference from the reference human genome at the genomic location in the human genome build Hg38 at chr12: 103310787. This variant is seen in >30% of the global population and has no known association to either changes in PAH expression or disease risk.

20 [0053] As used herein, the term “rAAV genome” refers to a recombinant AAV genome that is capable of integrating an editing element (*e.g.*, one or more nucleotides or an internucleotide bond) via homologous recombination into a target locus to correct a genetic defect in a PAH gene. In certain embodiments, the target locus is in the human PAH gene. The skilled artisan will appreciate that the portion of an rAAV genome comprising the 5'
25 homology arm, editing element, and 3' homology arm can be in the sense or antisense orientation relative to the target locus (*e.g.*, the human PAH gene).

[0054] As used herein, the term “editing element” refers to the portion of an rAAV genome that when integrated at a target locus modifies the target locus. An editing element can mediate insertion, deletion, or substitution of one or more nucleotides at the target locus.

30 [0055] As used herein, the term “target locus” refers to a region of a chromosome or an internucleotide bond (*e.g.*, a region or an internucleotide bond of the human PAH gene) that is modified by an editing element.

[0056] As used herein, the term “homology arm” refers to a portion of an rAAV genome positioned 5' or 3' of an editing element that is substantially identical to the genome

flanking a target locus. In certain embodiments, the target locus is in a human PAH gene, and the homology arm comprises a sequence substantially identical to the genome flanking the target locus.

5 [0057] As used herein, the term “AAV capsid protein” refers to an AAV VP1, VP2, or VP3 capsid protein. As used herein, the term “Clade F capsid protein” refers to an AAV VP1, VP2, or VP3 capsid protein that comprises an amino acid sequence having at least 90% identity with the VP1, VP2, or VP3 amino acid sequences set forth, respectively, in amino acids 1-736, 138-736, and 203-736 of SEQ ID NO: 1 herein.

10 [0058] As used herein, the “percentage identity” between two nucleotide sequences or between two amino acid sequences is calculated by multiplying the number of matches between the pair of aligned sequences by 100, and dividing by the length of the aligned region, including internal gaps. Identity scoring only counts perfect matches and does not consider the degree of similarity of amino acids to one another. Note that only internal gaps are included in the length, not gaps at the sequence ends.

15 [0059] As used herein, the term “a disease or disorder associated with a PAH gene mutation” refers to any disease or disorder caused by, exacerbated by, or genetically linked with variation of a PAH gene. In certain embodiments, the disease or disorder associated with a PAH gene mutation is phenylketonuria (PKU).

20 [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. Codon alteration can be conducted by any method known in the art (*e.g.*, as described in Mauro & Chappell (2014) Trends Mol Med. 20(11):604-13, which is incorporated by reference herein in its entirety). Such silent alteration is advantageous
25 in that it reduces the likelihood of integration of the rAAV genome into loci of other genes or pseudogenes paralogous to the target gene. Such silent alteration also reduces the homology between the editing element and the target gene, thereby reducing undesired integration mediated by the editing element rather than by a homology arm.

30 [0061] As used herein, the term “coding sequence” refers to the portion of a complementary DNA (cDNA) that encodes a polypeptide, starting at the start codon and ending at the stop codon. A gene may have one or more coding sequences due to alternative splicing and/or alternative translation initiation. A coding sequence may either be wild-type or silently altered. An exemplary wild-type PAH coding sequence is set forth in SEQ ID NO: 53.

[0062] 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
5 (*e.g.*, an SV40 polyadenylation sequence).

[0063] As used herein, the term “intron element” refers to a cis-acting nucleotide sequence, for example, a DNA sequence, that regulates (*e.g.*, controls, increases, or reduces) expression of a transgene. In certain embodiments, an intron element is a modified intron, *e.g.*, a synthetic intron sequence. In certain embodiments, an intron element is an exogenous intron
10 element and is derived from an intron exogenous to the transgene it may regulate. In certain embodiments, an intron element comprises a modified splice acceptor and/or splice donor resulting in more robust splicing activity. While not wishing to be bound by theory, it is hypothesized that introns can increase transgene expression, for example, by reducing transcriptional silencing and enhancing mRNA export from the nucleus to the cytoplasm. A
15 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
20 reference herein in their entirety.

[0064] 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,
25 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
30 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.

[0065] 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.

[0066] In the instant disclosure, nucleotide positions in a PAH gene are specified relative to the first nucleotide of the start codon. The first nucleotide of a start codon is position 1; the nucleotides 5' to the first nucleotide of the start codon have negative numbers; the nucleotides 3' to the first nucleotide of the start codon have positive numbers. As used herein, nucleotide 1 of the human PAH gene is nucleotide 5,473 of the NCBI Reference Sequence: NG_008690.1, and nucleotide -1 of the human PAH gene is nucleotide 5,472 of the NCBI Reference Sequence: NG_008690.1.

[0067] In the instant disclosure, exons and introns in a PAH gene are specified relative to the exon encompassing the first nucleotide of the start codon, which is nucleotide 5473 of the NCBI Reference Sequence: NG_008690.1. The exon encompassing the first nucleotide of the start codon is exon 1. Exons 3' to exon 1 are from 5' to 3': exon 2, exon 3, etc. Introns 3' to exon 1 are from 5' to 3': intron 1, intron 2, etc. Accordingly, the PAH gene comprises from 5' to 3': exon 1, intron 1, exon 2, intron 2, exon 3, etc. As used herein, exon 1 of the human PAH gene is nucleotides 5001-5532 of the NCBI Reference Sequence: NG_008690.1, and intron 1 of the human PAH gene is nucleotides 5533-9704 of the NCBI Reference Sequence: NG_008690.1.

[0068] As used herein, the term “integration” refers to introduction of an editing element into a target locus (*e.g.*, of a PAH gene) by homologous recombination between an rAAV genome and the target locus. Integration of an editing element can result in substitution, insertion and/or deletion of one or more nucleotides in a target locus (*e.g.*, of a PAH gene).

[0069] As used herein, the term “integration efficiency of the editing element into the target locus” refers to the percentage of cells in a transduced population in which integration of the editing element into the target locus has occurred.

[0070] As used herein, the term “allelic frequency of integration of the editing element into the target locus” refers to the percentage of alleles in a population of transduced cells in which integration of the editing element into the target locus has occurred.

[0071] As used herein, the term “standard AAV administration conditions” refers to transduction of human hepatocytes implanted into a mouse following hepatocyte ablation, wherein the AAV is administered intravenously at a dose of 1×10^{13} vector genomes per kilogram of body weight.

5 [0072] 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.

[0073] As used herein, the term “about” or “approximately” when referring to a measurable value, such as the expression level of an IDS protein, encompasses variations of
10 $\pm 20\%$ or $\pm 10\%$, $\pm 5\%$, $\pm 1\%$, or $\pm 0.1\%$ of a given value or range, as are appropriate to perform the methods disclosed herein.

II. Adeno-Associated Virus Compositions

[0074] In one aspect, provided herein are novel rAAV compositions useful for restoring
15 PAH expression in cells with reduced or otherwise defective PAH gene function. Such rAAV compositions are highly efficient at editing the genome of cells (*e.g.*, liver cells) in a subject to express PAH under the control of a liver-specific promoter, and do not require cleavage of the genome at the target locus by the action of an exogenous nuclease (*e.g.*, a meganuclease, a zinc finger nuclease, a transcriptional activator-like nuclease (TALEN), or an RNA-guided nuclease
20 such as a Cas9) to facilitate such editing. Accordingly, in certain embodiments, the rAAV compositions disclosed herein do not comprise or require an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

[0075] In certain embodiments, the rAAV disclosed herein comprises: (a) an AAV capsid comprising an AAV capsid protein (*e.g.*, an AAV Clade F capsid protein); and (b) an
25 rAAV genome comprising: (i) an editing element for editing a target locus in a PAH gene, comprising at least a portion of a PAH coding sequence operably linked to a transcriptional regulatory element; (ii) a 5' homology arm nucleotide sequence position 5' of the editing element, having homology to a first genomic region 5' to the target locus; and (iii) a 3' homology arm nucleotide sequence positioned 3' of the editing element, having homology to a
30 second genomic region 3' to the target locus. In certain embodiments, the rAAV disclosed herein has the potential to express a PAH transgene both via episomal expression and through insertion of the editing element into the genome at the target locus in the PAH gene (see, FIG. 1). In certain embodiments, the rAAV disclosed herein allows for the expression of PAH to

be maintained throughout the period of hepatic growth, during which episomal expression may be lost. Accordingly, the rAAV compositions are particularly useful for treating juvenile PKU.

[0076] A capsid protein from any capsid known in the art can be used in the rAAV compositions disclosed herein, including, without limitation, a capsid protein from an AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, or AAV9 serotype. For example, in certain embodiments, the 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: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17. In certain embodiments, the 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: 1, 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: 16 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 16 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 16 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 16 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid

protein corresponding to amino acid 681 of SEQ ID NO: 16 is M. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C. In certain embodiments, the capsid protein comprises the amino acid sequence of amino acids 203-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

[0077] For example, in certain embodiments, the 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: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17. In certain embodiments, the 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: 1, 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: 16 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 16 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 16 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 16 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 16 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 16 is K; the

amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C. In certain embodiments, the capsid protein comprises the amino acid sequence of amino acids 138-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

[0078] For example, in certain embodiments, the 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: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17. In certain embodiments, the 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: 1, 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: 16 is T; the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 16 is I; the amino acid in the capsid protein corresponding to amino acid 68 of SEQ ID NO: 16 is V; the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 16 is L; the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 16 is D; the amino

acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 16 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 16 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 16 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 16 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 16 is T, and the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 16 is Q. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 16 is I, and the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is Y. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 16 is K. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 16 is L, and the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 16 is S. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R. In certain embodiments, the amino acid in the capsid protein

corresponding to amino acid 346 of SEQ ID NO: 16 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R. In certain embodiments, the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C. In certain embodiments, the capsid protein comprises the amino acid sequence of amino acids 1-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

[0079] In certain embodiments, the AAV capsid comprises two or more of: (a) a capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 1, 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17; (b) a capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17; and (c) a capsid protein comprising the amino acid sequence of amino acids 1-736 of SEQ ID NO: 1, 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 capsid protein having an amino acid sequence consisting of amino acids 203-736 of SEQ ID NO: 1, 2, 3, 4, 6, 7, 10, 11, 12, 13, 15, 16, or 17; (b) a capsid protein having an amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17; and (c) a capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

[0080] In certain embodiments, the AAV capsid comprises one or more of: (a) a 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%, 99%, or 100% sequence identity with the sequence of amino acids 203-736 of SEQ ID NO: 8; (b) a 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%, 99%, or 100% sequence identity with the sequence of amino acids 138-736 of SEQ ID NO: 8; and (c) a 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%, 99%, or 100% sequence identity with the sequence of amino acids 1-736 of SEQ ID NO: 8. In certain embodiments, the AAV capsid comprises one or more of: (a) a capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 8; (b) a capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 8; and (c) a 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 capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 8; (b) a capsid protein comprising

the amino acid sequence of amino acids 138-736 of SEQ ID NO: 8; and (c) a 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 capsid protein having an amino acid sequence consisting of amino acids 203-736 of SEQ ID NO: 8; (b) a capsid protein having an amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 8; and (c) a capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 8.

[0081] In certain embodiments, the AAV capsid comprises one or more of: (a) a 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%, 99%, or 100% sequence identity with the sequence of amino acids 203-736 of SEQ ID NO: 11; (b) a 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%, 99%, or 100% sequence identity with the sequence of amino acids 138-736 of SEQ ID NO: 11; and (c) a 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%, 99%, or 100% 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 capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 11; (b) a capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 11; and (c) a 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 capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 11; (b) a capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 11; and (c) a 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 capsid protein having an amino acid sequence consisting of amino acids 203-736 of SEQ ID NO: 11; (b) a capsid protein having an amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 11; and (c) a capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 11.

[0082] In certain embodiments, the AAV capsid comprises one or more of: (a) a 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%, 99%, or 100% sequence identity with the sequence of amino acids 203-736 of SEQ ID NO: 13; (b) a 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%, 99%, or 100%

sequence identity with the sequence of amino acids 138-736 of SEQ ID NO: 13; and (c) a 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%, 99%, or 100% sequence identity with the sequence of amino acids 1-736 of SEQ ID NO: 13. In certain
5 embodiments, the AAV capsid comprises one or more of: (a) a capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 13; (b) a capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 13; and (c) a 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 capsid protein
10 comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 13; (b) a capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 13; and (c) a 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 capsid protein having an amino acid sequence consisting of amino acids 203-736 of SEQ ID NO: 13; (b) a capsid protein having
15 an amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 13; and (c) a capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 13.

[0083] In certain embodiments, the AAV capsid comprises one or more of: (a) a 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
20 identity with the sequence of amino acids 203-736 of SEQ ID NO: 16; (b) a 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 capsid protein comprising an amino acid sequence having at least 80%, 81%, 82%, 83%, 84%, 85%, 86%,
25 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 capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 16; (b) a capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 16; and (c) a capsid protein comprising the
30 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 capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 16; (b) a capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 16; and (c) a 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 capsid protein having an amino acid sequence consisting of amino acids 203-736 of SEQ ID NO: 16; (b) a capsid protein having an amino acid sequence consisting of amino acids 138-736 of SEQ ID NO: 16; and (c) a capsid protein having an amino acid sequence consisting of amino acids 1-736 of SEQ ID NO: 16.

5 [0084] rAAV genomes useful in the AAV compositions disclosed herein generally comprise: (i) an editing element for editing a target locus in a PAH gene, comprising at least a portion of a PAH coding sequence operably linked to a transcriptional regulatory element; (ii) a 5' homology arm nucleotide sequence position 5' of the editing element, having homology to a first genomic region 5' to the target locus; and (iii) a 3' homology arm nucleotide sequence
10 positioned 3' of the editing element, having homology to a second genomic region 3' to the target locus. In certain embodiments, the rAAV genome comprises a 5' inverted terminal repeat (5' ITR) nucleotide sequence 5' of the 5' homology arm nucleotide sequence, and a 3' inverted terminal repeat (3' ITR) nucleotide sequence 3' of the 3' homology arm nucleotide sequence.

[0085] Editing elements used in the rAAV genomes disclosed herein can mediate
15 insertion, deletion, or substitution of one or more nucleotides at the target locus.

[0086] In certain embodiments, when correctly integrated by homologous recombination at the target locus, the editing element inserts a nucleotide sequence comprising at least a portion of a PAH coding sequence into a PAH gene. In certain embodiments, the editing element comprises a PAH coding sequence (*e.g.*, a complete PAH coding sequence).

20 [0087] In certain embodiments, the PAH coding sequence encodes a wild-type PAH polypeptide (*e.g.*, having the amino acid sequence set forth in SEQ ID NO: 33). In certain embodiments, the PAH coding sequence is wild-type (*e.g.*, comprising the nucleotide sequence set forth in SEQ ID NO: 53). In certain embodiments, the PAH coding sequence is silently altered to be less than 100% (*e.g.*, less than 95%, 90%, 85%, 80%, 75%, 70%, 65%, 60%, 55%,
25 or 50%) identical to the corresponding exons of the wild-type PAH gene. In certain embodiments, the PAH coding sequence comprises or consists of a nucleotide sequence that is at least 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the nucleotide sequence set forth in SEQ ID NO: 28, 63, or 83. In certain embodiments, the PAH coding sequence comprises the
30 nucleotide sequence set forth in SEQ ID NO: 28. In certain embodiments, the PAH coding sequence consists of the nucleotide sequence set forth in SEQ ID NO: 28. In certain embodiments, the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 63. In certain embodiments, the PAH coding sequence consists of the nucleotide sequence set forth in SEQ ID NO: 63. In certain embodiments, the PAH coding sequence

comprises the nucleotide sequence set forth in SEQ ID NO: 83. In certain embodiments, the PAH coding sequence consists of the nucleotide sequence set forth in SEQ ID NO: 83.

[0088] In certain embodiments, rAAV genomes useful in the AAV compositions disclosed herein comprise a transcriptional regulatory element (TRE) operably linked to at least a portion of a PAH coding sequence. In certain embodiments, rAAV genomes useful in the AAV compositions disclosed herein comprise from 5' to 3': a TRE, and the at least a portion of a PAH coding sequence.

[0089] The rAAV 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 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 54, 55, or 56), SV40 promoter, chicken ACTB promoter (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 47 or 57), JeT promoter (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 58), smCBA promoter (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 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 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 39), minute virus of mouse (MVM) intron which comprises transcription factor binding sites (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 61), 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 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 41 or 62), human calmodulin 1 (CALM1) promoter (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%,

or 100% identical to SEQ ID NO: 44), 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, an rAAV 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 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 42 or 64). For example, an rAAV 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 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 48 or 65).

[0090] 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.

[0091] 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 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 66), human APOE/C-I hepatic control region (HCR) 1 or 2 (e.g., comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 25 or 68), human APOH promoter, and human SERPINA1 (hAAT) promoter (e.g., comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 26, 69 or 70) or a hepatic specific regulatory module thereof (e.g., comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 71). In certain embodiments, an hAAT promoter region comprises a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 72. 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.

[0092] 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 1 (SYN1) promoter. More brain-specific promoter elements are disclosed in WO 2016/100575A1, which is incorporated by reference herein in its entirety.

[0093] In certain embodiments, the rAAV 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 rAAV genome comprises a human HCR1 (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 25 or 68) and a human EF-1 α promoter (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 39), optionally wherein the human HCR1 is 5' to the human EF-1 α promoter. In certain embodiments, the rAAV genome comprises a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the sequence nucleotide set forth in SEQ ID NO: 60.

[0094] Similarly, combinations of two or more tissue-specific TREs can drive efficient and tissue-specific transcription. For example, in certain embodiments, the rAAV genome comprises a human HCR1 (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 25) and a hAAT promoter (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 26), optionally wherein the human HCR1 is 5' to the hAAT promoter. In certain embodiments, the

rAAV genome comprises a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleotide sequence set forth in SEQ ID NO: 27. In certain embodiments, the rAAV genome comprises a human HCR1 (*e.g.*, comprising the nucleotide sequence set forth in SEQ ID NO: 25) and a hAAT promoter (*e.g.*, comprising the nucleotide sequence set forth in SEQ ID NO: 26), optionally wherein the human HCR1 is 5' to the hAAT promoter. In certain embodiments, the rAAV genome comprises the nucleotide sequence set forth in SEQ ID NO: 27.

[0095] In certain embodiments, the rAAV genome comprises a hepatic specific regulatory module of hAAT promoter (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 71) and a human TTR promoter (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 66), optionally wherein the hepatic specific regulatory module is 5' to the human TTR promoter. In certain embodiments, the rAAV genome comprises a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleotide sequence set forth in SEQ ID NO: 67. In certain embodiments, the rAAV genome comprises a hepatic specific regulatory module of hAAT promoter (*e.g.*, comprising the nucleotide sequence set forth in SEQ ID NO: 71) and a human TTR promoter (*e.g.*, comprising the nucleotide sequence set forth in SEQ ID NO: 66), optionally wherein the hepatic specific regulatory module is 5' to the human TTR promoter. In certain embodiments, the rAAV genome comprises the nucleotide sequence set forth in SEQ ID NO: 67.

[0096] In certain embodiments, the rAAV genome further comprises an intron element 5' to the at least a portion of a PAH coding sequence. Such intron elements can increase transgene expression, for example, by reducing transcriptional silencing and enhancing mRNA export from the nucleus to the cytoplasm. In certain embodiments, the rAAV genome comprises from 5' to 3': a TRE, an intron element, and the at least a portion of a PAH coding sequence.

[0097] The intron element can comprise at least a portion of a native intron sequence of the PAH gene, or the intron element can be an exogenous intron element (*e.g.*, comprising at least an intron sequence from a different species or a different gene from the same species, and/or a synthetic intron sequence). In certain embodiments, the intron element is an

exogenous intron element comprising at least a portion of an intron sequence from a different species. In certain embodiments, the intron element is an exogenous intron element comprising at least a portion of an intron sequence from a different gene from the same species. In certain embodiments, the intron element is an exogenous intron element comprising a synthetic intron sequence. In certain embodiments, the intron element is an exogenous intron element comprising a combination of at least an intron sequence from a different species or a different gene from the same species, and/or a synthetic intron sequence.

[0098] A skilled worker will appreciate that intron elements 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.

[0099] In certain embodiments, the rAAV genome comprises an exogenous intron element. In certain embodiments, the rAAV comprises an SV40 intron element (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 29) or a minute virus of mouse (MVM) intron (*e.g.*, comprising a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 61). In certain embodiments, the rAAV genome comprises an SV40 intron element (*e.g.*, comprising the nucleotide sequence set forth in SEQ ID NO: 29) or a minute virus of mouse (MVM) intron element (*e.g.*, comprising the nucleotide sequence set forth in SEQ ID NO: 61).

[00100] In certain embodiments, the rAAV genome disclosed herein further comprises a transcription terminator (*e.g.*, a polyadenylation sequence). In certain embodiments, the transcription terminator is 3' to the at least a portion of a 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 at least a portion of a 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 a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 31, 34, or 35, or a nucleotide sequence complementary thereto). In certain embodiments, the polyadenylation sequence comprises the nucleotide sequence set forth in SEQ ID NO: 31.

5 In certain embodiments, the polyadenylation sequence consists of the nucleotide sequence set forth in SEQ ID NO: 31.

[00101] In certain embodiments, the rAAV genome comprises from 5' to 3': a TRE, an intron element, at least a portion of a PAH coding sequence, and a polyadenylation sequence.

In certain embodiments, the TRE has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%,
 10 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 25-27, 30, 36, 39, 40-42, 44, 46-49, 54-60, or 62-72; the intron element has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 29 or 61; the at least a portion of a PAH coding sequence has at least 80%, 81%, 82%, 83%,
 15 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 28; and/or the polyadenylation sequence has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to any one of SEQ ID NOs: 31, 34, or 35. In certain embodiments, the TRE comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 25-27, 30, 36, 39, 40-42, 44, 46-49, 54-60, and 62-72; the intron element comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 29 and 61; the at least a portion of a PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 28; and/or the polyadenylation sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 31, 34, and 35.

25 **[00102]** In certain embodiments, the TRE comprises or consists of the nucleotide sequence set forth in SEQ ID NO: 25, 26, or 27; the intron element comprises or consists of the nucleotide sequence set forth in SEQ ID NO: 29; the at least a portion of a PAH coding sequence comprises or consists of the nucleotide sequence set forth in SEQ ID NO: 28; and/or the polyadenylation sequence comprises or consists of the nucleotide sequence set forth in SEQ
 30 ID NO: 31. In certain embodiments, the TRE comprises from 5' to 3' the nucleotide sequence set forth in SEQ ID NO: 25, and the nucleotide sequence set forth in SEQ ID NO: 26 (*e.g.*, the TRE comprises the nucleotide sequence set forth in SEQ ID NO: 27); the intron element comprises the nucleotide sequence set forth in SEQ ID NO: 29; the at least a portion of a PAH

coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 28; and/or the polyadenylation sequence comprises the nucleotide sequence set forth in SEQ ID NO: 31.

5 [00103] Homology arms used in the rAAV genomes disclosed herein can be directed to any region of the PAH gene or a gene nearby on the genome. The precise identity and positioning of the homology arms are determined by the identity of the editing element and/or the target locus.

[00104] Homology arms employed in the rAAV genomes disclosed herein are substantially identical to the genome flanking a target locus (*e.g.*, a target locus in a PAH gene). In certain embodiments, the 5' homology arm has at least about 90% (*e.g.*, at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) nucleotide sequence identity to a first genomic region 5' to the target locus. In certain embodiments, the 5' homology arm has 100% nucleotide sequence identity to the first genomic region. In certain embodiments, the 3' homology arm has at least about 90% (*e.g.*, at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) nucleotide sequence identity to a second genomic region 3' to the target locus. In certain embodiments, the 3' homology arm has 100% nucleotide sequence identity to the second genomic region. In certain embodiments, the 5' and 3' homology arms are each at least about 90% (*e.g.*, at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) identical to the first and second genomic regions flanking the target locus (*e.g.*, a target locus in the PAH gene), respectively. In certain embodiments, the 5' and 3' homology arms are each 100% identical to the first and second genomic regions flanking the target locus (*e.g.*, a target locus in the PAH gene), respectively. In certain embodiments, differences in nucleotide sequences of the 5' homology arm and/or the 3' homology arm and the corresponding regions the genome flanking a target locus comprise, consist essentially of, or consist of non-coding differences in nucleotide sequences.

25 [00105] The skilled worker will appreciate that homology arms do not need to be 100% identical to the genomic sequence flanking the target locus to be able to mediate integration of an editing element into that target locus by homologous recombination. For example, the homology arms can comprise one or more genetic variations in the human population, and/or one or more modifications (*e.g.*, nucleotide substitutions, insertions, or deletions) designed to improve expression level or specificity. Human genetic variations include both inherited variations and *de novo* variations that are private to the target genome, and encompass simple nucleotide polymorphisms, insertions, deletions, rearrangements, inversions, duplications, micro-repeats, and combinations thereof. Such variations are known in the art, and can be found, for example, in the databases of dnSNP (see *Sherry et al. Nucleic Acids Res. 2001*;

29(1):308-11), the Database of Genomic Variants (see *Nucleic Acids Res.* 2014; 42(Database issue): D986-92), ClinVar (see *Nucleic Acids Res.* 2014; 42(Database issue): D980-D985), Genbank (see *Nucleic Acids Res.* 2016; 44(Database issue): D67-D72), ENCODE (genome.ucsc.edu/encode/terms.html), JASPAR (see *Nucleic Acids Res.* 2018; 46(D1): D260-D266), and PROMO (see *Messeguer et al. Bioinformatics* 2002; 18(2):333-334; *Farré et al. Nucleic Acids Res.* 2003; 31(13):3651-3653), each of which is incorporated herein by reference. The skilled worker will further appreciate that in situations where a homology arm is not 100% identical to the genomic sequence flanking the target locus, homologous recombination between the homology arm and the genome may alter the genomic sequence flanking the target locus such that it becomes identical to the sequence of the homology arm used.

[00106] In certain embodiments, the first genomic region 5' to the target locus is located in a first editing window, wherein the first editing window consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 37. In certain embodiments, the second genomic region 3' to the target locus is located in a second editing window, wherein the second editing window consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 38. In certain embodiments, the first genomic region 5' to the target locus is located in a first editing window, wherein the first editing window consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 37; and the second genomic region 3' to the target locus is located in a second editing window, wherein the second editing window consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 38.

[00107] In certain embodiments, the first and second editing windows are different. In certain embodiments, the first editing window is located 5' to the second editing window. In certain embodiments, the first genomic region consists of a sequence shorter than the sequence of the first editing window in which the first genomic region is located. In certain embodiments, the first genomic region consists of the sequence of the first editing window in which the first genomic region is located. In certain embodiments, the second genomic region consists of a sequence shorter than the sequence of the second editing window in which the second genomic region is located. In certain embodiments, the second genomic region consists of the sequence of the second editing window in which the second genomic region is located.

[00108] In certain embodiments, the first and second editing windows are the same. In certain embodiments, the target locus is an internucleotide bond or a nucleotide sequence in

the editing window, wherein the first genomic region consists of a first portion of the editing window 5' to the target locus, and the second genomic region consists of a second portion of the editing window 3' to the target locus. In certain embodiments, the first portion of the editing window consists of the sequence from the 5' end of the editing window to the nucleotide
5 adjacently 5' to the target locus. In certain embodiments, the second portion of the editing window consists of the sequence from the nucleotide adjacently 3' to the target locus to the 3' end of the editing window. In certain embodiments, the first portion of the editing window consists of the sequence from the 5' end of the editing window to the nucleotide adjacently 5' to the target locus, and the second portion of the editing window consists of the sequence from
10 the nucleotide adjacently 3' to the target locus to the 3' end of the editing window. In certain embodiments, the editing window consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 37 or 38. In certain embodiments, the first and second portions of the editing windows have substantially equal lengths (*e.g.*, the ratio of the length of the shorter portion to the length of the longer portion is greater than 0.5, 0.55, 0.6,
15 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 0.96, 0.97, 0.98, or 0.99).

[00109] In certain embodiments, the 5' homology arm has a length of about 50 to about 4000 nucleotides (*e.g.*, about 100 to about 3000, about 200 to about 2000, about 500 to about 1000 nucleotides). In certain embodiments, the 5' homology arm has a length of about 800 nucleotides. In certain embodiments, the 5' homology arm has a length of about 100
20 nucleotides. In certain embodiments, the 3' homology arm has a length of about 50 to about 4000 nucleotides (*e.g.*, about 100 to about 3000, about 200 to about 2000, about 500 to about 1000 nucleotides). In certain embodiments, the 3' homology arm has a length of about 800 nucleotides. In certain embodiments, the 3' homology arm has a length of about 100 nucleotides. In certain embodiments, each of the 5' and 3' homology arms independently has
25 a length of about 50 to about 4000 nucleotides (*e.g.*, about 100 to about 3000, about 200 to about 2000, about 500 to about 1000 nucleotides). In certain embodiments, each of the 5' and 3' homology arms independently has a length of about 800 nucleotides.

[00110] In certain embodiments, the 5' and 3' homology arms have substantially equal nucleotide lengths. In certain embodiments, the 5' and 3' homology arms have asymmetrical
30 nucleotide lengths. In certain embodiments, the asymmetry in nucleotide length is defined by a difference between the 5' and 3' homology arms of up to 90% in the length, such as up to an 80%, 70%, 60%, 50%, 40%, 30%, 20%, or 10% difference in the length.

[00111] In certain embodiments, the 5' homology arm has at least about 90% (*e.g.*, at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) nucleotide sequence

identity to the nucleotide sequence set forth in SEQ ID NO: 37. In certain embodiments, the 5' homology arm further comprises one or more genetic variations in the human population. In certain embodiments, the 5' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 23. In certain embodiments, the 5' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 23.

[00112] In certain embodiments, the 3' homology arm has at least about 90% (*e.g.*, at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) nucleotide sequence identity to the nucleotide sequence set forth in SEQ ID NO: 38. In certain embodiments, the 3' homology arm further comprises one or more genetic variations in the human population. In certain embodiments, the 3' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 24. In certain embodiments, the 3' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 24.

[00113] In certain embodiments, the 5' homology arm and the 3' homology arm each has at least about 90% (*e.g.*, at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) nucleotide sequence identity to the nucleotide sequences set forth in SEQ ID NOs: 37 and 38, respectively. In certain embodiments, the 5' homology arm and the 3' homology arm each has at least about 90% (*e.g.*, at least about 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) nucleotide sequence identity to the nucleotide sequences set forth in SEQ ID NOs: 23 and 24, respectively. In certain embodiments, the 5' homology arm and the 3' homology arm comprise the nucleotide sequences set forth in SEQ ID NOs: 37 and 38, 23 and 24, 37 and 24, or 23 and 38, respectively. In certain embodiments, the 5' homology arm and the 3' homology arm consist of the nucleotide sequences set forth in SEQ ID NOs: 37 and 38, 23 and 24, 37 and 24, or 23 and 38, respectively.

[00114] In certain embodiments, the rAAV genome comprises a nucleotide sequence at least 80% (*e.g.*, at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 99.5%) identical to SEQ ID NO: 43. In certain embodiments, the rAAV genome comprises the nucleotide sequence set forth in SEQ ID NO: 43. In certain embodiments, the rAAV genome consists of the nucleotide sequence set forth in SEQ ID NO: 43.

[00115] In certain embodiments, the rAAV 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 rAAV 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 rAAV genomes disclosed herein are set forth in SEQ ID NOs: 14, 18, 19, 20, 21, and 32, herein.

[00116] 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 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 14, or the 3' ITR nucleotide sequence has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 18. In certain embodiments, the 5' ITR nucleotide sequence has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 14, and the 3' ITR nucleotide sequence has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 18. In certain embodiments, the rAAV genome comprises a nucleotide sequence set forth in SEQ ID NO: 43, a 5' ITR nucleotide sequence having the sequence of SEQ ID NO: 14, and a 3' ITR nucleotide sequence having the sequence of SEQ ID NO: 18.

[00117] 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 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 20, or the 3' ITR nucleotide sequence has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 21. In certain embodiments, the 5' ITR nucleotide sequence has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 20, and the 3' ITR nucleotide sequence has at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 21. In certain embodiments, the rAAV genome comprises a nucleotide sequence set forth in any one of SEQ ID NO: 43, 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.

[00118] 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).

[00119] In certain embodiments, the 5' ITR or the 3' ITR is modified to reduce or abolish resolution by Rep protein ("non-resolvable ITR"). In certain embodiments, the non-resolvable ITR comprises an insertion, deletion, or substitution in the nucleotide sequence of the terminal resolution site. Such modification allows formation of a self-complementary, double-stranded DNA genome of the AAV after the rAAV 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 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 19. In certain embodiments, the 5' ITR consists of a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 19. In certain embodiments, the 5' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 19. In certain embodiments, the 3' ITR comprises a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 32. In certain embodiments, the 5' ITR consists of a nucleotide sequence at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to SEQ ID NO: 32. In certain embodiments, the 3' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 32. In certain embodiments, the 5' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 19, and the 3' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 32. In certain embodiments, the 5' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 19, and the 3' ITR consists of the nucleotide sequence set forth in SEQ ID NO: 32.

[00120] In certain embodiments, the 5' ITR is flanked by an additional nucleotide sequence derived from a wild-type AAV2 genomic sequence. In certain embodiments, the 5' ITR is flanked by an additional 46 bp sequence derived from a wild-type AAV2 sequence that is adjacent to a wild-type AAV2 ITR in an AAV2 genome. In certain embodiments, the additional 46 bp sequence is 3' to the 5' ITR in the rAAV genome. In certain embodiments, the 46 bp sequence consists of the nucleotide sequence set forth in SEQ ID NO: 74.

[00121] 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 in an AAV2 genome. *See, e.g., Savy et al. Human Gene Therapy Methods (2017) 28(5): 277-289* (which is hereby incorporated by reference herein in

its entirety). In certain embodiments, the additional 37 bp sequence is 5' to the 3' ITR in the rAAV genome. In certain embodiments, the 37 bp sequence consists of the nucleotide sequence set forth in SEQ ID NO: 73.

[00122] In certain embodiments, the 5' homology arm has at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% nucleotide sequence identity to the nucleotide sequence set forth in SEQ ID NO: 37. In certain embodiments, the 5' homology arm further comprises one or more genetic variations in the human population. In certain embodiments, the 5' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 23. In certain embodiments, the 5' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 23.

[00123] In certain embodiments, the 3' homology arm has at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% nucleotide sequence identity to the nucleotide sequence set forth in SEQ ID NO: 38. In certain embodiments, the 3' homology arm further comprises one or more genetic variations in the human population. In certain embodiments, the 3' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 24. In certain embodiments, the 3' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 24.

[00124] In certain embodiments, the rAAV genome comprises from 5' to 3': a 5' homology arm, a 5' ITR, a TRE, an intron element, at least a portion of a PAH coding sequence, a polyadenylation sequence, a 3' ITR, and/or a 3' homology arm. In certain embodiments, the 5' homology arm has at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 37 or 23; the 5' ITR has at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 14, 19, or 20; the TRE has at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 25, 26, 27, 30, 36, 39, 40, 41, 42, 44, 46, 47, 48, 49, 54, 55, 56, 57, 58, 59, 60, 62, 64, 65, 66, 67, 68, 69, 70, 72, or 72; the intron element has at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 29 or 61; the at least a portion of a PAH coding sequence has at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 28; the polyadenylation sequence has at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%,

99%, or 100% sequence identity to SEQ ID NO: 31, 34, or 35; the 3' ITR has at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 18, 21, or 32; and/or the 3' homology arm has at least about 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% sequence identity to SEQ ID NO: 38 or 24.

[00125] In certain embodiments, the 5' homology arm comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 37 and 23; the 5' ITR comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 14, 19, or 20; the TRE comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 25-27, 30, 36, 39, 40-42, 44, 46-49, 54-60, and 62-72; the intron comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 29 and 61; the at least a portion of a PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 28; the polyadenylation sequence comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 31, 34, and 35; the 3' ITR comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 18, 21, and 32; and/or the 3' homology arm comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs: 38 and 24.

[00126] In certain embodiments, the 5' homology arm comprises or consists of the nucleotide sequence set forth in SEQ ID NO: 23; the 5' ITR comprises or consists of the nucleotide sequence set forth in SEQ ID NO: 14; the TRE comprises or consists of the nucleotide sequence set forth in SEQ ID NO: 25, 26, or 27; the intron element comprises or consists of the nucleotide sequence set forth in SEQ ID NO: 29; the at least a portion of a PAH coding sequence comprises or consists of the nucleotide sequence set forth in SEQ ID NO: 28; the polyadenylation sequence comprises or consists of the nucleotide sequence set forth in SEQ ID NO: 31; the 3' ITR comprises or consists of the nucleotide sequence set forth in SEQ ID NO: 18; and/or the 5' homology arm comprises or consists of the nucleotide sequence set forth in SEQ ID NO: 24.

[00127] In certain embodiments, the rAAV genome comprises from 5' to 3': the 5' homology arm comprising the nucleotide sequence set forth in SEQ ID NO: 23; the 5' ITR comprising the nucleotide sequence set forth in SEQ ID NO: 14; the TRE comprising from 5' to 3' the nucleotide sequence set forth in SEQ ID NO: 25, and the nucleotide sequence set forth in SEQ ID NO: 26 (*e.g.*, the TRE comprises the nucleotide sequence set forth in SEQ ID NO: 27); the intron element comprising the nucleotide sequence set forth in SEQ ID NO: 29; the at least a portion of a PAH coding sequence comprising the nucleotide sequence set forth in SEQ

ID NO: 28; the polyadenylation sequence comprising the nucleotide sequence set forth in SEQ ID NO: 31; the 3' ITR comprising of the nucleotide sequence set forth in SEQ ID NO: 18; and the 5' homology arm comprising the nucleotide sequence set forth in SEQ ID NO: 24.

[00128] In certain embodiments, the rAAV genome comprises a nucleotide sequence at least 80% (*e.g.*, at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%) identical to SEQ ID NO: 45. In certain embodiments, the rAAV genome comprises the nucleotide sequence set forth in SEQ ID NO: 45. In certain embodiments, the rAAV genome consists of the nucleotide sequence set forth in SEQ ID NO: 45.

[00129] In another aspect, provided herein is a polynucleotide comprising a nucleic acid sequence that is at least 80% (*e.g.*, at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99%) identical to the nucleic acid sequence set forth in SEQ ID NO: 28, 43, 45, 51, or 52. In certain embodiments, the polynucleotide comprises or consists of the nucleic acid sequence set forth in SEQ ID NO: 43, 45, 51, or 52.

[00130] 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.*

[00131] 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 silently altered to have less than 100% (*e.g.*, less than 99%, 98%, 97%, 96%, 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 nucleotide sequence set

forth in SEQ ID NO: 28. The polynucleotide can comprise DNA, RNA, modified DNA, modified RNA, or a combination thereof. In certain embodiments, the polynucleotide is an expression vector.

[00132] The AAV compositions disclosed herein are particularly advantageous in that they are capable of editing a PAH gene in a cell with high efficiency both *in vivo* and *in vitro*. In certain embodiments, the integration efficiency of the editing element into the target locus is at least 1% (*e.g.*, at least 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%) when the AAV is administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions. In certain embodiments, the allelic frequency of integration of the editing element into the target locus is at least 0.5% (*e.g.*, at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%) when the AAV is administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions.

[00133] Any methods of determining the efficiency of editing of the PAH gene can be employed. In certain embodiments, individual cells are separated from the population of transduced cells and subject to single-cell PCR using PCR primers that can identify the presence of an editing element correctly integrated into the target locus of the PAH gene. Such method can further comprise single-cell PCR of the same cells using PCR primers that selectively amplify an unmodified target locus. In this way, the genotype of the cells can be determined. For example, if the single cell PCR showed that a cell has both an edited target locus and an unmodified target locus, then the cell would be considered heterozygous for the edited PAH gene.

[00134] Additionally or alternatively, in certain embodiments, linear amplification mediated PCR (LAM-PCR), quantitative PCR (qPCR) or digital droplet PCR (ddPCR) can be performed on DNA extracted from the population of transduced cells using primers and probes that only detect edited PAH alleles. Such methods can further comprise an additional qPCR or ddPCR (either in the same reaction or a separate reaction) to determine the number of total genomes in the sample and the number of unedited PAH alleles. These numbers can be used to determine the allelic frequency of integration of the editing element into the target locus.

[00135] Additionally or alternatively, in certain embodiments, the PAH locus can be amplified from DNA extracted from the population of transduced cells either by PCR using primers that bind to regions of the PAH gene flanking the target locus, or by LAM-PCR using

a primer that binds a region within the rAAV genome (*e.g.*, a region comprising an exogenous sequence non-native to the locus). The resultant PCR amplicons can be individually sequenced using single molecule next generation sequencing (NGS) techniques to determine the relative number of edited and unedited PAH alleles present in the population of transduced cells. These numbers can be used to determine the allelic frequency of integration of the editing element into the target locus.

[00136] In another aspect, the instant disclosure provides pharmaceutical compositions comprising an AAV as disclosed herein together with a pharmaceutically acceptable excipient, adjuvant, diluent, vehicle or carrier, or a combination thereof. A “pharmaceutically acceptable carrier” includes any material which, when combined with an active ingredient of a composition, allows the ingredient to retain biological activity and without causing disruptive physiological reactions, such as an unintended immune reaction. Pharmaceutically acceptable carriers include water, phosphate buffered saline, emulsions such as oil/water emulsion, and wetting agents. Compositions comprising such carriers are formulated by well-known conventional methods such as those set forth in *Remington’s Pharmaceutical Sciences, current Ed.*, Mack Publishing Co., Easton Pa. 18042, USA; *A. Gennaro (2000) “Remington: The Science and Practice of Pharmacy,” 20th edition, Lippincott, Williams, & Wilkins; Pharmaceutical Dosage Forms and Drug Delivery Systems (1999) H. C. Ansel et al, 7th ed., Lippincott, Williams, & Wilkins; and Handbook of Pharmaceutical Excipients (2000) A. H. Kibbe et al, 3rd ed. Amer. Pharmaceutical Assoc.*

III. Methods of Use

[00137] In another aspect, the instant disclosure provides methods for restoring PAH gene function in a cell. The methods generally comprise transducing the cell with an rAAV as disclosed herein. Such methods are highly efficient at editing the PAH gene function in a cell, and do not require cleavage of the genome at the target locus by the action of an exogenous nuclease (*e.g.*, a meganuclease, a zinc finger nuclease, a transcriptional activator-like nuclease (TALEN), or an RNA-guided nuclease such as a Cas9) to facilitate such correction. Accordingly, in certain embodiments, the methods disclosed herein involve transducing the cell with an rAAV as disclosed herein without co-transducing or co-administering an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

[00138] The methods disclosed herein can be applied to any cell harboring a mutation in the PAH gene. The skilled worker will appreciate that cells that actively express PAH are of particular interest. Accordingly, in certain embodiments, the method is applied to cells in

the liver, kidney, brain, pituitary gland, adrenal gland, pancreas, urinary bladder, gallbladder, colon, small intestine, or breast. In certain embodiments, the method is applied to hepatocytes and/or renal cells.

5 [00139] The methods disclosed herein can be performed *in vitro* for research purposes or can be performed *ex vivo* or *in vivo* for therapeutic purposes.

[00140] 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
10 generally comprising administering to the subject an effective amount of an rAAV as disclosed herein. The subject can be a human subject or a rodent subject (*e.g.*, a mouse) containing human liver cells. Suitable mouse subjects include without limitation, mice into which human liver cells (*e.g.*, human hepatocytes) have been engrafted. Any disease or disorder associated with a PAH gene mutation can be treated using the methods disclosed herein. Suitable diseases
15 or disorders include, without limitation, phenylketonuria. In certain embodiments, the cell is transduced without co-transducing or co-administering an exogenous nuclease or a nucleotide sequence that encodes an exogenous nuclease.

[00141] The methods disclosed herein are particularly advantageous in that they are capable of editing a PAH gene in a cell with high efficiency both *in vivo* and *in vitro*. In certain
20 embodiments, the integration efficiency of the editing element into the target locus is at least 1% (*e.g.*, at least 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%) when the AAV is administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions. In certain embodiments, the allelic
25 frequency of integration of the editing element into the target locus is at least 0.5% (*e.g.*, at least 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, or 95%) when the AAV is administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions.

30 [00142] The methods disclosed herein can be performed *in vitro* for research purposes or can be performed *ex vivo* or *in vivo* for therapeutic purposes.

[00143] 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 an rAAV 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 gene mutation. 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 (PKU).

[00144] In certain embodiments, the foregoing methods employ an rAAV comprising: (a) an AAV capsid protein comprising the amino acid sequence of amino acids 203-736 of SEQ ID NO: 16, and an rAAV genome comprising 5' to 3' following genetic elements: a 5' homology arm (*e.g.*, the 5' homology arm comprising the nucleotide sequence set forth in SEQ ID NO: 23), a 5' ITR (*e.g.*, the 5' ITR comprising the nucleotide sequence set forth in SEQ ID NO: 14), a transcriptional regulatory element (*e.g.*, a TRE comprising the nucleotide sequence set forth in SEQ ID NO: 27), an intron element (*e.g.*, the intron element comprising the nucleotide sequence set forth in SEQ ID NO: 29), at least a portion of a PAH coding sequence (*e.g.*, the PAH coding sequence comprising the nucleotide sequence set forth in SEQ ID NO: 28), a polyadenylation sequence (*e.g.*, the SV40 polyadenylation sequence of SEQ ID NO: 31), a 3' ITR (*e.g.*, the 3' ITR comprising the nucleotide sequence set forth in SEQ ID NO: 18), and a 3' homology arm (*e.g.*, the 3' homology arm comprising the nucleotide sequence set forth in SEQ ID NO: 24); (b) an AAV capsid protein comprising the amino acid sequence of amino acids 138-736 of SEQ ID NO: 16, and an rAAV genome comprising 5' to 3' following genetic elements: a 5' homology arm (*e.g.*, the 5' homology arm comprising the nucleotide sequence set forth in SEQ ID NO: 23), a 5' ITR (*e.g.*, the 5' ITR comprising the nucleotide sequence set forth in SEQ ID NO: 14), a transcriptional regulatory element (*e.g.*, a TRE comprising the nucleotide sequence set forth in SEQ ID NO: 27), an intron element (*e.g.*, the intron element comprising the nucleotide sequence set forth in SEQ ID NO: 29), at least a portion of a PAH coding sequence (*e.g.*, the PAH coding sequence comprising the nucleotide sequence set forth in SEQ ID NO: 28), a polyadenylation sequence (*e.g.*, the SV40 polyadenylation sequence of SEQ ID NO: 31), a 3' ITR (*e.g.*, the 3' ITR comprising the nucleotide sequence set forth in SEQ ID NO: 18), and a 3' homology arm (*e.g.*, the 3' homology arm comprising the nucleotide sequence set forth in SEQ ID NO: 24); and/or (c) an AAV capsid protein comprising the amino acid sequence of SEQ ID NO: 16, and an rAAV genome comprising 5' to 3' following genetic elements: a 5' homology arm (*e.g.*, the 5' homology arm comprising the nucleotide sequence set forth in SEQ ID NO: 23), a 5' ITR (*e.g.*, the 5' ITR comprising the nucleotide sequence set forth in SEQ ID NO: 14), a transcriptional regulatory element (*e.g.*, a TRE comprising the

nucleotide sequence set forth in SEQ ID NO: 27), an intron element (*e.g.*, the intron element comprising the nucleotide sequence set forth in SEQ ID NO: 29), at least a portion of a PAH coding sequence (*e.g.*, the PAH coding sequence comprising the nucleotide sequence set forth in SEQ ID NO: 28), a polyadenylation sequence (*e.g.*, the SV40 polyadenylation sequence of
5 SEQ ID NO: 31), a 3' ITR (*e.g.*, the 3' ITR comprising the nucleotide sequence set forth in SEQ ID NO: 18), and a 3' homology arm (*e.g.*, the 3' homology arm comprising the nucleotide sequence set forth in SEQ ID NO: 24).

[00145] 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
10 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.

[00146] An AAV composition disclosed herein can be administered to a subject by any
15 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.

20 IV. AAV Packaging Systems

[00147] In another aspect, the instant disclosure provides packaging systems for recombinant preparation of a recombinant adeno-associated virus (rAAV) disclosed herein. Such packaging systems generally comprise: first nucleotide encoding one or more AAV Rep proteins; a second nucleotide encoding a capsid protein of any of the AAVs as disclosed herein;
25 and a third nucleotide sequence comprising any of the rAAV genomes as disclosed herein, wherein the packaging system is operative in a cell for enclosing the rAAV genome in the capsid to form the AAV.

[00148] In certain embodiments, the packaging system comprises a first vector comprising the first nucleotide sequence encoding the one or more AAV Rep proteins and the
30 second nucleotide sequence encoding the AAV capsid protein, and a second vector comprising the third nucleotide sequence comprising the rAAV 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.).

[00149] 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.

[00150] In certain embodiments of the packaging system, the packaging system further comprises a fourth nucleotide sequence comprising one or more helper virus genes. In certain embodiments of the packaging system, the packaging system further comprises a third vector, *e.g.*, a helper virus vector, comprising the fourth nucleotide sequence comprising the one or more helper virus genes. The third vector may be an independent third vector, integral with the first vector, or integral with the second vector.

[00151] In certain embodiments of the packaging system, the helper virus is selected from the group consisting of adenovirus, herpes virus (including herpes simplex virus (HSV)), poxvirus (such as vaccinia virus), cytomegalovirus (CMV), and baculovirus. In certain embodiments of the packaging system, where the helper virus is adenovirus, the adenovirus genome comprises one or more adenovirus RNA genes selected from the group consisting of E1, E2, E4 and VA. In certain embodiments of the packaging system, where the helper virus is HSV, the HSV genome comprises one or more of HSV genes selected from the group consisting of UL5/8/52, ICPO, ICP4, ICP22 and UL30/UL42.

[00152] In certain embodiments of the packaging system, the first, second, and/or third vector are contained within one or more plasmids. In certain embodiments, the first vector and the third vector are contained within a first plasmid. In certain embodiments the second vector and the third vector are contained within a second plasmid.

[00153] In certain embodiments of the packaging system, the first, second, and/or third vector are contained within one or more recombinant helper viruses. In certain embodiments, the first vector and the third vector are contained within a recombinant helper virus. In certain embodiments, the second vector and the third vector are contained within a recombinant helper virus.

[00154] 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 herein under conditions operative for enclosing the rAAV genome in the capsid to form the rAAV as described herein. Exemplary methods for recombinant preparation of an rAAV include transient transfection (*e.g.*, with one or more transfection plasmids containing a first, and a second, and optionally a third vector as described herein), viral infection (*e.g.*, with one or more recombinant helper viruses, such as a adenovirus, poxvirus (such as vaccinia virus), herpes virus (including HSV, cytomegalovirus, or baculovirus), containing a first, and a second, and optionally a third vector as described herein), and stable producer cell line transfection or infection (*e.g.*, with a stable producer cell, such as a mammalian or insect cell, containing a Rep nucleotide sequence encoding one or more AAV Rep proteins and/or a Cap nucleotide sequence encoding one or more AAV capsid proteins as described herein, and with an rAAV genome as described herein being delivered in the form of a plasmid or a recombinant helper virus).

[00155] Accordingly, the instant disclosure provides a packaging system for preparation of a recombinant AAV (rAAV), wherein the packaging system comprises a first nucleotide sequence encoding one or more AAV Rep proteins; a second nucleotide sequence encoding a capsid protein of any one of the AAVs described herein; a third nucleotide sequence comprising an rAAV genome sequence of any one of the AAVs described herein; and optionally a fourth nucleotide sequence comprising one or more helper virus genes.

V. Examples

[00156] These examples are offered by way of illustration, and not by way of limitation.

Example 1: hPAH Correction Vector Designs

Mouse-specific PAH gene editing vector PAH-006m

[00157] The mouse-specific gene editing AAV vector, PAH-006m, is shown in FIG. 2A. This vector was designed to integrate a human PAH coding sequence into the mouse PAH gene locus using left and right homology arms that are specific to the mouse sequence. PAH-006m comprises, from 5' to 3', the following genetic elements: a 5' ITR element; a 5' homology arm; a silently altered human PAH coding sequence; an SV40 polyadenylation sequence; a 3' homology arm; and a 3' ITR element. The sequences of these elements are set forth in Table 1. PAH-006m does not comprise a heterologous promoter.

Mouse -specific PAH gene transfer/gene editing vector PAH-006m-LP-1

[00158] The mouse-specific gene transfer/gene editing AAV vector, PAH-006m-LP-1, is shown in FIG. 2B. This vector was designed to integrate a human PAH coding sequence into the mouse PAH gene locus using mouse left and right homology arms, and also to allow expression of human PAH under the control of a heterologous liver-specific promoter in a cell in the absence of genomic integration. PAH-006m-LP-1 comprises, from 5' to 3', the following genetic elements: a 5' ITR element; a 5' homology arm; a transcriptional regulatory element comprising a human apolipoprotein hepatic control region (HCR) element, a human alpha-1-antitrypsin (hAAT) promoter element, and a hAAT exon 1; an SV40 element; a silently altered human PAH coding sequence; an SV40 polyadenylation sequence; a 3' homology arm; and a 3' ITR element. The sequences of these elements are set forth in Table 1.

Human-specific PAH gene editing vector PAH-032h

[00159] The human-specific gene editing AAV vector, PAH-032h, is shown in FIG. 2C. This vector was designed to integrate a human PAH coding sequence into the human PAH gene locus using left and right homology arms that are specific to the human sequence. PAH-032h comprises, from 5' to 3', the following genetic elements: a 5' ITR element; a 5' homology arm; a splice acceptor; a 2A element; a silently altered human PAH coding sequence; an SV40 polyadenylation sequence; a 3' homology arm; and a 3' ITR element. The sequences of these elements are set forth in Table 1. PAH-032h does not comprise a heterologous promoter.

Human-specific PAH gene transfer/gene editing vector hPAH-hIIC-032-LP1-SD3

[00160] The human-specific gene transfer/gene editing AAV vector, hPAH-hIIC-032-LP1-SD3, is shown in FIG. 2D. This vector was designed to integrate a human PAH coding sequence into the human PAH gene locus using human left and right homology arms, and also to allow expression of human PAH under the control of a heterologous liver-specific promoter in a cell in the absence of genomic integration. hPAH-hIIC-032-LP1-SD3 comprises, from 5' to 3', the following genetic elements: a 5' ITR element; a 5' homology arm; a transcriptional regulatory element comprising a human apolipoprotein hepatic control region (HCR) element, a human alpha-1-antitrypsin (hAAT) promoter element, and a hAAT exon 1; an SV40 element; a silently altered human PAH coding sequence; an SV40 polyadenylation sequence; a 3' homology arm; and a 3' ITR element. The sequences of these elements are set forth in Table 1.

Table 1: Selected genetic elements in PAH correction vectors PAH-006m, PAH-006m-LP-1, PAH-032h, and hPAH-hI1C-032-LP1-SD3

	PAH-006m	PAH-006m-LP-1	PAH-032h	hPAH-hI1C-032-LP1-SD3
Genetic Element	SEQ ID NO:			
5' ITR element	14	14	14	14
5' homology arm	84	85	23	23
Splice acceptor	-	-	30	-
2A element	-	-	40	-
transcriptional regulatory element	-	27	-	27
SV40 intron element	-	29	-	29
silently altered human PAH coding sequence	83	28	83	28
SV40 polyadenylation sequence	31	31	31	31
3' homology arm	86	86	24	24
3' ITR element	18	18	18	18
AAV genome from 5' homology arm to 3' homology arm	77	79	81	43
AAV genome from 5' ITR to 3' ITR	78	80	82	45

“-” signifies an absence of this element

5 Example 2: Analysis of hPAH Correction Vectors in Mouse Models

Materials and Methods

PAH^{enu2} Mouse Model:

[00161] The PAH^{enu2} mouse model was used to establish dose response for PAH correction vectors packaged in AAVHSC15 capsid through measurement of vector genomes, percent target gene insertion, hPAH mRNA expression, and Phe/Tyr concentration.

[00162] The PAH^{enu2} mouse line was produced by germline ethylnitrosourea mutagenesis followed by a Phe clearance screen to isolate a mutant mouse line deficient in PAH activity (see McDonald JD, et al. PNAS 1990;87:1965–1967, which is hereby incorporated by reference in its entirety). It has a homozygous missense mutation (F263S in exon 7) in the region encoding the PAH active site, resulting in abolished activity of PAH. Mutations within exon 7 are common in humans with PAH deficiency, although PAH deficiency in humans is typically expressed in a compound heterozygous background. PAH^{enu2} mice exhibit present with blood Phe levels consistently above 1200 µmol/L.

FRG® Mouse Human Liver Xenograft Model:

[00163] Fah^{-/-}/Rag2^{-/-}/Il2rg^{-/-} (FRG) mice (Azuma *et al.* (2007) *Nat. Biotechnol.* 25(8): 903-910) containing human liver tissue were used to measure *in vivo* transduction of vector genomes, human DNA target gene insertion, and mRNA expression of hPAH in human liver tissue, following administration of correction vectors packaged in AAVHSC15 capsid across a range of doses.

[00164] The FRG® human liver xenograft model is an immunocompromised mouse strain harboring triple gene knockouts of Fah^{-/-}, IL2rg^{-/-} and Rag2^{-/-}. Because of the IL2rg^{-/-} and Rag2^{-/-} knockouts these mice lack B-cell, T-cell and NK-cells resulting in acceptance of engraftment with human cells. The Fah^{-/-} genotype renders these mice dependent on the liver protective drug, 2-(2-nitro-4-trifluoromethylbenzoyl)-1,3-cyclohexanedione (NTBC), such that its withdrawal leads to loss of hepatocytes and animal death within 4-8 weeks.

[00165] To generate humanized livers in FRG mice, human primary hepatocytes were implanted concurrent with NTBC withdrawal. Since the human hepatocytes have an intact FAH gene, only murine hepatocytes are affected, resulting in the gradual repopulation of the mouse liver compartment with human hepatocytes. The human repopulated livers show >90% human hepatocytes and restored hepatic function (Azuma *et al.* (2007) *Nat. Biotechnol.* 25(8): 903-910).

On-Target Insertion – ddPCR Linkage Assay:

[00166] To determine the level of target insertion per allele at the PAH gene locus, an assay that uses droplet digital PCR (ddPCR) was employed. In this assay, target insertion was determined by measuring genetic linkage between the hPAH transgene and the genomic target. Linkage is a measurement of how often two sequences are on the same strand of DNA. Determining the amount of linked to unlinked vector and genome sequences can be used to measure insertion efficiency, which is reported as percent insertion per allele.

[00167] To measure target insertion via linkage, two primer and probe sets were used, one targeting the silently altered hPAH payload and a second targeting the *Pah* genomic sequence beyond the homology arms. Each probe consisted of a fluorophore in either the FAM or HEX channel such that when DNA is analyzed in partitioned droplets, each droplet can be scored by its emission signature as producing one of three possible signals: FAM-positive (silently altered hPAH payload alone); HEX-positive (genomic alone); and HEX + FAM positive (contains silently altered hPAH payload and genomic sequence). Target insertion was measured by determining the proportion of droplets containing both the payload and the genomic sequence in excess of expectation due to probability, divided by the total number of *Pah* alleles tested.

[00168] As the relative concentration of vector to genome can vary across doses and models, it is important to ensure each measurement is within the linear range of detection for both the vector and the target genomic sequence. Therefore, prior to measuring target insertion, the relative abundance of vector and target genome was measured across a range of sample input concentrations from 1 ng, 5 ng, 10 ng, 100 ng. The ratio of vector to genome was used to determine the sample input concentration in which both the vector and genomes fall within the limits of detection by ddPCR which are between 0.25 – 5000 molecules per μ l.

On-Target Insertion – Next Generation Sequencing (NGS) Assay:

[00169] This approach was used to calculate target insertion using assays spanning both left and right integration sites of the human PAH editing construct. Each of the two amplicons share their respective outward primer located on the genomic DNA flanking each homology arm, while the inward primers were unique for unedited alleles or for the edited alleles respectively. Each number of wild-type (WT) and edited sequences were tallied by counting the number of sequences covering the junction between the homology arm and either the inserted gene or the unedited wild-type sequence. Having separate assays covering both the left and right side of the insertion site provides redundant quantitation of insertion. By using this NGS approach, the percentage of target insertions can be detected per total number of alleles, and in addition, sequences of the entire insertion site can be collected, enabling the detection of *de novo* mutations (e.g., incorrect insertion and deletion events and ITR integration).

[00170] Differences of uninserted and inserted amplification efficiency were accounted for using an 11-step standard for both left and right integration sites. Each control panel consists of purified uninserted amplicon and edited amplicon at the following ratios: 0% edited, 1%, 2%, 5%, 10%, 50%, 60%, 90%, 95%, 98%, 99% and 100% edited control sequence. Target insertion efficiency calculations were as follows: the sequencing reads specific to uninserted and inserted genomic loci are tallied and target insertion was calculated by (insertion read counts / total read counts * 100%) and fitted to the standard curve. As editing quantitation per sample was measured by two independent assays, each spanning opposite homology arms into the native genome, the consistency between each measurement was determined.

On-Target Accuracy – Next Generation Sequencing (NGS) Assay:

[00171] To determine if target gene insertion is accompanied by *de novo* mutation at the integration sites, PCR amplicons spanning genomic sequences beyond both left and right of the homologous arms and into the target insertion site were sequenced. A total of four amplicons were sequenced per sample. Insertion specific amplicons spanning each homology

arm (hPAH dependent PCR) or WT/uninserted alleles spanning each homology arm (no hPAH insertion) were assessed.

[00172] To identify the detection limit of the assay, an amplicon-based control panel was built by mixing two right homologous arm amplicons, varied by one base (a T to G variation), at 0%, 0.1%, 0.5%, 1%, 2%, 5%, 10%, 50, and 100%. The amplicon control panel went through the same processing steps. Based on the quantitative variant calling from the standard curve, false-positive variants were observed at a frequency of 0.1-0.25% at non-random positions. Sequencing coverage for almost all positions was above 10,000. A variant was considered to be a true positive with the following criteria: (1) at >0.5% variant frequency and pass filters; (2) appear in both technical sequencing duplicates; and (3) not observed in WT alleles given comparable sequence depth and quality.

Results

[00173] A study was performed in PAH^{enu2} mice using the mouse-specific correction vector, PAH-006m, which contains homology arms targeting the mouse PAH locus. PAH-006m does not contain a liver specific promoter, and as such, expression is driven by the endogenous promoter elements. Three male mice per group were administered either 1E+14 vg/kg PAH-006m packaged in AAVHSC15 capsid, or formulation buffer, and blood Phe level was measured weekly for 12 weeks. At the 12-week timepoint, mice were sacrificed to measure target insertion at the PAH locus, and mRNA expression levels.

[00174] Blood Phe levels were reduced to normal levels by 2 weeks post-administration of PAH-006m packaged in AAVHSC15 capsid, as shown in FIG. 3. Reduction of blood Phe was maintained at levels of <360 μ M for the duration of the 12-week study.

[00175] The level of on-target insertion at the PAH locus was measured using the ddPCR linkage assay as described in the Materials and Methods above. A range of 3-7% target insertion per allele was detected in the liver at the 12-week timepoint, as shown in FIG. 4. The level of silently altered hPAH mRNA was also measured (relative to mouse GAPDH) and is shown in FIG. 5. The expression of silently altered hPAH mRNA was detected (PAH/GAPDH mRNA ratio was 0.38 ± 0.06). Of note, the level of on-target insertion per allele was found to be directly correlated to the level of mRNA expression (R-squared 0.82; $P < 0.02$).

[00176] To determine if similar levels of on-target insertion and mRNA expression could be achieved in human liver, experiments were conducted in the FRG® mouse containing a human liver xenograft. Mice were treated with a single IV dose of 1E+14 vg/kg PAH-032h packaged in AAVHSC15 capsid. PAH-032h does not contain the liver specific promoter, and

as such, expression is driven by the endogenous human promoter elements, similar to construct PAH-006m.

[00177] Expression of hPAH mRNA derived from silently altered hPAH was evaluated in liver cells from the FRG® mouse. Expression was measured by RT-ddPCR and expressed as a ratio to GAPDH to allow for comparison across human and mouse hepatocytes. In the human hepatocytes isolated from the FRG® mouse model, 6-weeks after administration of PAH-032h packaged in AAVHSC15 capsid, hPAH mRNA derived from silently altered hPAH was expressed (hPAH/GAPDH mRNA ratio = 0.57 ± 0.11). Levels of silently altered hPAH mRNA in mouse hepatocytes were significantly lower (PAH/GAPDH mRNA ratio = 0.13 ± 0.08 , $p < 0.003$ vs human) and could be entirely accounted for by the presence of contaminating human hepatocytes in the preparations used (approximately 15-20% human). The proportion of mouse cells in human samples and human cells in mouse samples was calculated by measuring the relative abundance of human and mouse GAPDH gene in each sample.

[00178] Confirmation of human-specific target insertion into the human PAH locus was done in a separate study in FRG® mice where levels of on-target insertion (~6% using the linkage ddPCR method; FIG. 6) and levels of hPAH mRNA derived from silently altered hPAH (hPAH/GAPDH mRNA ratio = 0.45 ± 0.26) were achieved in human cells with treatment of PAH-032h packaged in AAVHSC15 capsid. Target gene insertion efficiency was characterized using the NGS assay in which quantitation of integration is based on two separate measurements that cover both the left and right homology arms, respectively, with coverage extending into the genomic targeted integration sites and past the end of each homology arm, as described in the Materials and Methods above. Untreated wild-type genomic DNA was used to detect the false-positive rate for this experiment. Integration of hPAH into the PAH locus was confirmed with orthogonal assays (linkage ddPCR and NGS) reporting integration efficiencies of 6%. These results are shown in FIG. 7.

[00179] FIG. 6 and FIG. 7 show targeted insertion frequency in hepatocytes isolated from treated FRG mice. FIG. 6 shows insertion of the human-targeting PAH-032h correction vector packaged in AAVHSC15 capsid into mouse and human PAH loci, as measured by the species specific ddPCR editing assays described herein. As a positive control linkage between a mouse-specific PAH locus probe and a probe for the homologous insertion site was used. In FIG. 6, An.1 HS refers to Animal 1 *Homo sapiens* Cells, An.1 MM refers to Animal 1 *Mus musculus* cells, An.1 MM Pos. refers to Animal 1 *Mus musculus* positive control, An.2 HS refers to Animal 2 *Homo sapiens* Cells, An.2 MM refers to Animal 2 *Mus musculus* cells, and An.2 MM Pos. refers to Animal 2 *Mus musculus* positive control. FIG. 7 shows the editing of

human hepatocytes as measured by quantitative NGS assays spanning both left and right homology arms. Editing efficiency was measured by read counts of edits specific to wild-type specific sequences. Untreated human DNA mixed with PAH-032h vector served as a negative control.

5 [00180] From these studies, it can be concluded that: (1) blood Phe concentration is reduced to a therapeutically relevant threshold ($\leq 360 \mu\text{M}$) following administration of the mouse construct PAH-006m packaged in AAVHSC15 capsid into PAH^{enu2} mice; (2) similar levels of mRNA expression were observed across species between the mouse-specific construct PAH-006m in PAH^{enu2} mice and the human specific construct PAH-032h in the human liver
10 tissue of FRG® mice, both packaged in AAVHSC15 capsid; (3) no significant target insertion was observed in mouse hepatocytes (using mouse-specific assays) following administration of the human specific construct PAH-032h packaged in AAVHSC15 capsid into FRG® mice, demonstrating the sequence/species specificity of the human homology arms; and (4) similar levels of on-target integration per allele were achieved across species between the mouse-
15 specific construct PAH-006m in the PAH^{enu2} mouse (4-7%) and the human specific construct PAH-032h in the FRG mouse (~6%).

Example 3: In Vivo Efficacy of PAH-006m-LP-1 in PAH^{enu2} Mice

[00181] The ability of the PAH-006m-LP-1 vector to reduce blood Phe levels in mice
20 was assessed and correlated with hPAH mRNA expression and target insertion at the mouse PAH locus. Specifically, four PAH^{enu2} mice received single intravenous administration of the murine specific PAH-006m-LP-1 vector packaged in AAVHSC15 capsid covering a ~2-log dose range, with dose levels of 5E+12, 1E+13, 1E+14, and 2E+14 vg/kg (“GE 5E12,” “GE 1E13,” “GE 1E14,” and “GE 2E14,” respectively). In addition, a control arm containing an
25 AAV gene transfer construct expressing hPAH but with no homology arms was administered at a dose of 1E+13 vg/kg (“GT (+) control 1E13”). Four PAH^{enu2} mice were also administered formulation buffer as a negative control (“FB”). Blood Phe concentration was measured over the course of the experiment.

[00182] A subset of animals dosed at 1E+14 vg/kg were sacrificed at the 2-week
30 timepoint to measure the level of target gene insertion in the liver. A level of $5 \pm 2\%$ target gene insertion per allele was detected. The blood Phe concentrations in the mice through week 41 are shown in FIG. 8A. FIG. 8A shows the full data set, while FIG. 8B shows the data out to 12 weeks with a reduced y-axis scale to allow differences between the dose levels to be seen

more clearly. Blood Phe concentration was reduced to normal levels by one week following administration of all dose levels, with the 1E+14 and 2E+14 doses resulting in the lowest levels.

[00183] On-target analysis was performed on animals dosed at 5E12 vg/kg (“5E+12”), 1E13 vg/kg (“1E+13”), and 1E14 vg/kg (“1E+14”), as indicated (FIG. 8C). On-target
5 integration was measured by next generation sequencing as described in Example 2. Integration rates were calculated based on the number of sequences with the integration divided by total number of sequences (with or without the integration event). In FIG. 8C, all data was collected at 43 weeks post-dosing, except “1E+14 vg/kg – 3 wks” where the data was collected at 3 weeks post-dosing. As shown, on-target integration was detected in animals dosed with
10 1E14 vg/kg PAH-006m-LP-1 vector packaged in AAVHSC15 capsid at both 3 and 43 weeks post-dosing.

Example 4: In Vitro Editing Specificity of PAH-032h in Non-Human Primate and Human Hepatocytes

[00184] Cynomolgus monkeys exhibit 94.2% sequence identity at the PAH locus to the human homology arm sequences of PAH-032h. To determine the specificity of PAH-032h for the human PAH locus, the ability of PAH-032h to edit the PAH locus in a mixture of primary cynomolgus monkey and primary human hepatocytes was assessed. A total of 5E+5 hepatocytes from a single cynomolgus monkey and a single human donor were treated with
15 1.5E+5 vg/cell of PAH-032h vector packaged in an AAVHSC15 capsid. Editing of the PAH locus was assessed by PCR using primers specific for the human edited allele and primers specific for the non-human primate edited allele. These PCR primers were designed to show either presence or absence of integration and were not quantitative. Accordingly, measurements of PCR efficiency were determined in positive control DNA across a standard
20 curve of dilutions. This PCR assay had a lower limit of detection of 15 copies of edited allele (2.618E-5 amol).

[00185] Targeted integration of the payload into the PAH locus of human hepatocytes was observed. In contrast, no integration of the payload into the homologous cynomolgus monkey loci was detected. These results indicate that the human homology arms were specific
30 for the human PAH locus in these assays.

Example 5: Comparison Between Gene Transfer/Gene Editing Vector and Episomal Transgene Expression Vector

[00186] To investigate the dose response of integrated transgene expression and episomal transgene expression, a head-to-head comparison was performed using the mouse-specific gene transfer/gene editing AAV vector, PAH-006m-LP-1, and an episomal transgene expression AAV vector (containing the same transcriptional regulatory element as PAH-006m-LP-1, but lacking homology arms), packaged in AAVHSC15 capsid. The only difference between the payloads of the two vectors was the presence of two silent nucleotide changes in the PAH coding sequence of PAH-006m-LP-1. Five male, 4-week-old PAH^{enu2} mice were intravenously administered 1E12 vg/kg (“1E+12”), 5E12 vg/kg (“5E+12”), 1E13 vg/kg (“1E+13”), 5E13 vg/kg (“5E+13”), or 1E14 vg/kg (“1E+14”) doses of either the mouse-specific gene transfer/gene editing vector or the episomal transgene expression vector, in each case packaged in AAVHSC15, or a formulation buffer control (FB). For the episomal transgene expression vector packaged in AAVHSC15, after an official titering procedure, the titer of the lot used was accordingly adjusted to 7.8E11 vg/kg (“7.8E+11”), 3.9E12 vg/kg (“3.9E+12 vg/kg”), 7.8E12 vg/kg (“7.8E+12”), 3.9E13 vg/kg (“3.9E+13”), and 7.8E13 vg/kg (“7.8E+13”). It was found that targeted integration in mice dosed with PAH-006m-LP-1 was detectable and stable. A dose response between phenotype correction and levels of integration was also found.

[00187] FIGs. 9A and 9B show 12-week time courses of the serum Phe levels in the mice that were dosed as described above. As shown, a reduction in serum Phe was observed in mice that received the episomal transgene expression vector packaged in AAVHSC15 capsid (FIG. 9A), or the gene transfer/gene editing vector packaged in AAVHSC15 capsid (FIG. 9B). Corresponding increases in blood Tyr concentration, a downstream metabolite of Phe, were also observed in mice that received the episomal transgene expression vector packaged in AAVHSC15 capsid (FIG. 9C), or the gene transfer/gene editing vector packaged in AAVHSC15 capsid (FIG. 9D).

[00188] FIGs. 9E and 9F show the vector genome levels in PAH^{enu2} mice dosed with the episomal transgene expression vector (FIG. 9E) or PAH-006m-LP1 (FIG. 9F), in each case packaged in AAVHSC15 capsid, at the indicated doses. Vector genomes per ug of DNA were measured using quantitative PCR using a coding sequence specific primer and probes relative to input genomic DNA target. A dose response in both vector genome copy number and mRNA was observed for both PAH-006m-LP-1 and the episomal transgene expression vector. The vector copy number for PAH-006m-LP-1 (FIG. 9F) was higher across all doses compared to

the episomal transgene expression vector (FIG. 9E). Differences in VG/ug between the transgene and the integration vector are observed and may reflect differences in stability between self-complementary and single-stranded vectors.

[00189] FIGs. 9G and 9H show the PAH transgene mRNA expression levels in PAH^{enu2} mice dosed with the episomal transgene expression vector (FIG. 9G) or PAH-006m-LP1 (FIG. 9H), in each case packaged in AAVHSC15 capsid, at the indicated doses. mRNA levels were measured by quantitative RT-PCR using coding sequence specific primers and probes relative to total RNA. As shown, mRNA expression was found to be dose-responsive, and expression level was found to be slightly higher in mice that received PAH-006m-LP-1 packaged in AAVHSC15 capsid as compared to mice that received the episomal transgene expression vector packaged in AAVHSC15 capsid.

[00190] To assess the quantity and fidelity of target locus integration in PAH^{enu2} mice, two methods were performed to detect integration events: ddPCR and next generation sequencing.

[00191] The level of on-target integration at the PAH locus was measured using the ddPCR linkage assay as described in Example 2. FIG. 9I shows the amount of on-target integration (measured in viral genomes per PAH allele detected by ddPCR) at 12 weeks post-dosing for the various doses of PAH-006m-LP1 and episomal transgene expression vector control indicated. As shown, a dose response of targeted integration detected by ddPCR was observed in mice that received PAH-006m-LP1 packaged in AAVHSC15 capsid (FIG. 9I; Groups 2, 3, 4, 5, and 6). In comparison, no targeted integration was detected by ddPCR in mice that received the episomal transgene expression vector packaged in AAVHSC15 capsid (FIG. 9I; Group 11).

[00192] FIG. 9J shows the frequency of on-target vector insertion detected (using the NGS approach described in Example 2) 12 weeks post-dosing with the indicated doses of PAH-006m-LP1 packaged in AAVHSC15. As shown, a dose response of targeted integration detected by NGS was observed in mice that received PAH-006m-LP1 packaged in AAVHSC15 capsid. In comparison, no targeted integration was detected by NGS in mice that received the episomal transgene expression vector packaged in AAVHSC15 capsid. Integration level was found to be dose-responsive and higher than 2% at doses greater than 5E13 vg/kg.

[00193] To evaluate the effect of integrated transgene expression and episomal transgene expression in proliferating cell populations, 10-week-old PAH^{enu2} mice were intravenously administered 5E12 vg/kg, 2E13 vg/kg, 6E13 vg/kg, or 1E15 vg/kg doses of either

the mouse-specific gene transfer/gene editing vector or the episomal transgene expression vector, in each case packaged in AAVHSC15, or a formulation buffer control (FB). For the episomal transgene expression vector packaged in AAVHSC15, after an official titering procedure, the titer of the lot used was accordingly adjusted to 3.92E12 vg/kg (“3.92E+12”), 1.57E13 vg/kg (“1.57E+13 vg/kg”), 4.71E13 vg/kg (“4.71E+13”), and 7.84E13 vg/kg (“7.84E+13”). About 2 weeks post-dosing, the mice underwent either a 70% partial hepatectomy (PHx) to induce rapid hepatocyte cycling, or a sham surgery (sham). FIGs. 9K and 9L show a time course out to 42 weeks (FIG. 9K) or 40 weeks (FIG. 9L) of the serum Phe levels in these mice. As shown, in mice treated with the episomal transgene expression vector (FIG. 9K), serum Phe levels were reduced at weeks 1 and 2. Following hepatectomy at the 2-week timepoint, the Phe levels at the 5E12 vg/kg and 2E13 vg/kg dose levels increased during the period of accelerated liver regeneration. The sham surgery animals treated with 2E13 vg/kg episomal transgene expression vector did not show a loss of this response, indicating that the increase in serum Phe levels in the episomal transgene expression vector treated hepatectomized mice may be due to vector dilution under conditions of accelerated liver growth that occurred following hepatectomy. Mice treated with the gene transfer/gene editing vector (FIG. 9L) showed a stable reduction in Phe levels throughout the period of accelerated liver growth in the partially hepatectomized mice. This result may be indicative of the benefit of integration at the target mouse *Pah* locus.

Example 6: Comparison Between Mouse-Specific Vector and Human-Specific Vector

[00194] To investigate dose range differences between the mouse-specific gene transfer/gene editing AAV vector (PAH-006m-LP-1; “mouse design”), and the human-specific gene transfer/gene editing AAV vector (hPAH-h11C-032-LP1-SD3; “human design”) was used in a head-to-head study. Four to five male, 4-week-old (FIGs. 10A-10H) or 10-week-old (FIGs. 10I-10P) PAH^{enu2} mice were intravenously administered 1E12 vg/kg, 5E12 vg/kg, 1E13 vg/kg, or 5E13 vg/kg of either the mouse design vector (FIGs. 10A, 10C, 10E, 10G, 10I, 10K, 10M, and 10O) or the human design vector (FIGs. 10B, 10D, 10F, 10H, 10J, 10L, 10N, and 10P), in each case packaged in AAVHSC15, or a formulation buffer control (FB).

[00195] The human design vector comprises human homology arm sequences and would not be expected to integrate into the mouse genome via homologous recombination due to sequence differences, accordingly the human design vector served as an episomal-only control in these experiments. As shown, 4-week-old mice dosed with the human design vector packaged in AAVHSC15 capsid (FIG. 10B) exhibited an improvement in serum Phe reduction,

similar to mice administered the mouse design vector (FIG. 10A). A corresponding increase in blood Tyr concentration was also observed in the 4-week-old mice dosed with the human design vector packaged in AAVHSC15 capsid (FIG. 10D) and in the mice dosed with the mouse design vector (FIG. 10C). As shown in FIGs. 10E and 10F, a dose response was observed for vector genome levels detected in the liver of 4-week-old mice dosed with the human design vector (FIG. 10F) or the mouse design vector (FIG. 10E) packaged in AAVHSC15 capsid. A dose response was also observed for mRNA levels detected in the liver of 4-week-old mice dosed with the human design vector (FIG. 10H) or the mouse design vector (FIG. 10G) packaged in AAVHSC15 capsid.

[00196] At 10 weeks old, mice dosed with the human design vector packaged in AAVHSC15 capsid (FIG. 10I) were found to have a similar reduction in serum Phe levels, compared to mice dosed with the mouse design vector (FIG. 10H). A corresponding increase in blood Tyr concentration was also observed in the 10-week-old mice dosed with the human design vector packaged in AAVHSC15 capsid (FIG. 10K) and in mice dosed with the mouse design vector (FIG. 10J). As shown in FIGs. 10M and 10N, a dose response was observed for vector genome levels detected in the liver of 10-week-old mice dosed with the human design vector (FIG. 10N) or the mouse design vector (FIG. 10M) packaged in AAVHSC15 capsid. A dose response was also observed for mRNA levels detected in the liver of 10-week-old mice dosed with the human design vector (FIG. 10P) or the mouse design vector (FIG. 10O) packaged in AAVHSC15 capsid.

Example 7: Effect of a Single Dose of Mouse-Specific Vector or Human-Specific Vector

[00197] To evaluate the effect of the age of PAHenu2 mice on the response to a single dose of the mouse-specific gene transfer/gene editing AAV vector (PAH-006m-LP-1; “mouse design”) or the human-specific gene transfer/gene editing AAV vector (hPAH-hIIC-032-LP1-SD3; “human design”), mice at 2-, 4-, and 10-weeks of age were selected. This allowed for a comparison between the effect of pediatric liver growth (2- and 4-week-old mice) and the liver of an adult (10-week-old mice). A single 1E14 vg/kg dose of the mouse design vector packaged in AAVHSC15 or the human design vector packaged in AAVHSC15 was intravenously administered to PAHenu2 mice at 2-, 4-, or 10-weeks of age. A total of four to five male mice in each age group were administered the respective packaged vectors. Formulation buffer was administered as control. As the human design vector does not integrate into the mouse *Pah* locus, it provides a control to demonstrate the effect of episomal expression on mice of different ages. In this example, after an official titering procedure, the titer of the lot used was

accordingly adjusted to 9.07E13 vg/kg for the mouse design and 1.06E14 vg/kg for the human design. The various dosing cohorts studied are set forth in Table 2.

Table 2: Dosing Cohorts

Group	Test Article	Dose Level (vg/kg)	Number of mice dosed / Sex (age at dosing)	Necropsy (weeks post-administration)
1	Formulation buffer	NA	5/M (2-weeks)	43 weeks
2	Mouse design vector packaged in AAVHSC15	1E+14 (9.07E13)	5/M (2-weeks)	
3			4/M (4-weeks)	
4			4/M (10-weeks)	
5			4/M (2-weeks)	2 weeks
6			4/M (2-weeks)	4 weeks
7			4/M (2-weeks)	8 weeks
8	Formulation buffer	NA	4/M (2-weeks)	42 weeks
9	Human design vector packaged in AAVHSC15	1E+14 (1.06E14)	4/M (2-weeks)	
10			4/M (4-weeks)	
11			4/M (10-weeks)	
12			4/M (2-weeks)	2 weeks
13			4/M (2-weeks)	4 weeks
14			4/M (2-weeks)	8 weeks
15	Mouse design vector packaged in AAVHSC15		4/M (4-weeks)	42 weeks

5

[00198] As set forth in Table 2, Groups 1-4 were sacrificed 43 weeks post-administration, and Groups 8-11, and 15 were sacrificed 42 weeks post-administration. Serum Phe and Tyr levels were examined over time for Groups 1-4, 8-11, and 15 (FIGs. 11A-11D). FIGs. 11E-11K show the vector genome levels and mRNA levels detected in the livers of mice at the necropsy time points indicated in Table 2. mRNA levels in the livers of Groups 5-7 mice were not measured.

10

[00199] In mice treated with the mouse design vector, serum Phe levels were reduced to normal levels within one week following administration of the vector (FIG. 11A). For the treated 4- and 10-week-old mice (Groups 3 and 4), the reduction in Phe levels was maintained through 43 weeks, demonstrating long term durability of response to treatment. For the 2-week-old mice, the results were variable. A corresponding increase in serum Tyr concentration was observed in the 4- and 10-week-old mice through 43 weeks (FIG. 11C). FIG. 11E shows

15

the vector genome levels detected in the liver of 2-, 4-, and 10-week-old mice treated with the mouse design vector. FIG. 11I shows the mRNA levels detected in the liver of 2-, 4-, and 10-week-old mice treated with the mouse design vector.

[00200] In mice treated with the human design vector, serum Phe levels were reduced by the one week time point following administration of the vector (FIG. 11B). For the treated 4- and 10-week-old mice (Groups 10 and 11), the reduction in Phe levels was maintained through 43 weeks, indicating long term durability of response to treatment. In the treated 2-week-old mice (Group 9), Phe levels maintained at clinically relevant levels of $\leq 360 \mu\text{M}$ through 30 weeks. A corresponding increase in serum Tyr concentration was observed in Groups 9 (through 30 weeks), 10 (through 42 weeks), and 11 (through 42 weeks) (FIG. 11D). These data indicate that a $1\text{E}14 \text{ vg/kg}$ dose of the human design vector packaged in AAVHSC15 is sufficient to provide sustained efficacy through the period of liver growth in a juvenile mouse, and that the effect is sustained for at least 30 weeks post-administration. FIG. 11F shows the vector genome levels detected in the liver of 2-, 4-, and 10-week-old mice treated with the human design vector. FIG. 11J shows the mRNA levels detected in the liver of 2-, 4-, and 10-week-old mice treated with the human design vector.

[00201] FIGs. 11G and 11H show that in 2-week-old mice treated with the mouse design vector (FIG. 11G; Groups 5-7) and 2-week-old mice treated with the human design vector (FIG. 11H; Groups 12-14), vector genome levels were sustained up to 8 weeks post-administration. FIG. 11K shows that mRNA levels of 2-week-old mice treated with the human design vector were sustained up to 8 weeks post-administration.

Example 8: Kinetics and Durability of Integration in Adult PAH^{enu2} Mice

[00202] To determine kinetics and durability of vector integration over time, 10-week-old male PAH^{enu2} mice were administered a mouse-specific gene transfer/gene editing AAV vector (PAH-006m-LP-1) packaged in AAVHSC15 capsid at a dose of $1\text{E}14 \text{ vg/kg}$. Six cohorts were sacrificed at 1, 2, 4, 8, and 16 weeks post-dosing (five mice per cohort). The kinetics of integration was investigated by monitoring serum Phe levels over time, vector genome levels over time, PAH transgene mRNA expression levels over time, and targeted integration levels over time, at the various time points. It was found that vector genome levels, PAH transgene mRNA expression levels, and integration frequencies were stable over time (FIGs. 12A-12C).

[00203] FIGs. 12A and 12B show a time course of vector genome levels (FIG. 12A) and PAH transgene mRNA expression levels (FIG. 12B) in PAH^{enu2} mice administered PAH-

006m-LP-1 packaged in AAVHSC15 capsid at a dose of 1E14 vg/kg. The vector genome and the mRNA levels over time were found to be stable up to 42 weeks. Vector genome levels were measured using quantitative PCR using a coding sequence specific primer and probes relative to input genomic DNA target. Measurement of expression was performed by quantitative RT-PCR using coding sequence specific primers and probes relative to total RNA.

[00204] A long-read next generation sequencing (NGS) method was developed to quantitate integrated versus unintegrated alleles. Long-read sequencing was selected due to the length of the homology arms requiring a read through of the integration sequence and into the genomic DNA. The method was able to distinguish between integrated alleles, wild-type alleles, and vector genomes that are present in a given sample. PCR is performed using three primers specific to i) the genomic region; ii) a region specific to the integrated allele; and iii) a region specific to the wild-type allele. The frequency of on-target vector integration was determined by competitively amplifying both wild-type and integrated alleles at the PAH region using DNA derived from the livers of treated animals. The long-read sequencing of the amplified products covered the homology arm, which was identical across genomic, integrated, and vector genome sequences. The adjoining sequences, which were also covered by long-read sequencing, determine whether the source of the read was from wild-type, integrated, or vector genomes. Contiguous sequences that included both the genomic DNA and the silently altered hPAH transgene were tallied as integrated alleles, while sequences that included only the genomic DNA without the silently altered hPAH transgene in the target integration site were tallied as wild-type alleles. Percentage of integrated sequences were calculated as the number of reads mapped to the integrated reference divided by the total number of reads mapped to the integrated reference, reads mapped to the wild-type reference, and reads mapped to vector genome concatemers. FIG. 12C shows a time course of targeted integration frequency in PAH^{enu2} mice administered PAH-006m-LP-1 packaged in AAVHSC15 capsid at a dose of 1E14 vg/kg. On-target integration was detected at one week post-dosing, achieved peak levels by about 2 weeks, and remained consistent through 42 weeks.

[00205] FIGs. 12D and 12E are graphs showing the effect on serum Phe (FIG. 12D) and Tyr (FIG. 12E) in PAH^{enu2} mice administered PAH-006m-LP-1 packaged in AAVHSC15 capsid at a dose of 1E14 vg/kg, up to 42 weeks post-injection.

Example 9: In Vivo Efficacy of hPAH-h11C-032-LP1-SD3

[00206] To determine kinetics and durability of integration of a human specific gene transfer/gene editing vector over time, male humanized liver FRG mice ("HuLiv" mice; see,

Example 2) were administered the human-specific gene transfer/gene editing vector, hPAH-hIIC-032-LP1-SD3, packaged in AAVHSC15 capsid, at a dose of 7E13 vg/kg (Groups 1 and 3) or 2E14 vg/kg (Groups 2 and 4), and sacrificed at 4 (Groups 1 and 2) or 12 weeks (Groups 3 and 4) post-dosing (six male mice per cohort). Control FRG mice were administered formulation buffer and sacrificed at 4 (Group 5) or 12 weeks (Group 6) post-administration.

[00207] FIGs. 13A and 13B show the vector genome levels and mRNA levels detected in human hepatocytes isolated from HuLiv mice at 4 and 12 weeks after administration of hPAH-hIIC-032-LP1-SD3 packaged in AAVHSC15, at the indicated doses. As shown, the level of vector genomes (FIG. 13A) and mRNA expression (FIG. 13B) in human hepatocytes were found to be dose responsive and stable over time up to 12 weeks. Vector genome levels were measured using quantitative PCR using a coding sequence specific primer and probes relative to input genomic DNA target. Measurement of expression was performed by quantitative RT-PCR using coding sequence specific primers and probes relative to total RNA. FIG. 13C shows the frequency of on-target integration detected in human hepatocytes isolated from HuLiv mice at 4 and 12 weeks after administration of hPAH-hIIC-032-LP1-SD3 packaged in AAVHSC15, at the indicated doses. As shown, the level of on-target integration in human hepatocytes was found to be dose responsive and stable over time up to 12 weeks. Measurement of targeted integration was carried out using the long-read NGS method described in Example 8.

[00208] Given that homologous recombination may be driven by short stretches of sequence similarity between the homology arms of the vector and an off-target location in the genome, a PCR-based assay was developed to specifically test for integration into genomic regions that contain sequence similar to the homology arms. The method bioinformatically predicted the most likely regions to undergo off-target homologous recombination in human cells transduced with vector containing homology arms specific to the human PAH locus. These predicted regions were selected based on the highest sequence similarity to the homology arms. A PCR-based method was designed to specifically test transduced samples for any occurrence of off-target integration in the predicted regions. Using this PCR-based method, integration to a level of 1:10000 DNA molecules was able to be detected.

[00209] Predicted off-target integration sites were selected based on two criteria: a minimum sequence length of 35 bp; and a minimum sequence identity of 60% relative to the homology arms. Six regions of the genome were identified that met these criteria via bioinformatic alignment against Genome Reference Consortium Human Build 38. PCR primers that allow specific amplification for off-target integration at 5 of these 6 regions were

designed. Region 3 was found to be in a highly repetitive region and was dropped from further analysis because specific primers could not be generated. As positive controls (“Pos”), specific control DNAs spiked into genomic DNA, were run for each specific region down to a dilution of 1:10,000, representing 0.01% and the limit of detection of the PCR-based method. For each
5 predicted region, integration-specific primer pairs targeting homologous recombination in either direction (inward (“In”) or outward (“Out”) from the region of homology) were tested independently to query for off-target integration. Off-target integration was defined as positive if a distinct PCR band of the correct size was identified in a sample.

[00210] FIGs. 14A-14D show the results of predicted off-target integration determined
10 by the PCR-based method described above, performed on genomic DNA samples isolated from 10-week-old male humanized liver FRG mice administered the human-specific gene transfer/gene editing vector, hPAH-hIIC-032-LP1-SD3, packaged in AAVHSC15 capsid, at a dose of 7E13 vg/kg (FIGs. 14A and 14C) or 2E14 vg/kg (FIGs. 14B and 14D), and sacrificed at 4 (FIGs. 14A and 14B) or 12 weeks (FIGs. 14C and 14D) post-dosing (six male mice per
15 cohort). As shown, no integration was detected at the predicted off-target integration sites.

* * *

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

[00212] All references (*e.g.*, publications or patents or patent applications) cited herein
25 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 recombinant adeno-associated virus (rAAV) comprising:
 - (a) an AAV capsid comprising an AAV capsid protein; and
 - (b) an rAAV genome comprising: (i) an editing element for editing a target locus in a PAH gene, comprising at least a portion of a PAH coding sequence operably linked to a transcriptional regulatory element; (ii) a 5' homology arm nucleotide sequence position 5' of the editing element, having homology to a first genomic region 5' to the target locus; and (iii) a 3' homology arm nucleotide sequence positioned 3' of the editing element, having homology to a second genomic region 3' to the target locus.
2. The rAAV of claim 1, wherein the editing element comprises a PAH coding sequence.
3. The rAAV of claim 2, wherein the PAH coding sequence encodes an amino acid sequence set forth in SEQ ID NO: 33.
4. The rAAV of claim 1 or 2, wherein the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 53.
5. The rAAV of claim 2, wherein the PAH coding sequence is silently altered.
6. The rAAV of claim 5, wherein the PAH coding sequence comprises a nucleotide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the nucleotide sequence set forth in SEQ ID NO: 28.
7. The rAAV of claim 5 or 6, wherein the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 28.
8. The rAAV of claim 5 or 6, wherein the PAH coding sequence comprises the nucleotide sequence set forth in SEQ ID NO: 83.
9. The rAAV of any one of claims 1-8, 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.

10. The rAAV of any one of claims 1-9, wherein the transcriptional regulatory element is endogenous to the PAH gene.
11. The rAAV of any one of claims 1-9, wherein the transcriptional regulatory element is exogenous to the PAH gene.
12. The rAAV of claim 11, wherein the transcriptional regulatory element is liver specific, optionally wherein the transcriptional regulatory element comprises one or more elements selected from the group consisting of a human albumin promoter, a human transthyretin (TTR) promoter, a human ApoE/C-I hepatic control region (HCR) 1 or 2, a human ApoH promoter, a human SERPINA1 (hAAT) promoter, and a hepatic specific regulatory module thereof.
13. The rAAV of claim 11 or 12, wherein the transcriptional regulatory element comprises a nucleotide sequence at least 90% identical to a sequence selected from the group consisting of SEQ ID NO: 25, 26, 27, and 69.
14. The rAAV of any one of claims 11-13, wherein the transcriptional regulatory element comprises a nucleotide sequence at least 90% identical to the nucleotide sequence set forth in SEQ ID NO: 27.
15. The rAAV of any one of claims 11-14, wherein the transcriptional regulatory element comprises the nucleotide sequence set forth in SEQ ID NO: 27.
16. The rAAV of any one of claims 11-15, wherein the nucleotide sequence of the transcriptional regulatory element consists of the nucleotide sequence set forth in SEQ ID NO: 27.
17. The rAAV of any one of claims 1-16, wherein the editing element further comprises an intron element positioned 5' to the PAH coding sequence and 3' to the transcriptional regulatory element.
18. The rAAV of claim 17, wherein the intron element is an exogenous intron element, optionally wherein the exogenous intron element is an SV40 intron element.

19. The rAAV of claim 18, wherein the SV40 intron element comprises a nucleotide sequence at least 90% identical to the nucleotide sequence set forth in SEQ ID NO: 29.
20. The rAAV of claim 18 or 19, wherein the SV40 intron element comprises the nucleotide sequence set forth in SEQ ID NO: 29.
21. The rAAV of any one of claims 18-20, wherein the nucleotide sequence of the SV40 intron element consists of the nucleotide sequence set forth in SEQ ID NO: 29.
22. The rAAV of any one of claims 1-21, wherein the editing element further comprises a polyadenylation sequence 3' to the PAH coding sequence.
23. The rAAV of claim 22, wherein the polyadenylation sequence is an exogenous polyadenylation sequence, optionally wherein the exogenous polyadenylation sequence is an SV40 polyadenylation sequence.
24. The rAAV of claim 23, wherein the SV40 polyadenylation sequence comprises a nucleotide sequence at least 90% identical to the nucleotide sequence set forth in SEQ ID NO: 31.
25. The rAAV of claim 23 or 24, wherein the SV40 polyadenylation sequence comprises the nucleotide sequence set forth in SEQ ID NO: 31.
26. The rAAV of any one of claims 23-25, wherein the nucleotide sequence of the SV40 polyadenylation sequence consists of the nucleotide sequence set forth in SEQ ID NO: 31.
27. The rAAV of any one of claims 1-26, wherein the nucleotide 5' to the target locus is in an intron of a PAH gene.
28. The rAAV of any one of claims 1-27, wherein the nucleotide 5' to the target locus is in intron 1 of a PAH gene.
29. The rAAV of any one of claims 1-28, wherein the nucleotide 3' to the target locus is in an intron of a PAH gene.

30. The rAAV of any one of claims 1-29, wherein the nucleotide 3' to the target locus is in intron 1 of a PAH gene.
31. The rAAV of any one of claims 1-30, wherein the PAH gene is a human PAH gene.
32. The rAAV of claim 31, wherein the human PAH gene is wild-type.
33. The rAAV of claim 31 or 32, wherein the human PAH gene is a variant PAH gene.
34. The rAAV of any one of claims 1-33, wherein the editing element comprises a nucleotide sequence selected from the group consisting of SEQ ID NO: 25, 26, 27, 28, 29, 31, 50, 51, 52, 69, or 70.
35. The rAAV of any one of claims 1-34, wherein the 5' homology arm nucleotide sequence is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the first genomic region.
36. The rAAV of any one of claims 1-35, wherein the 3' homology arm nucleotide sequence is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the second genomic region.
37. The rAAV of any one of claims 1-36, wherein the first genomic region is located in a first editing window, and the second genomic region is located in a second editing window.
38. The rAAV of claim 37, wherein the first editing window consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 37.
39. The rAAV of claim 37 or 38, wherein the second editing window consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 38.
40. The rAAV of any one of claims 1-39, wherein the first genomic region consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 37.

41. The rAAV of any one of claims 1-40, wherein the second genomic region consists of the region of the human genome corresponding to the nucleotide sequence set forth in SEQ ID NO: 38.
42. The rAAV of any one of claims 1-41, wherein each of the 5' and 3' homology arm nucleotide sequences independently has a length of about 100 to about 2000 nucleotides.
43. The rAAV of any one of claims 1-42, wherein the 5' homology arm comprises a nucleotide sequence that is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleotide sequence set forth in SEQ ID NO: 23.
44. The rAAV of any one of claims 1-43, wherein the 5' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 23.
45. The rAAV of any one of claims 1-44, wherein the nucleotide sequence of the 5' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 23.
46. The rAAV of any one of claims 1-45, wherein the 3' homology arm comprises a nucleotide sequence that is at least 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleotide sequence set forth in SEQ ID NO: 24.
47. The rAAV of any one of claims 1-46, wherein the 3' homology arm comprises the nucleotide sequence set forth in SEQ ID NO: 24.
48. The rAAV of any one of claims 1-47, wherein the nucleotide sequence of the 3' homology arm consists of the nucleotide sequence set forth in SEQ ID NO: 24.
49. The rAAV of any one of claims 1-48, wherein the rAAV genome comprises a nucleotide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the nucleotide sequence set forth in SEQ ID NO: 43.

50. The rAAV of any one of claims 1-49, wherein the rAAV genome comprises the nucleotide sequence set forth in SEQ ID NO: 43.

51. The rAAV of any one of claims 1-50, wherein the rAAV genome further comprises a 5' inverted terminal repeat (5' ITR) nucleotide sequence 5' of the 5' homology arm nucleotide sequence, and a 3' inverted terminal repeat (3' ITR) nucleotide sequence 3' of the 3' homology arm nucleotide sequence.

52. The rAAV of claim 51, wherein the 5' ITR nucleotide sequence is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleotide sequence set forth in SEQ ID NO: 14, and/or the 3' ITR nucleotide sequence is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the nucleotide sequence set forth in SEQ ID NO: 18.

53. The rAAV of any one of claims 1-52, wherein the rAAV genome comprises the nucleotide sequence set forth in SEQ ID NO: 73 and/or 74.

54. The rAAV of any one of claims 1-53, wherein the rAAV genome comprises the nucleotide sequence set forth in SEQ ID NO: 75 and/or 76.

55. The rAAV of any one of claims 1-54, wherein the rAAV genome comprises a nucleotide sequence that is at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, or 99% identical to the nucleotide sequence set forth in SEQ ID NO: 45.

56. The rAAV of any one of claims 1-55, wherein the rAAV genome comprises the nucleotide sequence set forth in SEQ ID NO: 45.

57. The rAAV of any one of claims 1-56, wherein the AAV capsid protein comprises an amino acid sequence that is at least 95% identical to the amino acid sequence of amino acids 203-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

58. The rAAV of claim 57, wherein: the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 16 is C; the amino acid in the capsid protein corresponding to

amino acid 296 of SEQ ID NO: 16 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 16 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 16 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 16 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G.

59. The rAAV of claim 58, wherein:

- (a) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G;
- (b) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M;
- (c) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R;
- (d) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; or
- (e) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C.

60. The rAAV of claim 58, wherein the capsid protein comprises the amino acid sequence of amino acids 203-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

61. The rAAV of any one of claims 1-60, wherein the AAV capsid protein comprises an amino acid sequence that is at least 95% identical to the amino acid sequence of amino acids 138-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

62. The rAAV of claim 61, wherein: the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 16 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 16 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 16 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 16 is S; the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 16 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G.

63. The rAAV of claim 62, wherein:

(a) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G;

(b) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is

R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M;

(c) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R;

(d) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; or

(e) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C.

64. The rAAV of claim 62, wherein the capsid protein comprises the amino acid sequence of amino acids 138-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 15, 16, or 17.

65. The rAAV of any one of claims 1-64, wherein the AAV capsid protein comprises an amino acid sequence that is at least 95% identical to the amino acid sequence of amino acids 1-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

66. The rAAV of claim 65, wherein: the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 16 is T; the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 16 is I; the amino acid in the capsid protein corresponding to amino acid 68 of SEQ ID NO: 16 is V; the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 16 is L; the amino acid in the capsid protein corresponding to amino acid 151 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 160 of SEQ ID NO: 16 is D; the amino acid in the capsid protein corresponding to amino acid 206 of SEQ ID NO: 16 is C; the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H; the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 16 is Q; the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A; the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N; the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 16 is S; the amino acid in the capsid protein corresponding to

amino acid 501 of SEQ ID NO: 16 is I; the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 590 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G or Y; the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M; the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R; the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 16 is K; the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C; or, the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G.

67. The rAAV of claim 66, wherein:

- (a) the amino acid in the capsid protein corresponding to amino acid 2 of SEQ ID NO: 16 is T, and the amino acid in the capsid protein corresponding to amino acid 312 of SEQ ID NO: 16 is Q;
- (b) the amino acid in the capsid protein corresponding to amino acid 65 of SEQ ID NO: 16 is I, and the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is Y;
- (c) the amino acid in the capsid protein corresponding to amino acid 77 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 690 of SEQ ID NO: 16 is K;
- (d) the amino acid in the capsid protein corresponding to amino acid 119 of SEQ ID NO: 16 is L, and the amino acid in the capsid protein corresponding to amino acid 468 of SEQ ID NO: 16 is S;
- (e) the amino acid in the capsid protein corresponding to amino acid 626 of SEQ ID NO: 16 is G, and the amino acid in the capsid protein corresponding to amino acid 718 of SEQ ID NO: 16 is G;
- (f) the amino acid in the capsid protein corresponding to amino acid 296 of SEQ ID NO: 16 is H, the amino acid in the capsid protein corresponding to amino acid 464 of SEQ ID NO: 16 is N, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 681 of SEQ ID NO: 16 is M;
- (g) the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 687 of SEQ ID NO: 16 is R;

(h) the amino acid in the capsid protein corresponding to amino acid 346 of SEQ ID NO: 16 is A, and the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R; or

(i) the amino acid in the capsid protein corresponding to amino acid 501 of SEQ ID NO: 16 is I, the amino acid in the capsid protein corresponding to amino acid 505 of SEQ ID NO: 16 is R, and the amino acid in the capsid protein corresponding to amino acid 706 of SEQ ID NO: 16 is C.

68. The rAAV of claim 66, wherein the capsid protein comprises the amino acid sequence of amino acids 1-736 of SEQ ID NO: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 15, 16, or 17.

69. The rAAV of any one of claims 1-68, wherein the integration efficiency of the editing element into the target locus is at least 1% when the AAV is administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions.

70. The rAAV of any one of claims 1-69 wherein the allelic frequency of integration of the editing element into the target locus is at least 0.5% when the AAV is administered to a mouse implanted with human hepatocytes in the absence of an exogenous nuclease under standard AAV administration conditions.

71. A pharmaceutical composition comprising an rAAV of any one of claims 1-70.

72. A polynucleotide comprising the nucleic acid sequence set forth in SEQ ID NO: 43, 45, 51, or 52.

73. A method for treating a subject having phenylketonuria (PKU), the method comprising administering to the subject an effective amount of the rAAV of any one of claims 1-70, the pharmaceutical composition of claim 71, or the polynucleotide of claim 72.

74. The method of claim 73, wherein the rAAV, pharmaceutical composition, or nucleic acid is administered intravenously.

75. The method of claim 73 or 74, wherein PKU is associated with a PAH gene mutation.

76. The method of any one of claims 73-75, wherein the subject is a human subject.
77. A packaging system for preparation of an rAAV, wherein the packaging system comprises:
- (a) a first nucleotide sequence encoding one or more AAV Rep proteins;
 - (b) a second nucleotide sequence encoding a capsid protein of the rAAV of any one of claims 1-70; and
 - (c) a third nucleotide sequence comprising an rAAV genome sequence of the rAAV of any one of claims 1-70.
78. The packaging system of claim 77, wherein the packaging system comprises a first vector comprising the first nucleotide sequence and the second nucleotide sequence, and a second vector comprising the third nucleotide sequence.
79. The packaging system of claim 77 or 78, further comprising a fourth nucleotide sequence comprising one or more helper virus genes.
80. The packaging system of claim 79, wherein the fourth nucleotide sequence is comprised within a third vector.
81. The packaging system of any one of claims 77-80, wherein the fourth nucleotide sequence comprises one or more genes from a virus selected from the group consisting of adenovirus, herpes virus, vaccinia virus, and cytomegalovirus (CMV).
82. The packaging system of any one of claims 77-81, wherein the first vector, second vector, and/or the third vector is a plasmid.
83. A method for recombinant preparation of an rAAV, the method comprising introducing the packaging system of any one of claims 77-82 into a cell under conditions whereby the rAAV is produced.
84. The rAAV of any one of claims 1-70, the pharmaceutical composition of claim 71, or the polynucleotide of claim 72, for use as a medicament.

85. The rAAV of any one of claims 1-70, the pharmaceutical composition of claim 71, or the polynucleotide of claim 72, for use in the treatment of PKU.

86. The rAAV of any one of claims 1-70, the pharmaceutical composition of claim 71, or the polynucleotide of claim 72, for use in a method of treating a subject having PKU, the method comprising administering to the subject an effective amount of the rAAV, the pharmaceutical composition, or the polynucleotide.

FIG. 1

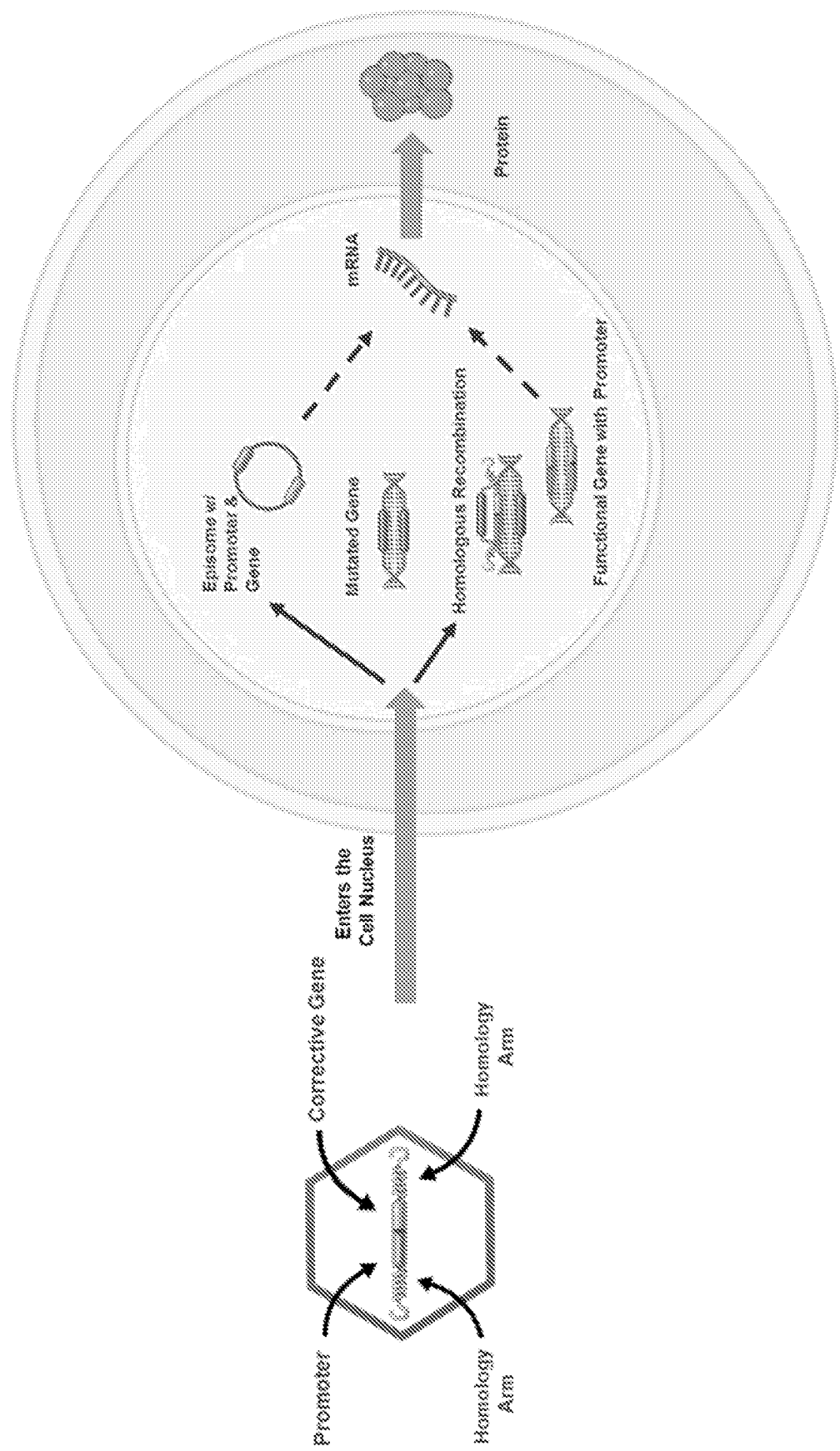


FIG. 2A

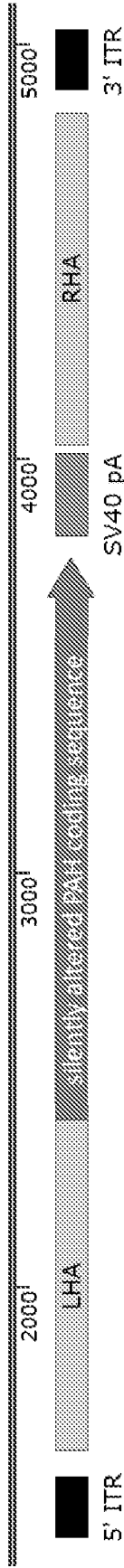


FIG. 2B

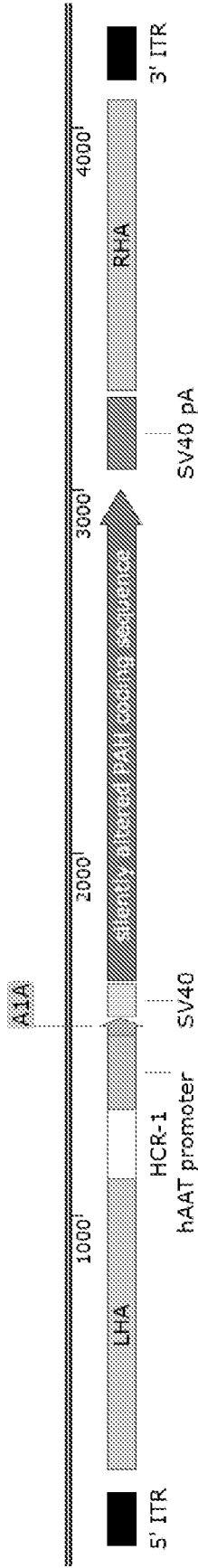


FIG. 2C

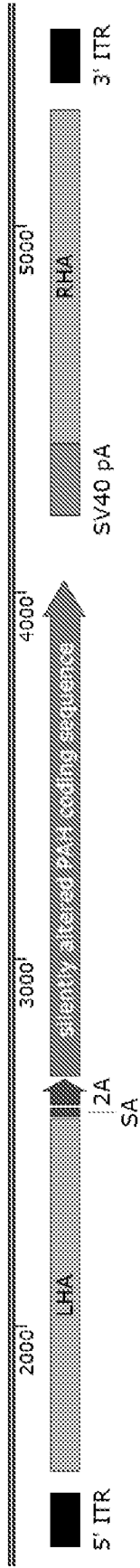


FIG. 2D

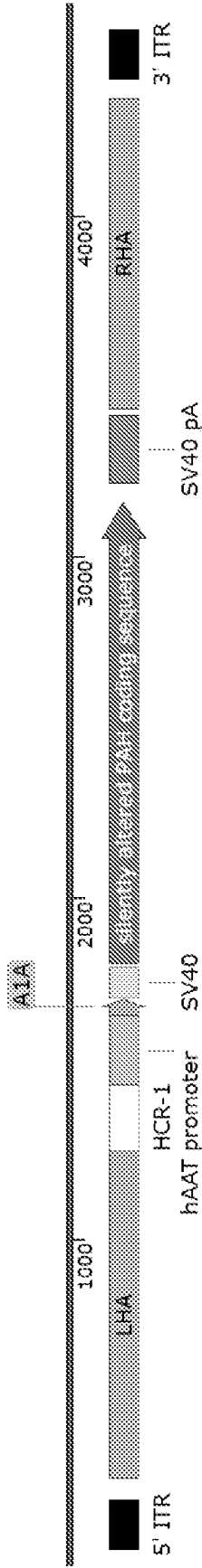


FIG. 3

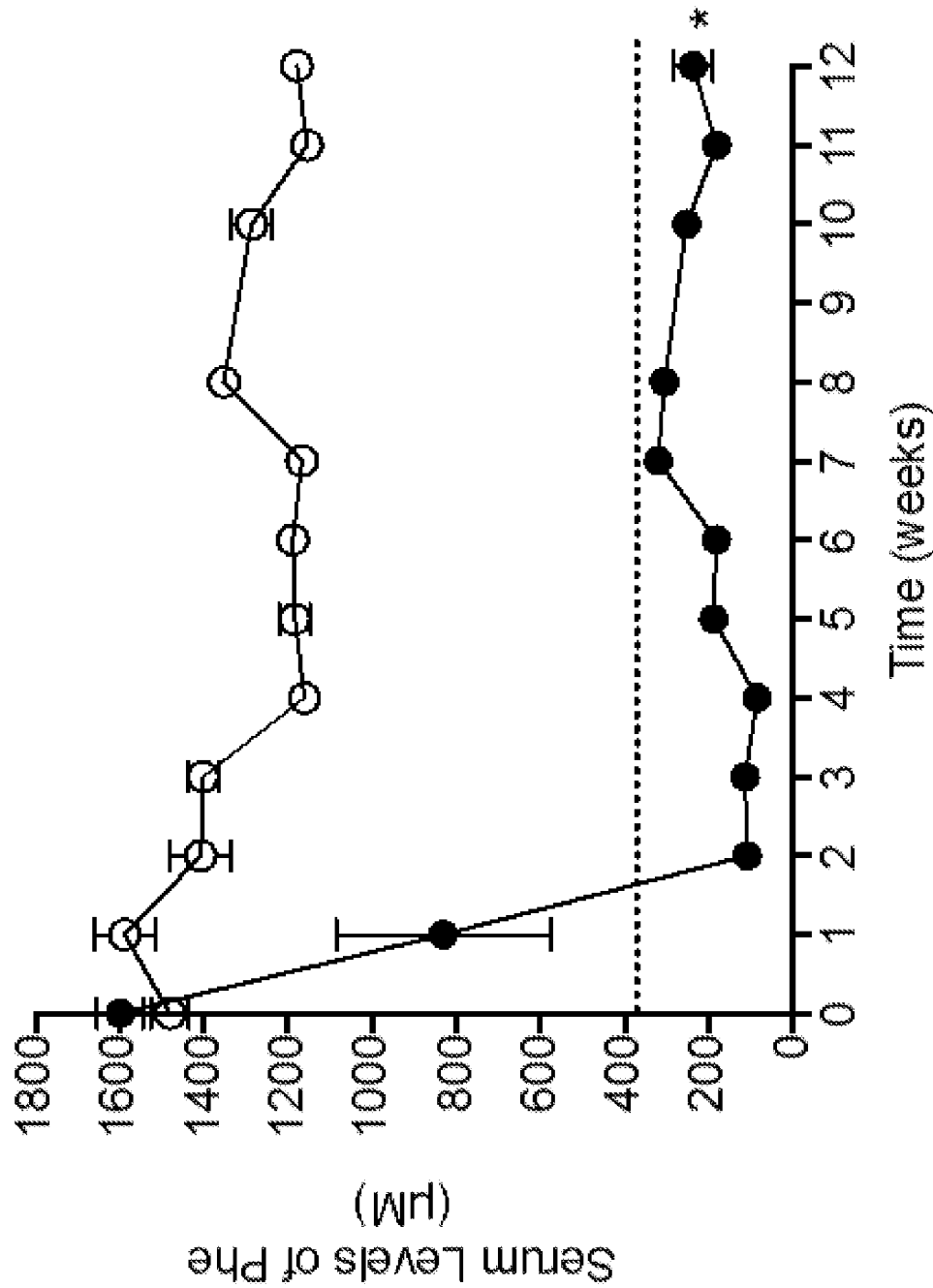


FIG. 4

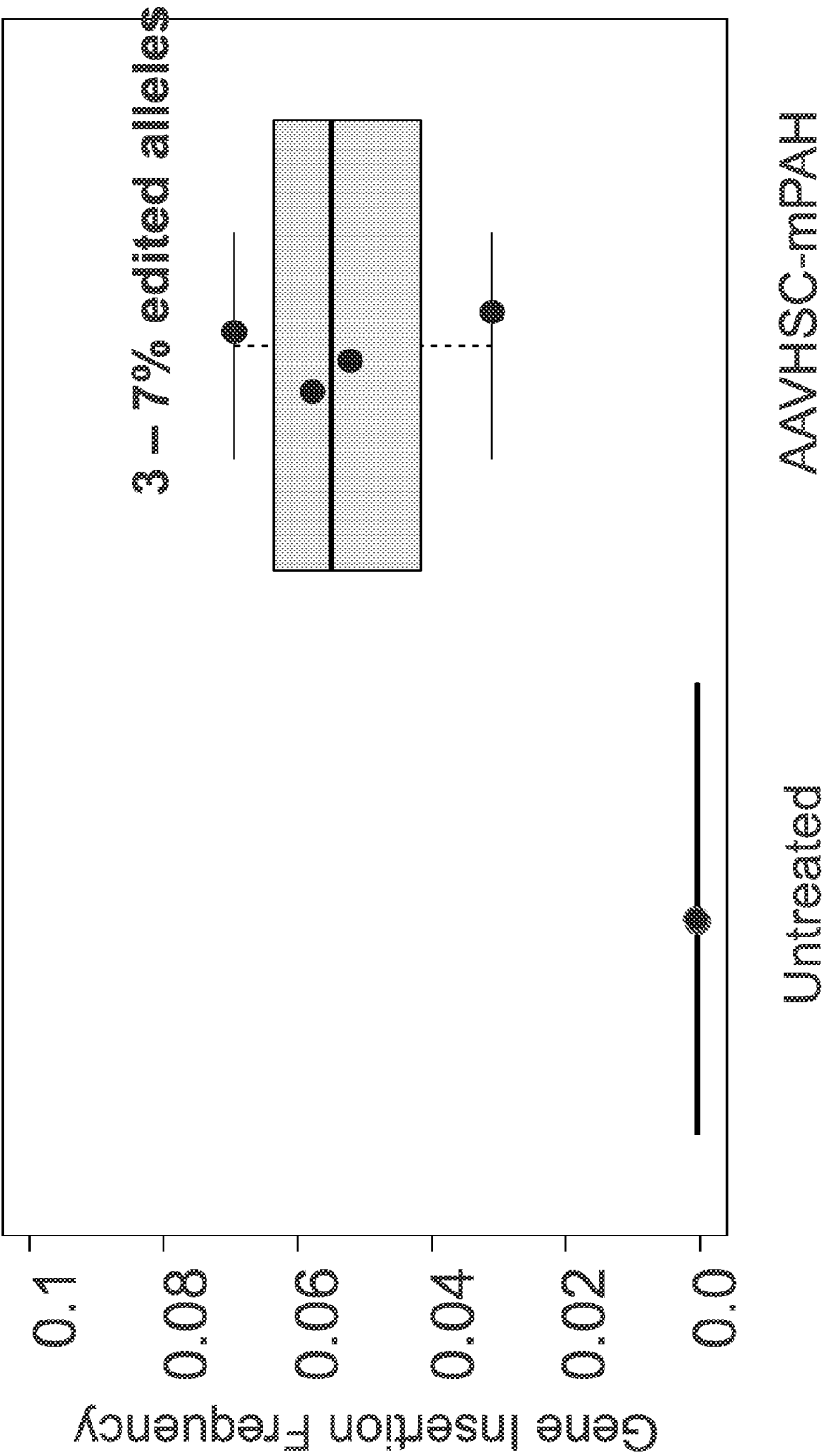


FIG. 5

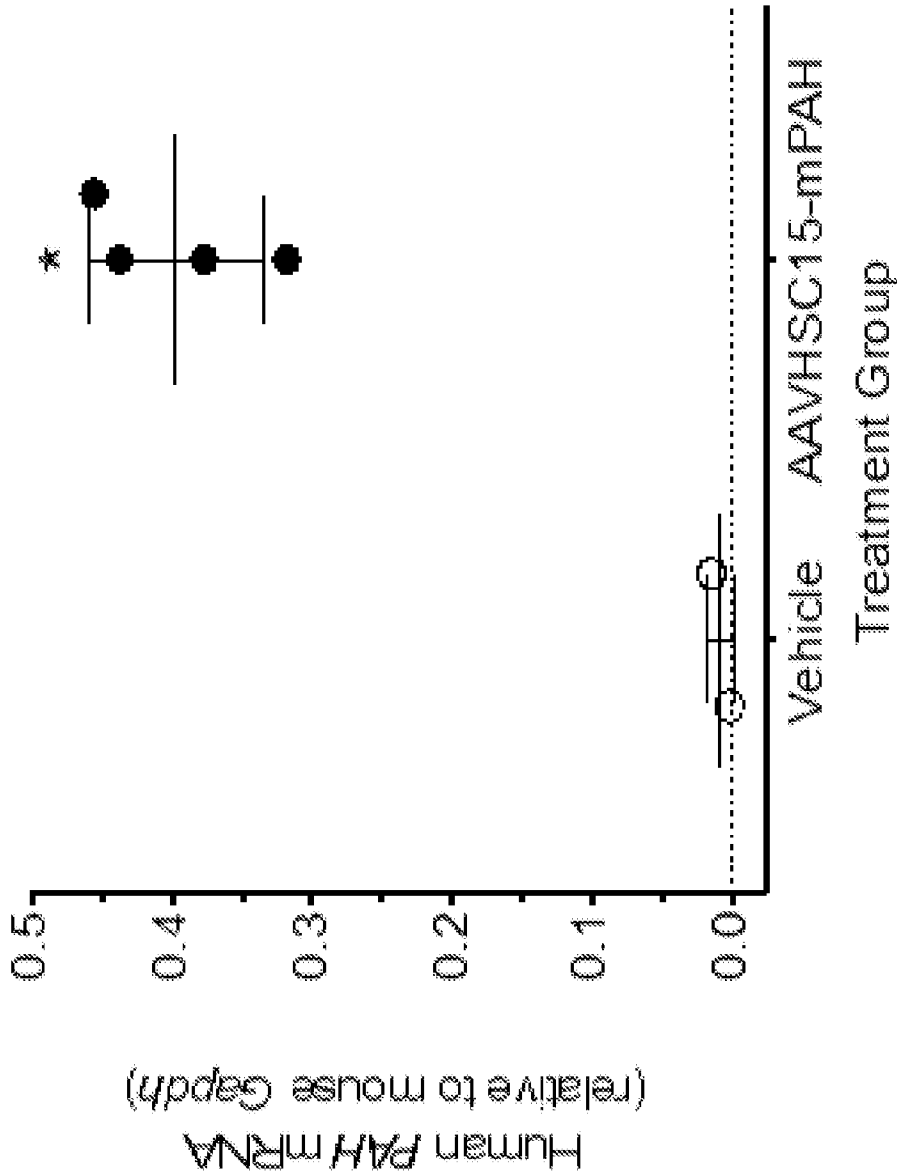


FIG. 6

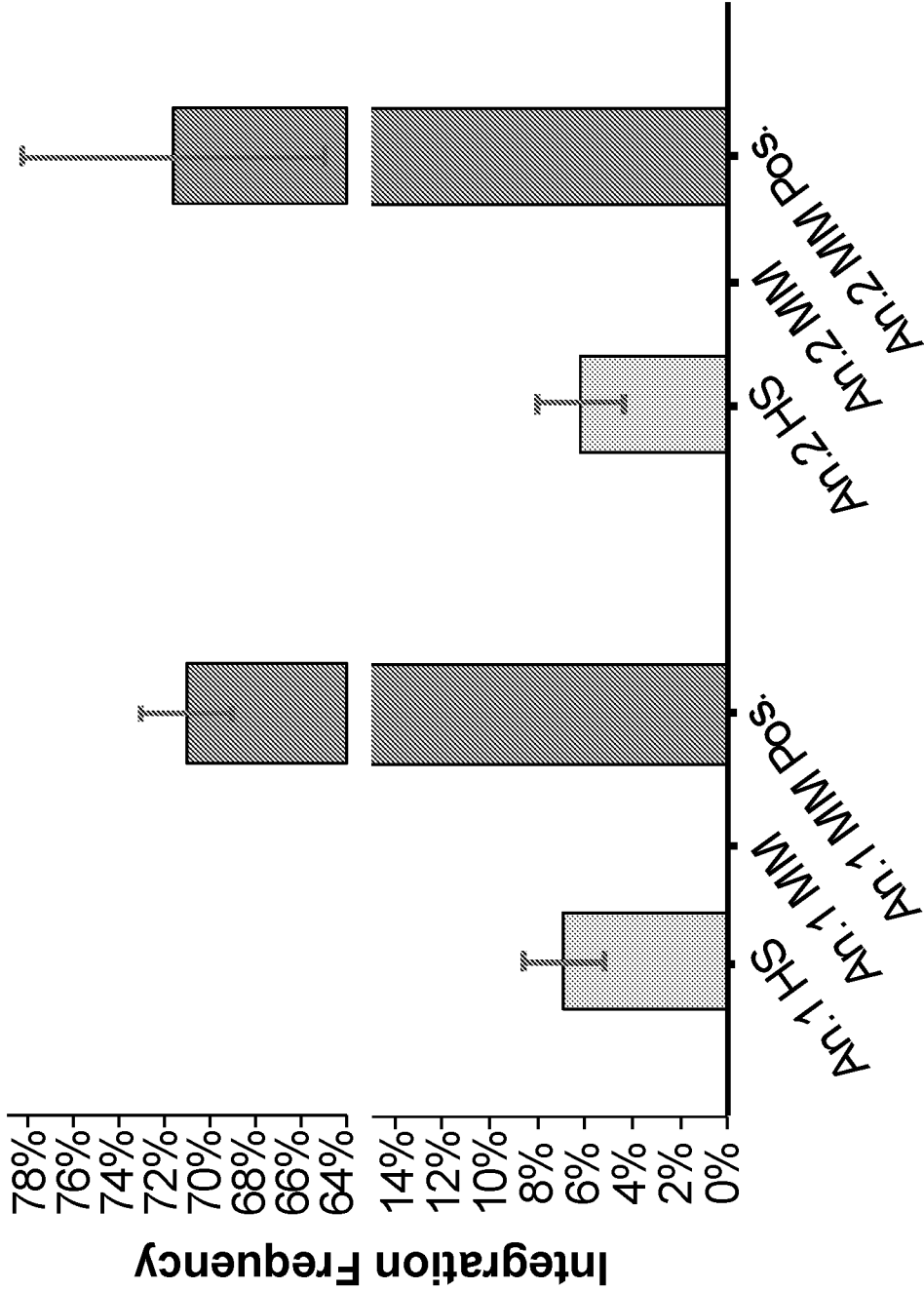


FIG. 7

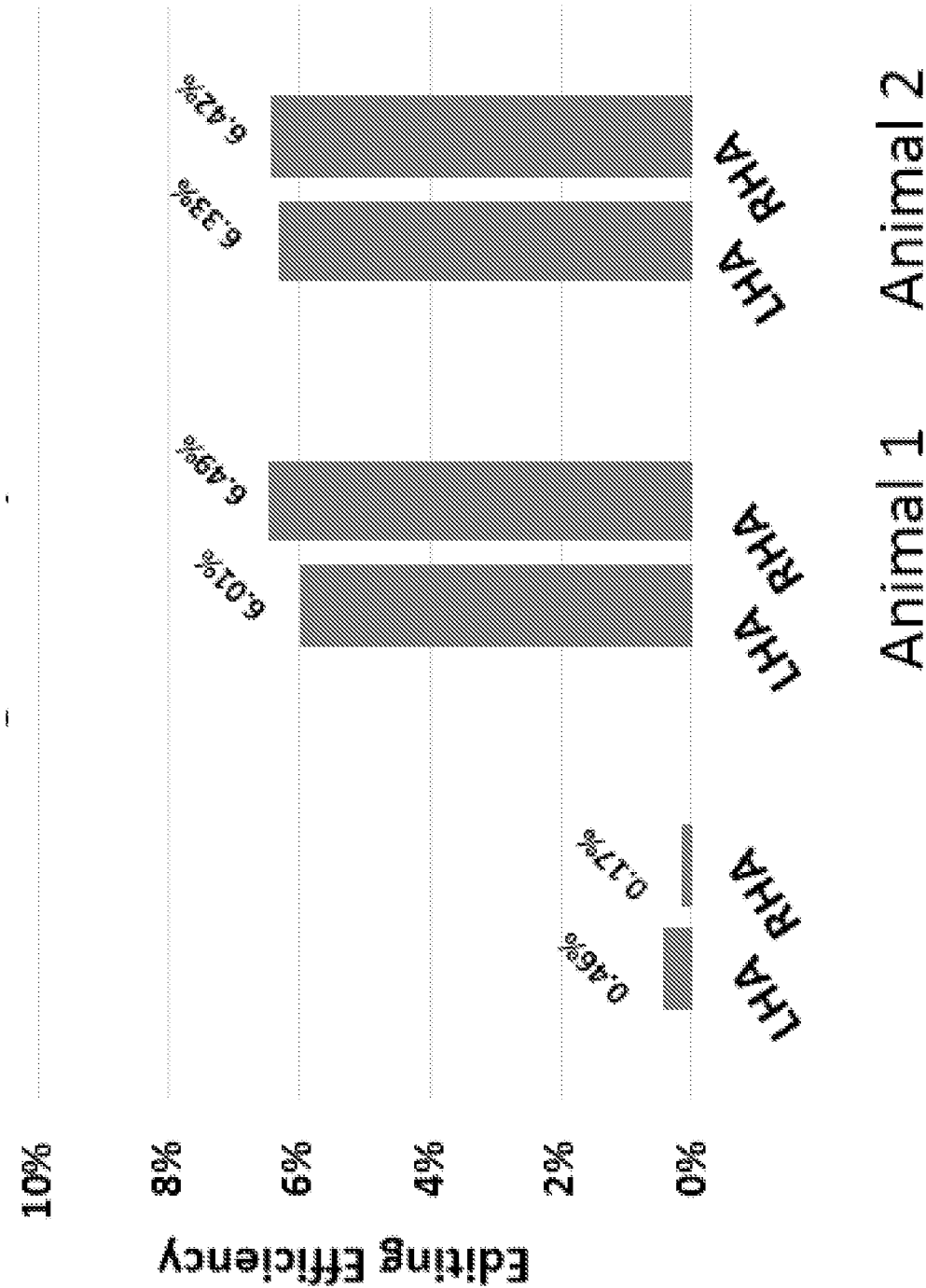


FIG. 8A

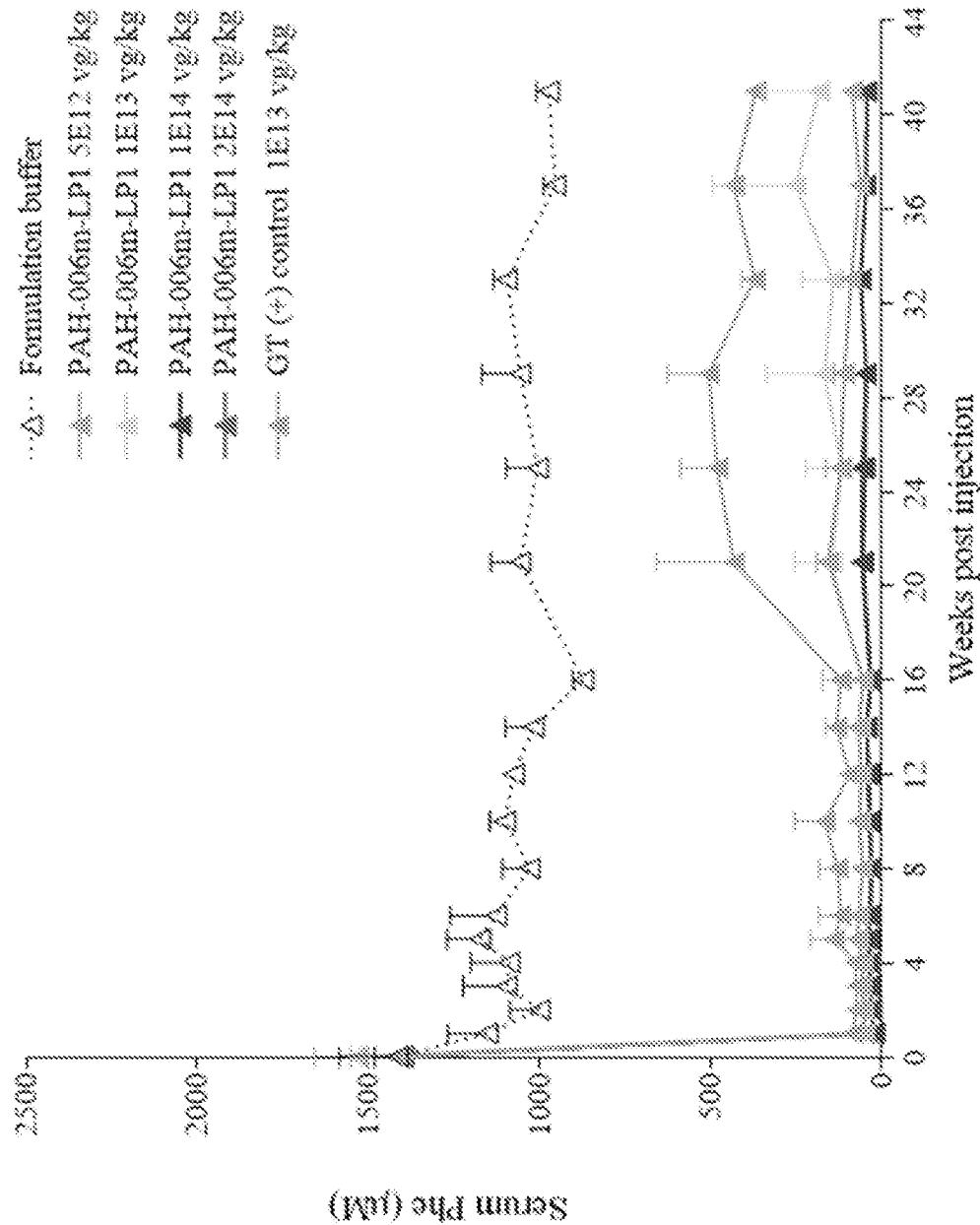


FIG. 8B

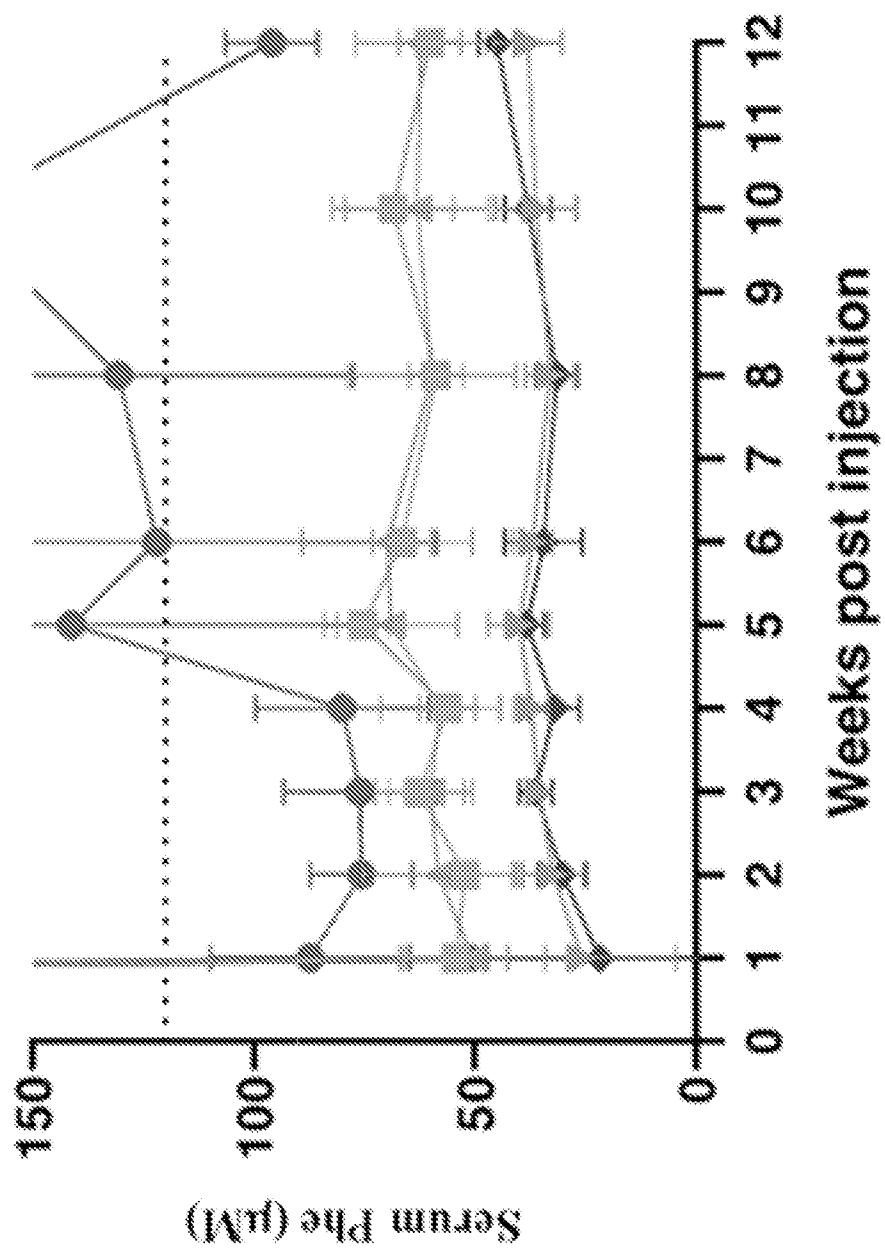


FIG. 8C

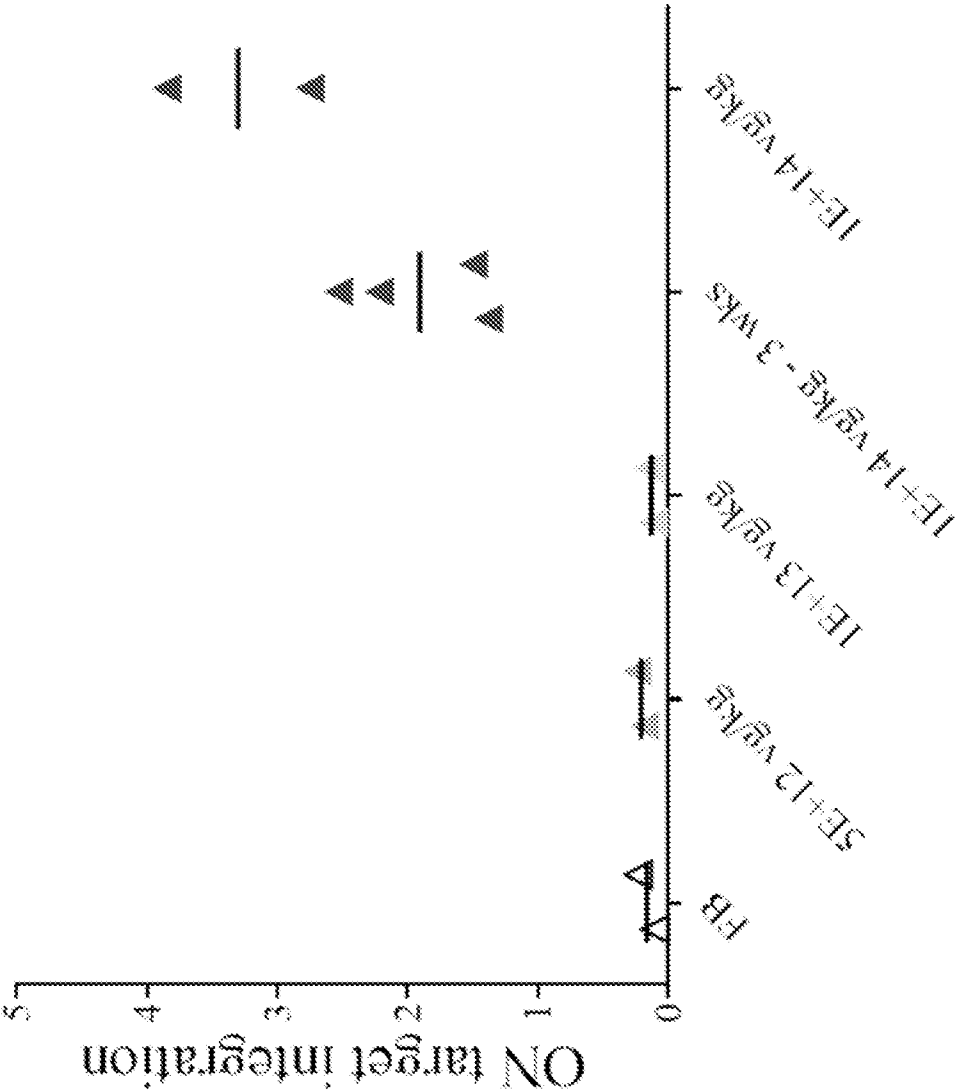


FIG. 9A

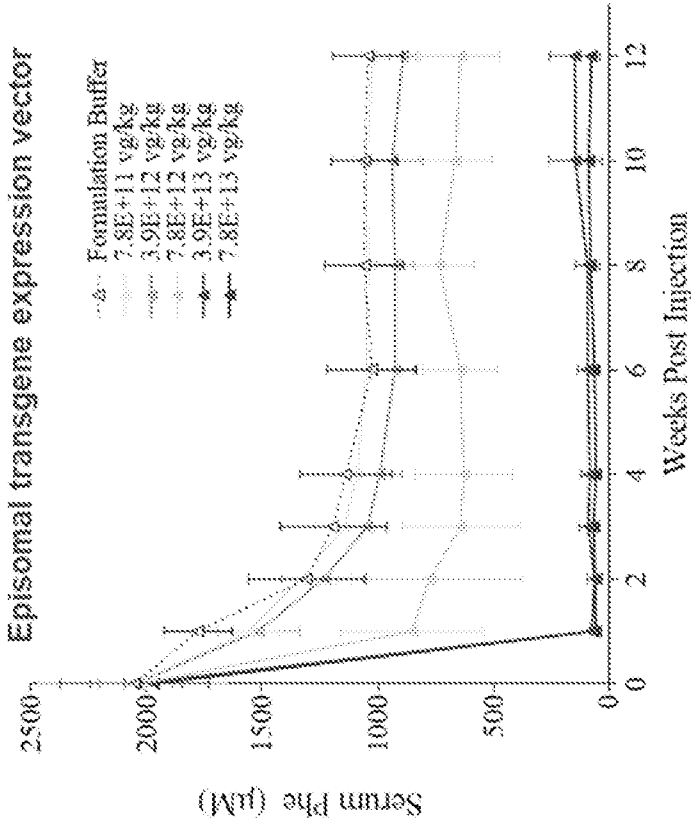


FIG. 9B

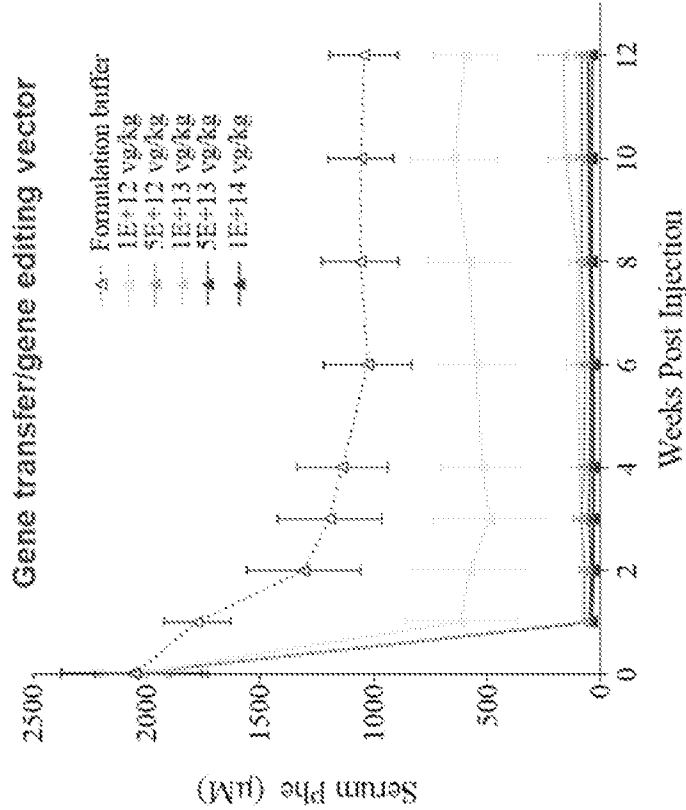


FIG. 9C

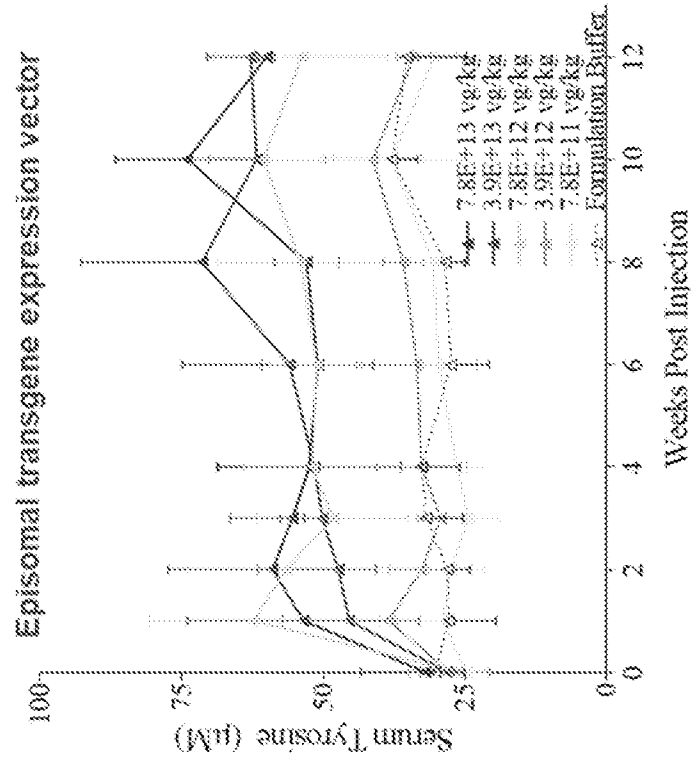


FIG. 9D

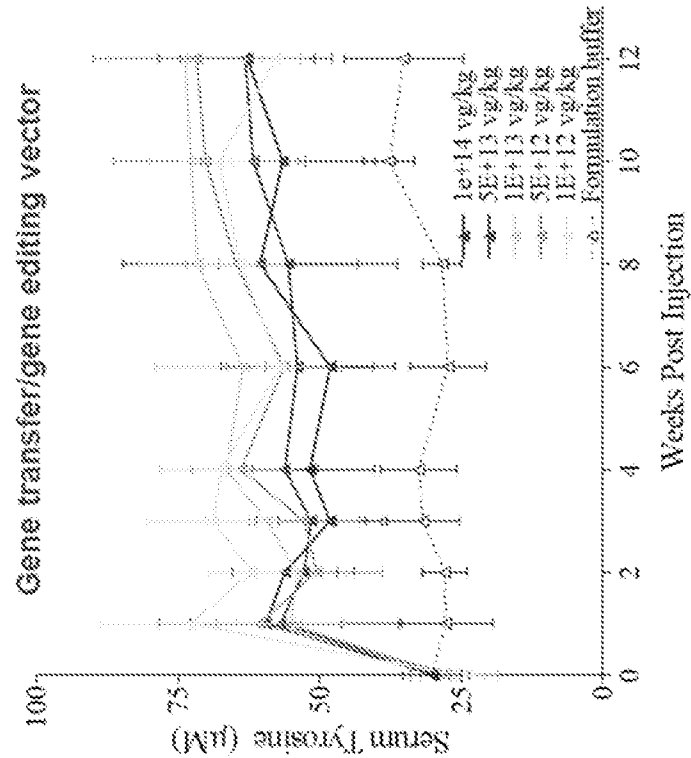


FIG. 9E

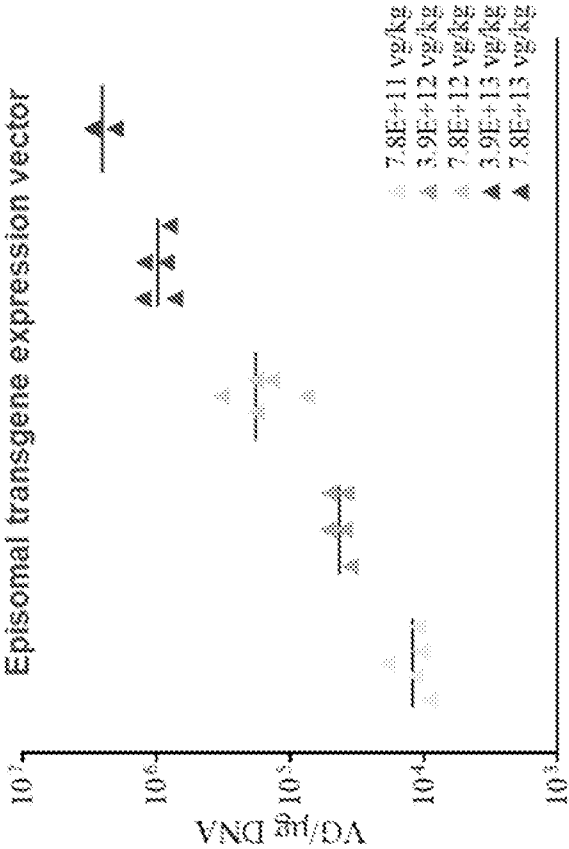


FIG. 9F

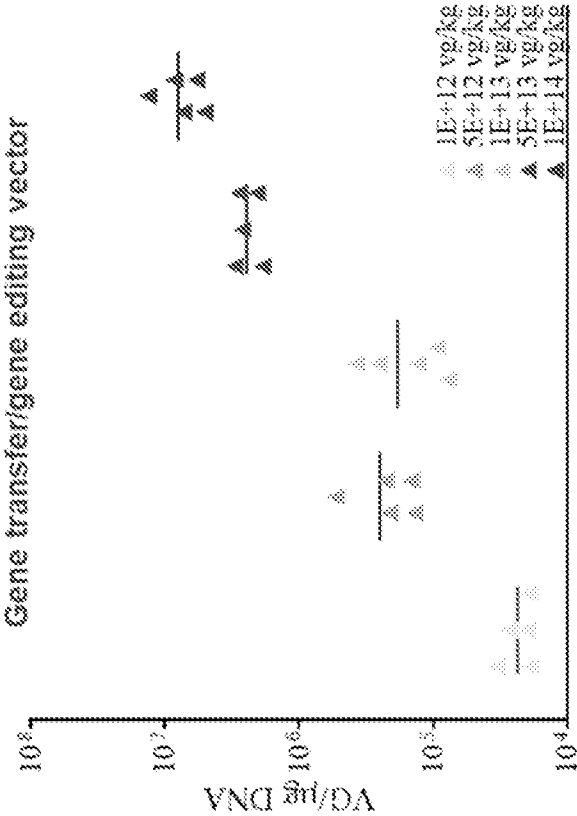


FIG. 9G

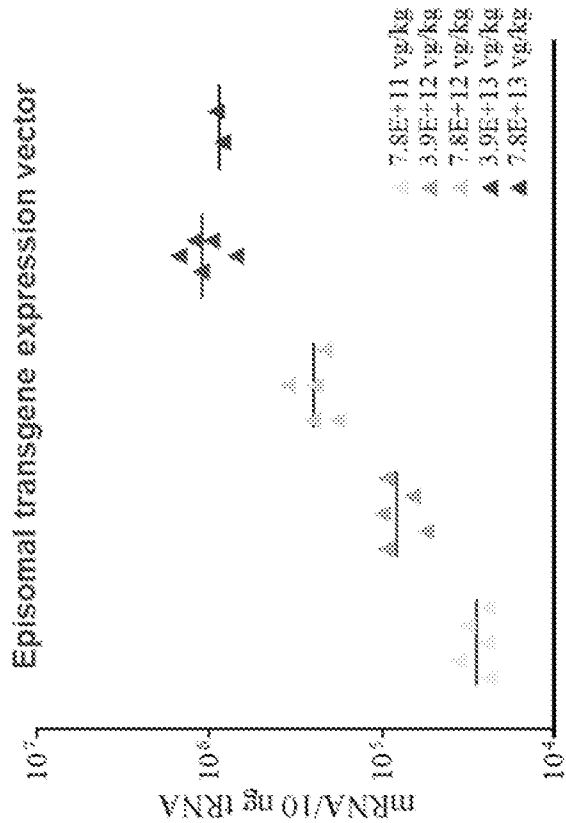


FIG. 9H

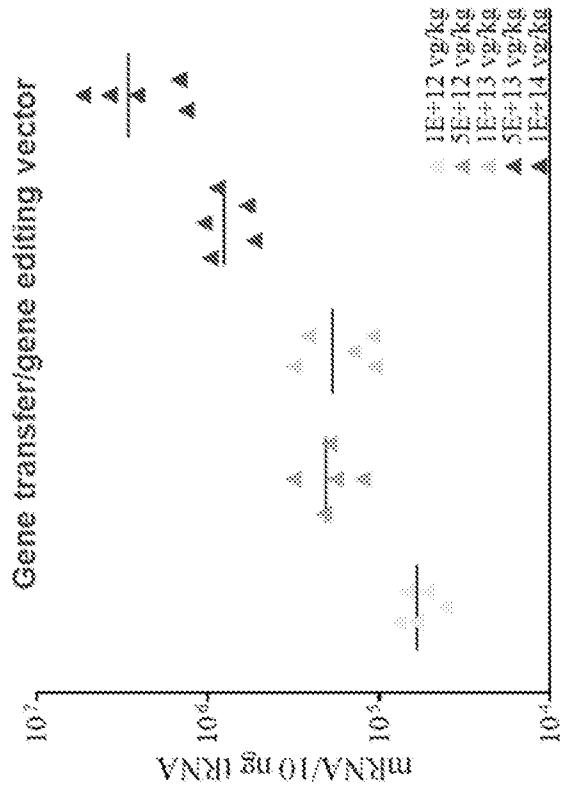


FIG. 9I

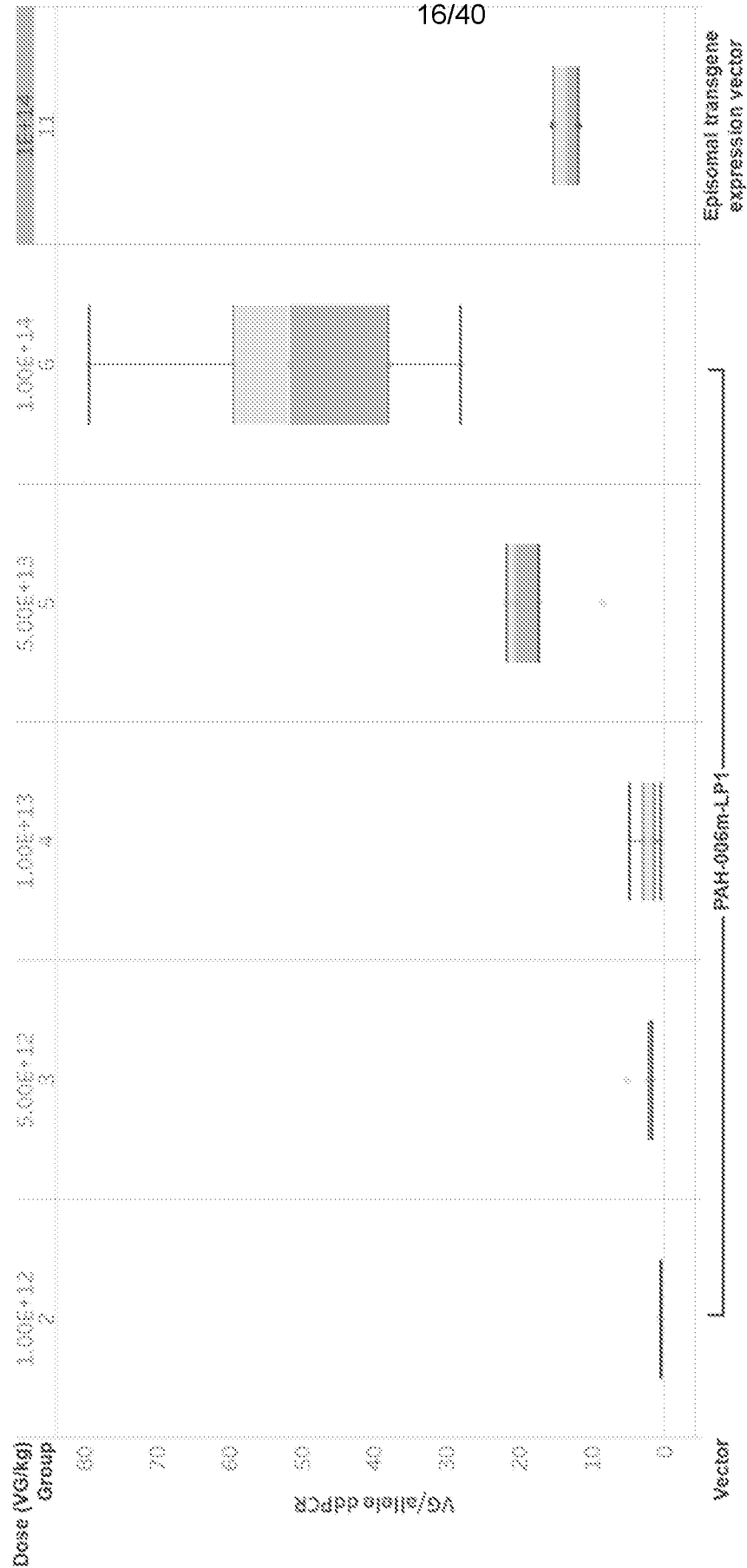
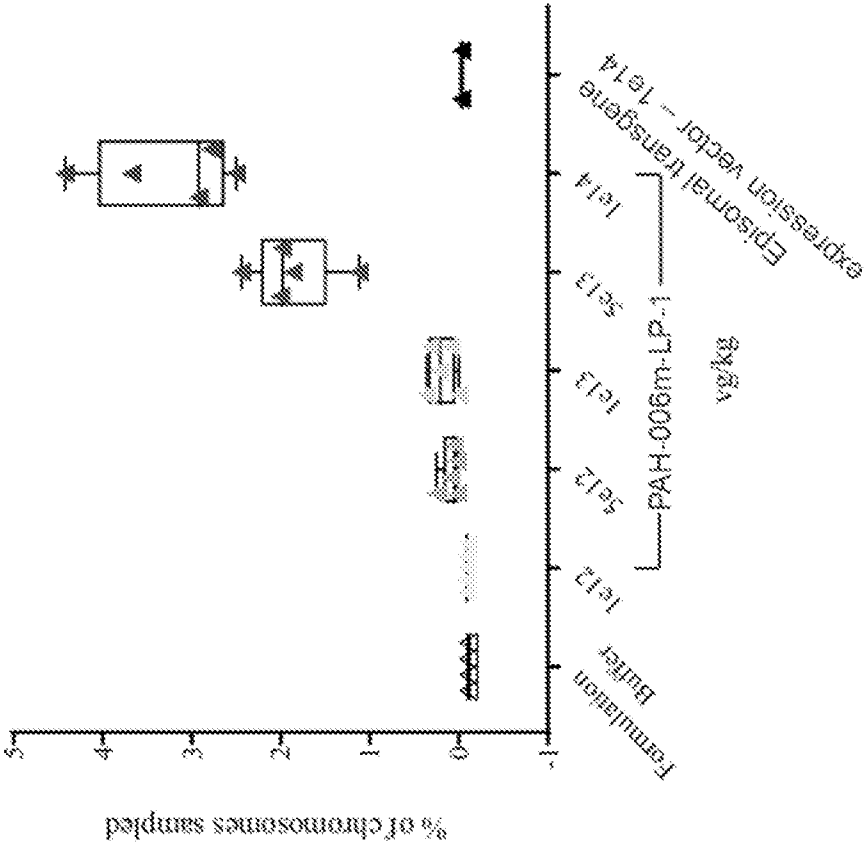


FIG. 9J

ON-Target Integration



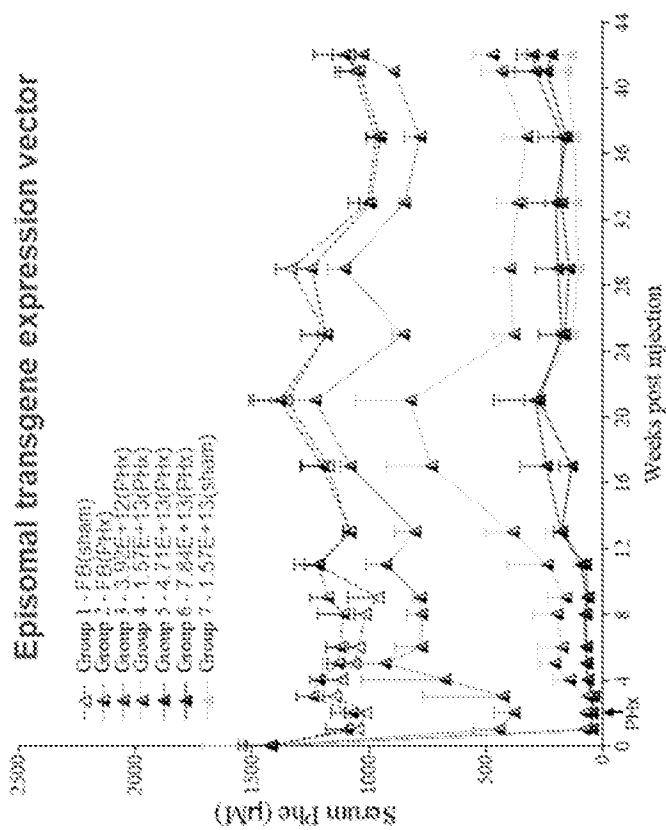
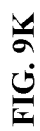


FIG. 10B

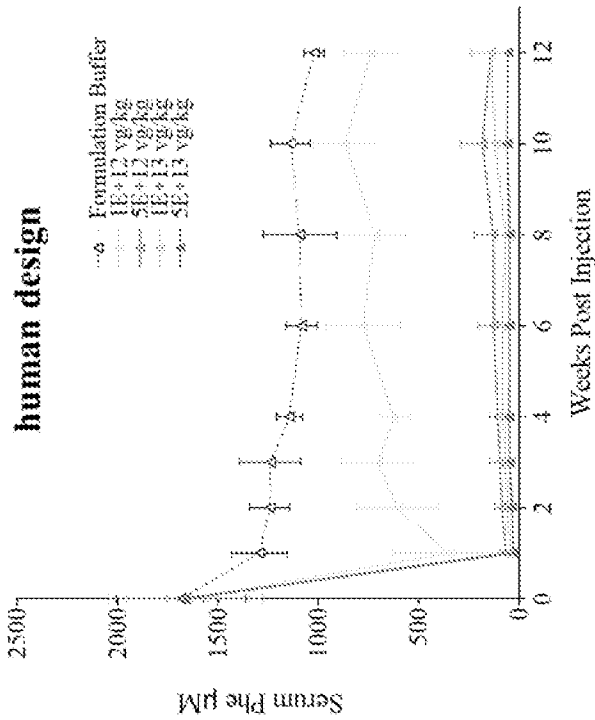


FIG. 10A

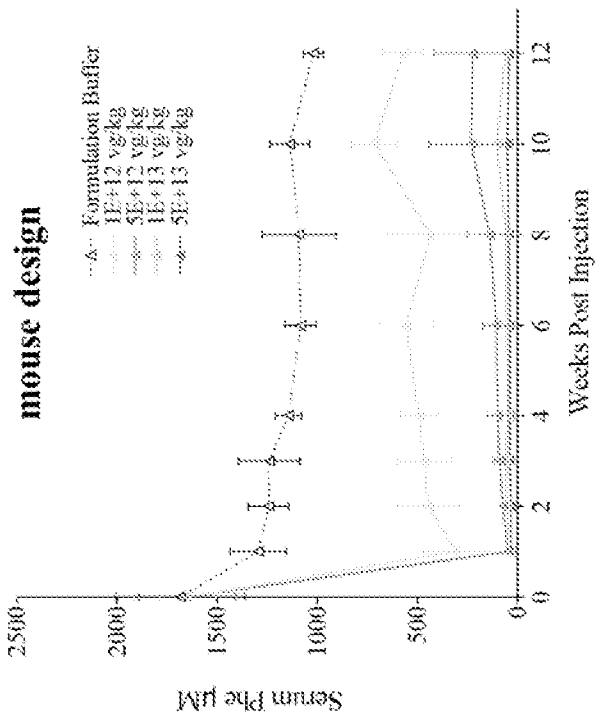


FIG. 10D

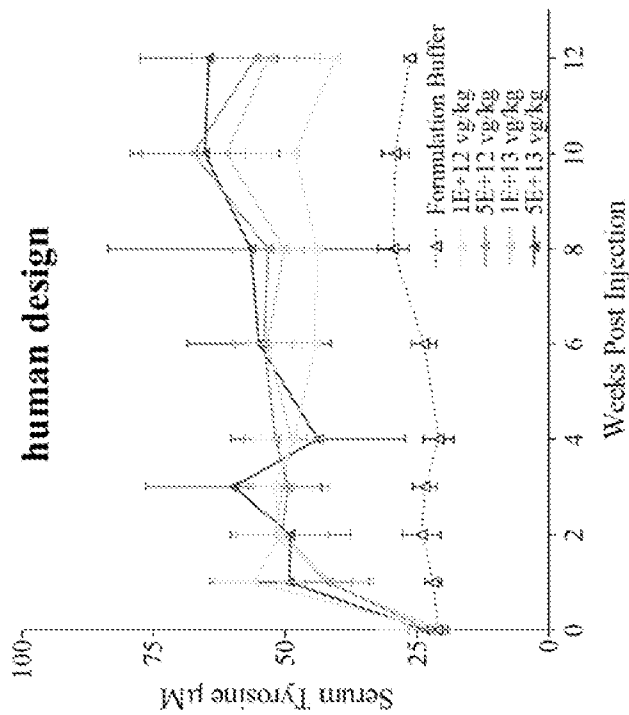


FIG. 10C

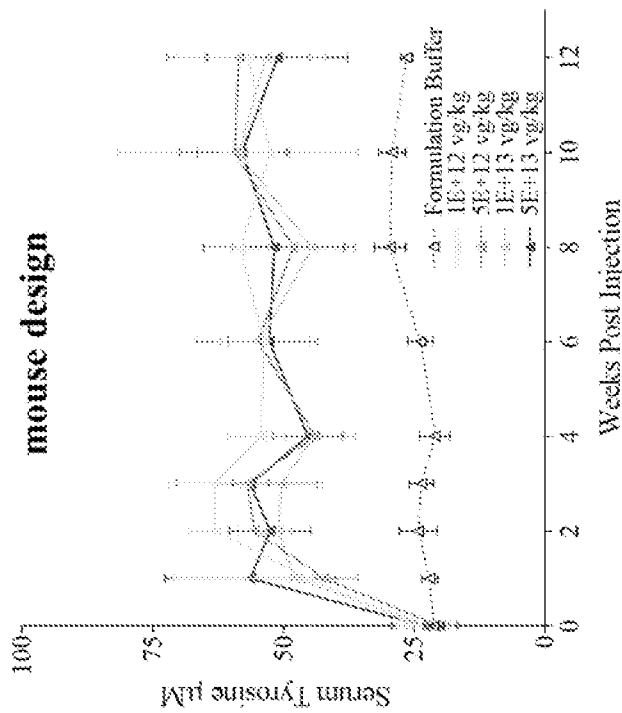


FIG. 10F

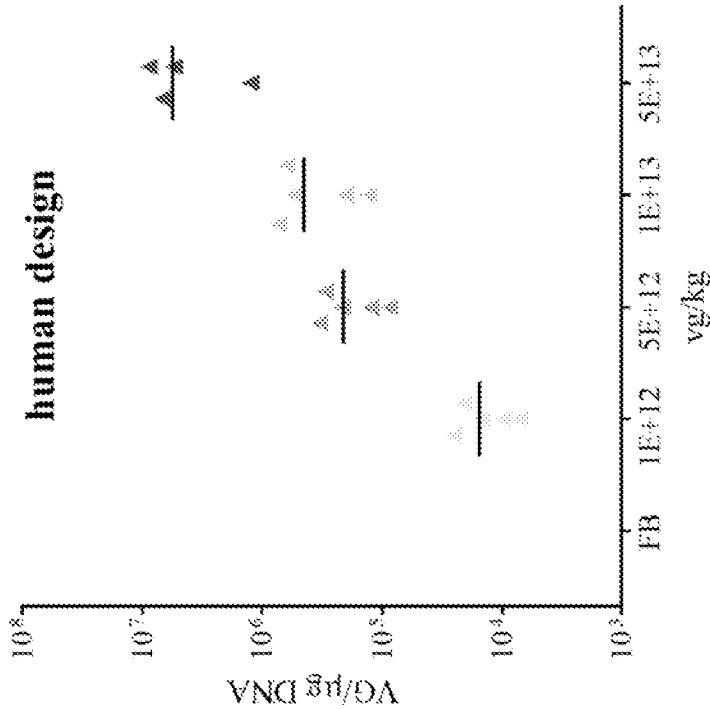


FIG. 10E

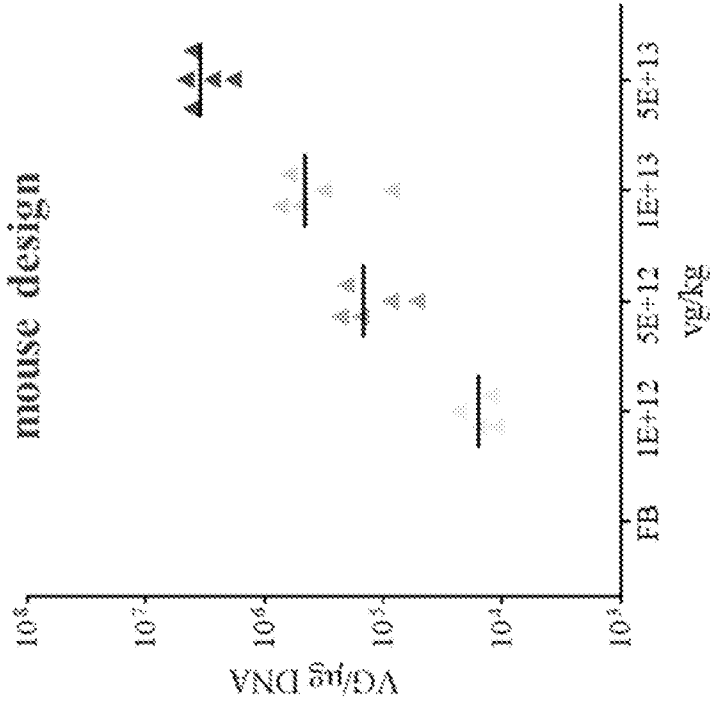


FIG. 10H

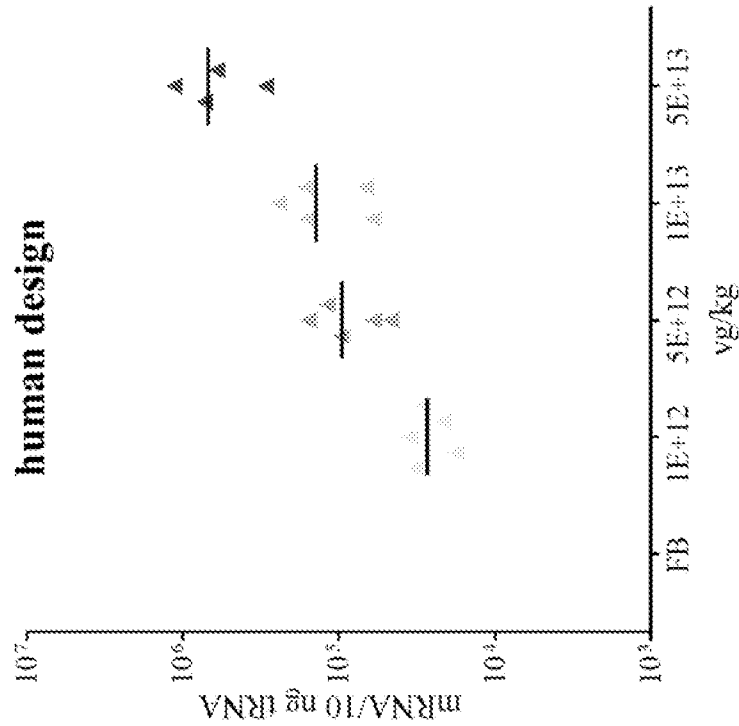


FIG. 10G

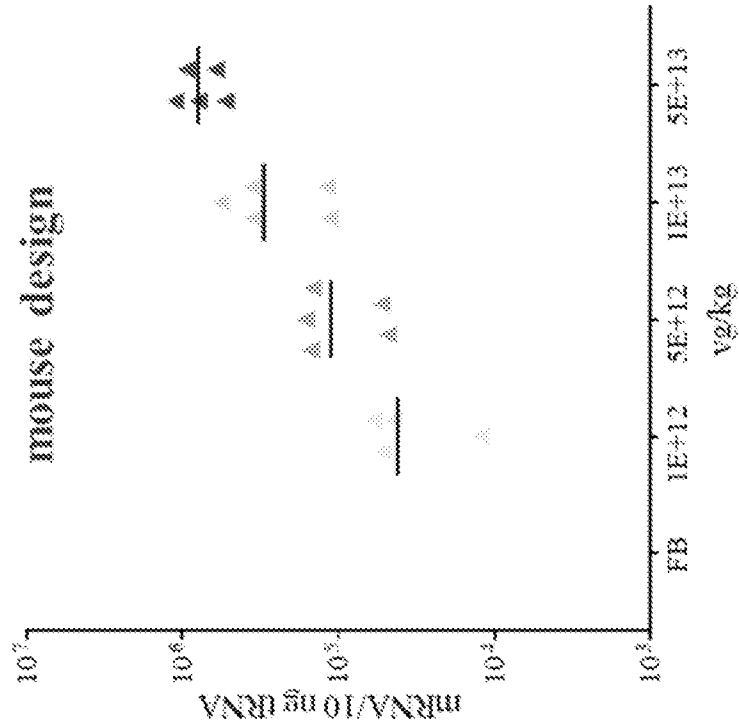


FIG. 10J

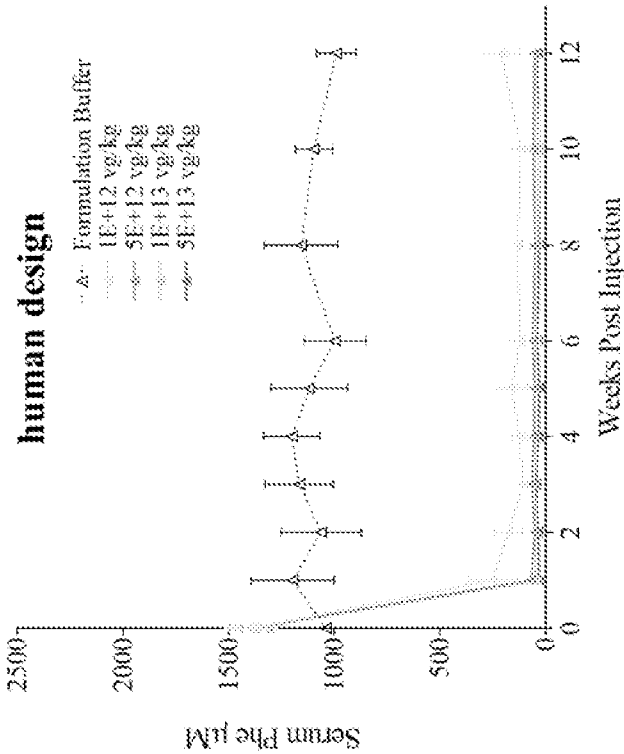


FIG. 10I

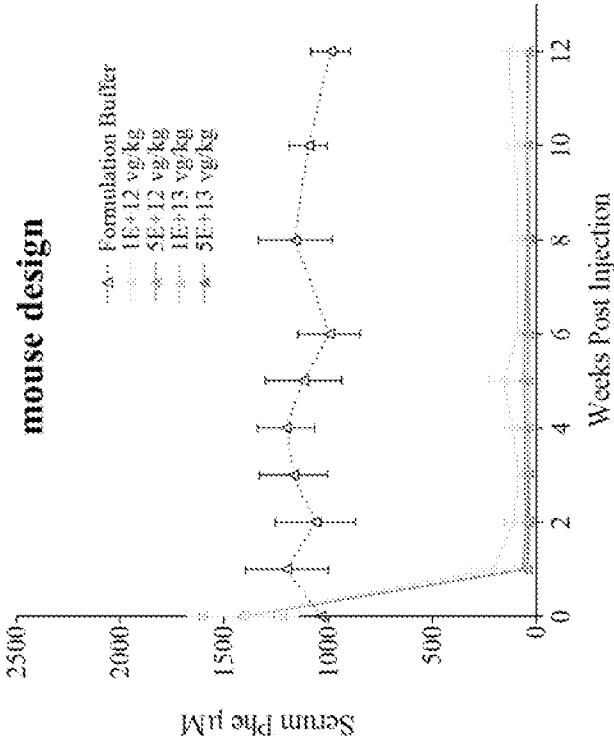


FIG. 10L

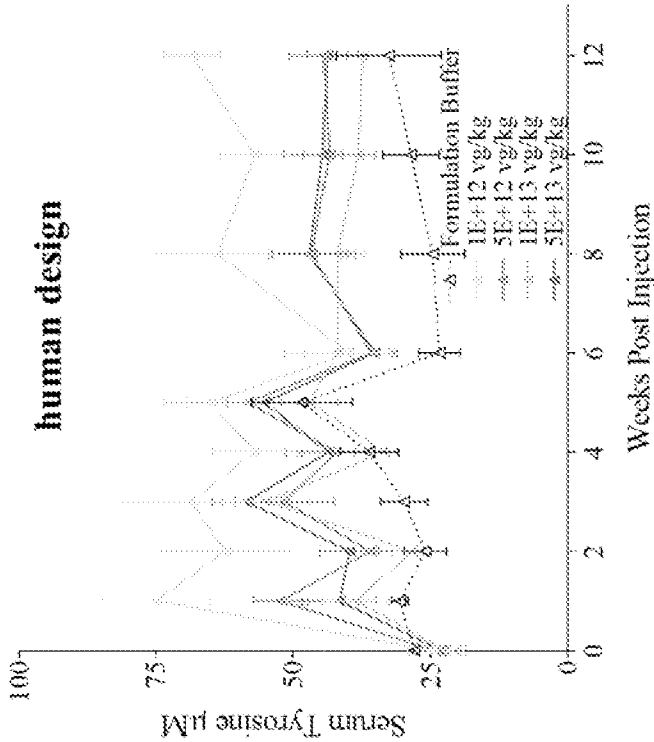


FIG. 10K

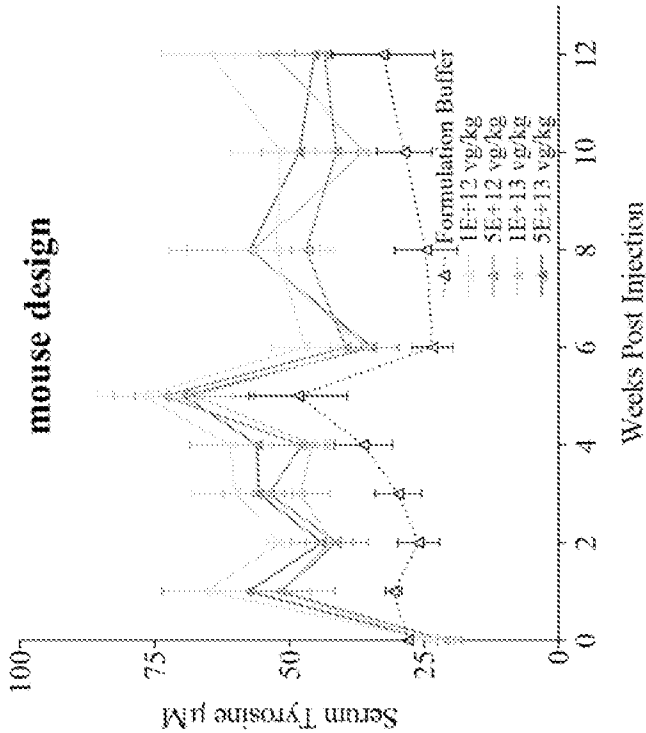


FIG. 10N

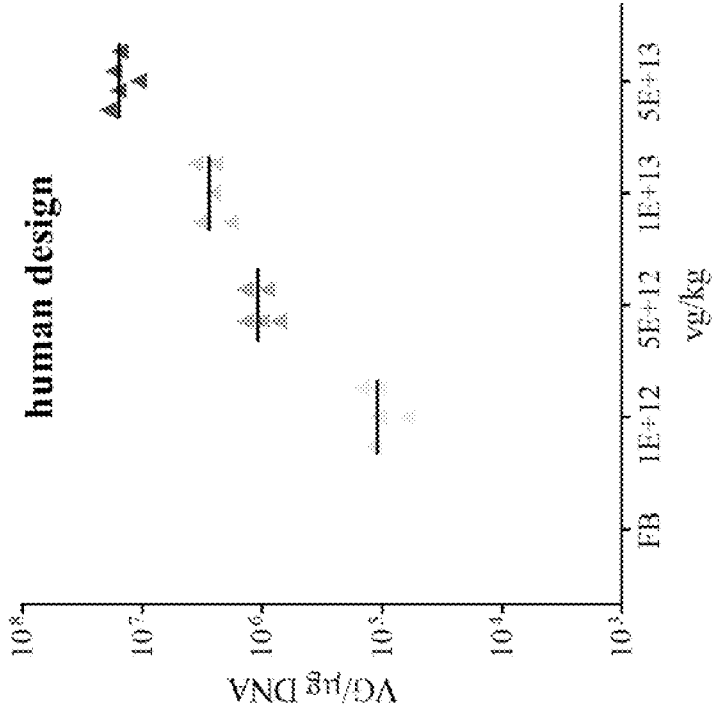


FIG. 10M

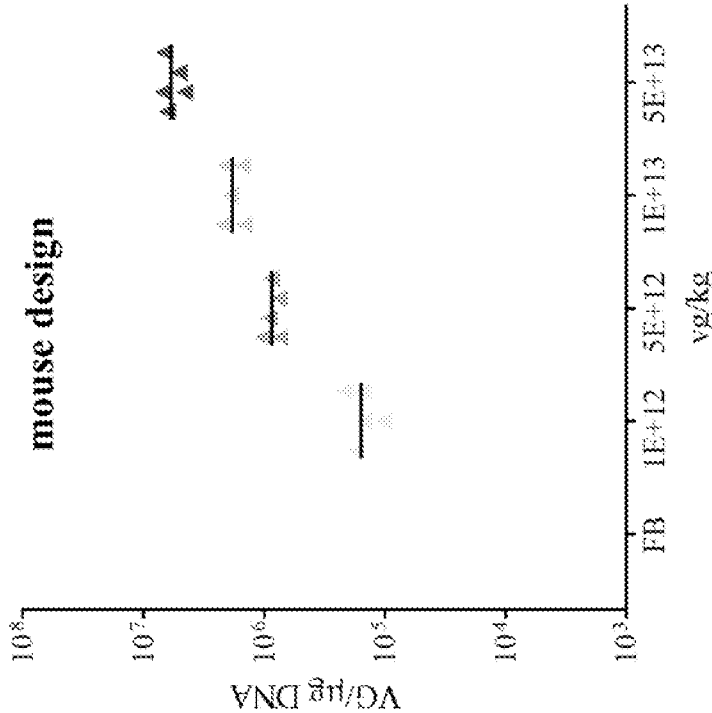


FIG. 10P

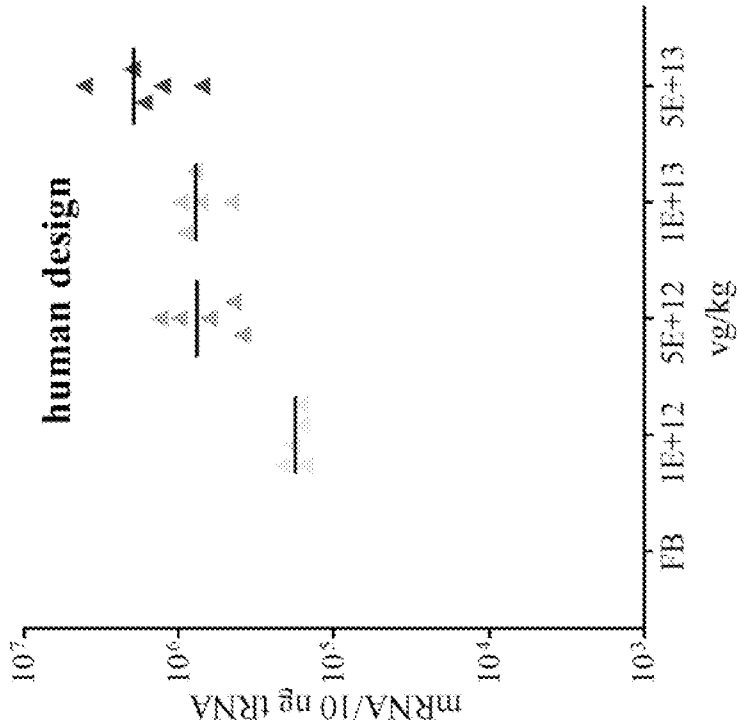


FIG. 10O

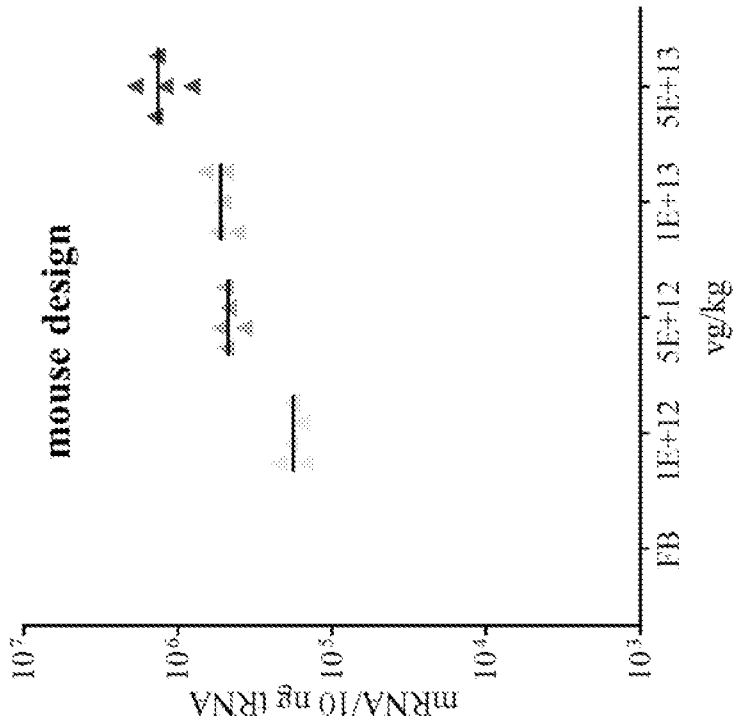


FIG. 11A

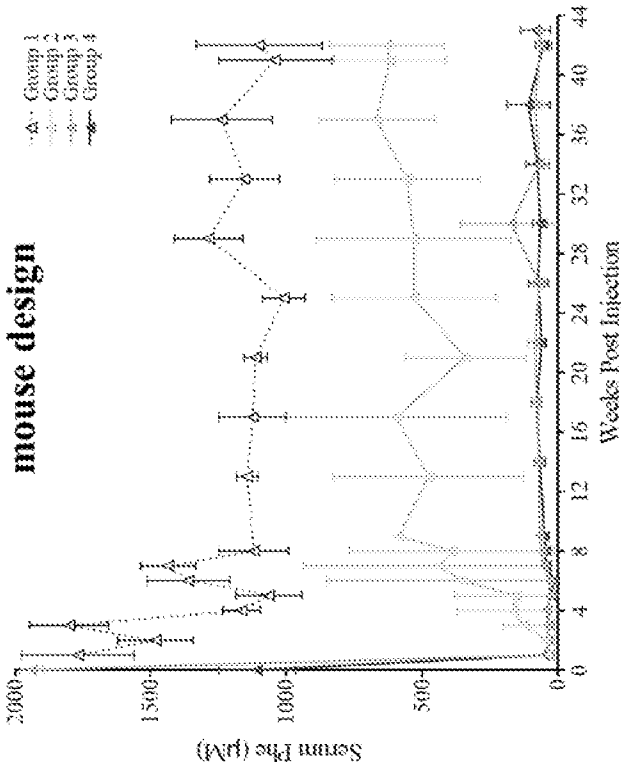


FIG. 11B

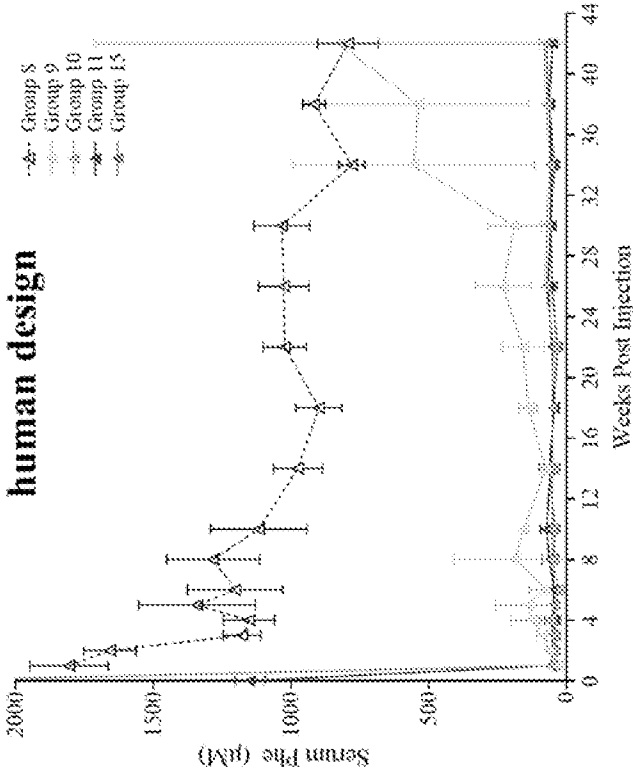


FIG. 11C

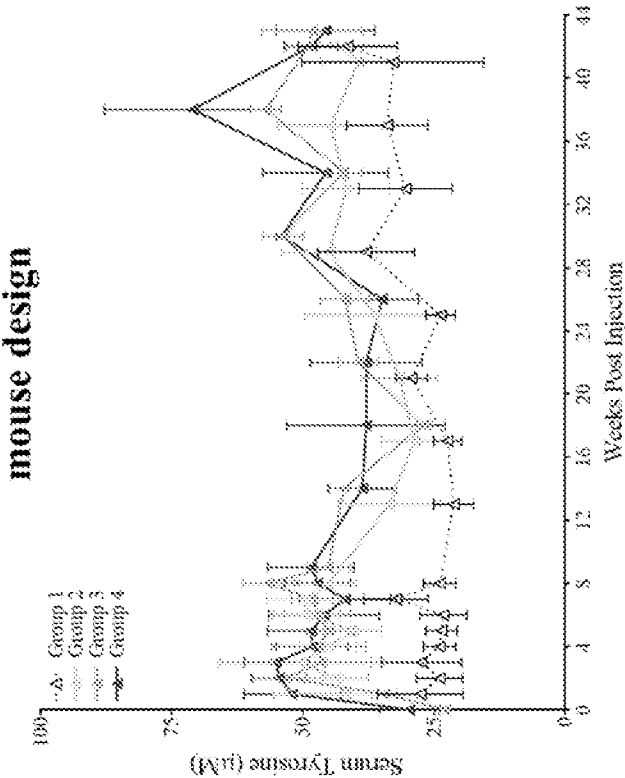


FIG. 11D

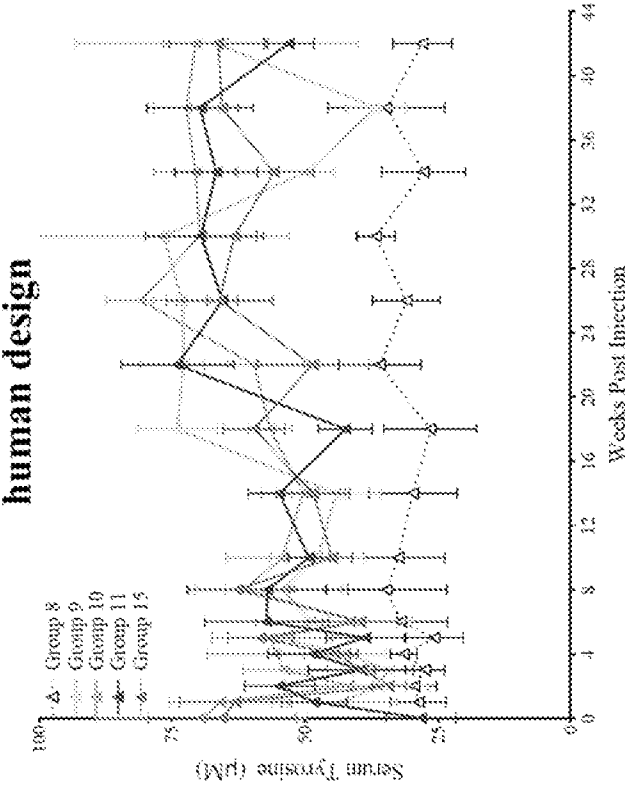


FIG. 11F

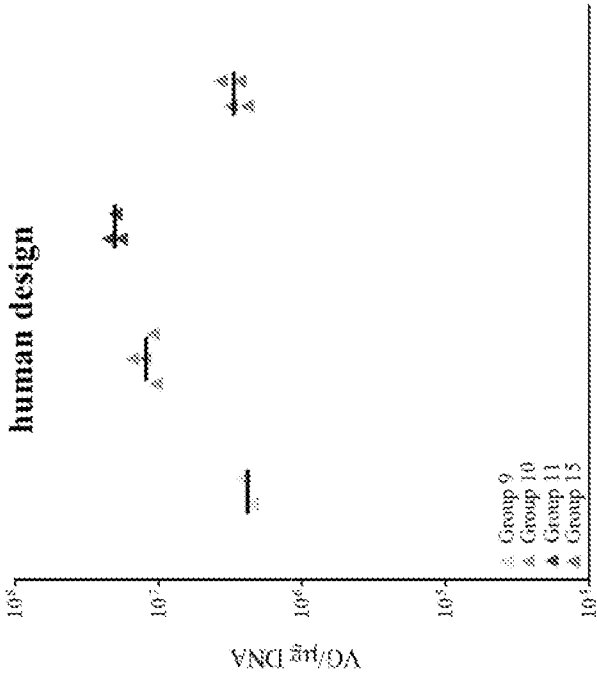


FIG. 11E

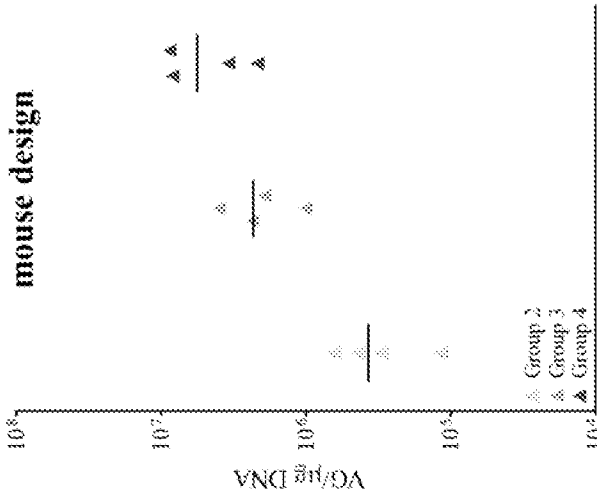


FIG. 11H

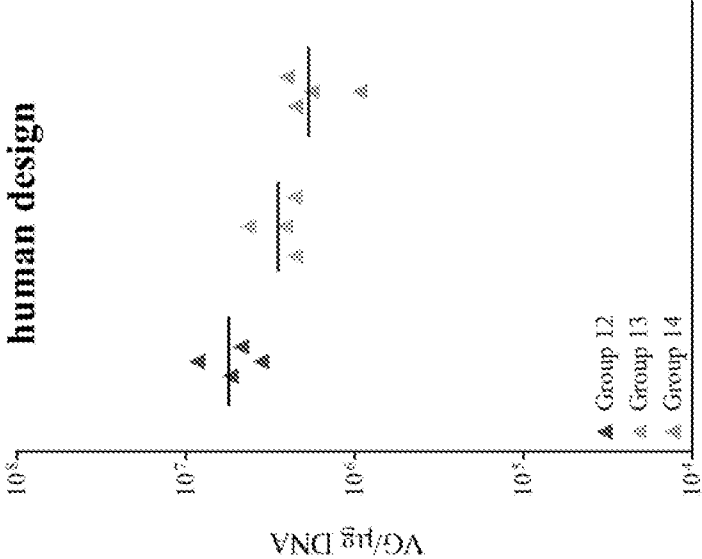


FIG. 11G

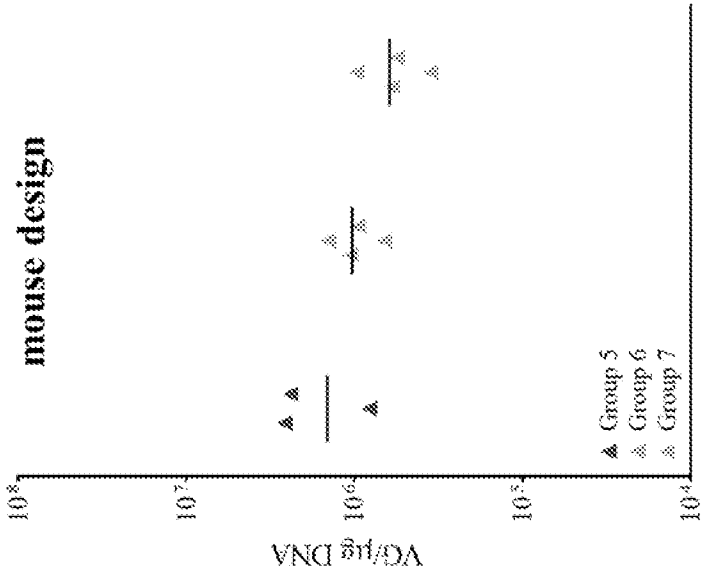


FIG. 11J

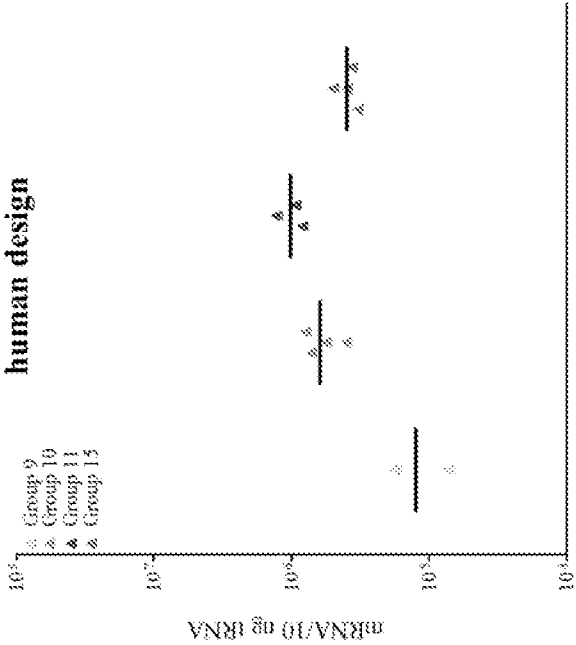


FIG. 11I

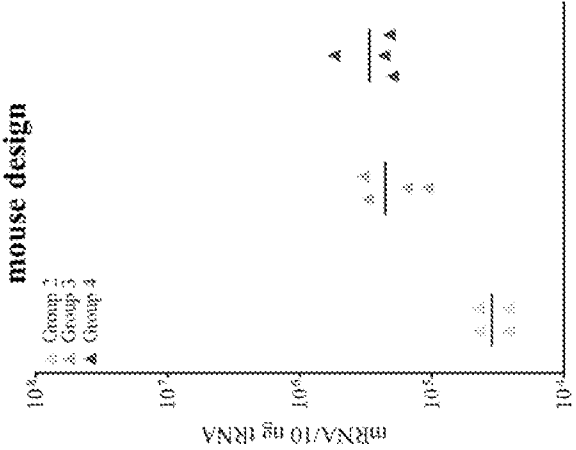


FIG. 11K

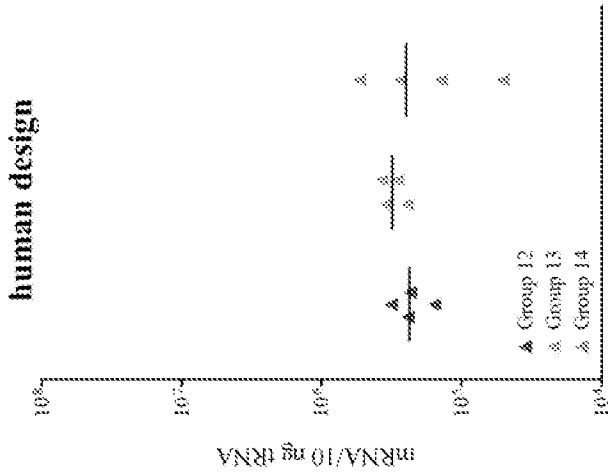


FIG. 12A

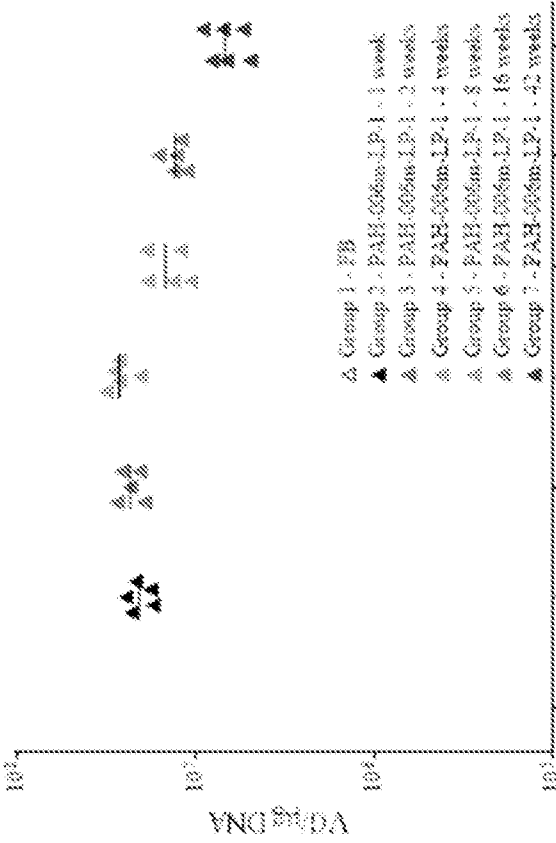


FIG. 12B

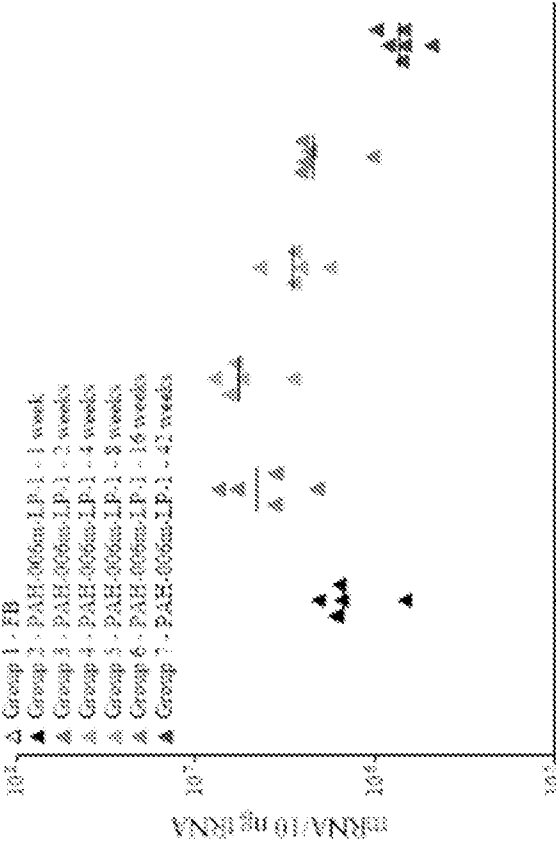


FIG. 12C

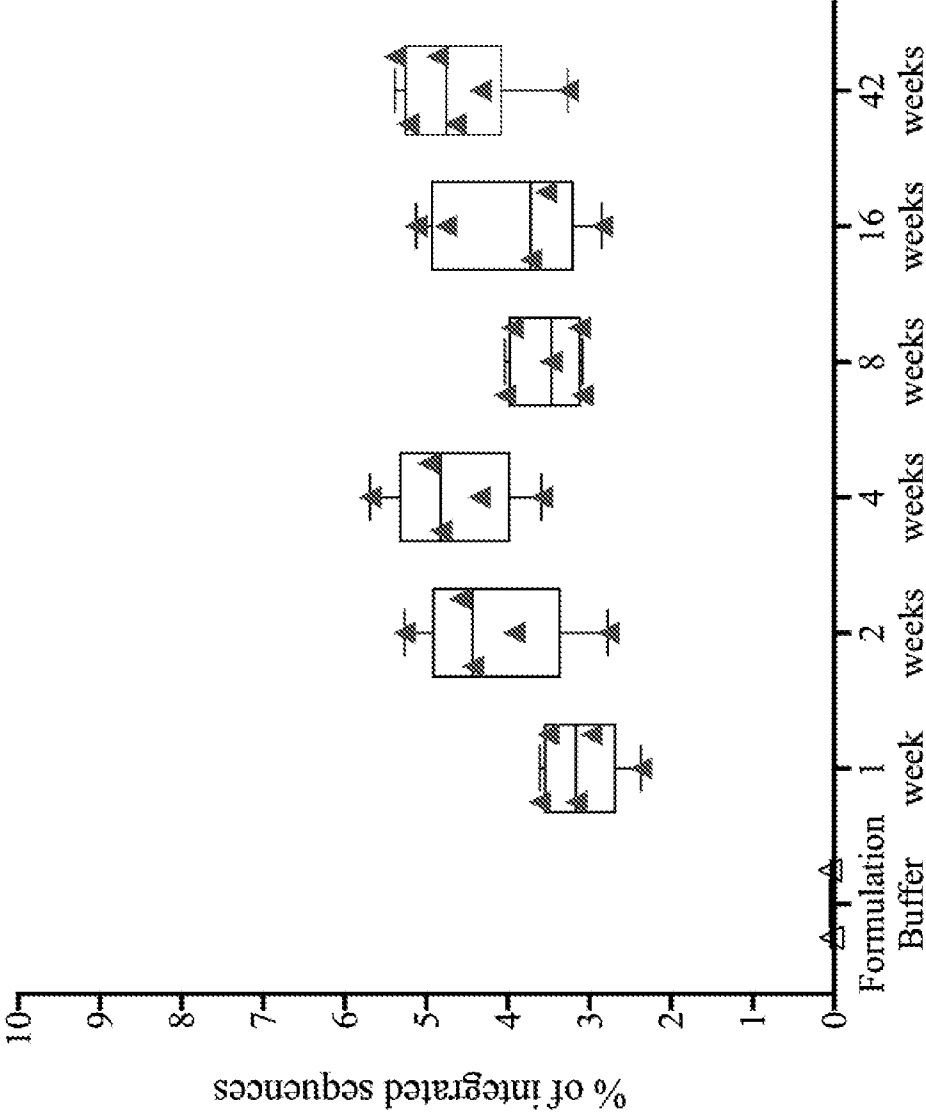


FIG. 12D

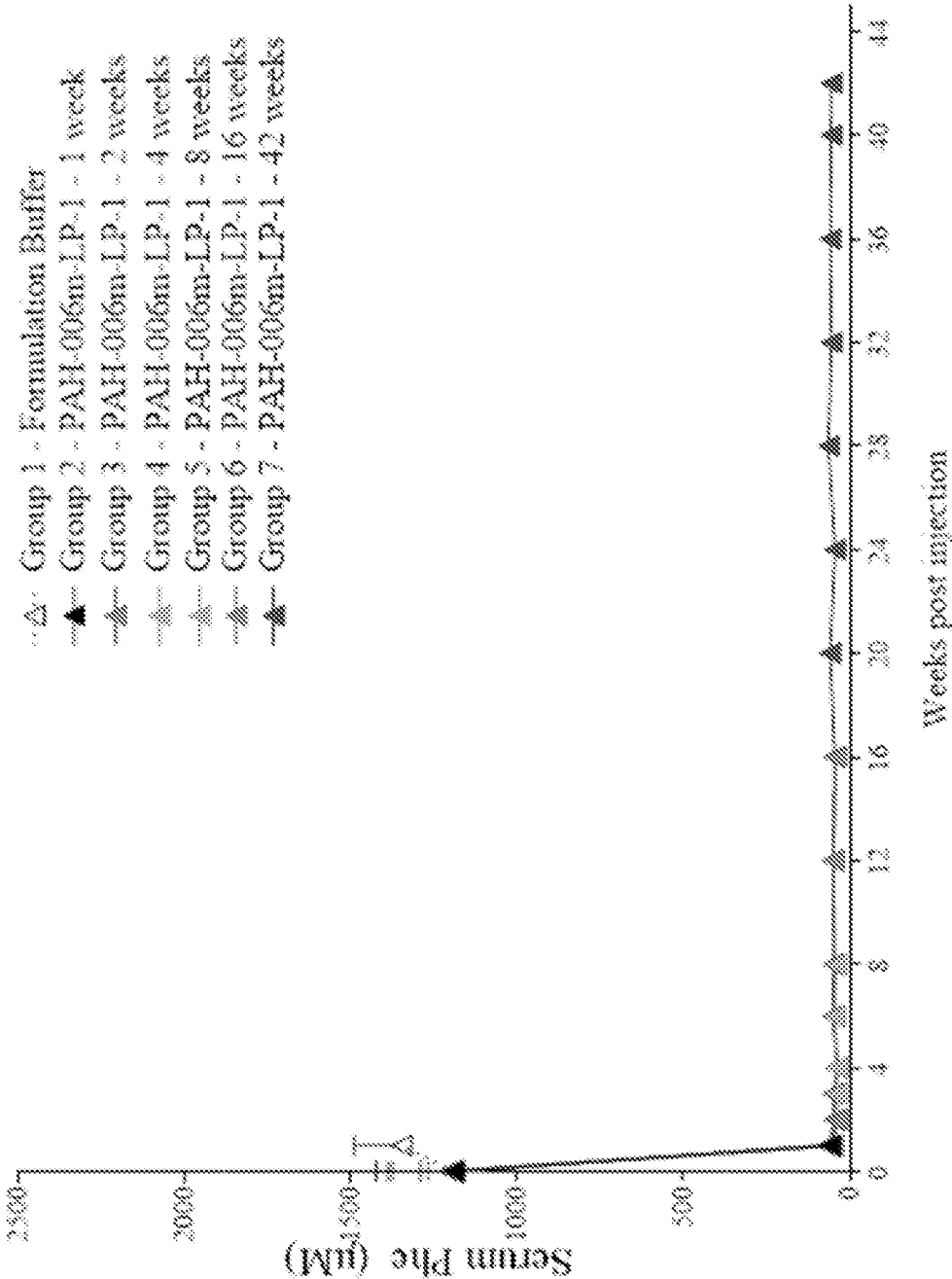


FIG. 12E

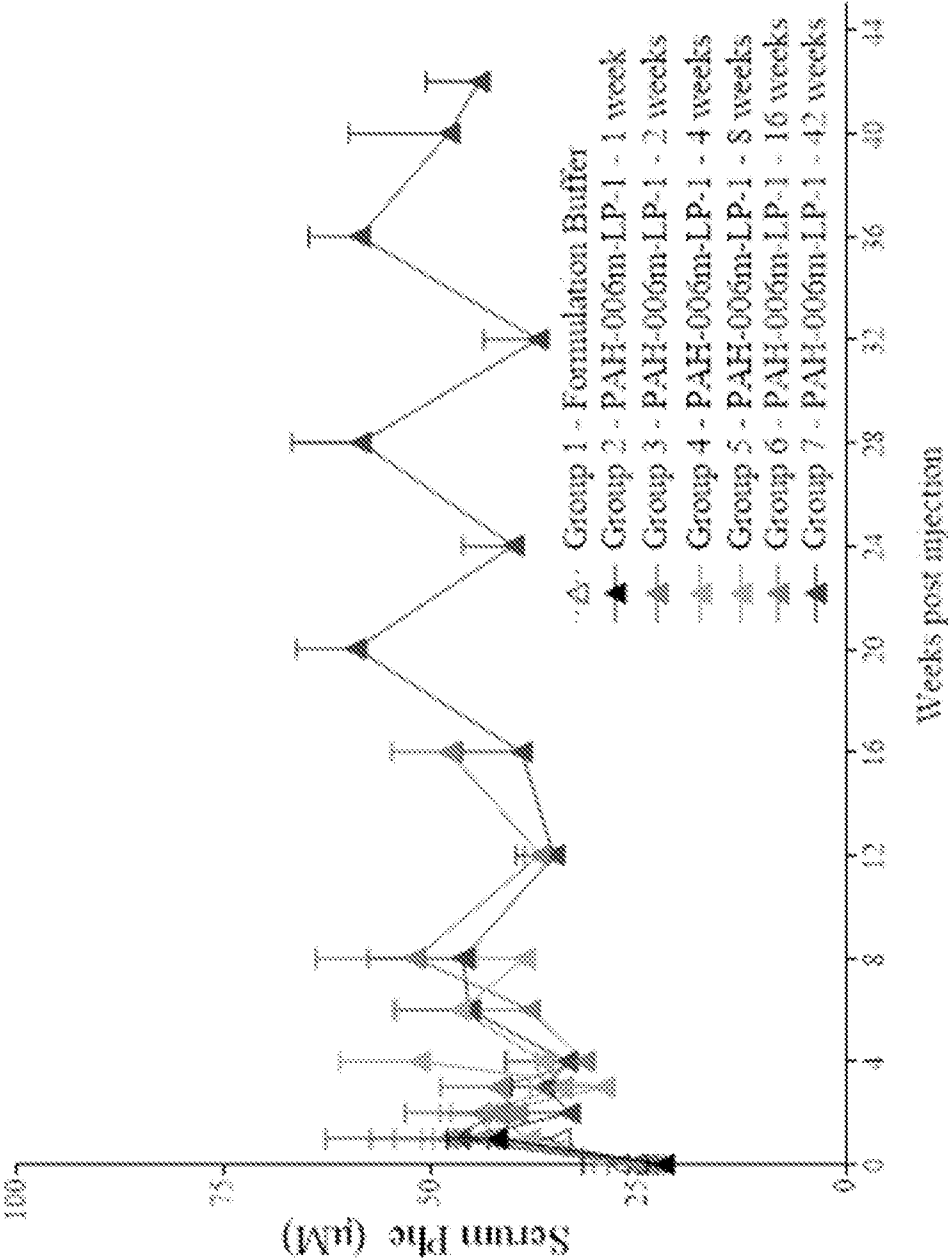


FIG. 13B

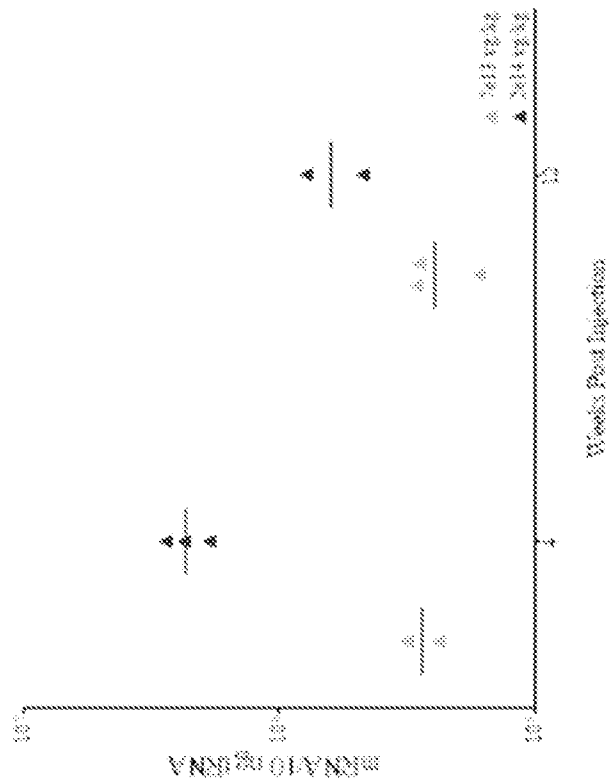


FIG. 13A

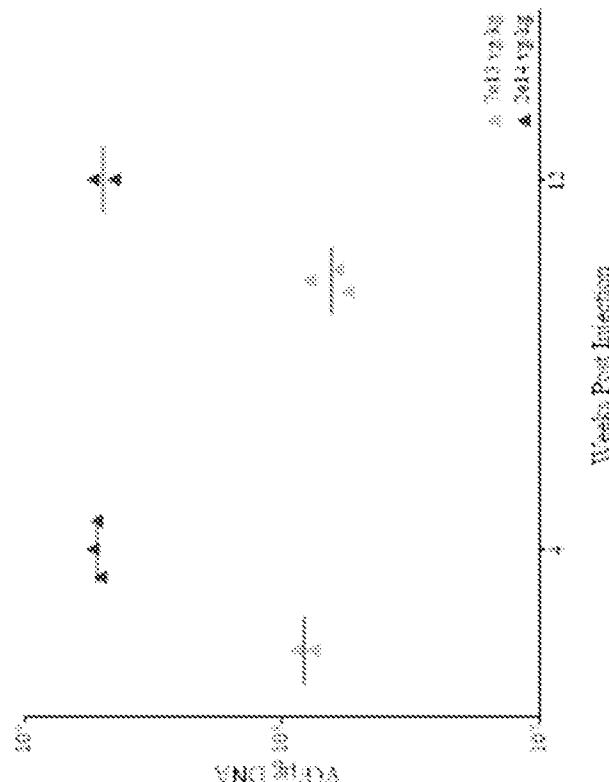


FIG. 13C

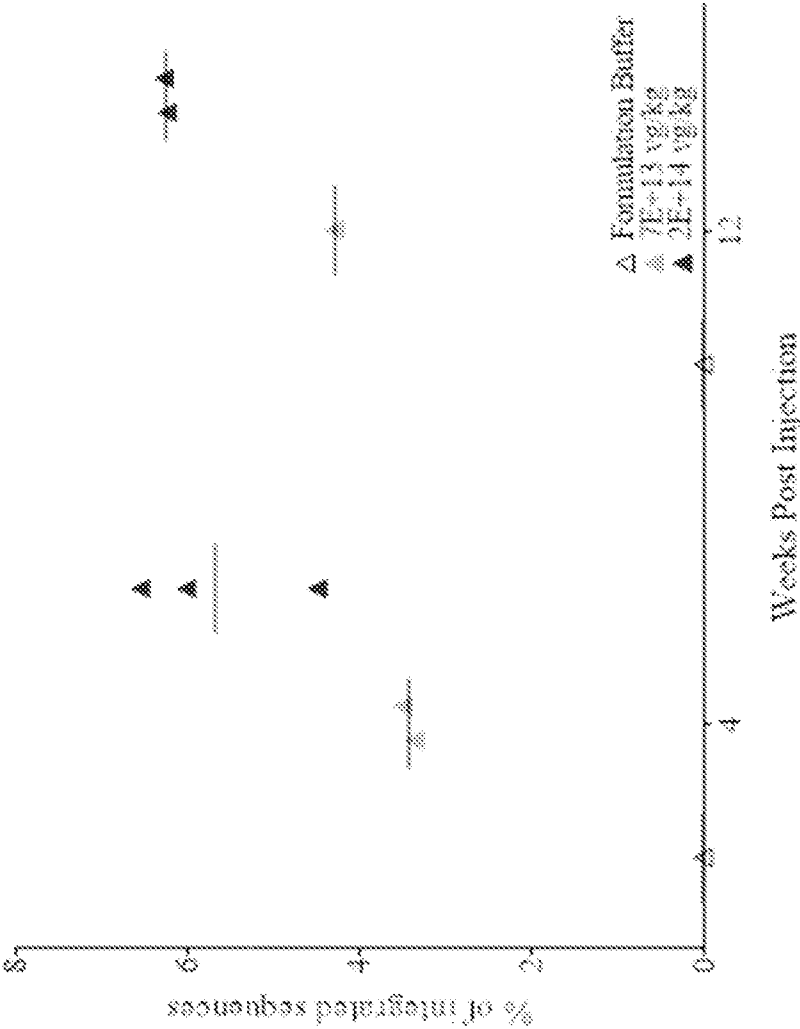


FIG. 14B

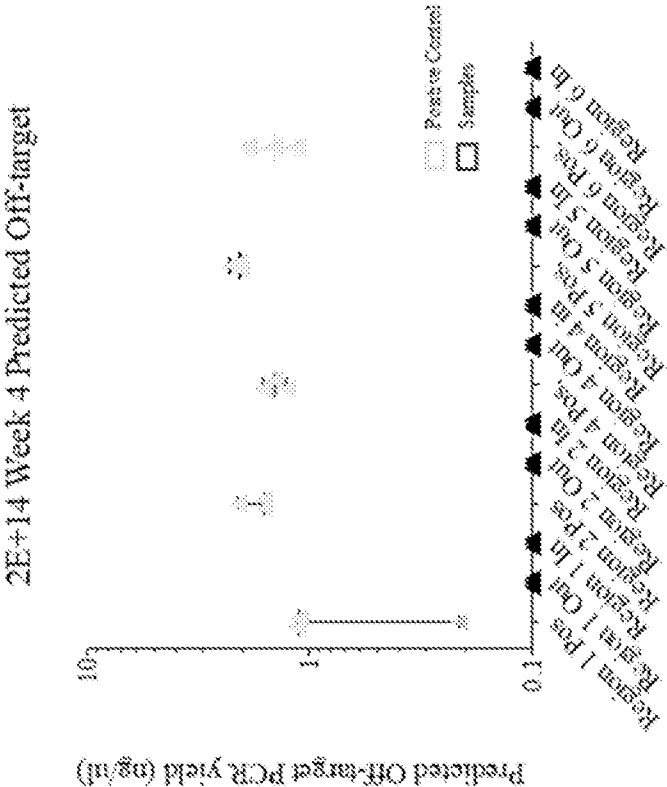


FIG. 14A

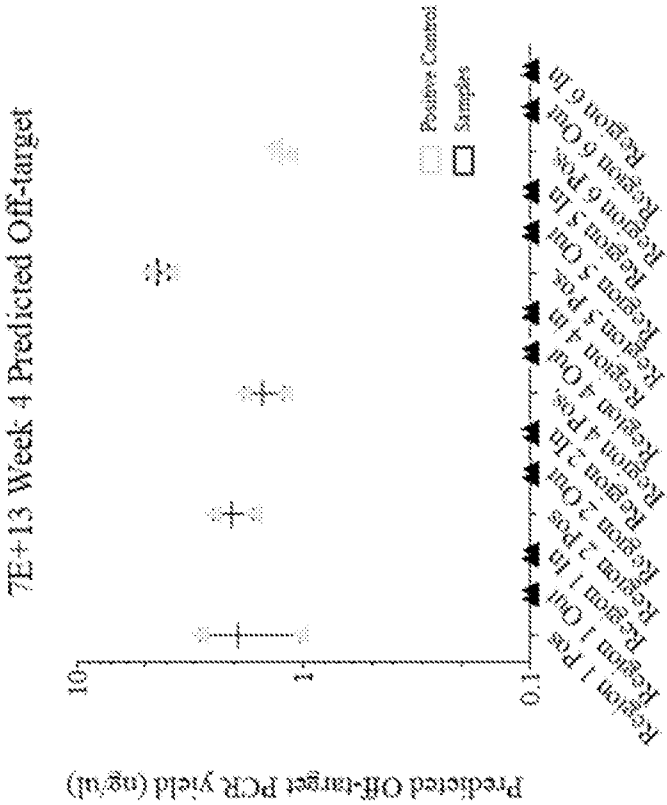


FIG. 14D

2E+14 Week 12 Predicted Off-target

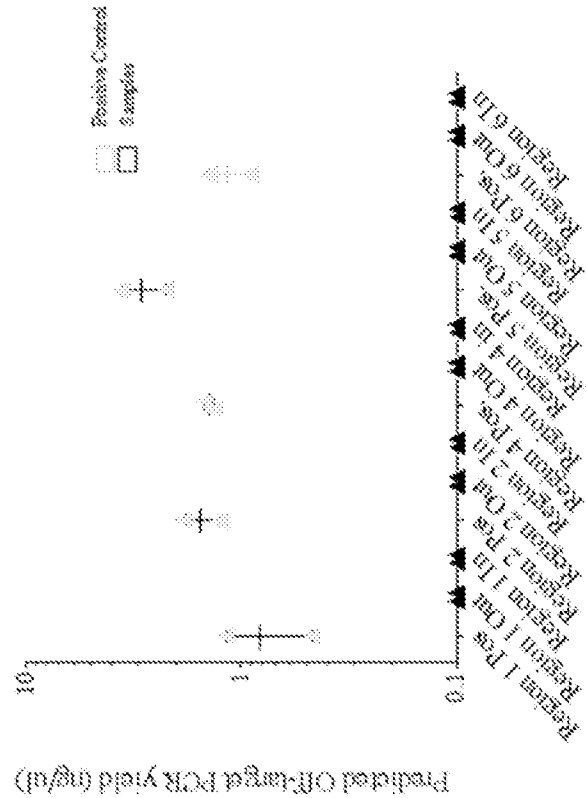
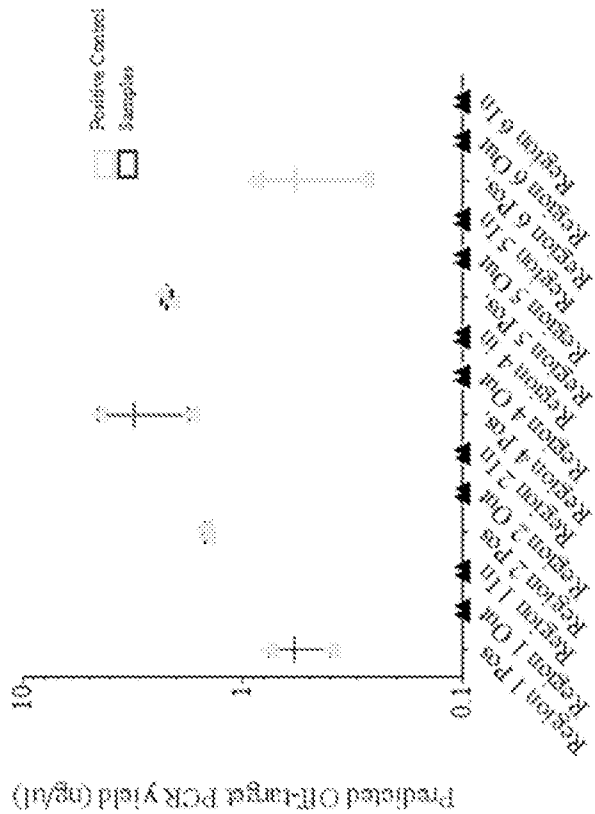


FIG. 14C

7E+13 Week 12 Predicted Off-target



SEQUENCE LISTING

<110> HOMOLOGY MEDICINES, INC.

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

<130> 404217-HMW-040WO (183161)

<150> US 63/030,341

<151> 2020-05-27

<150> US 63/117,252

<151> 2020-11-23

<160> 86

<170> PatentIn version 3.5

<210> 1

<211> 736

<212> PRT

<213> Artificial Sequence

<220>

<223> adeno-associated virus 9

<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> 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> 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> AAV isolate

<400> 4

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

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

Lys Ala Asn Gln Gln His Gln Asp Asn Ala Arg Gly Leu Val Leu Pro

35

40

45

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

<210> 5
<211> 736
<212> PRT
<213> Artificial Sequence

<220>
<223> AAV isolate

<400> 5

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

<210> 6
<211> 736
<212> PRT
<213> Artificial Sequence

<220>
<223> 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> 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> 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> 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> 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> 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> 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> 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
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
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
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> 145
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 14
 ttggccactc cctctctgcg cgctcgctcg ctcaactgagg ccgggcgacc aaaggtcgcc

cgacgccccg gctttgcccc ggccggcctca gtgagcgagc gagcgcgag agagggagtg	120
gccaaactcca tcactagggg ttcct	145

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

<220>
 <223> 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> 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> 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> Synthetic polynucleotide

<400> 18
aggaaccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctactgagg 60
ccgggcgacc aaaggtcgcc cgacgcccg gctttgccc ggcggcctca gtgagcgagc 120
gagcgcgag agagggagtg gccaa 145

<210> 19
<211> 106
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 19
ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgacctt 60
ggtcgcccgg cctcagtgag cgagcgagcg cgagagagg gaggtag 106

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

<220>
 <223> Synthetic polynucleotide

<400> 20
 ctctccccc tgcgcgttc gctcgctcgc tggctcgttt gggggggtgg cagctcaaag 60
 agctgccaga cgacggccct ctggccgtcg ccccccaaa cgagccagcg agcgagcgaa 120
 cgcgacaggg gggagagtgc cacactctca agcaaggggg ttttgta 167

<210> 21
 <211> 167
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 21
 tacaaaacct ctttgcttga gagtgtggca ctctccccc tgcgcgttc gctcgctcgc 60
 tggctcgttt gggggggtgg cagctcaaag agctgccaga cgacggccct ctggccgtcg 120
 ccccccaaa 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> 960
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 23
gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg 60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt 120

tctgaagtca ggcacagca aggcagtgtg cttagaggtt aacagaaggg aaaacaacaa	180
caacaaaaat ctaaattgaga atcctgactg ttccagctgg gggtaagggg ggcggattat	240
tcatataatt gttataaccag acggctgcag gcttagtcca attgcagaga actcgcttcc	300
caggctttctg agagtcccg gagtgcctaa acctgtctaa tcgacggggc ttgggtggcc	360
cgctgctccc tggctttctt cctttaccca gggcgggcag cgaagtgggt cctcctgcgt	420
ccccacacc ctccctcagc cctcccctc cggcccgctc tgggcagggt acctggagca	480
tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc	540
ggagatgcac cacgcaagag acaccctttg taactctctt ctctcccta gtgcgaggtt	600
aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc	660
agagacctca ctcccgggga gccagcatgt cactgcgggt cctggaaaac ccaggcttgg	720
gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg	780
gaatttgggc ctccagagaa agagatctga agactgctgg tgcttcctgg ttccataagc	840
tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa	900
atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa	960

<210> 24
 <211> 910
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 24	
ctgggatggg atgtggaatc cttctagatt tcttttgtaa tatttataaa gtgctctcag	60
caaggatatca aaatggcaaa attgtgagta actatcctcc ttccattttg ggaagaagat	120
gaggcatgaa gagaattcag acagaaactt actcagacca ggggaggcag aaactaagca	180
gagaggaaaa tgaccaagag ttagccctgg gcatggaatg tgaaagaacc ctaaactga	240
cttggaata atgcccagg tatattccat tctccgggat ttgttgcat tttcttgagg	300
tgaagaattg cagaatacat tctttaatgt gacctacata ttaccatg ggaggaagtc	360
tgctcctgga ctcttgagat tcagtcataa agcccaggcc agggaaataa tgtaagtctg	420
caggcccctg tcatcagtag gattaggagg aagagttctc agtagaaaac agggaggctg	480
gagagaaaag aatggttaat gttacgtta atataactag aaagactgca gaacttagga	540
ctgattttta ttgaatcct taaaaaaaaa atttcttatg aaaatagtag atggctctta	600
ggagacagaa ctattgttac agaggaacag cgtgagagtc agagtgatcc cagaacaggt	660
cctggctcca tcctgcacat agttttggtg ctgctggcaa tacggtcccc acaactgtgg	720
gaaggggtta ggggcaggga tctcatcagg aaagcatagg ggtttaaagt tctttataga	780
gcacttagaa gattgagaat ccacaaatta tattaataac aaacaaagta gtgtcgtgtt	840
atatagtaaa tgtgaatttg cagacacatt tagggaaaag ttataattaa aaaaataggc	900
tgatatata	910

<210> 25
<211> 192
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 25
ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc 60
tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc 120
cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt 180
agtgtgagag gg 192

<210> 26
<211> 255
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 26
aatgactcct ttcggtaagt gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc 60
agcgtaggcg ggcgactcag atcccagcca gtggacttag cccctgtttg ctctccgat 120
aactgggggtg accttggtta atattcacca gcagcctccc ccgttgcccc tctggatcca 180
ctgcttaaat acggacgagg acagggccct gtctcctcag cttcaggcac caccactgac 240
ctgggacagt gaatc 255

<210> 27
<211> 448
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 27
ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc 60
tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc 120
cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt 180
agtgtgagag gggaatgact ctttcggta agtgcagtgg aagctgtaca ctgcccaggc 240
aaagcgtccg ggcagcgtag gcgggcgact cagatcccag ccagtggact tagcccctgt 300
ttgtcctcc gataactggg gtgaccttg ttaatattca ccagcagcct cccccgttgc 360
ccctctggat cactgctta aatacggacg aggacagggc cctgtctcct cagcttcagg 420
caccaccact gacctgggac agtgaatc 448

<210> 28
<211> 1356
<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide

<400> 28

atgtccaccg ctgtgctgga gaaccctggg ctggggagga aactgtcaga cttcgggcag	60
gagacttcat acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc	120
ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg	180
aatctgaccc acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt	240
accacacttg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac	300
gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg	360
ttccctcgga caatccagga gctggataga ttgccaacc agatcctgtc ttacggagca	420
gagctggacg cagatcaccg tggcttcaag gaccagtgat atcgggcccg gagaaagcag	480
tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg	540
gaggaggaga agaagacctg gggcacagtg ttcaagacct tgaagagcct gtacaagaca	600
cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac	660
gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt	720
aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc	780
agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag	840
cccgacatct gtcacgagct gctgggccac gtgccctgt 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 gaaggtgaga	1200
aatttcgccg ccacaatccc taggcccttc agtgtgcgtt acgaccctta taccagagg	1260
atcgaggtgc tggataatac acagcagctg aagatcctgg ctgactcaat caatagcgaa	1320
atcggaatcc tgtgctccgc cctgcagaaa atcaaa	1356

<210> 29

<211> 93

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide

<400> 29

ctctaaggta aatataaaat ttttaagtgt ataatgtgtt aaactactga ttctaattgt	60
ttctctcttt tagattccaa cctttggaac tga	93

<210> 30

<211> 26

<212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 30
 ctgacctctt ctcttcctcc cacagg 26

<210> 31
 <211> 198
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 31
 gatccagaca tgataagata cattgatgag ttggacaaa ccacaactag aatgcagtga 60
 aaaaaatgct ttatttgtga aatttgtgat gctattgctt tatttgtaac cattataagc 120
 tgcaataaac aagttaacaa caacaattgc attcatttta tgtttcaggt tcagggggag 180
 gtgtgggagg ttttttaa 198

<210> 32
 <211> 143
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 32
 aggaaccctt agtgatggag ttggccactc cctctctgcg cgctcgctcg ctactgagg 60
 ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc 120
 gagcgcgag agagggagtg gcc 143

<210> 33
 <211> 452
 <212> PRT
 <213> Homo sapiens

<400> 33

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> 34
<211> 122
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 34
aacttgttta ttgcagctta taatggttac aaataaagca atagcatcac aaatttcaca 60
aataaagcat ttttttcact gcattctagt tgtggtttgt ccaaactcat caatgtatct 120
ta 122

<210> 35
<211> 133
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 35
tgctttatatt gtgaaatttg tgatgctatt gctttatttg taaccattat aagctgcaat 60
aaacaagtta acaacaacaa ttgcattcat tttatgtttc aggttcaggg ggaggtgtgg 120
gaggtttttt aaa 133

<210> 36
<211> 1676
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 36
ggcattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc 60
catatatgga gttccgcgtt acataactta cggtaaattg cccgcctggc tgaccgcca 120
acgacccccg ccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga 180
ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc 240
aagtgtatca tatgccaagt ccgcccccta ttgacgtcaa tgacggtaaa tggccgcct 300
ggcattatgc ccagtacatg accttacggg actttcctac ttggcagtac atctacgtat 360
tagtcatcgc tattaccatg gtcgagggtga gcccacgtt ctgcttact ctccccatct 420
ccccccctc cccaccccca attttgtatt tatttatttt ttaattattt tgtgcagcga 480
tgggggcggg gggggggggg gggcgcgcg caggcggggc ggggcggggc gaggggcggg 540
gcggggcgag gcggagaggt gcggcggcag ccaatcagag cggcgcgctc cgaaagtttc 600
cttttatggc gaggcggcgg gcggcggcgg cctataaaaa gcgaagcgcg cggcgggcgg 660
gagtcgctgc gcgctgcctt cgccccgtgc ccgctccgc cgccgcctcg cgccgccgcg 720
ccggctctg actgaccgcg ttactccac aggtgagcgg gcgggacggc cttctcctc 780
cgggctgtaa ttagcgcttg gtttaatgac ggcttgtttc ttttctgtgg ctgctgaaa 840
gccttgaggg gctccgggag ggccctttgt gcggggggag cggctcgggg ggtgcgtgcg 900
tgtgtgtgtg cgtggggagc gccgcgtgcg gctccgcgct gccgggcggc tgtgagcgt 960
gcgggcgcgg cgccgggctt tgtgcgtcc gcagtgtgcg cgaggggagc gcggccgggg 1020
gcggtgcccc gcggtgcggg gggggctgcg aggggaacaa aggctgcgtg cggggtgtgt 1080
gcgtgggggg gtgagcaggg ggtgtgggcg cgtcggtcgg gctgcaaccc cccctgcacc 1140
cccctccccg agttgctgag cacggcccgg cttcgggtgc ggggctccgt acggggcgtg 1200
gcgcggggct cgccgtgccg ggcggggggt ggcggcaggt gggggtgccg ggcggggcgg 1260
ggccgcctcg ggccggggag ggctcggggg aggggcgcgg cggcccccgg agcgccggcg 1320
gctgtcgagg cgggcgagc cgagccatt gccttttatg gtaatcgtgc gagaggcgcg 1380
agggacttcc tttgtccaa atctgtgcgg agccgaaatc tgggaggcgc cgccgcaccc 1440
cctctagcgg gcgcggggcg aagcgggtgc gcgccggcag gaaggaaatg ggcggggagg 1500
gccttcgtgc gtcgcgcgc cgccgtcccc ttctccctc ccagcctcgg ggctgtccgc 1560
ggggggacgg ctgccttcgg gggggacggg gcagggcggg gttcggcttc tggcgtgtga 1620
ccggcggctc tagagcctct gctaaccatg ttcatgcctt cttctttttc ctacag 1676

<210> 37
<211> 960
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 37
gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg 60

ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt	120
tctgaagtca gcgcacagca aggcagtgtg cttagagggtt aacagaaggg aaaacaacaa	180
caacaaaaat ctaaatgaga atcctgactg tttcagctgg gggtaagggg ggcggtattat	240
tcataataatt gttataaccag acggctgcag gcttagtcca attgcagaga actcgcttcc	300
caggcttctg agagtcccg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc	360
cgctgctccc tggcttcttc cttttaccca gggcgggcag cgaagtgggtg cctcctgcgt	420
ccccacacc ctccctcagc cctccctc cggccgtcc tgggcagggtg acctggagca	480
tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc	540
ggagatgcac cacgcaagag acaccctttg taactctctt ctctcccta gtgcgagggtt	600
aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc	660
agagacctca ctccgggga gccagcatgt ccactgcggt cctggaaaac ccaggcttgg	720
gcaggaaact ctctgacttt ggacagggtga gccacggcag cctgagctgc tcagttaggg	780
gaatttgggc ctccagagaa agagatccga agactgctgg tgcttcctgg tttcataagc	840
tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa	900
atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa	960

<210> 38
 <211> 911
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 38	
ctgggatggg atgtggaatc cttctagatt tcttttgtaa tatttataaa gtgctctcag	60
caaggtatca aaatggcaaa attgtgagta actatcctcc tttcattttg ggaagaagat	120
gaggcatgaa gagaattcag acagaaactt actcagacca ggggaggcag aaactaagca	180
gagaggaaaa tgaccaagag ttagccctgg gcatggaatg tgaaagaacc ctaaactga	240
cttggaata atgccaagg tatattccat tctccgggat ttgttggcat tttcttgagg	300
tgaagaattg cagaatacat tctttaatgt gacctacata tttacccatg ggaggaagtc	360
tgctcctgga ctcttgagat tcagtcataa agcccaggcc agggaaataa tgtaagtctg	420
caggcccctg tcatcagtag gattagggag aagagttctc agtagaaaac agggaggctg	480
gagagaaaag aatggttaat gttaacgtta atataactag aaagactgca gaacttagga	540
ctgattttta tttgaatcct taaaaaaaaa aatttcttat gaaaatagta catggctctt	600
aggagacaga acttattgta cagaggaaca gcgtgagagt cagagtgatc ccagaacagg	660
tcctggctcc atcctgcaca tagttttggg gctgctggca atacgggtccc cacaactgtg	720
ggaaggggtt aggggcaggg atctcatcag gaaagcatag gggtttaaag ttctttatag	780
agcacttaga agattgagaa tccacaaatt atattaataa caaaciaaagt agtgtcgtgt	840
tatatagtaa atgtgaattt gcagacacat ttagggaaaa gttataatta aaaaaatagg	900

<210> 39
 <211> 1168
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 39
 cgtgaggctc cggtgccgt cagtgggcag agcgcacatc gcccacagtc cccgagaagt 60
 tggggggagg ggtcggaat tgaaccggtg cctagagaag gtggcgcggg gtaaactggg 120
 aaagtgatgt cgtgtactgg ctccgccttt ttcccagggg tgggggagaa ccgtatataa 180
 gtgcagtagt cgccgtgaac gttctttttc gcaacggggtt tgccgccaga acacaggtaa 240
 gtgccgtgtg tggttccgc gggcctggcc tctttacggg ttatggccct tgcgtgcctt 300
 gaattacttc cacctggctc cagtacgtga ttcttgatcc cgagctggag ccaggggcgg 360
 gccttgcgt ttaggagccc cttcgcctcg tgcttgagtt gaggcctggc ctgggcgctg 420
 gggccgccgc gtgcgaatct ggtggcacct tcgcgctgt ctcgctgctt tcgataagtc 480
 tctagccatt taaaattttt gatgacctgc tgcgacgctt tttttctggc aagatagtct 540
 tgtaaatgcg ggccaggatc tgcacactgg tatctcggtt tttggggccg cgggcggcga 600
 cggggcccgt gcgctccagc gcacatgttc ggcgaggcgg ggcctgcgag cgcggccacc 660
 gagaatcgga cgggggtagt ctcaagctgg ccggcctgct ctggtgcctg gcctcgcgcc 720
 gccgtgtatc gccccgccct gggcggcaag gctggcccgg tcggcaccag ttgcgtgagc 780
 ggaaagatgg ccgcttcccg gccctgctcc aggggggtca aaatggagga cgcggcgctc 840
 gggagagcgg gcgggtgagt caccacaca aaggaaaggg gcctttccgt cctcagccgt 900
 cgcttcatgt gactccacgg agtaccgggc gccgtccagg cacctcgatt agttctggag 960
 cttttggagt acgtcgtctt taggttgggg ggagggggtt tatgcatgg agtttccca 1020
 cactgagtgg gtggagactg aagttaggcc agcttggcac ttgatgtaat tctccttga 1080
 atttgcctt tttgagtttg gatcttggtt cattctcaag cctcagacag tggttcaaag 1140
 tttttttctt ccatctcagg tgtcgtga 1168

<210> 40
 <211> 72
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 40
 agatctggca gcggagaggg cagaggaagt cttctaacat gcggtgacgt ggaggagaat 60
 cccggcccta gg 72

<210> 41

<211> 95
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 41
cctctgctaa ccatgttcat gccttcttct ttttctaca gctcctgggc aacgtgctgg 60
ttattgtgct gtctcatcat ttggcaaag aattc 95

<210> 42
<211> 1873
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 42
gatcttcaat attggccatt agccatatta ttcatlggtt atatagcata aatcaatatt 60
ggctattggc cattgcatac gttgtatcta tatcataata tgtacattta tattggctca 120
tgtccaatat gaccgcatg ttggcattga ttattgacta gttattaata gtaatcaatt 180
acgggggtcat tagttcatag cccatatatg gagttccgcg ttacataact tacggtaa 240
ggcccgctg gctgaccgcc caacgacccc cgccattga cgtcaataat gacgtatgtt 300
cccatagtaa cgccaatagg gactttccat tgacgtcaat ggggtggagta ttacggtaa 360
actgcccact tggcagtaca tcaagtgtat catatgcaa gtccgcccc tattgacgtc 420
aatgacggta aatggccgc ctggcattat gccagtaca tgaccttacg ggactttcct 480
acttggcagt acatctacgt attagtcac gctattacca tggctcaggt gagccccacg 540
ttctgcttca ctctcccat ctccccccc tccccacccc caattttgta tttatttatt 600
ttttaattat ttgtgcagc gatgggggcg gggggggggg gggggcgcg gccaggcggg 660
gcggggcggg gcgaggggcg gggcggggcg aggcggagag gtgcggcggc agccaatcag 720
agcggcgcg tccgaaagt tctttttatg gcgaggcggc ggcggcggc gccctataaa 780
aagcgaagcg gcggcgggc gggagtctg gcgacgtgc cttcgcccc tgccccgctc 840
cgccgcccgc tcgcccgc cgccccggct ctgactgacc gcgttactcc cacaggtag 900
cgggcgggac ggcccttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt 960
ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg 1020
ggagcggctc ggggggtgctg tgcgtgtgtg tgtgcgtggg gagcgccgcg tgcggccgc 1080
gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgagtgt 1140
gcgcgagggg agcgcggccg gggcggtgc cccgcggtgc ggggggggct gcgaggggaa 1200
caaaggctgc gtgcggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcggcggt 1260
cgggctgtaa cccccctg cccccctc ccgagttgc tgagcacggc ccggcttcgg 1320
gtgcggggct ccgtacgggg cgtaggcgg ggctcgccgt gccggcggg gggtagggc 1380
aggtgggggt gccggcggg gcggggccgc ctcgggccgg ggagggtcg ggggaggggc 1440

gcggcgcccc ccggagcgcc ggcggtgtc gaggcgcggc gagccgcagc cattgccttt	1500
tatggtaatc gtgcgagagg gcgcagggac ttctttgtc ccaaactgt gcggagccga	1560
aatctgggag gcgccgccgc acccctcta gcggcgcggg ggcaagcgg tgcggcgccg	1620
gcaggaagga aatgggcggg gagggccttc gtgcgtcgcc gcgccgccgt ccccttctcc	1680
ctctccagcc tcggggctgt ccgcgggggg acggctgcct tcggggggga cggggcaggg	1740
cggggttcgg cttctggcgt gtgaccggcg gctctagagc ctctgctaac catgttcattg	1800
ccttcttctt tttctacag ctctgggca acgtgctggt tattgtgctg tctcatcatt	1860
ttggcaaaga att	1873

<210> 43
 <211> 4046
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 43	
gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg	60
ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt	120
tctgaagtca gcgcacagca aggcagtgtg cttagagggt aacagaaggg aaaacaacaa	180
caacaaaaat ctaaatgaga atcctgactg ttctagctgg gggtaagggg ggcggtattat	240
tcataataatt gttataaccag acggctgcag gcttagtcca attgcagaga actcgcttcc	300
caggcttctg agagtcccg aagtgcctaa acctgtctaa tcgacggggc ttgggtggcc	360
cgctgctccc tggcttcttc cctttaccca gggcgggcag cgaagtgggt cctcctgcgt	420
ccccacacc ctccctcagc cctcccctc cggcccgtcc tgggcagggt acctggagca	480
tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc	540
ggagatgcac cacgcaagag acacccttg taactctctt ctctcccta gtgcgaggtt	600
aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc	660
agagacctca ctcccgggga gccagcatgt cactgcggt cctggaaaac ccaggcttgg	720
gcaggaaact ctctgacttt ggacaggtga gccacggcag cctgagctgc tcagttaggg	780
gaatttgggc ctccagagaa agagatctga agactgctgg tgcttcctgg ttccataagc	840
tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa	900
atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa	960
ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc	1020
tgaccttga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc	1080
cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt	1140
agtgtgagag gggaatgact ctttctggta agtgcagtgg aagctgtaca ctgcccaggc	1200
aaagcgtccg ggacgcgtag gcgggcgact cagatcccag ccagtggact tagcccctgt	1260

ttgctcctcc gataactggg gtgaccttgg ttaatatcca ccagcagcct cccccgttgc	1320
ccctctggat cactgtctta aatacggacg aggacagggc cctgtctcct cagcttcagg	1380
caccaccact gacctgggac agtgaatcct ctaaggtaaa tataaaattht ttaagtgtat	1440
aatgtgttaa actactgatt ctaattgttht ctctctthtta gattccaacc tttggaactg	1500
accgccacca tgtccaccgc tgtgtctggag aaccctgggc tggggaggaa actgtcagac	1560
ttcgggcagg agacttcata cattgaggat aactgtaacc agaattggcg cactctctctg	1620
atcttcagcc tgaaggagga agtgggcgcc ctggcaaagg tgctgcgcct gtttgaggag	1680
aacgacgtga atctgaccca catcgagtcc cggccttcta gactgaagaa ggacgagtac	1740
gagttcttht cccacctgga taagcggctc ctgccagccc tgacaaacat catcaagatc	1800
ctgaggcacg acatcggagc aaccgtgcac gagctgtctc gggacaagaa gaaggatacc	1860
gtgccctggt tccctcggac aatccaggag ctggatagat ttgccaacca gatcctgtct	1920
tacggagcag agctggacgc agatcacctt ggcttcaagg acccagtgtat tcggggcccg	1980
agaaagcagt ttgccgatat cgcctacaat tataggcacg gacagccaat cctctgcgtg	2040
gagtatatgg aggaggagaa gaagacctgg ggcacagtgt tcaagacctt gaagagcctg	2100
tacaagacac acgcctgcta cgagtataac cacatcttcc cctgtctgga gaagtattgt	2160
ggctttcacg aggacaatat cctcagctg gaggacgtga gccagttcct gcagacctgc	2220
acaggcttht ggctgaggcc agtggcagga ctgtctgagct cccgggactt cctgggagga	2280
ctggccttca gagtgttht ctgcaccag tacatcaggc acggctcaa gccaatgtat	2340
acaccagagc ccgacatctg tcacgagctg ctgggccacg tgcccctgth tagcgataga	2400
tccttcgccc agthtttcca ggagatcgga ctggcatctc tgggagcacc tgacgagtac	2460
atcgagaagc tggccaccat ctattggtht acagtggagt ttggcctgtg caagcagggc	2520
gatagcatca aggcctacgg agcaggactg ctgtctagct tcggcgagct gcagtattgt	2580
ctgtccgaga agccaaagct gctgcccctg gagctggaga agaccgcat ccagaactac	2640
accgtgacag agthccagcc cctgtactat gtggccgagt cthtttaacga tgccaaggag	2700
aagggtgagaa atthtcgccc cacaatccct aggcccttca gtgtgcgtht cgacccttat	2760
accagagga tcgagggtgt ggataataca cagcagctga agatcctggc tgactcaatc	2820
aatagcgaaa tcggaatcct gtgtctcgcc ctgcagaaaa tcaaatgaga attcaaggcc	2880
tctcgagcct ctagaactat agtgagtcgt attacgtaga tccagacatg ataagataca	2940
ttgatgagtht tggacaaacc acaactagaa tgcagtgaat aaaatgctth atttgtgaat	3000
thtgtgatgc tathgttht thtghtaacca ttataagctg caataaaca gthtaacaaca	3060
acaattgcat tcaththtath thtcaggtht agggggaggtht gtgggaggtht thtthaagctth	3120
tacgtacgat cgtcgactgg gatgggatgt ggaatcctt tagaththctth thtghtaatath	3180
tataaagthc tctcagcaag gtatcaaaat ggcaaaatht tgagthaacta tcctccttht	3240
aththtgggaa gaagatgagg catgaagaga attcagacag aaacttactc agaccagggg	3300
aggcagaaac taagcagaga ggaaaatgac caagagthttag ccttgggcat ggaatgtgaa	3360

agaaccctaa acgtgacttg gaaataatgc ccaaggtata ttccattctc cgggatttgt	3420
tggcattttc ttgaggtgaa gaattgcaga atacattctt taatgtgacc tacatattta	3480
cccatgggag gaagtctgct cctggactct tgagattcag tcataaagcc caggccaggg	3540
aaataatgta agtctgcagg cccctgtcat cagtaggatt agggagaaga gttctcagta	3600
gaaaacaggg aggctggaga gaaaagaatg gttaatgtta acgttaatat aactagaaag	3660
actgcagaac ttaggactga tttttatttg aatccttaaa aaaaaaattt cttatgaaaa	3720
tagtacatgg ctcttaggag acagaactta ttgtacagag gaacagcgtg agagtcagag	3780
tgatcccaga acaggtcctg gctccatcct gcacatagtt ttggtgctgc tggcaatacg	3840
gtccccacaa ctgtgggaag gggttagggg cagggatctc atcaggaaag cataggggtt	3900
taaagttctt tatagagcac ttagaagatt gagaatccac aaattatatt aataacaac	3960
aaagtagtgt cgtgttatat agtaaagtgt aatttgcaga cacatttagg gaaaagttat	4020
aattaaaaaa ataggctgta tatata	4046

<210> 44
 <211> 913
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 44	
ggcatcctaa aaaatattca gtggaaacgt aaaaacatta aagactgatt aaacatcgca	60
gcatgacaca gatttagcaa ctgagcataa ataatttgac tcggatactg ctccaaaatc	120
cgaagaggac caatttcttc caggaggaca actacctcgt cctctgcaga cccctctcct	180
cggcagctga aggagtgtgg ccaatctgcc tccacctccc cgcggaaccc ctactctcag	240
gacctcctgc agcaccctaa actggaagtg gccgctgcag acccaaggac gaggggcacg	300
cgggagccgg cagccctagt ggagcggttg gagatgttga ggtgggaggg tcaccaggt	360
ggggtgaggc tggggtaggt agcggagtga acggcttccg aagctctggg ccgccccag	420
gttgactaa gcaggcgctc tgtcttcgcc ccgcccagg gtgggcgtct cctgaggact	480
ccccgccaca cctgaccga gaccgcgcgc ccagcctaga acgcttcccc gaccagcgt	540
agggccgccg cgactggcgg gcgagggctg gcgggaggcc tggcgaaccc gggggcggga	600
ccaggcgggc aaggcccggc tgccgcagcg ccgctctgcg cgaggcggct ccgccgcggc	660
ggagggatac ggcgcacat atatatatcg cggggcgag actcgcgctc cggcagtgtt	720
gctgggagtg tcgtggacgc cgtgccgtta ctcgtagtca ggccggcgcg caggcggcgg	780
cggcggcata gcgcacagcg gccttagca gcagcagcag cagcagcggc atcggaggta	840
ccccgccgt cgagcccc gcgctggtgc agccaccctc gctccctctg ctcttcctcc	900
cttcgctcgc acc	913

<210> 45

<211> 4452
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 45
ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc 60
cgacgcccgg gctttgcccg ggccgcctca gtgagcgagc gagcgcgag agagggagtg 120
gccaactcca tctactagggg ttcctggagg ggtggagtcg tgacgtgaat tacgtcatag 180
ggttagggag gtcctgcata tgcggccgct tcaggagcag ttgtgcgaat agctggagaa 240
caccaggctg gatttaaacc cagatcgctc ttacatttgc tctttacctg ctgtgctcag 300
cgttcacgtg ccctctagct gtagttttct gaagtcagcg cacagcaagg cagtgtgctt 360
agaggttaac agaagggaaa acaacaaca caaaaatcta aatgagaatc ctgactgttt 420
cagctggggg taaggggggc ggattattca tataattgtt ataccagacg gtcgcaggct 480
tagtccaatt gcagagaact cgcttcccag gcttctgaga gtcccgaag tgcctaaacc 540
tgtctaatac acggggcttg ggtggcccgt cgctccctgg cttcttccct ttaccaggg 600
cgggcagcga agtggtgcct cctgcgtccc ccacaccctc cctcagcccc tcccctccgg 660
cccgtcctgg gcaggtgacc tggagcatcc ggccaggctgc cctggcctcc tgcgtcagga 720
caacgcccac gaggggcgtt actgtgcgga gatgcaccac gcaagagaca ccctttgtaa 780
ctctcttctc ctccctagtg cgaggttaaa accttcagcc ccacgtgctg tttgcaaacc 840
tgctgttacc tgaggcccta aaaagccaga gacctcactc ccggggagcc agcatgtcca 900
ctgcggctct ggaaaaccca ggcttgggca ggaaactctc tgactttgga caggtgagcc 960
acggcagcct gagctgtca gttaggggaa tttgggcctc cagagaaaga gatctgaaga 1020
ctgctggtgc ttcctggttt cataagctca gtaagaagtc tgaattcgtt ggaagctgat 1080
gagaatatcc aggaagtcaa cagacaaatg tcctcaaca ttgtttctaa gtaggagaac 1140
atctgtcctc ggtggctttc acaggaacct taaaatgggc aaacattgca agcagcaaac 1200
agcaaacaca cagccctccc tgctgtctga ccttgagact ggggcagagg tcagagacct 1260
ctctgggccc atgccacctc caacatccac tcgacccctt ggaatttcgg tggagaggag 1320
cagaggttgt cctggcgtgg tttaggtagt gtgagagggg aatgactcct ttcggttaagt 1380
gcagtggaag ctgtacactg cccaggcaaa gcgtccgggc agcgtaggcg ggcgactcag 1440
atcccagcca gtggacttag cccctgtttg ctctccgat aactgggggtg accttggtta 1500
atattcacca gcagcctccc ccgttgcccc tctggatcca ctgcttaaat acggacgagg 1560
acagggccct gtctcctcag cttcaggcac caccactgac ctgggacagt gaatcctcta 1620
aggtaaatat aaaattttta agtgtataat gtgttaaact actgattcta attgtttctc 1680
tcttttagat tccaaccttt ggaactgacc gccaccatgt ccaccgctgt gctggagaac 1740
cctgggctgg ggaggaaact gtcagacttc gggcaggaga cttcatacat tgaggataac 1800
tgtaaccaga atggcgccat ctctctgatc ttcagcctga aggaggaagt gggcgccctg 1860

gcaaaggtgc tgcgcctgtt tgaggagaac gacgtgaatc tgacccacat cgagtcccgg	1920
ccttctagac tgaagaagga cgagtacgag ttctttaccc acctggataa gcggtccctg	1980
ccagccctga caaacatcat caagatcctg aggcacgaca tcggagcaac cgtgcacgag	2040
ctgtctcggg acaagaagaa ggataccgtg ccctggttcc ctcggacaat ccaggagctg	2100
gatagatttg ccaaccagat cctgtcttac ggagcagagc tggacgcaga tcaccctggc	2160
ttcaaggacc cagtgtatcg ggcccggaga aagcagtttg ccgatatcgc ctacaattat	2220
aggcacggac agccaatccc tcgcgtggag tatatggagg aggagaagaa gacctggggc	2280
acagtgttca agaccctgaa gagcctgtac aagacacacg cctgctacga gtataaccac	2340
atcttcccc tgctggagaa gtattgtggc tttcacgagg acaatatccc tcagctggag	2400
gacgtgagcc agttcctgca gacctgcaca ggctttaggc tgaggccagt ggcaggactg	2460
ctgagctccc gggacttcct gggaggactg gccttcagag tgtttctactg caccagtac	2520
atcaggcacg gctccaagcc aatgtataca ccagagcccc acatctgtca cgagctgctg	2580
ggccacgtgc ccctgtttag cgatagatcc ttcgcccagt tttcccagga gatcggactg	2640
gcatctctgg gagcacctga cgagtacatc gagaagctgg ccaccatcta ttggttcaca	2700
gtggagtttg gcctgtgcaa gcagggcgat agcatcaagg cctacggagc aggactgctg	2760
tctagcttcg gcgagctgca gtattgtctg tccgagaagc caaagctgct gcccctggag	2820
ctggagaaga ccgccatcca gaactacacc gtgacagagt tccagcccct gtactatgtg	2880
gccgagtctt ttaacgatgc caaggagaag gtgagaaatt tcgccgccac aatccctagg	2940
cccttcagtg tgcgttacga cccttatacc cagaggatcg aggtgctgga taatacacag	3000
cagctgaaga tcctggctga ctcaatcaat agcgaaatcg gaatcctgtg ctccgccctg	3060
cagaaaaatca aatgagaatt caaggcctct cgagcctcta gaactatagt gagtcgtatt	3120
acgtagatcc agacatgata agatacattg atgagtttgg acaaaccaca actagaatgc	3180
agtgaaaaaa atgctttatt tgtgaaattt gtgatgctat tgctttattt gtaaccatta	3240
taagctgcaa taaacaagtt aacaacaaca attgcattca ttttatgttt caggttcagg	3300
gggagggtgtg ggagggtttt taagctttac gtacgatcgt cgactgggat gggatgtgga	3360
atccttctag atttcttttg taatatattt aaagtgtctt cagcaaggta tcaaaatggc	3420
aaaattgtga gtaactatcc tcctttcatt ttgggaagaa gatgaggcat gaagagaatt	3480
cagacagaaa cttactcaga ccaggggagg cagaaactaa gcagagagga aaatgaccaa	3540
gagttagccc tgggcatgga atgtgaaaga accctaaacg tgacttggaataaatgccca	3600
aggatatattc cattctccgg gatttgttgg cattttcttg aggtgaagaa ttgcagaata	3660
cattctttaa tgtgacctac atattttacc atgggaggaa gtctgctcct ggactcttga	3720
gattcagtca taaagcccag gccagggaaa taatgtaagt ctgcaggccc ctgtcatcag	3780
taggattagg gagaagagtt ctcagtagaa aacagggagg ctggagagaa aagaatggtt	3840
aatgttaacg ttaatatataac tagaaagact gcagaactta ggactgattt ttatttgaat	3900

ccttaaaaaa aaaatttctt atgaaaatag tacatggctc ttaggagaca gaacttattg	3960
tacagaggaa cagcgtgaga gtcagagtga tcccagaaca ggtcctggct ccatcctgca	4020
catagttttg gtgctgctgg caatacggtc cccacaactg tgggaagggg ttaggggcag	4080
ggatctcatc aggaaagcat aggggtttta agttctttat agagcactta gaagattgag	4140
aatccacaaa ttatattaat aacaaacaaa gtagtgtcgt gttatatagt aaatgtgaat	4200
ttgcagacac atttagggaa aagttataat taaaaaata ggctgtatat atacctgcag	4260
gtctagatac gtagataagt agcatggcgg gttaatcatt aactacaagg aaccctagt	4320
gatggagttg gccactccct ctctgcgcgc tcgctcgctc actgaggccg ggcgacaaaa	4380
ggtcgcccga cgccgggct ttgccgggc gccctcagtg agcgagcgag cgcgagaga	4440
gggagtggcc aa	4452

<210> 46
 <211> 380
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 46	
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 cgcccccta ttgacgtcaa tgacggtaaa tggccgcct	300
ggcattatgc ccagtacatg accttacggg actttcctac ttggcagtac atctacgtat	360
tagtcatcgc tattaccatg	380

<210> 47
 <211> 1246
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 47	
tcgaggtgag cccacgttc tgcttcactc tccccatctc cccccctcc ccacccccaa	60
ttttgtattht atttatthtt taattatthtt gtgcagcgat gggggcgggg gggggggggg	120
ggcgcgcgcc aggcggggcg gggcggggcg aggggcgggg cggggcgagg cggagaggtg	180
cggcggcagc caatcagagc ggcgcgctcc gaaagtttcc ttttatggcg aggcggcggc	240
ggcggcggcc ctataaaaag cgaagcgcgc ggcgggcggg agtcgctgcg cgctgccttc	300
gccccgtgcc ccgctccgcc gccgcctcgc gccgcccgcc ccggctctga ctgaccgcgt	360
tactcccaca ggtgagcggg cgggacggcc cttctcctcc gggctgtaat tagcgcttg	420
tttaatgacg gcttgtttct tttctgtggc tgcgtgaaag ccttgagggg ctccgggagg	480

gccctttgtg cggggggagc ggctcggggg gtgcgtgcgt gtgtgtgtgc gtggggagcg	540
ccgcgtgcgg ctccgcgctg cccggcggct gtgagcgctg cgggcgcggc gcggggcttt	600
gtgcgtccg cagtgtgcgc gaggggagcg cggccggggg cggtgccccg cggtcggggg	660
ggggctgcga ggggaacaaa ggctgcgtgc ggggtgtgtg cgtggggggg tgagcagggg	720
gtgtgggcgc gtcggtcggg ctgcaacccc ccctgcaccc ccctccccga gttgctgagc	780
acggccccgc ttcgggtgcg gggctccgta cggggcgtgg cgcggggctc gccgtgccgg	840
gcgggggggtg gcggcaggtg ggggtgccgg gcggggcggg gccgcctcgg gccggggagg	900
gctcggggga ggggcgcggc ggcccccgga gcgccggcgg ctgtcgaggc gcggcgagcc	960
gcagccattg ctttttatgg taatcgtgcg agagggcga gggacttcct ttgtccaaa	1020
tctgtcgga gccgaaatct gggaggcgcc gccgcacccc ctctagcggg cgcggggcga	1080
agcggtcgg cgccggcagg aaggaaatgg gcggggaggg ctttcgtgcg tcgccgcgcc	1140
gccgtcccct tctccctctc cagcctcggg gctgtccgcg gggggacggc tgccttcggg	1200
ggggacgggg cagggcgggg ttcggcttct ggcgtgtgac cggcgg	1246

<210> 48
 <211> 1061
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 48	
tagggaggtc ctgcacgtta cataacttac ggtaaattggc ccgcctggct gaccgccc	60
cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc caatagggac	120
tttccattga cgtcaatggg tggagtatctt acggtaaaact gcccacttgg cagtacatca	180
agtgtatcat atgccaagta cggccctat tgacgtcaat gacggtaa	240
gcattatgcc cagtacatga ctttatggga ctttcctact tggcagtaca tctacgtatt	300
agtcacgcgt attaccatgg tcgaggtag cccacgttc tgcttcactc tccccatctc	360
ccccccctcc ccacccccaa ttttgtatctt atttattttt taattatttt gtgcagcgat	420
gggggcgggg gggggggggg gcgcgcgcca ggccggggcgg ggccggggcga ggggcggggc	480
ggggcgaggc ggagaggtgc ggcggcagcc aatcagagcg gcgcgctccg aaagtcttct	540
tttatggcga ggcggcgggc gcggcgggcc tataaaaagc gaagcgcgcg gcggcgggga	600
gtcgtgcgc gctgccttcg ccccggtgcc cgctccgcc cgcctcgcg ccggcgccc	660
cggctctgac tgaccgcgtt actaaaacag gtaagtccgg cctccgcgc gggttttggc	720
gcctccgcg ggcgcccccc tcctcacggc gagcgtgcc acgtcagacg aaggcgcgag	780
cgagcgtcct gatccttccg cccggacgct caggacagcg gcccgctgct cataagactc	840
ggccttagaa cccagtatc agcagaagga catttttagga cgggacttgg gtgactctag	900
ggcactgggtt ttttttcag agagcggaac aggcgaggaa aagtagtccc ttctcggcga	960

ttctgcggag ggatctccgt ggggcggtga acgccgatga tgcctctact aaccatgttc	1020
atgttttctt ttttttcta caggtcctgg gtgacgaaca g	1061

<210> 49
 <211> 953
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 49	
aattcggtag cctagttatt aatagtaatc aattacgggg tcattagttc atagcccata	60
tatggagttc cgcgttacat aacttacggt aaatggcccg cctggctgac cgcccaacga	120
ccccgcccc ttgacgtcaa taatgacgta tggtcccata gtaacgcaa tagggacttt	180
ccattgacgt caatgggtgg actatttacg gtaactgcc cacttggcag tacatcaagt	240
gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaattggc ccgcctggca	300
ttatgcccag tacatgacct tatgggactt tcctacttgg cagtacatct acgtattagt	360
catcgctatt accatgggtcg aggtgagccc cacgttctgc ttcactctcc ccatctcccc	420
cccctcccca ccccaattt tgtatttatt tattttttaa ttattttgtg cagcgatggg	480
ggcggggggg gggggggggc gcgcgccagg cggggcgggg cggggcgagg ggcggggcgg	540
ggcgaggcgg agaggtgcgg cggcagccaa tcagagcggc gcgctccgaa agtttccttt	600
tatggcgagg cggcggcggc ggcggcccta taaaaagcga agcgcgcggc gggcgggagt	660
cgctcgacg ctgccttcgc cccgtgcccc gctccgccgc cgcctcgcgc cgcccgcccc	720
ggctctgact gaccgcgtta ctccacagg tgagcgggcg ggacggccct tctcctccgg	780
gctgtaatta gcgcttggtt taatgacggc ttgtttcttt tctgtggctg cgtgaaagcc	840
ttgaggggct ccgggagcta gagcctctgc taaccatgtt catgccttct tcttttcct	900
acagctcctg ggcaacgtgc tggttattgt gctgtctcat cattttggca aag	953

<210> 50
 <211> 541
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 50	
ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc	60
tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc	120
cactcgaccc cttggaattt cggtaggagag gagcagaggt tgcctggcg tggtttaggt	180
agtgtgagag gggaatgact ctttcggta agtgcagtgg aagctgtaca ctgcccaggc	240
aaagcgtccg ggcagcgtag gcgggcgact cagatcccag ccagtggact tagcccctgt	300
ttgctcctcc gataactggg gtgaccttgg ttaatatcca ccagcagcct ccccgttgc	360
ccctctggat ccactgctta aatacggacg aggacagggc cctgtctcct cagcttcagg	420

caccaccact gacctgggac agtgaatcct ctaaggtaaa tataaaatTT ttaagtgtat	480
aatgtgttaa actactgatt ctaattgttt ctctctttta gattccaacc tttggaactg	540
a	541

<210> 51
 <211> 1908
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 51	
ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc	60
tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc	120
cactcgaccc cttggaattt cgggtggagag gagcagaggt tgtcctggcg tggtttaggt	180
agtgtgagag gggaatgact ctttctggta agtgcagtgg aagctgtaca ctgcccaggc	240
aaagcgtccg ggagcgtag gcgggagcact 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 tgtgtcaggag aaccctgggc tggggaggaa actgtcagac	600
ttcgggcagg agacttcata cattgaggat aactgtaacc agaattggcg catctctctg	660
atcttcagcc tgaaggagga agtgggagcc 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 agatcacccct ggcttcaagg acccagtgtg tcggggcccg	1020
agaaagcagt ttgccgatat gcctacaat tataggcacg gacagccaat ccctcgcgtg	1080
gagtatatgg aggaggagaa gaagacctgg ggcacagtgt tcaagaccct gaagagcctg	1140
tacaagacac acgcctgcta cgagtataac cacatcttcc ccctgctgga gaagtattgt	1200
ggctttcacg aggacaatat cctcagctg gaggacgtga gccagttcct gcagacctgc	1260
acaggcttta ggctgaggcc agtggcagga ctgctgagct cccgggactt cctgggagga	1320
ctggccttca gagggtttca ctgcacccag tacatcaggc acggctcaa gccaatgtat	1380
acaccagagc ccgacatctg tcacgagctg ctgggccacg tgcccctgtt tagcgataga	1440
tccttcgccc agttttccca ggagatcgga ctggcatctc tgggagcacc tgacgagtac	1500
atcgagaagc tggccacat ctattggttc acagtggagt ttggcctgtg caagcagggc	1560

gatagcatca aggcctacgg agcaggactg ctgtctagct tcggcgagct gcagtattgt	1620
ctgtccgaga agccaaagct gctgcccctg gagctggaga agaccgcat ccagaactac	1680
accgtgacag agttccagcc cctgtactat gtggccgagt cttttaacga tgccaaggag	1740
aaggtgagaa atttcgccgc cacaatccct aggcccttca gtgtgctgta cgacccttat	1800
accagagga tcgaggtgct ggataatata cagcagctga agatcctggc tgactcaatc	1860
aatagcgaag tcggaatcct gtgctccgcc ctgcagaaaa tcaaatga	1908

<210> 52
 <211> 2156
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 52	
ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac acacagccct ccctgcctgc	60
tgaccttgga gctggggcag aggtcagaga cctctctggg cccatgccac ctccaacatc	120
cactcgaccc cttggaatct cggtggagag gagcagaggt tgtcctggcg tggtttaggt	180
agtgtgagag gggaaatgact cctttcggtg agtgcagtgg aagctgtaca ctgcccaggc	240
aaagcgtccg ggacagcgtg gcggggcact cagatcccag ccagtggact tagcccctgt	300
ttgctcctcc gataactggg gtgaccttgg ttaatatcca ccagcagcct cccccgttgc	360
ccctctggat ccactgctta aatacggacg aggacagggc cctgtctcct cagcttcagg	420
caccaccact gacctgggac agtgaatcct ctaaggtaaa tataaaatct 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 ccacactgga taagcggctc ctgccagccc tgacaaacat catcaagatc	840
ctgaggcacg acatcgagac aaccgtgcac gagctgtctc gggacaagaa gaaggatacc	900
gtgccctggt tccctcggac aatccaggag ctggatagat ttgccaacca gatcctgtct	960
tacggagcag agctggacgc agatcacctt ggcttcaagg acccagtgtg tcggggcccg	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 gagtgcttca ctgcacccag tacatcaggc acggctcaa gccaatgtat	1380
acaccagagc ccgacatctg tcacgagctg ctggggccac tgcccctgtt tagcgataga	1440

tccttcgccc agttttccca ggagatcgga ctggcatctc tgggagcacc tgacgagtac	1500
atcgagaagc tggccaccat ctattggttc acagtggagt ttggcctgtg caagcagggc	1560
gatagcatca aggcctacgg agcaggactg ctgtctagct tcggcgagct gcagtattgt	1620
ctgtccgaga agccaaagct gctgcccctg gagctggaga agaccgcat ccagaactac	1680
accgtgacag agttccagcc cctgtactat gtggccgagt cttttaacga tgccaaggag	1740
aaggtgagaa atttcgccgc cacaatccct aggcccttca gtgtgctgta cgacccttat	1800
acccagagga tcgagggtgct ggataataca cagcagctga agatcctggc tgactcaatc	1860
aatagcgaag tcggaatcct gtgctccgcc ctgcagaaaa tcaaatgaga attcaaggcc	1920
tctcgagcct ctagaactat agtgagtcgt attacgtaga tccagacatg ataagataca	1980
ttgatgagtt tggacaaacc acaactagaa tgcagtgaag aaaatgcttt atttgtgaaa	2040
tttgtgatgc tattgcttta tttgtaacca ttataagctg caataaaca gttacaaca	2100
acaattgcat tcattttatg tttcaggttc agggggaggt gtgggaggtt ttttaa	2156

<210> 53
 <211> 1356
 <212> DNA
 <213> Homo sapiens

<400> 53	
atgtccactg cggtcctgga aaaccaggc ttgggcagga aactctctga ctttggacag	60
gaaacaagct atattgaaga caactgcaat caaatgggtg ccatatcact gatcttctca	120
ctcaaagaag aagttgggtg attggccaaa gtattgcgct tatttgagga gaatgatgta	180
aacctgacct acattgaatc tagaccttct cgtttaaaaga aagatgagta tgaatttttc	240
acccatttgg ataaacgtag cctgcctgct ctgacaaaca tcatcaagat cttgaggcat	300
gacattgggtg cactgttcca tgagctttca cgagataaga agaaagacac agtgccctgg	360
ttccaagaa ccattcaaga gctggacaga ttgccaatc agattctcag ctatggagcg	420
gaactggatg ctgaccacct tggttttaaa gatcctgtgt accgtgcaag acggaagcag	480
tttctgaca ttgcctacaa ctaccgcat 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 taccgccgaa	840
cctgacatct gccatgagct gttgggacat gtgcccttgt tttcagatcg cagctttgcc	900
cagttttccc aggaaattgg ccttgcctct ctgggtgcac ctgatgaata cattgaaaag	960
ctcgccacaa ttactgggtt tactgtggag tttgggtctt gcaacaagg agactccata	1020
aaggcatatg gtgctgggct cctgtcatcc tttggtgaat tacagtactg cttatcagag	1080
aagccaaagc ttctccccct ggagctggag aagacagcca tccaaaatta cactgtcacg	1140

gagttccagc ccctgtatta cgtggcagag agttttaatg atgccaagga gaaagtaagg	1200
aacttttgctg ccacaatacc tcggcccttc tcagttcgct acgaccata cacccaaagg	1260
attgaggctt tggacaatac ccagcagctt aagattttgg ctgattccat taacagtga	1320
attggaatcc tttgcagtgc cctccagaaa ataaag	1356

<210> 54
 <211> 173
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 54	
actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc	60
aaaatcaacg ggactttcca aaatgtcgtg ataaccccg cccgttgacg caaatgggcg	120
gtaggcgtgt acggtgggag gtctatataa gcagagctcg tttagtgaac cgt	173

<210> 55
 <211> 100
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 55	
cmetaatcaa cgggactttc metaatgtcg taataacccc gccccgttga cgmetaatggg	60
cggtaggcgt gtacgggtgg aggtctatat aagcagagct	100

<210> 56
 <211> 380
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 56	
ggcattgatt attgactagt tattaatagt aatcaattac ggggtcatta gttcatagcc	60
catatatgga gttccgcgtt acataactta cggtaaattg cccgcctggc tgaccgcca	120
acgacccccg ccattgacg tcaataatga cgtatgttcc catagtaacg ccaataggga	180
ctttccattg acgtcaatgg gtggagtatt tacggtaaac tgcccacttg gcagtacatc	240
aagtgtatca tatgccaagt ccgcccccta ttgacgtcaa tgacggtaaa tggccgcct	300
ggcattatgc ccagtacatg accttacggg actttcctac ttggcagtac atctacgtat	360
tagtcatcgc tattaccatg	380

<210> 57
 <211> 1246
 <212> DNA
 <213> Artificial Sequence

<220>

<223> Synthetic polynucleotide

<400> 57

tcgaggtgag cccacagttc tgcttcactc tccccatctc cccccctcc ccacccccaa	60
ttttgtatattt atttattttt taattatttt gtgcagcgat gggggcgggg gggggggggg	120
ggcgcgcgcc aggcggggcg gggcggggcg aggggcgggg cggggcgagg cggagagggtg	180
cggcggcagc caatcagagc ggcgcgctcc gaaagtttcc ttttatggcg aggcggcggc	240
ggcgggcgcc ctataaaaag cgaagcgcg cggcgggcggg agtcgctgcg cgctgccttc	300
gccccgtgcc ccgctccgcc gccgcctcgc gccgcccgcc ccggctctga ctgaccgcgt	360
tactcccaca ggtgagcggg cgggacggcc cttctcctcc gggctgtaat tagcgcttgg	420
tttaatgacg gcttgtttct tttctgtggc tgcgtgaaag ccttgagggg ctccgggagg	480
gccctttgtg cggggggagc ggctcggggg gtgcgtgcgt gtgtgtgtgc gtggggagcg	540
ccgcgtgcgg ctccgcgctg cccggcggt gtgagcgctg cgggcgcggc gcggggcttt	600
gtgcgctccg cagtgtgcgc gaggggagcg cggccggggg cggtgccccg cggtgcgggg	660
ggggctgcga ggggaacaaa ggctgcgtgc ggggtgtgtg cgtggggggg tgagcagggg	720
gtgtgggcgc gtcggtcggg ctgcaacccc ccctgcaccc ccctccccga gttgctgagc	780
acggcccggc ttcgggtgcg gggctccgta cggggcgctg cgcggggctc gccgtgccgg	840
gcgggggggtg gcggcaggtg ggggtgccgg gcggggcggg gccgcctcgg gccggggagg	900
gctcggggga gggcgcgggc gggccccgga gcgcggcggg ctgtcgaggc gcggcgagcc	960
gcagccattg ctttttatgg taatcgtgcg agagggcga gggacttcct ttgtcccaaa	1020
tctgtgcgga gccgaaatct gggaggcgcc gccgcacccc ctctagcggg cgcggggcga	1080
agcgggtcgg cgccggcagg aaggaaatgg gcggggaggg ctttcgtgcg tcgccgcgcc	1140
gccgtcccct tctccctctc cagcctcggg gctgtccgcg gggggacggc tgccttcggg	1200
ggggacgggg cagggcgggg ttccggcttct ggcgtgtgac cggcgg	1246

<210> 58

<211> 201

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide

<400> 58

gaattcgggc ggagttaggc cggagccaat cagcgtgcgc cgttccgaaa gttgcctttt	60
atggctgggc ggagaatggg cggatgaacgc cgatgattat ataaggacgc gccgggtgtg	120
gcacagctag ttccgtcgca gccgggattt gggtcgcggt tcttgtttgt ggatccctgt	180
gatcgtgatc atcacttgtg a	201

<210> 59

<211> 953

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide

<400> 59

aattcggtag	cctagttatt	aatagtaatc	aattacgggg	tcattagttc	atagcccata	60
tatggagttc	cgcggttacat	aacttacggt	aaatggcccc	cctggctgac	cgcccaacga	120
ccccgcccc	ttgacgtcaa	taatgacgta	tgttcccata	gtaacgcaa	tagggacttt	180
ccattgacgt	caatgggtgg	actattttacg	gtaaactgcc	cacttggcag	tacatcaagt	240
gtatcatatg	ccaagtacgc	cccctattga	cgtaaatgac	ggtaaattggc	ccgcctggca	300
ttatgcccag	tacatgacct	tatgggactt	tcctacttgg	cagtacatct	acgtattagt	360
catcgctatt	accatggtcg	aggtgagccc	cacgttctgc	ttcactctcc	ccatctcccc	420
cccctcccca	cccccaattt	tgtattttatt	tatttttttaa	ttattttgtg	cagcgatggg	480
ggcggggggg	gggggggggc	gcgcgccagg	cggggcgggg	cggggcgagg	ggcggggcgg	540
ggcgaggcgg	agagggtcgg	ggcgagccaa	tcagagcggc	gcgctccgaa	agtttccttt	600
tatggcgagg	ggcgggcggc	ggcgggccta	taaaaagcga	agcgcgcggc	ggcggggagt	660
cgctgcgacg	ctgccttcgc	cccgtgcccc	gctccgccgc	cgcttcgcgc	cgcccgcccc	720
ggctctgact	gaccgcgtta	ctcccacagg	tgagcgggcg	ggacggccct	tctcctccgg	780
gctgtaatta	gcgcttggtt	taatgacggc	ttgtttcttt	tctgtggctg	cgtgaaagcc	840
ttgaggggct	ccgggagcta	gagcctctgc	taaccatgtt	catgccttct	tctttttcct	900
acagctcctg	ggcaacgtgc	tggttattgt	gctgtctcat	cattttggca	aag	953

<210> 60

<211> 1766

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide

<400> 60

gtaaaatttta	tggaatgtga	atcataattc	aatttttcaa	catgcgttag	gagggacatt	60
tcaaactctt	ttttacccta	gactttccta	ccatcaccca	gagtatccag	ccaggagggg	120
aggggctaga	gacaccagaa	gttttagcagg	gaggagggcg	tagggattcg	gggaatgaag	180
ggatgggatt	cagactaggg	ccaggaccca	gggatggaga	gaaagagatg	agagtggttt	240
gggggcttgg	tgacttagag	aacagagctg	caggctcaga	ggcacacagg	agtttctggg	300
ctcaccctgc	ccccttccaa	cccctcagtt	cccacctcc	agcagctgtt	tgtgtgctgc	360
ctctgaagtc	cacactgaac	aaacttcagc	ctactcatgt	ccctaaaatg	ggcaaacatt	420
gcaagcagca	aacagcaaac	acacagccct	ccctgcctgc	tgaccttgga	gctggggcag	480
aggtcagaga	cctctctggg	cccatgccac	ctccaacatc	cactcgaccc	cttggaattt	540
cgggtggagag	gagcagaggt	tgtcctggcg	tggttttaggt	agtgtgagag	ggcttaagcg	600
tgaggctccg	gtgcccgtca	gtgggcagag	gcacatcg	ccacagtccc	cgagaagttg	660

gggggagggg tcggcaattg aaccggtgcc tagagaaggt ggcgcggggt aaactgggaa	720
agtgatgtcg tgtactggct ccgccttttt cccgaggggtg ggggagaacc gtatataagt	780
gcagtagtcg ccgtgaacgt tctttttcgc aacgggtttg ccgccagaac acaggtaaagt	840
gccgtgtgtg gttcccgcg gcctggcctc tttacgggtt atggcccttg cgtgccttga	900
attacttcca cctggctcca gtacgtgatt cttgatcccg agctggagcc aggggcgggc	960
cttgcgcttt aggagcccct tcgcctcgtg cttgagttga ggcctggcct gggcgctggg	1020
gccgccgct 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 ggcctgtctt ggtgcctggc ctcgcgccgc	1320
cgtgtatcgc cccgccctgg gcggcaaggc tggcccggtc ggcaccagtt gcgtgagcgg	1380
aaagatggcc gcttcccggc cctgctccag ggggctcaaa atggaggacg cggcgctcgg	1440
gagagcgggc gggtagtca cccacacaaa ggaaaggggc ctttccgtcc tcagccgtcg	1500
cttcatgtga ctccacggag taccgggcgc cgtccaggca cctcgattag ttctggagct	1560
tttgagtagc gtcgtcttta ggttgggggg aggggtttta tgcgatggag tttccccaca	1620
ctgagtgggt ggagactgaa gttaggccag cttggcactt gatgtaattc tccttggaat	1680
ttgccctttt tgagtttga tcttggttca ttctcaagcc tcagacagtg gttcaaagtt	1740
tttttcttcc atttcaggtg tcgtga	1766

<210> 61
 <211> 82
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 61	
gtaaggggtt aagggatggt tggttggtgg ggtattaatg tttaattacc tggagcacct	60
gcctgaaatc actttttttc ag	82

<210> 62
 <211> 95
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 62	
cctctgctaa ccatgttcat gccttcttct ttttctaca gctcctgggc aacgtgctgg	60
ttattgtgct gtctcatcat tttggcaaag aattc	95

<210> 63
 <211> 1356

<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 63
atgtccaccg ctgtgctgga gaaccctggg ctggggagga aactgtcaga cttcgggcag 60
gagacttcat acattgagga taactgtaac cagaatggcg ccatctctct gatcttcagc 120
ctgaaggagg aagtgggctg cctggcaaag gtgctgcgcc tgtttgagga gaacgacgtg 180
aatctgaccc acatcgagtc ccggccttct agactgaaga aggacgagta cgagttcttt 240
accacacttg ataagcggtc cctgccagcc ctgacaaaca tcatcaagat cctgaggcac 300
gacatcggag caaccgtgca cgagctgtct cgggacaaga agaaggatac cgtgccctgg 360
ttccctcgga caatccagga gctggataga tttgccaacc agatcctgtc ttacggagca 420
gagctggacg cagatcaccg tggcttcaag gaccagtgat atcgggcccc gagaaagcag 480
tttgccgata tcgcctacaa ttataggcac ggacagccaa tccctcgcgt ggagtatatg 540
gaggaggaga agaagacctg gggcacagtg ttcaagacct tgaagagcct gtacaagaca 600
cacgcctgct acgagtataa ccacatcttc cccctgctgg agaagtattg tggctttcac 660
gaggacaata tccctcagct ggaggacgtg agccagttcc tgcagacctg cacaggcttt 720
aggctgaggc cagtggcagg actgctgagc tcccgggact tcctgggagg actggccttc 780
agagtgtttc actgcaccca gtacatcagg cacggctcca agccaatgta tacaccagag 840
cccgacatct gtcacgagct gctgggccac gtgcccctgt ttagcgatag atccttcgcc 900
cagttttccc aggagatcgg actggcatct ctgggagcac ctgacgagta catcgagaag 960
ctggccacca tctattgggt cacagtggag tttggcctgt gcaagcaggg cgatagcatc 1020
aaggcctacg gagcaggact gctgtctagc ttcggcgagc tgcagtattg tctgtccgag 1080
aagccaaagc tgctgcccct ggagctggag aagaccgcca tccagaacta caccgtgaca 1140
gagttccagc ccctgtacta tgtggccgag tcttttaacg atgccaagga gaaggtgaga 1200
aatttcgccg ccacaatccc taggcccttc agcgtgcggt acgaccctta taccagagg 1260
atcgaggtgc tggataatac acagcagctg aagatcctgg ctgactcaat caatagcgaa 1320
atcggaatcc tgtgctccgc cctgcagaaa atcaaa 1356

<210> 64
<211> 1873
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 64
gatcttcaat attggccatt agccatatta ttcatctggt atatagcata aatcaatatt 60
ggctattggc cattgcatac gttgtatcta tatcataata tgtacattta tattggctca 120
tgtccaatat gaccgccatg ttggcattga ttattgacta gttattaata gtaatcaatt 180

acggggtcat tagttcatag cccatatatg gagttccgcg ttacataact tacggtaa	240
ggcccgccctg gctgaccgcc caacgacccc cgcccattga cgtcaataat gacgtatgtt	300
cccatagtaa cgccaatagg gacttttccat tgacgtcaat ggggtggagta tttacggtaa	360
actgcccact tggcagtaca tcaagtgtat catatgccaa gtccgcccc tattgacgtc	420
aatgacggta aatggcccg cttggcattat gcccagtaca tgaccttacg ggactttcct	480
acttggcagt acatctacgt attagtcacg gctattacca tggtcgaggt gagccccacg	540
ttctgcttca ctctcccat ctccccccc tccccacccc caattttgta tttatttatt	600
ttttaattat tttgtgcagc gatggggcg gggggggggg gggggcgcg gccaggcggg	660
gcggggcggg gcgagggcg gggcggggcg aggcggagag gtgcggcggc agccaatcag	720
agcggcgcg tccgaaagt tctttttatg gcgaggcggc ggcgcgcggc gccctataaa	780
aagcgaagcg cgcgcgggc gggagtcgct gcgacgctgc cttcgcccc tgccccgctc	840
cgcccgccgc tcgcgccgc cgccccgct ctgactgacc gcgttactcc cacaggtag	900
cgggcgggac ggcccttct ctccgggctg taattagcg ttggtttaat gacggcttgt	960
ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg	1020
ggagcggctc ggggggtgcg tgcgtgtgtg tgtgcgtggg gagcgccgcg tgcggccgc	1080
gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgcagtgt	1140
gcgcgagggg agcgcggcgc ggggcgggtgc ccgcgggtgc ggggggggct gcgaggggaa	1200
caaaggctgc gtgcggggtg tgtgcgtggg ggggtgagca ggggtgtgg gcgcggcggt	1260
cgggctgtaa cccccctg cccccctc ccgagttgc tgagcacggc ccggcttcgg	1320
gtgcggggct ccgtacgggg cgtggcgcg ggctcgccgt gccgggcggg ggggtggcggc	1380
agggtgggggt gccgggcggg gcggggccgc ctcgggcccgg ggagggctcg ggggaggggc	1440
gcggcggccc ccggagcgcc ggcggtgtc gaggcgcggc gagccgcagc cattgccttt	1500
tatggtaatc gtgcgagagg gcgcagggac ttctttgtc ccaaacttgt gcggagccga	1560
aatctgggag gcgcccgcc acccctcta gcgggcgcgg ggcgaagcgg tgcggcgccg	1620
gcaggaagga aatgggcggg gagggccttc gtgcgtcgcc gcgccgccgt ccccttctcc	1680
ctctccagcc tcggggctgt ccgcgggggg acggctgcct tcggggggga cggggcaggg	1740
cggggttcgg cttctggcgt gtgaccggcg gctctagagc ctctgctaac catgttcatg	1800
ccttcttctt tttctacag ctctgggca acgtgctggt tattgtgctg tctcatcatt	1860
ttggcaaaga att	1873

<210> 65
 <211> 1061
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 65	
tagggaggtc ctgcacgtta cataacttac ggtaaattgc ccgcctggct gaccgccc	60

cgacccccgc ccattgacgt caataatgac gtatgttccc atagtaacgc caatagggac	120
tttccattga cgtcaatggg tggagtatgt acggtaaact gccacttgg cagtacatca	180
agtgtatcat atgccaagta cgccccctat tgacgtcaat gacggtaaata ggccccctg	240
gcattatgcc cagtacatga ctttatggga ctttctact tggcagtaca tctacgtatt	300
agtcacgct attaccatgg tcgaggtgag cccacgctt tgcttcactc tccccatctc	360
ccccccctcc ccacccccaa ttttgtatgt atttattttt taattatttt gtgcagcgat	420
gggggcgggg gggggggggg gcgcgcgcc ggcgggggcg ggcgggggcg ggggcggggc	480
ggggcgaggc ggagaggtgc ggcggcagcc aatcagagcg gcgcgctccg aaagtttcct	540
tttatggcga ggcggcggcg gcggcgggcc tataaaaagc gaagcgcgcg gcgggcggga	600
gtcgtcgcg gctgccttcg ccccggtccc cgctccgcc cgccctcgc cgccccgcc	660
cggctctgac tgaccgcgtt actaaaacag gtaagtccgg cctccgcgc gggttttggc	720
gcctccgcg ggcgcccc tcctcacggc gagcgtgccc acgtcagacg aaggcgcgag	780
cgagcgtcct gatccttccg cccggacgct caggacagcg gcccgctgct cataagactc	840
ggccttagaa cccagtatc agcagaagga catttttagga cgggacttgg gtgactctag	900
ggcactgggt ttctttccag agagcggaac aggcgaggaa aagtagtccc ttctcggcga	960
ttctgcggag ggatctccgt ggggcggtga acgccgatga tgcctctact aaccatgttc	1020
atgttttctt ttttttcta caggctctgg gtgacgaaca g	1061

<210> 66
 <211> 170
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 66	
cgatgctcta atctctctag acaaggttca tatttgtatg ggttacttat tctctctttg	60
ttgactaagt caataatcag aatcagcagg tttgcagtca gattggcagg gataagcagc	120
ctagctcagg agaagtgagt ataaaagccc caggctggga gcagccatca	170

<210> 67
 <211> 254
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 67	
gggggaggct gctggtgaat attaaccaag gtcacccag ttatcggagg agcaaacagg	60
ggctaagtcc acctcgagcc atggcgatgc tctaattctt ctagacaagg ttcatatttg	120
tatgggttac ttattctctc tttgttgact aagtcaataa tcagaatcag caggtttgca	180
gtcagattgg cagggataag cagcctagct caggagaagt gagtataaaa gccccaggct	240

<210> 68
<211> 592
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 68
gtaaatttta tggaatgtga atcataattc aatTTTTTcaa 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 ccccttcaa 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 gg 592

<210> 69
<211> 205
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 69
aatgactcct ttcggtaagt gcagtgaag ctgtacactg cccaggcaaa gcgtccgggc 60
agcgtaggcg ggcgactcag atcccagcca gtggacttag cccctgtttg ctctccgat 120
aactggggtg accttggtta atattcacca gcagcctccc ccgttgcccc tctggatcca 180
ctgcttaa at acggacgagg acagg 205

<210> 70
<211> 423
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 70
gctctaacc actctgatct cccagggcgg cagtaagtct tcagcatcag gcattttggg 60
gtgactcagt aaatggtaga tcttgctacc agtggacag ccactaagga ttctgcagtg 120
agagcagagg gccagctaag tgggtactctc ccagagactg tctgactcac gccacccct 180
ccaccttgga cacaggacgc tgtggtttct gagccaggta caatgactcc tttcggttaag 240
tgagtgga gctgtacact gccaggcaa agcgtccggg cagcgtaggc gggcgactca 300

gatcccagcc agtggactta gcccctgttt gctcctccga taactggggt gaccttggtt	360
aatattcacc agcagcctcc cccgttgccc ctctggatcc actgcttaaa tacggacgag	420
gac	423

<210> 71
 <211> 72
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 71	
gggggaggct gctggtgaat attaaccaag gtcacccag ttatcggagg agcaaacagg	60
ggctaagtcc ac	72

<210> 72
 <211> 913
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 72	
tagggaggtc ctgcacagaa ggggaggagg gggcagcagc tgtctgacca ctgttggtct	60
tgcaacttgt gtccccaggt taatttttaa aaagcagtca aaagtccaag tggcccttgg	120
cagcatttac tctctctggt tgctctggtt aataatctca ggagcacaaa cattcctgga	180
ggcaggagaa gaaatcaaca tcctggactt atcctctggg cctctcccca ccccaggat	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 agtatatgtt	480
tagccctata gctctaacc actctgatct cccagggcgg cagtaagtct tcagcatcag	540
gcattttggg gtgactcagt aaatggtaga tcttgctacc agtggaacag ccactaagga	600
ttctgcagtg agagcagagg gccagctaag tggactctc ccagagactg tctgactcac	660
gccacccctt ccaccttgga cacaggacgc tgtggtttct gagccaggta caatgactcc	720
tttcggtaag tgcaagtggaa gctgtacact gccaggcaa agcgtccggg cagcgtaggc	780
gggcgactca gatcccagcc agtggactta gcccctgttt gctcctccga taactggggt	840
gaccttggtt aatattcacc agcagcctcc cccgttgccc ctctggatcc actgcttaaa	900
tacggacgag gac	913

<210> 73
 <211> 37
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

 <400> 73
 gtagataagt agcatggcgg gttaatcatt aactaca 37

 <210> 74
 <211> 46
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic polynucleotide

 <400> 74
 ggaggggtgg agtcgtgacg tgaattacgt catagggtta gggagg 46

 <210> 75
 <211> 182
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic polynucleotide

 <400> 75
 gtagataagt agcatggcgg gttaatcatt aactacaagg aaccctagt gatggagttg 60
 gccactccct ctctgcgcgc tcgctcgtc actgaggccg ggcgaccaa ggtcgcccga 120
 cgcccgggct ttgcccgggc ggcctcagtg agcgagcgag cgcgcagaga gggagtggcc 180
 aa 182

 <210> 76
 <211> 191
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic polynucleotide

 <400> 76
 ttggccactc cctctctgcg cgctcgtcgc ctactgagg ccgggcgacc aaaggtcgcc 60
 cgacgcccgg gctttgccc ggcggcctca gtgagcgagc gagcgcgag agagggagtg 120
 gccaaactcca tctactagggg ttcttgagg ggtggagtgc tgacgtgaat tacgtcatag 180
 ggttagggag g 191

 <210> 77
 <211> 3230
 <212> DNA
 <213> Artificial Sequence

 <220>
 <223> Synthetic polynucleotide

 <400> 77
 agcattagct tccatttatg cagtgtaaat ggtgagaaca gccccgactg aataaccaga 60
 gcatcatctc gtctgtgtca ttcatgcaca taacatatct cagcgagggt gcccttctgt 120
 cctctttgca gagaccagc caccatacta gtacctagag aactggctgg atttcagccc 180

cgataacctcc gggcttttgc tcatgttcgc ctcatagggt catctgggtg gttgcctaag	240
gaaaagtatg tcatggagac taacttgctt ggcatagaat aaaaggtagg ttgagagtgg	300
agcgtgttta aattgcaatc ctgcctctat ttctgtgctt gcagggaaca gtcaccta	360
attgctatcc tccatcatca tcatgattat ttctggtttt tctctggttg cggagaatcc	420
atactccagg tattccaatg tctcagcatt gccaggcctg tctgagcgtc aggatgtagg	480
tagtctgggc tctctgcctt ctattcttgt ccaggatact ctgccaaaag aatcatgttg	540
tggctgccac cctcccaca aagcctcccg ctggggtcag tccaggactg gagttgggta	600
tggactgttc atgtctatcc actgctacgt cagggaaca cccactgaga gtgaccttgt	660
agactgcagt gggagacacc cttcaaaacc tctcctctcc tgtcctgaga gccaggta	720
aacctcagc cccgcatcct gaggcaaac ttttcctaac cctgctgcta agctagacac	780
ctcacttact gagagccagc atgtccaccg ctgtgctgga gaaccctggg ctggggagga	840
aactgtcaga cttcgggcag gagacttcat acattgagga taactgtaac cagaatggcg	900
ccatctctct gatcttcagc ctgaaggagg aagtgggcgc cctggcaaag gtgctgcgcc	960
tgtttgagga gaacgacgtg aatctgacct acatcgagtc ccggccttct agactgaaga	1020
aggacgagta cgagttcttt acccacctgg ataagcggtc cctgccagcc ctgacaaaca	1080
tcatcaagat cctgaggcac gacatcggag caaccgtgca cgagctgtct cgggacaaga	1140
agaaggatac cgtgccctgg ttccctcgga caatccagga gctggataga tttgccaacc	1200
agatcctgtc ttacggagca gagctggacg cagatcacc tggcttcaag gaccagtg	1260
atcgggcccg gagaaagcag tttgccgata tcgcctacaa ttataggcac ggacagccaa	1320
tccctcgcgt ggagtatatg gaggaggaga agaagacctg gggcacagt tcaagacct	1380
tgaagagcct gtacaagaca cacgcctgct acgagtataa ccacatcttc cccctgctgg	1440
agaagtattg tggctttcac gaggacaata tccctcagct ggaggacgtg agccagttcc	1500
tgcagacctg cacaggcttt aggctgaggc cagtggcagg actgctgagc tcccgggact	1560
tcctgggagg actggccttc agagtgtttc actgcacca gtacatcagg cacggctcca	1620
agccaatgta tacaccagag ccgacatct gtcacgagct gctgggccac gtgcccctgt	1680
ttagcgatag atccttcgcc cagttttccc aggagatcgg actggcatct ctgggagcac	1740
ctgacgagta catcgagaag ctggccacca tctattggtt cacagtggag tttggcctgt	1800
gcaagcaggg cgatagcatc aaggcctacg gagcaggact gctgtctagc ttcggcgagc	1860
tgcagtattg tctgtccgag aagccaaagc tgctgcccct ggagctggag aagaccgcca	1920
tccagaacta caccgtgaca gagttccagc ccctgtacta tgtggccgag tcttttaacg	1980
atgccaagga gaaggtaga aatttcgccg ccacaatccc taggccttc agcgtgcggt	2040
acgaccctta taccagagg atcgagggtc tggataatac acagcagctg aagatcctgg	2100
ctgactcaat caatagcgaa atcggaatcc tgtgctccgc cctgcagaaa atcaaatgag	2160
aattcaaggc ctctcgagcc tctagaacta tagtgagtcg tattacgtag atccagacat	2220

gataagatac attgatgagt ttggacaaac cacaactaga atgcagtgaa aaaaatgctt	2280
tatttgtgaa atttgtgatg ctattgcttt atttgaacc attataagct gcaataaaca	2340
agttaacaac aacaattgca ttcattttat gtttcaggtt cagggggagg tgtgggaggt	2400
tttttaagct ttacgtacga tcgtcgagca gctgttgtcc tggagaacgg agtcctgagc	2460
agaaaactct cagactttgg gcaggtaagc ctgttgggct tccactgcta ggagagaatt	2520
ggttccccac atgtgaaagc agtctgggaa atgctggtat ttccagtctc ctaaggctac	2580
taagaaatat gactttattt agaggcgagg aaaatgccca ggaagtcaac tgatgagact	2640
agtcttaaca agttgaggat acagaaagtt ggggatctga gctgctacca acatctgtgt	2700
gtctttgggt ggctcattgg tatcctctgc ctattggctt tatcttctgt aactgaaag	2760
gaaatggctg gtccttagtc acctgggggtg ggagtcccta tctctccagg gatacttatt	2820
caatcctttc ttctgggtat caaatgaca agcttgtaag aaactgtcct ctttcggctt	2880
tcaggaggtg atgtcgcag aagagaattt gggggggggg acttactcag aaccaaggag	2940
ggagaaatta aacagagagg gaaatgaaca ggagttagcc cggagcctga agcaccttgg	3000
ggattatgct gggggtggag ggaatccatt gtcctcccta gggagggtt gcagaacatg	3060
ttcttttctg tgatatttgt actttcccca gattgcaa atcatggtttgt aactgagat	3120
tcagtctctg gaggtaatat gccttttcta gcttttcctt ggacaggact aaggggttga	3180
gggttgcctg gagtcagaga aattaactgt gttaaagaag gttgatatga	3230

<210> 78
 <211> 3637
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 78	
ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggtcgcc	60
cgacgcccgg gctttgcccg ggccggcctca gtgagcgagc gagcgcgag agaggagtg	120
gccaactcca tctactagggg ttcttgaggg ggtggagtcg tgacgtgaat tacgtcatag	180
ggttagggag gtcctgcata tgcggccgca gcattagctt ccatttatgc agtgtaaatg	240
gtgagaacag ccccgactga ataccagag catcatctcg tctgtgtcat tcatgcacat	300
aacatatctc agcgagggtg cccttctgtc ctctttgcag agaccagcc accatactag	360
tacctagaga actggctgga tttcagcccc gatacctccg ggcttttgct catgttcgcc	420
tcatagggtc atctgggtgg ttgcctaagg aaaagtatgt catggagact aacttgcttg	480
gcattgaata aaaggtgagt tgagagtgga gcgtgtttaa attgcaatcc tgcctctatt	540
tctgtgcttg caggaacag tcatccttaa ttgctatcct ccatcatcat catgattatt	600
tctggttttt ctctggttgc ggagaatcca tactccagggt attccaatgt ctcagcattg	660
ccaggcctgt ctgagcgta ggatgtaggt agtctgggct ctctgccttc tattcttgtc	720
caggatactc tgccaaaaga atcatgttgt ggctgccacc cctccacaa agcctccgc	780

ttgggtcagt ccaggactgg agttgggtat ggactgttca tgtctatcca ctgctacgtc	840
agggcaacac cactgagag tgaccttgta gactgcagtg ggagacaccc ttcaaacct	900
ctcctctcct gtcctgagag ccagggttaaa accatcagcc ccgcatcctg agtgcaaaact	960
tttcctaacc ctgctgctaa gctagacacc tcacttactg agagccagca tgtccaccgc	1020
tgtgctggag aaccctgggc tggggaggaa actgtcagac ttcgggcagg agacttcata	1080
cattgaggat aactgtaacc agaatggcgc catctctctg atcttcagcc tgaaggagga	1140
agtgggcgcc ctggcaaagg tgctgcgcct gtttgaggag aacgacgtga atctgacca	1200
catcgagtcc cggccttcta gactgaagaa ggacgagtac gaggttctta cccacctgga	1260
taagcgggcc ctgccagccc tgacaaacat catcaagatc ctgaggcacg acatcggagc	1320
aaccgtgcac gagctgtctc gggacaagaa gaaggatacc gtgccctggc tccctcggac	1380
aatccaggag ctggatagat ttgccaacca gatcctgtct tacggagcag agctggacgc	1440
agatcacctt ggcttcaagg acccagtgtg tcggggcccg agaaagcagt ttgccgat	1500
cgcctacaat tataggcacg gacagccaat ccctcgcgtg gagtatatgg aggaggagaa	1560
gaagacctgg ggcacagtgt tcaagacctt gaagagcctg tacaagacac acgcctgcta	1620
cgagtataac cacatcttcc ccctgctgga gaagtattgt ggctttcacg aggacaatat	1680
ccctcagctg gaggacgtga gccagttcct gcagacctgc acaggcttta ggctgaggcc	1740
agtggcagga ctgctgagct cccgggactt cctgggagga ctggccttca gagtgtttca	1800
ctgcacccag tacatcaggc acggctccaa gccaatgtat acaccagagc ccgacatctg	1860
tcacgagctg ctggggccacg tgcccctggt tagcgataga tccttcgccc agttttccca	1920
ggagatcgga ctggcatctc tgggagcacc tgacgagtac atcgagaagc tggccaccat	1980
ctattggttc acagtggagt ttggcctgtg caagcagggc gatagcatca aggcctacgg	2040
agcaggactg ctgtctagct tcggcgagct gcagtattgt ctgtccgaga agccaaagct	2100
gctgcccctg gagctggaga agaccgccat ccagaactac accgtgacag agttccagcc	2160
cctgtactat gtggccgagt cttttaacga tgccaaggag aagggtgagaa atttcgccgc	2220
cacaatccct agggccctta gcgtgcggta cgacccttat acccagagga tcgaggtgct	2280
ggataataca cagcagctga agatcctggc tgactcaatc aatagcgaaa tcggaatcct	2340
gtgctccgcc ctgcagaaaa tcaaatgaga attcaaggcc tctcgagcct ctagaactat	2400
agtgagtcgt attacgtaga tccagacatg ataagataca ttgatgagtt tggacaaacc	2460
acaactagaa tgcagtgaac aaaatgcttt atttgtgaaa tttgtgatgc tattgcttta	2520
tttgtaacca ttataagctg caataaaca gttacaaca acaattgcat tcattttatg	2580
tttcaggttc agggggaggt gtgggaggtt ttttaagctt tacgtacgat cgtcgagcag	2640
ctgttgtcct ggagaacgga gtcctgagca gaaaactctc agactttggg caggtaagcc	2700
tgttgggctt cactgctag gagagaattg gttccccaca tgtgaaagca gtctgggaaa	2760
tgctgggtatt tccagtctcc taaggctact aagaaatatg actttattta gaggcgagga	2820

aaatgcccag gaagtcaact gatgagacta gtcttaacaa gttgaggata cagaaagttg	2880
gggatctgag ctgctaccaa catctgtgtg tctttgggtg gctcattggt atcctctgcc	2940
tattggcttt atcttctgta cactgaaagg aaatggctgg tccttagtca cctggggtgg	3000
gagtcctat ctctccaggg atacttattc aatcctttct tctgggtatc aaaatgacaa	3060
gcttgaaga aactgtcctc tttcggcttt caggaggtga tgtcgcatga agagaatttg	3120
ggggggggga cttactcaga accaaggagg gagaaattaa acagagaggg aaatgaacag	3180
gagttagccc ggagcctgaa gcaccttggg gattatgctg ggggtggagg gaatccattg	3240
tcctccctag ggagggcttg cagaacatgt tcttttctgt gatatttgta ctttcccag	3300
attgcaaadc atggtttgta cactgagatt cagtctctgg aggtaatatg ctttttctag	3360
cttttccttg gacaggacta aggggttgag ggttgcctgg agtcagagaa atttgtgtta	3420
aagaaggttg atatgaaacc tgcaggctca gatacgtaga taagtagcat ggcggttaa	3480
tcattaacta caaggaaccc ctagtgatgg agttggccac tccctctctg cgcgctcgt	3540
cgctcactga ggccgggcga ccaaaggctg cccgacgccc gggctttgcc cgggcggcct	3600
cagtgagcga gcgagcgcg agagaggag tggccaa	3637

<210> 79
 <211> 3776
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 79	
agcattagct tccatttatg cagtgtaaat ggtgagaaca gccccgactg aatacccaga	60
gcatcatctc gtctgtgtca ttcattgcaca taacatatct cagcgagggt gcccttctgt	120
cctctttgca gagaccagc caccatacta gtacctagag aactggctgg atttcagccc	180
cgatacctcc gggcttttgc tcatgttcgc ctcatagggt catctgggtg gttgcctaag	240
gaaaagtatg tcatggagac taacttgctt ggcatggaat aaaaggtagg ttgagagtgg	300
agcgtgttta aattgcaatc ctgcctctat ttctgtgctt gcagggaaca gtcattccta	360
attgctatcc tccatcatca tcatgattat ttctggtttt tctctggttg cggagaatcc	420
atactccagg tattccaatg tctcagcatt gccaggcctg tctgagcgtc aggatgtagg	480
tagtctgggc tctctgcctt ctattcttgt ccaggatact ctgccaaaag aatcatgttg	540
tggctgccac ccctcccaca aagcctcccg cttgggtcag tccaggactg gagttgggta	600
tggactgttc atgtctatcc actgctacgt cagggaaca cccactgaga gtgacctgt	660
agactgcagt gggagacacc cttcaaaacc tctcctctcc tgtcctgaga gccaggtaa	720
aaccatcagc cccgcatcct gactgcaaac ttttcctaac cctgctgcta agctagacac	780
cccacttact gagagccagc ccctaaaatg ggcaaacatt gcaagcagca aacagcaaac	840
acacagccct ccctgcctgc tgaccttgga gctggggcag aggtcagaga cctctctggg	900
cccatgccac ctccaacatc cactcgacct cttggaattt cgggtggagag gagcagaggt	960

tgtcctggcg	tggttttaggt	agtgtgagag	gggaatgact	cctttcggta	agtgacgtgg	1020
aagctgtaca	ctgcccaggc	aaagcgtccg	ggcagcgtag	gcgggcgact	cagatcccag	1080
ccagtggact	tagcccctgt	ttgtctctcc	gataactggg	gtgaccttgg	ttaatatcca	1140
ccagcagcct	cccccgttgc	ccctctggat	ccactgctta	aatacggacg	aggacagggc	1200
cctgtctcct	cagcttcagg	caccaccact	gacctgggac	agtgaatcct	ctaaggtaaa	1260
tataaaatth	ttaagtgtat	aatgtgttaa	actactgatt	ctaattgttt	ctctctttta	1320
gattccaacc	tttggaaactg	accgccacca	tgtccaccgc	tgtgctggag	aaccctgggc	1380
tggggaggaa	actgtcagac	ttcgggcagg	agacttcata	cattgaggat	aactgtaacc	1440
agaatggcgc	catctctctg	atcttcagcc	tgaaggagga	agtgggcgcc	ctggcaaagg	1500
tgctgcgcct	gtttgaggag	aacgacgtga	atctgaccca	catcgagtcc	cggccttcta	1560
gactgaagaa	ggacgagtac	gagttcttta	cccacctgga	taagcgggtcc	ctgccagccc	1620
tgacaaacat	catcaagatc	ctgaggcacg	acatcggagc	aaccgtgcac	gagctgtctc	1680
gggacaagaa	gaaggatacc	gtgccctggt	tccctcggac	aatccaggag	ctggatagat	1740
ttgccaacca	gacctgtctt	tacggagcag	agctggacgc	agatcacctt	ggcttcaagg	1800
accagtgta	tcgggcccgg	agaaagcagt	ttgccgatat	cgcctacaat	tataggcacg	1860
gacagccaat	ccctcgcgtg	gagtatatgg	aggaggagaa	gaagacctgg	ggcacagtgt	1920
tcaagaccct	gaagagcctg	tacaagacac	acgcctgcta	cgagtataac	cacatcttcc	1980
ccctgctgga	gaagtattgt	ggctttcacg	aggacaatat	ccctcagctg	gaggacgtga	2040
gccagttcct	gcagacctgc	acaggcttta	ggctgaggcc	agtggcagga	ctgctgagct	2100
cccgggactt	cctgggagga	ctggccttca	gagtgtttca	ctgcaccag	tacatcaggc	2160
acggctccaa	gccaatgtat	acaccagagc	ccgacatctg	tcacgagctg	ctgggccacg	2220
tgcccctgtt	tagcgataga	tccttcgccc	agttttccca	ggagatcgga	ctggcatctc	2280
tgggagcacc	tgacgagtac	atcgagaagc	tggccaccat	ctattggttc	acagtggagt	2340
ttggcctgtg	caagcagggc	gatagcatca	aggcctacgg	agcaggactg	ctgtctagct	2400
tcggcgagct	gcagtattgt	ctgtccgaga	agccaaagct	gctgcccctg	gagctggaga	2460
agaccgccat	ccagaactac	accgtgacag	agttccagcc	cctgtactat	gtggccgagt	2520
cttttaacga	tgccaaggag	aagggtgagaa	atttcgccgc	cacaatccct	aggcccttca	2580
gtgtgcgtta	cgacccttat	accagagga	tcgaggtgct	ggataataca	cagcagctga	2640
agatcctggc	tgactcaatc	aatagcgaaa	tcggaatcct	gtgctccgcc	ctgcagaaaa	2700
tcaaatgaga	attcaaggcc	tctcgagcct	ctagaactat	agtgagtcgt	attacgtaga	2760
tccagacatg	ataagataca	ttgatgagtt	tggacaaacc	acaactagaa	tgcagtgaaa	2820
aaaatgcttt	atttgtgaaa	tttgtgatgc	tattgcttta	tttgtaacca	ttataagctg	2880
caataaacia	gttaacaaca	acaattgcat	tcattttatg	tttcaggttc	agggggaggt	2940
gtgggaggtt	ttttaagctt	tacgtacgat	cgctgagcag	ctgttgcctt	ggagaacgga	3000

gtcctgagca gaaaactctc agacttttggg caggtaagcc tgttgggctt ccactgctag	3060
gagagaattg gttccccaca tgtgaaagca gtctgggaaa tgctgggtatt tccagtctcc	3120
taaggctact aagaaatatg actttattta gaggcgagga aaatgccag gaagtcaact	3180
gatgagacta gtcttaacaa gttgaggata cagaaagttg gggatctgag ctgctaccaa	3240
catctgtgtg tctttgggtg gctcattggt atcctctgcc tattggcttt atcttctgta	3300
cactgaaagg aaatggctgg tccttagtca cctgggggtgg gagtcctat ctctccaggg	3360
atacttattc aatcctttct tctgggtatc aaaatgacaa gcttgtaaga aactgtcctc	3420
tttcggcttt caggaggtga tgtcgcatga agagaatttg ggggggggga cttactcaga	3480
accaaggagg gagaaattaa acagagaggg aaatgaacag gagttagccc ggagcctgaa	3540
gcaccttggg gattatgctg ggggtggagg gaatccattg tcctccctag ggagggcttg	3600
cagaacatgt tcttttctgt gatatttgta ctttccccag attgcaaadc atggtttgta	3660
cactgagatt cagtctctgg aggtaatatg ctttttctag cttttccttg gacaggacta	3720
aggggttgag ggttgccctgg agtcagagaa atttgtgtta aagaaggttg atatga	3776

<210> 80
 <211> 4186
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 80	
ttggccactc cctctctgcg cgctcgctcg ctactgagg ccgggcgacc aaaggtcgcc	60
cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg	120
gccaactcca tcactagggg ttcttgagg ggtggagtcg tgacgtgaat tacgtcatag	180
ggttagggag gtcctgcata tgcggccgca gcattagctt ccatttatgc agtgtaaag	240
gtgagaacag ccccgactga ataccagag catcatctcg tctgtgtcat tcatgcacat	300
aacatatctc agcgaggtgg cccttctgtc ctctttgcag agaccagcc accatactag	360
tacctagaga actggctgga ttccagcccc gatacctccg ggcttttgct catgttcgcc	420
tcatagggtc atctgggtgg ttgcctaagg aaaagtatgt catggagact aacttgcttg	480
gcattgaata aaaggtgagt tgagagtgga gcgtgtttta attgcaatcc tgcctctatt	540
tctgtgcttg cagggaacag tcaccttaa ttgctatcct ccatcatcat catgattatt	600
tctggttttt ctctggttgc ggagaatcca tactccaggt attccaatgt ctgagcattg	660
ccaggcctgt ctgagcgta ggatgtaggt agtctgggct ctctgccttc tattcttgct	720
caggatactc tgccaaaaga atcatgttgt ggctgccacc cctcccacaa agcctccgc	780
ttgggtcagt ccaggactgg agttgggtat ggactgttca tgtctatcca ctgctacgtc	840
agggcaacac cactgagag tgaccttgta gactgcagtg ggagacaccc ttcaaacct	900
ctcctctct gtctgagag ccagggtaaa accatcagcc ccgcatcctg agtgcaaac	960
tttcctaacc ctgctgctaa gctagacacc ccacttactg agagccagcc cctaaaatgg	1020

gcaaacattg caagcagcaa acagcaaaca cacagccctc cctgcctgct gaccttggag	1080
ctggggcaga ggtcagagac ctctctgggc ccatgccacc tccaacatcc actcgacccc	1140
ttggaatttc ggtggagagg agcagaggtt gtcctggcgt ggtttaggta gtgtgagagg	1200
ggaatgactc ctttcggtaa gtgcagtgga agctgtacac tgcccaggca aagcgtccgg	1260
gcagcgtagg cgggcgactc agatcccagc cagtggactt agcccctgtt tgctcctccg	1320
ataactgggg tgaccttggg taatattcac cagcagcctc ccccgttgcc cctctggatc	1380
cactgcttaa atacggacga ggacagggcc ctgtctcctc agcttcaggc accaccactg	1440
acctgggaca gtgaatcctc taaggtaaata ataaaatttt taagtgtata atgtgttaaa	1500
ctactgattc taattgtttc tctcttttag attccaacct ttggaactga ccgccaccat	1560
gtccaccgct gtgctggaga accctgggct ggggaggaaa ctgtcagact tcgggcagga	1620
gacttcatac attgaggata actgtaacca gaatggcgcc atctctctga tcttcagcct	1680
gaaggaggaa gtgggcgccc tggcaaaggt gctgcgcctg tttgaggaga acgacgtgaa	1740
tctgaccac atcgagtccc ggccttctag actgaagaag gacgagtacg agttctttac	1800
ccacctggat aagcgggtccc tgccagccct gacaaacatc atcaagatcc tgaggcacga	1860
catcggagca accgtgcacg agctgtctcg ggacaagaag aaggataccg tgccctgggt	1920
ccctcggaca atccaggagc tggatagatt tgccaaccag atcctgtctt acggagcaga	1980
gctggacgca gatcacctg gcttcaagga cccagtgtat cgggcccgga gaaagcagtt	2040
tgccgatatc gcctacaatt ataggcacgg acagccaatc cctcgcgtgg agtatatgga	2100
ggaggagaag aagacctggg gcacagtgtt caagaccctg aagagcctgt acaagacaca	2160
cgctgctac gagtataacc acatcttccc cctgctggag aagtattgtg gctttcacga	2220
ggacaatatc cctcagctgg aggacgtgag ccagttcctg cagacctgca caggctttag	2280
gctgaggcca gtggcaggac tgctgagctc ccgggacttc ctgggaggac tggccttcag	2340
agtgtttcac tgcacccagt acatcaggca cggctccaag ccaatgtata caccagagcc	2400
cgacatctgt cacgagctgc tgggccacgt gcccctgttt agcgatagat ccttcgcca	2460
gttttcccag gagatcggac tggcatctct gggagcacct gacgagtaca tcgagaagct	2520
ggccaccatc tattggttca cagtggagtt tggcctgtgc aagcagggcg atagcatcaa	2580
ggcctacgga gcaggactgc tgtctagctt cggcgagctg cagtattgtc tgtccgagaa	2640
gccaaagctg ctgcccctgg agctggagaa gaccgccatc cagaactaca ccgtgacaga	2700
gttccagccc ctgtactatg tggccgagtc ttttaacgat gccaaaggaga aggtgagaaa	2760
tttcgccgcc acaatcccta ggcccttcag tgtgcgttac gacccttata cccagaggat	2820
cgaggtgctg gataatacac agcagctgaa gatcctggct gactcaatca atagcgaaat	2880
cggaatcctg tgctccgcc tgcaaaaaat caaatgagaa ttcaaggcct ctcgagcctc	2940
tagaactata gtgagtcgta ttacgtagat ccagacatga taagatacat tgatgagttt	3000
ggacaaacca caactagaat gcagtgaaaa aaatgcttta tttgtgaaat ttgtgatgct	3060

attgctttat ttgtaacat tataagctgc aataaacaag ttaacaacaa caattgcatt	3120
cattttatgt ttcaggttca gggggagggtg tgggagggtt ttttaagcttt acgtacgatc	3180
gtcgcagcagc tgttgtcctg gagaacggag tcctgagcag aaaactctca gactttgggc	3240
aggtaagcct gttgggcttc cactgctagg agagaattgg ttccccacat gtgaaagcag	3300
tctgggaaat gctggtat tccagtctcct aaggctacta agaaatatga ctttatttag	3360
aggcgaggaa aatgcccagg aagtcaactg atgagactag tcttaacaag ttgaggatac	3420
agaaagttag ggatctgagc tgctaccaac atctgtgtgt ctttgggtgg ctcattggta	3480
tcctctgcct attggcttta tcttctgtac actgaaagga aatggctggc ccttagtcac	3540
ctgggggtggg agtccctatc tctccaggga tacttattca atcctttctt ctgggtatca	3600
aaatgacaag cttgtaagaa actgtcctct ttcggctttc aggagggtgat gtcgcatgaa	3660
gagaatttgg ggggggggac ttactcagaa ccaaggaggg agaaattaaa cagagaggga	3720
aatgaacagg agttagcccc gagcctgaag caccttgggg attatgctgg ggggtggaggg	3780
aatccattgt cctccctagg gagggcttgc agaacatgtt cttttctgtg atatttgtac	3840
tttccccaga ttgcaaatca tggtttgtac actgagattc agtctctgga ggtaatatgc	3900
cttttctagc ttttccttgg acaggactaa ggggttgagg gttgcctgga gtcagagaaa	3960
tttgtgttaa agaaggttga tatgaaacct gcaggcttag atacgtagat aagtagcatg	4020
gcgggttaat cattaactac aaggaacccc tagtgatgga gttggccact ccctctctgc	4080
gcgctcgctc gctcactgag gccgggcgac caaaggctgc cgcagcccg ggctttgccc	4140
gggcggcctc agtgagcgag cgagcgcgca gagaggaggt ggccaa	4186

<210> 81
 <211> 3713
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 81	
gcttcaggag cagttgtgcg aatagctgga gaacaccagg ctggatttaa acccagatcg	60
ctcttacatt tgctctttac ctgctgtgct cagcgttcac gtgccctcta gctgtagttt	120
tctgaagtca gcgcacagca aggcagtggt cttagagggt aacagaaggg aaaacaacaa	180
caacaaaaat ctaaatagaga atcctgactg tttcagctgg gggtaagggg ggcggattat	240
tcatataatt gttataaccag acggctgcag gcttagtcca attgcagaga actcgcttcc	300
caggcttctg agagtcccgg aagtccttaa acctgtctaa tcgacggggc ttgggtggcc	360
cgtcgtctcc tggcttcttc cctttaccca gggcgggcag cgaagtgggt cctcctgcgt	420
ccccacacc ctccctcagc ccctcccctc cggcccgtcc tgggcagggt acctggagca	480
tccggcaggc tgccctggcc tcctgcgtca ggacaacgcc cacgaggggc gttactgtgc	540
ggagatgcac cacgcaagag acaccctttg taactctctt ctctcccta gtgcgagggt	600
aaaaccttca gccccacgtg ctgtttgcaa acctgcctgt acctgaggcc ctaaaaagcc	660

agagacctca ctcccgggga gccagcatgt cactgcggt cctggaaaac ccaggcttgg	720
gcaggaaact ctctgacttt ggacaggtga gccacggcag cctgagctgc tcagttaggg	780
gaatttgggc ctccagagaa agagatctga agactgctgg tgcttcctgg ttccataagc	840
tcagtaagaa gtctgaattc gttggaagct gatgagaata tccaggaagt caacagacaa	900
atgtcctcaa caattgtttc taagtaggag aacatctgtc ctcggtggct ttcacaggaa	960
aagcttctga cctcttctct tcctcccaca gggcggtacc agatctggca gcggagaggg	1020
cagaggaagt cttctaacat gcggtgacgt ggaggagaat cccggcccta ggggtacctc	1080
caccgctgtg ctggagaacc ctgggctggg gaggaaactg tcagacttcg ggcaggagac	1140
ttcatacatt gaggataact gtaaccagaa tggcgccatc tctctgatct tcagcctgaa	1200
ggaggaagtg ggcgccctgg caaaggtgct gcgcctgttt gaggagaacg acgtgaatct	1260
gaccacatc gactccggc cttctagact gaagaaggac gactacgagt tctttaccca	1320
cctggataag cggtccctgc cagccctgac aaacatcatc aagatcctga ggcacgacat	1380
cggagcaacc gtgcacgagc tgtctcgga caagaagaag gataccgtgc cctggttccc	1440
tcggacaatc caggagctgg atagatttgc caaccagatc ctgtcttacg gagcagagct	1500
ggacgcagat caccctggct tcaaggacct agtgtatcgg gcccgagaa agcagtttgc	1560
cgatatcgcc tacaattata ggcacggaca gccaatccct cgcgtggagt atatggagga	1620
ggagaagaag acctggggca cagtgttcaa gacctgaag agcctgtaca agacacacgc	1680
ctgctacgag tataaccaca tcttccccct gctggagaag tattgtggct ttcacagga	1740
caatatccct cagctggagg acgtgagcca gttcctgcag acctgcacag gctttaggct	1800
gaggccagtg gcaggactgc tgagctccc ggacttcctg ggaggactgg cttcagagt	1860
gtttcactgc acccagtaca tcaggcacgg ctccaagcca atgtatacac cagagcccga	1920
catctgtcac gagctgctgg gccacgtgcc cctgttttagc gatagatcct tcgcccagtt	1980
ttcccaggag atcggactgg catctctggg agcacctgac gactacatcg agaagctggc	2040
caccatctat tggttcacag tggagtttgg cctgtgcaag cagggcgata gcatcaaggc	2100
ctacggagca ggactgctgt ctgacttcgg cgagctgcag tattgtctgt ccgagaagcc	2160
aaagtgtctg cccctggagc tggagaagac cgccatccag aactacaccg tgacagagtt	2220
ccagcccctg tactatgtgg ccgagtcttt taacgatgcc aaggagaagg tgagaaattt	2280
cgccgccaca atccctaggc cttcagcgt gcggtacgac cttataccc agaggatcga	2340
ggtgctggat aatacacagc agctgaagat cctggctgac tcaatcaata gcgaaatcgg	2400
aatcctgtgc tccgccctgc agaaaatcaa aggtaagcct atccctaacc ctctcctcgg	2460
tctcgattct acgtgatctt gtggaaagga cgaaacaccg gggaattcaa ggcctctcga	2520
gcctctagaa tccccgagac gtttcgtctc gggatcacta tagtgagtgc tattacgtac	2580
acagtgcagg ggaaagaata gtagagatcc agacatgata agatacattg atgagtttgg	2640
acaaaccaca actagaatgc agtgaaaaaa atgctttatt tgtgaaattt gtgatgctat	2700

tgctttattt gtaaccatta taagctgcaa taaacaagtt aacaacaaca attgcattca	2760
ttttatgttt caggttcagg gggagggtgtg ggagggtttt taactgggat gggatgtgga	2820
atccttctag atttcttttg taatatattat aaagtgtctt cagcaaggta tcaaaatggc	2880
aaaattgtga gtaactatcc tcctttcatt ttgggaagaa gatgaggcat gaagagaatt	2940
cagacagaaa cttactcaga ccaggggagg cagaaactaa gcagagagga aaatgaccaa	3000
gagttagccc tgggcatgga atgtgaaaga accctaaacg tgacttgga ataatgcca	3060
aggatatattc cattctcgg gatttgttgg cattttcttg aggtgaagaa ttgcagaata	3120
cattctttaa tgtgacctac atatttacc atgggaggaa gtctgtctct ggactcttga	3180
gattcagtca taaagcccag gccagggaaa taatgtaagt ctgcaggccc ctgtcatcag	3240
taggattagg gagaagagtt ctcagtagaa aacagggagg ctggagagaa aagaatggtt	3300
aatgttaacg ttaatataac tagaaagact gcagaactta ggactgattt ttatttgaat	3360
ccttaaaaaa aaaatttctt atgaaaatag tacatggctc ttaggagaca gaacttattg	3420
tacagaggaa cagcgtgaga gtcagagtga tcccagaaca ggtcctggct ccatcctgca	3480
catagttttg gtgctgctgg caatacggct cccacaactg tgggaagggg ttaggggcag	3540
ggatctcatc aggaaagcat aggggtttta agttctttat agagcactta gaagattgag	3600
aatccacaaa ttatattaat aacaaacaaa gtagtgtcgt gttatatagt aaatgtgaat	3660
ttgcagacac atttagggaa aagttataat taaaaaata ggctgtatat ata	3713

<210> 82

<211> 4139

<212> DNA

<213> Artificial Sequence

<220>

<223> Synthetic polynucleotide

<400> 82

ttggccactc cctctctgcg cgctcgctcg ctcactgagg ccgggcgacc aaaggtcgcc	60
cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc gagcgcgag agagggagtg	120
gccaactcca tcactagggg ttcttgagg ggtggagtcg tgacgtgaat tacgtcatag	180
ggttagggag gtcctgcata tgcggccgct tcaggagcag ttgtgcgaat agctggagaa	240
caccaggctg gatttaaacc cagatcgctc ttacatttgc tctttacctg ctgtgctcag	300
cgttcacgtg ccctctagct gtagttttct gaagtcagcg cacagcaagg cagtgtgctt	360
agaggttaac agaagggaac acaacaacaa caaaaatcta aatgagaatc ctgactgttt	420
cagctggggg taaggggggc ggattattca tataattgtt ataccagacg gtcgcaggct	480
tagtccaatt gcagagaact cgcttcccag gcttctgaga gtcccgaag tgcctaaacc	540
tgtctaatac acggggcctg ggtggcccgt cgctccctgg cttcttcctt ttaccaggg	600
cgggcagcga agtgggtgcct cctgcgtccc ccacaccctc cctcagcccc tcccctccgg	660
cccgctctgg gcaggtgacc tggagcatcc ggacggctgc cctggcctcc tgcgtcagga	720
caacgcccac gaggggcgtt actgtgcgga gatgcaccac gcaagagaca ccctttgtaa	780

ctctcttctc ctccctagtg cgaggttaaa accttcagcc ccacgtgctg tttgcaaacc	840
tgccctgtacc tgaggcccta aaaagccaga gacctcactc ccggggagcc agcatgtcca	900
ctgcggctct ggaaaaccca ggcttgggca ggaaactctc tgactttgga cagggtgagcc	960
acggcagcct gagctgtca gtaggggaa tttgggcctc cagagaaaga gatctgaaga	1020
ctgctggtgc ttcctggttt cataagctca gtaagaagtc tgaattcgtt ggaagctgat	1080
gagaatatcc aggaagtcaa cagacaaatg tcctcaacaa ttgtttctaa gtaggagaac	1140
atctgtcctc ggtggctttc acaggaaaag cttctgacct cttctcttcc tcccacaggg	1200
cggtagccaga tctggcagcg gagagggcag aggaagtctt ctaacatgcg gtgacgtgga	1260
ggagaatccc ggccctaggg gtacctccac cgctgtgctg gagaaccctg ggctggggag	1320
gaaactgtca gacttcgggc aggagacttc atacattgag gataactgta accagaatgg	1380
cgccatctct ctgatcttca gcctgaagga ggaagtgggc gccctggcaa aggtgctgcg	1440
cctgtttgag gagaacgacg tgaatctgac ccacatcgag tcccggcctt ctagactgaa	1500
gaaggacgag tacgagttct ttaccacct ggataagcgg tccctgccag ccctgacaaa	1560
catcatcaag atcctgaggc acgacatcgg agcaaccgtg cacgagctgt ctcgggacaa	1620
gaagaaggat accgtgccct ggttccctcg gacaatccag gagctggata gatttgccaa	1680
ccagatcctg tcttacggag cagagctgga cgcagatcac cctggcttca aggaccagt	1740
gtatcgggcc cggagaaagc agtttgccga tatcgctac aattataggc acggacagcc	1800
aatccctcgc gtggagtata tggaggagga gaagaagacc tggggcacag tgttcaagac	1860
cctgaagagc ctgtacaaga cacacgcctg ctacgagtat aaccacatct tccccctgct	1920
ggagaagtat tgtggctttc acgaggacaa tatccctcag ctggaggacg tgagccagtt	1980
cctgcagacc tgcacaggct ttaggctgag gccagtggca ggactgctga gctcccggga	2040
cttcctggga ggactggcct tcagagtgtt tcactgcacc cagtacatca ggcacggctc	2100
caagccaatg tataccacag agcccgacat ctgtcacgag ctgctgggcc acgtgcccct	2160
gtttagcgat agatccttcg cccagttttc ccaggagatc ggactggcat ctctgggagc	2220
acctgacgag tacatcgaga agctggccac catctattgg ttcacagtgg agtttggcct	2280
gtgcaagcag ggcgatagca tcaaggccta cggagcagga ctgctgtcta gcttcggcga	2340
gctgcagtat tgtctgtccg agaagccaaa gctgctgccc ctggagctgg agaagaccgc	2400
catccagaac tacaccgtga cagagttcca gcccctgtac tatgtggccg agtcttttaa	2460
cgatgccaaag gagaaggtga gaaatttcgc cgccacaatc ctaggccct tcagcgtgcg	2520
gtacgaccct tatacccaga ggatcgaggt gctggataat acacagcagc tgaagatcct	2580
ggctgactca atcaatagcg aaatcggaat cctgtgtccc gccctgcaga aaatcaaagg	2640
taagcctatc ctaaccctc tcctcggctc cgattctacg tgatcttggt gaaaggacga	2700
aacaccgggg aattcaaggc ctctcgagcc tctagaatcc ccgagacgtt tcgtctcggg	2760
atcactatag tgagtcgtat tacgtacaca gtgcagggga aagaatagta gagatccaga	2820

catgataaga tacattgatg agtttggaca aaccacaact agaatgcagt gaaaaaaatg	2880
ctttatttgt gaaatttgtg atgctattgc tttatttcta accattataa gctgcaataa	2940
acaagttaac aacaacaatt gcattcattt tatgtttcag gttcaggggg aggtgtggga	3000
ggttttttaa ctgggatggg atgtggaatc cttctagatt tcttttgtaa tatttataaa	3060
gtgctctcag caaggtatca aaatggcaaa attgtgagta actatcctcc tttcattttg	3120
ggaagaagat gaggcataa gagaattcag acagaaactt actcagacca ggggaggcag	3180
aaactaagca gagaggaaaa tgaccaagag ttagccctgg gcatggaatg tgaaagaacc	3240
ctaaacgtga cttggaaata atgcccgaag tatattccat tctccgggat ttgttggcat	3300
tttcttgagg tgaagaattg cagaatacat tctttaatgt gacctacata tttacccatg	3360
ggaggaagtc tgctcctgga ctcttgagat tcagtcataa agcccaggcc agggaaataa	3420
tgtaagtctg caggccccctg tcatcagtag gattagggag aagagttctc agtagaaaac	3480
agggaggctg gagagaaaaa aatggttaat gttaacgtta atataactag aaagactgca	3540
gaacttagga ctgattttta tttgaatcct taaaaaaaaa atttcttatg aaaatagtac	3600
atggctctta ggagacagaa cttattgtac agaggaacag cgtgagagtc agagtgatcc	3660
cagaacaggt cctggctcca tcctgcacat agttttggtg ctgctggcaa tacggtcccc	3720
acaactgtgg gaaggggtta ggggcaggga tctcatcagg aaagcatagg ggtttaaagt	3780
tctttataga gcacttagaa gattgagaat ccacaaatta tattaataac aaacaaagta	3840
gtgtcgtgtt atatagtaaa tgtgaatttg cagacacatt tagggaaaag ttataattaa	3900
aaaaataggc tgtatatata ctagatacgt agataagtag cctgcaggtc tagatacgta	3960
gataagtagc atggcgggtt aatcattaac tacaaggaac ccctagtgat ggagttggcc	4020
actccctctc tgcgcgctcg ctcgctcact gaggcgggc gaccaaaggt cgcccgcgc	4080
ccgggctttg cccgggcggc ctcagtgagc gagcgagcgc gcagagaggg agtggccaa	4139

<210> 83
 <211> 1353
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 83	
tccaccgctg tgctggagaa ccctgggctg gggaggaaac tgtcagactt cgggcaggag	60
acttcataca ttgaggataa ctgtaaccag aatggcgcca tctctctgat cttcagcctg	120
aaggaggaag tgggcgcctt ggcaaagggtg ctgcgcctgt ttgaggagaa cgacgtgaat	180
ctgaccacac tcgagtcccg gccttctaga ctgaagaagg acgagtacga gttctttacc	240
cacctggata agcggtcctt gccagccctg acaaacatca tcaagatcct gaggcacgac	300
atcggagcaa ccgtgcacga gctgtctcgg gacaagaaga aggataccgt gccctggttc	360
cctcggacaa tccaggagct ggaatagattt gcccaaccaga tcctgtctta cggagcagag	420
ctggacgcag atcaccctgg cttcaaggac ccagtgtatc gggcccgag aaagcagttt	480

gccgatatcg cctacaatta taggcacgga cagccaatcc ctcgctgga gtatatggag	540
gaggagaaga agacctgggg cacagtgttc aagaccctga agagcctgta caagacacac	600
gcctgctacg agtataacca catcttcccc ctgctggaga agtattgtgg ctttcacgag	660
gacaatatcc ctcagctgga ggacgtgagc cagttcctgc agacctgcac aggctttagg	720
ctgaggccag tggcaggact gctgagctcc cgggacttcc tgggaggact ggccttcaga	780
gtgtttcact gcacccagta catcaggcac ggctccaagc caatgtatac accagagccc	840
gacatctgtc acgagctgct gggccacgtg cccctgttta gcgatagatc cttcgcccag	900
ttttcccagg agatcggact ggcatctctg ggagcacctg acgagtacat cgagaagctg	960
gccaccatct attggttcac agtggagttt ggctgtgca agcagggcga tagcatcaag	1020
gcctacggag caggactgct gtctagcttc ggcgagctgc agtattgtct gtccgagaag	1080
ccaaagctgc tgcccctgga gctggagaag accgccatcc agaactacac cgtgacagag	1140
ttccagcccc tgtactatgt ggccgagtct tttaacgatg ccaaggagaa ggtgagaaat	1200
ttcgccgcca caatccctag gcccttcagc gtgcggtacg acccttatac ccagaggatc	1260
gaggtgctgg ataatacaca gcagctgaag atcctggctg actcaatcaa tagcgaaatc	1320
ggaatcctgt gtcgccct gcagaaaatc aaa	1353

<210> 84
 <211> 800
 <212> DNA
 <213> Artificial Sequence

<220>
 <223> Synthetic polynucleotide

<400> 84	
agcattagct tccatttatg cagtgtaaat ggtgagaaca gccccgactg aatacccaga	60
gcatcatctc gtctgtgtca ttcattgcaca taacatatct cagcgagggt gcccttctgt	120
cctctttgca gagaccagc caccatacta gtacctagag aactggctgg atttcagccc	180
cgatacctcc gggcttttgc tcatgttcgc ctcattaggt catctgggtg gttgcctaag	240
gaaaagtatg tcatggagac taacttgctt ggcatggaat aaaaggtagg ttgagagtgg	300
agcgtgttta aattgcaatc ctgcctctat ttctgtgctt gcagggaaca gtcattccta	360
attgctatcc tccatcatca tcatgattat ttctggtttt tctctggttg cggagaatcc	420
atactccagg tattccaatg tctcagcatt gccaggcctg tctgagcgtc aggatgtagg	480
tagtctgggc tctctgcctt ctattcttgt ccaggatact ctgcaaaaag aatcatgttg	540
tggctgccac cctcccaca aagcctcccg ctggggtcag tccaggactg gagttgggta	600
tggactgttc atgtctatcc actgctacgt cagggaaca cccactgaga gtgaccttgt	660
agactgcagt gggagacacc cttcaaaacc tctcctctcc tgtcctgaga gccagggtta	720
aaccatcagc ccgcctcct gaggcaaac ttttcctaac cctgctgcta agctagacac	780
ctcacttact gagagccagc	800

<210> 85
<211> 800
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 85
agcattagct tccatttatg cagtgtaaat ggtgagaaca gccccgactg aatacccaga 60
gcatcatctc gtctgtgtca ttcatgcaca taacatatct cagcgagggtg gcccttctgt 120
cctcttttgca gagacccagc caccatacta gtacctagag aactggctgg atttcagccc 180
cgatacctcc gggcttttgc tcatgttcgc ctcataggggt catctgggtg gttgcctaag 240
gaaaagtatg tcatggagac taacttgctt ggcatggaat aaaagggtgag ttgagagtgg 300
agcgtgttta aattgcaatc ctgcctctat ttctgtgctt gcagggaaca gtcaccta 360
attgctatcc tccatcatca tcatgattat ttctggtttt tctctggttg cggagaatcc 420
atactccagg tattccaatg tctcagcatt gccaggcctg tctgagcgtc aggatgtagg 480
tagtctgggc tctctgcctt ctattcttgt ccaggatact ctgccaaaag aatcatgttg 540
tggctgccac cctcccaca aagcctcccg cttgggtcag tccaggactg gagttgggta 600
tggactgttc atgtctatcc actgctacgt cagggaaca cccactgaga gtgaccttgt 660
agactgcagt gggagacacc cttcaaaacc tctcctctcc tgtcctgaga gccaggttaa 720
aaccatcagc cccgcatcct gaggcaaac ttttcctaac cctgctgcta agctagacac 780
cccacttact gagagccagc 800

<210> 86
<211> 800
<212> DNA
<213> Artificial Sequence

<220>
<223> Synthetic polynucleotide

<400> 86
gcagctgttg tcctggagaa cggagtcctg agcagaaaac tctcagactt tgggcaggta 60
agcctgttgg gcttccactg ctaggagaga attggttccc cacatgtgaa agcagtctgg 120
gaaatgctgg tatttccagt ctccaaaggc tactaagaaa tatgacttta tttagaggcg 180
aggaaaatgc ccaggaagtc aactgatgag actagtctta acaagttgag gatacagaaa 240
gttggggatc tgagctgcta ccaacatctg tgtgtctttg ggtggctcat tggatatctc 300
tgcctattgg ctttatcttc tgtacactga aaggaaatgg ctggctctta gtcacctggg 360
gtgggagtcc ctatctctcc agggatactt attcaatcct ttcttctggg tatcaaaatg 420
acaagcttgt aagaaactgt cctctttcgg ctttcaggag gtgatgtcgc atgaagagaa 480
tttggggggg gggacttact cagaaccaag gagggagaaa ttaaacagag agggaaatga 540
acaggagtta gcccggagcc tgaagcacct tggggattat gctgggggtg gagggaatcc 600
attgtcctcc ctagggaggg cttgcagaac atgttctttt ctgtgatatt tgtactttcc 660

ccagattgca aatcatgggtt tgtacactga gattcagtct ctggaggtaa tatgcctttt	720
ctagcttttc cttggacagg actaaggggt tgagggttgc ctggagtcag agaaatttgt	780
gttaaagaag gttgatatga	800