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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
20 August 2020 (20.08.2020)

WIPO | PCT

(10) International Publication Number
WO 2020/167996 A1

(51) International Patent Classification:
C07K 14/705 (2006.01) *C12N 15/864* (2006.01)

(21) International Application Number:
PCT/US2020/017987

(22) International Filing Date:
12 February 2020 (12.02.2020)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
62/804,521 12 February 2019 (12.02.2019) US
62/934,928 13 November 2019 (13.11.2019) US

(71) Applicant: **ROCKET PHARMACEUTICALS, LTD.** [US/US]; The Empire State Building, 350 Fifth Avenue, Suite 7530, New York, New York 10118 (US).

(72) Inventors: **KERAVALA, Annahita**; c/o Rocket Pharmaceuticals, Ltd., The Empire State Building, 350 Fifth Avenue, Suite 7530, New York, New York 10118 (US). **PRABHAKAR, Raj**; c/o Rocket Pharmaceuticals, Ltd., The Empire State Building, 350 Fifth Avenue, Suite 7530, New York, New York 10118 (US). **SHAH, Gaurav**; c/o Rocket Pharmaceuticals, Ltd., The Empire State Building, 350 Fifth Avenue, Suite 7530, New York, New York 10118 (US). **WONG, Roderick**; c/o Rocket Pharmaceuticals, Ltd., The Empire State Building, 350 Fifth Avenue, Suite 7530, New York, New York 10118 (US). **YALAMANCHI, Naveen**; c/o Rocket Pharmaceuticals, Ltd., The Empire State Building, 350 Fifth Avenue, Suite 7530, New York, New York 10118 (US). **PRATUMSUWAN, Piratip**; c/o Rocket Pharmaceuticals, Ltd., The Empire State Building, 350 Fifth Avenue, Suite 7530, New York, New York 10118 (US).

(74) Agent: **CHRISTIANSEN, William** et al.; Cooley LLP, 1299 Pennsylvania Avenue, NW, Suite 700, Washington, District of Columbia 20004 (US).

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

(54) Title: GENE THERAPY VECTORS FOR TREATMENT OF DANON DISEASE

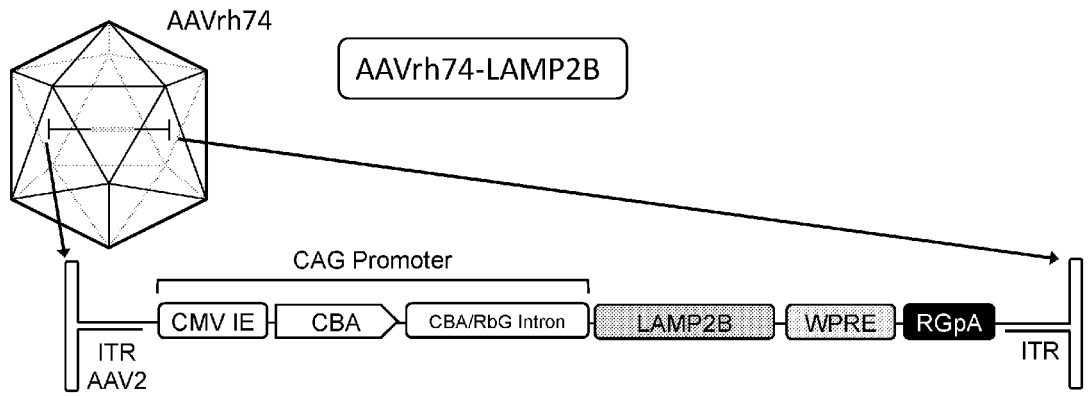

FIG. 1

(57) Abstract: The present disclosure provides gene therapy vectors comprising a polynucleotide sequence encoding a LAMP-2 polypeptide, methods of use thereof, pharmaceutical compositions, and more. In particular, the disclosure provides recombinant AAV vectors having AAVrh74 serotype expressing LAMP-2A, LAMP-2B, or LAMP-2C for use as a therapeutic in, for example, Danon Disease.

WO 2020/167996 A1

[Continued on next page]

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

Published:

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

GENE THERAPY VECTORS FOR TREATMENT OF DANON DISEASE

RELATED APPLICATIONS

This application claims priority to U.S. Provisional Patent Appl. No. 62/934,928, filed November 13, 2019, and U.S. Provisional Patent Appl. No. 62/804,521, filed February 12, 2019, each of which is incorporated herein by reference in its entirety.

SEQUENCE LISTING

This application is being filed electronically via EFS-Web and includes an electronically submitted sequence listing in .txt format. The .txt file contains a sequence listing entitled "ROPA_013_02WO_ST25.txt" created on February 11, 2020 and having a size of ~61 kilobytes. The sequence listing contained in this .txt file is part of the specification and is incorporated herein by reference in its entirety.

FIELD OF INVENTION

The invention relates generally to gene therapy for diseases associated with mutations in lysosome-associated membrane protein 2 (LAMP-2, also known as CD107b).

BACKGROUND

Lysosome-associated membrane protein 2 (LAMP-2, also known as CD107b) is a gene that encodes a lysosome-associated membrane glycoprotein. Alternative splicing of the gene produces three isoforms: LAMP-2A, LAMP-2B, and LAMP-2C. Loss-of-function mutations in LAMP-2 are associated with human diseases, including Danon disease, a familial cardiomyopathy associated with impaired autophagy. Danon disease is a rare but serious cardiac and skeletal myopathy leading to substantial morbidity and early mortality due to arrhythmia and cardiomyopathy. The X-linked nature of inheritance accounts for reported differences in phenotypic severity between men and women. Boucek et al. *Genetics in Medicine* 13:563-568 (2011). The disease is now understood to be caused by a primary deficiency in lysosome-associated membrane protein-2 (LAMP-2), which functions as a lysosomal membrane receptor in chaperone-mediated autophagy. Nishino et al. *Nature* 406:906-910 (2000).

SUMMARY OF THE INVENTION

The present disclosure provides such gene therapy vectors related to LAMP2, methods of use thereof, pharmaceutical compositions, and more. Although clinical use of adeno-associated virus (AAV) vectors is known, the selection of preferred serotype(s) of AAV for gene therapy remains challenging and unpredictable.

The present disclosure provides improved gene therapy vectors comprising a polynucleotide sequence encoding a LAMP-2 polypeptide, methods of use thereof, pharmaceutical compositions, and more. In particular, the disclosure provides recombinant AAV vectors having AAVrh74 serotype expressing LAMP-2A, LAMP-2B, or LAMP-2C for use as a therapeutic in, for example, Danon disease.

Other features and advantages of the invention will be apparent from and encompassed by the following detailed description and claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 depicts an embodiment of an AAV vector having AAVrh74 serotype.

FIG. 2 shows a bar graph of vector DNA quantification in organs most affected in Danon disease by qPCR.

FIG. 3A shows a bar graph of vector DNA quantification in regions of the heart by qPCR.

FIG. 3B shows a bar graph of vector DNA quantification in muscles by qPCR.

FIG. 4 shows a bar graph of mRNA quantification in organs most affected by Danon disease by RT-qPCR.

FIG. 5A shows a bar graph of mRNA quantification in regions of the heart by RT-qPCR.

FIG. 5B shows a bar graph of mRNA quantification in muscles by RT-qPCR.

FIG. 6A shows a micrograph of semi-quantitative mRNA analysis by RNAscope in an untreated left ventricle.

FIG. 6B shows micrographs of semi-quantitative mRNA analysis by RNAscope in treated left ventricles.

FIG. 7A shows a micrograph of semi-quantitative mRNA analysis by RNAscope in an untreated quadricep.

FIG. 7B shows micrographs of semi-quantitative mRNA analysis by RNAscope in treated quadriceps.

FIG. 8 shows micrographs of semi-quantitative mRNA analysis by RNAscope in treated gastrocnemius.

FIG. 9 shows a bar graph of protein quantification in tissues most affected in Danon disease by ELISA.

FIG. 10A shows a bar graph of protein quantification in regions of the heart by ELISA.

FIG. 10B shows a bar graph of protein quantification in muscles by ELISA.

FIGS. 11A-11D show line graphs of clinical pathology measurement in NHP serum over course of study. Clinical pathology levels were assessed as changes in (**FIG. 11A**) alanine aminotransferase, ALT; (**FIG. 11B**) aspartate aminotransferase, AST; (**FIG. 11C**) white blood cells, WBC; and (**FIG. 11D**) neutrophils over the study duration.

DETAILED DESCRIPTION OF THE INVENTION

The present disclosure provides AAVrh74-based gene therapy vectors that employ optimized expression cassettes to deliver a polynucleotide encoding one of the Lysosome-associated membrane protein 2 (LAMP2) proteins, also known as CD107b. Generally, the LAMP2 is a human LAMP2, though expression of any mammalian LAMP-2 is envisioned. The native *LAMP2* gene encodes by alternative splicing three variants: LAMP-2A, LAMP-2B, and LAMP-2C. LAMP-2B is associated with Danon disease. Although the disclosure concerns primarily Danon disease, LAMP2 is implicated in various other disease, including cancer. The disclosed vectors may be used to treat any of these diseases.

The disclosure further relates to AAVrh74 capsids or capsids having substantial homology to the AAVrh74 capsid and retaining the function of the AAVrh74 capsid. The disclosure provides the sequences listed in **Table 1**. Table 1 further provides polynucleotide sequences used in various embodiments. The sequences are not intended to limit the invention, as substitution or modification of these sequences with different promoters, enhancer, or other genetic elements is contemplated.

Table 1: Sequences

SEQ ID NO:	Type	Description
1	nucleotide	AAVrh74 capsid coding sequence
2	protein	AAVrh74 VP1
3	protein	AAVrh74 VP2
4	protein	AAVrh74 VP3
5	protein	LAMP-2B (wild-type)
6	nucleotide	LAMP-2B coding sequence (wild-type)
7	nucleotide	LAMP-2B engineered coding sequence
8	nucleotide	LAMP-2B engineered coding sequence
9	nucleotide	LAMP-2B engineered coding sequence
10	nucleotide	Engineered expression cassette
11	nucleotide	Engineered expression cassette
12	nucleotide	Engineered expression cassette
13	nucleotide	AAV inverted terminal repeat
14	nucleotide	AAV inverted terminal repeat
15	protein	LAMP-2A protein sequence
16	protein	LAMP-2B protein sequence
17	protein	LAMP-2C protein sequence
18	nucleotide	CAG promoter
19	nucleotide	WPRE
20	nucleotide	Kozak sequence
21	nucleotide	Kozak sequence
22	nucleotide	Kozak sequence
23	nucleotide	Kozak sequence
24	nucleotide	Kozak sequence
25	nucleotide	Kozak sequence
26	nucleotide	polyadenylation signal (full length)
27	nucleotide	bGH polyadenylation signal (bGHpA)
28	nucleotide	SV40 early/late polyadenylation signal
29	nucleotide	human growth hormone (HGH) polyadenylation signal

The disclosure provides recombinant adeno-associated virus (rAAV) gene therapy vectors. As used herein, an "rAAV gene therapy vector" refers to a complete virus including nucleic acid and protein components, including capsid proteins. In some embodiments, the capsid protein is encoded by a polynucleotide supplied on a plasmid *in trans* to the transfer plasmid. The polynucleotide sequence of wild-type AAVrh74 *cap* is as follows:

AAVrh74 capsid coding sequence (SEQ ID NO: 1)

ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAACCTCTCTGAGGGCA
TTCGCGAGTGGTGGGACCTGAAACCTGGAGCCCCGAAACCCAAAGCCAACCAGC
AAAAGCAGGACAACGGCCGGGGTCTGGTGCTTCCTGGCTACAAGTACCTCGGAC
CCTTCAACGGACTCGACAAGGGGGAGCCCGTCAACGCGGCGGACGCAGCGGCCC
TCGAGCACGACAAGGCCTACGACCAGCAGCTCCAAGCGGGTGACAATCCGTACC
TGCGGTATAATCACGCCGACGCCGAGTTTCAGGAGCGTCTGCAAGAAGATACGT
CTTTTGGGGGCAACCTCGGGCGCGCAGTCTTCCAGGCCAAAAAGCGGGTTCTCG
AACCTCTGGGCCTGGTTGAATCGCCGGTTAAGACGGCTCCTGGAAAGAAGAGAC
CGGTAGAGCCATCACCCCAGCGCTCTCCAGACTCCTCTACGGGCATCGGCAAGA
AAGGCCAGCAGCCCGCAAAAAAGAGACTCAATTTTGGGCAGACTGGCGACTCAG
AGTCAGTCCCCGACCCTCAACCAATCGGAGAACCACCAGCAGGCCCCTCTGGTCT
GGGATCTGGTACAATGGCTGCAGGCGGTGGCGCTCCAATGGCAGACAATAACGA
AGGCGCCGACGGAGTGGGTAGTTCCTCAGGAAATTGGCATTGCGATTCCACATG
GCTGGGCGACAGAGTCATCACCACCAGCACCCGCACCTGGGCCCTGCCCACCTA
CAACAACCACCTCTACAAGCAAATCTCCAACGGGACCTCGGGAGGAAGCACCAA
CGACAACACCTACTTCGGCTACAGCACCCCCTGGGGGTATTTTGACTTCAACAGA
TTCCACTGCCACTTTTCACCACGTGACTGGCAGCGACTCATCAACAACAACTGGG
GATTCCGGCCCAAGAGGCTCAACTTCAAGCTCTTCAACATCCAAGTCAAGGAGG
TCACGCAGAATGAAGGCACCAAGACCATCGCCAATAACCTTACCAGCACGATTC
AGGTCTTTACGGACTCGGAATACCAGCTCCCGTACGTGCTCGGCTCGGCGCACCA
GGGCTGCCTGCCTCCGTTCCCGGCGGACGTCTTCATGATTCCTCAGTACGGGTAC
CTGACTCTGAACAATGGCAGTCAGGCTGTGGGCCGGTCGTCCTTCTACTGCCTGG
AGTACTTTCCTTCTCAAATGCTGAGAACGGGCAACAACTTTGAATTCAGCTACAA
CTTCGAGGACGTGCCCTTCCACAGCAGCTACGCGCACAGCCAGAGCCTGGACCG
GCTGATGAACCCTCTCATCGACCAGTACTTGTACTACCTGTCCCGGACTCAAAGC
ACGGGCGGTACTGCAGGAACTCAGCAGTTGCTATTTTCTCAGGCCGGGCCTAACA
ACATGTCGGCTCAGGCCAAGAACTGGCTACCCGGTCCCTGCTACCGGCAGCAAC

GCGTCTCCACGACACTGTCGCAGAACAACAACAGCAACTTTGCCTGGACGGGTG
CCACCAAGTATCATCTGAATGGCAGAGACTCTCTGGTGAATCCTGGCGTTGCCAT
GGCTACCCACAAGGACGACGAAGAGCGATTTTTTCCATCCAGCGGAGTCTTAAT
GTTTGGGAAACAGGGAGCTGGAAAAGACAACGTGGACTATAGCAGCGTGATGCT
AACCAGCGAGGAAGAAATAAAGACCACCAACCCAGTGGCCACAGAACAGTACG
GCGTGGTGGCCGATAACCTGCAACAGCAAAACGCCGCTCCTATTGTAGGGGCCG
TCAATAGTCAAGGAGCCTTACCTGGCATGGTGTGGCAGAACCGGGACGTGTACC
TGCAGGGTCCCATCTGGGCCAAGATTCCTCATACGGACGGCAACTTTCATCCCTC
GCCGCTGATGGGAGGCTTTGGACTGAAGCATCCGCCTCCTCAGATCCTGATTAAA
AACACACCTGTTCCCGCGGATCCTCCGACCACCTTCAATCAGGCCAAGCTGGCTT
CTTTCATCACGCAGTACAGTACCGGCCAGGTCAGCGTGGAGATCGAGTGGGAGC
TGCAGAAGGAGAACAGCAAACGCTGGAACCCAGAGATTCAGTACACTTCCAACT
ACTACAAATCTACAAATGTGGACTTTGCTGTCAATACTGAGGGTACTTATTCCGA
GCCTCGCCCCATTGGCACCCGTTACCTCACCCGTAATCTGTAA

The disclosure further provides protein sequences for AAVrh74 VP1, VP2, and VP3, including SEQ ID NOs: 2-4, and homologs or functional variants thereof.

AAVrh74 VP1 (SEQ ID NO: 2)

MAAGGGAPMADNNEGADGVGSSSGNWHCDSTWLGDRVITTSTRTWALPTYNNHL
YKQISNGTSGGSTNDNTYFGYSTPWGYFDFNRFHCHFSPRDWQRLINNNWGFRPKR
LNFKLFNIQVKEVTQNEGTKTIANNLTSTIQVFTDSEYQLPYVLGSAHQGCLPPFPAD
VFMIPQYGYLTLNNGSQAVGRSSFYCLEYFPSQMLRTGNNFEFSYNFEDVPFHSSYA
HSQSLDRLMNPLIDQYLYYLSRTQSTGGTAGTQQLLFSQAGPNNMSAQAKNWLPGP
CYRQQRVSTTLSQNNNSNFAWTGATKYHLNGRDSLVNPGVAMATHKDDEERFFPS
SGVLMFGKQGAGKDNVDYSSVMLTSEEEIKTTNPVATEQYGVVADNLQQQNAAPI
VGAVNSQGALPGMVWQNRDVYLQGPIWAKIPHTDGNFHPSPLMGGFGLKHPPPQIL
IKNTPVPADPPTTFNQAKLASFITQYSTGQVSVEIEWELQKENSKRWNPEIQYTSNYY
KSTNVDFAVNTEGTYSEPRPIGTRYLTRNL

AAVrh74 VP2 (SEQ ID NO: 3)

STIQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRSSFYCL
EYFPSQMLRTGNNFEFSYNFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLSRTQSTG

GTAGTQQLLFSQAGPNNMSAQAKNWLPGPCYRQQRVSTTLSQNNNSNFAWTGATK
YHLNGRDSLVNPGVAMATHKDDEERFFPSSGVLMFGKQGAGKDNVDYSSVMLTSE
EEIKTTNPVATEQYGVVADNLQQQNAAPIVGAVNSQGALPGMVWQNRDVYLQGPI
WAKIPHTDGNFHPSPLMGGFGLKHPPPQILIKNTPVPADPPTTFNQAKLASFITQYSTG
QVSVEIEWELQKENSKRWNPEIQYTSNYYKSTNVDFAVNTEGTYSEPRPIGTRYLTR
NL

AAVrh74 VP3 (SEQ ID NO: 4)

RTGNNFEFSYNFEDVPFHSSYAHSQSLDRLMNPLIDQYLYYLSRTQSTGGTAGTQQL
LFSQAGPNNMSAQAKNWLPGPCYRQQRVSTTLSQNNNSNFAWTGATKYHLNGRDS
LVNPGVAMATHKDDEERFFPSSGVLMFGKQGAGKDNVDYSSVMLTSEEEIKTTNPV
ATEQYGVVADNLQQQNAAPIVGAVNSQGALPGMVWQNRDVYLQGPIWAKIPHTD
GNFHPSPLMGGFGLKHPPPQILIKNTPVPADPPTTFNQAKLASFITQYSTGQVSVEIEW
ELQKENSKRWNPEIQYTSNYYKSTNVDFAVNTEGTYSEPRPIGTRYLTRNL

In certain cases, the AAVrh74 capsid comprises the amino acid sequence set forth in SEQ ID NO: 2. In some embodiments, the rAAV vector comprises a polypeptide that comprises, or consists essentially of, or yet further consists of a sequence, *e.g.*, at least 65%, at least 70%, at least 75%, at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more typically 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to amino acid sequence of AAVrh74 VP1 which is set forth in SEQ ID NO: 2. In some embodiments, the rAAV vector comprises a polypeptide that comprises, or consists essentially of, or yet further consists of a sequence, *e.g.*, at least 65%, at least 70%, at least 75%, at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more typically 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to amino acid sequence of AAVrh74 VP2 which is set forth in SEQ ID NO: 3. In some embodiments, the rAAV vector comprises a polypeptide that comprises, or consists essentially of, or yet further consists of a sequence, *e.g.*, at least 65%, at least 70%, at least 75%, at least 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, or 89%, more typically 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% or more identical to amino acid sequence of AAVrh74 VP3 which is set forth in SEQ ID NO: 4.

The wild-type polypeptide sequence of LAMP-2B (SEQ ID NO: 5) and the wild-type polynucleotide sequence of LAMP-2B (SEQ ID NO: 6) are, respectively:

MVCFRLFPVPGSGLVLVCLVLGAVRSYALELNLTDSENATCLYAKWQMNFT
VRYETTNKTYKTVTISDHGTVTYNGSICGDDQNGPKIAVQFGPGFSWIANFTK
AASTYSIDSVSFSYNTGDNTTFPDAEDKGILTVDELLAIRIPLNDLFRCNSLSTL
EKNDVVQHYWDVLVQAFVQNGTVSTNEFLCDKDKTSTVAPTIHTTVPSPTTT
PTPKEKPEAGTYSVNNGNDTCLLATMGLQLNITQDKVASVININPNTTHSTGS
CRSHTALLRLNSSTIKYLDFVFAVKNENRFYLKEVNISMYLVNGSVFSIANNN
LSYWDAPLGSSYMCNKEQTVSVSGAFQINTFDLRVQPFNVTQGKYSTAQECS
LDDDTILIPIIVGAGLSGLIIVIVIAYVIGRRKSYAGYQT (SEQ ID NO: 5); and

ATGGTGTGCTTCCGCCTCTTCCCGGTTCCGGGCTCAGGGCTCGTTCTGGTCTGCCT
AGTCCTGGGAGCTGTGCGGTCTTATGCATTGGAACTTAATTTGACAGATTCAGAA
AATGCCACTTGCCTTTATGCAAAATGGCAGATGAATTTCACAGTTCGCTATGAAA
CTACAAATAAAACTTATAAAACTGTAACCATTTCAGACCATGGCACTGTGACATA
TAATGGAAGCATTTGTGGGGATGATCAGAATGGTCCCAAAATAGCAGTGCAGTT
CGGACCTGGCTTTTCCTGGATTGCGAATTTTACCAAGGCAGCATCTACTTATTCA
ATTGACAGCGTCTCATTTTCCTACAACACTGGTGATAACACAACATTTCCTGATG
CTGAAGATAAAGGAATTCTTACTGTTGATGAACTTTTGGCCATCAGAATTCCATT
GAATGACCTTTTTAGATGCAATAGTTTATCAACTTTGGAAAAGAATGATGTTGTC
CAACACTACTGGGATGTTCTTGTACAAGCTTTTGTCCAAAATGGCACAGTGAGCA
CAAATGAGTTCCTGTGTGATAAAGACAAAACTTCAACAGTGGCACCCACCATAC
ACACCACTGTGCCATCTCCTACTACAACACCTACTCCAAAGGAAAAACCAGAAG
CTGGAACCTATTCAGTTAATAATGGCAATGATACTTGTCTGCTGGCTACCATGGG
GCTGCAGCTGAACATCACTCAGGATAAGGTTGCTTCAGTTATTAACATCAACCCC
AATACAACTCACTCCACAGGCAGCTGCCGTTCTCACACTGCTCTACTTAGACTCA
ATAGCAGCACCATTAAGTATCTAGACTTTGTCTTTGCTGTGAAAAATGAAAACCG
ATTTTATCTGAAGGAAGTGAACATCAGCATGTATTTGGTTAATGGCTCCGTTTTC
AGCATTGCAAATAACAATCTCAGCTACTGGGATGCCCCCCTGGGAAGTTCTTATA
TGTGCAACAAAGAGCAGACTGTTTCAGTGTCTGGAGCATTTCAGATAAATACCTT
TGATCTAAGGGTTCAGCCTTTCAATGTGACACAAGGAAAGTATTCTACAGCCCAA
GAGTGTTCGCTGGATGATGACACCATTCTAATCCCAATTATAGTTGGTGCTGGTC
TTTCAGGCTTGATTATCGTTATAGTGATTGCTTACGTAATTGGCAGAAGAAAAAG
TTATGCTGGATATCAGACTCTGTAA (SEQ ID NO:6).

In an embodiment, the transgene shares at least 95% identity to the polynucleotide sequence of SEQ ID NO: 5. In an embodiment, the transgene shares at least 99% identity to the polynucleotide sequence of SEQ ID NO: 5. In an embodiment, the transgene comprises the polynucleotide sequence of SEQ ID NO: 5. In embodiment, the transgene shares at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or complete identity to SEQ ID NO:5.

In an embodiment, the transgene encodes a polypeptide that shares at least 95% identity to the amino acid sequence of SEQ ID NO: 6. In an embodiment, the transgene encodes a polypeptide shares at least 99% identity to the amino acid sequence of SEQ ID NO: 6. In an embodiment, the polypeptide encoded by the transgene comprises the amino acid sequence of SEQ ID NO: 6. In embodiment, the polypeptide encoded by the transgene shares at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or complete identity to SEQ ID NO:6.

Disclosed herein are modifications to the gene sequence of LAMP-2B including: codon-optimization, CpG depletion, removal of cryptic splice sites, and reduction of alternative open-reading frames (ORFs). In embodiments, the disclosure provides a transgene encoding an isoform of lysosome-associated membrane protein 2 (LAMP-2) or a functional variant thereof. In embodiments, the transgene shares at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or complete identity to a sequence selected from SEQ ID NO: 7-9. The disclosure provides at least three variant gene sequences for LAMP-2B (SEQ ID NO: 7-9):

ATGGTCTGCTTCAGACTGTTCCCTGTCCCTGGATCTGGTCTGGTGCTTGTGTGCTT
GGTGCTGGGTGCTGTGAGATCCTATGCCCTTGAGCTGAACCTGACTGACTCAGAA
AATGCCACTTGCCTGTATGCCAAGTGGCAGATGAACTTCACTGTGAGATATGAGA
CTACCAACAAGACCTACAAGACTGTGACCATCTCAGACCATGGCACTGTCACCTA
CAATGGATCAATCTGTGGTGATGATCAGAATGGCCCAAAGATAGCAGTGCAGTT
TGGGCCCGGTTTTTCCTGGATTGCTAACTTCACCAAGGCAGCCTCCACCTACAGC
ATTGACTCAGTCAGCTTCAGCTACAACACTGGGGATAACACCACCTTCCCTGACG
CAGAGGACAAGGGAATCCTTACTGTGGACGAACTCCTGGCAATCAGAATCCCCC
TTAACGACCTGTTCAGATGCAACTCCCTTTCAACCCTTGAAAAGAATGATGTGGT
GCAACACTATTGGGACGTCCTGGTGCAAGCCTTTGTGCAGAATGGGACAGTGAG
TACCAACGAGTTCCTCTGTGACAAGGACAAGACCAGCACTGTGGCCCCCACTATC
CACACCACTGTGCCCAGCCCTACCACTACCCCCACCCCTAAAGAGAAGCCAGAA

GCTGGAACCTACTCAGTCAACAATGGAAATGACACATGCCTCCTTGCCACCATGG
GACTGCAGCTGAACATCACTCAGGACAAGGTGGCCTCAGTGATTAACATCAACC
CTAACACCACTCATAGCACTGGGAGCTGCAGATCACATACAGCTCTGCTGAGGCT
CAACTCCTCCACCATCAAGTACCTGGACTTTGTGTTTGCTGTGAAGAATGAGAAC
AGGTTCTACCTCAAGGAAGTGAACATTTCCATGTACCTGGTCAATGGTTCAGTGT
TCTCTATTGCCAACAACAATCTGAGCTACTGGGATGCACCCCTGGGATCCTCCTA
CATGTGCAACAAGGAGCAGACTGTGAGTGTGTCAGGTGCTTTTCAGATCAACACT
TTTGACCTGAGGGTGCAGCCCTTCAATGTGACTCAGGGAAAGTACTCCACTGCAC
AAGAGTGTTCCTTGGATGATGACACTATCCTCATCCCCATTATTGTGGGAGCTGG
ACTGTCAGGATTGATTATAGTGATTGTGATTGCTTATGTGATTGGAAGGAGAAAG
AGCTATGCTGGCTACCAGACCCTGTAA (SEQ ID NO: 7);

ATGGTGTGCTTTAGACTGTTTCCTGTGCCTGGTTCAGGGCTGGTCCTGGTCTGTCT
GGTGCTGGGGGCTGTCAGAAGCTATGCCTTGGAGCTGAACCTCACTGATAGTGA
AAATGCCACTTGTCTGTATGCTAAGTGGCAGATGAACTTCACTGTGAGATATGAA
ACCACCAACAAGACTTACAAAACAGTGACCATCTCAGATCATGGAACTGTGACC
TACAACGGCAGCATTTGTGGAGACGACCAGAACGGACCAAAAATCGCTGTCCAA
TTTGGGCCTGGATTCTCCTGGATTGCCAATTTCACTAAAGCTGCCTCCACATATTC
AATTGACTCAGTGTCCTTCTCCTACAACACTGGGGACAACACTACTTTCCCTGAT
GCTGAAGATAAGGGAATCTTGACAGTGGATGAGCTGCTGGCTATCAGGATCCCT
TTGAATGACCTGTTTAGGTGTAATTCACTGAGCACTCTGGAGAAGAACGACGTGG
TGCAGCACTACTGGGACGTGCTGGTGCAGGCCTTTGTGCAGAACGGCACTGTGTC
CACCAACGAATTCCTGTGTGATAAGGACAAAACTTCCACTGTGGCACCTACAATT
CACACTACTGTGCCTTCACCTACCACCACTCCAACTCCAAAGGAAAAGCCTGAAG
CAGGAACCTACTCTGTGAACAATGGCAATGATACCTGTCTGTTGGCCACCATGGG
CCTCCAACTGAACATTACTCAGGACAAGGTGGCCTCAGTGATTAACATTAACCCC
AACACTACCCACTCCACTGGCAGCTGTAGATCACACACAGCCTTGCTCAGACTGA
ATAGCAGCACCATCAAGTATTTGGATTTTGTGTTTGCAGTGAAGAATGAAAACAG
GTTCTACCTGAAGGAAGTCAACATCTCAATGTACCTGGTGAACGGCTCAGTGTTC
AGCATTGCCAACAACAACCTCTCCTATTGGGACGCTCCACTGGGGAGCAGCTAC
ATGTGTAACAAGGAACAGACTGTGTCAGTGTCAGGAGCCTTCCAGATTAACACC
TTTGATCTGAGGGTCCAACCCTTTAATGTCACTCAAGGAAAGTATAGCACTGCCC
AGGAGTGCTCCCTGGATGATGACACCATTCTGATTCCAATCATTGTGGGTGCAGG

ACTTTCTGGGCTTATTATTGTGATTGTGATTGCCTATGTGATTGGCAGAAGGAAA
TCCTATGCAGGGTACCAAACTCTGTAA (SEQ ID NO: 8); or

ATGGTCTGTTTTAGGCTGTTCCCTGTCCCTGGTTCAGGACTGGTCTTAGTGTGTCT
GGTGCTTGGAGCTGTCAGAAGCTATGCCCTGGAGCTGAACCTGACTGACTCAGA
AAATGCCACTTGCCTGTATGCCAAGTGGCAGATGAACTTCACTGTCAGATATGAA
ACCACCAACAAGACCTATAAGACTGTGACCATCTCAGACCATGGCACTGTGACTT
ACAATGGGTCAATTTGTGGAGATGACCAGAATGGCCCTAAGATAGCTGTCCAGT
TTGGTCCAGGATTCAGCTGGATTGCCAACTTCACCAAGGCAGCCAGCACCTACAG
CATTGACTCTGTGTCCTTCTCCTACAACACAGGAGACAACACCACTTTCCCTGAT
GCAGAGGACAAAGGTATCCTGACTGTGGATGAGTTGCTGGCAATCAGGATCCCA
CTGAACGATCTGTTCAGGTGCAACTCACTGTCCACTCTGGAAAAGAATGATGTGG
TGCAGCACTATTGGGATGTGCTAGTCCAGGCCTTTGTCCAGAATGGGACTGTGTC
AACTAATGAGTTCCTGTGTGACAAGGACAAGACAAGCACTGTAGCCCCCACTAT
CCATACCACAGTACCTAGCCCCACCACTACTCCAACCCCCAAGGAGAAGCCTGA
GGCTGGCACCTACTCAGTGAACAATGGGAATGACACCTGTTTGCTGGCCACTATG
GGACTCCAACTGAACATCACCCAGGACAAAGTGGCCTCTGTGATCAATATCAAT
CCCAACACCACCCACAGCACTGGGTCCTGCAGAAGCCACACTGCCCTCCTGAGG
CTCAACTCATCAACTATCAAGTACTTGGATTTTGTGTTTGCAGTGAAGAATGAGA
ACAGATTCTACCTCAAAGAGGTCAACATTTCAATGTACCTGGTGAATGGGAGTGT
GTTCTCCATTGCTAACAACAACCTGAGCTACTGGGATGCCCCTCTGGGCTCCTCA
TACATGTGCAACAAGGAACAGACTGTGAGTGTGTCAGGGGCCTTCCAGATCAAC
ACTTTTGACCTGAGAGTGCAGCCCTTTAATGTGACACAGGGAAAGTACAGCACT
GCTCAGGAGTGCAGCCTGGATGATGACACTATCCTGATCCCTATCATTGTGGGGG
CAGGCCTGTCTGGACTCATTATTGTGATTGTGATTGCCTATGTGATAGGGAGAAG
GAAGTCTTATGCTGGATACCAGACCCTGTAA (SEQ ID NO: 9).

In an embodiment, the transgene shares at least 95% identity to a sequence selected from SEQ ID NO: 7-9. In an embodiment, the transgene shares at least 99% identity to a sequence selected from SEQ ID NO: 7-9. In an embodiment, the transgene comprises a sequence selected from SEQ ID NO: 7-9. In embodiment, the transgene shares at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or complete identity to SEQ ID NO: 7. In embodiment, the transgene shares at least 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or complete identity to SEQ ID NO: 8. In embodiment, the transgene shares at least 85%, 86%, 87%, 88%, 89%,

90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or complete identity to SEQ ID NO: 9.

In some cases, the transgene has a polynucleotide sequence that is different from the polynucleotide sequence of a reference sequence, *e.g.*, a "native" or "wild-type" LAMP-2B sequence. In some embodiments, the transgene shares at most 70%, 71%, 72%, 73%, 74%, 75%, 76%, 77%, 78%, 79%, 80%, 81%, 82%, 83%, 84%, 85%, 86%, 87%, 88%, 89%, 90%, 91%, 92%, 93%, 94%, or 95% identity with a reference sequence. In some embodiments, the reference sequence is SEQ ID NO: 6. For example, SEQ ID NO: 7 shares 78.5% identity to SEQ ID NO: 6.

In some embodiments, the transgene is similar to or identical to a subsequence of any one of SEQ ID NOs: 5 or 7-9. In some embodiments, the transgene comprises a subsequence of any one of SEQ ID NOs: 5 or 7-9. In various embodiments, the subsequence may comprise any set of consecutive nucleotides (nt) in the full sequence having a length of at least about 50 nt, at least about 100 nt, at least about 150 nt, at least about 250 nt, at least about 200 nt, at least about 350 nt, at least about 450 nt, at least about 400 nt, at least about 450 nt, at least about 550 nt, at least about 600 nt, at least about 650 nt, at least about 600 nt, at least about 650 nt, at least about 700 nt, at least about 750 nt, at least about 800 nt, at least about 850 nt, at least about 900 nt, at least about 950 nt, at least about 1000 nt, at least about 1050 nt, at least about 1100 nt, at least about 1150 nt, or at least about 1200 nt.

In some embodiments, the transgene encodes a polypeptide similar to or identical to a subsequence of any one of SEQ ID NOs: 6 or 16-18. In some embodiments, the transgene encodes a polypeptide comprises a subsequence of any one of SEQ ID NOs: 6 or 16-18. In some embodiments, the subsequence may comprises any set of consecutive amino acids (aa) in the full sequence having a length of at least about 20 aa, at least about 30 aa, at least about 50 aa, at least about 70 aa, at least about 80 aa, at least about 100 aa, at least about 120 aa, at least about 130 aa, at least about 150 aa, at least about 170 aa, at least about 180 aa, at least about 200 aa, at least about 220 aa, at least about 230 aa, at least about 250 aa, at least about 270 aa, at least about 280 aa, at least about 300 aa, at least about 320 aa, at least about 330 aa, at least about 350 aa, at least about 370 aa, at least about 380 aa, or at least about 400 aa.

In some embodiments, the transgene encodes a LAMP-2 polypeptide comprising an N-terminal truncation 1 to 10 amino acids (aa), 1 to 20 aa, 1 to 30 aa, 1 to 40 aa, or 1 to 50 aa, or an N-terminal truncation 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20,

21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 48, 50, or more aa; and/or a C-terminal truncation 1 to 10 amino acids (aa), 1 to 20 aa, 1 to 30 aa, 1 to 40 aa, or 1 to 50 aa, or a C-terminal truncation 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 48, 50, or more aa.

In some embodiments, the subsequence of the LAMP2 polypeptide comprises a functional variant of LAMP-2A, LAMP-2B, or LAMP-2C. As used herein, a "functional variant" refers to polypeptide sharing sequence similarity to a reference LAMP-2A, LAMP-2B, or LAMP-2C and having at least one biological property of LAMP-2A, LAMP-2B, or LAMP-2C. The biological property may include the ability to specifically interact with one or more binding partners, the ability to bind an anti-LAMP2 antibody, and/or the ability to complement a defect in LAMP2 activity in a cell, tissue, and/or organism.

In some embodiments, the subsequence of the LAMP2 polypeptide comprises a functional fragment of LAMP-2A, LAMP-2B, or LAMP-2C. As used herein, a "functional fragment" refers to polypeptide sharing sequence similarity to a subsequence of a reference LAMP-2A, LAMP-2B, or LAMP-2C and having at least one biological property of LAMP-2A, LAMP-2B, or LAMP-2C. The biological property may include the ability to specifically interact with one or more binding partners, the ability to bind an anti-LAMP2 antibody, and/or the ability to complement a defect in LAMP2 activity in a cell, tissue, and/or organism.

In an embodiment, the transgene is codon-optimized for expression in a human host cell. In an embodiment, the transgene coding sequence is modified, or "codon optimized" to enhance expression by replacing infrequently represented codons with more frequently represented codons. The coding sequence is the portion of the mRNA sequence that encodes the amino acids for translation. During translation, each of 61 trinucleotide codons are translated to one of 20 amino acids, leading to a degeneracy, or redundancy, in the genetic code. However, different cell types, and different animal species, utilize tRNAs (each bearing an anticodon) coding for the same amino acids at different frequencies. When a gene sequence contains codons that are infrequently represented by the corresponding tRNA, the ribosome translation machinery may slow, impeding efficient translation. Expression can be improved via "codon optimization" for a particular species, where the coding sequence is altered to encode the same protein sequence, but utilizing codons that are highly represented,

and/or utilized by highly expressed human proteins (Cid-Arregui et al., 2003; J. Virol. 77: 4928).

In some embodiments, the coding sequence of the transgene is modified to replace codons infrequently expressed in mammal or in primates with codons frequently expressed in primates. For example, in some embodiments, the transgene encodes a polypeptide having at least 85% sequence identity to a reference polypeptide (*e.g.* wild-type LAMP-2B; SEQ ID NO: 16)—for example, at least 90% sequence identity, at least 95% sequence identity, at least 98% identity, or at least 99% identity to the reference polypeptide—wherein at least one codon of the coding sequence has a higher tRNA frequency in humans than the corresponding codon in the sequence disclosed above or herein.

In an embodiment, the transgene comprises fewer alternative open reading frames than SEQ ID: 6. In an embodiment, the transgene is modified to enhance expression by termination or removal of open reading frames (ORFs) that do not encode the desired transgene. An open reading frame (ORF) is the nucleic acid sequence that follows a start codon and does not contain a stop codon. ORFs may be in the forward or reverse orientation, and may be "in frame" or "out of frame" compared with the gene of interest. Such open reading frames have the potential to be expressed in an expression cassette alongside the gene of interest, and could lead to undesired adverse effects. In some embodiments the transgene has been modified to remove open reading frames by further altering codon usage. This is done by eliminating one or more start codons (ATG, TTG, CTG) and/or introducing one or more stop codons (TAG, TAA, or TGA) in reverse orientation or out-of-frame to the desired ORF, while preserving the encoded amino acid sequence and, optionally, maintaining highly utilized codons in the gene of interest (*i.e.*, avoiding codons with frequency < 20%).

In variations of the present disclosure, the transgene coding sequence may be optimized by either of codon optimization and removal of non-transgene ORFs or using both techniques. In some cases, one removes or minimizes non-transgene ORFs after codon optimization in order to remove ORFs introduced during codon optimization.

In an embodiment, the transgene contains fewer CpG sites than SEQ ID: 6. Without being bound by theory, it is believed that the presence of CpG sites in a polynucleotide sequence is associated with the undesirable immunological responses of the host against a viral vector comprising the polynucleotide sequence. In some embodiments, the transgene is

designed to reduce the number of CpG sites. Exemplary methods are provided in U.S. Patent Application Publication No. US20020065236A1.

In an embodiment, the transgene contains fewer cryptic splice sites than SEQ ID: 6. For the optimization, GeneArt® software may be used, *e.g.*, to increase the GC content and/or remove cryptic splice sites in order to avoid transcriptional silencing and, therefore, increase transgene expression. Alternatively, any optimization method known in the art may be used. Removal of cryptic splice sites is described, for example, in International Patent Application Publication No. WO2004015106A1.

Also disclosed herein are expression cassettes and gene therapy vectors encoding LAMP-2B, *e.g.*, a codon-optimized LAMP-2B sequence disclosed herein, comprising: a consensus optimal Kozak sequence, a full-length polyadenylation (polyA) sequence (or substitution of full-length polyA for a truncated polyA), and minimal or no upstream (*i.e.* 5') start codons (*i.e.* ATG sites).

In some cases, the expression cassette contains two or more of a first inverted terminal repeat, an enhancer/promoter region, a consensus optimal Kozak sequence, a transgene (*e.g.*, a transgene encoding a LAMP-2B disclosed herein), a 3' untranslated region including a full-length polyA sequence, and a second inverted terminal repeat.

In an embodiment, the expression cassette comprises a Kozak sequence operatively linked to the transgene. In an embodiment, the Kozak sequence is a consensus optimal Kozak sequence comprising or consisting of SEQ ID NO: 20.

GCCGCCACCATGG (SEQ ID NO: 20)

In various embodiments, the expression cassette comprises an alternative Kozak sequence operatively linked to the transgene. In an embodiment, the Kozak sequence is an alternative Kozak sequence comprising or consisting of any one of SEQ ID NOs. 21-25.

(gcc)gccRccAUGG (SEQ ID NO: 21)

AGNNAUGN (SEQ ID NO: 22)

ANNAUGG (SEQ ID NO: 23)

ACCAUGG (SEQ ID NO: 24)

GACACCAUGG (SEQ ID NO: 25)

In SEQ ID NO: 21, a lower-case letter denotes the most common base at a position where the base can nevertheless vary; an upper-case letter indicates a highly conserved base; 'R' indicates adenine or guanine. In SEQ ID NO: 21, the sequence in parentheses (gcc) is optional. IN SEQ ID NOs: 22-23, 'N' denotes any base.

A variety of sequences can be used in place of this consensus optimal Kozak sequence as the translation-initiation site and it is within the skill of those in the art to identify and test other sequences. *See* Kozak M. An analysis of vertebrate mRNA sequences: intimations of translational control. *J. Cell Biol.* 115 (4): 887–903 (1991).

In an embodiment, the expression cassette comprises a full-length polyA sequence operatively linked to the transgene. In an embodiment, the full-length polyA sequence comprises SEQ ID NO: 26.

TGGCTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTC
TCTCACTCGGAAGGACATATGGGAGGGCAAATCATTTAAAACATCAGAATGAGT
ATTTGGTTTAGAGTTTGGCAACATATGCCCATATGCTGGCTGCCATGAACAAAGG
TTGGCTATAAAGAGGTCATCAGTATATGAAACAGCCCCCTGCTGTCCATTCCTTA
TTCCATAGAAAAGCCTTGACTTGAGGTTAGATTTTTTTTATATTTTGTTTTGTGTT
ATTTTTTTCTTTAACATCCCTAAAATTTTCCTTACATGTTTTACTAGCCAGATTTTT
CCTCCTCTCCTGACTACTCCCAGTCATAGCTGTCCCTCTTCTCTTATGGAGATC
(SEQ ID NO: 26)

Various alternative polyA sequences may be used in expression cassettes of the present disclosure, including without limitation, bovine growth hormone polyadenylation signal (bGHpA) (SEQ ID NO: 27), the SV40 early/late polyadenylation signal (SEQ ID NO: 28), and human growth hormone (HGH) polyadenylation signal (SEQ ID NO: 29).

TCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTC
CTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATT
GCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGG
ACAGCAAGGGGGAGGATTGGGAGGACAATAGCAGGCATGCTGGGGATGCGGTG
GGCTCTATGGCTTCTG (SEQ ID NO: 27)

CAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGT
GAAAAAAATGCTTTATTTGTGAAATTTGTGATGCTATTGCTTTATTTGTAACCATT
ATAAGCTGCAATAAACAAGTTAACAACAACAATTGCATTCATTTTATGTTTCAGG

TTCAGGGGGAGATGTGGGAGGTTTTTTAAAGCAAGTAAAACCTCTACAAATGTG GTA (SEQ ID NO: 28)

CTGCCCGGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAG TTGCCACTCCAGTGCCCACCAGCCTTGTCCTAATAAAATTAAGTTGCATCATTTTG TCTGACTAGGTGTCCTTCTATAATATTATGGGGTGGAGGGGGGTGGTATGGAGCA AGGGGCCCAAGTTGGGAAGAAACCTGTAGGGCCTGC (SEQ ID NO: 29)

In some embodiments, the expression cassette comprises an active fragment of a polyA sequence. In particular embodiments, the active fragment of the polyA sequence comprises or consists of less than 20 base pair (bp), less than 50 bp, less than 100 bp, or less than 150 bp, *e.g.*, of any of the polyA sequences disclosed herein.

In some cases, expression of the transgene is increased by ensuring that the expression cassette does not contain competing ORFs. In an embodiment, the expression cassette comprises no start codon within 20, 30, 40, 50, 60, 70, 80, 90, 100, 200, or 300 base pairs 5' of the start codon of the transgene. In an embodiment, the expression cassette comprises no start codon 5' of the start codon of the transgene. In some embodiments, the expression cassette comprises no alternative transcripts. In some embodiments, the expression cassette comprises no alternative transcripts, except small transcripts, *e.g.* 300 base pairs or less.

In an embodiment, the expression cassette comprises operatively linked, in the 5' to 3' direction, a first inverted terminal repeat, an enhancer/promoter region, introns, a consensus optimal Kozak sequence, the transgene, a 3' untranslated region including a full-length polyA sequence, and a second inverted terminal repeat, where the expression cassette comprises no start codon 5' to the start codon of the transgene.

In an embodiment, the enhancer/promoter region comprises, in the 5' to 3' direction: a CMV IE Enhancer and a Chicken Beta-Actin Promoter. In an embodiment, the enhancer/promoter region comprises a CAG promoter. As used herein "CAG promoter" refers to a polynucleotide sequence comprising a CMV early enhancer element, a chicken beta-actin promoter, the first exon and first intron of the chicken beta-actin gene, and a splice acceptor from the rabbit beta-globin gene.

In an embodiment, the expression cassette shares at least 95% identity to a sequence selected from SEQ ID NOs: 10-12. In an embodiment, the expression cassette shares complete identity to a sequence selected from SEQ ID NOs: 10-12, or shares at least 80%, at least 85%, at least 90%, at least 95%, at least 98%, or at least 99% identity to a sequence

selected from SEQ ID NOs: 10-12. In certain embodiments, the expression cassette comprises one or more modifications as compared to a sequence selected from SEQ ID NOs: 10-12. In particular embodiments, the one or more modifications comprises one or more of: removal of one or more (*e.g.*, all) upstream ATG sequences, replacement of the Kozak sequence with an optimized consensus Kozak sequence or another Kozak sequence, including but not limited to any of those disclosed herein, and/or replacement of the polyadenylation sequence with a full-length polyadenylation sequence or another polyadenylation sequence, including but not limited to any of those disclosed herein. An illustrative configuration of genetic elements within these exemplary expression cassettes is depicted in **FIG. 1**.

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCG
ACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCC
AACTCCATCACTAGGGGTTCCTTGTAGTTAATGATTAACCCGCCATGCTACTTAT
CTACCAGGGTAATGGGGATCCTCTAGAACTATAGCTAGTCGACATTGATTATTGA
CTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGA
GTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGAC
CCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGA
CTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGT
ACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAA
TGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGC
AGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTT
CTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTAT
TTTTTAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGCGCGCGCC
AGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCG
GCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGG
CGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGC
GCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGG
CTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTC
CGGGCTGTAATTAGCGCTTGGTTTAATGACGGCTTGTTTCTTTTCTGTGGCTGCGT
GAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGG
GGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGCGCTGCCC
GGCGGCTGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCAGTGTG
CGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCGA
GGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGGGGGTG

TGGGCGCGTCGGTCGGGCTGCAACCCCCCCTGCACCCCCCTCCCCGAGTTGCTGA
GCACGGCCCGGCTTCGGGTGCGGGGCTCCGTACGGGGCGTGGCGCGGGGCTCGC
CGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGC
CTCGGGCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCCGGAGCGCCGGC
GGCTGTCGAGGCGCGGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAG
AGGGCGCAGGGACTTCCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGG
CGCCGCCGCACCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGG
AAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCCTTCTC
CCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTCGGGGGGGACGG
GGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTA
ACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATT
GTGCTGTCTCATCATTTTGGCAAAGAATTCGAGCGGCCGCCAGCCGCCACCATGG
TCTGCTTCAGACTGTTCCCTGTCCCTGGATCTGGTCTGGTGCTTGTGTGCTTGGTG
CTGGGTGCTGTGAGATCCTATGCCCTTGAGCTGAACCTGACTGACTCAGAAAATG
CCACTTGCCTGTATGCCAAGTGGCAGATGAACTTCACTGTGAGATATGAGACTAC
CAACAAGACCTACAAGACTGTGACCATCTCAGACCATGGCACTGTCACCTACAA
TGGATCAATCTGTGGTGATGATCAGAATGGCCCAAAGATAGCAGTGCAGTTTGG
GCCCGGTTTTTCCTGGATTGCTAACTTCACCAAGGCAGCCTCCACCTACAGCATT
GACTCAGTCAGCTTCAGCTACAACACTGGGGATAACACCACCTTCCCTGACGCAG
AGGACAAGGGAATCCTTACTGTGGACGAACTCCTGGCAATCAGAATCCCCCTTA
ACGACCTGTTCAGATGCAACTCCCTTTCAACCCTTGAAAAGAATGATGTGGTGCA
ACACTATTGGGACGTCCTGGTGCAAGCCTTTGTGCAGAATGGGACAGTGAGTAC
CAACGAGTTCCTCTGTGACAAGGACAAGACCAGCACTGTGGCCCCCACTATCCA
CACCACTGTGCCCAGCCCTACCACTACCCCCACCCCTAAAGAGAAGCCAGAAGC
TGGAACCTACTCAGTCAACAATGGAAATGACACATGCCTCCTTGCCACCATGGG
ACTGCAGCTGAACATCACTCAGGACAAGGTGGCCTCAGTGATTAACATCAACCC
TAACACCACTCATAGCACTGGGAGCTGCAGATCACATACAGCTCTGCTGAGGCTC
AACTCCTCCACCATCAAGTACCTGGACTTTGTGTTTGCTGTGAAGAATGAGAACA
GGTTCTACCTCAAGGAAGTGAACATTTCCATGTACCTGGTCAATGGTTCAGTGTT
CTCTATTGCCAACAACAATCTGAGCTACTGGGATGCACCCCTGGGATCCTCCTAC
ATGTGCAACAAGGAGCAGACTGTGAGTGTGTCAGGTGCTTTTCAGATCAACACTT
TTGACCTGAGGGTGCAGCCCTTCAATGTGACTCAGGGAAAGTACTCCACTGCACA
AGAGTGTTCCTTGGATGATGACACTATCCTCATCCCCATTATTGTGGGAGCTGGA
CTGTCAGGATTGATTATAGTGATTGTGATTGCTTATGTGATTGGAAGGAGAAAGA

GCTATGCTGGCTACCAGACCCTGTAAAAGGGCGAATTCCAGCACACGCGTCCTA
GGAGCTCGAGTACTACTGGCGGCCGTTACTAGTGGATCCGCGGTACAAGTAAGC
ATGCAAGCTTCGAGGACGGGGTGAACTACGCCTGAATCAAGCTTATCGATAAAT
TCGAGCATCTTACCGCCATTTATTCCCATATTTGTTCTGTTTTTCTTGATTTGGGTA
TACATTTAAATGTTAATAAAACAAAATGGTGGGGCAATCATTTACATTTTTAGGG
ATATGTAATTACTAGTTCAGGTGTATTGCCACAAGACAAACATGTTAAGAAACTT
TCCCGTTATTTACGCTCTGTTCCTGTTAATCAACCTCTGGATTACAAAATTTGTGA
AAGATTGACTGATATTCTTAACTATGTTGCTCCTTTTACGCTGTGTGGATATGCTG
CTTTAATGCCTCTGTATCATGCTATTGCTTCCCGTACGGCTTTCGTTTTCTCCTCCT
TGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCCGTCAA
CGTGGCGTGGTGTGCTCTGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCATTG
CCACCACCTGTCAACTCCTTTCTGGGACTTTCGCTTTCCCCCTCCCGATCGCCACG
GCAGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTAGGTTGCTGG
GCACTGATAATTCCGTGGTGTTGTCGGGGAAGGGCCTCGATACCGTCGATATCGA
TCCTGGCTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGT
GTCTCTCACTCGGAAGGACATATGGGAGGGCAAATCATTTAAAACATCAGAATG
AGTATTTGGTTTAGAGTTTGGCAACATATGCCCATATGCTGGCTGCCATGAACAA
AGGTTGGCTATAAAGAGGTCATCAGTATATGAAACAGCCCCCTGCTGTCCATTCC
TTATTCCATAGAAAAGCCTTGACTTGAGGTTAGATTTTTTTTATATTTTGTTTTGT
GTTATTTTTTTCTTTAACATCCCTAAAATTTTCCTTACATGTTTTACTAGCCAGATT
TTTCCTCCTCTCCTGACTACTCCCAGTCATAGCTGTCCCTCTTCTCTTATGGAGAT
CGAAGCAATTCGTTGATCTGAATTTCGACCACCCATAATAGATCTCCCATTACCC
TGGTAGATAAGTAGCATGGCGGGTTAATCATTAACTACAAGGAACCCCTAGTGA
TGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACC
AAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGC
GCGCAG (SEQ ID NO: 10)

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCG
ACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCC
AACTCCATCACTAGGGGTTCCTTGTAGTTAATGATTAACCCGCCATGCTACTTAT
CTACCAGGGTAATGGGGATCCTCTAGAACTATAGCTAGTCGACATTGATTATTGA
CTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGA
GTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGAC
CCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGA

CTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGT
ACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAA
TGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGC
AGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTT
CTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTAT
TTTTTAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGCGCGCGCC
AGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCG
GCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGG
CGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGC
GCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGG
CTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTC
CGGGCTGTAATTAGCGCTTGGTTTAATGACGGCTTGTTTCTTTTCTGTGGCTGCGT
GAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGG
GGTGCGTGCGTGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGCGCTGCCC
GGCGGCTGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCAGTGTG
CGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCGA
GGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGGGGGTG
TGGGCGCGTCGGTCGGGCTGCAACCCCCCCTGCACCCCCCTCCCCGAGTTGCTGA
GCACGGCCCGGCTTCGGGTGCGGGGCTCCGTACGGGGCGTGGCGCGGGGCTCGC
CGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGC
CTCGGGCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCCGGAGCGCCGGC
GGCTGTCGAGGCGCGGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAG
AGGGCGCAGGGACTTCCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGG
CGCCGCCGCACCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGG
AAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCCTTCTC
CCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTCGGGGGGGACGG
GGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTA
ACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATT
GTGCTGTCTCATCATTTTGGCAAAGAATTCGAGCGGCCGCCAGCCGCCACCATGG
TGTGCTTTAGACTGTTTCCTGTGCCTGGTTCAGGGCTGGTCCTGGTCTGTCTGGTG
CTGGGGGCTGTCAGAAGCTATGCCTTGGAGCTGAACCTCACTGATAGTGAAAAT
GCCACTTGTCTGTATGCTAAGTGGCAGATGAACTTCACTGTGAGATATGAAACCA
CCAACAAGACTTACAAAACAGTGACCATCTCAGATCATGGAACTGTGACCTACA
ACGGCAGCATTTGTGGAGACGACCAGAACGGACCAAAAATCGCTGTCCAATTTG

GGCCTGGATTCTCCTGGATTGCCAATTTCACTAAAGCTGCCTCCACATATTCAATT
GACTCAGTGTCCTTCTCCTACAACACTGGGGACAACACTACTTTCCCTGATGCTG
AAGATAAGGGAATCTTGACAGTGGATGAGCTGCTGGCTATCAGGATCCCTTTGA
ATGACCTGTTTAGGTGTAATTCACTGAGCACTCTGGAGAAGAACGACGTGGTGC
AGCACTACTGGGACGTGCTGGTGCAGGCCTTTGTGCAGAACGGCACTGTGTCCAC
CAACGAATTCCTGTGTGATAAGGACAAAACTTCCACTGTGGCACCTACAATTCAC
ACTACTGTGCCTTCACCTACCACCACTCCAACTCCAAAGGAAAAGCCTGAAGCA
GGAACCTACTCTGTGAACAATGGCAATGATACCTGTCTGTTGGCCACCATGGGCC
TCCAACTGAACATTACTCAGGACAAGGTGGCCTCAGTGATTAACATTAACCCCAA
CACTACCCACTCCACTGGCAGCTGTAGATCACACACAGCCTTGCTCAGACTGAAT
AGCAGCACCATCAAGTATTTGGATTTTGTGTTTGCAGTGAAGAATGAAAACAGGT
TCTACCTGAAGGAAGTCAACATCTCAATGTACCTGGTGAACGGCTCAGTGTTCAG
CATTGCCAACAACAACCTCTCCTATTGGGACGCTCCACTGGGGAGCAGCTACATG
TGTAACAAGGAACAGACTGTGTCAGTGTCAGGAGCCTTCCAGATTAACACCTTTG
ATCTGAGGGTCCAACCCTTTAATGTCACTCAAGGAAAGTATAGCACTGCCCAGG
AGTGCTCCCTGGATGATGACACCATTCTGATTCCAATCATTGTGGGTGCAGGACT
TTCTGGGCTTATTATTGTGATTGTGATTGCCTATGTGATTGGCAGAAGGAAATCCT
ATGCAGGGTACCAAACTCTGTAAAAGGGCGAATTCCAGCACACGCGTCCTAGGA
GCTCGAGTACTACTGGCGGCCGTTACTAGTGGATCCGCGGTACAAGTAAGCATG
CAAGCTTCGAGGACGGGGTGAACTACGCCTGAATCAAGCTTATCGATAAATTCG
AGCATCTTACCGCCATTTATTCCCATATTTGTTCTGTTTTTCTTGATTTGGGTATAC
ATTTAAATGTTAATAAAACAAAATGGTGGGGCAATCATTTACATTTTTAGGGATA
TGTAATTACTAGTTCAGGTGTATTGCCACAAGACAAACATGTTAAGAAACTTTCC
CGTTATTTACGCTCTGTTCCTGTTAATCAACCTCTGGATTACAAAATTTGTGAAAG
ATTGACTGATATTCTTAACTATGTTGCTCCTTTTACGCTGTGTGGATATGCTGCTT
TAATGCCTCTGTATCATGCTATTGCTTCCCGTACGGCTTTCGTTTTCTCCTCCTTGT
ATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCCGTCAACGT
GGCGTGGTGTGCTCTGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCATTGCCA
CCACCTGTCAACTCCTTTCTGGGACTTTCGCTTTCCCCCTCCCGATCGCCACGGCA
GAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTAGGTTGCTGGGCA
CTGATAATTCCGTGGTGTTGTCGGGGAAGGGCCTCGATACCGTCGATATCGATCC
TGGCTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTC
TCTCACTCGGAAGGACATATGGGAGGGCAAATCATTTAAAACATCAGAATGAGT
ATTTGGTTTAGAGTTTGGCAACATATGCCCATATGCTGGCTGCCATGAACAAAGG

TTGGCTATAAAGAGGTCATCAGTATATGAAACAGCCCCCTGCTGTCCATTCCTTA
TTCCATAGAAAAGCCTTGACTTGAGGTTAGATTTTTTTTATATTTTGTTTTGTGTT
ATTTTTTTCTTTAACATCCCTAAAATTTTCCTTACATGTTTTACTAGCCAGATTTTT
CCTCCTCTCCTGACTACTCCCAGTCATAGCTGTCCCTCTTCTCTTATGGAGATCGA
AGCAATTCGTTGATCTGAATTTCGACCACCCATAATAGATCTCCCATTACCCTGG
TAGATAAGTAGCATGGCGGGTTAATCATTAACTACAAGGAACCCCTAGTGATGG
AGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACCAAA
GGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGCGCG
CAG (SEQ ID NO: 11)

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCG
ACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCC
AACTCCATCACTAGGGGTTCCTTGTAGTTAATGATTAACCCGCCATGCTACTTAT
CTACCAGGGTAATGGGGATCCTCTAGAACTATAGCTAGTCGACATTGATTATTGA
CTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGA
GTTCCGCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGAC
CCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGGA
CTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGT
ACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAA
TGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGC
AGTACATCTACGTATTAGTCATCGCTATTACCATGGTCGAGGTGAGCCCCACGTT
CTGCTTCACTCTCCCCATCTCCCCCCCCTCCCCACCCCCAATTTTGTATTTATTTAT
TTTTTAATTATTTTGTGCAGCGATGGGGGCGGGGGGGGGGGGGGGCGCGCGCC
AGGCGGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCG
GCGGCAGCCAATCAGAGCGGCGCGCTCCGAAAGTTTCCTTTTATGGCGAGGCGG
CGGCGGCGGCGGCCCTATAAAAAGCGAAGCGCGCGGCGGGCGGGAGTCGCTGC
GCGCTGCCTTCGCCCCGTGCCCCGCTCCGCCGCCGCCTCGCGCCGCCCGCCCCGG
CTCTGACTGACCGCGTTACTCCCACAGGTGAGCGGGCGGGACGGCCCTTCTCCTC
CGGGCTGTAATTAGCGCTTGGTTTAATGACGGCTTGTTTCTTTTCTGTGGCTGCGT
GAAAGCCTTGAGGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGG
GGTGCGTGCGTGTGTGTGCGTGGGGAGCGCCGCGTGCGGCTCCGCGCTGCCC
GGCGGCTGTGAGCGCTGCGGGCGCGGCGCGGGGCTTTGTGCGCTCCGCAGTGTG
CGCGAGGGGAGCGCGGCCGGGGGCGGTGCCCCGCGGTGCGGGGGGGGCTGCGA
GGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGGGGGTGAGCAGGGGGTG

TGGGCGCGTCGGTCGGGCTGCAACCCCCCCTGCACCCCCCTCCCCGAGTTGCTGA
GCACGGCCCGGCTTCGGGTGCGGGGCTCCGTACGGGGCGTGGCGCGGGGCTCGC
CGTGCCGGGCGGGGGGTGGCGGCAGGTGGGGGTGCCGGGCGGGGCGGGGCCGC
CTCGGGCCGGGGAGGGCTCGGGGGAGGGGCGCGGCGGCCCCCGGAGCGCCGGC
GGCTGTCGAGGCGCGGCGAGCCGCAGCCATTGCCTTTTATGGTAATCGTGCGAG
AGGGCGCAGGGACTTCCTTTGTCCCAAATCTGTGCGGAGCCGAAATCTGGGAGG
CGCCGCCGCACCCCCTCTAGCGGGCGCGGGGCGAAGCGGTGCGGCGCCGGCAGG
AAGGAAATGGGCGGGGAGGGCCTTCGTGCGTCGCCGCGCCGCCGTCCCCTTCTC
CCTCTCCAGCCTCGGGGCTGTCCGCGGGGGGACGGCTGCCTTCGGGGGGGACGG
GGCAGGGCGGGGTTCGGCTTCTGGCGTGTGACCGGCGGCTCTAGAGCCTCTGCTA
ACCATGTTCATGCCTTCTTCTTTTTCCTACAGCTCCTGGGCAACGTGCTGGTTATT
GTGCTGTCTCATCATTTTGGCAAAGAATTCGAGCGGCCGCCAGCCGCCACCATGG
TCTGTTTTAGGCTGTTCCCTGTCCCTGGTTCAGGACTGGTCTTAGTGTGTCTGGTG
CTTGGAGCTGTCAGAAGCTATGCCCTGGAGCTGAACCTGACTGACTCAGAAAAT
GCCACTTGCCTGTATGCCAAGTGGCAGATGAACTTCACTGTCAGATATGAAACCA
CCAACAAGACCTATAAGACTGTGACCATCTCAGACCATGGCACTGTGACTTACA
ATGGGTCAATTTGTGGAGATGACCAGAATGGCCCTAAGATAGCTGTCCAGTTTGG
TCCAGGATTCAGCTGGATTGCCAACTTCACCAAGGCAGCCAGCACCTACAGCATT
GACTCTGTGTCCTTCTCCTACAACACAGGAGACAACACCACTTTCCCTGATGCAG
AGGACAAAGGTATCCTGACTGTGGATGAGTTGCTGGCAATCAGGATCCCACTGA
ACGATCTGTTCAGGTGCAACTCACTGTCCACTCTGGAAAAGAATGATGTGGTGCA
GCACTATTGGGATGTGCTAGTCCAGGCCTTTGTCCAGAATGGGACTGTGTCAACT
AATGAGTTCCTGTGTGACAAGGACAAGACAAGCACTGTAGCCCCCACTATCCAT
ACCACAGTACCTAGCCCCACCACTACTCCAACCCCCAAGGAGAAGCCTGAGGCT
GGCACCTACTCAGTGAACAATGGGAATGACACCTGTTTGCTGGCCACTATGGGA
CTCCAACTGAACATCACCCAGGACAAAGTGGCCTCTGTGATCAATATCAATCCCA
ACACCACCCACAGCACTGGGTCCTGCAGAAGCCACACTGCCCTCCTGAGGCTCA
ACTCATCAACTATCAAGTACTTGGATTTTGTGTTTGCAGTGAAGAATGAGAACAG
ATTCTACCTCAAAGAGGTCAACATTTCAATGTACCTGGTGAATGGGAGTGTGTTC
TCCATTGCTAACAACAACCTGAGCTACTGGGATGCCCCTCTGGGCTCCTCATACA
TGTGCAACAAGGAACAGACTGTGAGTGTGTCAGGGGCCTTCCAGATCAACACTT
TTGACCTGAGAGTGCAGCCCTTTAATGTGACACAGGGAAAGTACAGCACTGCTC
AGGAGTGCAGCCTGGATGATGACACTATCCTGATCCCTATCATTGTGGGGGCAG
GCCTGTCTGGACTCATTATTGTGATTGTGATTGCCTATGTGATAGGGAGAAGGAA

GTCTTATGCTGGATACCAGACCCTGTAAAAGGGCGAATTCCAGCACACGCGTCCT
AGGAGCTCGAGTACTACTGGCGGCCGTTACTAGTGGATCCGCGGTACAAGTAAG
CATGCAAGCTTCGAGGACGGGGTGAACTACGCCTGAATCAAGCTTATCGATAAA
TTCGAGCATCTTACCGCCATTTATTCCCATATTTGTTCTGTTTTTCTTGATTTGGGT
ATACATTTAAATGTTAATAAAACAAAATGGTGGGGCAATCATTTACATTTTTAGG
GATATGTAATTACTAGTTCAGGTGTATTGCCACAAGACAAACATGTTAAGAAACT
TTCCCGTTATTTACGCTCTGTTCCTGTTAATCAACCTCTGGATTACAAAATTTGTG
AAAGATTGACTGATATTCTTAACTATGTTGCTCCTTTTACGCTGTGTGGATATGCT
GCTTTAATGCCTCTGTATCATGCTATTGCTTCCCGTACGGCTTTCGTTTTCTCCTCC
TTGTATAAATCCTGGTTGCTGTCTCTTTATGAGGAGTTGTGGCCCGTTGTCCGTCA
ACGTGGCGTGGTGTGCTCTGTGTTTGCTGACGCAACCCCCACTGGCTGGGGCATT
GCCACCACCTGTCAACTCCTTTCTGGGACTTTCGCTTTCCCCCTCCCGATCGCCAC
GGCAGAACTCATCGCCGCCTGCCTTGCCCGCTGCTGGACAGGGGCTAGGTTGCTG
GGCACTGATAATTCCGTGGTGTTGTCGGGGAAGGGCCTCGATACCGTCGATATCG
ATCCTGGCTAATAAAGGAAATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTG
TGTCTCTCACTCGGAAGGACATATGGGAGGGCAAATCATTTAAAACATCAGAAT
GAGTATTTGGTTTAGAGTTTGGCAACATATGCCCATATGCTGGCTGCCATGAACA
AAGGTTGGCTATAAAGAGGTCATCAGTATATGAAACAGCCCCCTGCTGTCCATTC
CTTATTCCATAGAAAAGCCTTGACTTGAGGTTAGATTTTTTTTATATTTTGTTTTGT
GTTATTTTTTTCTTTAACATCCCTAAAATTTTCCTTACATGTTTTACTAGCCAGATT
TTTCCTCCTCTCCTGACTACTCCCAGTCATAGCTGTCCCTCTTCTCTTATGGAGAT
CGAAGCAATTCGTTGATCTGAATTTCGACCACCCATAATAGATCTCCCATTACCC
TGGTAGATAAGTAGCATGGCGGGTTAATCATTAACTACAAGGAACCCCTAGTGA
TGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCACTGAGGCCGGGCGACC
AAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTCAGTGAGCGAGCGAGC
GCGCAG (SEQ ID NO: 12).

In an embodiment, the vector is an adeno-associated virus (AAV) vector. In an embodiment, the expression cassette comprises ITR sequences selected from SEQ ID NOs: 13 and 14.

CTGCGCGCTCGCTCGCTCACTGAGGCCGCCCGGGCAAAGCCCGGGCGTCGGGCG
ACCTTTGGTCGCCCGGCCTCAGTGAGCGAGCGAGCGCGCAGAGAGGGAGTGGCC
AACTCCATCACTAGGGGTTCCT (SEQ ID NO: 13)

AGGAACCCCTAGTGATGGAGTTGGCCACTCCCTCTCTGCGCGCTCGCTCGCTCAC
TGAGGCCGGGCGACCAAAGGTCGCCCGACGCCCGGGCTTTGCCCGGGCGGCCTC
AGTGAGCGAGCGAGCGCGCAG (SEQ ID NO: 14)

In related embodiments, the disclosure provides gene therapy vectors comprising an expression cassette disclosed herein. Generally, the gene therapy vectors described herein comprise an expression cassette comprising a polynucleotide encoding one or more isoforms of lysosome-associated membrane protein 2 (LAMP-2), that allows for the expression of LAMP-2 to partially or wholly rectify deficient LAMP-2 protein expression levels and/or autophagic flux in a subject in need thereof (*e.g.*, a subject having Danon disease or another disorder characterized by deficient autophagic flux at least in part due to deficient LAMP-2 expression).

LAMP-2A protein sequence

MVCFRLFPVPGSGLVLVCLVLGAVRSYALELNLTDSENATCLYAKWQMNFTVRYET
TNKTYKTVTISDHGTVTYNGSICGDDQNGPKIAVQFGPGFSWIANFTKAASTYSIDSV
SFSYNTGDNTTFPDAEDKGILTVDELLAIRIPLNDLFRCNSLSTLEKNDVVQHYWDVL
VQAFVQNGTVSTNEFLCDKDKTSTVAPTIHTTVPSPTTTPTPKEKPEAGTYSVNNGN
DTCLLATMGLQLNITQDKVASVININPNTTHSTGSCRSHTALLRLNSSTIKYLDFVFA
VKNENRFYLKEVNISMYLVNGSVFSIANNNLSYWDAPLGSSYMCNKEQTVSVSGAF
QINTFDLRVQPFNVTQGKYSTAQDCSADDDNFLVPIAVGAALAGVLILVLLAYFIGL
KHHHAGYEQF (SEQ ID NO: 15)

LAMP-2B protein sequence

MVCFRLFPVPGSGLVLVCLVLGAVRSYALELNLTDSENATCLYAKWQMNFTVRYET
TNKTYKTVTISDHGTVTYNGSICGDDQNGPKIAVQFGPGFSWIANFTKAASTYSIDSV
SFSYNTGDNTTFPDAEDKGILTVDELLAIRIPLNDLFRCNSLSTLEKNDVVQHYWDVL
VQAFVQNGTVSTNEFLCDKDKTSTVAPTIHTTVPSPTTTPTPKEKPEAGTYSVNNGN
DTCLLATMGLQLNITQDKVASVININPNTTHSTGSCRSHTALLRLNSSTIKYLDFVFA
VKNENRFYLKEVNISMYLVNGSVFSIANNNLSYWDAPLGSSYMCNKEQTVSVSGAF
QINTFDLRVQPFNVTQGKYSTAQECSLDDDTILIPIIVGAGLSGLIIVIVIAYVIGRRKSY
AGYQTL (SEQ ID NO: 16)

LAMP-2C protein sequence

MVCFRLFPVPGSGLVLVCLVLGAVRSYALELNLTDSENATCLYAKWQMNFTVRYET
TNKTYKTVTISDHGTVTYNGSICGDDQNGPKIAVQFGPGFSWIANFTKAASTYSIDSV

SFSYNTGDNTTFPDAEDKGILTVDELLAIRIPLNDLFRCNSLSTLEKNDVVQHYWDVL
VQAFVQNGTVSTNEFLCDKDKTSTVAPTIHTTVPSPTTTPTPKEKPEAGTYSVNNGN
DTCLLATMGLQLNITQDKVASVININPNTTHSTGSCRSHTALLRLNSSTIKYLDFVFA
VKNENRFYLKEVNISMYLVNGSVFSIANNNLSYWDAPLGSSYMCNKEQTVSVSGAF
QINTFDLRVQPFNVTQGKYSTAEECSADSDLNFLIPVAVGVALGFLIIVVFISYMIGRR
KSRTGYQSV (SEQ ID NO: 17)

In particular embodiments, the expression cassette comprises a polynucleotide sequence encoding LAMP-2 disclosed herein, *e.g.*, SEQ ID NOs: 15-17 or a sequence having at least 90%, at least 95%, at least 98%, or at least 99% identity to any of SEQ ID NOs: 15-17. The gene therapy vectors can be viral or non-viral vectors. Illustrative non-viral vectors include, *e.g.*, naked DNA, cationic liposome complexes, cationic polymer complexes, cationic liposome-polymer complexes, and exosomes. Examples of viral vector include, but are not limited, to adenoviral, retroviral, lentiviral, herpesvirus and adeno-associated virus (AAV) vectors.

In some embodiments, the expression cassette comprising a polynucleotide sequence encoding one or more, two or more, or all three of SEQ ID NOs: 15-17. In some embodiments, the polynucleotide sequence comprising the native introns of the LAMP-2 gene, enabling expression of more than one isoform in the same cell using one vector. In some embodiments, artificial introns, splice acceptors, and/or splice donors are using to optimize the length of the polynucleotide and/or optimize the ratio of isoforms expressed by the polynucleotide encoding two or more, or all three of SEQ ID NOs: 15-17.

In some embodiments, the expression cassette, AAV capsid gene, and/or helper genes are delivered to cells using transduction, transfection, electroporation, lipofection, and any other methods known in the art. In some embodiments, the the expression cassette, AAV capsid gene, and/or helper genes are delivered in a liposome or a lipid nanoparticle (LNP). The expression cassette, AAV capsid gene, and/or helper genes may be provide as DNA, *e.g.* on one or more plasmids, bacmids, or other DNA molecules. In some embodiments, expression cassette, AAV capsid gene, and/or helper genes are delivery as RNA molecules. In some embodiments, the RNA molecules comprise one or more mRNA molecules, *e.g.*, one or more *in vitro* transcribed mRNA molecules. In some embodiments, the mRNA molecules are modified mRNA molecules. Illustrative modifications include lock nucleic acids, phoshothiolate linkages, and modified nucleosides (*e.g.* pseudouridine, 5-methylcytosine, or 5-methylcytidine). In some embodiments, the modified mRNA comprises a cap, *e.g.* an

ARCA cap. The expression cassette, AAV capsid gene, and/or helper genes may be delivered *in vitro* or *in vivo*. In some embodiments, the AAV capsid gene comprises one or more of an AAV9 capsid gene and an AAVrh74 capsid gene.

Gene delivery viral vectors useful in the practice of the present invention can be constructed utilizing methodologies well known in the art of molecular biology. Typically, viral vectors carrying transgenes are assembled from polynucleotides encoding the transgene, suitable regulatory elements and elements necessary for production of viral proteins, which mediate cell transduction. Such recombinant viruses may be produced by techniques known in the art, *e.g.*, by transfecting packaging cells or by transient transfection with helper plasmids or viruses. Typical examples of virus packaging cells include but are not limited to HeLa cells, SF9 cells (optionally with a baculovirus helper vector), 293 cells, etc. A Herpesvirus-based system can be used to produce AAV vectors, as described in US20170218395A1. Detailed protocols for producing such replication-defective recombinant viruses may be found for instance in W095/14785, W096/22378, U.S. Pat. No. 5,882,877, U.S. Pat. No. 6,013,516, U.S. Pat. No. 4,861,719, U.S. Pat. No. 5,278,056 and W094/19478, the complete contents of each of which is hereby incorporated by reference.

AAV is a 4.7 kb, single stranded DNA virus. Recombinant vectors based on AAV are associated with excellent clinical safety, since wild-type AAV is nonpathogenic and has no etiologic association with any known diseases. In addition, AAV offers the capability for highly efficient gene delivery and sustained transgene expression in numerous tissues. By an "AAV vector" is meant a vector derived from an adeno-associated virus serotype, including without limitation, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAVrh.10, AAVrh74, *etc*. AAV vectors can have one or more of the AAV wild-type genes deleted in whole or part, *e.g.*, the rep and/or cap genes, but retain functional flanking inverted terminal repeat (ITR) sequences. Functional ITR sequences are necessary for the rescue, replication and packaging of the AAV virion. Thus, an AAV vector is defined herein to include at least those sequences required in cis for replication and packaging (*e.g.*, functional ITRs) of the virus. The ITRs need not be the wild-type nucleotide sequences, and may be altered, *e.g.* by the insertion, deletion or substitution of nucleotides, as long as the sequences provide for functional rescue, replication and packaging. AAV vectors may comprise other modifications, including but not limited to one or more modified capsid proteins (*e.g.*, VP1, VP2 and/or VP3). For example, a capsid protein may be modified to alter tropism and/or reduce immunogenicity. AAV expression vectors are constructed using

known techniques to at least provide as operatively linked components in the direction of transcription, control elements including a transcriptional initiation region, the DNA of interest (*i.e.* the LAMP-2 gene) and a transcriptional termination region.

Adeno-associated virus (AAV) is single stranded DNA virus. The AAV genome is built of single-stranded deoxyribonucleic acid (ssDNA), either positive- or negative-sensed, which is about 4.7 kilobase long. The genome comprises inverted terminal repeats (ITRs) at both ends of the DNA strand, and two open reading frames (ORFs): *rep* and *cap*. The first, *rep*, is composed of four overlapping genes encoding Rep proteins required for the AAV life cycle, and the second, *cap*, encodes three capsid proteins: VP1, VP2 and VP3. The *cap* gene is expressed as a messenger RNA (mRNA) from the p40 promoter of AAV. The mRNA is alternatively spliced into 2.3 kb and 2.6 kb transcripts, with the 2.3 kb transcript being more abundant. VP1 is expressed only from the 2.6 kb transcript and the VP1 protein is 87 kilodaltons (kDa) in molecular weight. VP2 is expressed from an open reading frame that begins with an ACG codon, rather than a canonical AUG codon, due to the presence of an optimal Kozak sequence for translation initiation. VP2 is 72 kDa. VP3, only 62 kDa, is expressed from the ATG sequence presence in the 2.3 kb transcript, as well as the 2.6 kb transcript. The relative abundances of VP1:VP2:VP3 are 1:1:10. VP1, VP2, and VP3 interact together to form a capsid of an icosahedral symmetry.

Recombinant vectors based on AAV are associated with excellent clinical safety, since wild-type AAV is nonpathogenic and has no etiologic association with any known diseases. In addition, AAV offers the capability for highly efficient gene delivery and sustained transgene expression in numerous tissues. Various serotypes of AAV are known, including, AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAV9, AAV10, AAVrh.10, AAVrh74, *etc.*. AAV vectors can have one or more of the AAV wild-type genes deleted in whole or part, *e.g.*, the *rep* and/or *cap* genes, but retain functional flanking inverted terminal repeat (ITR) sequences. The serotype of a recombinant AAV vector is determined by its capsid. International Patent Publication No. WO2003042397A2 discloses various capsid sequences including those of AAV1, AAV2, AAV3, AAV8, AAV9, and AAVrh10. International Patent Publication No. WO2013078316A1 discloses the polypeptide sequence of the VP1 from AAVrh74. Numerous diverse naturally occurring or genetically modified AAV capsid sequences are known in the art.

The present disclosure also provides pharmaceutical compositions comprising an expression cassette or vector (*e.g.*, gene therapy vector) disclosed herein and one or more

pharmaceutically acceptable carriers, diluents or excipients. In particular embodiments, the pharmaceutical composition comprises an AAV vector comprising an expression cassette disclosed herein, *e.g.*, wherein the expression cassette comprises a codon-optimized transgene encoding LAMP-2B, *e.g.*, any of SEQ ID NOs: 7-9. Provided are pharmaceutical compositions, *e.g.*, for use in preventing or treating a disorder characterized by deficient autophagic flux (*e.g.*, Danon disease) which comprises a therapeutically effective amount of a vector that comprises a nucleic acid sequence of a polynucleotide that encodes one or more isoforms of LAMP-2.

The pharmaceutical compositions that contain the expression cassette or vector may be in any form that is suitable for the selected mode of administration, for example, for intraventricular, intramyocardial, intracoronary, intravenous, intra-arterial, intra-renal, intraurethral, epidural or intramuscular administration. The gene therapy vector comprising a polynucleotide encoding one or more LAMP-2 isoforms can be administered, as sole active agent, or in combination with other active agents, in a unit administration form, as a mixture with conventional pharmaceutical supports, to animals and human beings. In some embodiments, the pharmaceutical composition comprises cells transduced *ex vivo* with any of the gene therapy vectors of the disclosure.

In various embodiments, the pharmaceutical compositions contain vehicles (*e.g.*, carriers, diluents and excipients) that are pharmaceutically acceptable for a formulation capable of being injected. These may be in particular isotonic, sterile, saline solutions (monosodium or disodium phosphate, sodium, potassium, calcium or magnesium chloride and the like or mixtures of such salts). Illustrative pharmaceutical forms suitable for injectable use include, *e.g.*, sterile aqueous solutions or dispersions; formulations including sesame oil, peanut oil or aqueous propylene glycol; and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersions.

In another aspect, the disclosure provides methods of preventing, mitigating, ameliorating, reducing, inhibiting, eliminating and/or reversing one or more symptoms of Danon disease or another autophagy disorder in a subject in need thereof, comprising administering to the subject a gene therapy vector of the disclosure. The term “Danon disease” refers to an X-linked dominant skeletal and cardiac muscle disorder with multisystem clinical manifestations. Danon disease mutations lead to an absence of lysosome-associated membrane protein 2 (LAMP-2) protein expression. Major clinical features include skeletal and cardiac myopathy, cardiac conduction abnormalities, cognitive

difficulties, and retinal disease. Men are typically affected earlier and more severely than women.

In an embodiment, the vector is administered via a route selected from the group consisting of parenteral, intravenous, intra-arterial, intracardiac, intracoronary, intramyocardial, intrarenal, intraurethral, epidural, and intramuscular. In an embodiment, the vector is administered multiple times. In an embodiment, the vector is administered by intramuscular injection of the vector. In an embodiment, the vector is administered by injection of the vector into skeletal muscle. In an embodiment, the expression cassette comprises a muscle-specific promoter, optionally a muscle creatine kinase (MCK) promoter or a MCK/SV40 hybrid promoter as described in Takeshita et al. Muscle creatine kinase/SV40 hybrid promoter for muscle-targeted long-term transgene expression. *Int J Mol Med.* 2007 Feb;19(2):309-15. In an embodiment, the vector is administered by intracardiac injection.

In an embodiment, the disclosure provides a method of treating a disease or disorder, optionally Danon disease, in a subject in need thereof, comprising contacting cells with a gene therapy vector according to the present disclosure and administering the cells to the subject. In an embodiment, the cells are stem cells, optionally pluripotent stem cells. In an embodiment, the stem cells are capable of differentiation into cardiac tissue. In an embodiment, the stem cells are capable of differentiation into muscle tissue, *e.g.*, cardiac muscle tissue and/or skeletal muscle tissue. In an embodiment, the stem cells are autologous. In an embodiment, the stem cells are induced pluripotent stem cells (iPSCs).

In an embodiment, the autophagy disorder is selected from the group consisting of end-stage heart failure, myocardial infarction, drug toxicities, diabetes, end-stage renal failure, and aging. In an embodiment, the subject is a mammal, *e.g.*, a human. In an embodiment, the subject is exhibiting symptoms of Danon disease or another autophagy disorder. In an embodiment, the subject has been identified as having reduced or non-detectable LAMP-2 expression. In an embodiment, the subject has been identified as having a mutated LAMP-2 gene.

Subjects/patients amenable to treatment using the methods described herein include individuals at risk of a disease or disorder characterized by insufficient autophagic flux (*e.g.*, Danon disease as well as other known disorders of autophagy including, but not limited to, systolic and diastolic heart failure, myocardial infarction, drug toxicities (for example,

anthracyclines, chloroquine, and its derivatives), diabetes, end-stage renal disease, and aging) but not showing symptoms, as well as subjects presently showing symptoms. Such subject may have been identified as having a mutated LAMP-2 gene or as having reduced or non-detectable levels of LAMP-2 expression.

In some embodiments, the subject is exhibiting symptoms of a disease or disorder characterized by insufficient autophagic flux (*e.g.*, Danon disease as well as other known disorders of autophagy including, but not limited to, systolic and diastolic heart failure, myocardial infarction, drug toxicities, diabetes, end-stage renal disease, and aging). The symptoms may be actively manifesting, or may be suppressed or controlled (*e.g.*, by medication) or in remission. The subject may or may not have been diagnosed with the disorder, *e.g.*, by a qualified physician.

Definitions

The terms "lysosome-associated membrane protein 2" and "LAMP-2" interchangeably refer to nucleic acids and polypeptide polymorphic variants, alleles, mutants, and interspecies homologs that: (1) have an amino acid sequence that has greater than about 90% amino acid sequence identity, for example, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98% or 99% or greater amino acid sequence identity, preferably over a region of at least about 25, 50, 100, 200, 300, 400, or more amino acids, or over the full-length, to an amino acid sequence encoded by a LAMP-2 nucleic acid (*see, e.g.*, GenBank Accession Nos. NM_002294.2 (isoform A). NM_013995.2 (isoform B), NM_001122606.1 (isoform C)) or to an amino acid sequence of a LAMP-2 polypeptide (*see e.g.*, GenBank Accession Nos. NP_002285.1 (isoform A), NP_054701.1 (isoform B), NP_001116078.1 (isoform C)); (2) bind to antibodies, *e.g.*, polyclonal antibodies, raised against an immunogen comprising an amino acid sequence of a LAMP-2 polypeptide (*e.g.*, LAMP-2 polypeptides described herein); or an amino acid sequence encoded by a LAMP-2 nucleic acid (*e.g.*, LAMP-2 polynucleotides described herein), and conservatively modified variants thereof; (3) specifically hybridize under stringent hybridization conditions to an anti-sense strand corresponding to a nucleic acid sequence encoding a LAMP-2 protein, and conservatively modified variants thereof; (4) have a nucleic acid sequence that has greater than about 90%, preferably greater than about 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or higher nucleotide sequence identity, preferably over a region of at least about 25, 50, 100, 200, 500, 1000, 2000 or more nucleotides, or over the full-length, to a LAMP-2 nucleic acid (*e.g.*,

LAMP-2 polynucleotides, as described herein, and LAMP-2 polynucleotides that encode LAMP-2 polypeptides, as described herein).

The terms "identical" or percent "identity," in the context of two or more nucleic acids or polypeptide sequences, refer to two or more sequences or subsequences that are the same or have a specified percentage of amino acid residues or nucleotides that are the same (*i.e.* , share at least about 80% identity, for example, at least about 85%, 86%, 87%, 88%, 89%, 90%, 91 %, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99% identity over a specified region to a reference sequence, *e.g.*, LAMP-2 polynucleotide or polypeptide sequence as described herein, when compared and aligned for maximum correspondence over a comparison window, or designated region as measured using one of the following sequence comparison algorithms or by manual alignment and visual inspection. Such sequences are then said to be "substantially identical." This definition also refers to the compliment of a test sequence. Preferably, the identity exists over a region that is at least about 25 amino acids or nucleotides in length, for example, over a region that is 50, 100, 200, 300, 400 amino acids or nucleotides in length, or over the full-length of a reference sequence.

For sequence comparison, typically one sequence acts as a reference sequence, to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are entered into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. Default program parameters can be used, or alternative parameters can be designated. The sequence comparison algorithm then calculates the percent sequence identities for the test sequences relative to the reference sequence, based on the program parameters. For sequence comparison of nucleic acids and proteins to LAMP-2 nucleic acids and proteins, the BLAST and BLAST 2.0 algorithms and the default parameters are used.

A "comparison window", as used herein, includes reference to a segment of any one of the number of contiguous positions selected from the group consisting of from 20 to 600, usually about 50 to about 200, more usually about 100 to about 150 in which a sequence may be compared to a reference sequence of the same number of contiguous positions after the two sequences are optimally aligned. Methods of alignment of sequences for comparison are well-known in the art. Optimal alignment of sequences for comparison can be conducted, *e.g.*, by the local homology algorithm of Smith & Waterman, Adv. Appl. Math. 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, J. Mol. Biol. 48:443 (1970), by the search for similarity method of Pearson & Lipman, Proc. Nat'l. Acad. Sci. USA

85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, Wl), or by manual alignment and visual inspection (see, *e.g.*, Ausubel et al., eds., Current Protocols in Molecular Biology (1995 supplement)). Examples of algorithms that are suitable for determining percent sequence identity and sequence similarity are the BLAST and BLAST 2.0 algorithms, which are described in Altschul et al., J. Mol. Biol. 215:403-410 (1990) and Altschul et al., Nucleic Acids Res. 25:3389-3402 (1977), respectively. Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (on the worldwide web at ncbi.nlm.nih.gov/).

An indication that two nucleic acid sequences or polypeptides are substantially identical is that the polypeptide encoded by the first nucleic acid is immunologically cross reactive with the antibodies raised against the polypeptide encoded by the second nucleic acid, as described below. Thus, a polypeptide is typically substantially identical to a second polypeptide, for example, where the two peptides differ only by conservative substitutions. Another indication that two nucleic acid sequences are substantially identical is that the two molecules or their complements hybridize to each other under stringent conditions. Yet another indication that two nucleic acid sequences are substantially identical is that the same primers can be used to amplify the sequence.

As used herein, "administering" refers to local and systemic administration, *e.g.*, including enteral, parenteral, pulmonary, and topical/transdermal administration. Routes of administration for compounds (*e.g.*, polynucleotide encoding one or more LAMP- 2 isoforms) that find use in the methods described herein include, *e.g.*, oral (per os (P.O.)) administration, nasal or inhalation administration, administration as a suppository, topical contact, transdermal delivery (*e.g.*, via a transdermal patch), intrathecal (IT) administration, intravenous ("iv") administration, intraperitoneal ("ip") administration, intramuscular ("im") administration, intralesional administration, or subcutaneous ("sc") administration, or the implantation of a slow-release device *e.g.*, a mini-osmotic pump, a depot formulation, etc. , to a subject. Administration can be by any route including parenteral and transmucosal (*e.g.*, oral, nasal, vaginal, rectal, or transdermal). Parenteral administration includes, *e.g.*, intravenous, intramuscular, intraarterial, intrarenal, intraurethral, intracardiac, intracoronary, intramyocardial, intradermal, epidural, subcutaneous, intraperitoneal, intraventricular, ionophoretic and intracranial. Other modes of delivery include, but are not limited to, the use

of liposomal formulations, intravenous infusion, transdermal patches, etc. In some embodiments, the dose of rAAV gene therapy vector administered is about 1E+11 vector genomes (vg)/kg to about 1E+12 vg/kg, about 1E+12 vg/kg to about 2E+12 vg/kg, about 2E+12 vg/kg to about 3E+12 vg/kg, about 3E+12 vg/kg to about 3E+13 vg/kg, or about 3E+13 vg/kg to about 3E+14 vg/kg. In some embodiments, the dose of rAAV gene therapy vector administered is about 3E+12 vg/kg to about 3E+14 vg/kg.

The terms "systemic administration" and "systemically administered" refer to a method of administering a compound or composition to a mammal so that the compound or composition is delivered to sites in the body, including the targeted site of pharmaceutical action, via the circulatory system. Systemic administration includes, but is not limited to, oral, intranasal, rectal and parenteral (*e.g.*, other than through the alimentary tract, such as intramuscular, intravenous, intra-arterial, transdermal and subcutaneous) administration.

The term "co-administering" or "concurrent administration", when used, for example with respect to the compounds (*e.g.*, LAMP-2 polynucleotides) and/or analogs thereof and another active agent, refers to administration of the compound and/or analogs and the active agent such that both can simultaneously achieve a physiological effect. The two agents, however, need not be administered together. In certain embodiments, administration of one agent can precede administration of the other. Simultaneous physiological effect need not necessarily require presence of both agents in the circulation at the same time. However, in certain embodiments, co-administering typically results in both agents being simultaneously present in the body (*e.g.*, in the plasma) at a significant fraction (*e.g.*, 20% or greater, *e.g.*, 30% or 40% or greater, *e.g.*, 50% or 60% or greater, *e.g.*, 70% or 80% or 90% or greater) of their maximum serum concentration for any given dose.

The term "effective amount" or "pharmaceutically effective amount" refer to the amount and/or dosage, and/or dosage regime of one or more compounds (*e.g.*, gene therapy vectors) necessary to bring about the desired result *e.g.*, increased expression of one or more LAMP-2 isoforms in an amount sufficient to reduce the ultimate severity of a disease characterized by impaired or deficient autophagy (*e.g.*, Danon disease). In some embodiments, the effective amount is about 1E+11 vg/kg to about 1E+12 vg/kg, about 1E+12 vg/kg to about 2E+12 vg/kg, about 2E+12 vg/kg to about 3E+12 vg/kg, about 3E+12 vg/kg to about 3E+13 vg/kg, or about 3E+13 vg/kg to about 3E+14 vg/kg of rAAV gene therapy vector. In some embodiments, the effective amount is about 3E+12 vg/kg to about 3E+14 vg/kg of rAAV gene therapy vector.

The phrase "cause to be administered" refers to the actions taken by a medical professional (*e.g.*, a physician), or a person controlling medical care of a subject, that control and/or permit the administration of the agent(s)/compound(s) at issue to the subject. Causing to be administered can involve diagnosis and/or determination of an appropriate therapeutic or prophylactic regimen, and/or prescribing particular agent(s)/compounds for a subject. Such prescribing can include, for example, drafting a prescription form, annotating a medical record, and the like.

As used herein, the terms "treating" and "treatment" refer to delaying the onset of, retarding or reversing the progress of, reducing the severity of, or alleviating or preventing either the disease or condition to which the term applies, or one or more symptoms of such disease or condition. The terms "treating" and "treatment" also include preventing, mitigating, ameliorating, reducing, inhibiting, eliminating and/or reversing one or more symptoms of the disease or condition.

The term "mitigating" refers to reduction or elimination of one or more symptoms of that pathology or disease, and/or a reduction in the rate or delay of onset or severity of one or more symptoms of that pathology or disease, and/or the prevention of that pathology or disease. In certain embodiments, the reduction or elimination of one or more symptoms of pathology or disease can include, *e.g.*, measurable and sustained increase in the expression levels of one or more isoforms of LAMP-2.

As used herein, the phrase "consisting essentially of refers to the genera or species of active pharmaceutical agents recited in a method or composition, and further can include other agents that, on their own do not have substantial activity for the recited indication or purpose.

The terms "subject," "individual," and "patient" interchangeably refer to a mammal, preferably a human or a non-human primate, but also domesticated mammals (*e.g.*, canine or feline), laboratory mammals (*e.g.*, mouse, rat, rabbit, hamster, guinea pig) and agricultural mammals (*e.g.*, equine, bovine, porcine, ovine). In various embodiments, the subject can be a human (*e.g.*, adult male, adult female, adolescent male, adolescent female, male child, female child).

The terms "gene transfer" or "gene delivery" refer to methods or systems for reliably inserting foreign DNA into host cells. Such methods can result in transient expression of non-integrated transferred DNA, extrachromosomal replication and expression of transferred

replicons (*e.g.* episomes), or integration of transferred genetic material into the genomic DNA of host cells.

The term "vector" is used herein to refer to a nucleic acid molecule capable transferring or transporting another nucleic acid molecule. The transferred nucleic acid is generally linked to, *e.g.*, inserted into, the vector nucleic acid molecule. A vector may include sequences that direct autonomous replication or reverse transcription in a cell, or may include sequences sufficient to allow integration into host cell DNA. "vectors" include gene therapy vectors. As used herein, the term "gene therapy vector" refers to a vector capable of use in performing gene therapy, *e.g.*, delivering a polynucleotide sequence encoding a therapeutic polypeptide to a subject. Gene therapy vectors may comprise a nucleic acid molecule ("transgene") encoding a therapeutically active polypeptide, *e.g.*, a LAMP-2B or other gene useful for replacement gene therapy when introduced into a subject. Useful vectors include, but are not limited to, viral vectors.

As used herein, the term "expression cassette" refers to a DNA segment that is capable in an appropriate setting of driving the expression of a polynucleotide (a "transgene") encoding a therapeutically active polypeptide (*e.g.*, LAMP-2B) that is incorporated in said expression cassette. When introduced into a host cell, an expression cassette *inter alia* is capable of directing the cell's machinery to transcribe the transgene into RNA, which is then usually further processed and finally translated into the therapeutically active polypeptide. The expression cassette can be comprised in a gene therapy vector. Generally, the term expression cassette excludes polynucleotide sequences 5' to the 5' ITR and 3' to the 3' ITR.

All patents, patent publications, and other publications referenced and identified in the present specification are individually and expressly incorporated herein by reference in their entirety for all purposes.

EXAMPLES

EXAMPLE 1: Pre-Clinical and Clinical Evaluation of AAVrh74-LAMP-2B

A plasmid vector including the gene expression cassette as depicted in **FIG. 1** is generated. The transgene is modified to encode LAM2B-HA-FLAG, so that the protein may be detected using either anti-HA or anti-FLAG antibodies. AAVrh74-LAMP2B viral vector is generated using a three-plasmid, helper virus-free system to generate recombinant AAV particles containing serotype rh74 capsid proteins and viral genomes that have AAV2 ITRs

flanking a human LAMP-2B expression cassette. The viral vector is tested in non-human primates.

Pharmacology and toxicology studies are conducted in LAMP-2B$^{-/-}$ and wild-type mice. Based on the preclinical safety and efficacy data observed in mice and non-human primate studies, clinical studies in patients with Danon disease are performed.

EXAMPLE 2: DNA, RNA, and Protein Expression in Non-Human Primates Following Intravenous Administration of 1×10^{13} vg/kg AAV9.LAMP2B and AAVrh74.LAMP2B

Non-human primate studies of AAV9 versus AAVrh74 vectors were performed in paired male and female African Green Monkeys (AGM). Subjects received either AAV9.LAMP2B-HA-Flag or AAVrh74.LAMP2B-HA-Flag. "AAV9.LAMP2B-HA-Flag" is an AAV9 serotype adeno-associated virus vector encoding LAMP2B C-terminally fused to an HA-Flag tag. "AAVrh74.LAMP2B-HA-Flag" is an AAVrh74 serotype adeno-associated virus vector encoding LAMP2B C-terminally fused to an HA-Flag tag. One subject was given vehicle control. The vectors were administered by intravenous injection of 2 mL of 1.85×10^{13} vector genomes (vg)/mL as determined by quantitative polymerase chain reaction (qPCR) using a plasmid containing the WPRE sequence to generate a reference curve. This injection achieved the target dose of vector, which was about 1.0×10^{13} vg/kg. Due to lower body weight, female subjects received about 1.2×10^{13} vg/kg of their respective vectors. This experiment is summarized in **Table 2**.

Table 2

Animal	Sex	Treatment	Weight (kg)	Dose (vg/kg)
B059	M	Vehicle buffer	3.88	n/a
A991	F	AAV9	2.97	1.2×10^{13}
A602	M	AAV9	3.56	1.0×10^{13}
A710	F	AAVrh74	3.27	1.2×10^{13}
A981	M	AAVrh74	3.69	1.0×10^{13}

Subjects were humanely sacrificed two months after injection and tissues were collected for DNA, RNA, and protein analysis. The following tissues were examined: heart

(left atrium, right atrium, left ventricle and right ventricle); skeletal muscle (quadricep and gastrocnemius); liver (left, right, middle and quadrate lobes); brain (frontal lobe, parietal lobe, temporal lobe, occipital lobe, cortex, hippocampus, medulla, and cerebellum); and gonads.

Vector DNA – Quantitative PCR

DNA was extracted from frozen tissues using Qiagen DNeasy® kit. DNA purity (A260/A280) and concentration were evaluated on a NanoDrop One™ spectrophotometer (Thermo). Quantitative PCR (qPCR) was performed on 20 ng DNA using TaqMan Universal Master Mix II (Thermo, 4440038) on a real-time PCR system (QuantStudio5, Thermo) using the following primers/probes:

WPRE (cassette):

Forward primer: 5`-ATCATGCTATTGCTTCCCGTA-3` (SEQ ID NO:30)

Reverse primer: 5`-GGGCCACAACTCCTCATAAA-3` (SEQ ID NO:31)

Probe: 5`-CCTCCTTGTATAAATCCTGGTTGCTGTCT-3` (SEQ ID NO:32)

RNaseP (housekeeping gene, Thermo)

A standard curve was generated using plasmid DNA containing the WPRE sequence. **FIG. 2** shows a bar graph of vector DNA quantification in organs most affected in Danon disease by qPCR.

LAMP2B mRNA – Quantitative RT-PCR

RNA was extracted from heart and muscle tissues using the RNeasy Fibrous Tissue kit (Qiagen), and from liver and brain using RNeasy Lipid Tissue kit (Qiagen). Purity (A260/A280) and concentration were determined on a NanoDrop One spectrophotometer. RNA was converted to cDNA using the Superscript IV VILO master mix (Thermo). qPCR was performed on 10 ng of RNA in TaqMan Universal Master Mix II (Thermo) on a real-time PCR system (QuantStudio5, Thermo) using the following primers/probes:

WPRE (cassette):

Forward primer: 5`-ATCATGCTATTGCTTCCCGTA-3` (SEQ ID NO:33)

Reverse primer: 5`-GGGCCACAACTCCTCATAAA-3` (SEQ ID NO:34)

Probe: 5`-CCTCCTTGTATAAATCCTGGTTGCTGTCT-3` (SEQ ID NO:35)

HPRT-1 (Housekeeping gene, Thermo)

A standard curve was generated using plasmid DNA containing the WPRE sequence. **FIG. 3A** shows a bar graph of vector DNA quantification in regions of the heart by qPCR. **FIG. 3B** shows a bar graph of vector DNA quantification in muscles by qPCR. **FIG. 4** shows a bar graph of mRNA quantification in organs most affected by Danon disease by RT-qPCR. **FIG. 5A** shows a bar graph of mRNA quantification in regions of the heart by RT-qPCR. **FIG. 5B** shows a bar graph of mRNA quantification in muscles by RT-qPCR.

LAMP2B mRNA - RNAscope

5mm tissue cubes fixed in 10% neutral buffered formalin, embedded in paraffin and sectioned. Transgene mRNA was detected using WPRE-O3 ZZ probe (ACD) with RNAscope 2.5 LS RED. Semi-quantitative visual assessment of one section from each tissue was performed with cells with ≥1 dot per cell considered positive. The percentage of cells positive were binned into five categories: 0%, 1-25%, 26-50%, 51-75% or 100%.

FIG. 6A shows a micrograph of semi-quantitative mRNA analysis by RNAscope in an untreated left ventricle. **FIG. 6B** shows micrographs of semi-quantitative mRNA analysis by RNAscope in treated left ventricles. **FIG. 7A** shows a micrograph of semi-quantitative mRNA analysis by RNAscope in an untreated quadricep. **FIG. 7B** shows micrographs of semi-quantitative mRNA analysis by RNAscope in treated quadriceps. **FIG. 8** shows micrographs of semi-quantitative mRNA analysis by RNAscope in treated gastrocnemius. Results are summarized for vehicle control (**Table 3A**), AAV9 (**Table 3A**), and AAVrh74 (**Table 3C**).

Table 3A

Animal	Tissue	Location	Treatment	% Expressing Cells
B059	Heart	left ventricle	Untreated, vehicle	0%
B059	Muscle	Quadricep	Untreated, vehicle	0%
B059	Liver	Left Lobe	Untreated, vehicle	0%

Table 3B

Animal	Tissue	Location	Treatment	% Expressing Cells
A991	Heart	left ventricle	Treated, AAV9	26-50%
A991	Heart	right ventricle	Treated, AAV9	26-50%
A991	Heart	left atrium	Treated, AAV9	1-25%
A991	Heart	right atrium	Treated, AAV9	26-50%
A991	Muscle	quadricep	Treated, AAV9	0%
A991	Muscle	gastrocnemius	Treated, AAV9	0%
A991	Liver	left lobe	Treated, AAV9	26-50%
A991	Liver	right lobe	Treated, AAV9	51-75%
A602	Heart	left ventricle	Treated, AAV9	1-25%
A602	Heart	right ventricle	Treated, AAV9	1-25%
A602	Heart	left atrium	Treated, AAV9	26-50%
A602	Heart	right atrium	Treated, AAV9	1-25%
A602	Muscle	quadricep	Treated, AAV9	1-25%
A602	Muscle	gastrocnemius	Treated, AAV9	1-25%
A602	Liver	left lobe	Treated, AAV9	51-75%
A602	Liver	right lobe	Treated, AAV9	51-75%

Table 3C

Animal	Tissue	Location	Treatment	% Expressing Cells
A710	Heart	left ventricle	Treated, AAVrh.74	1-25%
A710	Heart	right ventricle	Treated, AAVrh.74	1-25%
A710	Heart	left atrium	Treated, AAVrh.74	1-25%
A710	Heart	right atrium	Treated, AAVrh.74	1-25%
A710	Muscle	quadricep	Treated, AAVrh.74	0%
A710	Muscle	gastrocnemius	Treated, AAVrh.74	1-25%
A710	Liver	left lobe	Treated, AAVrh.74	51-75%
A710	Liver	right lobe	Treated, AAVrh.74	51-75%
A981	Heart	left ventricle	Treated, AAVrh.74	26-50%

Animal	Tissue	Location	Treatment	% Expressing Cells
A981	Heart	right ventricle	Treated, AAVrh.74	26-50%
A981	Heart	left atrium	Treated, AAVrh.74	0%
A981	Heart	right atrium	Treated, AAVrh.74	26-50%
A981	Muscle	quadricep	Treated, AAVrh.74	0%
A981	Muscle	gastrocnemius	Treated, AAVrh.74	0%
A981	Liver	left lobe	Treated, AAVrh.74	51-75%
A981	Liver	right lobe	Treated, AAVrh.74	26-51%

LAMP2B protein - ELISA

Approximately 125 mg of tissue was homogenized in 500 µL of lysis buffer using 0.9-2.00 mm stainless steel beads (Next Advance) and a Next Advance Bullet Blender 24. The lysis buffer contains 300mM NaCl, 20mM EDTA, 100mM Tris pH 8.0, 2% NP-40 and 0.2% SDS with Complete™ EDTA-free protease inhibitor and PhosSTOP™ phosphatase inhibitor. Total protein was assessed by BCA (Thermo). 100 mg of total protein was loaded per well. A standard curve was constructed using purified human LAMP2 protein (Origene). ELISA was performed with a mouse monoclonal antibody (H4B4, Novus Biologicals) as the capture antibody, a goat polyclonal antibody (R&D Systems) as the detection antibody, HRP-linked antibody: Donkey anti-goat (Millipore) as the secondary antibody. Plates were developed with TMB (Thermo) and quantified on a spectrophotometer (Spectramax M5c).

FIG. 9 shows a bar graph of protein quantification in tissues most affected in Danon disease by ELISA. **FIG. 10A** shows a bar graph of protein quantification in regions of the heart by ELISA. **FIG. 10B** shows a bar graph of protein quantification in muscles by ELISA.

Clinical Pathology

Pathological effects of the vectors were assessed. **FIGS. 11A-11D** show line graphs of clinical pathology measurement in NHP serum over course of study. Clinical pathology levels were assessed as changes in (**FIG. 11A**) alanine aminotransferase, ALT; (**FIG. 11B**) aspartate aminotransferase, AST; (**FIG. 11C**) white blood cells, WBC; and (**FIG. 11D**) neutrophils over the study duration. B059 is the vehicle control. A991 and A602 are AAV9-treated animals. A710 and A981 are AAVrh74-treated animals.

Conclusions

AAV-based gene therapy using a LAMP2B transgene was well tolerated in non-human primates at vector dose 1.0×10^{13} vg/kg. This result is an important and unexpected result because experiments with AAV-based gene therapy for some other transgenes have demonstrated pathological effects at doses equal to or lower than 1.0×10^{13} vg/kg. Furthermore, both AAV9 and AAVrh74 were well tolerated. Elevated levels of certain markers were observed in A602 and A710 animals at day 21, but these outliers may be due to experimental error or pathology that was self-resolving.

Both vectors localized to and transduce target tissues for treatment of Danon disease (heart and muscle), but not as much in the brain or gonads. Expression in the gonads would be undesirable for safety reasons. As expected significant amounts of vector accumulate in the liver, which is desirable because liver is a tissue affected by Danon disease. Vector is present in each quadrant of the heart and in both quadricep and gastrocnemius muscles. This is a desirable result for treatment of Danon disease.

Localization of a serotype of AAV vector (*e.g.* AAV9) is not predictive of localization of others (*e.g.* AAVrh74). This experiment demonstrates that AAVrh74 achieves desirable localization for treatment of Danon disease, or other diseases with etiology linked to heart and muscle tissues. Both LAMP2B transgene mRNA and protein are expressed in the same sets of tissues in both AAV9 and AAVrh74 groups. Expression is comparable between vector serotypes. The number of animals in the study is too few to discern statistically significant trends in expression levels between AAV9 and AAVrh74. RNAscope suggests that a similar fraction of cells are infected in heart, muscle, and liver tissues in AAV9 and AAVrh74 groups.

These data demonstrate that AAVrh74 may be used as a vector to deliver LAMP2B to tissues relevant to treatment of Danon disease. AAVrh74 was non-inferior to AAV9 in these experiments.

CLAIMS

What is claimed is:

1. A recombinant adeno-associated virus (rAAV) gene therapy vector, comprising a polynucleotide comprising a 5' ITR, an expression cassette, and a 3' ITR; and a capsid protein,

wherein the expression cassette comprises a transgene encoding a lysosome-associated membrane protein 2 (LAMP-2) or a functional variant thereof,

wherein the expression cassette is flanked by the 5' ITR and the 3' ITR, and

wherein the capsid protein comprises an AAVrh.74 capsid protein or a functional variant thereof.

2. The rAAV gene therapy vector of claim 1, wherein the LAMP-2 is selected from LAMP-2A, LAMP-2B and LAMP-2C.

3. The rAAV gene therapy vector of claim 1, wherein the capsid protein has at least 95% sequence identity to an amino acid sequence selected from SEQ ID NOs: 2-4.

4. The rAAV gene therapy vector of claim 3, wherein the capsid protein shares at least 95% sequence identity to SEQ ID NOs: 2.

5. The rAAV gene therapy vector of claim 4, wherein the capsid protein shares at least 97% sequence identity to SEQ ID NOs: 2.

6. The rAAV gene therapy vector of claim 5, wherein the capsid protein shares at least 99% sequence identity to SEQ ID NOs: 2.

7. The rAAV gene therapy vector of any one of claims 1-6, wherein the capsid protein is an AAVrh.74 capsid protein.

8. The rAAV gene therapy vector of any one of claims 1-7, wherein the 5' ITR and the 3' ITR are each respectively the 5' ITR of AAV2 and the 3' ITR of AAV2, or variants thereof.

9. The rAAV gene therapy vector of claim 8, wherein the 5' ITR shares at least 98% identity to SEQ ID NO: 13 and the 3' ITR shares at least 98% identity to SEQ ID NO: 14.

10. The rAAV gene therapy vector of any one of claims 1-9, wherein the transgene is codon-optimized for expression in a human host cell.

11. The rAAV gene therapy vector of any one of claims 1-10, wherein the expression cassette contains fewer CpG sites than SEQ ID: 6.

12. The rAAV gene therapy vector of any one of claims 1-11, wherein the expression cassette contains fewer cryptic splice sites than SEQ ID: 6.

13. The rAAV gene therapy vector of any one of claims 1-12, wherein the expression cassette encodes fewer alternative open reading frames than SEQ ID: 6.

14. The rAAV gene therapy vector of any one of claims 1 to 13, wherein the transgene shares at least 95% identity to a sequence selected from SEQ ID NO: 7-9.

15. The rAAV gene therapy vector of claim 14, wherein the transgene shares at least 99% identity to a sequence selected from SEQ ID NO: 7-9.

16. The rAAV gene therapy vector of claim 15, wherein the transgene comprises a sequence selected from SEQ ID NO: 7-9.

17. The rAAV gene therapy vector of any one of claims 1 to 16, where the expression cassette comprises a consensus optimal Kozak sequence operatively linked to the transgene, wherein the consensus optimal Kozak sequence comprises SEQ ID NO: 20.

18. The rAAV gene therapy vector of any one of claims 1 to 17, where the expression cassette comprises a full-length polyA sequence operatively linked to the transgene, wherein the full-length polyA sequence comprises SEQ ID NO: 26.

19. The rAAV gene therapy vector of any one of claims 1 to 18, where the expression cassette comprises no start codon 5' to the start codon of the transgene.

20. The rAAV gene therapy vector of any one of claims 1 to 19, wherein the expression cassette comprises operatively linked, in the 5' to 3' direction, a first inverse terminal repeat, an enhancer/promoter region, an intron, a consensus optimal Kozak sequence, the transgene, a 3' untranslated region including a full-length polyA sequence, and a second inverse terminal repeat, where the expression cassette comprises no start codon 5' to the start codon of the transgene.

21. The rAAV gene therapy vector of claim 20, wherein the enhancer/promoter region comprises in the 5' to 3' direction a CMV IE Enhancer and a Chicken Beta-Actin Promoter, and optionally wherein the enhancer/promoter region further comprises a first exon and first intron of a chicken beta-actin gene and a splice acceptor of a rabbit beta-globin gene.

22. The rAAV gene therapy vector of claim 20, wherein the enhancer/promoter region comprises a tissues-specific promoter capable of mediating increased expression in cardiac tissue and/or skeletal muscle tissue compared to liver tissue.

23. The rAAV gene therapy vector of any one of claims 1 to 21, wherein the expression cassette shares at least 95% identity to a sequence selected from SEQ ID NOs: 10-12.

24. The rAAV gene therapy vector of claim 23, wherein the expression cassette comprises a sequence selected from SEQ ID NOs: 10-12.

25. A pharmaceutical composition comprising the rAAV gene therapy vector of any one of claims 1 to 24.

26. A method of treating or preventing Danon disease or another autophagy disorder in a subject in need thereof, comprising administering to the subject the rAAV gene therapy vector of any one of claims 1 to 24 or the pharmaceutical composition of claim 25.

27. The method of claim 26, wherein the rAAV gene therapy vector or pharmaceutical composition is administered via a route selected from the group consisting of intravenous, intra-arterial, intracardiac, intracoronary, intramyocardial, intrarenal, intraurethral, epidural, and intramuscular.

28. The method of claim 26 or claim 27, wherein the autophagy disorder is selected from the group consisting of end-stage heart failure, myocardial infarction, drug toxicities, diabetes, end-stage renal failure, and aging.

29. The method of any one of claims 26 to 28, wherein the subject is a human.

30. The method of any one of claims 26 to 29, wherein the subject is exhibiting symptoms of Danon disease or another autophagy disorder.

31. The method of any one of claims 26 to 30, wherein the subject has been identified as having reduced or non-detectable expression of endogenous LAMP-2.

32. The method of any one of claims 26 to 31, wherein the subject has been identified as having a mutated LAMP-2 gene.

33. The method of any one of claims 26 to 32, wherein the rAAV gene therapy vector is administered at a dose of about 3×10^{12} vg/kg to about 3×10^{14} vg/kg.

34. The method of any one of claims 26 to 33, wherein the rAAV gene therapy vector is administered at a dose of about 3×10^{12} vg/kg to about 1.2×10^{13} vg/kg.

35. The method of any one of claims 26 to 33, wherein the rAAV gene therapy vector is administered at a dose of about 1.0×10^{13} vg/kg.

36. The method of any one of claims 26 to 35, wherein the dose of rAAV gene therapy vector does not cause clinical pathology when administered, optionally when administered at a dose of about 1.0×10^{13} vg/kg.

37. The method of any one of claims 26 to 36, wherein administration of the rAAV gene therapy vector transduces one or more of heart, muscle, and liver.

38. The method of any one of claims 26 to 37, wherein administration of the rAAV gene therapy vector causes LAMP2B mRNA expression in one or more of heart, muscle, and liver.

39. The method of any one of claims 26 to 38, wherein administration of the rAAV gene therapy vector causes LAMP2B protein expression in one or more of heart, muscle, and liver.

40. The method of any one of claims 26 to 39, wherein administration of the rAAV gene therapy vector causes infection with the rAAV gene therapy vector of at least about 10%, at least about 20%, or at least about 30% of cells in one or more of heart, muscle, and liver.

41. The method of any one of claims 26 to 40, wherein administration of the rAAV gene therapy vector causes transduction of the rAAV gene therapy vector in gonads at less 0.1 vector genomes (vg) per diploid genome.

42. The method of any one of claims 26 to 41, wherein administration of the rAAV gene therapy vector causes LAMP2B mRNA expression in gonads at less than 2×10^4 mRNA copies per µg total RNA.

43. The method of any one of claims 26 to 42, wherein administration of the rAAV gene therapy vector causes no LAMP2B protein expression in brain and/or gonads.

44. The method of any one of claims 26 to 43, wherein administration of the rAAV gene therapy vector transduces and/or causes transgene expression at about the same level as an AAV9 gene therapy vector having the same expression cassette.

45. A method of delivering a LAMP-2 polynucleotide encoding a LAMP-2 protein to a cell, comprising contacting the cell with the rAAV gene therapy vector of any one of claims 1 to 24 or the pharmaceutical composition of claim 25, wherein the cell is optionally selected from a heart cell, a lung cell, and/or a muscle cell.

46. A method of transducing cells, comprising contacting the cells with the rAAV gene therapy vector of any one of claims 1 to 24 or the pharmaceutical composition of claim 25, wherein the cell is optionally selected from a heart cell, a lung cell, and/or a muscle cell.

47. A method of delivering a LAMP-2 polynucleotide encoding a LAMP-2 protein to a tissue and/or expressing a LAMP-2 protein in a tissue, comprising contacting the tissue with the rAAV gene therapy vector of any one of claims 1 to 24 or the pharmaceutical composition of claim 25, wherein the tissue is optionally selected from heart tissue, lung tissue, and/or muscle tissue.

48. A method of delivering a LAMP-2 polynucleotide encoding a LAMP-2 protein to a subject and/or expressing a LAMP-2 protein in a subject, comprising administering to the subject the rAAV gene therapy vector of any one of claims 1 to 24 or the pharmaceutical composition of claim 25.

49. The method of claim 48, wherein the rAAV gene therapy vector or pharmaceutical composition is administered via a route selected from the group consisting of intravenous, intra-arterial, intracardiac, intracoronary, intramyocardial, intrarenal, intraurethral, epidural, and intramuscular.

50. The method of claim 48 or claim 49, wherein the subject suffers from or is at risk for an autophagy disorder selected from the group consisting of Danon disease, end-stage heart failure, myocardial infarction, drug toxicities, diabetes, end-stage renal failure, and aging.

51. The method of any one of claims 48 to 50, wherein the subject is a human.

52. The method of any one of claims 48 to 51, wherein the subject is exhibiting symptoms of the autophagy disorder.

53. The method of any one of claims 48 to 52, wherein the subject has been identified as having reduced or non-detectable expression of endogenous LAMP-2.

54. The method of any one of claims 48 to 53, wherein the subject has been identified as having a mutated LAMP-2 gene.

55. The method of any one of claims 48 to 54, wherein the rAAV gene therapy vector is administered at a dose of about 3×10^{12} vg/kg to about 3×10^{14} vg/kg.

56. The method of any one of claims 48 to 55, wherein the rAAV gene therapy vector is administered at a dose of about 3×10^{12} vg/kg to about 1.2×10^{13} vg/kg.

57. The method of any one of claims 48 to 56, wherein the rAAV gene therapy vector is administered at a dose of about 1.0×10^{13} vg/kg.

58. The method of any one of claims 48 to 57, wherein the dose of rAAV gene therapy vector does not cause clinical pathology when administered, optionally when administered at a dose of about 1.0×10^{13} vg/kg.

59. The method of any one of claims 48 to 58, wherein administration of the rAAV gene therapy vector transduces one or more of heart, muscle, and liver.

60. The method of any one of claims 48 to 59, wherein administration of the rAAV gene therapy vector causes LAMP2B mRNA expression in one or more of heart, muscle, and liver.

61. The method of any one of claims 48 to 60, wherein administration of the rAAV gene therapy vector causes LAMP2B protein expression in one or more of heart, muscle, and liver.

62. The method of any one of claims 48 to 61, wherein administration of the rAAV gene therapy vector causes infection with the rAAV gene therapy vector of at least about 10%, at least about 20%, or at least about 30% of cells in one or more of heart, muscle, and liver.

63. The method of any one of claims 48 to 62, wherein administration of the rAAV gene therapy vector causes transduction of the rAAV gene therapy vector in gonads at less 0.1 vector genomes (vg) per diploid genome.

64. The method of any one of claims 48 to 63, wherein administration of the rAAV gene therapy vector causes LAMP2B mRNA expression in gonads at less than 2×10^4 mRNA copies per μg total RNA.

65. The method of any one of claims 48 to 64, wherein administration of the rAAV gene therapy vector causes no LAMP2B protein expression in brain and/or gonads.

66. The method of any one of claims 48 to 65, wherein administration of the rAAV gene therapy vector transduces and/or causes transgene expression at about the same level as an AAV9 gene therapy vector having the same expression cassette.

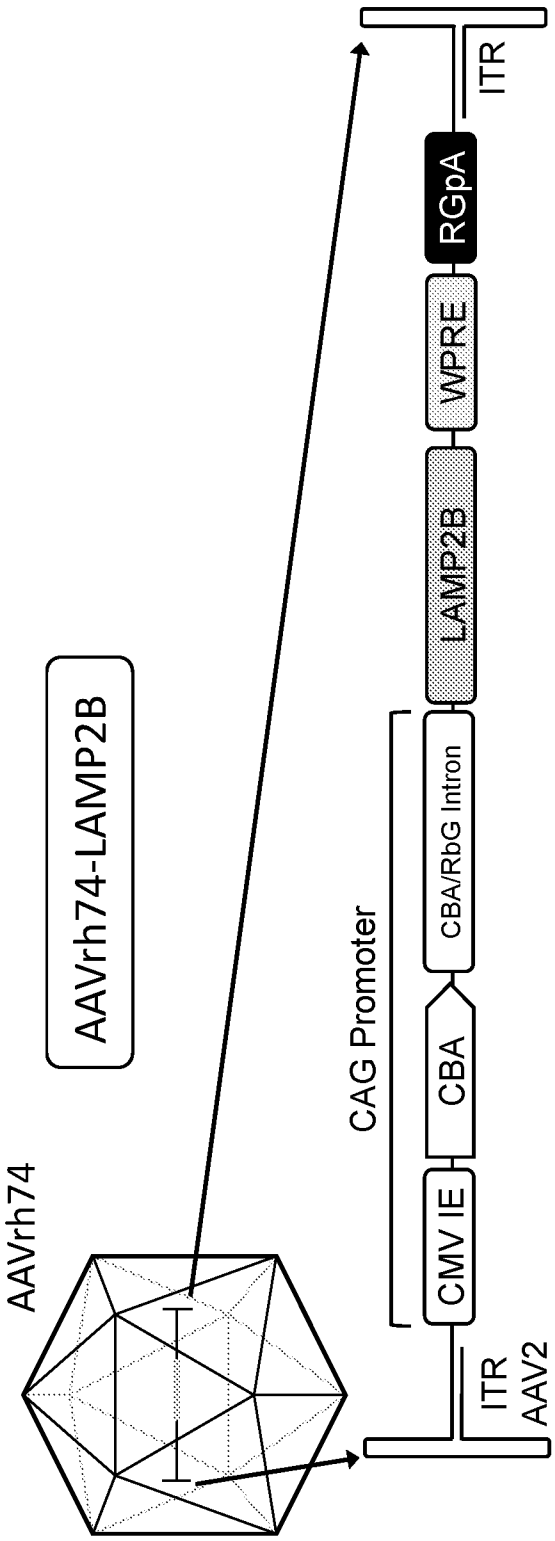
AAVrh74-LAMP2B
AAVrh74
CAG Promoter
CMV IE
CBA
CBA/RbG Intron
LAMP2B
WPRE
RGpA
ITR
ITR
AAV2

FIG. 1

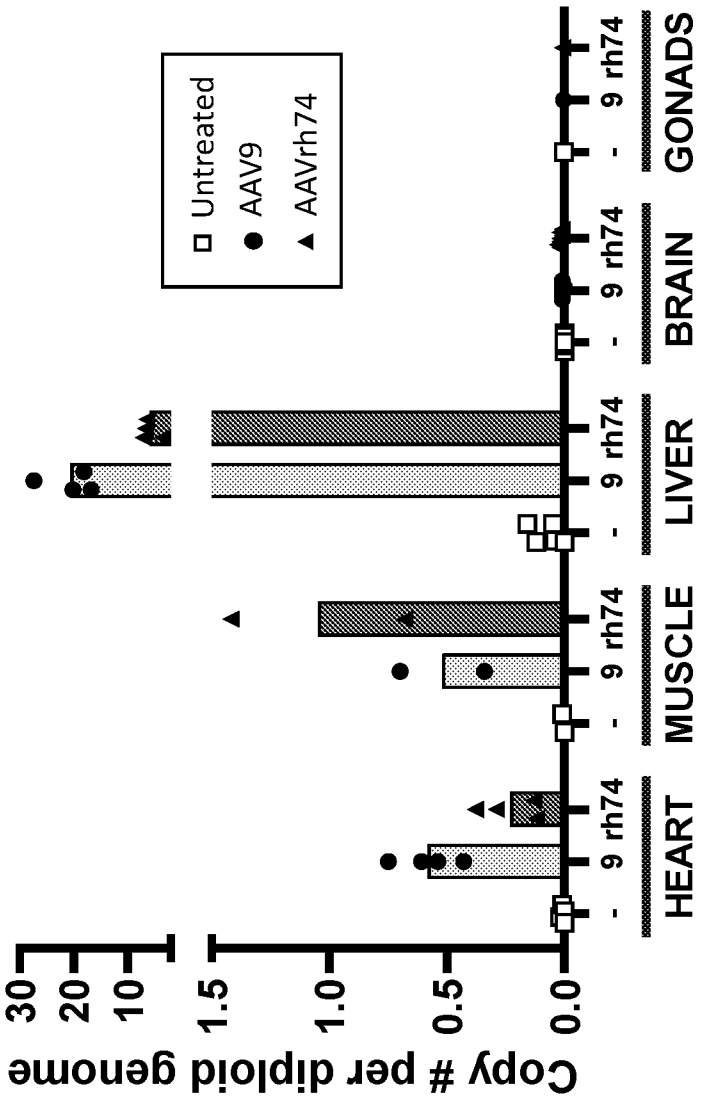
Copy # per diploid genome
30
20
10
1.5
1.0
0.5
0.0
Untreated
AAV9
AAVrh74
- 9 rh74
HEART
MUSCLE
LIVER
BRAIN
GONADS

FIG. 2

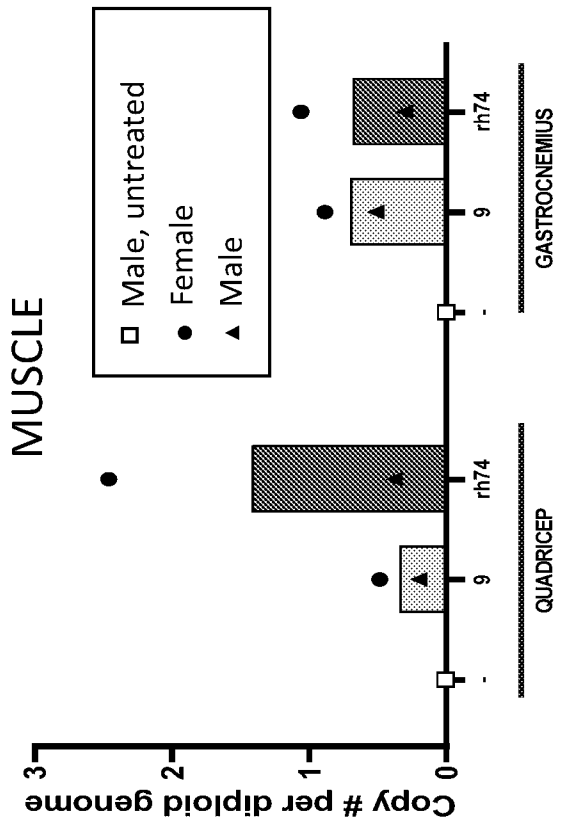
MUSCLE
Male, untreated
Female
Male
Copy # per diploid genome
0
1
2
3
-
9
rh74
QUADRICEP
GASTROCNEMIUS

FIG. 3B

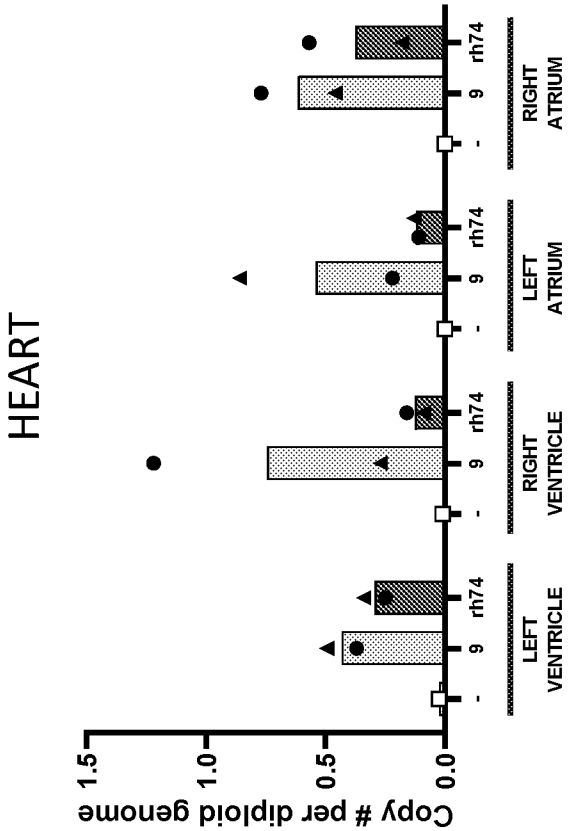
HEART
Copy # per diploid genome
0.0
0.5
1.0
1.5
-
9
rh74
LEFT VENTRICLE
RIGHT VENTRICLE
LEFT ATRIUM
RIGHT ATRIUM

FIG. 3A

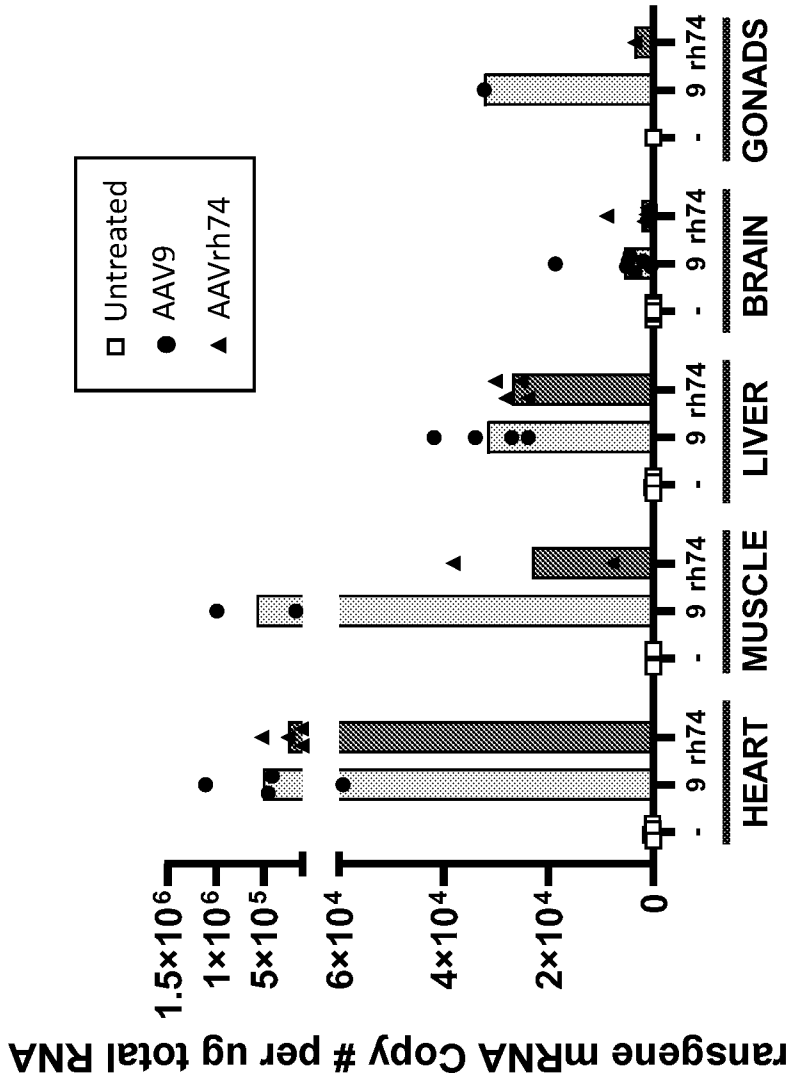
Transgene mRNA Copy # per ug total RNA
1.5×10^6
1×10^6
5×10^5
6×10^4
4×10^4
2×10^4
0
Untreated
AAV9
AAVrh74
- 9 rh74
HEART
MUSCLE
LIVER
BRAIN
GONADS

FIG. 4

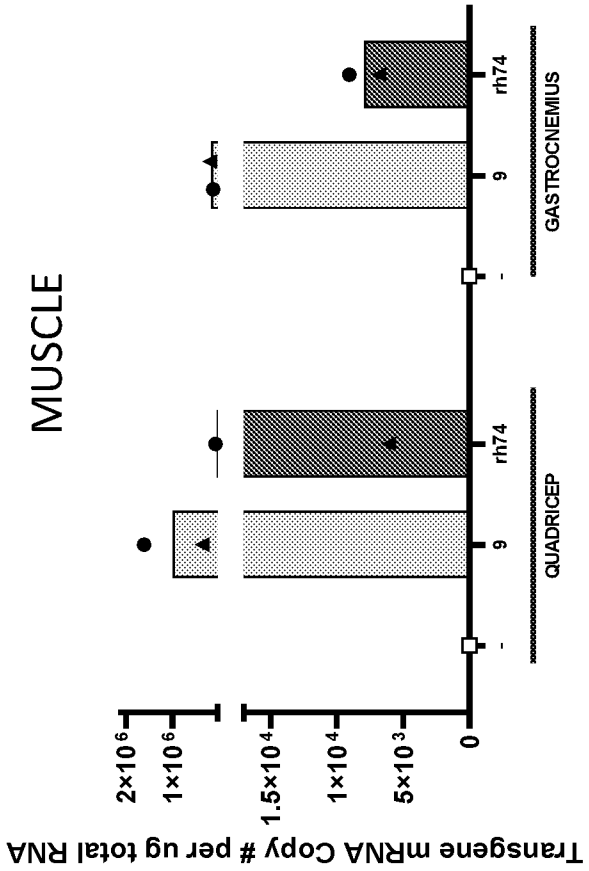
MUSCLE
Transgene mRNA Copy # per ug total RNA
2×10^6
1×10^6
1.5×10^4
1×10^4
5×10^3
0
-
9
rh74
QUADRICEP
GASTROCNEMIUS

FIG. 5B

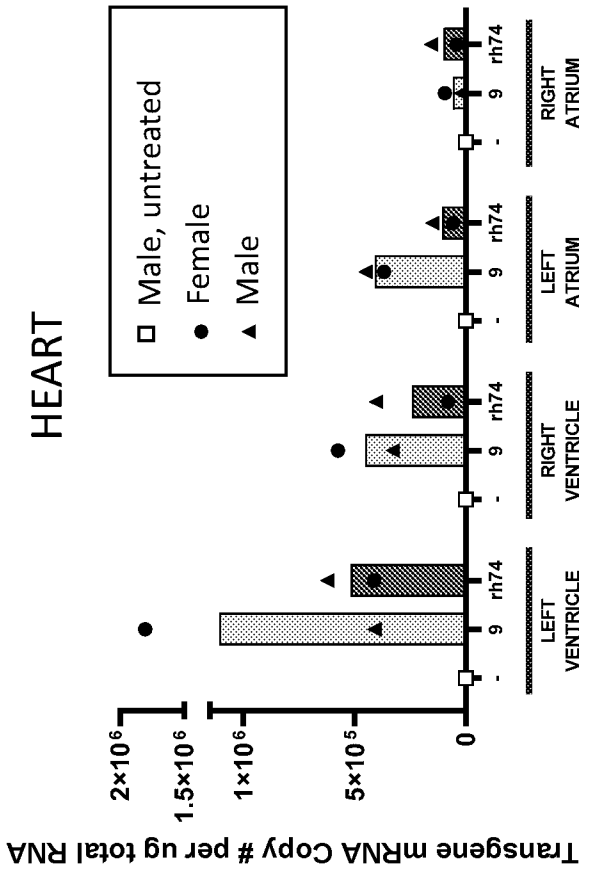
HEART
Male, untreated
Female
Male
Transgene mRNA Copy # per ug total RNA
2×10^6
1.5×10^6
1×10^6
5×10^5
0
-
9
rh74
LEFT VENTRICLE
RIGHT VENTRICLE
LEFT ATRIUM
RIGHT ATRIUM

FIG. 5A

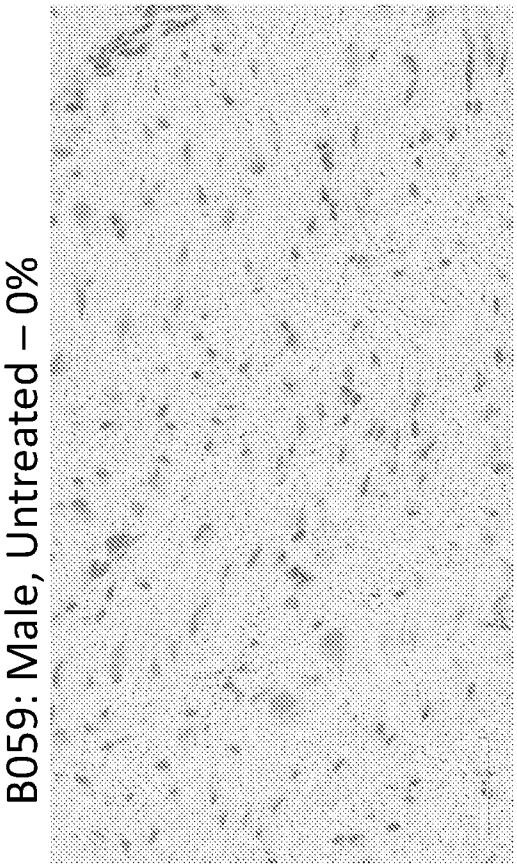

FIG. 6A

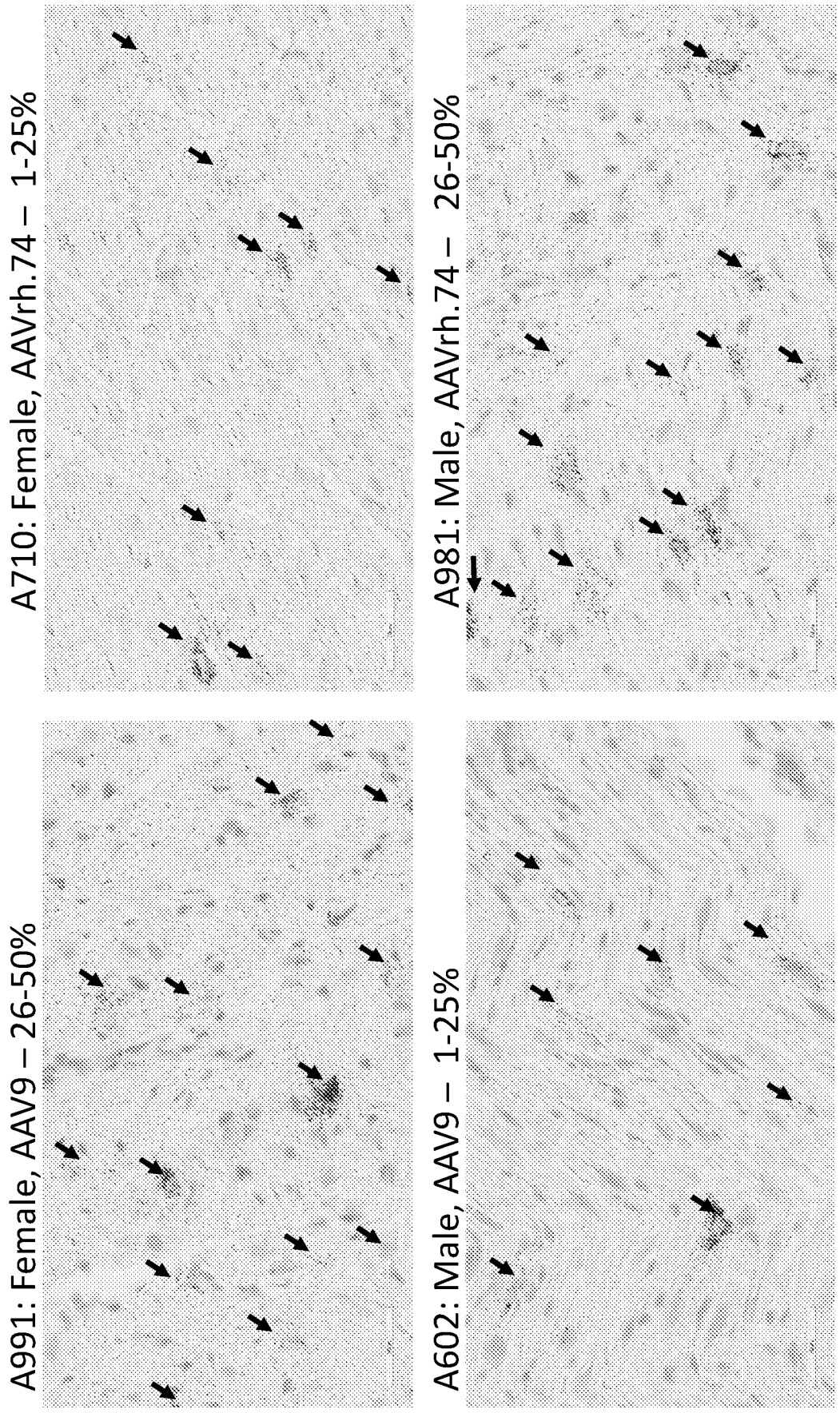
A710: Female, AAVrh.74 – 1-25%
A981: Male, AAVrh.74 – 26-50%
A991: Female, AAV9 – 26-50%
A602: Male, AAV9 – 1-25%

FIG. 6B

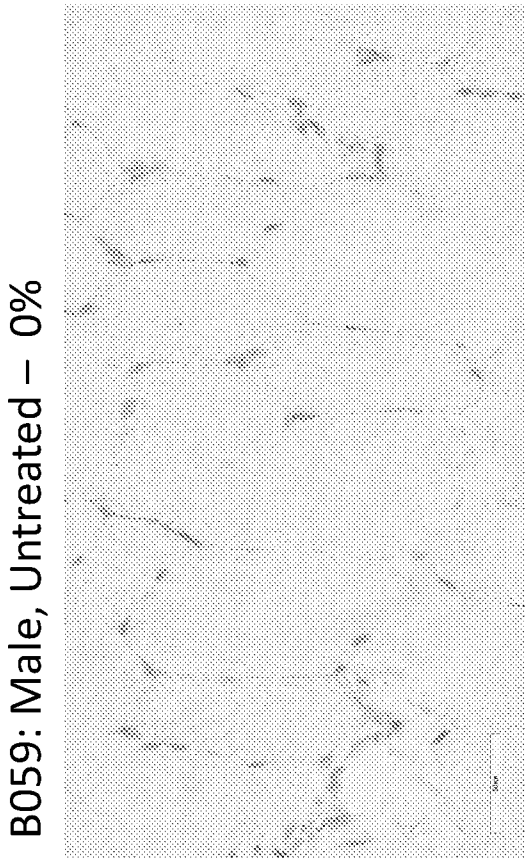

FIG. 7A

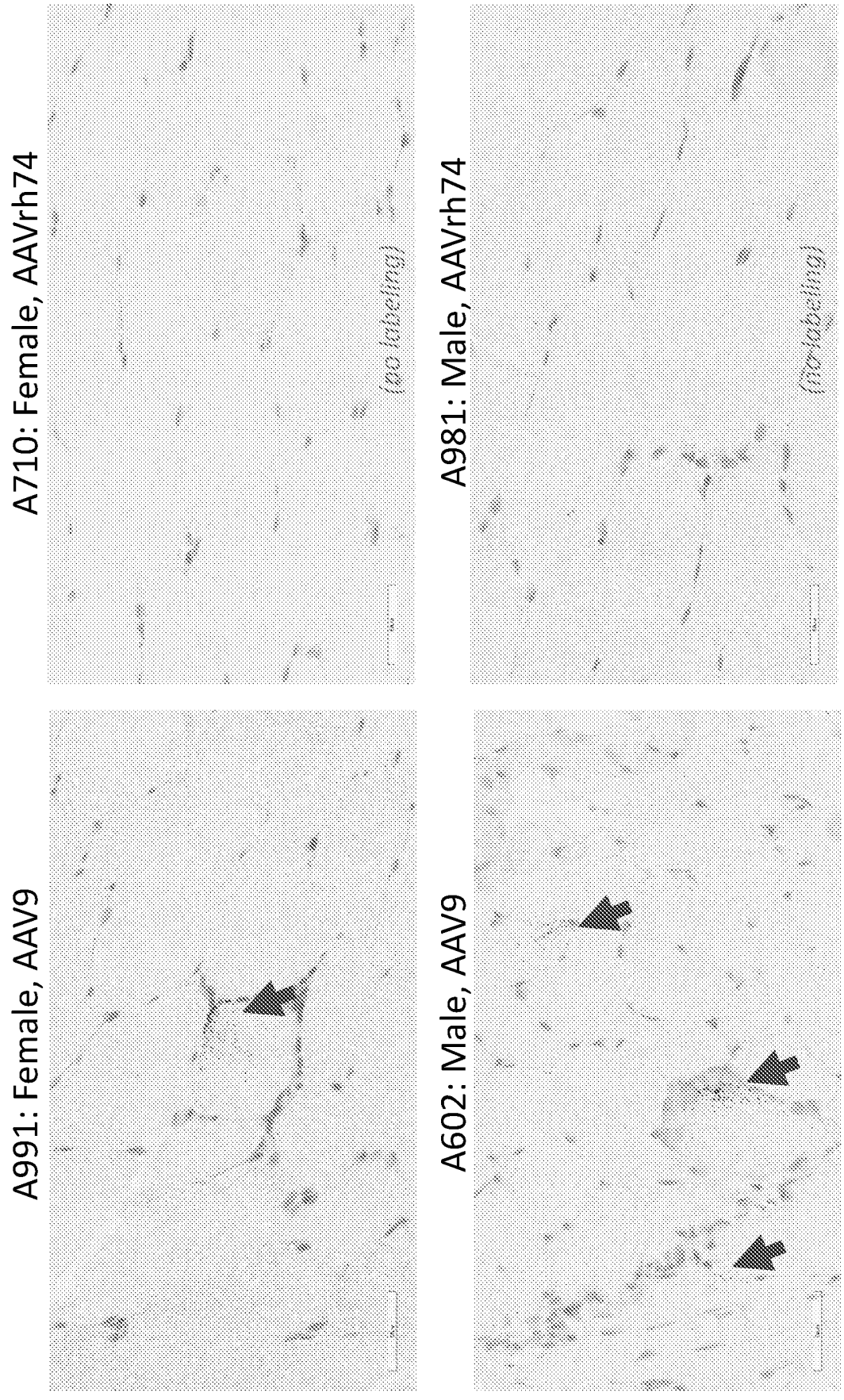
A710: Female, AAVrh74
(no labeling)
A981: Male, AAVrh74
(no labeling)
A991: Female, AAV9
A602: Male, AAV9

FIG. 7B

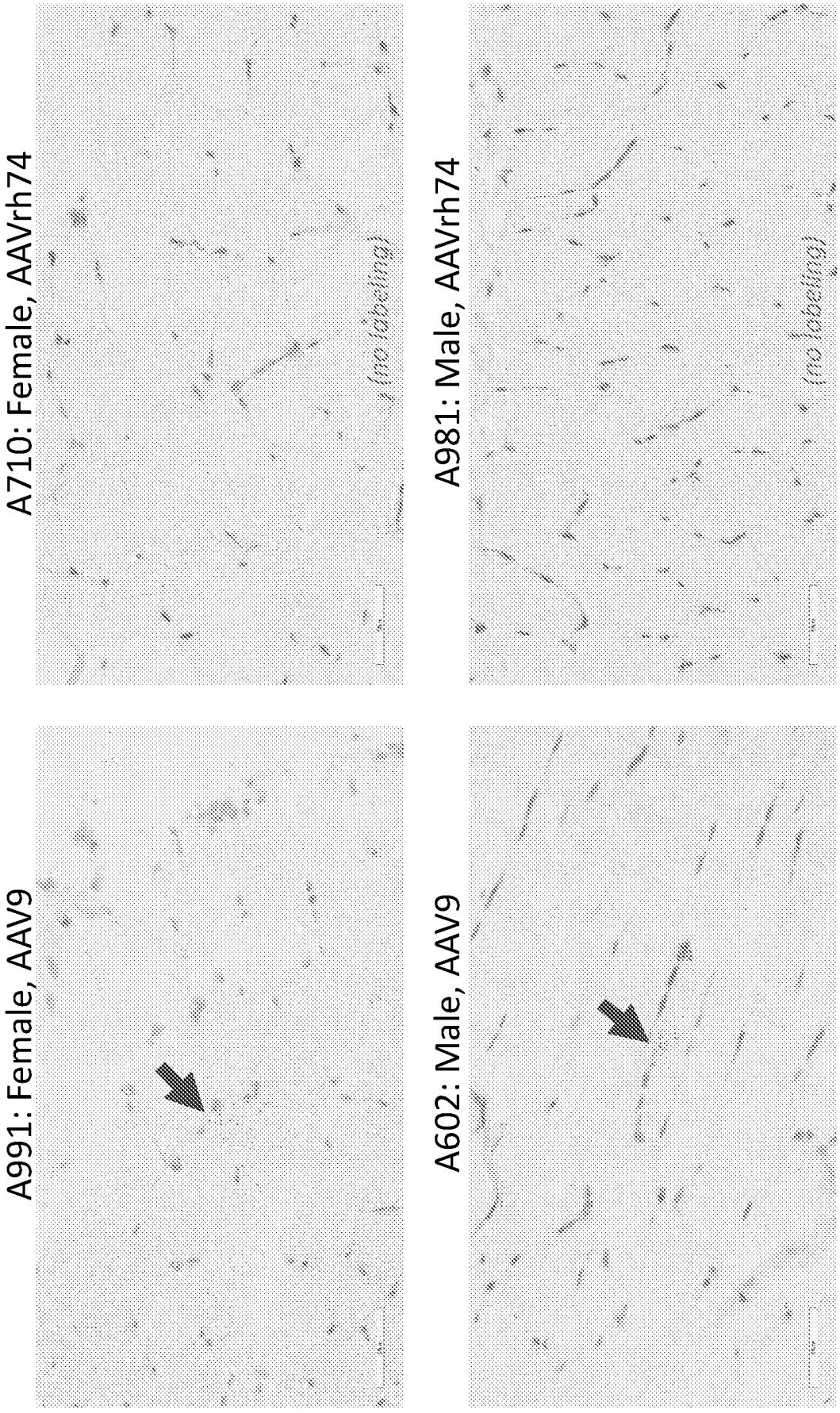
A710: Female, AAVrh74
(no labeling)
A981: Male, AAVrh74
(no labeling)
A991: Female, AAV9
A602: Male, AAV9

FIG. 8

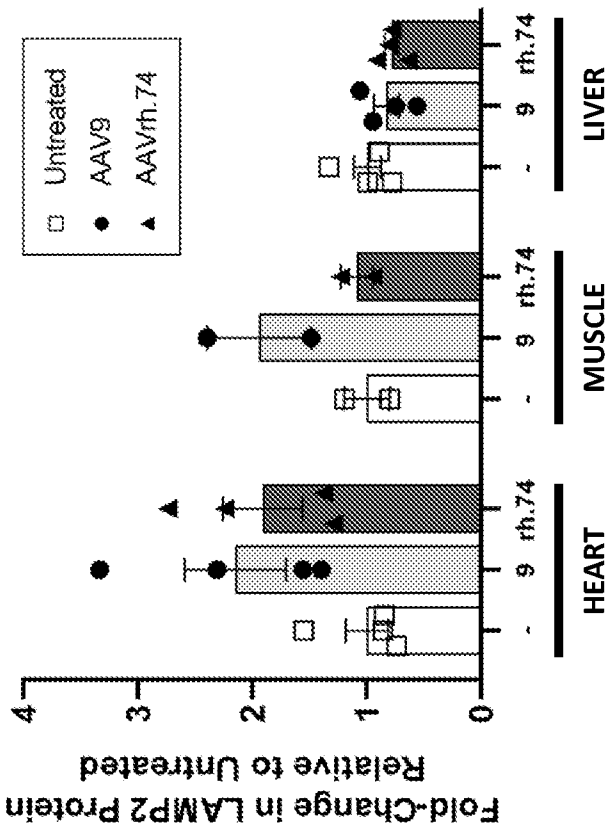
Fold-Change in LAMP2 Protein
Relative to Untreated
0
1
2
3
4
Untreated
AAV9
AAVrh.74
-
9
rh.74
HEART
MUSCLE
LIVER

FIG. 9

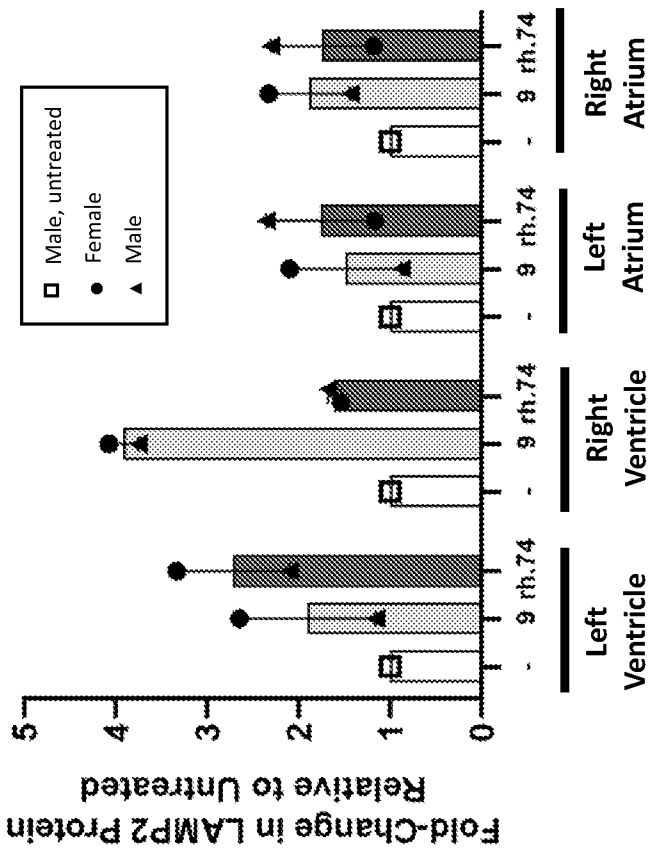
Fold-Change in LAMP2 Protein
Relative to Untreated
0
1
2
3
4
5
Male, untreated
Female
Male
- 9 rh.74
- 9 rh.74
- 9 rh.74
- 9 rh.74
Left Ventricle
Right Ventricle
Left Atrium
Right Atrium

FIG. 10A

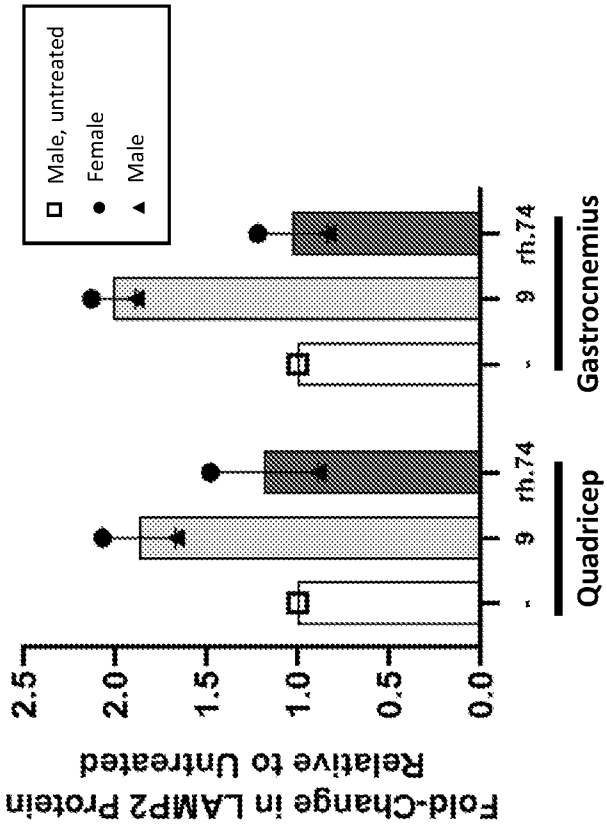
Male, untreated
Female
Male
Fold-Change in LAMP2 Protein
Relative to Untreated
2.5
2.0
1.5
1.0
0.5
0.0
- 9 rh.74
Quadricep
- 9 rh.74
Gastrocnemius

FIG. 10B

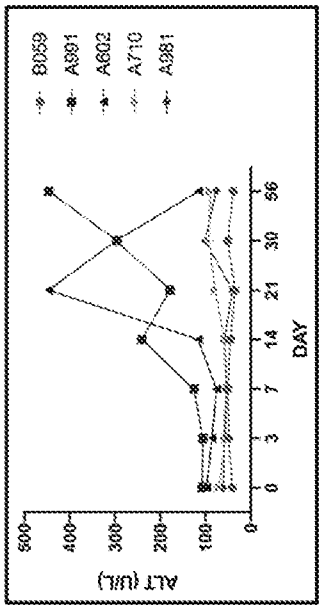

FIG. 11A

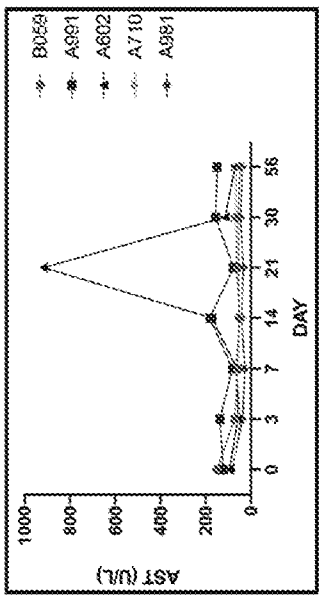

FIG. 11B

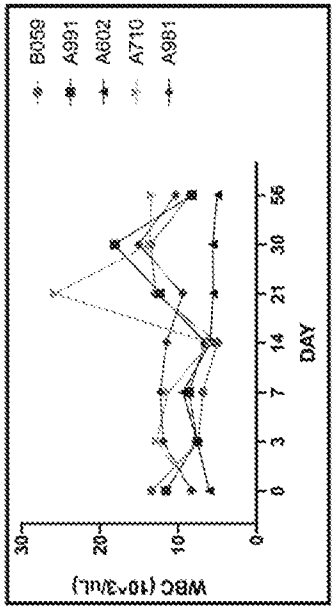

FIG. 11C

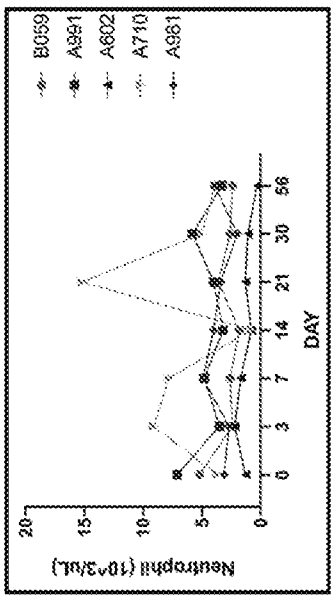

FIG. 11D

SEQUENCE LISTING

<110> Rocket Pharmaceuticals
Keravala, Annahita
Prabhakar, Raj
Shah, Gaurav
Wong, Roderick
Yalamanchi, Naveen
Pratumsuwan, Piratip

<120> GENE THERAPY VECTORS FOR TREATMENT OF DANON DISEASE

<130> ROPA-013/02WO 329592-2080

<150> US 62/804,521
<151> 2019-02-12

<150> US 62/934,928
<151> 2019-11-13

<160> 35

<170> PatentIn version 3.5

<210> 1
<211> 2217
<212> DNA
<213> Non-human primate Adeno-Associated Virus

<400> 1
atggctgccg atggttatct tccagattgg ctcgaggaca acctctctga gggcattcgc 60

gagtggtggg acctgaaacc tggagccccg aaacccaaag ccaaccagca aaagcaggac 120

aacggccggg gtctggtgct tcctggctac aagtacctcg gacccttcaa cggactcgac 180

aaggggggagc ccgtcaacgc ggcggacgca gcggccctcg agcacgacaa ggcctacgac 240

cagcagctcc aagcgggtga caatccgtac ctgcggtata atcacgccga cgccgagttt 300

caggagcgtc tgcaagaaga tacgtctttt gggggcaacc tcgggcgcgc agtcttccag 360

gccaaaaagc gggttctcga acctctgggc ctggttgaat cgccggttaa gacggctcct 420

ggaaagaaga gaccggtaga gccatcaccc cagcgctctc cagactcctc tacgggcatc 480

ggcaagaaag gccagcagcc cgcaaaaaag agactcaatt ttgggcagac tggcgactca 540

gagtcagtcc ccgaccctca accaatcgga gaaccaccag caggcccctc tggtctggga 600

tctggtacaa tggctgcagg cggtggcgct ccaatggcag acaataacga aggcgccgac 660

ggagtgggta gttcctcagg aaattggcat tgcgattcca catggctggg cgacagagtc 720

atcaccacca gcacccgcac ctgggccctg cccacctaca acaaccacct ctacaagcaa 780

atctccaacg ggacctcggg aggaagcacc aacgacaaca cctacttcgg ctacagcacc 840

ccctgggggt attttgactt caacagattc cactgccact tttcaccacg tgactggcag 900

cgactcatca acaacaactg gggattccgg cccaagaggc tcaacttcaa gctcttcaac 960

atccaagtca aggaggtcac gcagaatgaa ggcaccaaga ccatcgccaa taaccttacc 1020

agcacgattc aggtctttac ggactcggaa taccagctcc cgtacgtgct cggctcggcg 1080

caccagggct gcctgcctcc gttcccggcg gacgtcttca tgattcctca gtacgggtac 1140

ctgactctga acaatggcag tcaggctgtg ggccggtcgt ccttctactg cctggagtac 1200

tttccttctc aaatgctgag aacgggcaac aactttgaat tcagctacaa cttcgaggac 1260

gtgcccttcc acagcagcta cgcgcacagc cagagcctgg accggctgat gaaccctctc 1320

atcgaccagt acttgtacta cctgtcccgg actcaaagca cgggcggtac tgcaggaact 1380

cagcagttgc tattttctca ggccgggcct aacaacatgt cggctcaggc caagaactgg 1440

ctacccggtc cctgctaccg gcagcaacgc gtctccacga cactgtcgca gaacaacaac 1500

agcaactttg cctggacggg tgccaccaag tatcatctga atggcagaga ctctctggtg 1560

aatcctggcg ttgccatggc tacccacaag gacgacgaag agcgattttt tccatccagc 1620

ggagtcttaa tgtttgggaa acagggagct ggaaaagaca acgtggacta tagcagcgtg 1680

atgctaacca gcgaggaaga aataaagacc accaacccag tggccacaga acagtacggc 1740

gtggtggccg ataacctgca acagcaaaac gccgctccta ttgtaggggc cgtcaatagt 1800

caaggagcct tacctggcat ggtgtggcag aaccgggacg tgtacctgca gggtcccatc 1860

tgggccaaga ttcctcatac ggacggcaac tttcatccct cgccgctgat gggaggcttt 1920

ggactgaagc atccgcctcc tcagatcctg attaaaaaca cacctgttcc cgcggatcct 1980

ccgaccacct tcaatcaggc caagctggct tctttcatca cgcagtacag taccggccag 2040

gtcagcgtgg agatcgagtg ggagctgcag aaggagaaca gcaaacgctg gaacccagag 2100

attcagtaca cttccaacta ctacaaatct acaaatgtgg actttgctgt caatactgag 2160

ggtacttatt ccgagcctcg ccccattggc acccgttacc tcacccgtaa tctgtaa 2217

<210> 2
<211> 535
<212> PRT

<213> Non-human primate adeno-associated Virus

<400> 2

```
Met Ala Ala Gly Gly Gly Ala Pro Met Ala Asp Asn Asn Glu Gly Ala
1               5                  10                  15


Asp Gly Val Gly Ser Ser Ser Gly Asn Trp His Cys Asp Ser Thr Trp
            20                  25                  30


Leu Gly Asp Arg Val Ile Thr Thr Ser Thr Arg Thr Trp Ala Leu Pro
        35                  40                  45


Thr Tyr Asn Asn His Leu Tyr Lys Gln Ile Ser Asn Gly Thr Ser Gly
    50                  55                  60


Gly Ser Thr Asn Asp Asn Thr Tyr Phe Gly Tyr Ser Thr Pro Trp Gly
65                  70                  75                  80


Tyr Phe Asp Phe Asn Arg Phe His Cys His Phe Ser Pro Arg Asp Trp
                85                  90                  95


Gln Arg Leu Ile Asn Asn Asn Trp Gly Phe Arg Pro Lys Arg Leu Asn
            100                 105                 110


Phe Lys Leu Phe Asn Ile Gln Val Lys Glu Val Thr Gln Asn Glu Gly
        115                 120                 125


Thr Lys Thr Ile Ala Asn Asn Leu Thr Ser Thr Ile Gln Val Phe Thr
    130                 135                 140


Asp Ser Glu Tyr Gln Leu Pro Tyr Val Leu Gly Ser Ala His Gln Gly
145                 150                 155                 160


Cys Leu Pro Pro Phe Pro Ala Asp Val Phe Met Ile Pro Gln Tyr Gly
                165                 170                 175


Tyr Leu Thr Leu Asn Asn Gly Ser Gln Ala Val Gly Arg Ser Ser Phe
            180                 185                 190


Tyr Cys Leu Glu Tyr Phe Pro Ser Gln Met Leu Arg Thr Gly Asn Asn
        195                 200                 205
```

```
Phe Glu Phe Ser Tyr Asn Phe Glu Asp Val Pro Phe His Ser Ser Tyr
    210                 215                 220

Ala His Ser Gln Ser Leu Asp Arg Leu Met Asn Pro Leu Ile Asp Gln
225                 230                 235                 240

Tyr Leu Tyr Tyr Leu Ser Arg Thr Gln Ser Thr Gly Gly Thr Ala Gly
                245                 250                 255

Thr Gln Gln Leu Leu Phe Ser Gln Ala Gly Pro Asn Asn Met Ser Ala
            260                 265                 270

Gln Ala Lys Asn Trp Leu Pro Gly Pro Cys Tyr Arg Gln Gln Arg Val
        275                 280                 285

Ser Thr Thr Leu Ser Gln Asn Asn Asn Ser Asn Phe Ala Trp Thr Gly
    290                 295                 300

Ala Thr Lys Tyr His Leu Asn Gly Arg Asp Ser Leu Val Asn Pro Gly
305                 310                 315                 320

Val Ala Met Ala Thr His Lys Asp Asp Glu Glu Arg Phe Phe Pro Ser
                325                 330                 335

Ser Gly Val Leu Met Phe Gly Lys Gln Gly Ala Gly Lys Asp Asn Val
            340                 345                 350

Asp Tyr Ser Ser Val Met Leu Thr Ser Glu Glu Glu Ile Lys Thr Thr
        355                 360                 365

Asn Pro Val Ala Thr Glu Gln Tyr Gly Val Val Ala Asp Asn Leu Gln
    370                 375                 380

Gln Gln Asn Ala Ala Pro Ile Val Gly Ala Val Asn Ser Gln Gly Ala
385                 390                 395                 400

Leu Pro Gly Met Val Trp Gln Asn Arg Asp Val Tyr Leu Gln Gly Pro
                405                 410                 415
```

```
Ile Trp Ala Lys Ile Pro His Thr Asp Gly Asn Phe His Pro Ser Pro
            420                 425                 430

Leu Met Gly Gly Phe Gly Leu Lys His Pro Pro Pro Gln Ile Leu Ile
        435                 440                 445

Lys Asn Thr Pro Val Pro Ala Asp Pro Pro Thr Thr Phe Asn Gln Ala
    450                 455                 460

Lys Leu Ala Ser Phe Ile Thr Gln Tyr Ser Thr Gly Gln Val Ser Val
465                 470                 475                 480

Glu Ile Glu Trp Glu Leu Gln Lys Glu Asn Ser Lys Arg Trp Asn Pro
                485                 490                 495

Glu Ile Gln Tyr Thr Ser Asn Tyr Tyr Lys Ser Thr Asn Val Asp Phe
            500                 505                 510

Ala Val Asn Thr Glu Gly Thr Tyr Ser Glu Pro Arg Pro Ile Gly Thr
        515                 520                 525

Arg Tyr Leu Thr Arg Asn Leu
    530                 535
```

<210> 3
<211> 398
<212> PRT
<213> Non-human primate adeno-associated virus

<400> 3

```
Ser Thr Ile Gln Val Phe Thr Asp Ser Glu Tyr Gln Leu Pro Tyr Val
1               5                   10                  15

Leu Gly Ser Ala His Gln Gly Cys Leu Pro Pro Phe Pro Ala Asp Val
            20                  25                  30

Phe Met Ile Pro Gln Tyr Gly Tyr Leu Thr Leu Asn Asn Gly Ser Gln
        35                  40                  45

Ala Val Gly Arg Ser Ser Phe Tyr Cys Leu Glu Tyr Phe Pro Ser Gln
    50                  55                  60
```

```
Met Leu Arg Thr Gly Asn Asn Phe Glu Phe Ser Tyr Asn Phe Glu Asp
65                  70                  75                  80

Val Pro Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu Asp Arg Leu
                85                  90                  95

Met Asn Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser Arg Thr Gln
            100                 105                 110

Ser Thr Gly Gly Thr Ala Gly Thr Gln Gln Leu Leu Phe Ser Gln Ala
        115                 120                 125

Gly Pro Asn Asn Met Ser Ala Gln Ala Lys Asn Trp Leu Pro Gly Pro
    130                 135                 140

Cys Tyr Arg Gln Gln Arg Val Ser Thr Thr Leu Ser Gln Asn Asn Asn
145                 150                 155                 160

Ser Asn Phe Ala Trp Thr Gly Ala Thr Lys Tyr His Leu Asn Gly Arg
                165                 170                 175

Asp Ser Leu Val Asn Pro Gly Val Ala Met Ala Thr His Lys Asp Asp
            180                 185                 190

Glu Glu Arg Phe Phe Pro Ser Ser Gly Val Leu Met Phe Gly Lys Gln
        195                 200                 205

Gly Ala Gly Lys Asp Asn Val Asp Tyr Ser Ser Val Met Leu Thr Ser
    210                 215                 220

Glu Glu Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Gln Tyr Gly
225                 230                 235                 240

Val Val Ala Asp Asn Leu Gln Gln Gln Asn Ala Ala Pro Ile Val Gly
                245                 250                 255

Ala Val Asn Ser Gln Gly Ala Leu Pro Gly Met Val Trp Gln Asn Arg
            260                 265                 270

Asp Val Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His Thr Asp
```

```
        275                 280                 285

Gly Asn Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Leu Lys His
    290                 295                 300

Pro Pro Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala Asp Pro
305                 310                 315                 320

Pro Thr Thr Phe Asn Gln Ala Lys Leu Ala Ser Phe Ile Thr Gln Tyr
                325                 330                 335

Ser Thr Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln Lys Glu
            340                 345                 350

Asn Ser Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn Tyr Tyr
        355                 360                 365

Lys Ser Thr Asn Val Asp Phe Ala Val Asn Thr Glu Gly Thr Tyr Ser
    370                 375                 380

Glu Pro Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
385                 390                 395


<210>  4
<211>  332
<212>  PRT
<213>  Non-human primate adeno-associated virus

<400>  4

Arg Thr Gly Asn Asn Phe Glu Phe Ser Tyr Asn Phe Glu Asp Val Pro
1               5                   10                  15

Phe His Ser Ser Tyr Ala His Ser Gln Ser Leu Asp Arg Leu Met Asn
            20                  25                  30

Pro Leu Ile Asp Gln Tyr Leu Tyr Tyr Leu Ser Arg Thr Gln Ser Thr
        35                  40                  45

Gly Gly Thr Ala Gly Thr Gln Gln Leu Leu Phe Ser Gln Ala Gly Pro
    50                  55                  60
```

```
Asn Asn Met Ser Ala Gln Ala Lys Asn Trp Leu Pro Gly Pro Cys Tyr
65                  70                  75                  80

Arg Gln Gln Arg Val Ser Thr Thr Leu Ser Gln Asn Asn Asn Ser Asn
                85                  90                  95

Phe Ala Trp Thr Gly Ala Thr Lys Tyr His Leu Asn Gly Arg Asp Ser
            100                 105                 110

Leu Val Asn Pro Gly Val Ala Met Ala Thr His Lys Asp Asp Glu Glu
        115                 120                 125

Arg Phe Phe Pro Ser Ser Gly Val Leu Met Phe Gly Lys Gln Gly Ala
    130                 135                 140

Gly Lys Asp Asn Val Asp Tyr Ser Ser Val Met Leu Thr Ser Glu Glu
145                 150                 155                 160

Glu Ile Lys Thr Thr Asn Pro Val Ala Thr Glu Gln Tyr Gly Val Val
                165                 170                 175

Ala Asp Asn Leu Gln Gln Gln Asn Ala Ala Pro Ile Val Gly Ala Val
            180                 185                 190

Asn Ser Gln Gly Ala Leu Pro Gly Met Val Trp Gln Asn Arg Asp Val
        195                 200                 205

Tyr Leu Gln Gly Pro Ile Trp Ala Lys Ile Pro His Thr Asp Gly Asn
    210                 215                 220

Phe His Pro Ser Pro Leu Met Gly Gly Phe Gly Leu Lys His Pro Pro
225                 230                 235                 240

Pro Gln Ile Leu Ile Lys Asn Thr Pro Val Pro Ala Asp Pro Pro Thr
                245                 250                 255

Thr Phe Asn Gln Ala Lys Leu Ala Ser Phe Ile Thr Gln Tyr Ser Thr
            260                 265                 270

Gly Gln Val Ser Val Glu Ile Glu Trp Glu Leu Gln Lys Glu Asn Ser
        275                 280                 285
```

Lys Arg Trp Asn Pro Glu Ile Gln Tyr Thr Ser Asn Tyr Tyr Lys Ser
290 295 300

Thr Asn Val Asp Phe Ala Val Asn Thr Glu Gly Thr Tyr Ser Glu Pro
305 310 315 320

Arg Pro Ile Gly Thr Arg Tyr Leu Thr Arg Asn Leu
325 330

<210> 5
<211> 409
<212> PRT
<213> Homo sapiens

<400> 5

Met Val Cys Phe Arg Leu Phe Pro Val Pro Gly Ser Gly Leu Val Leu
1 5 10 15

Val Cys Leu Val Leu Gly Ala Val Arg Ser Tyr Ala Leu Glu Leu Asn
20 25 30

Leu Thr Asp Ser Glu Asn Ala Thr Cys Leu Tyr Ala Lys Trp Gln Met
35 40 45

Asn Phe Thr Val Arg Tyr Glu Thr Thr Asn Lys Thr Tyr Lys Thr Val
50 55 60

Thr Ile Ser Asp His Gly Thr Val Thr Tyr Asn Gly Ser Ile Cys Gly
65 70 75 80

Asp Asp Gln Asn Gly Pro Lys Ile Ala Val Gln Phe Gly Pro Gly Phe
85 90 95

Ser Trp Ile Ala Asn Phe Thr Lys Ala Ala Ser Thr Tyr Ser Ile Asp
100 105 110

Ser Val Ser Phe Ser Tyr Asn Thr Gly Asp Asn Thr Thr Phe Pro Asp
115 120 125

Ala Glu Asp Lys Gly Ile Leu Thr Val Asp Glu Leu Leu Ala Ile Arg

```
    130                 135                 140

Ile Pro Leu Asn Asp Leu Phe Arg Cys Asn Ser Leu Ser Thr Leu Glu
145                 150                 155                 160

Lys Asn Asp Val Val Gln His Tyr Trp Asp Val Leu Val Gln Ala Phe
                165                 170                 175

Val Gln Asn Gly Thr Val Ser Thr Asn Glu Phe Leu Cys Asp Lys Asp
            180                 185                 190

Lys Thr Ser Thr Val Ala Pro Thr Ile His Thr Thr Val Pro Ser Pro
        195                 200                 205

Thr Thr Thr Pro Thr Pro Lys Glu Lys Pro Glu Ala Gly Thr Tyr Ser
    210                 215                 220

Val Asn Asn Gly Asn Asp Thr Cys Leu Leu Ala Thr Met Gly Leu Gln
225                 230                 235                 240

Leu Asn Ile Thr Gln Asp Lys Val Ala Ser Val Ile Asn Ile Asn Pro
                245                 250                 255

Asn Thr Thr His Ser Thr Gly Ser Cys Arg Ser His Thr Ala Leu Leu
            260                 265                 270

Arg Leu Asn Ser Ser Thr Ile Lys Tyr Leu Asp Phe Val Phe Ala Val
        275                 280                 285

Lys Asn Glu Asn Arg Phe Tyr Leu Lys Glu Val Asn Ile Ser Met Tyr
    290                 295                 300

Leu Val Asn Gly Ser Val Phe Ser Ile Ala Asn Asn Asn Leu Ser Tyr
305                 310                 315                 320

Trp Asp Ala Pro Leu Gly Ser Ser Tyr Met Cys Asn Lys Glu Gln Thr
                325                 330                 335

Val Ser Val Ser Gly Ala Phe Gln Ile Asn Thr Phe Asp Leu Arg Val
            340                 345                 350
```

```
Gln Pro Phe Asn Val Thr Gln Gly Lys Tyr Ser Thr Ala Gln Glu Cys
        355                 360                 365

Ser Leu Asp Asp Asp Thr Ile Leu Ile Pro Ile Ile Val Gly Ala Gly
    370                 375                 380

Leu Ser Gly Leu Ile Ile Val Ile Val Ile Ala Tyr Val Ile Gly Arg
385                 390                 395                 400

Arg Lys Ser Tyr Ala Gly Tyr Gln Thr
                405
```

<210> 6
<211> 1233
<212> DNA
<213> Homo sapiens

<400> 6

```
atggtgtgct tccgcctctt cccggttccg ggctcagggc tcgttctggt ctgcctagtc       60

ctgggagctg tgcggtctta tgcattggaa cttaatttga cagattcaga aaatgccact      120

tgcctttatg caaaatggca gatgaatttc acagttcgct atgaaactac aaataaaact      180

tataaaactg taaccatttc agaccatggc actgtgacat ataatggaag catttgtggg      240

gatgatcaga atggtcccaa aatagcagtg cagttcggac ctggcttttc ctggattgcg      300

aattttacca aggcagcatc tacttattca attgacagcg tctcattttc ctacaacact      360

ggtgataaca caacatttcc tgatgctgaa gataaaggaa ttcttactgt tgatgaactt      420

ttggccatca gaattccatt gaatgacctt tttagatgca atagtttatc aactttggaa      480

aagaatgatg ttgtccaaca ctactgggat gttcttgtac aagcttttgt ccaaaatggc      540

acagtgagca caaatgagtt cctgtgtgat aaagacaaaa cttcaacagt ggcacccacc      600

atacacacca ctgtgccatc tcctactaca acacctactc caaaggaaaa accagaagct      660

ggaacctatt cagttaataa tggcaatgat acttgtctgc tggctaccat ggggctgcag      720

ctgaacatca ctcaggataa ggttgcttca gttattaaca tcaaccccaa tacaactcac      780

tccacaggca gctgccgttc tcacactgct ctacttagac tcaatagcag caccattaag      840

tatctagact ttgtctttgc tgtgaaaaat gaaaaccgat tttatctgaa ggaagtgaac      900

atcagcatgt atttggttaa tggctccgtt ttcagcattg caaataacaa tctcagctac      960
```

tgggatgccc ccctgggaag ttcttatatg tgcaacaaag agcagactgt ttcagtgtct 1020

ggagcatttc agataaatac ctttgatcta agggttcagc ctttcaatgt gacacaagga 1080

aagtattcta cagcccaaga gtgttcgctg gatgatgaca ccattctaat cccaattata 1140

gttggtgctg gtctttcagg cttgattatc gttatagtga ttgcttacgt aattggcaga 1200

agaaaaagtt atgctggata tcagactctg taa 1233

<210> 7
<211> 1233
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab - Engineered Lamp-2B coding sequence

<400> 7
atggtctgct tcagactgtt ccctgtccct ggatctggtc tggtgcttgt gtgcttggtg 60

ctgggtgctg tgagatccta tgcccttgag ctgaacctga ctgactcaga aaatgccact 120

tgcctgtatg ccaagtggca gatgaacttc actgtgagat atgagactac caacaagacc 180

tacaagactg tgaccatctc agaccatggc actgtcacct acaatggatc aatctgtggt 240

gatgatcaga atggcccaaa gatagcagtg cagtttgggc ccggtttttc ctggattgct 300

aacttcacca aggcagcctc cacctacagc attgactcag tcagcttcag ctacaacact 360

ggggataaca ccaccttccc tgacgcagag gacaagggaa tccttactgt ggacgaactc 420

ctggcaatca gaatcccccc taacgacctg ttcagatgca actccctttc aacccttgaa 480

aagaatgatg tggtgcaaca ctattgggac gtcctggtgc aagcctttgt gcagaatggg 540

acagtgagta ccaacgagtt cctctgtgac aaggacaaga ccagcactgt ggcccccact 600

atccacacca ctgtgcccag ccctaccact acccccaccc ctaaagagaa gccagaagct 660

ggaacctact cagtcaacaa tggaaatgac acatgcctcc ttgccaccat gggactgcag 720

ctgaacatca ctcaggacaa ggtggcctca gtgattaaca tcaaccctaa caccactcat 780

agcactggga gctgcagatc acatacagct ctgctgaggc tcaactcctc caccatcaag 840

tacctggact ttgtgtttgc tgtgaagaat gagaacaggt tctacctcaa ggaagtgaac 900

atttccatgt acctggtcaa tggttcagtg ttctctattg ccaacaacaa tctgagctac 960

tgggatgcac ccctgggatc ctcctacatg tgcaacaagg agcagactgt gagtgtgtca 1020

ggtgcttttc agatcaacac ttttgacctg agggtgcagc ccttcaatgt gactcaggga 1080

aagtactcca ctgcacaaga gtgttccttg gatgatgaca ctatcctcat ccccattatt 1140

gtgggagctg gactgtcagg attgattata gtgattgtga ttgcttatgt gattggaagg 1200

agaaagagct atgctggcta ccagaccctg taa 1233

<210> 8
<211> 1233
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab - Engineered Lamp-2B coding sequence

<400> 8
atggtgtgct ttagactgtt tcctgtgcct ggttcagggc tggtcctggt ctgtctggtg 60

ctgggggctg tcagaagcta tgccttggag ctgaacctca ctgatagtga aaatgccact 120

tgtctgtatg ctaagtggca gatgaacttc actgtgagat atgaaaccac caacaagact 180

tacaaaacag tgaccatctc agatcatgga actgtgacct acaacggcag catttgtgga 240

gacgaccaga acggaccaaa aatcgctgtc caatttgggc ctggattctc ctggattgcc 300

aatttcacta aagctgcctc cacatattca attgactcag tgtccttctc ctacaacact 360

ggggacaaca ctactttccc tgatgctgaa gataagggaa tcttgacagt ggatgagctg 420

ctggctatca ggatcccttt gaatgacctg tttaggtgta attcactgag cactctggag 480

aagaacgacg tggtgcagca ctactgggac gtgctggtgc aggcctttgt gcagaacggc 540

actgtgtcca ccaacgaatt cctgtgtgat aaggacaaaa cttccactgt ggcacctaca 600

attcacacta ctgtgccttc acctaccacc actccaactc caaaggaaaa gcctgaagca 660

ggaacctact ctgtgaacaa tggcaatgat acctgtctgt tggccaccat gggcctccaa 720

ctgaacatta ctcaggacaa ggtggcctca gtgattaaca ttaaccccaa cactacccac 780

tccactggca gctgtagatc acacacagcc ttgctcagac tgaatagcag caccatcaag 840

tatttggatt ttgtgtttgc agtgaagaat gaaaacaggt tctacctgaa ggaagtcaac 900

atctcaatgt acctggtgaa cggctcagtg ttcagcattg ccaacaacaa cctctcctat 960

tgggacgctc cactggggag cagctacatg tgtaacaagg aacagactgt gtcagtgtca 1020

ggagccttcc agattaacac ctttgatctg agggtccaac cctttaatgt cactcaagga 1080

aagtatagca ctgcccagga gtgctccctg gatgatgaca ccattctgat tccaatcatt 1140

gtgggtgcag gactttctgg gcttattatt gtgattgtga ttgcctatgt gattggcaga 1200

aggaaatcct atgcagggta ccaaactctg taa 1233

<210> 9
<211> 1233
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab - Engineered Lamp-2B coding sequence

<400> 9
atggtctgtt ttaggctgtt ccctgtccct ggttcaggac tggtcttagt gtgtctggtg 60

cttggagctg tcagaagcta tgccctggag ctgaacctga ctgactcaga aaatgccact 120

tgcctgtatg ccaagtggca gatgaacttc actgtcagat atgaaaccac caacaagacc 180

tataagactg tgaccatctc agaccatggc actgtgactt acaatgggtc aatttgtgga 240

gatgaccaga atggccctaa gatagctgtc cagtttggtc caggattcag ctggattgcc 300

aacttcacca aggcagccag cacctacagc attgactctg tgtccttctc ctacaacaca 360

ggagacaaca ccactttccc tgatgcagag gacaaaggta tcctgactgt ggatgagttg 420

ctggcaatca ggatcccact gaacgatctg ttcaggtgca actcactgtc cactctggaa 480

aagaatgatg tggtgcagca ctattgggat gtgctagtcc aggcctttgt ccagaatggg 540

actgtgtcaa ctaatgagtt cctgtgtgac aaggacaaga caagcactgt agcccccact 600

atccatacca cagtacctag ccccaccact actccaaccc ccaaggagaa gcctgaggct 660

ggcacctact cagtgaacaa tgggaatgac acctgtttgc tggccactat gggactccaa 720

ctgaacatca cccaggacaa agtggcctct gtgatcaata tcaatcccaa caccacccac 780

agcactgggt cctgcagaag ccacactgcc ctcctgaggc tcaactcatc aactatcaag 840

tacttggatt ttgtgtttgc agtgaagaat gagaacagat tctacctcaa agaggtcaac 900

atttcaatgt acctggtgaa tgggagtgtg ttctccattg ctaacaacaa cctgagctac 960

tgggatgccc ctctgggctc ctcatacatg tgcaacaagg aacagactgt gagtgtgtca 1020

ggggccttcc agatcaacac ttttgacctg agagtgcagc cctttaatgt gacacaggga 1080

aagtacagca ctgctcagga gtgcagcctg gatgatgaca ctatcctgat ccctatcatt 1140

gtgggggcag gcctgtctgg actcattatt gtgattgtga ttgcctatgt gatagggaga 1200

aggaagtctt atgctggata ccagaccctg taa 1233

<210> 10
<211> 4549
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab - Engineered expression cassette

<400> 10
ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120

aggggttcct tgtagttaat gattaacccg ccatgctact tatctaccag ggtaatgggg 180

atcctctaga actatagcta gtcgacattg attattgact agttattaat agtaatcaat 240

tacggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa 300

tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt 360

tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta 420

aactgcccac ttggcagtac atcaagtgta tcatatgcca agtacgcccc ctattgacgt 480

caatgacggt aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc 540

tacttggcag tacatctacg tattagtcat cgctattacc atggtcgagg tgagccccac 600

gttctgcttc actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat 660

tttttaatta ttttgtgcag cgatgggggc gggggggggg gggggcgcgc cgccaggcgg 720

ggcggggcgg ggcgaggggc ggggcggggc gaggcggaga ggtgcggcgg cagccaatca 780

gagcggcgcg ctccgaaagt ttccttttat ggcgaggcgg cggcggcggc ggccctataa 840

aaagcgaagc gcgcggcggg cgggagtcgc tgcgcgctgc cttcgccccg tgccccgctc 900

cgccgccgcc tcgcgccgcc cgccccggct ctgactgacc gcgttactcc cacaggtgag 960

cgggcgggac ggcccttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt 1020

ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg 1080

gagcggctcg gggggtgcgt gcgtgtgtgt gtgcgtgggg agcgccgcgt gcggctccgc 1140

gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgcagtgt 1200

gcgcgagggg agcgcggccg ggggcggtgc cccgcggtgc gggggggct gcgaggggaa 1260

caaaggctgc gtgcggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcgtcggt 1320

cgggctgcaa ccccccctgc acccccctcc ccgagttgct gagcacggcc cggcttcggg 1380

tgcggggctc cgtacggggc gtggcgcggg gctcgccgtg ccgggcgggg ggtggcggca 1440

ggtgggggtg ccgggcgggg cggggccgcc tcgggccggg gagggctcgg gggaggggcg 1500

cggcggcccc cggagcgccg gcggctgtcg aggcgcggcg agccgcagcc attgcctttt 1560

atggtaatcg tgcgagaggg cgcagggact tcctttgtcc caaatctgtg cggagccgaa 1620

atctgggagg cgccgccgca ccccctctag cgggcgcggg gcgaagcggt gcggcgccgg 1680

caggaaggaa atgggcgggg agggccttcg tgcgtcgccg cgccgccgtc cccttctccc 1740

tctccagcct cggggctgtc cgcgggggga cggctgcctt cgggggggac ggggcagggc 1800

ggggttcggc ttctggcgtg tgaccggcgg ctctagagcc tctgctaacc atgttcatgc 1860

cttcttcttt ttcctacagc tcctgggcaa cgtgctggtt attgtgctgt ctcatcattt 1920

tggcaaagaa ttcgagcggc cgccagccgc caccatggtc tgcttcagac tgttccctgt 1980

ccctggatct ggtctggtgc ttgtgtgctt ggtgctgggt gctgtgagat cctatgccct 2040

tgagctgaac ctgactgact cagaaaatgc cacttgcctg tatgccaagt ggcagatgaa 2100

cttcactgtg agatatgaga ctaccaacaa gacctacaag actgtgacca tctcagacca 2160

tggcactgtc acctacaatg gatcaatctg tggtgatgat cagaatggcc caaagatagc 2220

agtgcagttt gggcccggtt tttcctggat tgctaacttc accaaggcag cctccaccta 2280

cagcattgac tcagtcagct tcagctacaa cactggggat aacaccacct tccctgacgc 2340

agaggacaag ggaatcctta ctgtggacga actcctggca atcagaatcc cccttaacga 2400

cctgttcaga tgcaactccc tttcaaccct tgaaaagaat gatgtggtgc aacactattg 2460

ggacgtcctg gtgcaagcct ttgtgcagaa tgggacagtg agtaccaacg agttcctctg 2520

tgacaaggac aagaccagca ctgtggcccc cactatccac accactgtgc ccagccctac 2580

cactaccccc acccctaaag agaagccaga agctggaacc tactcagtca acaatggaaa 2640

tgacacatgc ctccttgcca ccatgggact gcagctgaac atcactcagg acaaggtggc 2700

ctcagtgatt aacatcaacc ctaacaccac tcatagcact gggagctgca gatcacatac 2760

agctctgctg aggctcaact cctccaccat caagtacctg gactttgtgt ttgctgtgaa 2820

gaatgagaac aggttctacc tcaaggaagt gaacatttcc atgtacctgg tcaatggttc 2880

agtgttctct attgccaaca acaatctgag ctactgggat gcacccctgg gatcctccta 2940

catgtgcaac aaggagcaga ctgtgagtgt gtcaggtgct tttcagatca acacttttga 3000

cctgagggtg cagcccttca atgtgactca gggaaagtac tccactgcac aagagtgttc 3060

cttggatgat gacactatcc tcatccccat tattgtggga gctggactgt caggattgat 3120

tatagtgatt gtgattgctt atgtgattgg aaggagaaag agctatgctg gctaccagac 3180

cctgtaaaag ggcgaattcc agcacacgcg tcctaggagc tcgagtacta ctggcggccg 3240

ttactagtgg atccgcggta caagtaagca tgcaagcttc gaggacgggg tgaactacgc 3300

ctgaatcaag cttatcgata aattcgagca tcttaccgcc atttattccc atatttgttc 3360

tgtttttctt gatttgggta tacatttaaa tgttaataaa acaaaatggt ggggcaatca 3420

tttacatttt tagggatatg taattactag ttcaggtgta ttgccacaag acaaacatgt 3480

taagaaactt tcccgttatt tacgctctgt tcctgttaat caacctctgg attacaaaat 3540

ttgtgaaaga ttgactgata ttcttaacta tgttgctcct tttacgctgt gtggatatgc 3600

tgctttaatg cctctgtatc atgctattgc ttcccgtacg gctttcgttt tctcctcctt 3660

gtataaatcc tggttgctgt ctctttatga ggagttgtgg cccgttgtcc gtcaacgtgg 3720

cgtggtgtgc tctgtgtttg ctgacgcaac ccccactggc tggggcattg ccaccacctg 3780

tcaactcctt tctgggactt tcgctttccc cctcccgatc gccacggcag aactcatcgc 3840

cgcctgcctt gcccgctgct ggacaggggc taggttgctg ggcactgata attccgtggt 3900

gttgtcgggg aagggcctcg ataccgtcga tatcgatcct ggctaataaa ggaaatttat 3960

tttcattgca atagtgtgtt ggaatttttt gtgtctctca ctcggaagga catatgggag 4020

ggcaaatcat ttaaaacatc agaatgagta tttggtttag agtttggcaa catatgccca 4080

tatgctggct gccatgaaca aaggttggct ataaagaggt catcagtata tgaaacagcc 4140

ccctgctgtc cattccttat tccatagaaa agccttgact tgaggttaga tttttttat 4200

attttgtttt gtgttatttt tttctttaac atccctaaaa ttttccttac atgttttact 4260

agccagattt ttcctcctct cctgactact cccagtcata gctgtccctc ttctcttatg 4320

gagatcgaag caattcgttg atctgaattt cgaccaccca taatagatct cccattaccc 4380

tggtagataa gtagcatggc gggttaatca ttaactacaa ggaaccccta gtgatggagt 4440

tggccactcc ctctctgcgc gctcgctcgc tcactgaggc cgggcgacca aaggtcgccc 4500

gacgcccggg ctttgcccgg gcggcctcag tgagcgagcg agcgcgcag 4549

<210> 11
<211> 4549
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab - Engineered expression cassette

<400> 11
ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120

aggggttcct tgtagttaat gattaacccg ccatgctact tatctaccag ggtaatgggg 180

atcctctaga actatagcta gtcgacattg attattgact agttattaat agtaatcaat 240

tacggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa 300

tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt 360

tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta 420

aactgcccac ttggcagtac atcaagtgta tcatatgcca agtacgcccc ctattgacgt 480

caatgacggt aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc 540

tacttggcag tacatctacg tattagtcat cgctattacc atggtcgagg tgagccccac 600

gttctgcttc actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat 660

tttttaatta ttttgtgcag cgatgggggc gggggggggg gggggcgcgc cgccaggcgg 720

ggcggggcgg ggcgaggggc ggggcggggc gaggcggaga ggtgcggcgg cagccaatca 780

gagcggcgcg ctccgaaagt ttccttttat ggcgaggcgg cggcggcggc ggccctataa 840

aaagcgaagc gcgcggcggg cgggagtcgc tgcgcgctgc cttcgccccg tgccccgctc 900

cgccgccgcc tcgcgccgcc cgccccggct ctgactgacc gcgttactcc cacaggtgag 960

cgggcgggac ggcccttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt 1020

ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg 1080

gagcggctcg gggggtgcgt gcgtgtgtgt gtgcgtgggg agcgccgcgt gcggctccgc 1140

gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgcagtgt 1200

gcgcgagggg agcgcggccg ggggcggtgc cccgcggtgc gggggggct gcgaggggaa 1260

caaaggctgc gtgcggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcgtcggt 1320

cgggctgcaa ccccccctgc acccccctcc ccgagttgct gagcacggcc cggcttcggg 1380

tgcggggctc cgtacggggc gtggcgcggg gctcgccgtg ccgggcgggg ggtggcggca 1440

ggtgggggtg ccgggcgggg cggggccgcc tcgggccggg gagggctcgg gggaggggcg 1500

cggcggcccc cggagcgccg gcggctgtcg aggcgcggcg agccgcagcc attgcctttt 1560

atggtaatcg tgcgagaggg cgcagggact tcctttgtcc caaatctgtg cggagccgaa 1620

atctgggagg cgccgccgca ccccctctag cgggcgcggg gcgaagcggt gcggcgccgg 1680

caggaaggaa atgggcgggg agggccttcg tgcgtcgccg cgccgccgtc cccttctccc 1740

tctccagcct cggggctgtc cgcgggggga cggctgcctt cgggggggac ggggcagggc 1800

ggggttcggc ttctggcgtg tgaccggcgg ctctagagcc tctgctaacc atgttcatgc 1860

cttcttcttt ttcctacagc tcctgggcaa cgtgctggtt attgtgctgt ctcatcattt 1920

tggcaaagaa ttcgagcggc cgccagccgc caccatggtg tgctttagac tgtttcctgt 1980

gcctggttca gggctggtcc tggtctgtct ggtgctgggg gctgtcagaa gctatgcctt 2040

ggagctgaac ctcactgata gtgaaaatgc cacttgtctg tatgctaagt ggcagatgaa 2100

cttcactgtg agatatgaaa ccaccaacaa gacttacaaa acagtgacca tctcagatca 2160

tggaactgtg acctacaacg gcagcatttg tggagacgac cagaacggac caaaaatcgc 2220

tgtccaattt gggcctggat tctcctggat tgccaatttc actaaagctg cctccacata 2280

ttcaattgac tcagtgtcct tctcctacaa cactggggac aacactactt tccctgatgc 2340

tgaagataag ggaatcttga cagtggatga gctgctggct atcaggatcc ctttgaatga 2400

cctgtttagg tgtaattcac tgagcactct ggagaagaac gacgtggtgc agcactactg 2460

ggacgtgctg gtgcaggcct ttgtgcagaa cggcactgtg tccaccaacg aattcctgtg 2520

tgataaggac aaaacttcca ctgtggcacc tacaattcac actactgtgc cttcacctac 2580

caccactcca actccaaagg aaaagcctga agcaggaacc tactctgtga acaatggcaa 2640

tgatacctgt ctgttggcca ccatgggcct ccaactgaac attactcagg acaaggtggc 2700

ctcagtgatt aacattaacc ccaacactac ccactccact ggcagctgta gatcacacac 2760

agccttgctc agactgaata gcagcaccat caagtatttg gattttgtgt ttgcagtgaa 2820

gaatgaaaac aggttctacc tgaaggaagt caacatctca atgtacctgg tgaacggctc 2880

agtgttcagc attgccaaca acaacctctc ctattgggac gctccactgg ggagcagcta 2940

catgtgtaac aaggaacaga ctgtgtcagt gtcaggagcc ttccagatta acacctttga 3000

tctgagggtc caacccttta atgtcactca aggaaagtat agcactgccc aggagtgctc 3060

cctggatgat gacaccattc tgattccaat cattgtgggt gcaggacttt ctgggcttat 3120

tattgtgatt gtgattgcct atgtgattgg cagaaggaaa tcctatgcag ggtaccaaac 3180

tctgtaaaag ggcgaattcc agcacacgcg tcctaggagc tcgagtacta ctggcggccg 3240

ttactagtgg atccgcggta caagtaagca tgcaagcttc gaggacgggg tgaactacgc 3300

ctgaatcaag cttatcgata aattcgagca tcttaccgcc atttattccc atatttgttc 3360

tgtttttctt gatttgggta tacatttaaa tgttaataaa acaaaatggt ggggcaatca 3420

tttacatttt tagggatatg taattactag ttcaggtgta ttgccacaag acaaacatgt 3480

taagaaactt tcccgttatt tacgctctgt tcctgttaat caacctctgg attacaaaat 3540

ttgtgaaaga ttgactgata ttcttaacta tgttgctcct tttacgctgt gtggatatgc 3600

tgctttaatg cctctgtatc atgctattgc ttcccgtacg gctttcgttt tctcctcctt 3660

gtataaatcc tggttgctgt ctctttatga ggagttgtgg cccgttgtcc gtcaacgtgg 3720

cgtggtgtgc tctgtgtttg ctgacgcaac ccccactggc tggggcattg ccaccacctg 3780

tcaactcctt tctgggactt tcgctttccc cctcccgatc gccacggcag aactcatcgc 3840

cgcctgcctt gcccgctgct ggacaggggc taggttgctg ggcactgata attccgtggt 3900

gttgtcgggg aagggcctcg ataccgtcga tatcgatcct ggctaataaa ggaaatttat 3960

tttcattgca atagtgtgtt ggaatttttt gtgtctctca ctcggaagga catatgggag 4020

ggcaaatcat ttaaaacatc agaatgagta tttggtttag agtttggcaa catatgccca 4080

tatgctggct gccatgaaca aaggttggct ataaagaggt catcagtata tgaaacagcc 4140

ccctgctgtc cattccttat tccatagaaa agccttgact tgaggttaga tttttttttat 4200

attttgtttt gtgttatttt tttctttaac atccctaaaa ttttccttac atgttttact 4260

agccagattt ttcctcctct cctgactact cccagtcata gctgtccctc ttctcttatg 4320

gagatcgaag caattcgttg atctgaattt cgaccaccca taatagatct cccattaccc 4380

tggtagataa gtagcatggc gggttaatca ttaactacaa ggaaccccta gtgatggagt 4440

tggccactcc ctctctgcgc gctcgctcgc tcactgaggc cgggcgacca aaggtcgccc 4500

gacgcccggg ctttgcccgg gcggcctcag tgagcgagcg agcgcgcag 4549

<210> 12
<211> 4549
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab - Engineered expression cassette

<400> 12
ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120

aggggttcct tgtagttaat gattaacccg ccatgctact tatctaccag ggtaatgggg 180

atcctctaga actatagcta gtcgacattg attattgact agttattaat agtaatcaat 240

tacggggtca ttagttcata gcccatatat ggagttccgc gttacataac ttacggtaaa 300

tggcccgcct ggctgaccgc ccaacgaccc ccgcccattg acgtcaataa tgacgtatgt 360

tcccatagta acgccaatag ggactttcca ttgacgtcaa tgggtggagt atttacggta 420

aactgcccac ttggcagtac atcaagtgta tcatatgcca agtacgcccc ctattgacgt 480

caatgacggt aaatggcccg cctggcatta tgcccagtac atgaccttat gggactttcc 540

tacttggcag tacatctacg tattagtcat cgctattacc atggtcgagg tgagccccac 600

gttctgcttc actctcccca tctccccccc ctccccaccc ccaattttgt atttatttat 660

tttttaatta ttttgtgcag cgatgggggc gggggggggg gggggcgcgc cgccaggcgg 720

ggcggggcgg ggcgaggggc ggggcggggc gaggcggaga ggtgcggcgg cagccaatca 780

gagcggcgcg ctccgaaagt ttccttttat ggcgaggcgg cggcggcggc ggccctataa 840

aaagcgaagc gcgcggcggg cgggagtcgc tgcgcgctgc cttcgccccg tgccccgctc 900

cgccgccgcc tcgcgccgcc cgccccggct ctgactgacc gcgttactcc cacaggtgag 960

cgggcgggac ggcccttctc ctccgggctg taattagcgc ttggtttaat gacggcttgt 1020

ttcttttctg tggctgcgtg aaagccttga ggggctccgg gagggccctt tgtgcggggg 1080

gagcggctcg gggggtgcgt gcgtgtgtgt gtgcgtgggg agcgccgcgt gcggctccgc 1140

gctgcccggc ggctgtgagc gctgcgggcg cggcgcgggg ctttgtgcgc tccgcagtgt 1200

gcgcgagggg agcgcggccg ggggcggtgc cccgcggtgc gggggggct gcgaggggaa 1260

caaaggctgc gtgcggggtg tgtgcgtggg ggggtgagca gggggtgtgg gcgcgtcggt 1320

cgggctgcaa ccccccctgc acccccctcc ccgagttgct gagcacggcc cggcttcggg 1380

tgcggggctc cgtacggggc gtggcgcggg gctcgccgtg ccgggcgggg ggtggcggca 1440

ggtgggggtg ccgggcgggg cggggccgcc tcgggccggg gagggctcgg gggaggggcg 1500

cggcggcccc cggagcgccg gcggctgtcg aggcgcggcg agccgcagcc attgcctttt 1560

atggtaatcg tgcgagaggg cgcagggact tcctttgtcc caaatctgtg cggagccgaa 1620

atctgggagg cgccgccgca ccccctctag cgggcgcggg gcgaagcggt gcggcgccgg 1680

caggaaggaa atgggcgggg agggccttcg tgcgtcgccg cgccgccgtc cccttctccc 1740

tctccagcct cggggctgtc cgcgggggga cggctgcctt cgggggggac ggggcagggc 1800

ggggttcggc ttctggcgtg tgaccggcgg ctctagagcc tctgctaacc atgttcatgc 1860

cttcttcttt ttcctacagc tcctgggcaa cgtgctggtt attgtgctgt ctcatcattt 1920

tggcaaagaa ttcgagcggc cgccagccgc caccatggtc tgttttaggc tgttccctgt 1980

ccctggttca ggactggtct tagtgtgtct ggtgcttgga gctgtcagaa gctatgccct 2040

ggagctgaac ctgactgact cagaaaatgc cacttgcctg tatgccaagt ggcagatgaa 2100

cttcactgtc agatatgaaa ccaccaacaa gacctataag actgtgacca tctcagacca 2160

tggcactgtg acttacaatg ggtcaatttg tggagatgac cagaatggcc ctaagatagc 2220

tgtccagttt ggtccaggat tcagctggat tgccaacttc accaaggcag ccagcaccta 2280

cagcattgac tctgtgtcct tctcctacaa cacaggagac aacaccactt tccctgatgc 2340

agaggacaaa ggtatcctga ctgtggatga gttgctggca atcaggatcc cactgaacga 2400

tctgttcagg tgcaactcac tgtccactct ggaaaagaat gatgtggtgc agcactattg 2460

ggatgtgcta gtccaggcct ttgtccagaa tgggactgtg tcaactaatg agttcctgtg 2520

tgacaaggac aagacaagca ctgtagcccc cactatccat accacagtac ctagccccac 2580

cactactcca acccccaagg agaagcctga ggctggcacc tactcagtga acaatgggaa 2640

tgacacctgt ttgctggcca ctatgggact ccaactgaac atcacccagg acaaagtggc 2700

ctctgtgatc aatatcaatc ccaacaccac ccacagcact gggtcctgca gaagccacac 2760

tgccctcctg aggctcaact catcaactat caagtacttg gattttgtgt ttgcagtgaa 2820

gaatgagaac agattctacc tcaaagaggt caacatttca atgtacctgg tgaatgggag 2880

tgtgttctcc attgctaaca acaacctgag ctactgggat gcccctctgg gctcctcata 2940

catgtgcaac aaggaacaga ctgtgagtgt gtcaggggcc ttccagatca acacttttga 3000

cctgagagtg cagcccttta atgtgacaca gggaaagtac agcactgctc aggagtgcag 3060

cctggatgat gacactatcc tgatccctat cattgtgggg gcaggcctgt ctggactcat 3120

tattgtgatt gtgattgcct atgtgatagg gagaaggaag tcttatgctg gataccagac 3180

cctgtaaaag ggcgaattcc agcacacgcg tcctaggagc tcgagtacta ctggcggccg 3240

ttactagtgg atccgcggta caagtaagca tgcaagcttc gaggacgggg tgaactacgc 3300

ctgaatcaag cttatcgata aattcgagca tcttaccgcc atttattccc atatttgttc 3360

tgtttttctt gatttgggta tacatttaaa tgttaataaa acaaaatggt ggggcaatca 3420

tttacatttt tagggatatg taattactag ttcaggtgta ttgccacaag acaaacatgt 3480

taagaaactt tcccgttatt tacgctctgt tcctgttaat caacctctgg attacaaaat 3540

ttgtgaaaga ttgactgata ttcttaacta tgttgctcct tttacgctgt gtggatatgc 3600

tgctttaatg cctctgtatc atgctattgc ttcccgtacg gctttcgttt tctcctcctt 3660

gtataaatcc tggttgctgt ctctttatga ggagttgtgg cccgttgtcc gtcaacgtgg 3720

cgtggtgtgc tctgtgtttg ctgacgcaac ccccactggc tggggcattg ccaccacctg 3780

tcaactcctt tctgggactt tcgctttccc cctcccgatc gccacggcag aactcatcgc 3840

cgcctgcctt gcccgctgct ggacaggggc taggttgctg ggcactgata attccgtggt 3900

gttgtcgggg aagggcctcg ataccgtcga tatcgatcct ggctaataaa ggaaatttat 3960

tttcattgca atagtgtgtt ggaatttttt gtgtctctca ctcggaagga catatgggag 4020

ggcaaatcat ttaaaacatc agaatgagta tttggtttag agtttggcaa catatgccca 4080

tatgctggct gccatgaaca aaggttggct ataaagaggt catcagtata tgaaacagcc 4140

ccctgctgtc cattccttat tccatagaaa agccttgact tgaggttaga tttttttat 4200

attttgtttt gtgttatttt tttctttaac atccctaaaa ttttccttac atgttttact 4260

agccagattt ttcctcctct cctgactact cccagtcata gctgtccctc ttctcttatg 4320

gagatcgaag caattcgttg atctgaattt cgaccaccca taatagatct cccattaccc 4380

tggtagataa gtagcatggc gggttaatca ttaactacaa ggaaccccta gtgatggagt 4440

tggccactcc ctctctgcgc gctcgctcgc tcactgaggc cgggcgacca aaggtcgccc 4500

gacgcccggg ctttgcccgg gcggcctcag tgagcgagcg agcgcgcag 4549

<210> 13
<211> 130
<212> DNA
<213> Adeno-associated viurs

<400> 13
ctgcgcgctc gctcgctcac tgaggccgcc cgggcaaagc ccgggcgtcg ggcgaccttt 60

ggtcgcccgg cctcagtgag cgagcgagcg cgcagagagg gagtggccaa ctccatcact 120

aggggttcct 130

<210> 14
<211> 130
<212> DNA
<213> Adeno-associated virus

<400> 14
aggaacccct agtgatggag ttggccactc cctctctgcg cgctcgctcg ctcactgagg 60

ccgggcgacc aaaggtcgcc cgacgcccgg gctttgcccg ggcggcctca gtgagcgagc 120

gagcgcgcag 130

<210> 15
<211> 410
<212> PRT
<213> Homo sapiens

<400> 15

Met Val Cys Phe Arg Leu Phe Pro Val Pro Gly Ser Gly Leu Val Leu
1 5 10 15

Val Cys Leu Val Leu Gly Ala Val Arg Ser Tyr Ala Leu Glu Leu Asn
20 25 30

Leu Thr Asp Ser Glu Asn Ala Thr Cys Leu Tyr Ala Lys Trp Gln Met
35 40 45

Asn Phe Thr Val Arg Tyr Glu Thr Thr Asn Lys Thr Tyr Lys Thr Val
50 55 60

```
Thr Ile Ser Asp His Gly Thr Val Thr Tyr Asn Gly Ser Ile Cys Gly
65                  70                  75                  80

Asp Asp Gln Asn Gly Pro Lys Ile Ala Val Gln Phe Gly Pro Gly Phe
                85                  90                  95

Ser Trp Ile Ala Asn Phe Thr Lys Ala Ala Ser Thr Tyr Ser Ile Asp
            100                 105                 110

Ser Val Ser Phe Ser Tyr Asn Thr Gly Asp Asn Thr Thr Phe Pro Asp
        115                 120                 125

Ala Glu Asp Lys Gly Ile Leu Thr Val Asp Glu Leu Leu Ala Ile Arg
    130                 135                 140

Ile Pro Leu Asn Asp Leu Phe Arg Cys Asn Ser Leu Ser Thr Leu Glu
145                 150                 155                 160

Lys Asn Asp Val Val Gln His Tyr Trp Asp Val Leu Val Gln Ala Phe
                165                 170                 175

Val Gln Asn Gly Thr Val Ser Thr Asn Glu Phe Leu Cys Asp Lys Asp
            180                 185                 190

Lys Thr Ser Thr Val Ala Pro Thr Ile His Thr Thr Val Pro Ser Pro
        195                 200                 205

Thr Thr Thr Pro Thr Pro Lys Glu Lys Pro Glu Ala Gly Thr Tyr Ser
    210                 215                 220

Val Asn Asn Gly Asn Asp Thr Cys Leu Leu Ala Thr Met Gly Leu Gln
225                 230                 235                 240

Leu Asn Ile Thr Gln Asp Lys Val Ala Ser Val Ile Asn Ile Asn Pro
                245                 250                 255

Asn Thr Thr His Ser Thr Gly Ser Cys Arg Ser His Thr Ala Leu Leu
            260                 265                 270

Arg Leu Asn Ser Ser Thr Ile Lys Tyr Leu Asp Phe Val Phe Ala Val
```

```
        275                 280                 285

Lys Asn Glu Asn Arg Phe Tyr Leu Lys Glu Val Asn Ile Ser Met Tyr
    290                 295                 300

Leu Val Asn Gly Ser Val Phe Ser Ile Ala Asn Asn Asn Leu Ser Tyr
305                 310                 315                 320

Trp Asp Ala Pro Leu Gly Ser Ser Tyr Met Cys Asn Lys Glu Gln Thr
                325                 330                 335

Val Ser Val Ser Gly Ala Phe Gln Ile Asn Thr Phe Asp Leu Arg Val
            340                 345                 350

Gln Pro Phe Asn Val Thr Gln Gly Lys Tyr Ser Thr Ala Gln Asp Cys
        355                 360                 365

Ser Ala Asp Asp Asp Asn Phe Leu Val Pro Ile Ala Val Gly Ala Ala
    370                 375                 380

Leu Ala Gly Val Leu Ile Leu Val Leu Leu Ala Tyr Phe Ile Gly Leu
385                 390                 395                 400

Lys His His His Ala Gly Tyr Glu Gln Phe
                405                 410
```

<210> 16
<211> 410
<212> PRT
<213> Homo sapiens

<400> 16

```
Met Val Cys Phe Arg Leu Phe Pro Val Pro Gly Ser Gly Leu Val Leu
1               5                   10                  15

Val Cys Leu Val Leu Gly Ala Val Arg Ser Tyr Ala Leu Glu Leu Asn
            20                  25                  30

Leu Thr Asp Ser Glu Asn Ala Thr Cys Leu Tyr Ala Lys Trp Gln Met
        35                  40                  45
```

```
Asn Phe Thr Val Arg Tyr Glu Thr Thr Asn Lys Thr Tyr Lys Thr Val
    50                  55                  60

Thr Ile Ser Asp His Gly Thr Val Thr Tyr Asn Gly Ser Ile Cys Gly
65                  70                  75                  80

Asp Asp Gln Asn Gly Pro Lys Ile Ala Val Gln Phe Gly Pro Gly Phe
                85                  90                  95

Ser Trp Ile Ala Asn Phe Thr Lys Ala Ala Ser Thr Tyr Ser Ile Asp
            100                 105                 110

Ser Val Ser Phe Ser Tyr Asn Thr Gly Asp Asn Thr Thr Phe Pro Asp
        115                 120                 125

Ala Glu Asp Lys Gly Ile Leu Thr Val Asp Glu Leu Leu Ala Ile Arg
    130                 135                 140

Ile Pro Leu Asn Asp Leu Phe Arg Cys Asn Ser Leu Ser Thr Leu Glu
145                 150                 155                 160

Lys Asn Asp Val Val Gln His Tyr Trp Asp Val Leu Val Gln Ala Phe
                165                 170                 175

Val Gln Asn Gly Thr Val Ser Thr Asn Glu Phe Leu Cys Asp Lys Asp
            180                 185                 190

Lys Thr Ser Thr Val Ala Pro Thr Ile His Thr Thr Val Pro Ser Pro
        195                 200                 205

Thr Thr Thr Pro Thr Pro Lys Glu Lys Pro Glu Ala Gly Thr Tyr Ser
    210                 215                 220

Val Asn Asn Gly Asn Asp Thr Cys Leu Leu Ala Thr Met Gly Leu Gln
225                 230                 235                 240

Leu Asn Ile Thr Gln Asp Lys Val Ala Ser Val Ile Asn Ile Asn Pro
                245                 250                 255

Asn Thr Thr His Ser Thr Gly Ser Cys Arg Ser His Thr Ala Leu Leu
            260                 265                 270
```

```
Arg Leu Asn Ser Ser Thr Ile Lys Tyr Leu Asp Phe Val Phe Ala Val
        275                 280                 285


Lys Asn Glu Asn Arg Phe Tyr Leu Lys Glu Val Asn Ile Ser Met Tyr
    290                 295                 300


Leu Val Asn Gly Ser Val Phe Ser Ile Ala Asn Asn Asn Leu Ser Tyr
305                 310                 315                 320


Trp Asp Ala Pro Leu Gly Ser Ser Tyr Met Cys Asn Lys Glu Gln Thr
                325                 330                 335


Val Ser Val Ser Gly Ala Phe Gln Ile Asn Thr Phe Asp Leu Arg Val
            340                 345                 350


Gln Pro Phe Asn Val Thr Gln Gly Lys Tyr Ser Thr Ala Gln Glu Cys
        355                 360                 365


Ser Leu Asp Asp Asp Thr Ile Leu Ile Pro Ile Ile Val Gly Ala Gly
    370                 375                 380


Leu Ser Gly Leu Ile Ile Val Ile Val Ile Ala Tyr Val Ile Gly Arg
385                 390                 395                 400


Arg Lys Ser Tyr Ala Gly Tyr Gln Thr Leu
                405                 410


<210>  17
<211>  411
<212>  PRT
<213>  Homo sapiens

<400>  17

Met Val Cys Phe Arg Leu Phe Pro Val Pro Gly Ser Gly Leu Val Leu
1               5                   10                  15


Val Cys Leu Val Leu Gly Ala Val Arg Ser Tyr Ala Leu Glu Leu Asn
            20                  25                  30


Leu Thr Asp Ser Glu Asn Ala Thr Cys Leu Tyr Ala Lys Trp Gln Met
```

```
        35                  40                  45

Asn Phe Thr Val Arg Tyr Glu Thr Thr Asn Lys Thr Tyr Lys Thr Val
    50                  55                  60

Thr Ile Ser Asp His Gly Thr Val Thr Tyr Asn Gly Ser Ile Cys Gly
65                  70                  75                  80

Asp Asp Gln Asn Gly Pro Lys Ile Ala Val Gln Phe Gly Pro Gly Phe
                85                  90                  95

Ser Trp Ile Ala Asn Phe Thr Lys Ala Ala Ser Thr Tyr Ser Ile Asp
            100                 105                 110

Ser Val Ser Phe Ser Tyr Asn Thr Gly Asp Asn Thr Thr Phe Pro Asp
        115                 120                 125

Ala Glu Asp Lys Gly Ile Leu Thr Val Asp Glu Leu Leu Ala Ile Arg
    130                 135                 140

Ile Pro Leu Asn Asp Leu Phe Arg Cys Asn Ser Leu Ser Thr Leu Glu
145                 150                 155                 160

Lys Asn Asp Val Val Gln His Tyr Trp Asp Val Leu Val Gln Ala Phe
                165                 170                 175

Val Gln Asn Gly Thr Val Ser Thr Asn Glu Phe Leu Cys Asp Lys Asp
            180                 185                 190

Lys Thr Ser Thr Val Ala Pro Thr Ile His Thr Thr Val Pro Ser Pro
        195                 200                 205

Thr Thr Thr Pro Thr Pro Lys Glu Lys Pro Glu Ala Gly Thr Tyr Ser
    210                 215                 220

Val Asn Asn Gly Asn Asp Thr Cys Leu Leu Ala Thr Met Gly Leu Gln
225                 230                 235                 240

Leu Asn Ile Thr Gln Asp Lys Val Ala Ser Val Ile Asn Ile Asn Pro
                245                 250                 255
```

```
Asn Thr Thr His Ser Thr Gly Ser Cys Arg Ser His Thr Ala Leu Leu
            260                 265                 270

Arg Leu Asn Ser Ser Thr Ile Lys Tyr Leu Asp Phe Val Phe Ala Val
        275                 280                 285

Lys Asn Glu Asn Arg Phe Tyr Leu Lys Glu Val Asn Ile Ser Met Tyr
    290                 295                 300

Leu Val Asn Gly Ser Val Phe Ser Ile Ala Asn Asn Asn Leu Ser Tyr
305                 310                 315                 320

Trp Asp Ala Pro Leu Gly Ser Ser Tyr Met Cys Asn Lys Glu Gln Thr
                325                 330                 335

Val Ser Val Ser Gly Ala Phe Gln Ile Asn Thr Phe Asp Leu Arg Val
            340                 345                 350

Gln Pro Phe Asn Val Thr Gln Gly Lys Tyr Ser Thr Ala Glu Glu Cys
        355                 360                 365

Ser Ala Asp Ser Asp Leu Asn Phe Leu Ile Pro Val Ala Val Gly Val
    370                 375                 380

Ala Leu Gly Phe Leu Ile Ile Val Val Phe Ile Ser Tyr Met Ile Gly
385                 390                 395                 400

Arg Arg Lys Ser Arg Thr Gly Tyr Gln Ser Val
                405                 410
```

<210> 18
<211> 1730
<212> DNA
<213> Artificial Sequence

<220>
<223> Made in Lab - promoter sequence

<400> 18
ctagtcgaca ttgattattg actagttatt aatagtaatc aattacgggg tcattagttc 60

atagcccata tatggagttc cgcgttacat aacttacggt aaatggcccg cctggctgac 120

cgcccaacga cccccgccca ttgacgtcaa taatgacgta tgttcccata gtaacgccaa 180

tagggacttt ccattgacgt caatgggtgg agtatttacg gtaaactgcc cacttggcag 240

tacatcaagt gtatcatatg ccaagtacgc cccctattga cgtcaatgac ggtaaatggc 300

ccgcctggca ttatgcccag tacatgacct tatgggactt tcctacttgg cagtacatct 360

acgtattagt catcgctatt accatggtcg aggtgagccc cacgttctgc ttcactctcc 420

ccatctcccc cccctcccca cccccaattt tgtatttatt tatttttaa ttattttgtg 480

cagcgatggg ggcgggggg ggggggggc gcgcgccagg cggggcgggg cggggcgagg 540

ggcggggcgg ggcgaggcgg agaggtgcgg cggcagccaa tcagagcggc gcgctccgaa 600

agtttccttt tatggcgagg cggcggcggc ggcggcccta taaaaagcga agcgcgcggc 660

gggcgggagt cgctgcgcgc tgccttcgcc ccgtgccccg ctccgccgcc gcctcgcgcc 720

gcccgccccg gctctgactg accgcgttac tcccacaggt gagcgggcgg gacggccctt 780

ctcctccggg ctgtaattag cgcttggttt aatgacggct tgtttctttt ctgtggctgc 840

gtgaaagcct tgaggggctc cgggagggcc ctttgtgcgg ggggagcggc tcgggggtg 900

cgtgcgtgtg tgtgtgcgtg gggagcgccg cgtgcggctc cgcgctgccc ggcggctgtg 960

agcgctgcgg gcgcggcgcg gggctttgtg cgctccgcag tgtgcgcgag gggagcgcgg 1020

ccgggggcgg tgccccgcgg tgcgggggg gctgcgaggg gaacaaaggc tgcgtgcggg 1080

gtgtgtgcgt gggggggtga gcaggggtg tgggcgcgtc ggtcgggctg caacccccc 1140

tgcacccccc tccccgagtt gctgagcacg gcccggcttc gggtgcgggg ctccgtacgg 1200

ggcgtggcgc ggggctcgcc gtgccgggcg gggggtggcg gcaggtgggg gtgccgggcg 1260

gggcggggcc gcctcgggcc ggggagggct cgggggaggg gcgcggcggc ccccggagcg 1320

ccggcggctg tcgaggcgcg gcgagccgca gccattgcct tttatggtaa tcgtgcgaga 1380

gggcgcaggg acttcctttg tcccaaatct gtgcggagcc gaaatctggg aggcgccgcc 1440

gcacccccctc tagcgggcgc ggggcgaagc ggtgcggcgc cggcaggaag gaaatgggcg 1500

gggagggcct tcgtgcgtcg ccgcgccgcc gtcccccttct ccctctccag cctcggggct 1560

gtccgcgggg ggacggctgc cttcgggggg gacggggcag ggcggggttc ggcttctggc 1620

gtgtgaccgg cggctctaga gcctctgcta accatgttca tgccttcttc tttttcctac 1680

agctcctggg caacgtgctg gttattgtgc tgtctcatca ttttggcaaa 1730

<210> 19
<211> 596
<212> DNA
<213> Woodchuck hepatitis virus

<400> 19
attcgagcat cttaccgcca tttattccca tatttgttct gtttttcttg atttgggtat 60

acatttaaat gttaataaaa caaaatggtg gggcaatcat ttacattttt agggatatgt 120

aattactagt tcaggtgtat tgccacaaga caaacatgtt aagaaacttt cccgttattt 180

acgctctgtt cctgttaatc aacctctgga ttacaaaatt tgtgaaagat tgactgatat 240

tcttaactat gttgctcctt ttacgctgtg tggatatgct gctttaatgc ctctgtatca 300

tgctattgct tcccgtacgg ctttcgtttt ctcctccttg tataaatcct ggttgctgtc 360

tctttatgag gagttgtggc ccgttgtccg tcaacgtggc gtggtgtgct ctgtgtttgc 420

tgacgcaacc cccactggct ggggcattgc caccacctgt caactccttt ctgggacttt 480

cgctttcccc ctcccgatcg ccacggcaga actcatcgcc gcctgccttg cccgctgctg 540

gacaggggct aggttgctgg gcactgataa ttccgtggtg ttgtcgggga agggcc 596

<210> 20
<211> 13
<212> DNA
<213> Artificial Sequence

<220>
<223> Kozak consensus sequence

<400> 20
gccgccacca tgg 13

<210> 21
<211> 13
<212> RNA
<213> Artificial Sequence

<220>
<223> Kozak consensus sequence

<400> 21
gccgccrcca ugg 13

<210> 22
<211> 8

<212> RNA
<213> Artificial Sequence

<220>
<223> Kozak consensus sequence

<220>
<221> misc_feature
<222> (3)..(4)
<223> n is a, c, g, or u

<220>
<221> misc_feature
<222> (8)..(8)
<223> n is a, c, g, or u

<400> 22
agnnaugn 8

<210> 23
<211> 7
<212> RNA
<213> Artificial Sequence

<220>
<223> Kozak consensus sequence

<220>
<221> misc_feature
<222> (2)..(3)
<223> n is a, c, g, or u

<400> 23
annaugg 7

<210> 24
<211> 7
<212> RNA
<213> Artificial Sequence

<220>
<223> Kozak consensus sequence

<400> 24
accaugg 7

<210> 25
<211> 10
<212> RNA

<213> Artificial Sequence

<220>
<223> Kozak consensus sequence

<400> 25
gacaccaugg 10

<210> 26
<211> 387
<212> DNA
<213> Mus musculus

<400> 26
tggctaataa aggaaattta ttttcattgc aatagtgtgt tggaattttt tgtgtctctc 60

actcggaagg acatatggga gggcaaatca tttaaaacat cagaatgagt atttggttta 120

gagtttggca acatatgccc atatgctggc tgccatgaac aaaggttggc tataaagagg 180

tcatcagtat atgaaacagc cccctgctgt ccattcctta ttccatagaa aagccttgac 240

ttgaggttag atttttttta tattttgttt tgtgttattt ttttctttaa catccctaaa 300

attttcctta catgttttac tagccagatt tttcctcctc tcctgactac tcccagtcat 360

agctgtccct cttctcttat ggagatc 387

<210> 27
<211> 235
<212> DNA
<213> Bos taurus

<400> 27
tcgactgtgc cttctagttg ccagccatct gttgtttgcc cctcccccgt gccttccttg 60

accctggaag gtgccactcc cactgtcctt tcctaataaa atgaggaaat tgcatcgcat 120

tgtctgagta ggtgtcattc tattctgggg ggtggggtgg ggcaggacag caaggggggag 180

gattgggagg acaatagcag gcatgctggg gatgcggtgg gctctatggc ttctg 235

<210> 28
<211> 222
<212> DNA
<213> Simian viurs 40

<400> 28
cagacatgat aagatacatt gatgagtttg gacaaaccac aactagaatg cagtgaaaaa 60

aatgctttat ttgtgaaatt tgtgatgcta ttgctttatt tgtaaccatt ataagctgca 120

ataaacaagt taacaacaac aattgcattc attttatgtt tcaggttcag ggggagatgt 180

gggaggtttt ttaaagcaag taaaacctct acaaatgtgg ta 222

<210> 29
<211> 202
<212> DNA
<213> Homo sapiens

<400> 29
ctgcccgggt ggcatccctg tgacccctcc ccagtgcctc tcctggccct ggaagttgcc 60

actccagtgc ccaccagcct tgtcctaata aaattaagtt gcatcatttt gtctgactag 120

gtgtccttct ataatattat ggggtggagg ggggtggtat ggagcaaggg gcccaagttg 180

ggaagaaacc tgtagggcct gc 202

<210> 30
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 30
atcatgctat tgcttcccgt a 21

<210> 31
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 31
gggccacaac tcctcataaa 20

<210> 32
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 32

cctccttgta taaatcctgg ttgctgtct 29

<210> 33
<211> 21
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 33
atcatgctat tgcttcccgt a 21

<210> 34
<211> 20
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 34
gggccacaac tcctcataaa 20

<210> 35
<211> 29
<212> DNA
<213> Artificial Sequence

<220>
<223> Primer

<400> 35
cctccttgta taaatcctgg ttgctgtct 29