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

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
14 March 2024 (14.03.2024)

WIPO | PCT

(10) International Publication Number
WO 2024/055020 A2

(51) International Patent Classification:
C12N 15/86 (2006.01) *C07K 14/075* (2006.01)

(21) International Application Number:
PCT/US2023/073808

(22) International Filing Date:
08 September 2023 (08.09.2023)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
63/405,360 09 September 2022 (09.09.2022) US

(71) Applicant: **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 5th Floor, Oakland, California 94607-5200 (US).

(72) Inventors: **MALI, Prashant**; c/o University of California, San Diego, 9500 Gilman Drive MC 0910, La Jolla, California 92093 (US). **PORTELL, Andrew**; c/o University of California, San Diego, 9500 Gilman Drive MC 0910, La Jolla, California 92093 (US). **FORD, Kyle**; c/o University of California, San Diego, 9500 Gilman Drive MC 0910, La Jolla, California 92093 (US). **SUHARDJO, Amanda**; c/o University of California, San Diego, 9500 Gilman Drive MC 0910, La Jolla, California 92093 (US).

(74) Agent: **BAKER, Joseph R., Jr.**; 12526 High Bluff Drive, Suite 210, San Diego, California 92130 (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, CV, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IQ, IR, IS, IT, JM, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, MG, MK, MN, MU, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, WS, ZA, ZM, ZW.

(84) Designated States *(unless otherwise indicated, for every kind of regional protection available)*: ARIPO (BW, CV, GH, GM, KE, LR, LS, MW, MZ, NA, RW, SC, 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, ME, 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:
— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) Title: REPROGRAMMING TROPISM VIA DISPLAYED PEPTIDES TILING RECEPTOR-LIGANDS

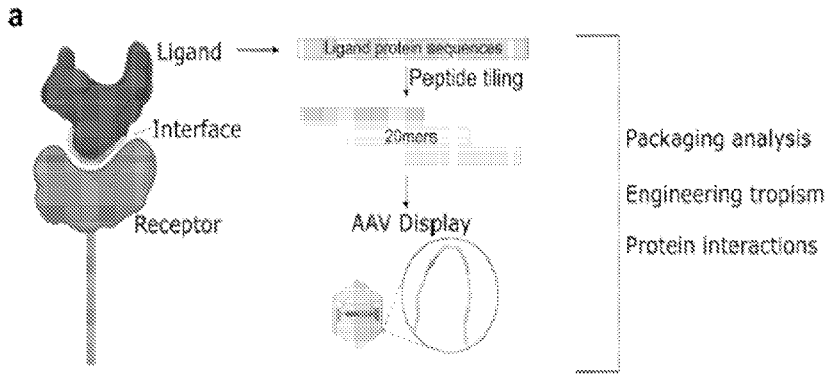

FIG. 1A

(57) Abstract: Described herein are methods for engineering proteins and viruses to improve tropism, and proteins and viruses made by using said methods.

WO 2024/055020 A2

REPROGRAMMING TROPISM VIA DISPLAYED PEPTIDES TILING RECEPTOR-LIGANDS

CROSS REFERENCE TO RELATED APPLICATIONS

[0001] This application claims priority under 35 U.S.C. §119 from Provisional Application Serial No. 63/405,360, filed September 9, 2022, the disclosures of which are incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONOSORED RESEARCH

[0002] This invention was made with Government support under OD032742, CA222826, and GM123313 awarded by the National Institutes of Health and W81XWH-22-1-0401 awarded by the Department of Defense. The government has certain rights in the invention.

TECHNICAL FIELD

[0003] Described herein are methods for engineering proteins and viruses to improve tropism, and proteins and viruses made by using said methods.

INCORPORATION BY REFERENCE OF SEQUENCE LISTING

[0004] Accompanying this filing is a Sequence Listing entitled "00015-417WO1_SL.xml", created on September 8, 2023 and having 872,924 bytes of data, machine formatted on IBM-PC, MS-Windows operating system. The sequence listing is hereby incorporated herein by reference in its entirety for all purposes.

BACKGROUND

[0005] Nucleic acid and protein based therapeutics to modulate healthy and diseased states are poised to enable the next frontier of human medicine. However their successful deployment is contingent on the ability to deliver them efficiently and in a safe and targeted manner. In this regard, a host of viral and non-viral delivery formulations have been developed, but the ability to directly modulate their tropism remains challenging.

SUMMARY

[0006] The disclosure provides a method of improving tropism of a virus or other delivery agent, the method comprising identifying ligand protein sequences derived from all known receptor-interacting ligands; systematically tile the ligand peptides into 5-50 or 10-20 or 20 amino acid peptides which are inserted into surface-exposed

loops of AAV capsids; assessing the engineered capsids for their packaging capacity, in vivo tropism, and enhanced protein interactions. In one embodiment, the virus is an adeno-associated virus (AAV). In a further embodiment, the AAV is selected from the group consisting of AAV1, AAV2, AAV5, AAV6, AAV7, AAV8, and AAV9. In still a further embodiment, the AAV is AAV5 or AAV9. In another embodiment, peptide sequences were generated via pooled oligonucleotide synthesis and inserted into 4 distinct loop regions: AAV5-Loop1 (N443), AAV5-Loop2 (S576), AAV9-Loop1 (Q456), and AAV9-Loop2 (A587) to generate over 1 million AAV variants. In another embodiment, a 20'mer ligand peptide is inserted into one or both of 2 surface-exposed loops of AAV5 (SEQ ID NO: 2) or AAV9 (SEQ ID NO: 4).

[0007] The disclosure provides a recombinant vector comprising capsid proteins containing one or two peptides inserted into one or both of two surface-exposed loops in the capsid protein, wherein the vector has a desired tropism or immune-orthogonality and wherein the peptides are independently selected from the group consisting of SEQ ID NOs:5 to 820 and 870-911. In one embodiment, the vector is an adeno-associated virus (AAV). In a further embodiment, the AAV vector is an AAV5 serotype. In still a further embodiment, the AAV5 comprises a capsid protein having a sequence as set forth in SEQ ID NO:2. In another embodiment, the AAV vector is an AAV9 serotype. In a further embodiment, the AAV9 comprises a capsid protein having a sequence as set forth in SEQ ID NO: 4. In another embodiment, the vector has tropism to pancreas, heart, brain, lung, liver, kidney, muscle, spleen or intestine. In another or further embodiment, a peptide of SEQ ID NOs:530-820, 870-910 or 911 is inserted into loop 1 and/or loop 2 of an AAV5 capsid or an AAV9 capsid. In another embodiment, the vector is immune orthogonal. In a further embodiment, the vector is an adeno-associated virus (AAV). In still a further embodiment, the AAV comprises a capsid protein of any one of SEQ ID NOs:822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, or 866, or a sequence that it at least 85% to 99% identical to any of the foregoing sequences.

[0008] The disclosure also provides a method of making a delivery vehicle or vector with a desired tropism, the method comprising selecting a peptide sequence from any one of SEQ ID NOs: 5-820 or 870-911 and (i) cloning a nucleic acid sequence encoding the peptide into a coding sequence for a capsid protein at an exposed loop site to obtain a recombinant capsid coding sequence and producing the vector using the recombinant capsid coding sequence or (ii) inserting the peptide into an exposed surface of the delivery vehicle. In one embodiment, the vector is an adeno-associated virus (AAV). In a further embodiment, the AAV is selected from the group consisting of AAV1, AAV2, AAV5, AAV6, AAV7, AAV8, and AAV9. In still a further embodiment, the AAV is AAV5 or AAV9. In a further embodiment, the AAV5 capsid coding sequence comprises SEQ ID NO:1. In another embodiment, the AAV9 capsid coding sequence comprises SEQ ID NO:2. The disclosure also provide an AAV vector comprising a capsid protein modified by any of the foregoing embodiments.

[0009] The disclosure also provides an AAV vector comprising a capsid protein wherein the capsid protein expresses a peptide of any one of SEQ ID NO: 5-820 or 870-911 in surface exposed loop 1 and/or loop 2 of the capsid protein. In one embodiment, the wild-type capsid protein sequence comprises SEQ ID NO:2 or 4. In another embodiment, the AAV vector has a AAV5 or AAV9 serotype. In still another embodiment, the vector has a desired tropism. In a further embodiment, the AAV vector has a tropism to pancreas, heart, brain, lung, liver, kidney, muscle, spleen or intestine.

[0010] The disclosure also provides a viral vector having a capsid protein comprising a heterologous targeting peptide of 10-30 amino acids in length inserted into a surface exposed portion of the capsid protein and wherein the targeting peptide is set forth in any one of SEQ ID NOs: 5-820 or 870-911. In one embodiment, the heterologous targeting peptide is about 15-25 amino acids in length. In a further embodiment, the heterologous targeting peptide is about 20 amino acids in length. In another embodiment, the viral vector is an adeno-associated virus (AAV). In still another embodiment, the viral vector is a lentiviral vector. In another embodiment, the capsid protein is a VP1 capsid protein. In still another embodiment, the capsid protein is a VP2 capsid protein. In yet another

embodiment, the capsid protein is a VP3 capsid protein. In another embodiment, the heterologous targeting peptide is inserted into an AAV capsid protein at loop 1 and/or loop 2. In yet another embodiment, the viral vector is an AAV5. In another embodiment, the viral vector is an AAV9. In still another embodiment, the heterologous targeting peptide is flanked by a linker peptide at the N-terminal and C-terminal ends of the heterologous targeting peptide. In another embodiment, the heterologous targeting peptide targets the viral vector to hepatocytes or liver tissue. In yet still another embodiment, the heterologous targeting peptide targets the viral vector to neuronal cells or brain tissue. In still another embodiment, the heterologous targeting peptide targets the viral vector to pancreatic cells or pancreas tissue. In yet another embodiment, the heterologous targeting peptide targets the viral vector to cardiac cells or heart tissue. In another embodiment, the heterologous targeting peptide targets the viral vector to lung tissue. In still yet another embodiment, the heterologous targeting peptide targets the viral vector to intestinal tissue. In yet another embodiment, the heterologous targeting peptide targets the viral vector to spleen tissue. In still another embodiment, the heterologous targeting peptide targets the viral vector to renal cells or kidney tissue. In another embodiment, the heterologous targeting peptide targets the viral vector to muscle cells or tissue.

[0011] The disclosure also provides an adeno-associated virus (AAV) capsid protein comprising a heterologous targeting peptide cloned into loop 1 and/or loop 2 of the capsid protein, wherein the heterologous targeting peptide is about 10-30 amino acids in length and is contained within or comprises any one of the peptides of SEQ ID NO:5-820 or 870-911. In one embodiment, the capsid protein is a VP1 capsid protein. In another embodiment, the capsid protein is a VP2 capsid protein. In still another embodiment, the capsid protein is a VP3 capsid protein. In another embodiment, the heterologous targeting peptide is about 15-25 amino acids in length. In yet another embodiment, the heterologous targeting peptide is about 20 amino acids in length. In a still another embodiment, the heterologous targeting peptide is flanked by a linker peptide at the

N-terminal and C-terminal ends of the heterologous targeting peptide. In another embodiment, the heterologous targeting peptide targets hepatocytes or liver tissue. In still another embodiment, the heterologous targeting peptide targets neuronal cells or brain tissue. In another embodiment, the heterologous targeting peptide targets pancreatic cells or pancreas tissue. In yet another embodiment, the heterologous targeting peptide targets cardiac cells or heart tissue. In another embodiment, the heterologous targeting peptide targets lung tissue. In yet another embodiment, the heterologous targeting peptide targets intestinal tissue. In another embodiment, the heterologous targeting peptide targets spleen tissue. In another embodiment, the heterologous targeting peptide targets renal cells or kidney tissue. In yet another embodiment, the heterologous targeting peptide targets muscle cells or tissue. The disclosure also provides a recombinant AAV (rAAV) comprising a capsid protein of any of the foregoing embodiments.

[0012] The disclosure provides a recombinant AAV (rAAV) comprising a capsid protein having a targeting peptide in loop 1 and/or loop 2 wherein the targeting peptide is independently selected from SEQ ID Nos: 5-820 or 870-911. In one embodiment, a targeting peptide is present in both Loop 1 and Loop 2. In a further embodiment, the targeting peptide has the same tropism. In another embodiment, the recombinant AAV further comprises a heterologous polynucleotide for gene delivery. In a further embodiment, the heterologous polynucleotide is a therapeutic gene. In still another or further embodiment, of any of the foregoing embodiments, the rAAV is present in a pharmaceutical composition.

[0013] The disclosure also provides a method for delivering a transgene to a subject comprising: administering a recombinant AAV (rAAV) to a subject, wherein the rAAV comprises: (i) a capsid protein of the disclosure, and (ii) at least one transgene, and wherein the rAAV infects cells of a target tissue of the subject. In one embodiment, the at least one transgene encodes a protein. In a further embodiment, the protein is an immunoglobulin heavy chain or light chain or fragment thereof. In another embodiment, the at least one transgene encodes a small interfering nucleic acid. In a further embodiment, the small interfering nucleic acid is a miRNA.

In another embodiment, the small interfering nucleic acid is a miRNA sponge or TuD RNA that inhibits the activity of at least one miRNA in the subject or animal. In yet another embodiment, the miRNA is expressed in a cell of the target tissue. In still another embodiment, the target tissue is skeletal muscle, heart, liver, pancreas, brain or lung. In another embodiment, the transgene expresses a transcript that comprises at least one binding site for a miRNA, wherein the miRNA inhibits activity of the transgene, in a tissue other than the target tissue, by hybridizing to the binding site. In yet another embodiment, the at least one transgene encodes a gene product that mediates genome editing. In another embodiment, the transgene comprises a tissue specific promoter or inducible promoter. In a further embodiment, the tissue specific promoter is a liver-specific thyroxin binding globulin (TBG) promoter, an insulin promoter, a glucagon promoter, a somatostatin promoter, a pancreatic polypeptide (PPY) promoter, a synapsin-1 (Syn) promoter, a creatine kinase (MCK) promoter, a mammalian desmin (DES) promoter, a a-myosin heavy chain (a-MHC) promoter, or a cardiac Troponin T (cTnT) promoter. In another embodiment, the rAAV is administered intravenously, intravascularly, transdermally, intraocularly, intrathecally, orally, intramuscularly, subcutaneously, intranasally, or by inhalation. In still another embodiment, the subject is selected from a mouse, a rat, a rabbit, a dog, a cat, a sheep, a pig, and a non-human primate. In yet another embodiment, the subject is a human.

[0014] The disclosure also provides an isolated nucleic acid encoding an AAV capsid protein containing an amino acid sequence selected from the group consisting of SEQ ID No: 5-820 and 870-911.

[0015] The disclosure also provides a delivery vehicle for delivery of a small molecule drug or biological agent having a desired tropism, wherein the delivery vehicle comprises a peptide or peptide fragment of at least 10-20 amino acids of any one of SEQ ID NOs:5-820 or 870-911. In one embodiment, the delivery vehicle is selected from the group consisting of a liposome, a nanoparticle, a bacteria, a bacteriophage, a virus-like particle (VLP), a erythrocyte ghost, and an exosome. In another embodiment, the biological agent comprises an siRNA, an antisense molecule, a

protein or polypeptide, insulin, a vaccine, or an antibody. In still another embodiment, the small molecule drug comprises a chemotherapeutic agent, an anti-inflammatory, a steroid, and an antibiotic.

[0016] The disclosure also provides a biological agent having a desired tropism, the biological agent linked to a peptide or peptide fragment of at least 10-20 amino acids of any one of SEQ ID NOs:5-820 or 870-911. In one embodiment, the biological agent is a nucleic acid, a protein, a polypeptide, a peptide, an antibody, an antibody fragment, a non-immunoglobulin binding agent, or an enzyme.

DESCRIPTION OF DRAWINGS

[0017] **Figure 1A-B Design of AAV libraries displaying peptides tiling receptor-ligands. (a)** Schematic of approach for rationally engineering and characterizing AAV variants. Ligand protein sequences derived from all known receptor-interacting ligands are systematically tiled into 20 amino acid peptides which are inserted into surface-exposed loops of AAV capsids. These engineered variants were then assessed for their packaging capacity, *in vivo* tropism, and enhanced protein interactions. **(b)** Protein class distribution of the ligands which were tiled to compose the screening library. These include known receptor-interacting ligands (orange), cell membrane permeable proteins (green) and protein domains (blue), and stop codon containing negative controls (purple). Peptide sequences were generated via pooled oligonucleotide synthesis and inserted into 4 distinct loop regions: AAV5-Loop1 (N443), AAV5-Loop2 (S576), AAV9-Loop1 (Q456), and AAV9-Loop2 (A587) to generate over 1 million AAV variants. Capsid surface residues are colored according to their distance from the core of the capsid with the insertion site flanking residues shown in red.

[0018] **Figure 2A-F. AAV library packaging analyses reveal biophysical features contributing to capsid fitness. (a)** Schematic illustrating recombinant production of pooled AAV libraries in HEK293T cells. **(b)** Normalized abundance for each inserted peptide in the plasmid libraries versus DNA isolated from recombinantly produced AAV variant capsid libraries. Dotted-red line shows where the plasmid abundance is equal to the capsid abundance. **(c)** Normalized abundance for negative control peptides containing stop

codons, in both the plasmid libraries and recombinantly produced AAV capsid libraries **(d)** Distribution of peptide biophysical parameters relevant to packaging. Peptide charge, alpha-helical content, flexibility, and hydrophobicity distributions are shown for peptides which are enriched in the capsid pool ("Packaging") and those which are depleted ("Non Packaging"). Statistical significance between groups was calculated via a T-test (****$p<0.0001$). **(e)** Biophysical parameters of inserted peptides were used as features to train a support vector machine classifier predicting which AAV variants successfully package into capsids. The receiver operating characteristic curve is shown for the resulting model, with an area under the curve of 0.89. **(f)** UMAP embedding for each AAV variant, colored by packaging status. Inserted peptide charge, alpha-helical content, flexibility, and hydrophobicity were used as input features for the embedding.

[0019] Figure 3A-D. *In vivo* screen analyses enable predictive computational models of tropism. (a) Overview of *in vivo* screening methodology. The four AAV variant libraries were injected retro-orbitally into C57/BL6 mice in duplicate. Two weeks post-injection, nine organs were harvested from each mouse and the inserted peptide-containing region of the AAV capsid was amplified and subjected to next generation sequencing. **(b)** Replicate Pearson correlation for the $\log_2$ fold change ($\log_2$FC) values (organ vs. capsid) for each library and organ. **(c)** Overview of screen results. Significantly enriched AAV variants in a particular organ are defined as those with a $\log_2$FC > 1 and an FDR-adjusted p-value < 0.05. Bar plots show the number of significantly enriched variants detected per organ, as well as a comparison of peptide hits for each loop insertion site, for both AAV5 and AAV9. **(d)** Overview of classification model predicting AAV tissue tropism from peptide sequence alone. Inserted peptide sequences were converted to a binary one-hot encoding (for each peptide, 20 rows corresponding to position, and 20 columns corresponding to presence of a particular amino acid). This one-hot encoding scheme was then used as input to a convolutional neural network (CNN) to predict organ targeting. Model performance was separately evaluated on each organ, via accuracy, area under the receiver operator characteristic curve (AUROC), F1 score, and

Matthews Correlation Coefficient (MCC). Models were trained on ⅔ of the data, and the remaining ⅓ was held out as a validation dataset to evaluate performance. Figure discloses SEQ ID NOs:867-869, respectively, in order of appearance.

[0020] Figure 4A-D. *In vivo* screen identified AAV variants demonstrate reprogrammed tropism. (a) Heatmap showing log_2FC values for each AAV variant which was significantly enriched in at least one organ. Rows are individual variants, and columns are organs (n=2 per organ). **(b)** UMAP embedding of significantly enriched AAV variants. Each dot represents a variant, colored by the organ with max log_2FC. **(c)** AAVs were chosen for validation from the pool of significant hits on the basis of their tissue specificity, broad tropism, and/or internal consistency. Internal consistency was quantified by counting the number of similar (>50% homology) inserted peptides also detected as hits for a given organ. AAV variants were characterized structurally via transmission electron microscopy, and functionally via delivery of the mCherry transgene *in vivo*. Heatmaps depict all variants chosen for validation (n=21), with the left heatmap showing the Z-normalized log2FC values from the pooled *in vivo* screen (n=2), and the right heatmap showing the Z-normalized mCherry expression quantified by RT-qPCR, relative to AAV9 (n=2). **(d)** Comparison between liver RT-qPCR quantification of mCherry delivery, versus protein level quantification via fluorescent microscopy.

[0021] Figure 5A-D. Characterization and mechanistic exploration of AAV variants with displayed ligand peptides. (a) Full characterization experiments for the variant AAV9.DKK1. In the top left, lung screen count values for all AAV9Loop2 variants with inserted DKK1 derived peptides are shown. The x-axis indicates the position in the DKK1 structure a given peptide starts on. Shown in blue are the lung counts, and shown in orange are the capsid counts. Red arrow indicates the location of the peptide inserted in AAV9.DKK1. Bar plot shows the RT-qPCR quantification (n=2) of mCherry transgene expression via individual validation, normalized to that of AAV9. Also shown are electron microscope images for AAV variant capsids, as well as fluorescent microscopy of mCherry protein expression levels in the lung and liver. **(b)** Knockout

experiments to validate AAV9.DKK1 receptor dependency. Cas9-containing lentivirus was produced with either non-targeting controls (NTC) and LRP6 targeting sgRNA (n=2 each). HEK293T cells were then transduced with lentivirus and positively selected with puromycin. Lentivirally transduced cells were then plated in a 24-well plate and transduced with either wild-type AAV9 ($4x10^9$ viral genomes) or AAV9.DKK1 virus ($1x10^9$ viral genomes). Flow cytometry was then used to quantify the mean fluorescence intensity (MFI) of the infected cells. Shown to the right is the crystal structure of a 7-mer peptide DKK1 peptide (contained within AAV9.DKK1) in complex with LRP6 (*58*). **(c)** Full characterization experiments for the variant AAV9.PDGFC. In the top left, muscle screen count values for all AAV9Loop1 variants with inserted PDGFC derived peptides are shown. The x-axis indicates the position in the PDGFC structure a given peptide starts on. Shown in blue are the muscle counts, and shown in orange are the capsid counts. Red line indicates the location of the peptide inserted in AAV9.PDGFC. Bar plot shows the RT-qPCR quantification (n=2) of mCherry transgene expression via individual validation, normalized to that of AAV9. Also shown are electron micrographs for variant capsids, as well as fluorescent microscopy of mCherry protein expression levels in the heart and muscle. **(d)** Receptor overexpression studies for the variant AAV9.PDGFC. Known receptors for the PDGFC ligand were cloned into an overexpression plasmid and transfected into HEK293T cells in a 24-well plate. After 24 hours, either AAV9 ($4x10^9$ viral genomes) or AAV9.PDGFC ($4x10^9$ viral genomes) were used to transduce the cells. Following 24 hours of transduction, the cells were collected and mCherry expression was quantified via flow cytometry. Bar plots show the MFI normalized to the average MFI of the AAV9 transduced cells with an empty vector overexpressed. Statistical significance between groups was calculated via a T-test (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

[0022] Figure 6A-D. Inserted peptide facilitates efficient liver de-targeting across multiple AAV scaffolds in a mouse strain-independent manner. (a) Significantly enriched AAV variants from all capsids are projected into 2 dimensions via UMAP. The distance between points is calculated explicitly via the Levenshtein distance

of the amino acid sequences of the inserted peptides, and subsequently embedded via UMAP. AAV variants are colored by their log_2 fold change in the brain. Select variant clusters which are highly enriched in the brain are highlighted. AAV variants are also colored by capsid insertion site in the embedding on the right. **(b)** Brain and capsid counts for all AAV5Loop2 and AAV9Loop1 variants with inserted APOA1 derived peptides. Shown in blue are the brain counts, and shown in orange are the capsid counts. Red arrow indicates the location of the peptide inserted in AAV5.APOA1 and AAV9.APOA1 respectively. Also shown are electron micrographs for AAV5.APOA1 and AAV9.APOA1. **(c)** RT-qPCR values (transduction relative to AAV9) for individual *in vivo* validations of AAV5.APOA1 and AAV9.APOA1 in C57/BL6 mice, confirming liver detargeting. **(d)** Performance of AAV5.APOA1 in BALB/c mice. Bar graph shows RT-qPCR values (transduction relative to AAV9 in BALB/c), for AAV5.APOA1 across 8 organs, confirming liver detargeting is strain independent.

[0023] Figure 7A-E. Novel immune-orthogonal AAV serotype mining, characterization and tropism reprogramming via displayed ligand peptides. (a) Schematic of the computational pipeline used to identify novel AAV serotypes for testing. Basic local alignment search tool (BLAST) was used to identify 687 initial capsids with sequence homology to the AAV2 *cap* gene. This initial list was then filtered to exclude truncated genomes, redundant samples, human and non-mammalian serotypes, as well as close orthologs. The final resulting list contained 23 AAV capsid sequences for subsequent investigation. **(b)** Hierarchical clustering dendrogram of AAV capsid sequences. Shown in red are previously identified AAV serotypes currently in use. Novel AAVs identified and used for downstream testing are shown in black. **(c)** Two stage filtering of novel AAVs. AAVs were first assessed by measuring their ability to package, and then by their ability to transduce the liver (using an mCherry transgene) *in vivo*. All values shown relative to the orthologous wild-type AAV5. **(d)** The four novel AAVs which could infect the liver (AAV MM2, AAV MG2, AAV MG1 and AAV CH1), were tested for immune cross-reactivity with AAV8. Mice were immunized with the indicated AAV, and then 3 weeks post-injection tested for antibody cross reactivity via an ELISA. **(e)** The PDGFC peptide from AAV9.PDGFC was

inserted onto loop1 of AAV MG2 to yield AAV MG2.PDGFC. AAV MG2 and AAV MG2.PDGFC were injected into C57BL/6 mice, quantifying muscle transduction via RT-qPCR after three weeks. Bar plots show muscle transduction relative to wild-type AAV MG2. Statistical significance between groups was calculated via a T-test (*p<0.05, **p<0.01, ***p<0.001, ****p<0.0001).

DETAILED DESCRIPTION

[0024] As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural referents unless the context clearly dictates otherwise. Thus, for example, reference to "a cell" includes a plurality of such cells and reference to "the fragment" includes reference to one or more fragments and equivalents thereof known to those skilled in the art, and so forth.

[0025] Also, the use of "or" means "and/or" unless stated otherwise. Similarly, "comprise," "comprises," "comprising" "include," "includes," and "including" are interchangeable and not intended to be limiting.

[0026] It is to be further understood that where descriptions of various embodiments use the term "comprising," those skilled in the art would understand that in some specific instances, an embodiment can be alternatively described using language "consisting essentially of" or "consisting of."

[0027] Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood to one of ordinary skill in the art to which this disclosure belongs. Although many methods and reagents are similar or equivalent to those described herein, the exemplary methods and materials are disclosed herein.

[0028] All publications mentioned herein are incorporated herein by reference in full for the purpose of describing and disclosing the methodologies, which might be used in connection with the description herein. Moreover, with respect to any term that is presented in one or more publications that is similar to, or identical with, a term that has been expressly defined in this disclosure, the definition of the term as expressly provided in this disclosure will control in all respects.

[0029] It should be understood that this disclosure is not limited to the particular methodology, protocols, and reagents, *etc.*, described herein and as such may vary. The terminology used herein is for the purpose of describing particular embodiments or aspects only and is not intended to limit the scope of the present disclosure.

[0030] Other than in the operating examples, or where otherwise indicated, all numbers expressing quantities of ingredients or reaction conditions used herein should be understood as modified in all instances by the term "about". The term "about" when used to describe the present invention, in connection with percentages means ±1%. The term "about", as used herein can mean within an acceptable error range for the particular value as determined by one of ordinary skill in the art, which can depend in part on how the value is measured or determined, *e.g.*, the limitations of the measurement system. Alternatively, "about" can mean a range of plus or minus 20%, plus or minus 10%, plus or minus 5%, or plus or minus 1% of a given value. Alternatively, particularly with respect to biological systems or processes, the term can mean within an order of magnitude, within 5-fold, or within 2-fold, of a value. Where particular values are described in the application and claims, unless otherwise stated the term "about" meaning within an acceptable error range for the particular value can be assumed. Also, where ranges and/or subranges of values are provided, the ranges and/or subranges can include the endpoints of the ranges and/or subranges. In some cases, variations can include an amount or concentration of 20%, 10%, 5%, 1 %, 0.5%, or even 0.1 % of the specified amount.

[0031] For the recitation of numeric ranges herein, each intervening number there between with the same degree of precision is explicitly contemplated. For example, for the range of 6-9, the numbers 7 and 8 are contemplated in addition to 6 and 9, and for the range 6.0-7.0, the number 6.0, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, and 7.0 are explicitly contemplated.

[0032] The term "adeno-associated virus" or "AAV" as used herein refers to a member of the class of viruses associated with this name and belonging to the genus dependoparvovirus, family Parvoviridae.

Multiple serotypes of this virus are known to be suitable for gene delivery; all known serotypes can infect cells from various tissue types. Non-limiting exemplary serotypes useful in the methods disclosed herein include any of the 11 or 12 serotypes, *e.g.*, AAV2, AAV5, and AAV8, or variant serotypes such as AAV-DJ. The AAV structural particle is composed of 60 protein molecules made up of VP1, VP2 and VP3. Each particle contains approximately 5 VP1 proteins, 5 VP2 proteins and 50 VP3 proteins ordered into an icosahedral structure. Non-limiting exemplary VP1 sequences useful in the methods disclosed herein are provided below.

[0033] The term "amino acid" refers to naturally occurring and synthetic amino acids, as well as amino acid analogs and amino acid mimetics that function in a manner similar to the naturally occurring amino acids. Naturally occurring amino acids are those encoded by the genetic code, as well as those amino acids that are later modified, *e.g.*, hydroxyproline, γ-carboxyglutamate, and O-phosphoserine. In some embodiments, an amino acid analog refers to compounds that have the same basic chemical structure as a naturally occurring amino acid, *i.e.*, a carbon that is bound to a hydrogen, a carboxyl group, an amino group, and an R group, *e.g.*, homoserine, norleucine, methionine sulfoxide, methionine methyl sulfonium. Such analogs have modified R groups (*e.g.*, norleucine) or modified peptide backbones, but retain the same basic chemical structure as a naturally occurring amino acid. In some embodiments, an amino acid mimetic refers to chemical compounds that have a structure that is different from the general chemical structure of an amino acid, but that functions in a manner similar to a naturally occurring amino acid. The terms "non-naturally occurring amino acid" and "unnatural amino acid" refer to amino acid analogs, synthetic amino acids, and amino acid mimetics which are not found in nature. In certain instances, one or more D-amino acids can be used in various peptide compositions of the disclosure. The disclosure provides various peptides that are useful for treating various diseases and infections. These peptides can comprise naturally occurring amino acid. In other embodiments, the peptides can comprise non-natural amino acids. The use of non-natural amino acids can improve the peptides stability, decrease degradation and/or improve biological

activity. For example, in some embodiments, one or more D-amino acids. In other embodiments, retroinverso peptides are contemplated using various amino acid configurations.

[0034] Amino acids may be referred to herein by either their commonly known three letter symbols or by the one-letter symbols recommended by the IUPAC-IUB Biochemical Nomenclature Commission. Nucleotides, likewise, may be referred to by their commonly accepted single-letter codes.

[0035] The term "Cas9" refers to a CRISPR-associated, RNA-guided endonuclease such as *Streptococcus pyogenes* Cas9 (spCas9; see Accession Number Q99ZW2.1, the sequence of which is incorporated herein by reference) and orthologs and biological equivalents thereof. Biological equivalents of Cas9 include, but are not limited to, C2c1 from *Alicyclobacillus acideterrestris* and Cpf1 (which performs cutting/cleaving functions analogous to Cas9) from various bacterial species including *Acidaminococcus spp.* and *Francisella novicida* U112. Cas9 may refer to an endonuclease that causes double stranded breaks in DNA, a nickase variant such as a RuvC or HNH mutant that causes a single stranded break in DNA, as well as other variations such as deadCas-9 ("dCas9"), which lack endonuclease activity. Cas9 may also refer to "split-Cas9" in which Cas9 is split into two halves - C-terminal Cas9 (C-Cas9) and an N-terminal Cas-9 (N-Cas9) - which can be fused with two intein moieties. *See, e.g.*, U.S. Pat. No. 9,074,199 B1; Zetsche *et al.* (2015) Nat Biotechnol. 33(2):139-42; Wright *et al.* (2015) PNAS 112(10) 2984-89. Non-limiting examples of commercially available sources of SpCas9 comprising plasmids can be found under the following AddGene reference numbers:

42230: PX330; SpCas9 and single guide RNA;

48138: PX458; SpCas9-2A-EGFP and single guide RNA;

62988: PX459; SpCas9-2A-Puro and single guide RNA;

48873: PX460; SpCas9n (D10A nickase) and single guide RNA;

48140: PX461; SpCas9n-2A-EGFP (D10A nickase) and single guide RNA;

62987: PX462; SpCas9n-2A-Puro (D10A nickase) and single guide RNA; and

48137: PX165; SpCas9;

all of which are incorporated herein by reference.

[0036] As used herein, the term "CRISPR" refers to Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR). CRISPR may also refer to a technique or system of sequence-specific genetic manipulation relying on the CRISPR pathway. A CRISPR recombinant expression system can be programmed to cleave a target polynucleotide using a CRISPR endonuclease and a guideRNA. A CRISPR system can be used to cause double stranded or single stranded breaks in a target polynucleotide. A CRISPR system can also be used to recruit proteins or label a target polynucleotide. In some aspects, CRISPR-mediated gene editing utilizes the pathways of nonhomologous end-joining (NHEJ) or homologous recombination to perform the edits. These applications of CRISPR technology are known and widely practiced in the art. See, *e.g.*, U.S. Pat. No. 8,697,359 and Hsu *et al.* (2014) Cell 156(6): 1262-1278.

[0037] As used herein, the term "delivery vehicle" refers to a composition useful for delivering a payload to a cell, tissue or subject. The delivery vehicle can deliver various payloads including biological agents and small molecule drugs. Exemplary, but non-limiting, delivery vehicles include a liposome, a nanoparticle, a bacteria, a bacteriophage, a virus-like particle (VLP), a erythrocyte ghost, and an exosome.

[0038] As used herein, the term "domain" can refer to a particular region of a larger molecule (*e.g.*, a particular region of a protein or polypeptide), which can be associated with a particular function. For example, "a domain which binds to a cognate" can refer to the domain of a protein that binds one or more receptors or other protein moieties. Similarly, a corresponding coding sequence for a particular polypeptide domain can be referred to as a polynucleotide domain.

[0039] The term "encode" as it is applied to polynucleotides can refer to a polynucleotide which is said to "encode" a polypeptide if, in its native state or when manipulated by methods well known to those skilled in the art, it can be transcribed and/or translated to produce the mRNA for the polypeptide and/or a fragment thereof. In some cases, the antisense strand is the complement of such a nucleic acid, and the encoding sequence can be deduced therefrom.

[0040] The terms "equivalent" or "biological equivalent" are used interchangeably when referring to a particular molecule, biological, or cellular material and intend those having minimal homology while still maintaining desired structure or functionality.

[0041] As used herein, "expression" can refer to the process by which polynucleotides are transcribed into mRNA and/or the process by which the transcribed mRNA is subsequently being translated into peptides, polypeptides, or proteins. If the polynucleotide is derived from genomic DNA, expression can include splicing of the mRNA in a eukaryotic cell.

[0042] As used herein, the term "functional" may be used to modify any molecule, biological, or cellular material to intend that it accomplishes a particular, specified effect.

[0043] The term "gRNA" or "guide RNA" as used herein refers to the guide RNA sequences used to target specific genes for correction employing the CRISPR technique. Techniques of designing gRNAs and donor therapeutic polynucleotides for target specificity are well known in the art. For example, Doench, J., *et al.* Nature biotechnology 2014; 32(12): 1262-7, Mohr, S. *et al.* (2016) FEBS Journal 283: 3232-38, and Graham, D., *et al.* Genome Biol. 2015; 16: 260. gRNA comprises or alternatively consists essentially of, or yet further consists of a fusion polynucleotide comprising CRISPR RNA (crRNA) and trans-activating CRIPSPR RNA (tracrRNA); or a polynucleotide comprising CRISPR RNA (crRNA) and trans-activating CRIPSPR RNA (tracrRNA). In some aspects, a gRNA is synthetic (Kelley, M. *et al.* J of Biotechnology 233 (2016) 74-83).

[0044] "Homology" or "identity" or "similarity" can refer to sequence similarity between two peptides or between two nucleic acid molecules. Homology can be determined by comparing a position in each sequence which can be aligned for purposes of comparison. For example, when a position in the compared sequence is occupied by the same base or amino acid, then the molecules are homologous at that position. A degree of homology between sequences is a function of the number of matching or homologous positions shared by the sequences. An "unrelated" or "non-homologous" sequence shares less than 40% identity, or alternatively less than 25% identity, with one of the sequences of the disclosure.

[0045] Homology refers to a percent (%) identity of a sequence to a reference sequence. As a practical matter, any particular sequence can be at least 50%, 60%, 70%, 80%, 85%, 90%, 92%, 95%, 96%, 97%, 98% or 99% identical to any sequence described herein. Whether such particular peptide, polypeptide or nucleic acid sequence has a particular identity/homology can be determined conventionally using known computer programs such the Bestfit program (Wisconsin Sequence Analysis Package, Version 8 for Unix, Genetics Computer Group, University Research Park, 575 Science Drive, Madison, Wis. 53711). When using Bestfit or any other sequence alignment program to determine whether a particular sequence is, for instance, 95% identical to a reference sequence, the parameters can be set such that the percentage of identity is calculated over the full length of the reference sequence and that gaps in homology of up to 5% of the total reference sequence are allowed. A number of sequences are provided herein, it is contemplated that sequences having at least 50%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 98%, 99% and 100% to any one of the sequences herein find use in any of the compositions and methods described herein.

[0046] For example, in a specific embodiment the identity between a reference sequence (query sequence, *i.e.*, a sequence of the disclosure) and a subject sequence, also referred to as a global sequence alignment, can be determined using the FASTDB computer program based on the algorithm of Brutlag *et al.* (Comp. App. Biosci. 6:237-245 (1990)). In some cases, parameters for a particular embodiment in which identity is narrowly construed, used in a FASTDB amino acid alignment, can include: Scoring Scheme=PAM (Percent Accepted Mutations) 0, k-tuple=2, Mismatch Penalty=1, Joining Penalty=20, Randomization Group Length=0, Cutoff Score=1, Window Size=sequence length, Gap Penalty=5, Gap Size Penalty=0.05, Window Size=500 or the length of the subject sequence, whichever is shorter. According to this embodiment, if the subject sequence is shorter than the query sequence due to N- or C-terminal deletions, not because of internal deletions, a manual correction can be made to the results to take into consideration the fact that the FASTDB program does not account for N- and C-terminal truncations of the

subject sequence when calculating global percent identity. For subject sequences truncated at the N- and C-termini, relative to the query sequence, the percent identity can be corrected by calculating the number of residues of the query sequence that are lateral to the N- and C-terminal of the subject sequence, which are not matched/aligned with a corresponding subject residue, as a percent of the total bases of the query sequence. A determination of whether a residue is matched/aligned can be determined by results of the FASTDB sequence alignment. This percentage can be then subtracted from the percent identity, calculated by the FASTDB program using the specified parameters, to arrive at a final percent identity score. This final percent identity score can be used for the purposes of this embodiment. In some cases, only residues to the N- and C-termini of the subject sequence, which are not matched/aligned with the query sequence, are considered for the purposes of manually adjusting the percent identity score. That is, only query residue positions outside the farthest N- and C-terminal residues of the subject sequence are considered for this manual correction. For example, a 90 residue subject sequence can be aligned with a 100 residue query sequence to determine percent identity. The deletion occurs at the N-terminus of the subject sequence and therefore, the FASTDB alignment does not show a matching/alignment of the first 10 residues at the N-terminus. The 10 unpaired residues represent 10% of the sequence (number of residues at the N- and C-termini not matched/total number of residues in the query sequence) so 10% is subtracted from the percent identity score calculated by the FASTDB program. If the remaining 90 residues were perfectly matched the final percent identity can be 90%. In another example, a 90 residue subject sequence is compared with a 100 residue query sequence. This time the deletions are internal deletions so there are no residues at the N- or C-termini of the subject sequence which are not matched/aligned with the query. In this case the percent identity calculated by FASTDB is not manually corrected. Once again, only residue positions outside the N- and C-terminal ends of the subject sequence, as displayed in the FASTDB alignment, which are not matched/aligned with the query sequence are manually corrected for.

[0047] "Hybridization" can refer to a reaction in which one or more polynucleotides react to form a complex that is stabilized via hydrogen bonding between the bases of the nucleotide residues. The hydrogen bonding can occur by Watson-Crick base pairing, Hoogstein binding, or in any other sequence-specific manner. The complex can comprise two strands forming a duplex structure, three or more strands forming a multi-stranded complex, a single self-hybridizing strand, or any combination of these. A hybridization reaction can constitute a step in a more extensive process, such as the initiation of a PC reaction, or the enzymatic cleavage of a polynucleotide by a ribozyme.

[0048] Examples of stringent hybridization conditions include: incubation temperatures of about 25°C to about 37°C; hybridization buffer concentrations of about 6x SSC to about 10x SSC; formamide concentrations of about 0% to about 25%; and wash solutions from about 4x SSC to about 8x SSC. Examples of moderate hybridization conditions include: incubation temperatures of about 40°C to about 50°C; buffer concentrations of about 9x SSC to about 2x SSC; formamide concentrations of about 30% to about 50%; and wash solutions of about 5x SSC to about 2x SSC. Examples of high stringency conditions include: incubation temperatures of about 55°C to about 68°C; buffer concentrations of about 1x SSC to about 0.1x SSC; formamide concentrations of about 55% to about 75%; and wash solutions of about 1x SSC, 0.1x SSC, or deionized water. In general, hybridization incubation times are from 5 minutes to 24 hours, with 1, 2, or more washing steps, and wash incubation times are about 1, 2, or 15 minutes. SSC is 0.15 M NaCl and 15 mM citrate buffer. It is understood that equivalents of SSC using other buffer systems can be employed.

[0049] As used herein, the term "immune orthogonal" refers to a lack of immune cross-reactivity between two or more antigens. In some embodiments, the antigens are proteins (*e.g.*, Cas9). In some embodiments, the antigens are viral antigens associated with a particular viral vector (*e.g.*, AAV). As is recognized in the art, antigens typically include antigenic determinants having a particular sequence of 3 dimensional structure. Moreover, an antigenic determinant can comprise a domain or subsequence of a

larger polypeptide or molecular sequence. In some embodiments, antigens that are immune orthogonal do not share an amino acid sequence of greater than 5, greater than 6, greater than 7, greater than 8, greater than 9, greater than 10, greater than 11, greater than 12, greater than 13, greater than 14, greater than 15, or greater than 16 consecutive amino acids. In some embodiments, antigens that are immune orthogonal do not share any highly immunogenic peptides. In some embodiments, antigens that are immune orthogonal do not share affinity for a major histocompatibility complex (*e.g.*, MHC class I or class II). Antigens that are immune orthogonal are amenable for sequential dosing to evade a host immune system.

[0050] The term "immunosilent" refers to an epitope or foreign peptide, polypeptide or protein that does not elicit an immune response from a host upon administration. In some embodiments, the peptide, polypeptide or protein does not elicit an adaptive immune response. In some embodiments, the peptide, polypeptide or protein does not elicit an innate immune response. In some embodiments, the peptide, polypeptide or protein does not elicit either an adaptive or an innate immune response. In some embodiments, an immunosilent peptide, polypeptide or protein has reduced immunogenicity.

[0051] The term "isolated" as used herein can refer to molecules or biologicals or cellular materials being substantially free from other materials. In one aspect, the term "isolated" can refer to nucleic acid, such as DNA or RNA, or protein or polypeptide (*e.g.*, an antibody or derivative thereof), or cell or cellular organelle, or tissue or organ, separated from other DNAs or RNAs, or proteins or polypeptides, or cells or cellular organelles, or tissues or organs, respectively, that are present in the natural source. The term "isolated" also can refer to a nucleic acid or peptide that is substantially free of cellular material, viral material, or culture medium when produced by recombinant DNA techniques, or chemical precursors or other chemicals when chemically synthesized. Moreover, an "isolated nucleic acid" is meant to include nucleic acid fragments which are not naturally occurring as fragments and may not be found in the natural state. In some cases, the term "isolated" is also used herein to refer to polypeptides which are

isolated from other cellular proteins and is meant to encompass both purified and recombinant polypeptides. In some cases, the term "isolated" is also used herein to refer to cells or tissues that are isolated from other cells or tissues and is meant to encompass both cultured and engineered cells or tissues.

[0052] "Messenger RNA" or "mRNA" is a nucleic acid molecule that is transcribed from DNA and then processed to remove non-coding sections known as introns. In some cases, the resulting mRNA is exported from the nucleus (or another locus where the DNA is present) and translated into a protein. The term "pre-mRNA" can refer to the strand prior to processing to remove non-coding sections. mRNA has "U" in place of "T" in cDNA coding sequences.

[0053] The term "ortholog" is used in reference of another gene or protein and intends a homolog of said gene or protein that evolved from the same ancestral source or which are evolved artificially using molecular biology and genetic engineering. Orthologs may or may not retain the same function as the gene or protein to which they are orthologous. Non-limiting examples of Cas9 orthologs include *S. aureus* Cas9 ("spCas9"), *S. thermophiles* Cas9, *L. pneumophilia* Cas9, *N. lactamica* Cas9, *N. meningitides* Cas9, *B. longum* Cas9, *A. muciniphila* Cas9, and *O. laneus* Cas9.

[0054] The term "payload" refers to a therapeutic and diagnostic agents that can be loaded into or onto a delivery vehicle. Such payload include biological and small molecule entities. Exemplary payload agents include small molecule drugs, biological molecules, viruses, therapeutic agents, prodrugs, gene silencing agents, chemotherapeutics, diagnostic agents, and/or components of gene editing systems. Examples of biological molecules include, but are not limited to, nucleic acids (e.g., DNA, RNA, mRNA, modified mRNA, small RNAs, siRNA, miRNA, genes, and transgenes), peptides/proteins (including antibodies, enzymes, transcription factors, etc.), viruses, hormones, carbohydrates, lipids, and vitamins. Examples of gene silencing agents, include siRNA, chRNAs, miRs, ribozymes, morpholinos, and esiRNAs. Examples of gene editing systems include, but are not limited to, CRISPR-Cas systems, zinc finger nucleases, and TALENs. Examples of diagnostic agents, include but are not limited to, dyes and stains, radioactive tracers, and contrast

agents. Examples, of anticancer agents and chemotherapeutics that can be used with or loaded into a delivery vehicle include, but are not limited to, alkylating agents such as thiotepa and CYTOXAN® cyclosphosphamide; alkyl sulfonates such as busulfan, improsulfan and piposulfan; aziridines such as benzodopa, carboquone, meturedopa, and uredopa; ethylenimines and methylamelamines including altretamine, triethylenemelamine, trietylenephosphoramide, triethiylenethiophosphoramide and tiimethylolomelamine; acetogenins (e.g., bullatacin and bullatacinone); a camptothecin (including the synthetic analogue topotecan); bryostatin; callystatin; CC-1065 (including its adozelesin, carzelesin and bizelesin synthetic analogues); cryptophycins (particularly cryptophycin 1 and cryptophycin 8); dolastatin; duocarmycin (including the synthetic analogues, KW-2189 and CB1-TM1); eleutherobin; pancratistatin; a sarcodictyin; spongistatin; nitrogen mustards such as chlorambucil, chlornaphazine, cholophosphamide, estramustine, ifosfamide, mechlorethamine, mechlorethamine oxide hydrochloride, melphalan, novembichin, phenesterine, prednimustine, trofosfamide, uracil mustard; nitrosureas such as carmustine, chlorozotocin, fotemustine, lomustine, nimustine, and ranimnustine; vinca alkaloids; epipodophyllotoxins; antibiotics such as the enediyne antibiotics (e.g., calicheamicin, especially calicheamicin gammall and calicheamicin omegall; L-asparaginase; anthracenedione substituted urea; methyl hydrazine derivatives; dynemicin, including dynemicin A; bisphosphonates, such as clodronate; an esperamicin; as well as neocarzinostatin chromophore and related chromoprotein enediyne antiobiotic chromophores), aclacinomysins, actinomycin, authramycin, azaserine, bleomycins, cactinomycin, carabicin, carminomycin, carzinophilin, chromomycinis, dactinomycin, daunorubicin, detorubicin, 6-diazo-5-oxo-L-norleucine, ADRIAMYCIN® doxorubicin (including morpholino-doxorubicin, cyanomorpholino-doxorubicin, 2-pyrrolino-doxorubicin and deoxydoxorubicin), epirubicin, esorubicin, idarubicin, marcellomycin, mitomycins such as mitomycin C, mycophenolic acid, nogalamycin, olivomycins, peplomycin, potfiromycin, puromycin, quelamycin, rodorubicin, streptonigrin, streptozocin, tubercidin, ubenimex, zinostatin, zorubicin; anti-metabolites such as methotrexate and 5-fluorouracil (5-FU); folic

acid analogs such as denopterin, methotrexate, pteropterin, trimetrexate; purine analogs such as fludarabine, 6-mercaptopurine, thiamiprine, thioguanine; pyrimidine analogs such as ancitabine, azacitidine, 6-azauridine, carmofur, cytarabine, dideoxyuridine, doxifluridine, enocitabine, floxuridine; androgens such as calusterone, dromostanolone propionate, epitiostanol, mepitiostane, testolactone; anti-adrenals such as aminoglutethimide, mitotane, trilostane; folic acid replenisher such as frolinic acid; aceglatone; aldophosphamide glycoside; aminolevulinic acid; eniluracil; amsacrine; bestrabucil; bisantrene; edatraxate; defofamine; demecolcine; diaziquone; elfornithine; elliptinium acetate; an epothilone; etoglucid; gallium nitrate; hydroxyurea; lentinan; lonidainine; maytansinoids such as maytansine and ansamitocins; mitoguazone; mitoxantrone; mopidanmol; nitiaerine; pentostatin; phenamet; pirarubicin; losoxantione; podophyllinic acid; 2-ethylhydrazide; procarbazine; PSK® polysaccharide complex (JHS Natural Products, Eugene, Oreg.); razoxane; rhizoxin; sizofiran; spirogermanium; tenuazonic acid; triaziquone; 2,2 2''-trichlorotiiethylamine; trichothecenes (especially T-2 toxin, verracurin A, roridin A and anguidine); urethan; vindesine; dacarbazine; mannomustine; mitobronitol; mitolactol; pipobroman; gacytosine; arabinoside ("Ara-C"); cyclophosphamide; thiotepa; taxoids, e.g., TAXOL® paclitaxel (Bristol-Myers Squibb Oncology, Princeton, N.J.), ABRAXANE® Cremophor-free, albumin-engineered nanoparticle formulation of paclitaxel (American Pharmaceutical Partners, Schaumberg, Ill.), and TAXOTERE® (docetaxel) (Rhone-Poulenc Rorer, Antony, France); chloranbucil; GEMZAR® (gemcitabine); 6-thioguanine; mercaptopurine; methotrexate; platinum coordination complexes such as cisplatin, oxaliplatin and carboplatin; vinblastine; platinum; etoposide (VP-16); ifosfamide; mitoxantrone; vincristine; NAVELBINE® vinorelbine; novantrone; teniposide; edatrexate; daunomycin; aminopterin; xeloda; ibandronate; irinotecan (e.g., CPT-11); topoisomerase inhibitor RFS 2000; difluoromethylornithine (DFMO); retinoids such as retinoic acid; capecitabine; leucovorin (LV); irenotecan; adrenocortical suppressant; adrenocorticosteroids; progestins; estrogens; androgens; gonadotropin-releasing hormone analogs; and

pharmaceutically acceptable salts, acids or derivatives of any of the above. Also included anticancer agents are anti-hormonal agents that act to regulate or inhibit hormone action on tumors such as anti-estrogens and selective estrogen receptor modulators (SERMs), including, for example, tamoxifen (including NOLVADEX® tamoxifen), raloxifene, droloxifene, 4-hydroxytamoxifen, trioxifene, keoxifene, LY117018, onapristone, and FARESTON-toremifene; aromatase inhibitors that inhibit the enzyme aromatase, which regulates estrogen production in the adrenal glands, such as, for example, 4(5)-imidazoles, aminoglutethimide, MEGASE® megestrol acetate, AROMASL® exemestane, formestanie, fadrozole, RIVISOR® vorozole, FEMARA® letrozole, and ARTMIDEX® anastrozole; and anti-androgens such as flutamide, nilutamide, bicalutamide, leuprolide, and goserelin; as well as troxacitabine (a 1,3-dioxolane nucleoside cytosine analog); antisense oligonucleotides, particularly those which inhibit expression of genes in signaling pathways implicated in abherant cell proliferation, such as, for example, PKC-alpha, Ralf and H-Ras; ribozymes such as a VEGF-A expression inhibitor (e.g., ANGIOZYME® ribozyme) and a HER2 expression inhibitor; vaccines such as gene therapy vaccines, for example, ALLOVECTIN® vaccine, LEUVECTIN® vaccine, and VAXID® vaccine; PROLEUKIN® rJL-2; LURTOTECAN® topoisomerase 1 inhibitor; ABARELLX® rmRH; antibodies such as trastuzumab and pharmaceutically acceptable salts, acids or derivatives of any of the above. Examples of protein payloads include mammalian proteins, such as, e.g., growth hormone (GH), including human growth hormone, bovine growth hormone, and other members of the GH supergene family; growth hormone releasing factor; parathyroid hormone; thyroid stimulating hormone; lipoproteins; alpha-1-antitrypsin; insulin A-chain; insulin B-chain; proinsulin; follicle stimulating hormone; calcitonin; luteinizing hormone; glucagon; clotting factors such as factor VIIIC, factor IX tissue factor, and von Willebrands factor; anti-clotting factors such as Protein C; atrial natriuretic factor; lung surfactant; a plasminogen activator, such as urokinase or tissue-type plasminogen activator (t-PA); bombazine; thrombin; alpha tumor necrosis factor, beta tumor necrosis factor; enkephalinase; RANTES (regulated on activation normally T-cell expressed and secreted); human macrophage

inflammatory protein (MIP-1-alpha); serum albumin such as human serum albumin; mullerian-inhibiting substance; relaxin A-chain; relaxin B-chain; prorelaxin; mouse gonadotropin-associated peptide; DNase; inhibin; activin; vascular endothelial growth factor (VEGF); receptors for hormones or growth factors; an integrin; protein A or D; rheumatoid factors; a neurotrophic factor such as bone-derived neurotrophic factor (BDNF), neurotrophin-3, -4, -5, or -6 (NT-3, NT-4, NT-5, or NT-6), or a nerve growth factor such as NGF-beta; platelet-derived growth factor (PDGF); fibroblast growth factor such as aFGF and bFGF; epidermal growth factor (EGF); transforming growth factor (TGF) such as TGF-alpha and TGF-beta, including TGF-beta1, TGF-beta2, TGF-beta3, TGF-beta4, or TGF-beta5; insulin-like growth factor-I and -II (IGF-I and IGF-II); des(1-3)-IGF-I (brain IGF-D; insulin-like growth factor binding proteins; CD proteins such as CD3, CD4, CD8, CD19 and CD20; osteoinductive factors; immunotoxins; a bone morphogenetic protein (BMP); T-cell receptors; surface membrane proteins; decay accelerating factor (DAF); a viral antigen such as, for example, a portion of the AIDS envelope; transport proteins; homing receptors; addressins; regulatory proteins; immunoadhesins; antibodies; and biologically active fragments or variants of any of the above-listed polypeptides.

[0055] The members of the GH supergene family include growth hormone, prolactin, placental lactogen, erythropoietin, thrombopoietin, interleukin-2, interleukin-3, interleukin-4, interleukin-5, interleukin-6, interleukin-7, interleukin-9, interleukin-10, interleukin-11, interleukin-12 (p35 subunit), interleukin-13, interleukin-15, oncostatin M, ciliary neurotrophic factor, leukemia inhibitory factor, alpha interferon, beta interferon, gamma interferon, omega interferon, tau interferon, granulocyte-colony stimulating factor, granulocyte-macrophage colony stimulating factor, macrophage colony stimulating factor, cardiotrophin-1 and other proteins identified and classified as members of the family.

[0056] Other payload agents that can be incorporated in the delivery vehicle include gastrointestinal therapeutic agents such as aluminum hydroxide, calcium carbonate, magnesium carbonate, sodium carbonate and the like; non-steroidal antifertility agents;

parasympathomimetic agents; psychotherapeutic agents; major tranquilizers such as chloropromazine HCl, clozapine, mesoridazine, metiapine, reserpine, thioridazine and the like; minor tranquilizers such as chlordiazepoxide, diazepam, meprobamate, temazepam and the like; rhinological decongestants; sedative-hypnotics such as codeine, phenobarbital, sodium pentobarbital, sodium secobarbital and the like; other steroids such as testosterone and testosterone propionate; sulfonamides; sympathomimetic agents; vaccines; vitamins and nutrients such as the essential amino acids, essential fats and the like; antimalarials such as 4-aminoquinolines, 8-aminoquinolines, pyrimethamine and the like; anti-migraine agents such as mazindol, phentermine and the like; anti-Parkinson agents such as L-dopa; anti-spasmodics such as atropine, methscopolamine bromide and the like; antispasmodics and anticholinergic agents such as bile therapy, digestants, enzymes and the like; antitussives such as dextromethorphan, noscapine and the like; bronchodilators; cardiovascular agents such as anti-hypertensive compounds, Rauwolfia alkaloids, coronary vasodilators, nitroglycerin, organic nitrates, pentaerythritotetranitrate and the like; electrolyte replacements such as potassium chloride; ergotalkaloids such as ergotamine with and without caffeine, hydrogenated ergot alkaloids, dihydroergocristine methanesulfate, dihydroergocomine methanesulfonate, dihydroergokroyptine methanesulfate and combinations thereof; alkaloids such as atropine sulfate, Belladonna, hyoscine hydrobromide and the like; analgesics; narcotics such as codeine, dihydrocodienone, meperidine, morphine and the like; non-narcotics such as salicylates, aspirin, acetaminophen, d-propoxyphene and the like.

[0057] Antibiotic payloads include, for example, the cephalosporins, chlorarnphenical, gentamicin, kanamycin A, kanamycin B, the penicillins, ampicillin, streptomycin A, antimycin A, chloropamtheniol, metronidazole, oxytetracycline penicillin G, the tetracyclines, and the like.

[0058] Other payload agents can include vaccines or antigenic agents. For example, payload antigens derived from microorganisms such as Neisseria gonorrhea, Mycobacterium tuberculosis, Herpes virus (humonis, types 1 and 2), Candida albicans, Candida

tropicalis, Trichomonas vaginalis, Haemophilus vaginalis, Group B Streptococcus sp., Microplasma hominis, Hemophilus ducreyi, Granuloma inguinale, Lymphopathia venereum, Treponema pallidum, Brucella abortus. Brucella melitensis, Brucella suis, Brucella canis, Campylobacter fetus, Campylobacter fetus intestinalis, Leptospira pomona, Listeria monocytogenes, Brucella ovis, equine herpes virus 1, equine arteritis virus, IBR-IBP virus, BVD-MB virus, Chlamydia psittaci, Trichomonas foetus, Toxoplasma gondii, Escherichia coli, Actinobacillus equuli, Salmonella abortus ovis, Salmonella abortus equi, Pseudomonas aeruginosa, Corynebacterium equi, Corynebacterium pyogenes, Actinobaccilus seminis, Mycoplasma bovigenitalium, Aspergillus fumigatus, Absidia ramosa, Trypanosoma equiperdum, Babesia caballi, Clostridium tetani, Clostridium botulinum and the like can be loaded. In other embodiments, the payload can comprise neutralizing antibodies that counteract the above microorganisms.

[0059] In other embodiments, the payload can comprise enzymes such as ribonuclease, neuramidinase, trypsin, glycogen phosphorylase, sperm lactic dehydrogenase, sperm hyaluronidase, adenossinetriphosphatase, alkaline phosphatase, alkaline phosphatase esterase, amino peptidase, trypsin chymotrypsin, amylase, muramidase, acrosomal proteinase, diesterase, glutamic acid dehydrogenase, succinic acid dehydrogenase, beta-glycophosphatase, lipase, ATP-ase alpha-peptate gamma-glutamylotranspeptidase, sterol-3-beta-ol-dehydrogenase, DPN-di-aprorase.

[0060] Peptide-payload conjugates are also encompassed by the disclosure, wherein a peptide of the disclosure is linked or fused directly to a payload molecule as set forth herein such that the peptide-payload conjugate can be directly delivery (without loading into a delivery vehicle) to target a desired tissue based upon the peptide's tropism.

[0061] The term "promoter" as used herein refers to any sequence that regulates the expression of a coding sequence, such as a gene. Promoters may be constitutive, inducible, repressible, or tissue-specific, for example. A "promoter" is a control sequence that is a region of a polynucleotide sequence at which initiation and rate of transcription are controlled. It may contain genetic elements at

which regulatory proteins and molecules may bind such as RNA polymerase and other transcription factors. Non-limiting exemplary promoters include CMV promoter and U6 promoter.

[0062] The term "protein", "peptide" and "polypeptide" are used interchangeably and in their broadest sense to refer to a compound of two or more subunit amino acids, amino acid analogs or peptidomimetics. The subunits can be linked by peptide bonds. In another embodiment, the subunit can be linked by other bonds, *e.g.*, ester, ether, *etc.* A protein or peptide can contain at least two amino acids and no limitation is placed on the maximum number of amino acids which can comprise a protein's or peptide's sequence. As mentioned above, the term "amino acid" can refer to either natural and/or unnatural or synthetic amino acids, including glycine and both the D and L optical isomers, amino acid analogs and peptidomimetics. As used herein, the term "fusion protein" can refer to a protein comprised of domains from more than one naturally occurring or recombinantly produced protein, where generally each domain serves a different function. In this regard, the term "linker" can refer to a peptide fragment that is used to link these domains together - optionally to preserve the conformation of the fused protein domains and/or prevent unfavorable interactions between the fused protein domains which can compromise their respective functions.

[0063] The terms "polynucleotide" and "oligonucleotide" are used interchangeably and refer to a polymeric form of nucleotides of any length, either deoxyribonucleotides or ribonucleotides or analogs thereof. Polynucleotides can have any three-dimensional structure and can perform any function, known or unknown. The following are non-limiting examples of polynucleotides: a gene or gene fragment (for example, a probe, primer, EST or SAGE tag), exons, introns, messenger RNA (mRNA), transfer RNA, ribosomal RNA, RNAi, ribozymes, cDNA, recombinant polynucleotides, branched polynucleotides, plasmids, vectors, isolated DNA of any sequence, isolated RNA of any sequence, nucleic acid probes and primers. A polynucleotide can comprise modified nucleotides, such as methylated nucleotides and nucleotide analogs. If present, modifications to the nucleotide structure can be imparted before or after assembly of the

polynucleotide. The sequence of nucleotides can be interrupted by non-nucleotide components. A polynucleotide can be further modified after polymerization, such as by conjugation with a labeling component. The term also can refer to both double and single stranded molecules. Unless otherwise specified or required, any embodiment of this disclosure that is a polynucleotide can encompass both the double stranded form and each of two complementary single stranded forms known or predicted to make up the double stranded form.

[0064] The term "polynucleotide sequence" can be the alphabetical representation of a polynucleotide molecule. This alphabetical representation can be input into databases in a computer having a central processing unit and used for bioinformatics applications such as functional genomics and homology searching.

[0065] Similarly, the term "polypeptide sequence", "peptide sequence" or "protein sequence" can be the alphabetical representation of a polypeptide molecule. This alphabetical representation can be input into databases in a computer having a central processing unit and used for bioinformatics applications such as functional proteomics and homology searching.

[0066] As used herein, the term "recombinant expression system" refers to a genetic construct or constructs for the expression of certain genetic material formed by recombination.

[0067] As used herein, the term "recombinant protein" can refer to a polypeptide or peptide which is produced by recombinant DNA techniques, wherein generally, DNA encoding the polypeptide or peptide is inserted into a suitable expression vector which is in turn used to transform a host cell to produce the heterologous polypeptide or peptide.

[0068] The term "sequencing" as used herein, can comprise bisulfite-free sequencing, bisulfite sequencing, TET-assisted bisulfite (TAB) sequencing, ACE-sequencing, high-throughput sequencing, Maxam-Gilbert sequencing, massively parallel signature sequencing, Polony sequencing, 454 pyrosequencing, Sanger sequencing, Illumina sequencing, SOLiD sequencing, Ion Torrent semiconductor sequencing, DNA nanoball sequencing, Heliscope single

molecule sequencing, single molecule real time (SMRT) sequencing, nanopore sequencing, shot gun sequencing, RNA sequencing, Enigma sequencing, or any combination thereof.

[0069] As used herein, the term "subject" is intended to mean any animal. In some embodiments, the subject may be a mammal; in further embodiments, the subject may be a bovine, equine, feline, murine, porcine, canine, human, or rat.

[0070] As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection (*e.g.*, using commercially available reagents such as, for example, LIPOFECTIN® (Invitrogen Corp., San Diego, CA), LIPOFECTAMINE® (Invitrogen), FUGENE® (Roche Applied Science, Basel, Switzerland), JETPEI™ (Polyplus-transfection Inc., New York, NY), EFFECTENE® (Qiagen, Valencia, CA), DREAMFECT™ (OZ Biosciences, France) and the like), or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989), and other laboratory manuals. Standard recombinant DNA and molecular cloning techniques used herein are well known in the art and are described in Sambrook, J., Fritsch, E.F. and Maniatis, T., Molecular Cloning: A Laboratory Manual, 2nd ed.; Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y., (1989) and by Silhavy, T.J., Bennan, M.L. and Enquist, L.W., Experiments with Gene Fusions; Cold Spring Harbor Laboratory: Cold Spring Harbor, N.Y., (1984); and by Ausubel, F.M. et. al., Current Protocols in Molecular Biology, Greene Publishing and Wiley-Interscience (1987) each of which are hereby incorporated by reference in its entirety. Additional useful methods are described in manuals including Advanced Bacterial Genetics (Davis, Roth and Botstein, Cold Spring Harbor Laboratory, 1980), Experiments with Gene Fusions (Silhavy, Berman and Enquist, Cold Spring Harbor Laboratory, 1984), Experiments in Molecular Genetics (Miller, Cold Spring Harbor Laboratory, 1972) Experimental Techniques in Bacterial Genetics (Maloy, in Jones and Bartlett, 1990), and A Short Course in

Bacterial Genetics (Miller, Cold Spring Harbor Laboratory 1992) each of which are hereby incorporated by reference in its entirety.

[0071] The terms "treat", "treating" and "treatment", as used herein, refers to ameliorating symptoms associated with a disease or disorder (*e.g.*, cancer, Covid-19 *etc.*), including preventing or delaying the onset of the disease or disorder symptoms, and/or lessening the severity or frequency of symptoms of the disease or disorder.

[0072] As used herein, the term "vector" can refer to a nucleic acid construct deigned for transfer between different hosts, including but not limited to a plasmid, a virus, a cosmid, a phage, a BAC, a YAC, *etc.* In some embodiments, a "viral vector" is defined as a recombinantly produced virus or viral particle that comprises a polynucleotide to be delivered into a host cell, either *in vivo*, *ex vivo* or *in vitro*. In some embodiments, plasmid vectors can be prepared from commercially available vectors. In other embodiments, viral vectors can be produced from baculoviruses, retroviruses, adenoviruses, AAVs, *etc.* according to techniques known in the art. In one embodiment, the viral vector is a lentiviral vector. Examples of viral vectors include retroviral vectors, adenovirus vectors, adeno-associated virus vectors, alphavirus vectors and the like. Infectious tobacco mosaic virus (TMV)-based vectors can be used to manufacturer proteins and have been reported to express Griffithsin in tobacco leaves (O'Keefe *et al.* (2009) Proc. Nat. Acad. Sci. USA 106(15):6099-6104). Alphavirus vectors, such as Semliki Forest virus-based vectors and Sindbis virus-based vectors, have also been developed for use in gene therapy and immunotherapy. See, Schlesinger & Dubensky (1999) Curr. Opin. Biotechnol. 5:434-439 and Ying *et al.* (1999) Nat. Med. 5(7):823-827. In aspects where gene transfer is mediated by a retroviral vector, a vector construct can refer to the polynucleotide comprising the retroviral genome or part thereof, and a gene of interest. Further details as to modern methods of vectors for use in gene transfer can be found in, for example, Kotterman *et al.* (2015) Viral Vectors for Gene Therapy: Translational and Clinical Outlook Annual Review of Biomedical Engineering 17. Vectors that contain both a promoter and a cloning site into which a polynucleotide can be operatively linked are well

known in the art. Such vectors are capable of transcribing RNA *in vitro* or *in vivo* and are commercially available from sources such as Agilent Technologies (Santa Clara, Calif.) and Promega Biotech (Madison, Wis.). In one aspect, the promoter is a pol III promoter.

[0073] Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors." In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. In the present specification, "plasmid" and 'Vector" can be used interchangeably. However, the disclosure is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions. Typically, the vector or plasmid contains sequences directing transcription and translation of a relevant gene or genes, a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene which harbors transcriptional initiation controls and a region 3' of the DNA fragment which controls transcription termination. Both control regions may be derived from genes homologous to the transformed host cell, although it is to be understood that such control regions may also be derived from genes that are not native to the species chosen as a production host.

[0074] Typically, the vector or plasmid contains sequences directing transcription and translation of a gene fragment, a selectable marker, and sequences allowing autonomous replication or chromosomal integration. Suitable vectors comprise a region 5' of the gene which harbors transcriptional initiation controls and a region 3' of the DNA fragment which controls transcription termination. Both control regions may be derived from genes homologous to the transformed host cell, although it is to be

understood that such control regions may also be derived from genes that are not native to the species chosen as a production host.

[0075] Initiation control regions or promoters, which are useful to drive expression of the relevant pathway coding regions in the desired host cell are numerous and familiar to those skilled in the art. Virtually any promoter capable of driving these genetic elements is suitable for the present invention including, but not limited to, lac, ara, tet, trp, IPL, IPR, T7, tac, and trc (useful for expression in Escherichia coli and Pseudomonas); the amy, apr, npr promoters and various phage promoters useful for expression in *Bacillus subtilis*, and *Bacillus licheniformis*; nisA (useful for expression in gram positive bacteria, Eichenbaum *et al.* Appl. Environ. Microbiol. 64(8):2763-2769 (1998)); and the synthetic P11 promoter (useful for expression in *Lactobacillus plantarum*, Rud *et al.*, Microbiology 152:1011-1019 (2006)). Termination control regions may also be derived from various genes native to the preferred hosts.

[0076] Adeno-associated viruses (AAVs) are common gene therapy vectors, however, their effectiveness is hindered by poor target tissue transduction and off-target delivery. Hypothesizing that naturally occurring receptor-ligand interactions could be repurposed to engineer tropism. The disclosure provides a method wherein all annotated protein ligands known to bind human receptors were fragmented into tiling 20-mer peptides and displayed onto the surface loops of AAV5 and AAV9 capsids at two sites. The resulting capsid libraries, comprising >1 million AAV variants, were screened across 9 tissues in C57BL/6 mice. Tracking variant abundance, >250,000 variants were identified which packaged into capsids, and >15,000 variants which efficiently transduced at least one mouse organ. Twenty-one (21) AAV variants were validated with 74.3% of the organ tropism predictions accurately reproducing, confirming overall screen efficacy. Systematic ligand tiling enabled prediction of putative AAV-receptor interactions, were successfully validated by targeted genetic perturbations. Comprehensive peptide tiling also enabled examination of homologous peptide activity. Interestingly, functional peptides that were observed tended to be derived from specific domains on ligands. Notably, certain peptides also

displayed consistent activity across mice strains, capsid insertion contexts, and capsid serotypes, including novel immune orthogonal serotypes. Further analyses of displayed peptides revealed that biophysical attributes were highly predictive of AAV variant packaging, and there was a machine learnable relationship between peptide sequence and tissue tropism. The disclosed comprehensive ligand peptide tiling and display approach can enable engineering of tropism across diverse viral, viral-like and non-viral delivery platforms, and shed light into basic receptor-ligand biology.

[0077] The disclosure provides a plurality of peptides useful for targeting a desired tissue. The disclosure also provides peptides that are immunosilent. The disclosure also provides vectors, delivery vehicles and peptide-payload conjugates comprising one or more of the peptides of the disclosure having a desired tropism.

[0078] Rational screening strategies have immense potential to expand the molecular tools available for clinical gene therapy applications. While AAV engineering efforts have been conducted for over a decade, advances in DNA synthesis have enabled us to create here a data-driven library of AAV variants leveraging existing functional biomolecules from nature. Using natural biomolecules as a defined source of inserted peptides has multiple benefits over random hexamers (and similar methods). First, natural biomolecules have been pre-filtered for biological functionality by millennia of evolutionary selection pressure. Second, a defined library allows for robust quantification of the fitness of each AAV variant, enabling facile stratification of AAV variants by infectivity across organs of interest. While this method was primarily applied to engineer AAVs, mining nature for functional biomolecules has applications in a wide range of protein engineering challenges, such as engineering orthogonal viral (including lentiviruses) and non-viral (including lipid nanoparticles) vectors or identifying biologic inhibitors of critical protein/protein interactions.

[0079] Additionally, it is believed that the rational ligand tiling approach described herein could lead to important insights into basic AAV-receptor binding and transduction more generally. Previous efforts have either utilized large-scale genetic screens or

unique observations about a particular AAV variant's transduction behavior. As the AAV variants described herein display peptides derived from natural receptor-interacting ligands, it is believed that the AAV variants may maintain some of this binding capacity which partially explains their altered transduction profiles.

[0080] The disclosure focuses on adeno-associated viruses (AAVs) which have emerged as a leading vector for gene delivery in clinical applications. While multiple AAV-mediated therapies have achieved regulatory approval, efficient directing of treatment to target tissues is challenging with systemic injection. To overcome this, high viral titers are often used in treatments, which has in-turn been associated with potential for hepatotoxicity in clinical trials. Localized injections are also problematic, often requiring invasive procedures with the potential for organ damage and long recovery times. Due to these delivery challenges, some gene therapeutics have elected to pursue, where feasible, *ex vivo* treatment designs to overcome targeting issues, but this in-turn can lead to dependency on complex lab procedures, and high manufacturing costs.

[0081] There are thus numerous ongoing research efforts to improve *in vivo* therapeutic targeting, AAV variants have been engineered to specifically target tissues such as the brain and muscle. This has predominantly been established using strategies of iteratively screening random peptides inserted into the AAV capsid, or capsid shuffling, or randomly mutagenizing the capsid sequence as a whole, or direct chemical engineering. Although mutagenizing AAV capsids via random oligomers has yielded functional capsids with novel properties a stochastic mutational screening strategy limits the ability to predict future functional variants, and thus rational and programmable engineering of viral phenotypes remains an elusive goal.

[0082] Towards rational engineering of viral function, deep mutational libraries and associated screens of function have enabled systematic mapping of capsid mutation fitness, providing critical information which can be used to predict future variant activity. Additionally, defined libraries of pooled oligonucleotides have been used to insert gene fragments derived from proteins with known

affinity to synapses into the AAV capsid, with the goal of improving retrograde axonal transport. While these methodologies have provided important insights for AAV engineering, much is still unknown regarding how AAV genotype impacts packaging and tissue transduction. Consequently, there is a critical need for systematic datasets mapping AAV genotype to clinically relevant properties such as organ specificity. Given the clinical danger of hepatotoxicity and other efficacy issues related to off-target transduction, leveraging screening technologies to yield highly specific AAV capsids has great value to the medical and scientific community.

[0083] With a goal to develop a rational and potentially programmable strategy for enabling tissue targeting, it was hypothesized that receptor-ligand interactions, which mediate the spectrum of naturally occurring cell-protein and cell-cell interactions, could be repurposed for engineering AAV tropism. However, as neither the interaction interfaces nor the cell type specificity profiles of receptor-ligand interactions are fully mapped to comprehensively interrogate as well as engineer these. To study and address this, all annotated protein ligands known to bind human receptors were fragmented into tiling short peptides and displayed onto surface loops of AAV capsids, an approach termed "AAV-PepTile". Short peptides grafted into stabilizing molecular scaffolds can recapitulate local protein domain structure, and as protein-protein interface sites are typically 1200-2000 $Å^2$, with specific peptide hot loops ranging from 4-8 amino acids (AAs) contributing maximally to the binding energy of protein-protein interactions, it was hypothesized that 20 amino acid peptide insertions into the AAV capsid could drastically alter its binding ability, and therefore, its transduction profile *in vivo*. While insertion of still longer peptides could, in principle, more faithfully recapitulate local ligand structures, AAVs do not generally tolerate larger than 20-25 amino acid insertions without severely compromising capsid packaging efficiency. Based on this, experiments proceeded using a 20 amino acid peptide + flanking 2 amino acid linker insertions, all together a library of >1 million AAV variants was constructed by inserting corresponding oligonucleotide pool synthesized gene fragments coding for potential

receptor-ligands and cell membrane permeable proteins into one of two surface loops on AAV5 and AAV9. Unlike random peptide libraries, ligand tiling enabled robust quantitation of tissue transduction rates for all variants screened. Furthermore, systematic examination of the activity of similar peptides generate predictions of putative receptor interactions driving AAV variant tropism. Quantifying transduction rates across nine organs, extremely specific variants targeting the brain and lung were identified, as well as muscle and heart targeting variants with broader organ transduction. The resulting data linking AAV variant genotype to packaging efficacy and tissue specificity expands the understanding of the AAV fitness landscape, and provides a unique resource from which further data-driven engineering efforts can be built.

[0084] The general schematic for the AAV screens is outlined in **Fig. 3a**. Briefly, in the primary screen, a 20 amino acid peptide library consisting of tiled fragments from all known receptor-interacting ligands along with other interesting protein classes flanked by glycine-serine linkers were displayed into the 2 surface-exposed loops of two clinically relevant AAV serotypes, AAV5 and AAV9 with their sequences and specific amino acid site of insertion shown in **Table 1.**

[0085] **Table 1:** Showing both the DNA and amino acid sequences and insertion sites of the two wild-type AAV serotypes utilized in the screens, AAV5 and AAV9. Bolded and underlined are the flanking residues of the of the 'Loop 1' site and shown in bold/ital/underline are the flanking residues of the 'Loop 2' site.

Name	DNA Sequence	AA Sequence

AAV5	ATGTCTTTTGTTGATCACCCTCCAGATTGGTTGGAAGAAGTT GGTGAAGGTCTTCGCGAGTTTTTGGGCCTTGAAGCGGGCCCA CCGAAACCAAAACCCAATCAGCAGCATCAAGATCAAGCCCG TGGTCTTGTGCTGCCTGGTTATAACTATCTCGGACCCGGAAA CGGTCTCGATCGAGGAGAGCCTGTCAACAGGGCAGACGAGG TCGCGCGAGAGCACGACATCTCGTACAACGAGCAGCTTGAG GCGGGAGACAACCCCTACCTCAAGTACAACCACGCGGACGC CGAGTTTCAGGAGAAGCTCGCCGACGACACATCCTTCGGGG GAAACCTCGGAAAGGCAGTCTTTCAGGCCAAGAAAAGGGTT CTCGAACCTTTTGGCCTGGTTGAAGAGGGTGCTAAGACGGC CCCTACCGGAAAGCGGATAGACGACCACTTTCCAAAAAGAA AGAAGGCTCGGACCGAAGAGGACTCCAAGCCTTCCACCTCG TCAGACGCCGAAGCTGGACCCAGCGGATCCCAGCAGCTGCA AATCCCAGCCCAACCAGCCTCAAGTTTGGGAGCTGATACAA TGTCTGCGGGAGGTGGCGGCCCATTGGGCGACAATAACCAA GGTGCCGATGGAGTGGGCAATGCCTCGGGAGATTGGCATTG CGATTCCACGTGGATGGGGGACAGAGTCGTCACCAAGTCCA CCCGAACCTGGGTGCTGCCCAGCTACAACAACCACCAGTAC CGAGAGATCAAAAGCGGCTCCGTCGACGGAAGCAACGCCAA CGCCTACTTTGGATACAGCACCCCCTGGGGGTACTTTGACTT TAACCGCTTCCACAGCCACTGGAGCCCCCGAGACTGGCAAA GACTCATCAACAACTACTGGGGCTTCAGACCCCGGTCCCTCA GAGTCAAAATCTTCAACATTCAAGTCAAAGAGGTCACGGTG CAGGACTCCACCACCACCATCGCCAACAACCTCACCTCCACC GTCCAAGTGTTTACGGACGACGACTACCAGCTGCCCTACGTC GTCGGCAACGGGACCGAGGGATGCCTGCCGGCCTTCCCTCC GCAGGTCTTTACGCTGCCGCAGTACGGTTACGCGACGCTGA ACCGCGACAACACAGAAAATCCCACCGAGAGGAGCAGCTTC TTCTGCCTAGAGTACTTTCCCAGCAAGATGCTGAGAACGGGC AACAACTTTGAGTTTACCTACAACTTTGAGGAGGTGCCCTTC CACTCCAGCTTCGCTCCCAGTCAGAACCTGTTCAAGCTGGCC AACCCGCTGGTGGACCAGTACTTGTACCGCTTCGTGAGCACA AAT**AACACT**GGCGGAGTCCAGTTCAACAAGAACCTGGCCGG GAGATACGCCAACACCTACAAAAACTGGTTCCCGGGGCCCA TGGGCCGAACCCAGGGCTGGAACCTGGGCTCCGGGGTCAAC CGCGCCAGTGTCAGCGCCTTCGCCACGACCAATAGGATGGA GCTCGAGGGCGCGAGTTACCAGGTGCCCCCGCAGCCGAACG GCATGACCAACAACCTCCAGGGCAGCAACACCTATGCCCTG GAGAACACTATGATCTTCAACAGCCAGCCGGCGAACCCGGG CACCACCGCCACGTACCTCGAGGGCAACATGCTCATCACCA GCGAGAGCGAGACGCAGCCGGTGAACCGCGTGGCGTACAAC GTCGGCGGGCAGATGGCCACCAACAACCAGAGC***TCCACC***AC TGCCCCCGCGACCGGCACGTACAACCTCCAGGAAATCGTGC CCGGCAGCGTGTGGATGGAGAGGGACGTGTACCTCCAAGGA CCCATCTGGGCCAAGATCCCAGAGACGGGGGCGCACTTTCA CCCCTCTCCGGCCATGGGCGGATTCGGACTCAAACACCCACC GCCCATGATGCTCATCAAGAACACGCCTGTGCCCGGAAATA TCACCAGCTTCTCGGACGTGCCCGTCAGCAGCTTCATCACCC AGTACAGCACCGGGCAGGTCACCGTGGAGATGGAGTGGGAG CTCAAGAAGGAAAACTCCAAGAGGTGGAACCCAGAGATCCA GTACACAAACAACTACAACGACCCCCAGTTTGTGGACTTTGC CCCGGACAGCACCGGGGAATACAGAACCACCAGACCTATCG GAACCCGATACCTTACCCGACCCCTTTAA (SEQ ID NO:1)	MSFVDHPPDWLEEVGEGLREFLGL EAGPPKPKPNQQHQDQARGLVLPG YNYLGPGNGLDRGEPVNRADEVAR EHDISYNEQLEAGDNPYLKYNHAD AEFQEKLADDTSFGGNLGKAVFQA KKRVLEPFGLVEEGAKTAPTGKRID DHFPKRKKARTEEDSKPSTSSDAEA GPSGSQQLQIPAQPASSLGADTMSA GGGGPLGDNNQGADGVGNASGDW HCDSTWMGDRVVTKSTRTWVLPS YNNHQYREIKSGSVDGSNANAYFG YSTPWGYFDFNRFHSHWSPRDWQR LINNYWGFRPRSLRVKIFNIQVKEV TVQDSTTTIANNLTSTVQVFTDDDY QLPYVVGNGTEGCLPAFPPQVFTLP QYGYATLNRDNTENPTERSSFFCLE YFPSKMLRTGNNFEFTYNFEEVPFH SSFAPSQNLFKLANPLVDQYLYRFV STN**NT**GGVQFNKNLAGRYANTYK NWFPGPMGRTQGWNLGSGVNRAS VSAFATTNRMELEGASYQVPPQPN GMTNNLQGSNTYALENTMIFNSQP ANPGTTATYLEGNMLITSESETQPV NRVAYNVGGQMATNNQS***ST***TAPA TGTYNLQEIVPGSVWMERDVYLQG PIWAKIPETGAHFHPSPAMGGFGLK HPPPMMLIKNTPVPGNITSFSDVPVS SFITQYSTGQVTVEMEWELKKENS KRWNPEIQYTNNYNDPQFVDFAPD STGEYRTTRPIGTRYLTRPL (SEQ ID NO:2)

AAV9	ATGGCTGCCGATGGTTATCTTCCAGATTGGCTCGAGGACAAC CTTAGTGAAGGAATTCGCGAGTGGTGGGCTTTGAAACCTGG AGCCCCTCAACCCAAGGCAAATCAACAACATCAAGACAACG CTCGAGGTCTTGTGCTTCCGGGTTACAAATACCTTGGACCCG GCAACGGACTCGACAAGGGGGAGCCGGTCAACGCAGCAGA CGCGGCGGCCCTCGAGCACGACAAGGCCTACGACCAGCAGC TCAAGGCCGGAGACAACCCGTACCTCAAGTACAACCACGCC GACGCCGAGTTCCAGGAGCGGCTCAAAGAAGATACGTCTTT TGGGGGCAACCTCGGGCGAGCAGTCTTCCAGGCCAAAAAGA GGCTTCTTGAACCTCTTGGTCTGGTTGAGGAAGCGGCTAAGA CGGCTCCTGGAAAGAAGAGGCCTGTAGAGCAGTCTCCTCAG GAACCGGACTCCTCCGCGGGTATTGGCAAATCGGGTGCACA GCCCGCTAAAAAGAGACTCAATTTCGGTCAGACTGGCGACA CAGAGTCAGTCCCAGACCCTCAACCAATCGGAGAACCTCCC GCAGCCCCCTCAGGTGTGGGATCTCTTACAATGGCTTCAGGT GGTGGCGCACCAGTGGCAGACAATAACGAAGGTGCCGATGG AGTGGGTAGTTCCTCGGGAAATTGGCATTGCGATTCCCAATG GCTGGGGGACAGAGTCATCACCACCAGCACCCGAACCTGGG CCCTGCCCACCTACAACAATCACCTCTACAAGCAAATCTCCA ACAGCACATCTGGAGGATCTTCAAATGACAACGCCTACTTC GGCTACAGCACCCCCTGGGGGTATTTTGACTTCAACAGATTC CACTGCCACTTCTCACCACGTGACTGGCAGCGACTCATCAAC AACAACTGGGGATTCCGGCCTAAGCGACTCAACTTCAAGCT CTTCAACATTCAGGTCAAAGAGGTTACGGACAACAATGGAG TCAAGACCATCGCCAATAACCTTACCAGCACGGTCCAGGTCT TCACGGACTCAGACTATCAGCTCCCGTACGTGCTCGGGTCGG CTCACGAGGGCTGCCTCCCGCCGTTCCCAGCGGACGTTTTCA TGATTCCTCAGTACGGGTATCTGACGCTTAATGATGGAAGCC AGGCCGTGGGTCGTTCGTCCTTTTACTGCCTGGAATATTTCC CGTCGCAAATGCTAAGAACGGGTAACAACTTCCAGTTCAGC TACGAGTTTGAGAACGTACCTTTCCATAGCAGCTACGCTCAC AGCCAAAGCCTGGACCGACTAATGAATCCACTCATCGACCA ATACTTGTACTATCTCTCAAAGACTATTAACGGTTCTGGA**CA** **GAAT**CAACAAACGCTAAAATTCAGTGTGGCCGGACCCAGCA ACATGGCTGTCCAGGGAAGAAACTACATACCTGGACCCAGC TACCGACAACAACGTGTCTCAACCACTGTGACTCAAAACAA CAACAGCGAATTTGCTTGGCCTGGAGCTTCTTCTTGGGCTCT CAATGGACGTAATAGCTTGATGAATCCTGGACCTGCTATGGC CAGCCACAAAGAAGGAGAGGACCGTTTCTTTCCTTTGTCTGG ATCTTTAATTTTTGGCAAACAAGGAACTGGAAGAGACAACG TGGATGCGGACAAAGTCATGATAACCAACGAAGAAGAAATT AAAACTACTAACCCGGTAGCAACGGAGTCCTATGGACAAGT GGCCACAAACCACCAGAGT***GCCCAA***GCACAGGCGCAGACCG GCTGGGTTCAAAACCAAGGAATACTTCCGGGTATGGTTTGG CAGGACAGAGATGTGTACCTGCAAGGACCCATTTGGGCCAA AATTCCTCACACGGACGGCAACTTTCACCCTTCTCCGCTGAT GGGAGGGTTTGGAATGAAGCACCCGCCTCCTCAGATCCTCA TCAAAAACACACCTGTACCTGCGGATCCTCCAACGGCCTTCA ACAAGGACAAGCTGAACTCTTTCATCACCCAGTATTCTACTG GCCAAGTCAGCGTGGAGATCGAGTGGGAGCTGCAGAAGGA AAACAGCAAGCGCTGGAACCCGGAGATCCAGTACACTTCCA ACTATTACAAGTCTAATAATGTTGAATTTGCTGTTAATACTG AAGGTGTATATAGTGAACCCCGCCCCATTGGCACCAGATAC CTGACTCGTAATCTGTAA (SEQ ID NO:3)	MAADGYLPDWLEDNLSEGIREWW ALKPGAPQPKANQQHQDNARGLVL PGYKYLGPGNGLDKGEPVNAADA AALEHDKAYDQQLKAGDNPYLKY NHADAEFQERLKEDTSFGGNLGRA VFQAKKRLLEPLGLVEEAAKTAPG KKRPVEQSPQEPDSSAGIGKSGAQP AKKRLNFGQTGDTESVPDPQPIGEP PAAPSGVGSLTMASGGGAPVADNN EGADGVGSSSGNWHCDSQWLGDR VITTSTRTWALPTYNNHLYKQISNS TSGGSSNDNAYFGYSTPWGYFDFN RFHCHFSPRDWQRLINNNWGFRPK RLNFKLFNIQVKEVTDNNGVKTIAN NLTSTVQVFTDSDYQLPYVLGSAHE GCLPPFPADVFMIPQYGYLTLNDGS QAVGRSSFYCLEYFPSQMLRTGNNF QFSYEFENVPFHSSYAHSQSLDRLM NPLIDQYLYYLSKTINGSG**QN**QQTL KFSVAGPSNMAVQGRNYIPGPSYR QQRVSTTVTQNNNSEFAWPGASSW ALNGRNSLMNPGPAMASHKEGEDR FFPLSGSLIFGKQGTGRDNVDADKV MITNEEEIKTTNPVATESYGQVATN HQS***AO***AQAQTGWVQNQGILPGMV WQDRDVYLQGPIWAKIPHTDGNFH PSPLMGGFGMKHPPPQILIKNTPVP ADPPTAFNKDKLNSFITQYSTGQVS VEIEWELQKENSKRWNPEIQYTSNY YKSNNVEFAVNTEGVYSEPRPIGTR YLTRNL (SEQ ID NO:4)

[0086] These libraries were used to make AAV capsid particles which were injected systemically into mice in duplicate. After two weeks, the heart, lung, liver, small intestine, spleen, pancreas,

kidney, brain, and skeletal muscle were extracted. Genomic DNA was isolated from these tissues and the variable peptide region of interest was amplified out and subjected to next generation sequencing (NGS) to quantify AAV abundance in each tissue by measuring the $\log_2$ fold change (log2FC) value in each organ relative to the injected pool.

[0087] A candidate list of "hit peptides" in the AAV9Loop1 and AAV9Loop2 scaffolds which were synthesized and subjected to deep mutagenesis to yield AAV libraries with millions of variants that were then profiled in a secondary screen. As done in the primary screen, AAV capsid particles were produced from these libraries and injected into mice where the tropism was quantified. Additionally, to further evaluate the species-dependent nature of the engineered viral particles, the AAV capsid libraries were pooled from the primary and secondary screens and injected into rhesus macaques as an important preclinical non-human primate (NHP) model. Similar to the experiments done in mice, various organs were isolated from the NHPs 3 weeks after injection, the variable peptide region was amplified out and the AAV abundance quantified in the liver, lung, and 4 lobes of the brain. moreover, to apply additional selected pressure, the variable peptide regions from the heart, lung, pancreas, intestine, brain, and muscle were amplified out and cloned back into the AAV9Loop1 and AAV9Loop2 scaffolds. This tertiary screen was then subjected to two weeks on *in vivo* selective pressure in mice before the tropism was quantified by NGS. For all screens, a "hit" was defined as an engineered AAV variant that exhibited a log2FC relative to the capsid > 1 in both replicates and a p-value < 0.1 in at least one organ.

[0088] Shown in **Table 2** are the peptides that were present in the starting library of the primary screen that showed prominent *in vivo* activity across all four scaffolds in which they were grafted. In addition to this list, there are a set of peptides that were consistently identified as hits across both the primary, secondary, tertiary, and NHP screens. These screen-independent hits for both AAV9Loop1 and AAV9Loop2 are shown in **Table 3.**

[0089] **Table 2:** Table for "wild-type" peptides that were present in the primary screen library and showed *in vivo* activity across all

4 scaffolds in which they were grafted (i.e. AAV5Loop1, AAV5Loop2, AAV9Loop1, AAV9Loop2). Shown is the amino acid sequence.

Amino Acid Sequence	Amino Acid Sequence	Amino Acid Sequence
SSQAGIPVRAPRSPRAPEPA (SEQ ID NO: 5)	TSRMSTSQIPGRVASSGLQS (SEQ ID NO: 6)	KQLNQQHSQIKEIENQLRRT (SEQ ID NO: 7)
VLGAPRAGSGAHTAVISPQD (SEQ ID NO: 8)	QGASVNSSSQTAAQVTKGVK (SEQ ID NO: 9)	KALAAAGYDVEKNNSRIKLS (SEQ ID NO: 10)
LEALSAGPGTSGSHLDRSTV (SEQ ID NO: 11)	TSTMPNSSQSSHSSSGTLSS (SEQ ID NO: 12)	GIKSLVSKGTLVQTKGTGAS (SEQ ID NO: 13)
VVYTGHDTKLMQNSTKAPLK (SEQ ID NO: 14)	STMPNSSQSSHSSSGTLSSL (SEQ ID NO: 15)	RNNMIQTKPTGTQQSTNTAV (SEQ ID NO: 16)
HDTKLMQNSTKAPLKRSNVE (SEQ ID NO: 17)	ANGDKEIGNIISDAMKKVGR (SEQ ID NO: 18)	AGLTRISKEILTQGQINSTS (SEQ ID NO: 19)
QTAHVGVGISGNEGMQATNN (SEQ ID NO: 20)	SKYVNKEIQNAVNGVKQIKT (SEQ ID NO: 21)	NHFGTASSTGNLVVKDPTRV (SEQ ID NO: 22)
TAHVGVGISGNEGMQATNNS (SEQ ID NO: 23)	EQLGGTVTNPGGSGTSSRLE (SEQ ID NO: 24)	YNSAGTGPSSATVNVTTRKP (SEQ ID NO: 25)
ALAAGGYDVEKNNSRIKLGL (SEQ ID NO: 26)	VDFRIHVENQTRARDDVSRK (SEQ ID NO: 27)	LSLTQSDISHIGSMRVEGIV (SEQ ID NO: 28)
SDGSLEDDDYHATTETIITM (SEQ ID NO: 29)	DFRIHVENQTRARDDVSRKQ (SEQ ID NO: 30)	SSRGTTGNTQVQSPSNATTA (SEQ ID NO: 31)
HVAMEMKEDINTKKEDMMFH (SEQ ID NO: 32)	AGVAHAITAACTQGNLSDCG (SEQ ID NO: 33)	RGTTGNTQVQSPSNATTATG (SEQ ID NO: 34)
AIAAGATARATTASRLPSSA (SEQ ID NO: 35)	GSVGTQGRACNKTAPQASGC (SEQ ID NO: 36)	GTTGNTQVQSPSNATTATGV (SEQ ID NO: 37)
TSPATPSAGGAASSRTPNRI (SEQ ID NO: 38)	AEDVRGNLRGNTEGLQKSLA (SEQ ID NO: 39)	IKVSNTLESRLDLSAKQKMP (SEQ ID NO: 40)
SPATPSAGGAASSRTPNRIS (SEQ ID NO: 41)	QPAVVHLQGQGSAIQVKNDL (SEQ ID NO: 42)	VSSLPTTNSTTNSATKITAN (SEQ ID NO: 43)
QDAASSSSSSSSSATTTTPE (SEQ ID NO: 44)	NNGSMRNTSQDREREVDNNR (SEQ ID NO: 45)	SLPTTNSTTNSATKITANTT (SEQ ID NO: 46)
SLKASSTMAKSENLVKSHVQ (SEQ ID NO: 47)	KENGGARLAEYHAKATEHLS (SEQ ID NO: 48)	AVSRVADTIMSGPSNSQQVP (SEQ ID NO: 49)
KASSTMAKSENLVKSHVQLQ (SEQ ID NO: 50)	KQINDYVEKGTQGKIVDLVK (SEQ ID NO: 51)	MIVAAIQAEKNRAGSSRQSI (SEQ ID NO: 52)
SSTMAKSENLVKSHVQLQNY (SEQ ID NO: 53)	SNAIKNLPPPLGGAAGHPGS (SEQ ID NO: 54)	KSEVSTDLIQANPKQGALAK (SEQ ID NO: 55)
STMAKSENLVKSHVQLQNYS (SEQ ID NO: 56)	PEFSNDIKKRKVDDKDSSHY (SEQ ID NO: 57)	YGGGATNANARAKGGTGQTG (SEQ ID NO: 58)
TMAKSENLVKSHVQLQNYSK (SEQ ID NO: 59)	FSNDIKKRKVDDKDSSHYDS (SEQ ID NO: 60)	GGATNANARAKGGTGQTGGL (SEQ ID NO: 61)
SALAGQRQGTQAESNLSSKF (SEQ ID NO: 62)	NVINNAILRANLQKGENPSH (SEQ ID NO: 63)	ATNANARAKGGTGQTGGLTQ (SEQ ID NO: 64)
ALAGQRQGTQAESNLSSKFQ (SEQ ID NO: 65)	VASVSLLQLGQETDKLKTRN (SEQ ID NO: 66)	ISSDSIGVLGYQKTVDHTKV (SEQ ID NO: 67)
FQFSSNKEQNGVQDPQHERI (SEQ ID NO: 68)	EVHNAKTKPREEQYNSTYRV (SEQ ID NO: 69)	PLTPANNLPDTQSSGPAHSK (SEQ ID NO: 70)
SSFQGSKRGLNSSFETSPKK (SEQ ID NO: 71)	AARPYSATAAAATGGMANVK (SEQ ID NO: 72)	TSVFLKNNKIAGVNNRTEYV (SEQ ID NO: 73)

VEVHNAKTKPREEQFNSTYR (SEQ ID NO: 74)	TNSNNHLAADSQAYDQKNIR (SEQ ID NO: 75)	LKNNKIAGVNNRTEYVETTS (SEQ ID NO: 76)
EFSVDETGQMSATAKGRVRL (SEQ ID NO: 77)	KRSSQSTLERSEQQIRAASS (SEQ ID NO: 78)	APANTHNNNIEGQAIVYQNK (SEQ ID NO: 79)
AIVKANANAAAINNLASSIQ (SEQ ID NO: 80)	RFVQNALNGNGDPNNMDKAV (SEQ ID NO: 81)	KNTALDKNYLLNNKIPSNNV (SEQ ID NO: 82)
NAVARVADTIGSGPSNSQAV (SEQ ID NO: 83)	PLTPVNNLPDTRSSGPAHSK (SEQ ID NO: 84)	

[0090] **Table 3:** Table for "wild-type" peptides that were present as hits in both the primary, secondary, tertiary, and NHP screens. Shown are the peptide amino acid sequences

Amino Acid Sequence	Amino Acid Sequence	Amino Acid Sequence
LGIEQGVGVPGGSPGSRHRQ (SEQ ID NO: 85)	YISFNPKENKITVKSKNLPE (SEQ ID NO: 86)	SATLNSVLNSNAIKNLPPPL (SEQ ID NO: 87)
GGGGPGGSGGVGGGSGGYTY (SEQ ID NO: 88)	MAKSENLVKSHVQLQNYSKV (SEQ ID NO: 89)	VTSGKSIDDNSTKSEKQTPR (SEQ ID NO: 90)
SAGPGTSGSHLDRSTVSSFS (SEQ ID NO: 91)	MIETYNQTSPRSAATGLPIS (SEQ ID NO: 92)	VATKGLGVHAKSSDWGLQAG (SEQ ID NO: 93)
AARDSLSLSSGSGGGSASAS (SEQ ID NO: 94)	LPEAVLALYNSTRDRVAGES (SEQ ID NO: 95)	LENNKTMNRAENGGRPPHHP (SEQ ID NO: 96)
ILTQGQINSTSGLNKESFRY (SEQ ID NO: 97)	HCRQEATNATRVIHDNYDQL (SEQ ID NO: 98)	MWVDDQEVINKASNSNKIRL (SEQ ID NO: 99)
RALGTSPELPTGVTGSSGTR (SEQ ID NO: 100)	VTNNVKDVTKLVANLPKDYM (SEQ ID NO: 101)	KVQGLVPAGGSSSNSTREPP (SEQ ID NO: 102)
LKENGGARLAEYHAKATEHL (SEQ ID NO: 103)	NSVLNSNAIKNLPPPLGGAA (SEQ ID NO: 104)	GIVVYTGHDTKLMQNSTKAP (SEQ ID NO: 105)
GLRGGAGPPGPEGGKGAAGP (SEQ ID NO: 106)	EPPPELENNKTMNRAENGGR (SEQ ID NO: 107)	ETGWTPGLRSLENDGTKLLD (SEQ ID NO: 108)
RTLVSSKPQAHGAPAPPSGS (SEQ ID NO: 109)	ATLPQCQAANKTCPTNYMWN (SEQ ID NO: 110)	WETGAVPGLRSLENDGTRLL (SEQ ID NO: 111)
VDNEDDHGRGDLGRLKKQKD (SEQ ID NO: 112)	KVVPEGIRMNKTVAVRTLDP (SEQ ID NO: 113)	WRSRYPHKPEINSTTHPGAD (SEQ ID NO: 114)
MGDMANSSIEFHPKPQQQRD (SEQ ID NO: 115)	GSHGTQDRQCNKTSNGSDSC (SEQ ID NO: 116)	NFSAAFSVNYDTKSGPKNMT (SEQ ID NO: 117)
EGTICKRARGDDMDDYCNGK (SEQ ID NO: 118)	LLREGNTVRSFRAAAAETLE (SEQ ID NO: 119)	TSGLLETNFTASARTTGSGL (SEQ ID NO: 120)
QDNTCEEYHTVGPHSCHIPK (SEQ ID NO: 121)	LPGDPTSNATSRGAEGPLPP (SEQ ID NO: 122)	QRSGASASGVYTIQVSNATK (SEQ ID NO: 123)
GSSGTGGTATWKPGSSGPGS (SEQ ID NO: 124)	TLGADLVDSTDNTKINRGIF (SEQ ID NO: 125)	RTSGLLETNFTASARTTGSG (SEQ ID NO: 126)
LQGAGNSAAGRGAIADTFRS (SEQ ID NO: 127)	WNRQSSTSVIETNKTSVELS (SEQ ID NO: 128)	VPEGIRMNKTVAVRTLDPER (SEQ ID NO: 129)
MKGASRSPEDSSPDAARIRV (SEQ ID NO: 130)	GSFGTRDRTCNVTSHGIDGC (SEQ ID NO: 131)	EHGAVSQEEVIRAYDTTKKK (SEQ ID NO: 132)
GAPRAGSGAHTAVISPQDPT (SEQ ID NO: 133)	SIGVLGYQKTVDHTKVNSKL (SEQ ID NO: 134)	SNRVRVSIQTGTAFQPDAAM (SEQ ID NO: 135)

TFFRGLESQTKLVNGQSHIS (SEQ ID NO: 136)	AGVVCTNETRSTHTLDLSRE (SEQ ID NO: 137)	FNDMANAAKGLEWISNKISK (SEQ ID NO: 138)
GSAGPPGATGFPGAAGRVGP (SEQ ID NO: 139)	LFGTTISNNIKYGRDDVTDE (SEQ ID NO: 140)	VAEIPFNSTNKYQLSIHETE (SEQ ID NO: 141)
MSTSQIPGRVASSGLQSVVH (SEQ ID NO: 142)	EILELAGNASRDNKKTRIIP (SEQ ID NO: 143)	RTILPMDQEFTYRVKNREQA (SEQ ID NO: 144)
TPGPQGIAGQRGVVGLPGQR (SEQ ID NO: 145)	GTAGRVCNLTSRGMDSCEVM (SEQ ID NO: 146)	LTATLGADLVDSTDNTKINR (SEQ ID NO: 147)
VEGTARGVVVATGDRTVMGR (SEQ ID NO: 148)	YWGSVTGSQSNAVSPTPAPR (SEQ ID NO: 149)	VEVGGSAKYLNNTKKYQNQS (SEQ ID NO: 150)
PFPGLGTAAAPAQGGAHLKQ (SEQ ID NO: 151)	FNSHINASNNSEPSRMNMTS (SEQ ID NO: 152)	QSAKDNNIKFRAIGHSDNSV (SEQ ID NO: 153)
ARARELSSGQKDQGQAANML (SEQ ID NO: 154)	ENNKTMNRAENGGRPPHHPF (SEQ ID NO: 155)	PELENNKTMNRAENGGRPPH (SEQ ID NO: 156)
KPQKQLTAQSTPSTSGVNSQ (SEQ ID NO: 157)	KHSTIFENLANKADRDQYEL (SEQ ID NO: 158)	DNRTRQHIPYREDKNLTGTA (SEQ ID NO: 159)
RLEALKENGGARLAEYHAKA (SEQ ID NO: 160)	PPPELENNKTMNRAENGGRP (SEQ ID NO: 161)	PRGAAPPGQKGQANATRIPA (SEQ ID NO: 162)
GVDNEDDHGRGDLGRLKKQK (SEQ ID NO: 163)	GGSSSNSTREPPASGRVALV (SEQ ID NO: 164)	MLTCPEANATKVCGSDGVTY (SEQ ID NO: 165)
TAGSAMGAASLTLSAQSRTL (SEQ ID NO: 166)	LNSVLNSNAIKNLPPPLGGA (SEQ ID NO: 167)	TWLVRHTENVTKTKTPEDYA (SEQ ID NO: 168)
GRPGLPGAAGARGNDGARGS (SEQ ID NO: 169)	YRWNRQSSTSVIETNKTSVE (SEQ ID NO: 170)	TGHDTKLMQNSTKAPLKRSN (SEQ ID NO: 171)
VRHIPDRGAPTRASEPASAA (SEQ ID NO: 172)	CVVADGNSTRSPETNGLLCG (SEQ ID NO: 173)	SQTIPQISNRTLNTLVQDNS (SEQ ID NO: 174)
PSSAGSWNSGSSGPGSTGNR (SEQ ID NO: 175)	MIQTLTNKSKIQNVGGTYEL (SEQ ID NO: 176)	GSSSNSTREPPASGRVALVK (SEQ ID NO: 177)
RGSSSRQSSSSAMSSSSASS (SEQ ID NO: 178)	CRHERDAGVVCTNETRSTHT (SEQ ID NO: 179)	SDSIGVLGYQKTVDHTKVNS (SEQ ID NO: 180)
DLSGITGARNLAVSQVVHKA (SEQ ID NO: 181)	LLGTCDQAPKSGSHEVRCNL (SEQ ID NO: 182)	LVRHTENVTKTKTPEDYADS (SEQ ID NO: 183)
VYADSFVIRGDEVRQIAPGQ (SEQ ID NO: 184)	SHGTQDRQCNKTSNGSDSCD (SEQ ID NO: 185)	PELKTPLTANITKSGNTFRP (SEQ ID NO: 186)
VRLVLTNPEGRKEGKEVTFQ (SEQ ID NO: 187)	TQGRACNKTAPQASGCDLMC (SEQ ID NO: 188)	GTQGRECNRTSEGADGCNLL (SEQ ID NO: 189)
VRLVLTNPEGRKEGKEVTFQ (SEQ ID NO: 187)	DKMFSQNDTRCAGPEAVKGQ (SEQ ID NO: 190)	GLANGVRRYASQTELYDAPG (SEQ ID NO: 191)
LPKPQSNLPDTKASGPAHSK (SEQ ID NO: 192)	SSDSNNLKLNNVRLPRENMS (SEQ ID NO: 193)	LERAASRANTVRSFHHEESL (SEQ ID NO: 194)
TVQAARTPGSLEGGSQPWRP (SEQ ID NO: 195)	TDANQLWTLKRDNTIRSNGK (SEQ ID NO: 196)	LLTVTSGKSIDDNSTKSEKQ (SEQ ID NO: 197)
GAGVAGRITGSMSLASGKSL (SEQ ID NO: 198)	EKYIDQEELNKTKPIWTRNP (SEQ ID NO: 199)	ENVDFRIHVENQTRARDDVS (SEQ ID NO: 200)
TAGAAMGAASLTLSAQSRTL (SEQ ID NO: 201)	ASSTMAKSENLVKSHVQLQN (SEQ ID NO: 202)	PEAVLALYNSTRDRVAGESA (SEQ ID NO: 203)
ENGGARLAEYHAKATEHLST (SEQ ID NO: 204)	LGTAGRVCNLTSRGMDSCEV (SEQ ID NO: 205)	PGDPTSNATSRGAEGPLPPP (SEQ ID NO: 206)
AARLEALKENGGARLAEYHA (SEQ ID NO: 207)	GTQDRQCNKTSNGSDSCDLM (SEQ ID NO: 208)	EFSNDIKKRKVDDKDSSHYD (SEQ ID NO: 209)

SLAGRAPAGAAAVSASGHGR (SEQ ID NO: 210)	IAPVVSSRVSPGTRPTAMGS (SEQ ID NO: 211)	LREGNTVRSFRAAAAETLER (SEQ ID NO: 212)
VEGKVQGLVPAGGSSSNSTR (SEQ ID NO: 213)	PAGGSSSNSTREPPASGRVA (SEQ ID NO: 214)	AKSENLVKSHVQLQNYSKVE (SEQ ID NO: 215)
STSTMPNSSQSSHSSSGTLS (SEQ ID NO: 216)	GPEFSNDIKKRKVDDKDSSH (SEQ ID NO: 217)	RMDIDVVEKTNKIKDRYWDS (SEQ ID NO: 218)
PKPTSNVHVSVAGSHSQAAN (SEQ ID NO: 219)	QQPYITNRTFMLAKEASLAD (SEQ ID NO: 220)	RALTQALPAPTGQNTQVSSH (SEQ ID NO: 221)
RDSPHSERYVSALTTPARLS (SEQ ID NO: 222)	EAVLALYNSTRDRVAGESAE (SEQ ID NO: 223)	PSTDAVLGNSTAVNRGDSSI (SEQ ID NO: 224)
LAGQRQGTQAESNLSSKFQF (SEQ ID NO: 225)	KLTDMEAQLLNQTSRMDAQM (SEQ ID NO: 226)	TITTNEIRNKLSYSFDGAGG (SEQ ID NO: 227)
RGGSQSVYGVDQTHSGLPLH (SEQ ID NO: 228)	GQQYPGQGPPSGQPPYGGHQ (SEQ ID NO: 229)	LALAPNSTARRQLASLLEGT (SEQ ID NO: 230)
IVGDAGVRGTAQQPRLFVGM (SEQ ID NO: 231)	AEILELAGNASRDNKKTRII (SEQ ID NO: 232)	GDADNDMVKIENLTKVYKSR (SEQ ID NO: 233)
KTPLTANITKSGNTFRPEVH (SEQ ID NO: 234)	VTWLVRHTENVTKTKTPEDY (SEQ ID NO: 235)	PGTQGRECNRTSEGADGCNL (SEQ ID NO: 236)
KTPLTANITKSGNTFRPEVH (SEQ ID NO: 234)	RVTNNVKDVTKLVANLPKDY (SEQ ID NO: 237)	SMGLQLNLTYERKDNTTVTR (SEQ ID NO: 238)
SFVIRGDEVRQIAPGQTGKI (SEQ ID NO: 239)	VGTQGRLCNRTSPGADGCDT (SEQ ID NO: 240)	YISNRVRVSIQTGTAFQPDA (SEQ ID NO: 241)
GRNVQEERSAGATANLPLRS (SEQ ID NO: 242)	YLTAEILELAGNASRDNKKT (SEQ ID NO: 243)	SSSNSTREPPASGRVALVKV (SEQ ID NO: 244)
IGNVGAPGAKGARGSAGPPG (SEQ ID NO: 245)	IRKLTDMEAQLLNQTSRMDA (SEQ ID NO: 246)	DMIQTLTNKSKIQNVGGTYE (SEQ ID NO: 247)
LLQGAGNSAAGRGAIADTFR (SEQ ID NO: 248)	SLGTAGRVCNLTSRGMDSCE (SEQ ID NO: 249)	VLALYNSTRDRVAGESAEPE (SEQ ID NO: 250)
EPYADQLRTQVNTQAEQLRR (SEQ ID NO: 251)	HPELKTPLTANITKSGNTFR (SEQ ID NO: 252)	YLGTSLGNITRDQKILNPSA (SEQ ID NO: 253)
LSVGRGGSPRLQPDSRALHY (SEQ ID NO: 254)	NSTWDPQSNSIAKYNYPNGK (SEQ ID NO: 255)	DVLYAGHNSTWDPQSNSIAK (SEQ ID NO: 256)
AAGSTMGAASMTLTVQARQL (SEQ ID NO: 257)	VKIQRVNNSSTNDNLVRKND (SEQ ID NO: 258)	LTVTSGKSIDDNSTKSEKQT (SEQ ID NO: 259)
AAMGAASLTLSAQSRTLLAG (SEQ ID NO: 260)	QAAELQNDVRQAVEGAIGRV (SEQ ID NO: 261)	NNRGKANSTGTLVITDPTRI (SEQ ID NO: 262)
GATNANARAKGGTGQTGGLT (SEQ ID NO: 263)	VQGLVPAGGSSSNSTREPPA (SEQ ID NO: 264)	RWNRQSSTSVIETNKTSVEL (SEQ ID NO: 265)
VLVEAAASQRSSSLASNRLK (SEQ ID NO: 266)	LDLNMTRSHSGGELESLARG (SEQ ID NO: 267)	KSALVLSAESSSMGGTLTVQ (SEQ ID NO: 268)
RRAAPEASGTPSSDAVSRLE (SEQ ID NO: 269)	WKAAIPSALDTNSSKSTSSF (SEQ ID NO: 270)	RITMTNSVDDALINSTKIYS (SEQ ID NO: 271)
VVSSRVSPGTRPTAMGSFSS (SEQ ID NO: 272)	EENKIRQEDLTKIISSAQKV (SEQ ID NO: 273)	IPSALDTNSSKSTSSFPCPA (SEQ ID NO: 274)
TVSGKITHSPQAHVNPQEKR (SEQ ID NO: 275)	LLLAGDRASSNPNATSSSSQ (SEQ ID NO: 276)	EISSDSIGVLGYQKTVDHTK (SEQ ID NO: 277)
FKGVDNEDDHGRGDLGRLKK (SEQ ID NO: 278)	VLNSNAIKNLPPPLGGAAGH (SEQ ID NO: 279)	QDNKIVVKNATKSNVNNAVN (SEQ ID NO: 280)
LTGRQDTSRMSTSQIPGRVA (SEQ ID NO: 281)	PFNHIKSSLLGSTSDSNLNK (SEQ ID NO: 282)	LPATLPQCQAANKTCPTNYM (SEQ ID NO: 283)

SRMSTSQIPGRVASSGLQSV (SEQ ID NO: 284)	SAISNLIRKHVSEARLVEDI (SEQ ID NO: 285)	NRTSGLLETNFTASARTTGS (SEQ ID NO: 286)
GRQDTSRMSTSQIPGRVASS (SEQ ID NO: 287)	QRVNNSSTNDNLVRKNDQVY (SEQ ID NO: 288)	VLFGTTISNNIKYGRDDVTD (SEQ ID NO: 289)
LGGTVTNPGGSGTSSRLEPR (SEQ ID NO: 290)	TAAHPELKTPLTANITKSGN (SEQ ID NO: 291)	GKSIDDNSTKSEKQTPREHS (SEQ ID NO: 292)
ARLEALKENGGARLAEYHAK (SEQ ID NO: 293)	LKASSTMAKSENLVKSHVQL (SEQ ID NO: 294)	VVLANASRIHYDGGAGTNTV (SEQ ID NO: 295)
ATAGSAMGAASLTLSAQSRT (SEQ ID NO: 296)	ELKTPLTANITKSGNTFRPE (SEQ ID NO: 297)	FATDRTSMPSANIIRSFKNE (SEQ ID NO: 298)
PLPSVTKLTEDRWNEHQKMQ (SEQ ID NO: 299)	YSLKASSTMAKSENLVKSHV (SEQ ID NO: 300)	TSGKSIDDNSTKSEKQTPRE (SEQ ID NO: 301)
STYMAASGLNASTYDQVGRS (SEQ ID NO: 302)	IGVLGYQKTVDHTKVNSKLS (SEQ ID NO: 303)	KEGGPDHNYTKEKIKIVEGI (SEQ ID NO: 304)
SYSQMQSQAGGNNTGSTPLR (SEQ ID NO: 305)	SSMSNHTKERVTMTKVTLEN (SEQ ID NO: 306)	KPCAPGTFSNTTSSTDICRP (SEQ ID NO: 307)
AATLEQPASSSHAQGTHLRL (SEQ ID NO: 308)	NDWKHNITQREARALELASQ (SEQ ID NO: 309)	SMSNHTKERVTMTKVTLENF (SEQ ID NO: 310)
RQAVEGAIGRVADTIRSGPS (SEQ ID NO: 311)	LLAPLNDTRVVHAAKAALAA (SEQ ID NO: 312)	RSGASASGVYTIQVSNATKP (SEQ ID NO: 313)
AMAGSAMGATSLTLSAQSRT (SEQ ID NO: 314)	EILELAGNAARDNKKTRIIP (SEQ ID NO: 315)	ISNRVRVSIQTGTAFQPDAA (SEQ ID NO: 316)
LKEGTICKRARGDDMDDYCN (SEQ ID NO: 317)	LVEVGGSAKYLNNTKKYQNQ (SEQ ID NO: 318)	AAHPELKTPLTANITKSGNT (SEQ ID NO: 319)
DLTGRQDTSRMSTSQIPGRV (SEQ ID NO: 320)	QSAVEILLTIDDTKRAGMKE (SEQ ID NO: 321)	GLVEVGGSAKYLNNTKKYQN (SEQ ID NO: 322)
GSTYMAASGLNASTYDQVGR (SEQ ID NO: 323)	VGYLDRNLEKNSTKQEILAA (SEQ ID NO: 324)	TPGLRSLENDGTKLLDRIRD (SEQ ID NO: 325)
VTVSGKITHSPQAHVNPQEK (SEQ ID NO: 326)	ATLNSVLNSNAIKNLPPPLG (SEQ ID NO: 327)	QAQKDKMFSQNDTRCAGPEA (SEQ ID NO: 328)
PGVGLTPSAAQTARQHPKMH (SEQ ID NO: 329)	FDNSTQMNMIRDTLGNPTVK (SEQ ID NO: 330)	EYPERPASRANTVRSFHHEE (SEQ ID NO: 331)
SVGTQGRACNKTAPQASGCD (SEQ ID NO: 332)	DQEVINKASNSNKIRLEKGR (SEQ ID NO: 333)	PLLAPLNDTRVVHAAKAALA (SEQ ID NO: 334)
QNGKSYSLKASSTMAKSENL (SEQ ID NO: 335)	LVQIRSNFSKPDLMASLLPQ (SEQ ID NO: 336)	LQSAKDNNIKFRAIGHSDNS (SEQ ID NO: 337)
TGGSGGFHLAQQASASGSVS (SEQ ID NO: 338)	LNNAILRANLPKSKGNPAAY (SEQ ID NO: 339)	TPLTANITKSGNTFRPEVHL (SEQ ID NO: 340)
LTQGDGVAKLSINTHPSQKP (SEQ ID NO: 341)	ATLGADLVDSTDNTKINRGI (SEQ ID NO: 342)	TMWVDDQEVINKASNSNKIR (SEQ ID NO: 343)
RSLAGRAPAGAAAVSASGHG (SEQ ID NO: 344)	KDMIQTLTNKSKIQNVGGTY (SEQ ID NO: 345)	VVPEGIRMNKTVAVRTLDPE (SEQ ID NO: 346)
ATAGSAMGAASLTVSAQSRT (SEQ ID NO: 347)	SGLLETNFTASARTTGSGLL (SEQ ID NO: 348)	RVNNSSTNDNLVRKNDQVYI (SEQ ID NO: 349)
NGGARLAEYHAKATEHLSTL (SEQ ID NO: 350)	YTGHDTKLMQNSTKAPLKRS (SEQ ID NO: 351)	NQDNKIVVKNATKSNVNNAV (SEQ ID NO: 352)
HQASLSKQPTSQPRGDPTGP (SEQ ID NO: 353)	FRIHVENQTRARDDVSRKQL (SEQ ID NO: 354)	DDVSNQTSCRLAGLKPGTVY (SEQ ID NO: 355)
LVEAAASQRSSSLASNRLKI (SEQ ID NO: 356)	SLPATLPQCQAANKTCPTNY (SEQ ID NO: 357)	LVKSHVQLQNYSKVERHPVT (SEQ ID NO: 358)

AAPMKEANIRGQGGLAYPGV (SEQ ID NO: 359)	ELENNKTMNRAENGGRPPHH (SEQ ID NO: 360)	GKSVSGDVELTNIIKNSSFK (SEQ ID NO: 361)
FGPVLDKATRESAFVHAIAS (SEQ ID NO: 362)	PPGQKGQANATRIPAKTPPA (SEQ ID NO: 363)	VDDQEVINKASNSNKIRLEK (SEQ ID NO: 364)
SATYDDEHHSTLVSTRHYRA (SEQ ID NO: 365)	GLSTINQTRLDFHFSSDRTA (SEQ ID NO: 366)	CSSMSNHTKERVTMTKVTLE (SEQ ID NO: 367)
GAPGTPGPQGIAGQRGVVGL (SEQ ID NO: 368)	RNGPEFSNDIKKRKVDDKDS (SEQ ID NO: 369)	MFSQNDTRCAGPEAVKGQTL (SEQ ID NO: 370)
SGPSVLGTGTGGSGGFHLAQ (SEQ ID NO: 371)	VRHTENVTKTKTPEDYADSS (SEQ ID NO: 372)	SGKSIDDNSTKSEKQTPREH (SEQ ID NO: 373)
GVAGNIVRHIPDRGAPTRAS (SEQ ID NO: 374)	VKDRTHNVTEKVTQVLSLGA (SEQ ID NO: 375)	LISNLSNTIEKINTLGLDSK (SEQ ID NO: 376)
KPTSNVHVSVAGSHSQAANF (SEQ ID NO: 377)	NRQSSTSVIETNKTSVELSL (SEQ ID NO: 378)	SIPTPSNRVKTQQKSNLELL (SEQ ID NO: 379)
TSNVHVSVAGSHSQAANFAA (SEQ ID NO: 380)	SVLNSNAIKNLPPPLGGAAG (SEQ ID NO: 381)	LLTFSDVTSNQSSAVLGKSI (SEQ ID NO: 382)
GPMGPRGAPGERGRPGLPGA (SEQ ID NO: 383)	DKRRNGPEFSNDIKKRKVDD (SEQ ID NO: 384)	YGERSLLQTNQISNFIDRKN (SEQ ID NO: 385)
KTIVKAGPEYGQGMNPISRL (SEQ ID NO: 386)	GTQGRLCNRTSPGADGCDTM (SEQ ID NO: 387)	AVLALYNSTRDRVAGESAEP (SEQ ID NO: 388)
GFENEESDYHTLPRARIMQR (SEQ ID NO: 389)	RQEATNATRVIHDNYDQLEV (SEQ ID NO: 390)	VPAGGSSSNSTREPPASGRV (SEQ ID NO: 391)
RQDTSRMSTSQIPGRVASSG (SEQ ID NO: 392)	IQTLTNKSKIQNVGGTYELH (SEQ ID NO: 393)	DTKLMQNSTKAPLKRSNVEK (SEQ ID NO: 394)
TGRQDTSRMSTSQIPGRVAS (SEQ ID NO: 395)	KIQRVNNSSTNDNLVRKNDQ (SEQ ID NO: 396)	QRITMTNSVDDALINSTKIY (SEQ ID NO: 397)
DTQRRTSMGGTQQQFVEGVR (SEQ ID NO: 398)	WVDDQEVINKASNSNKIRLE (SEQ ID NO: 399)	SSDSIGVLGYQKTVDHTKVN (SEQ ID NO: 400)
REALMASGSGGHNTMPSGGN (SEQ ID NO: 401)	NAIKNLPPPLGGAAGHPGSA (SEQ ID NO: 402)	PLNDTRVVHAAKAALAAFNA (SEQ ID NO: 403)
LSRVNAAVANRDSANGFINK (SEQ ID NO: 404)	SHVQLQNYSKVERHPVTALE (SEQ ID NO: 405)	QGLVPAGGSSSNSTREPPAS (SEQ ID NO: 406)
GPIGNVGAPGAKGARGSAGP (SEQ ID NO: 407)	LLASLQKSPEDEEKNFDQTR (SEQ ID NO: 408)	SVGTQGRLCNRTSPGADGCD (SEQ ID NO: 409)
NPRGSSSRQSSSSAMSSSSA (SEQ ID NO: 410)	NGPEFSNDIKKRKVDDKDSS (SEQ ID NO: 411)	KNNKIAGVNNRTEYVETTST (SEQ ID NO: 412)
QAKFVSNQTELDREVLVGKV (SEQ ID NO: 413)	LVGYLDRNLEKNSTKQEILA (SEQ ID NO: 414)	WLVRHTENVTKTKTPEDYAD (SEQ ID NO: 415)
PGAKGANGAPGIAGAPGFPG (SEQ ID NO: 416)	RGDADNDMVKIENLTKVYKS (SEQ ID NO: 417)	TIFDTVNGKLVKKVNLDTTH (SEQ ID NO: 418)
GSSSRQSSSSAMSSSSASSS (SEQ ID NO: 419)	SALDTNSSKSTSSFPCPAGH (SEQ ID NO: 420)	SFNSHINASNNSEPSRMNMT (SEQ ID NO: 421)
FVIRGDEVRQIAPGQTGKIA (SEQ ID NO: 422)	CAPGTFSNTTSSTDICRPHQ (SEQ ID NO: 423)	ISNRVGVSIRTGTAFQPDPA (SEQ ID NO: 424)
GARGSAGPPGATGFPGAAGR (SEQ ID NO: 425)	QTLTNKSKIQNVGGTYELHV (SEQ ID NO: 426)	ANEKHMLALAPNSTARRQLA (SEQ ID NO: 427)
GGGATNANARAKGGTGQTGG (SEQ ID NO: 428)	VVDDVSNQTSCRLAGLKPGT (SEQ ID NO: 429)	ISSDSNNLKLNNVRLPRENM (SEQ ID NO: 430)
TVLVEAAASQRSSSLASNRL (SEQ ID NO: 431)	HGTQDRQCNKTSNGSDSCDL (SEQ ID NO: 432)	DSIGVLGYQKTVDHTKVNSK (SEQ ID NO: 433)

GRGSAGGNAHSPLGVPGGGL (SEQ ID NO: 434)	FGTRDRTCNVTSHGIDGCDL (SEQ ID NO: 435)	IQRVNNSSTNDNLVRKNDQV (SEQ ID NO: 436)
RLVTTGVLKQTKGVGASGSF (SEQ ID NO: 437)	SESIPTPSNRVKTQQKSNLE (SEQ ID NO: 438)	VHAKSSDWGLQAGYIPVNPN (SEQ ID NO: 439)
DINGFTTGRRGDLATIHGMN (SEQ ID NO: 440)	GSVGTQGRLCNRTSPGADGC (SEQ ID NO: 441)	LRGDADNDMVKIENLTKVYK (SEQ ID NO: 442)
QGLSNPSTLSAQPQLSRSHS (SEQ ID NO: 443)	SAGTGPSSATVNVTTRKPPP (SEQ ID NO: 444)	GHDTKLMQNSTKAPLKRSNV (SEQ ID NO: 445)
GARGPPGPAGANGAPGLRGG (SEQ ID NO: 446)	QTIPQISNRTLNTLVQDNSY (SEQ ID NO: 447)	LINDYVKNGTRGKITDLIKD (SEQ ID NO: 448)
QAVPTGSFMMVNSSGRASGQ (SEQ ID NO: 449)	HCVVADGNSTRSPETNGLLC (SEQ ID NO: 450)	STQNKRANRTEENVSDGSPN (SEQ ID NO: 451)
GPAGKEGGKGPRGETGPAGR (SEQ ID NO: 452)	DDQEVINKASNSNKIRLEKG (SEQ ID NO: 453)	VLVSRGGSQSVYGVDQTHSG (SEQ ID NO: 454)
QMGPRGLPGERGRPGAPGPA (SEQ ID NO: 455)	HNPRSSLVLAATSGNSGTT (SEQ ID NO: 456)	ERAASRANTVRSFHHEESLE (SEQ ID NO: 457)
HVPEGQTIKGRVGTVGYMAP (SEQ ID NO: 458)	AAIPSALDTNSSKSTSSFPC (SEQ ID NO: 459)	SLNLSGVPSQDKTRVEPPQY (SEQ ID NO: 460)
KFKGVDNEDDHGRGDLGRLK (SEQ ID NO: 461)	RAASRANTVRSFHHEESLEE (SEQ ID NO: 462)	LDRNLEKNSTKQEILAALEK (SEQ ID NO: 463)
YKAVGSLDPSADLSSQRGKV (SEQ ID NO: 464)	HSDAVFTDNYTALRKQMAVK (SEQ ID NO: 465)	TATLGADLVDSTDNTKINRG (SEQ ID NO: 466)
RSADSHGHPRENRTIGATMV (SEQ ID NO: 467)	CQDCPLLAPLNDTRVVHAAK (SEQ ID NO: 468)	KAAIPSALDTNSSKSTSSFP (SEQ ID NO: 469)
TLRPGEVTALVGPNGSGKST (SEQ ID NO: 470)	PPELENNKTMNRAENGGRPP (SEQ ID NO: 471)	NVDFRIHVENQTRARDDVSR (SEQ ID NO: 472)
RAHQNSQTHQASLSKQPTSQ (SEQ ID NO: 473)	VQLQNYSKVERHPVTALEKM (SEQ ID NO: 474)	VKSHVQLQNYSKVERHPVTA (SEQ ID NO: 475)
AMGAGVAGRITGSMSLASGK (SEQ ID NO: 476)	TLNSVLNSNAIKNLPPPLGG (SEQ ID NO: 477)	GLVPAGGSSSNSTREPPASG (SEQ ID NO: 478)
RPRGGGAGGSSVGTGGGGTG (SEQ ID NO: 479)	GITAFNHPLNLTKQQLSEVA (SEQ ID NO: 480)	DFDVIRNVIKDNATFSRKNP (SEQ ID NO: 481)
TPKGQLPGGKAPPKAGSVPS (SEQ ID NO: 482)	KSHVQLQNYSKVERHPVTAL (SEQ ID NO: 483)	YPERPASRANTVRSFHHEEH (SEQ ID NO: 484)
AGPAQTLIRPQDMKGASRSP (SEQ ID NO: 485)	SGASASGVYTIQVSNATKPR (SEQ ID NO: 486)	IVVYTGHDTKLMQNSTKAPL (SEQ ID NO: 487)
AVHVPEGQTIKGRVGTVGYM (SEQ ID NO: 488)	DVWGKGTTVTVSSASPTSPK (SEQ ID NO: 489)	RKNGPIDISNTDNFVRKNDL (SEQ ID NO: 490)
SKKQLSLKLSSVNAADTAVY (SEQ ID NO: 491)	NRVTNNVKDVTKLVANLPKD (SEQ ID NO: 492)	RRNGPEFSNDIKKRKVDDKD (SEQ ID NO: 493)
GFPGPKGNDGAPGKNGERGG (SEQ ID NO: 494)	LRGDFSSANNRDNTYNRVSE (SEQ ID NO: 495)	LPGMTNQDNKIVVKNATKSN (SEQ ID NO: 496)
KPGSSGPGSTGSWNSGSSGT (SEQ ID NO: 497)	LNSNAIKNLPPPLGGAAGHP (SEQ ID NO: 498)	NSQTIPQISNRTLNTLVQDN (SEQ ID NO: 499)
GSGPSVLGTGTGGSGGFHLA (SEQ ID NO: 500)	LLQSAKDNNIKFRAIGHSDN (SEQ ID NO: 501)	LNLSGVPSQDKTRVEPPQYM (SEQ ID NO: 502)
ALKENGGARLAEYHAKATEH (SEQ ID NO: 503)	ERPASRANTVRSFHHEEHLE (SEQ ID NO: 504)	PGMTNQDNKIVVKNATKSNV (SEQ ID NO: 505)
GVQLGKRYNGSDPASGPSVQ (SEQ ID NO: 506)	GSLGTAGRVCNLTSRGMDSC (SEQ ID NO: 507)	GYLDRNLEKNSTKQEILAAL (SEQ ID NO: 508)

GIEQGVGVPGGSPGSRHRQK (SEQ ID NO: 509)	LVPAGGSSSNSTREPPASGR (SEQ ID NO: 510)	ESIPTPSNRVKTQQKSNLEL (SEQ ID NO: 511)
LKTPLTANITKSGNTFRPEV (SEQ ID NO: 512)	RFDVSAMEKNASNLVKAEFR (SEQ ID NO: 513)	VYLGTSLGNITRDQKILNPS (SEQ ID NO: 514)
LKTPLTANITKSGNTFRPEV (SEQ ID NO: 512)	KYIDQEELNKTKPIWTRNPD (SEQ ID NO: 515)	EILELAGNAARDNKKTRIIP (SEQ ID NO: 315)
QRIAVETAATARKSSPPASS (SEQ ID NO: 516)	YISNRVGVSIRTGTAFQPDP (SEQ ID NO: 517)	HVQLQNYSKVERHPVTALEK (SEQ ID NO: 518)
AAGVAHAITAACTQGNLSDC (SEQ ID NO: 519)	SFGTRDRTCNVTSHGIDGCD (SEQ ID NO: 520)	TAFNHPLNLTKQQLSEVALM (SEQ ID NO: 521)
LEYPERPASRANTVRSFHHE (SEQ ID NO: 522)	QSSTSVIETNKTSVELSLPF (SEQ ID NO: 523)	MLALAPNSTARRQLASLLEG (SEQ ID NO: 524)
ETGLSTINQTRLDFHFSSDR (SEQ ID NO: 525)	AGGSSSNSTREPPASGRVAL (SEQ ID NO: 526)	KDNNIKFRAIGHSDNSVPPF (SEQ ID NO: 527)
GTFSNTTSSTDICRPHQICN (SEQ ID NO: 528)	RDLFFNEIKARSGQDDKTVY (SEQ ID NO: 529)	

[0091] Shown in **Table 4A-B** are the top performing peptides across all of these described screens. These include the mutant versions of the peptides that were originally identified in the primary screen. The information includes the amino acid sequence of the peptide, the AAV scaffold in which that peptide was inserted, the organ where that peptide was a hit, the log2FC in that organ, and the screen in which that peptide was identified. While the peptides shown here were highlighted for their ability to transduce a particular organ, many of these also display other interesting properties such as their increased ability to form functional capsids (high titers), their significant de-targeting away from the liver, and their ability to target unique combinations of organs simultaneously. It is expected that these peptides and mutants thereof to have the ability to modulate tropism of AAV capsids, such as AAV5, AAV9, and the 23 immune orthogonal capsids (shown in Table 5) and beyond as well as in tropism of other delivery vehicles and peptide-payload conjugates.

[0092] **Table 4A**: Table for mutant peptides that were identified as the top hits in the primary, secondary, tertiary, and NHP screens. Shown are the peptide amino acid sequences, the scaffold where they were grafted, the log2FC in the organ they were a hit, the organ they were a hit in, and the screen in which that log2FC was observed. (P=pancreas; H=heart; B=Brain; L=Lung; Lv=liver; K=kidney; M=muscle; S=spleen; I=intestine):

Amino Acid Sequence	Scaffold	Log2FC	Organ	Screen Identified In	Amino Acid Sequence	Scaffold	Log2FC	Organ	Screen Identified In
YSIGVLGYQK TVDHTKVNSK (SEQ ID NO: 530)	AAV9 Loop2	12.88569	H	3°	GFENEESDYHTL PRAGIMQK (SEQ ID NO: 531)	AAV9 Loop1	16.11414	P	3°
YSIGVLGYQK TEDHTKVNSK (SEQ ID NO: 532)	AAV9 Loop2	12.68698	H	3°	YSIGVLGYQKTE DHTKVNSK (SEQ ID NO: 532)	AAV9 Loop2	14.12717	P	3°
GFENEESDYH TLPRAGIMQK (SEQ ID NO: 531)	AAV9 Loop1	11.76037	H	3°	YSIGVLGYQKT VDHTKVNSK (SEQ ID NO: 530)	AAV9 Loop2	13.93828	P	3°
YSIGVLGNQK TVDHTKVNSK (SEQ ID NO: 533)	AAV9 Loop2	11.64864	H	3°	YSIGVLGNQKT VDHTKVNSK (SEQ ID NO: 533)	AAV9 Loop2	13.13025	P	3°
DAPLGTGANK DLELPWLEQQ (SEQ ID NO: 534)	AAV9 Loop2	11.28843	H	3°	DAPLGTGANKD LELPWLEQQ (SEQ ID NO: 534)	AAV9 Loop2	12.72721	P	3°
YSIGVLSYQKT VDHTKVNSK (SEQ ID NO: 535)	AAV9 Loop2	10.70343	H	3°	YSIGVLGYQKT VDHTKVNGK (SEQ ID NO: 536)	AAV9 Loop2	11.96429	P	3°
YSIGVLVYQK TVDHTKVNSK (SEQ ID NO: 537)	AAV9 Loop2	10.63466	H	3°	YSIGVLSYQKTV DHTKVNSK (SEQ ID NO: 535)	AAV9 Loop2	11.86451	P	3°
YSIGVLGYQK TVDHTKVNGK (SEQ ID NO: 536)	AAV9 Loop2	10.60146	H	3°	YSIGVLVYQKT VDHTKVNSK (SEQ ID NO: 537)	AAV9 Loop2	11.61047	P	3°
FTGEPSLLPVS PVGHLHASL (SEQ ID NO: 538)	AAV9 Loop2	10.28004	H	2°	YSIGVQGYQKT VDHTKVNSK (SEQ ID NO: 539)	AAV9 Loop2	11.05567	P	3°
YSIGVQGYQK TVDHTKVNSK (SEQ ID NO: 539)	AAV9 Loop2	9.748812	H	3°	LGKGTGGSGGF HRAQQASGS (SEQ ID NO: 540)	AAV9 Loop1	11.0131	P	3°
KKDRVTEALN ATRAAVEEGF (SEQ ID NO: 541)	AAV9 Loop2	9.689435	H	2°	HDTNLMQNSTK AHLKRSNVE (SEQ ID NO: 542)	AAV9 Loop2	10.80511	P	3°

HDTNLMQNSTKAHLKRSNVE (SEQ ID NO: 542)	AAV9 Loop2	9.358717	H	3°	QDNTCEEYHTMGPHSSHIPK (SEQ ID NO: 543)	AAV9 Loop1	10.63549	P	3°
LNSNAIKNLPTPLGGAAGHQ (SEQ ID NO: 544)	AAV9 Loop2	9.262807	H	2°	FSIGVLGYQKTVDHTKVNSK (SEQ ID NO: 545)	AAV9 Loop2	10.49796	P	3°
DKRRNGPEFSNYIKKRKVDD (SEQ ID NO: 546)	AAV9 Loop2	9.215325	H	2°	YSIGVLGYQKSVDHTKVNSK (SEQ ID NO: 547)	AAV9 Loop2	10.26756	P	3°
FSIGVLGYQKTVDHTKVNSK (SEQ ID NO: 545)	AAV9 Loop2	9.095547	H	3°	YSIGVLGYQKTVDHTKVNST (SEQ ID NO: 548)	AAV9 Loop2	10.06449	P	3°
WGQGSTVAASSGSASAPTLF (SEQ ID NO: 549)	AAV9 Loop1	9.070575	H	2°	LSATQSVVSHIGSKRVEGIV (SEQ ID NO: 550)	AAV9 Loop1	9.54629	P	2°
FGPSAALDSQSGLLKGQEPW (SEQ ID NO: 551)	AAV9 Loop2	8.9294	H	2°	SDSICVLGYQKTVDHTKVNS (SEQ ID NO: 552)	AAV9 Loop2	9.334344	P	2°
YSIGVLGYQKSVDHTKVNSK (SEQ ID NO: 547)	AAV9 Loop2	8.773237	H	3°	AGGSSSNSTREPPASGLVAF (SEQ ID NO: 553)	AAV9 Loop2	9.009486	P	2°
YKRRNGPEFSNDIKKRKVDD (SEQ ID NO: 554)	AAV9 Loop2	8.732732	H	2°	LNSNAIKNLPTPLGGAAGHQ (SEQ ID NO: 544)	AAV9 Loop2	8.997317	P	2°
YSIGVLGYQKTVDHTKVNST (SEQ ID NO: 548)	AAV9 Loop2	8.694392	H	3°	YSIGVLGNQKTVDHTKENSK (SEQ ID NO: 555)	AAV9 Loop2	8.819996	P	3°
DINGLTTGRRGDMAAIHGMN (SEQ ID NO: 556)	AAV9 Loop1	8.657613	H	2°	NVDFSINVENQTRARDDVSR (SEQ ID NO: 557)	AAV9 Loop2	8.789112	P	2°
AGGSSSNSTREPPASGLVAF (SEQ ID NO: 553)	AAV9 Loop2	8.589748	H	2°	DSIGVLGYKKTVDNTKVNSK (SEQ ID NO: 558)	AAV9 Loop2	8.735838	P	2°
VLNSNAIKNLHQPLGGAAGH (SEQ ID NO: 559)	AAV9 Loop2	8.588456	H	2°	ETGLSAINQTRLDFNNSSDR (SEQ ID NO: 560)	AAV9 Loop2	8.688695	P	2°

YQEVINKASN SNKIRLEKGR (SEQ ID NO: 561)	AAV9 Loop2	8.463129	H	2°	ISSDSICALGYQ KTVDHTKV (SEQ ID NO: 562)	AAV9 Loop2	8.609108	P	2°
VQAGGSSSNS TREQPASGRV (SEQ ID NO: 563)	AAV9 Loop2	8.435494	H	2°	YSIGVLGYQKT VDHTKVNSK (SEQ ID NO: 530)	AAV9 Loop2	8.596609	P	2°
NVDFSINVEN QTRARDDVSR (SEQ ID NO: 557)	AAV9 Loop2	8.420162	H	2°	SSDSIGVHGYQK TVDHTKVN (SEQ ID NO: 564)	AAV9 Loop2	8.58793	P	2°
QVHPDSGISSE AMGIVNSFV (SEQ ID NO: 565)	AAV9 Loop1	8.418562	H	2°	SSTSTMPNSSRS SHSSSWTL (SEQ ID NO: 566)	AAV9 Loop2	8.409485	P	2°
VTAGGSSSNS TREQPASGRV (SEQ ID NO: 567)	AAV9 Loop2	8.371044	H	2°	TAAHPELKTQL TANITKSGN (SEQ ID NO: 568)	AAV9 Loop2	8.392862	P	2°
LVPACCSSSNS TREPPASGR (SEQ ID NO: 569)	AAV9 Loop2	8.321773	H	2°	LVPACCSSSNST REPPASGR (SEQ ID NO: 569)	AAV9 Loop2	8.382255	P	2°
VPAGGSSSNS TREQHASGRV (SEQ ID NO: 570)	AAV9 Loop2	8.313722	H	2°	SSTSTMPNSSRS SNSSSGTL (SEQ ID NO: 571)	AAV9 Loop2	8.336041	P	2°
LQALVREMSG SPASGIPAEV (SEQ ID NO: 572)	AAV9 Loop2	8.269379	H	2°	IARVADTISSGP VNSESFPT (SEQ ID NO: 573)	AAV9 Loop1	8.320148	P	2°
APPTELHRTR KEGVATYAAA (SEQ ID NO: 574)	AAV9 Loop1	8.226368	H	2°	MIQTLTNKSKIQ NVGWTYEF (SEQ ID NO: 575)	AAV9 Loop2	8.315111	P	2°
VLNSNAIKNLP QQLGGAAGH (SEQ ID NO: 576)	AAV9 Loop2	8.132094	H	2°	VQAGGSSSNST REQPASGRV (SEQ ID NO: 563)	AAV9 Loop2	8.297221	P	2°
DKEVINKASN SNKIRLEKGR (SEQ ID NO: 577)	AAV9 Loop2	8.128914	H	2°	TATMGADLVDS TDNTKINRG (SEQ ID NO: 578)	AAV9 Loop2	8.289843	P	2°
RTSMGGTQQQ FVKGARMEEI (SEQ ID NO: 579)	AAV9 Loop2	8.128729	H	2°	TATLGADLVDS TDNTKINRC (SEQ ID NO: 580)	AAV9 Loop2	8.269284	P	2°

LGGTVANPGG SGTSSRLEPK (SEQ ID NO: 581)	AAV9 Loop1	8.06852	H	2°	SSDSICVLGYQK TVDHTKVN (SEQ ID NO: 582)	AAV9 Loop2	8.247059	P	2°
ACGSSSNSTRE PPASGLVAL (SEQ ID NO: 583)	AAV9 Loop2	8.035724	H	2°	TLGADLVDSTD NTKINLGIF (SEQ ID NO: 584)	AAV9 Loop2	8.244473	P	2°
MIQTLTNKSKI QNVGWTYEF (SEQ ID NO: 575)	AAV9 Loop2	7.993694	H	2°	ISSDSIGVLGYK KTVDHTKV (SEQ ID NO: 585)	AAV9 Loop2	8.243375	P	2°
SVLNSNAIKNL PQQLGGAAG (SEQ ID NO: 586)	AAV9 Loop2	7.993393	H	2°	VTAGGSSSNSTR EQPASGRV (SEQ ID NO: 567)	AAV9 Loop2	8.237233	P	2°
LVPAGGSSSN STREQHASGR (SEQ ID NO: 587)	AAV9 Loop2	7.979569	H	2°	ENVDFRINVEN QTRARDDVS (SEQ ID NO: 588)	AAV9 Loop2	8.225842	P	2°
TAAHPELKTQ LTANITKSGN (SEQ ID NO: 568)	AAV9 Loop2	7.958291	H	2°	AKSAVEILLTID DTKRAGMK (SEQ ID NO: 589)	AAV9 Loop2	8.199508	P	2°
DFRIHVENQTL ALDDVSRKQ (SEQ ID NO: 590)	AAV9 Loop2	7.887135	H	2°	SDSIGVLCYQKT VDHTKVNS (SEQ ID NO: 591)	AAV9 Loop2	8.1967	P	2°
PACGSSSNSTR EPPASGLVA (SEQ ID NO: 592)	AAV9 Loop2	7.859241	H	2°	ISSDSIGALGYK KTVDHTKV (SEQ ID NO: 593)	AAV9 Loop2	8.187129	P	2°
RRVGDSSGAA AAHKVLPRGL (SEQ ID NO: 594)	AAV9 Loop1	7.852159	H	2°	DSIGVLGYQKT VDHTNVNSN (SEQ ID NO: 595)	AAV9 Loop2	8.183675	P	2°
PAGCSSSNSTR EPPASGLVA (SEQ ID NO: 596)	AAV9 Loop2	7.847743	H	2°	LASIQKSPEDEE KNYDQTRF (SEQ ID NO: 597)	AAV9 Loop2	8.182704	P	2°
VPAGGTSSNS TREPPASGLV (SEQ ID NO: 598)	AAV9 Loop2	7.841973	H	2°	LVPAGGSSSNST REQHASGR (SEQ ID NO: 587)	AAV9 Loop2	8.174143	P	2°
LGGTVTHPGG SGTSSRLEPK (SEQ ID NO: 599)	AAV9 Loop1	7.817083	H	2°	GSATIYWHDIRT EMDETRKN (SEQ ID NO: 600)	AAV9 Loop2	8.165904	P	2°

RRNGPEFSNYIKKRKVDDKD (SEQ ID NO: 601)	AAV9 Loop2	7.767658	H	2°	SDSIGVLGYQKTVDHTNVNS (SEQ ID NO: 602)	AAV9 Loop2	8.159857	P	2°
KAAIPSALDTNSSKSGSSFP (SEQ ID NO: 603)	AAV9 Loop2	7.767145	H	2°	DFRIHVENQTLALDDVSRKQ (SEQ ID NO: 590)	AAV9 Loop2	8.154733	P	2°
SVLNTNAIKNLPTPLGGAAG (SEQ ID NO: 604)	AAV9 Loop2	7.718368	H	2°	GFENEESDYHTLPRAGIMQK (SEQ ID NO: 531)	AAV9 Loop1	16.04491	K	3°
SRNGPEFSNDIKKRKVDDKD (SEQ ID NO: 605)	AAV9 Loop2	7.717071	H	2°	YSIGVLGYQKTEDHTKVNSK (SEQ ID NO: 532)	AAV9 Loop2	12.57863	K	3°
LPEAVLALHNSTRDLVAGES (SEQ ID NO: 606)	AAV9 Loop2	7.713016	H	2°	YSIGVLGYQKTVDHTKVNSK (SEQ ID NO: 530)	AAV9 Loop2	12.39794	K	3°
GSEGTQGRACNKTAPQASCC (SEQ ID NO: 607)	AAV9 Loop2	7.712568	H	2°	YSIGVLGNQKTVDHTKVNSK (SEQ ID NO: 533)	AAV9 Loop2	11.62058	K	3°
VLALNNSTSDRVAGESAEPE (SEQ ID NO: 608)	AAV9 Loop2	7.688014	H	2°	CKFLKEGTICKRARCDDMDD (SEQ ID NO: 609)	AAV9 Loop2	11.23581	K	2°
VQAGGSSSNSTRELPASGRV (SEQ ID NO: 610)	AAV9 Loop2	7.679988	H	2°	DAPLGTGANKDLELPWLEQQ (SEQ ID NO: 534)	AAV9 Loop2	11.05473	K	3°
LNTVLNSNAIKNLPQPLGGA (SEQ ID NO: 611)	AAV9 Loop2	7.659771	H	2°	YSIGVLSYQKTVDHTKVNSK (SEQ ID NO: 535)	AAV9 Loop2	10.72641	K	3°
VKGAGGESRVLPSGQGINNL (SEQ ID NO: 612)	AAV9 Loop2	7.633291	H	2°	YSIGVLGYQKTVDHTKVNGK (SEQ ID NO: 536)	AAV9 Loop2	10.58271	K	3°
QGLVQAGGSSSNSTRELPAS (SEQ ID NO: 613)	AAV9 Loop2	7.627424	H	2°	NRAVFKGGKDNIPVLKRDVA (SEQ ID NO: 614)	AAV9 Loop2	10.37732	K	2°
VLNSNAIKNLPPPLCCAAGH (SEQ ID NO: 615)	AAV9 Loop2	7.622054	H	2°	NRAVFKGCQDNIPVLKRDVA (SEQ ID NO: 616)	AAV9 Loop2	10.22739	K	2°

GDKANSSIEFH SKPQHQRDV (SEQ ID NO: 617)	AAV9 Loop1	7.610955	H	2°	YSIGVLVYQKT VDHTKVNSK (SEQ ID NO: 537)	AAV9 Loop2	10.19363	K	3°
LKTQLTANITK SGNTFRHEV (SEQ ID NO: 618)	AAV9 Loop2	7.592926	H	2°	VMLVLTNPEGR KEGKEVTFQ (SEQ ID NO: 619)	AAV9 Loop2	10.19295	K	2°
VPACGSSSNST REPPASGLV (SEQ ID NO: 620)	AAV9 Loop2	7.588375	H	2°	ARLVLTNHEGR KEGKEVTFQ (SEQ ID NO: 621)	AAV9 Loop2	10.14889	K	2°
RRNGPEFSNDI KKRKVDYKD (SEQ ID NO: 622)	AAV9 Loop2	7.586276	H	2°	NCKFLKEGTICK RARCDDMD (SEQ ID NO: 623)	AAV9 Loop2	10.14535	K	2°
VPAGGSSSNS TLEPPASGLV (SEQ ID NO: 624)	AAV9 Loop2	7.576296	H	2°	QDNTCEEYHTM GPHSSHIPK (SEQ ID NO: 543)	AAV9 Loop1	9.670684	K	3°
HVQLQNYSKV ERHHVTALEN (SEQ ID NO: 625)	AAV9 Loop2	7.555717	H	2°	DSIGVLGYQKT GDHTKVNSK (SEQ ID NO: 626)	AAV9 Loop2	9.528444	K	3°
GGLAYPGVRT HGTLESVNWP (SEQ ID NO: 627)	AAV9 Loop2	7.552054	H	2°	YSIGVQGYQKT VDHTKVNSK (SEQ ID NO: 539)	AAV9 Loop2	9.522257	K	3°
SVLNSNAIKNL PPPLGCAAW (SEQ ID NO: 628)	AAV9 Loop2	7.549378	H	2°	VRLVLTNHEGR KEGKEVSFQ (SEQ ID NO: 629)	AAV9 Loop2	9.485177	K	2°
LKTQLTANTT KSGNTFRPEV (SEQ ID NO: 630)	AAV9 Loop2	7.546106	H	2°	VRLVLTNPEGR KECKEVTFQ (SEQ ID NO: 631)	AAV9 Loop2	9.479958	K	2°
DINGLTTGRR GDMAAIHGM N (SEQ ID NO: 556)	AAV9 Loop1	18.96709	L	NHP	NCKFLKEGTICK RARGYDMD (SEQ ID NO: 632)	AAV9 Loop2	9.428763	K	2°
GFENEESDYH TLPRAGIMQK (SEQ ID NO: 531)	AAV9 Loop1	13.97277	L	3°	HDTNLMQNSTK AHLKRSNVE (SEQ ID NO: 542)	AAV9 Loop2	9.402727	K	3°
SMGTQGRAC NKTAPQASGC D (SEQ ID NO: 633)	AAV9 Loop1	12.78786	L	NHP	SESIPTLSNRVK TQQKSNLE (SEQ ID NO: 634)	AAV9 Loop2	9.379991	K	3°

QDNTCEEYHT MGPHSSHIPK (SEQ ID NO: 543)	AAV9 Loop1	12.39187	L	NHP	NCKFLKECTICK RARGDDMD (SEQ ID NO: 635)	AAV9 Loop2	9.329294	K	2°
KLARTIRTART QAEEREMIQ (SEQ ID NO: 636)	AAV9 Loop1	10.76914	L	2°	ARLVLTNPEGR KDGKEVTFQ (SEQ ID NO: 637)	AAV9 Loop2	9.299425	K	2°
MVNSSGRASG QMAHQLLPTL (SEQ ID NO: 638)	AAV9 Loop1	10.36477	L	2°	SNCRHEIDAGV VCTNETRST (SEQ ID NO: 639)	AAV9 Loop2	9.29233	K	2°
QDNTCEEYHT MGPHSSHIPK (SEQ ID NO: 543)	AAV9 Loop1	10.08961	L	3°	CKFLKEGTICKR ARGDDVDV (SEQ ID NO: 640)	AAV9 Loop2	9.1154	K	2°
LNSNAIKNLPT PLGGAAGHQ (SEQ ID NO: 544)	AAV9 Loop2	9.993713	L	2°	FSIGVLGYQKTV DHTKVNSK (SEQ ID NO: 545)	AAV9 Loop2	9.097607	K	3°
GLESEESEYHT LPRARIMQR (SEQ ID NO: 641)	AAV9 Loop1	9.917488	L	NHP	CRHERDAGVVC TNETRSTNT (SEQ ID NO: 642)	AAV9 Loop2	9.06325	K	2°
QDNTCEEYHT MGPHSNHIPK (SEQ ID NO: 643)	AAV9 Loop1	9.798377	L	3°	KLSSTSRLKSVG LNLEGNGE (SEQ ID NO: 644)	AAV9 Loop1	9.03007	K	2°
SSQAGVPVRV PRSPRAPESA (SEQ ID NO: 645)	AAV9 Loop1	9.748563	L	2°	VRLVLTNPEGR KEGKEVSFK (SEQ ID NO: 646)	AAV9 Loop2	8.979787	K	2°
AGGSSSNSTR EPPASGLVAF (SEQ ID NO: 553)	AAV9 Loop2	9.500178	L	2°	VRRVLTNPEGR KEGMEVTFQ (SEQ ID NO: 647)	AAV9 Loop2	8.966907	K	2°
LQSAKDNNIK FRAIGHSDSG (SEQ ID NO: 648)	AAV9 Loop1	9.472418	L	3°	NCKFLKEGTICK RARGDYMD (SEQ ID NO: 649)	AAV9 Loop2	8.957443	K	2°
QGISNSSTSSM PNSSQSSHS (SEQ ID NO: 650)	AAV9 Loop1	9.337745	L	2°	AGGSSSNSTREP PASGLVAF (SEQ ID NO: 553)	AAV9 Loop2	8.949626	K	2°
FLEVSEQILAK IPSGKNKQT (SEQ ID NO: 651)	AAV9 Loop1	9.324263	L	2°	ARLVLTNTEGR KEGKEVTFQ (SEQ ID NO: 652)	AAV9 Loop2	8.901878	K	2°

FRGVAGGTALSRSAVHESTG (SEQ ID NO: 653)	AAV9 Loop2	9.322438	L	NHP	NRAVFKGGQDNIHVLKRDVA (SEQ ID NO: 654)	AAV9 Loop2	8.878608	K	2°
NVDFSINVENQTRARDDVSR (SEQ ID NO: 557)	AAV9 Loop2	9.203686	L	2°	SNCRHERYAGVVCTNETRST (SEQ ID NO: 655)	AAV9 Loop2	8.839399	K	2°
VLNSNAIKNLHQPLGGAAGH (SEQ ID NO: 559)	AAV9 Loop2	9.100841	L	2°	YSIGVLGYQKSVDHTKVNSK (SEQ ID NO: 547)	AAV9 Loop2	8.809637	K	3°
GHDTKLMQNSTKAPLKSSTV (SEQ ID NO: 656)	AAV9 Loop1	9.066517	L	2°	DINGLTTGRRGDMAAIHGMN (SEQ ID NO: 556)	AAV9 Loop1	21.38375	B	NHP
LREIRSYQRTTEPLIRKLPF (SEQ ID NO: 657)	AAV9 Loop1	9.013501	L	2°	QDNTCEEYHTMGPHSSHIPK (SEQ ID NO: 543)	AAV9 Loop1	15.55224	B	NHP
MIQTLTNKSKIQNVGWTYEF (SEQ ID NO: 575)	AAV9 Loop2	8.963172	L	2°	GFENEESDYHTLPRAGIMQK (SEQ ID NO: 531)	AAV9 Loop1	14.22356	B	3°
DFRIHVENQTLALDDVSRKQ (SEQ ID NO: 590)	AAV9 Loop2	8.919031	L	2°	GLESEESEYHTLPRARIMQR (SEQ ID NO: 641)	AAV9 Loop1	12.48049	B	NHP
LVPACCSSSNSTREPPASGR (SEQ ID NO: 569)	AAV9 Loop2	8.830824	L	2°	QDDTCEEYHSAQPNSCHISK (SEQ ID NO: 658)	AAV9 Loop1	12.37509	B	2°
LLPSSTVTPRAPHLSHSSQP (SEQ ID NO: 659)	AAV9 Loop1	8.786784	L	2°	YSIGVLGYQKTEDHTKVNSK (SEQ ID NO: 532)	AAV9 Loop2	12.3493	B	3°
GHDTKLMQNSTKAPLNRSYV (SEQ ID NO: 660)	AAV9 Loop1	8.762197	L	2°	YSIGVLGYQKTVDHTKVNSK (SEQ ID NO: 530)	AAV9 Loop2	12.17751	B	3°
ELIRTIRSARTQAEERGVIQ (SEQ ID NO: 661)	AAV9 Loop1	8.745947	L	2°	WNGQDNTCEEYHSVGPHSCH (SEQ ID NO: 662)	AAV9 Loop1	11.82604	B	2°
VPAGGSSSNSTREQHASGRV (SEQ ID NO: 570)	AAV9 Loop2	8.731578	L	2°	GFENEESDYHTLPRAGIMQK (SEQ ID NO: 531)	AAV9 Loop1	11.59255	B	2°

DFSIHVENQTRARDDVSSKQ (SEQ ID NO: 663)	AAV9 Loop2	8.72787	L	2°	SSDNIGVLGYQKTVDHTKVK (SEQ ID NO: 664)	AAV9 Loop2	11.55392	B	NHP
PGVLPLLTVDPGSAKQRPPR (SEQ ID NO: 665)	AAV9 Loop1	8.712366	L	2°	NGGARLAEYHAKTTEHQSTL (SEQ ID NO: 666)	AAV9 Loop1	11.53923	B	2°
LVPAGGSSSNSTREQHASGR (SEQ ID NO: 587)	AAV9 Loop2	8.69303	L	2°	GLENEESDYHTLPRARIMQK (SEQ ID NO: 667)	AAV9 Loop1	11.42221	B	2°
EYPERQASRANTVRSFNHEE (SEQ ID NO: 668)	AAV9 Loop2	8.656612	L	2°	YSIGVLGNQKTVDHTKVNSK (SEQ ID NO: 533)	AAV9 Loop2	11.35652	B	3°
HQANLSKQPTSQPRGDPTGP (SEQ ID NO: 669)	AAV9 Loop1	8.654464	L	NHP	DARHIVTQSDDPHFHEKKQI (SEQ ID NO: 670)	AAV9 Loop1	11.27287	B	2°
VTAGGSSSNSTREQPASGRV (SEQ ID NO: 567)	AAV9 Loop2	8.626233	L	2°	VARHVGSQSDDPHFHEKKQI (SEQ ID NO: 671)	AAV9 Loop1	11.11897	B	2°
IRLGLSHKADMQTREVLMKL (SEQ ID NO: 672)	AAV9 Loop1	8.606854	L	2°	DAPLGTGANKDLELPWLEQQ (SEQ ID NO: 534)	AAV9 Loop2	11.04062	B	3°
VLNSNAIKNLPQQLGGAAGH (SEQ ID NO: 576)	AAV9 Loop2	8.595369	L	2°	KWVDDQEVINKASNSNKIRP (SEQ ID NO: 673)	AAV9 Loop2	11.02815	B	3°
VQAGGSSSNSTREQPASGRV (SEQ ID NO: 563)	AAV9 Loop2	8.594954	L	2°	GFENEESEYHTLPRARIMQK (SEQ ID NO: 674)	AAV9 Loop1	10.77684	B	2°
SVLNSNAIKNLPQQLGGAAG (SEQ ID NO: 586)	AAV9 Loop2	8.5523	L	2°	GVDNEDYHGRGDLGRLKTQK (SEQ ID NO: 675)	AAV9 Loop1	10.66715	B	2°
TAAHPELKTQLTANITKSGN (SEQ ID NO: 568)	AAV9 Loop2	8.493584	L	2°	SSDSIGVLGYQKTEDHTKVK (SEQ ID NO: 676)	AAV9 Loop2	10.45263	B	NHP
GMRLPRGAPVDIPSSDLTGR (SEQ ID NO: 677)	AAV9 Loop1	8.489128	L	2°	PAFRNEEYHKSVLENQATFF (SEQ ID NO: 678)	AAV9 Loop1	10.41258	B	2°

ACGSSSNSTRE PPASGLVAL (SEQ ID NO: 583)	AAV9 Loop2	8.4853	L	2°	YLESGSINEKDY HSRLNIDE (SEQ ID NO: 679)	AAV9 Loop1	10.40359	B	2°
VDFRIHVENQ TLARDDVSRN (SEQ ID NO: 680)	AAV9 Loop2	8.477405	L	2°	NSVLNSNAIKNL PQPLGGTA (SEQ ID NO: 681)	AAV9 Loop2	10.30484	B	NHP
LLRECNTVRS FLAAAAETLE (SEQ ID NO: 682)	AAV9 Loop2	8.444189	L	2°	GFENEESDYHTL PRARIVQK (SEQ ID NO: 683)	AAV9 Loop1	10.27759	B	2°
MIQTLTNKSKI QNVGVTYEF (SEQ ID NO: 684)	AAV9 Loop2	8.411206	L	2°	YSIGVLVYQKT VDHTKVNSK (SEQ ID NO: 537)	AAV9 Loop2	10.24646	B	3°
GHDTKLMQNS TKAPLKRPDV (SEQ ID NO: 685)	AAV9 Loop1	8.376499	L	2°	YSIGVLGYQKT VDHTKVNGK (SEQ ID NO: 536)	AAV9 Loop2	10.22341	B	3°
NRQSSTSVLET NKTSVVLSS (SEQ ID NO: 686)	AAV9 Loop2	8.337899	L	NHP	GDMANSSIEFHP KPRQQREV (SEQ ID NO: 687)	AAV9 Loop1	10.20711	B	2°
VVYTGRDTKI MQNSTKAPLK (SEQ ID NO: 688)	AAV9 Loop1	8.333714	L	2°	YSIGVLSYQKTV DHTKVNSK (SEQ ID NO: 535)	AAV9 Loop2	10.19875	B	3°
HVHEGQTVKG RVGTVGYLAP (SEQ ID NO: 689)	AAV9 Loop1	8.315208	L	2°	HKENGGARLVE YHAKASEHL (SEQ ID NO: 690)	AAV9 Loop1	10.17384	B	2°
LKTQLTANITK SGNTFRHEV (SEQ ID NO: 618)	AAV9 Loop2	8.303464	L	2°	GFENEESDYHA LPRARIMQK (SEQ ID NO: 691)	AAV9 Loop1	10.10022	B	2°
LVPKAAGNSN GGFVKPNPND (SEQ ID NO: 692)	AAV9 Loop1	8.297592	L	2°	SESIPTLSNRVK TQQKSNLE (SEQ ID NO: 634)	AAV9 Loop2	10.09714	B	3°
HDTNLMQNST KAHLKRSNVE (SEQ ID NO: 542)	AAV9 Loop2	8.29573	L	2°	RSADSYGHPSE NRTIGTTMV (SEQ ID NO: 693)	AAV9 Loop1	10.07147	B	2°
LRLADREHSS GESSFEYTVT (SEQ ID NO: 694)	AAV9 Loop1	8.276815	L	2°	GIESEESEYHTL PRARIMQR (SEQ ID NO: 695)	AAV9 Loop1	9.942643	B	NHP

SSTSTMPNSSR SSHSSSWTL (SEQ ID NO: 566)	AAV9 Loop2	8.265266	L	2°	LNSNAIKNLPTP LGGAAGHQ (SEQ ID NO: 544)	AAV9 Loop2	9.829831	B	2°
PACGSSSNSTR EPPASGLVA (SEQ ID NO: 592)	AAV9 Loop2	8.226133	L	2°	KFKGVDNVDYH GRGDLGRPE (SEQ ID NO: 696)	AAV9 Loop1	9.75871	B	2°
LKTQLTANTT KSGNTFRPEV (SEQ ID NO: 630)	AAV9 Loop2	8.191223	L	2°	GDSIGVLGYQK TVDHTKVNR (SEQ ID NO: 697)	AAV9 Loop2	9.70596	B	NHP
VLNSNAIKNLP PPLCCAAGH (SEQ ID NO: 615)	AAV9 Loop2	8.172553	L	2°	HKENGGAGLAE YHAKATEHL (SEQ ID NO: 698)	AAV9 Loop1	9.701024	B	2°
DFRINVENKT RARDDVSRKQ (SEQ ID NO: 699)	AAV9 Loop2	8.16575	L	2°	GFENVESDYHT LPRARIMQK (SEQ ID NO: 700)	AAV9 Loop1	9.690835	B	2°
HDTKLMQNST KAHLKRSNVE (SEQ ID NO: 701)	AAV9 Loop2	8.163909	L	2°	QDNTCEEYHTM GPHSSHIPK (SEQ ID NO: 543)	AAV9 Loop1	9.67355	B	3°
ELENNKTMNR AENGGRPPLN (SEQ ID NO: 702)	AAV9 Loop2	8.137323	L	2°	RIKLTYIHKGSA MQDLAEVS (SEQ ID NO: 703)	AAV9 Loop1	9.622513	B	2°
NVDIRINVENQ TRARDDVSR (SEQ ID NO: 704)	AAV9 Loop2	8.121921	L	2°	AGGSSSNSTREP PASGLVAF (SEQ ID NO: 553)	AAV9 Loop2	9.56453	B	2°
ENNKTMNRAE NGGRPPNHHF (SEQ ID NO: 705)	AAV9 Loop2	8.112242	L	2°	GGGGPGGSGGV GGDSGGYTY (SEQ ID NO: 706)	AAV9 Loop1	9.522066	B	NHP
PELVNNKTMN RAENGGRPPN (SEQ ID NO: 707)	AAV9 Loop2	8.102343	L	2°	QDNTCEEYHSV GPHSCRIPK (SEQ ID NO: 708)	AAV9 Loop1	9.510913	B	2°
LPEAVLALHN STRDLVAGES (SEQ ID NO: 606)	AAV9 Loop2	8.101657	L	2°	SLLESTSDSNLN KYSTIHKI (SEQ ID NO: 709)	AAV9 Loop1	9.442508	B	2°
GGSSSNSTRVP PASGLVALV (SEQ ID NO: 710)	AAV9 Loop2	8.092907	L	2°	ESLVGRKESPHS RSGSIVSS (SEQ ID NO: 711)	AAV9 Loop1	9.42889	B	2°

EPQPVLENNK TMNRAENGG R (SEQ ID NO: 712)	AAV9 Loop2	8.088374	L	2°	PIGNQEKNVPLP VIAELPPK (SEQ ID NO: 713)	AAV9 Loop1	9.412134	B	2°
AESLELAGNA TRDNKKTRII (SEQ ID NO: 714)	AAV9 Loop1	8.057996	L	2°	GFENEESDYHTL PRAGIMQM (SEQ ID NO: 715)	AAV9 Loop1	9.404316	B	NHP
ENNKTMNRAE NGGRPHNHPF (SEQ ID NO: 716)	AAV9 Loop2	8.057221	L	2°	SSDSIGVLGYQK TVDHTKEK (SEQ ID NO: 717)	AAV9 Loop2	9.301232	B	NHP
VDFRIHVENQ TRARDDESSK (SEQ ID NO: 718)	AAV9 Loop2	8.056751	L	2°	YSIGVQGYQKT VDHTKVNSK (SEQ ID NO: 539)	AAV9 Loop2	9.298736	B	3°
MIQTITNKSKI QNVGGTYES (SEQ ID NO: 719)	AAV9 Loop2	8.048419	L	2°	YGSTTDDDDHH STLVSSRHY (SEQ ID NO: 720)	AAV9 Loop1	9.288509	B	2°
VPAGGSSSNS TLEPPASGLV (SEQ ID NO: 624)	AAV9 Loop2	8.046119	L	2°	EDGSARLAEYH AKATERLGT (SEQ ID NO: 721)	AAV9 Loop1	9.284193	B	2°
LKTTLTANITK SGNTFRHEV (SEQ ID NO: 722)	AAV9 Loop2	8.037933	L	2°	RNGPEVSNDIKK RKVDDKDS (SEQ ID NO: 723)	AAV9 Loop2	9.261346	B	3°
SALDTNSSKST SSFPCPAVH (SEQ ID NO: 724)	AAV9 Loop2	8.03392	L	2°	GFENEESDYHTL PRARTMQK (SEQ ID NO: 725)	AAV9 Loop1	9.205369	B	2°
DINGLTTGRR GDMAAIHGM N (SEQ ID NO: 556)	AAV9 Loop1	10.13873	L	NHP	IDAQGEPHGYR QGDPSYRSF (SEQ ID NO: 726)	AAV9 Loop1	9.160974	B	2°
GFENEESDYH TLPRAGIMQK (SEQ ID NO: 531)	AAV9 Loop1	10.00615	Lv	3°	RHLESQSDDPHF HEKKQIPF (SEQ ID NO: 727)	AAV9 Loop1	9.117673	B	2°
LNSNAIKNLPT PLGGAAGHQ (SEQ ID NO: 544)	AAV9 Loop2	9.488507	Lv	2°	QDNSCEEYHTA GPHSGHIPK (SEQ ID NO: 728)	AAV9 Loop1	9.094751	B	2°
KENGGSRLAE YHAKATEHLS (SEQ ID NO: 729)	AAV9 Loop1	8.79956	Lv	NHP	NVDFSINVENQT RARDDVSR (SEQ ID NO: 557)	AAV9 Loop2	9.086539	B	2°

SSQAGVPVRV PRSPRAPESA (SEQ ID NO: 645)	AAV9 Loop1	8.419343	Lv	NHP	TAAHPELKTQL TANITKSGN (SEQ ID NO: 568)	AAV9 Loop2	9.047351	B	2°
VLNSNAIKNL HQPLGGAAGH (SEQ ID NO: 559)	AAV9 Loop2	8.328063	Lv	2°	QTFDDPHVSSFH GSKRGLNS (SEQ ID NO: 730)	AAV9 Loop1	9.022515	B	2°
QRGKGLSHW VPEHPALPTAS (SEQ ID NO: 731)	AAV9 Loop2	8.279242	Lv	2°	SSTSTMPNSSRS SHSSSWTL (SEQ ID NO: 566)	AAV9 Loop2	8.998521	B	2°
QRSGASASGV YTIKASNATK (SEQ ID NO: 732)	AAV9 Loop2	8.262213	Lv	2°	YGQDNTCEEYH TVGPHSSPI (SEQ ID NO: 733)	AAV9 Loop1	8.996019	B	2°
QRSWASASGV YTIQASNATK (SEQ ID NO: 734)	AAV9 Loop2	8.220422	Lv	2°	AKSFDWGLQAG YITVNSNLS (SEQ ID NO: 735)	AAV9 Loop2	8.992627	B	2°
QKGIQKPKSG TKGNYDDDW K (SEQ ID NO: 736)	AAV9 Loop2	8.190674	Lv	2°	GFENEESDYHTL PRARITQK (SEQ ID NO: 737)	AAV9 Loop1	8.978805	B	2°
DFRIHVENQTL ALDDVSRKQ (SEQ ID NO: 590)	AAV9 Loop2	8.158255	Lv	2°	HDTNLMQNSTK AHLKRSNVE (SEQ ID NO: 542)	AAV9 Loop2	8.946112	B	3°
VDFRIHVENQ TLARDDVSRN (SEQ ID NO: 680)	AAV9 Loop2	7.790053	Lv	2°	VLNSNAIKNLH QPLGGAAGH (SEQ ID NO: 559)	AAV9 Loop2	8.939443	B	2°
ICVLGYQKTV DHTKVNSKLS (SEQ ID NO: 738)	AAV9 Loop2	7.766946	Lv	2°	VKENGGARLAE YHTKATEHL (SEQ ID NO: 739)	AAV9 Loop1	8.91688	B	2°
AGGSSSNSTR EPPASGLVAF (SEQ ID NO: 553)	AAV9 Loop2	7.746055	Lv	2°	LKENGGAILAE YHTKATEHL (SEQ ID NO: 740)	AAV9 Loop1	8.908043	B	2°
VLNSNAIKNLP QQLGGAAGH (SEQ ID NO: 576)	AAV9 Loop2	7.745215	Lv	2°	GFEDEESDYHTL PRARITQR (SEQ ID NO: 741)	AAV9 Loop1	8.900953	B	2°
RWLQAQKDK MLSKNDTRSA G (SEQ ID NO: 742)	AAV9 Loop2	7.694831	Lv	2°	LPEAVLALHNST RDLVAGES (SEQ ID NO: 606)	AAV9 Loop2	8.872191	B	2°

NMTKATVNY RKEKTHMMS AV (SEQ ID NO: 743)	AAV9 Loop2	7.644042	Lv	2°	FGSATLDDEHH STLVSTSHY (SEQ ID NO: 744)	AAV9 Loop1	8.868846	B	2°
QRSWASASGV NTIQVSNATK (SEQ ID NO: 745)	AAV9 Loop2	7.642708	Lv	2°	FGSATIDVEHHS TLASTRHY (SEQ ID NO: 746)	AAV9 Loop1	8.865535	B	2°
DFSIHVENQTR ARDDVSSKQ (SEQ ID NO: 663)	AAV9 Loop2	7.631215	Lv	2°	LALYNSTRDRK AGVSAEPEP (SEQ ID NO: 747)	AAV9 Loop2	8.860514	B	2°
LGAPALVQRP QGNRPGLAKA (SEQ ID NO: 748)	AAV9 Loop1	7.630549	Lv	NHP	GGARLVEYHAK ATENLSTLS (SEQ ID NO: 749)	AAV9 Loop1	8.834379	B	2°
SIGVLGYKKT VDHTKVNSKP (SEQ ID NO: 750)	AAV9 Loop2	7.603106	Lv	2°	GSATYEDEHHS TLVSTRQYG (SEQ ID NO: 751)	AAV9 Loop1	8.830555	B	2°
QRSWASASGA YTIQVSNATK (SEQ ID NO: 752)	AAV9 Loop2	7.588547	Lv	2°	GFENEESDYHTL PRAGIMQK (SEQ ID NO: 531)	AAV9 Loop1	17.18655	M	3°
YWGSVTGSQS NAVSHTQAPR (SEQ ID NO: 753)	AAV9 Loop2	7.563096	Lv	2°	YSIGVLGYQKT VDHTKVNSK (SEQ ID NO: 530)	AAV9 Loop2	14.02683	M	3°
QAYKAPVVVE GKVQGLVQA G (SEQ ID NO: 754)	AAV9 Loop2	7.544197	Lv	2°	YSIGVLGYQKTE DHTKVNSK (SEQ ID NO: 532)	AAV9 Loop2	13.96413	M	3°
IMELAGNAAR DNKTTRILPR (SEQ ID NO: 755)	AAV9 Loop1	7.541857	Lv	NHP	YSIGVLGNQKT VDHTKVNSK (SEQ ID NO: 533)	AAV9 Loop2	13.02269	M	3°
LYRSLAGRAP AWAAAVSAS G (SEQ ID NO: 756)	AAV9 Loop2	7.53098	Lv	2°	DAPLGTGANKD LELPWLEQQ (SEQ ID NO: 534)	AAV9 Loop2	12.71293	M	3°
SICVLGYQKT VDHTKANSKL (SEQ ID NO: 757)	AAV9 Loop2	7.484248	Lv	2°	YSIGVLSYQKTV DHTKVNSK (SEQ ID NO: 535)	AAV9 Loop2	12.17668	M	3°
SVSRALTQAL PAPTCQNTQV (SEQ ID NO: 758)	AAV9 Loop2	7.436854	Lv	2°	YSIGVLGYQKT VDHTKVNGK (SEQ ID NO: 536)	AAV9 Loop2	11.92955	M	3°

VLNSNAIKNLPPPLCCAAGH (SEQ ID NO: 615)	AAV9 Loop2	7.41241	Lv	2°	YSIGVLVYQKTVDHTKVNSK (SEQ ID NO: 537)	AAV9 Loop2	11.56475	M	3°
LKTQLTANITKSGNTFRHEV (SEQ ID NO: 618)	AAV9 Loop2	7.409261	Lv	2°	QDNTCEEYHTMGPHSSHIPK (SEQ ID NO: 543)	AAV9 Loop1	11.10573	M	3°
QKGIKKPKSGTQGNYDDDWK (SEQ ID NO: 759)	AAV9 Loop2	7.408743	Lv	2°	YSIGVQGYQKTVDHTKVNSK (SEQ ID NO: 539)	AAV9 Loop2	10.96218	M	3°
VPTGSFMMVNSSGRASGKKA (SEQ ID NO: 760)	AAV9 Loop2	7.399719	Lv	2°	HDTNLMQNSTKAHLKRSNVE (SEQ ID NO: 542)	AAV9 Loop2	10.94556	M	3°
GFENEESDYHTLPRAGIMQK (SEQ ID NO: 531)	AAV9 Loop1	16.28212	I	3°	FSIGVLGYQKTVDHTKVNSK (SEQ ID NO: 545)	AAV9 Loop2	10.34177	M	3°
YSIGVLGYQKTVDHTKVNSK (SEQ ID NO: 530)	AAV9 Loop2	13.38331	I	3°	YSIGVLGYQKSVDHTKVNSK (SEQ ID NO: 547)	AAV9 Loop2	10.06095	M	3°
YSIGVLGYQKTEDHTKVNSK (SEQ ID NO: 532)	AAV9 Loop2	13.26262	I	3°	RSDLTTLTESYANRQGLNKN (SEQ ID NO: 761)	AAV9 Loop1	9.866847	M	3°
YSIGVLGNQKTVDHTKVNSK (SEQ ID NO: 533)	AAV9 Loop2	12.32805	I	3°	LNSNAIKNLPTPLGGAAGHQ (SEQ ID NO: 544)	AAV9 Loop2	9.7937	M	2°
DAPLGTGANKDLELPWLEQQ (SEQ ID NO: 534)	AAV9 Loop2	11.7147	I	3°	YSIGVLGYQKTVDHTKVNST (SEQ ID NO: 548)	AAV9 Loop2	9.791349	M	3°
YSIGVLSYQKTVDHTKVNSK (SEQ ID NO: 535)	AAV9 Loop2	11.44192	I	3°	SSQAGVPVRAPRSPRAPESA (SEQ ID NO: 762)	AAV9 Loop1	9.761722	M	3°
YSIGVLGYQKTVDHTKVNGK (SEQ ID NO: 536)	AAV9 Loop2	11.27777	I	3°	TAGSTMGAASITQTVLARNL (SEQ ID NO: 763)	AAV9 Loop1	9.655512	M	2°
YSIGVLVYQKTVDHTKVNSK (SEQ ID NO: 537)	AAV9 Loop2	11.27179	I	3°	KENGGARLAEYHAKSTEHLS (SEQ ID NO: 764)	AAV9 Loop1	9.605415	M	2°

QDNTCEEYHT MGPHSSHIPK (SEQ ID NO: 543)	AAV9 Loop1	10.76647	I	3°	TAAHPELKTQL TANITKSGN (SEQ ID NO: 568)	AAV9 Loop2	9.196748	M	2°
QDNTCEEYHT MGPHSNHIPK (SEQ ID NO: 643)	AAV9 Loop1	10.69909	I	3°	WVDDQKVINKA SNSNKIRLE (SEQ ID NO: 765)	AAV9 Loop2	9.097937	M	3°
HDTNLMQNST KAHLKRSNVE (SEQ ID NO: 542)	AAV9 Loop2	10.69498	I	3°	AGGSSSNSTREP PASGLVAF (SEQ ID NO: 553)	AAV9 Loop2	9.074173	M	2°
YSIGVQGYQK TVDHTKVNSK (SEQ ID NO: 539)	AAV9 Loop2	10.26071	I	3°	ENVDFRINVEN QTRARDDVS (SEQ ID NO: 588)	AAV9 Loop2	9.004632	M	2°
LLSASSVGSA HGDEGAGYTD (SEQ ID NO: 766)	AAV9 Loop2	10.06101	I	2°	GMIQTANVGVG ISGNEGMQA (SEQ ID NO: 767)	AAV9 Loop2	8.954396	M	2°
FSIGVLGYQKT VDHTKVNSK (SEQ ID NO: 545)	AAV9 Loop2	9.780598	I	3°	EQLITQEPKEAN STTSKKQS (SEQ ID NO: 768)	AAV9 Loop2	8.880195	M	2°
YSIGVLGYQK SVDHTKVNSK (SEQ ID NO: 547)	AAV9 Loop2	9.476502	I	3°	PGAVLALYNST SDRVAGESA (SEQ ID NO: 769)	AAV9 Loop2	8.795954	M	2°
YSIGVLGYQK TVDHTKVNST (SEQ ID NO: 548)	AAV9 Loop2	9.285147	I	3°	NVDFSINVENQT RARDDVSR (SEQ ID NO: 557)	AAV9 Loop2	8.79169	M	2°
ISEERFRQLEKI KTVGSTYM (SEQ ID NO: 770)	AAV9 Loop1	8.875858	I	2°	LVPACCSSSNST REPPASGR (SEQ ID NO: 569)	AAV9 Loop2	8.739936	M	2°
AKSAVEILLTI DDTKRAGMK (SEQ ID NO: 589)	AAV9 Loop2	8.42562	I	2°	LVPAGGSSSNST REQHASGR (SEQ ID NO: 587)	AAV9 Loop2	8.722982	M	2°
IQGQLQPSLKT GSEEIRGQY (SEQ ID NO: 771)	AAV9 Loop1	8.372161	I	2°	DFRIHVENQTLA LDDVSRKQ (SEQ ID NO: 590)	AAV9 Loop2	8.702759	M	2°
ERVPVHATVV GDVVSPEMGE (SEQ ID NO: 772)	AAV9 Loop1	8.282936	I	2°	QNAVNVVKQIK TLIEKTNEE (SEQ ID NO: 773)	AAV9 Loop2	8.696219	M	2°

SVAGSLSLAA NSAAQKTHQD (SEQ ID NO: 774)	AAV9 Loop1	8.172114	I	2°	VTAGGSSSNSTR EQPASGRV (SEQ ID NO: 567)	AAV9 Loop2	8.685301	M	2°
GLESEESEYHT LPRARIMQR (SEQ ID NO: 641)	AAV9 Loop1	8.167663	I	3°	VQAGGSSSNST REQPASGRV (SEQ ID NO: 563)	AAV9 Loop2	8.685292	M	2°
ARLLGSLDPS ADLSSQRGKV (SEQ ID NO: 775)	AAV9 Loop1	8.138603	I	2°	YSIGVLGNQKT VDHTKENSK (SEQ ID NO: 555)	AAV9 Loop2	8.670167	M	3°
HYLGLARRAP AGAAAVSASG (SEQ ID NO: 776)	AAV9 Loop2	8.117226	I	2°	VPAGGSSSNSTR EQHASGRV (SEQ ID NO: 570)	AAV9 Loop2	8.591786	M	2°
AGGSSSNSTR EPPASGLVAF (SEQ ID NO: 553)	AAV9 Loop2	8.115973	I	2°	NVDIRINVENQT RARDDVSR (SEQ ID NO: 704)	AAV9 Loop2	8.58057	M	2°
KQQDSLPDTK AGGPAHSKVV (SEQ ID NO: 777)	AAV9 Loop1	8.098824	I	2°	FGSASYDEEHHS TLVSTRHY (SEQ ID NO: 778)	AAV9 Loop1	8.453823	M	2°
KWKTTFDLKK TRMEPFHFKT (SEQ ID NO: 779)	AAV9 Loop1	8.013461	I	2°	ACGSSSNSTREP PASGLVAL (SEQ ID NO: 583)	AAV9 Loop2	8.411724	M	2°
SDSICVLGYQ KTVDHTKVNS (SEQ ID NO: 552)	AAV9 Loop2	8.003536	I	2°	QNAVNWVKQIK TLIEKTNEE (SEQ ID NO: 780)	AAV9 Loop2	8.324026	M	2°
SFMIRGDEVR QTASGQTGKI (SEQ ID NO: 781)	AAV9 Loop2	7.959401	I	2°	LVTAGGSSSNST REQPASGR (SEQ ID NO: 782)	AAV9 Loop2	8.295296	M	2°
GMIQTANVGV GISGNEGMQA (SEQ ID NO: 767)	AAV9 Loop2	7.936933	I	2°	ENVDFRIHVEN KTRARDDVS (SEQ ID NO: 783)	AAV9 Loop2	8.291462	M	2°
RGLQVTVLGE NKTGNSPVER (SEQ ID NO: 784)	AAV9 Loop2	7.901321	I	2°	PELKTQLTANIT KSGDTFRP (SEQ ID NO: 785)	AAV9 Loop2	8.279723	M	2°
LAMAGSAVGT TSLTLSAQSR (SEQ ID NO: 786)	AAV9 Loop2	7.893966	I	2°	LVKIRSNSSKPD LMASLLPQ (SEQ ID NO: 787)	AAV9 Loop2	8.275692	M	2°

TAAHPELKTQ LTANITKSGN (SEQ ID NO: 568)	AAV9 Loop2	7.789394	I	2°	VLNSNAIKNLH QPLGGAAGH (SEQ ID NO: 559)	AAV9 Loop2	8.260826	M	2°
TGLLPAAGEK TCGPLGNLAE (SEQ ID NO: 788)	AAV9 Loop2	7.788899	I	2°	TGGSSSNSTREP HASGRVAL (SEQ ID NO: 789)	AAV9 Loop2	8.259557	M	2°
YSIGVLGNQK TVDHTKENSK (SEQ ID NO: 555)	AAV9 Loop2	7.687701	I	3°	DFSIHVENQTRA RDDVSSKQ (SEQ ID NO: 663)	AAV9 Loop2	8.246807	M	2°
AKSAVEILLTI DDTKRAGMK (SEQ ID NO: 589)	AAV9 Loop2	9.094878	S	2°	SVLNSNAIKNLP QQLGGAAG (SEQ ID NO: 586)	AAV9 Loop2	8.218652	M	2°
YSIGVLGYQK TEDHTKVNSK (SEQ ID NO: 532)	AAV9 Loop2	8.994118	S	3°	LKTPLTTNITKS GNTFLPEV (SEQ ID NO: 790)	AAV9 Loop2	8.202921	M	2°
GFENEESDYH TLPRAGIMQK (SEQ ID NO: 531)	AAV9 Loop1	8.825712	S	3°	YSFGVLGYQKT VDHTKVNSK (SEQ ID NO: 791)	AAV9 Loop2	8.199302	M	3°
CKFLKEGTICK RARCDDMDD (SEQ ID NO: 609)	AAV9 Loop2	8.791321	S	2°	LLRECNTVRSFL AAAAETLE (SEQ ID NO: 682)	AAV9 Loop2	8.170173	M	2°
YSIGVLGYQK TVDHTKVNSK (SEQ ID NO: 530)	AAV9 Loop2	8.783985	S	3°	MIQTLTNKSKIQ NVGWTYEF (SEQ ID NO: 575)	AAV9 Loop2	8.154686	M	2°
AELKPANLPV TKPFTVGDNK (SEQ ID NO: 792)	AAV9 Loop2	8.550499	S	2°	QTIPQISNRTRN TMVQDNSY (SEQ ID NO: 793)	AAV9 Loop2	8.139015	M	2°
PSFCTQKPRKF YCELPDISL (SEQ ID NO: 794)	AAV9 Loop2	8.151695	S	2°	YWGSVTGSQSN AVSHTQAPR (SEQ ID NO: 753)	AAV9 Loop2	8.124375	M	2°
YSIGVLGNQK TVDHTKVNSK (SEQ ID NO: 533)	AAV9 Loop2	8.113986	S	3°	LVQIRSNSSKPD LMASMLPQ (SEQ ID NO: 795)	AAV9 Loop2	8.111179	M	2°
GMIQTANVGV GISGNEGMQA (SEQ ID NO: 767)	AAV9 Loop2	7.999101	S	2°	VNNSSTNDNLV RKNDKVYIN (SEQ ID NO: 796)	AAV9 Loop2	8.098509	M	2°

EKMLVLDAEQRVTAGEALAN (SEQ ID NO: 797)	AAV9 Loop2	7.961541	S	2°	VPAGGSSSNSTLEPPASGLV (SEQ ID NO: 624)	AAV9 Loop2	8.07234	M	2°
GSQASGIPAKVSATVSSPGS (SEQ ID NO: 798)	AAV9 Loop2	7.915445	S	2°	PELENNKTMSRAENGGRPPN (SEQ ID NO: 799)	AAV9 Loop2	8.048644	M	2°
WLVAKRVQDHTTVVGDVVSQ (SEQ ID NO: 800)	AAV9 Loop2	7.88424	S	2°	GNVDFRIHVENQTRALDDVS (SEQ ID NO: 801)	AAV9 Loop2	8.030347	M	2°
YKVLKQVHPDSGVSSQAMGI (SEQ ID NO: 802)	AAV9 Loop1	7.823109	S	3°	QTIPKISNRTLNTQVQDNSY (SEQ ID NO: 803)	AAV9 Loop2	7.983687	M	2°
SPANGSLGPTLNLSSWESRL (SEQ ID NO: 804)	AAV9 Loop2	7.701629	S	2°	GLVQAGGSSSNSTREPHASG (SEQ ID NO: 805)	AAV9 Loop2	7.956418	M	2°
QAGIPVRAPRSPRAQEHALE (SEQ ID NO: 806)	AAV9 Loop2	7.658135	S	2°	VPACGSSSNSTREPPASGLV (SEQ ID NO: 620)	AAV9 Loop2	7.928663	M	2°
IGILSQGVASGMAGSAMSSK (SEQ ID NO: 807)	AAV9 Loop2	7.654589	S	2°	TAAHPELKTTLTANITKSGN (SEQ ID NO: 808)	AAV9 Loop2	7.920023	M	2°
NCKFLKEGTICKRARCDDMD (SEQ ID NO: 623)	AAV9 Loop2	7.646112	S	2°	LKTQLTANITKSGNTFRHEV (SEQ ID NO: 618)	AAV9 Loop2	7.902972	M	2°
LKQVHPDTCISSKAMGIMNS (SEQ ID NO: 809)	AAV9 Loop2	7.54635	S	2°	MSQVPLQNYSKVERHPVTAK (SEQ ID NO: 810)	AAV9 Loop1	7.901576	M	2°
HSWTTLYNPSLESRVTMSVD (SEQ ID NO: 811)	AAV9 Loop2	7.529922	S	2°	VPACCSSSNSTREPPASGRV (SEQ ID NO: 812)	AAV9 Loop2	7.898005	M	2°
QNTAIILLKEDSLDATRAAI (SEQ ID NO: 813)	AAV9 Loop2	7.521096	S	2°	AKSSDWGLQAGYIHVNPNLS (SEQ ID NO: 814)	AAV9 Loop2	7.881375	M	2°
DAPLGTGANKDLELPWLEQQ (SEQ ID NO: 534)	AAV9 Loop2	7.515163	S	3°	YSIGVLGYQKTVDLTKVNSK (SEQ ID NO: 815)	AAV9 Loop2	7.866734	M	3°

SDSICVLGYQ KTVDHTKVNS (SEQ ID NO: 552)	AAV9 Loop2	7.50278	S	2°	DNVDFRIHVEN QTRARDDVS (SEQ ID NO: 816)	AAV9 Loop2	7.853419	M	2°
LLAGGLHAW KVIKGEEISVV (SEQ ID NO: 817)	AAV9 Loop2	7.465502	S	2°					
EGKVQGLVQA GGSSSNSTRE (SEQ ID NO: 818)	AAV9 Loop2	7.406399	S	2°					
QSAVEILLTID DTKMAGMKE (SEQ ID NO: 819)	AAV9 Loop2	7.390435	S	2°					
YSIGVLSYQKT VDHTKVNSK (SEQ ID NO: 535)	AAV9 Loop2	7.340184	S	3°					
ETGWTPGLRS LENYGTKLLD (SEQ ID NO: 820)	AAV9 Loop2	7.318566	S	2°					

[0093] **Table 4B**: These sequences were identified by looking at peptides derived from domains that exhibit consistent organ transfectivity across many variants. Shown in the table below is the amino acid sequence, the AAV scaffold in which the peptide was grafted, the log2fold change value and the respective organ that it transduces, as well as the screen in which the peptide was identified (P=pancreas; H=heart; B=Brain; L=Lung; Lv=liver; K=kidney; M=muscle; S=spleen; I=intestine):

Amino Acid Sequence	Scaffold	Log2FC	Organ	Screen Identified	Amino Acid Sequence	Scaffold	Log2FC	Organ	Screen Identified In
QQVSSPHTKAPL HKDSANGF (SEQ ID NO:870)	AAV9 Loop1	8.3	B	2°	TQNKRANRTEE NVSDGSPTA (SEQ ID NO:891)	AAV9 Loop2	5.11	L	NHP
VAGHFESQSDD PHFRGKKQL (SEQ ID NO:871)	AAV9 Loop1	8.79	B	2°	FGPVLDNATRK SASVHAIAS (SEQ ID NO:892)	AAV9 Loop1	4.69	L	2°
GVHAKSSDWGL QAGYIPVNP	AAV9 Loop1	11.1 4	B	1°	DNRTRQHIQYR EDKNLTGTA	AAV9 Loop2	7.12	L	2°

(SEQ ID NO:872)					(SEQ ID NO:893)				
RWYGQDNTCEEYHTVGPHSC (SEQ ID NO:873)	AAV9 Loop1	5.34	B	1°	KPQVTRGDVFTMSEDEHTVY (SEQ ID NO:894)	AAV9 Loop2	5.08	L	NHP
FGSATYDDEHHSTLVSTRHY (SEQ ID NO:874)	AAV9 Loop1	6.55	B	1°	AGMDASDRQVIISSNLTRPK (SEQ ID NO:895)	AAV9 Loop1	7.4	L	2°
YLESRSINEKDYHSRRYIDE (SEQ ID NO:875)	AAV9 Loop1	7.63	B	1°	TDRTSLPSANITRSFKNEDL (SEQ ID NO:896)	AAV9 Loop2	3.01	L	2°
GDMANSSIEFHPKPQQQRDV (SEQ ID NO:876)	AAV9 Loop1	7.43	B	1°	NRDGTVLESNRTNSLRKINS (SEQ ID NO:897)	AAV9 Loop1	4.22	L	3°
PQSWGRGSAGGNTHSPLGVP (SEQ ID NO:877)	AAV9 Loop1	4.7	H	2°	NLPGDPTSNATSRGAEGPLP (SEQ ID NO:898)	AAV9 Loop2	3.76	L	1°
FKGVDNEDDHGRGHLGRLKK (SEQ ID NO:878)	AAV9 Loop1	2.9	H	2°	PMPDQAVPTGSFMMVNSSRR (SEQ ID NO:899)	AAV9 Loop1	4.67	M	2°
GKGNYAERAGAGAPAYLEAV (SEQ ID NO:879)	AAV9 Loop1	6.32	H	2°	QQQSYLDSGIHSGATTTAPS (SEQ ID NO:900)	AAV9 Loop1	3.67	M	1°
LVSKGPLVQTKGTGASGSFR (SEQ ID NO:880)	AAV9 Loop1	6.93	H	2°	GTQGRLCNKTSECMDGCELM (SEQ ID NO:901)	AAV9 Loop2	4.67	M	2°
GFAIGIVGAAGERGTAQQPR (SEQ ID NO:881)	AAV9 Loop1	6.4	H	2°	DHTGRHDTPRMSTSQTPGRA (SEQ ID NO:902)	AAV9 Loop1	6.91	M	2°
ATLERPASSSHAQGTHLQLR (SEQ ID NO:882)	AAV9 Loop1	6.66	H	2°	VAEIPFNSTNKYQLYIHETE (SEQ ID NO:903)	AAV9 Loop2	5.03	M	2°
LATAGSAMGAASLTVSAQSR (SEQ ID NO:883)	AAV9 Loop2	6.33	H	1°	LVQIRSNSSKPDLMASLLPK (SEQ ID NO:904)	AAV9 Loop2	6.44	M	2°
LGAVGSTMGAGSLTLTEQAR (SEQ ID NO:884)	AAV9 Loop1	4.47	H	2°	VLKQGHPDTGISSKAMGIIY (SEQ ID NO:905)	AAV9 Loop1	6.17	M	2°
TAGAAMGAAGPTLSAQSRSL (SEQ ID NO:885)	AAV9 Loop1	6.59	H	2°	ISNRVLVSNQTGTAFQPDAA (SEQ ID NO:906)	AAV9 Loop2	7.11	M	2°
IGAGVAGGITGSISLANGKS (SEQ ID NO:886)	AAV9 Loop1	7.7	H	2°	AMGAGVVGRTTGSMSLASGK (SEQ ID NO:907)	AAV9 Loop1	2.06	M	3°
AVGGAIGRVANTIRSGPSNS (SEQ ID NO:887)	AAV9 Loop1	7.79	H	2°	RLTAGTGGREVGARGEPSGI (SEQ ID NO:908)	AAV9 Loop2	4.29	P	2°
SQGISSSSTSTMPNSSQSSH (SEQ	AAV9 Loop2	9.99	H	1°	ILQRGSGTAAVDFTKKDNTA	AAV9 Loop2	5.54	P	2°

ID NO:888)					(SEQ ID NO:909)				
VKDRSHNVTEN VTQVLSLGA (SEQ ID NO:889)	AAV9 Loop2	4.86	L	2°	WGQGKLVSVSP DSTQGPSVF (SEQ ID NO:910)	AAV9 Loop1	5.09	P	3°
EILELAGNATRD SNKTRIIP (SEQ ID NO:890)	AAV9 Loop1	6.44	L	2°	SNLHDTKASGP AHSKEVPTL (SEQ ID NO:911)	AAV9 Loop2	2.76	P	2°

[0094] **Table 5:** table of the identified 23 immune-orthogonal AAV serotypes with their DNA and amino acid (AA) sequences.

Name	DNA Sequence	AA Sequence
BT1	ATGTCTTTTGTTGACCACCCTCCAGATTGGTTGGA ATCGATCGGCGACGGCTTTCGTGAATTTCTCGGCC TTGAGGCGGGTCCCCCGAAACCCAAGGCCAATCA ACAGAAGCAAGATAACGCTCGAGGTCTTGTGCTTC CTGGGTACAAGTATCTTGGTCCTGGGAACGGCCTT GATAAGGGCGATCCTGTCAATTTTGCTGACGAGGT TGCCCGAGAGCACGACCTCTCCTACCAGAAACAG CTTGAGGCGGGCGATAACCCTTACCTCAAGTACAA CCACGCGGACGCAGAGTTTCAGGAGAAACTCGCT TCTGACACTTCTTTTGGAGGAAACCTTGGGAAGGC TGTTTTCCAGGCTAAAAAGAGGCTTCTCGAACCTC TTGGCCTGGTTGAGACGCCGGATAAAACGGCGCCT GCGGCAAAAAAGAGGCCTCTAGAGCAGAGTCCTC AAGAGCCAGACTCCTCGAGCGGATTTGGCAAAAA GGGCAAACAGCCTGCCAGAAAGAGACTCAATTTT GACGACGAACCTGGAGCCGGAGACGGGCCTCCCC CAGAAGGACCATCTTCCGGAGCTATGTCTACTGAG ACTGAAATGCGTGCAGCAGCTGGCGGAAATGGTG GCGATGCGGGACAAGGTGCCGAGGGAGTGGGTAA TGCCTCCGGTGATTGGCATTGCGATTCCACTTGGT CAGAGAGCCACGTCACCACCACCTCAACCCGCAC CTGGGTCCTGCCGACCTACAACAACCACCTGTACC TGCGGCTCGGCACGAGCAACGCCAGCGACACCTT CAACGGATTCTCCACCCCCTGGGGATACTTTGACT TCAACCGCTTCCACTGCCACTTCTCGCCAAGAGAC TGGCAAAGGCTCATCAACAACCACTGGGGGCTGC GCCCCAAAAGCATGCAAGTCCGCATCTTCAACATC CAAGTTAAGGAGGTCACGACGTCTAACGGAGAGA CGACCGTATCAAACAACCTCACCAGCACGGTCCA GATCTTTGCGGACAGCACGTACGAGCTCCCGTACG TGATGGATGCGGGTCAGGAGGGCAGCTTGCCCCCT TTCCCCAACGACGTGTTCATGGTGCCCCAGTACGG GTACTGCGGACTGGTAACCGGAGGCAGCTCTCAA AACCAGACAGACAGAAACGCCTTCTACTGTCTGG AGTACTTTCCCAGCCAGATGCTGAGAACAGGAAA CAACTTTGAGATGGTGTACAAGTTTGAGAACGTGC CCTTCCACTCCATGTACGCTCACAGCCAGAGCCTG GATAGGCTGATGAACCCGCTGCTGGACCAGTACCT GTGGGAACTCCAGTCTACCACATCTGGAGGAAATC TCAACCAGGGCACGTCAGCCACAAACTTTGCCAA	MSFVDHPPDWLESIGDGFREFLGLEAGPP KPKANQQKQDNARGLVLPGYKYLGPGN GLDKGDPVNFADEVAREHDLSYQKQLE AGDNPYLKYNHADAEFQEKLASDTSFG GNLGKAVFQAKKRLLEPLGLVETPDKTA PAAKKRPLEQSPQEPDSSSGFGKKGKQP ARKRLNFDDEPGAGDGPPPEGPSSGAMS TETEMRAAAGGNGGDAGQGAEGVGNA SGDWHCDSTWSESHVTTTSTRTWVLPTY NNHLYLRLGTSNASDTFNGFSTPWGYFD FNRFHCHFSPRDWQRLINNHWGLRPKSM QVRIFNIQVKEVTTSNGETTVSNNLTSTV QIFADSTYELPYVMDAGQEGSLPPFPND VFMVPQYGYCGLVTGGSSQNQTDRNAF YCLEYFPSQMLRTGNNFEMVYKFENVPF HSMYAHSQSLDRLMNPLLDQYLWELQS TTSGGNLNQGTSATNFAKLTKTNFSGYR KNWLPGPMMKQQRFSKTANQNYKIPQG GNNSLLHYETRTTLDGRWSNFAPGTAM ATAANDATDFSQAQLIFAGPNITGNTTTD ANNLMFTSEDELRATNPRDTDLFGHLAT NQQNATTVPTADDVDGVGVYPGMVWQ DRDIYYQGPIWAKIPHTDGHFHPSPLIGG FGLKSPPPQIFIKNTPVPANPATTFSPARIN SFITQYSTGQVAVKIEWEIQKERSKRWNP EVQFTSNYGAQDSLLWAPDNAGAYKEP RAIGSRYLTNHL (SEQ ID NO: 822)

	GCTGACCAAAACAAACTTTTCTGGCTACCGCAAAA ACTGGCTTCCGGGACCCATGATGAAGCAGCAGAG ATTCTCCAAGACTGCCAATCAAAACTACAAGATTC CCCAAGGGGGAAATAACAGTCTGCTCCATTATGA GACCAGAACTACCCTCGACGGAAGATGGAGCAAT TTTGCCCCGGGAACGGCCATGGCAACCGCGGCCA ACGATGCCACCGACTTCTCCCAGGCCCAGCTCATC TTTGCGGGGCCCAACATCACCGGCAATACCACGAC AGACGCCAACAACCTGATGTTCACTTCAGAAGATG AACTCAGGGCCACCAACCCCCGGGACACTGACCT GTTTGGCCACCTGGCAACCAACCAGCAAAACGCC ACCACCGTTCCTACCGCAGACGACGTGGACGGAG TCGGCGTGTACCCGGGAATGGTGTGGCAGGACAG AGACATTTATTACCAAGGGCCCATCTGGGCCAAGA TTCCCCATACGGATGGACACTTTCACCCATCTCCC CTCATTGGCGGATTTGGACTGAAAAGCCCGCCTCC ACAAATATTCATCAAAAACACCCCTGTACCCGCCA ATCCCGCGACGACCTTCTCTCCGGCCAGAATCAAC AGCTTCATCACCCAGTACAGCACCGGACAGGTGG CTGTCAAAATAGAATGGGAAATCCAGAAGGAGCG GTCCAAGAGATGGAACCCAGAGGTCCAGTTCACG TCCAACTACGGAGCACAGGACTCGCTTCTCTGGGC TCCCGACAACGCCGGAGCCTACAAAGAGCCCAGG GCCATTGGATCCCGATACCTCACCAACCACCTCTA G (SEQ ID NO: 821)	
BT2	ATGTCTTTTGTTGACCACCCTCCAGATTGGTTGGA ATCGATCGGCGACGGCTTTCGTGAATTTCTCGGCC TTGAGGCGGGTCCCCCGAAACCCAAGGCCAATCA ACAGAAGCAAGATAACGCTCGAGGTCTTGTGCTTC CTGGGTACAAGTATCTTGGTCCTGGGAACGGCCTT GATAAGGGCGATCCTGTCAATTTTGCTGACGAGGT TGCCCGAGAGCACGACCTCTCCTACCAGAAACAG CTTGAGGCGGGCGATAACCCTTACCTCAAGTACAA CCACGCGGACGCAGAGTTTCAGGAGAAACTCGCT TCTGACACTTCTTTTGGGGGAAACCTTGGGAAGGC TGTTTTCCAGGCTAAAAAGAGGATTCTCGAACCTC TTGGCCTGGTTGAGACGCCGGATAAAACGGCGCCT GCGGCAAAAAAGAGGCCTCTAGAGCAGAGTCCTC AAGAGCCAGACTCCTCGAGCGGAGTTGGCAAGAA AGGCAAACAGCCTGCCAGAAAGAGACTCAACTTT GACGACGAACCTGGAGCCGGAGACGGGCCTCCCC CAGAAGGACCATCTTCCGGAGCTATGTCTACTGAG ACTGAAATGCGTGCAGCAGCTGGCGGAAATGGTG GCGATGCGGGACAAGGTGCCGAGGGAGTGGGTAA TGCCTCCGGTGATTGGCATTGCGATTCCACTTGGT CAGAGAGCCACGTCACCACCACCTCAACCCGCAC CTGGGTCCTGCCGACCTACAACAACCACCTGTACC TGCGGCTCGGCTCGAGCAACGCCAGCGACACCTTC AACGGATTCTCCACCCCCTGGGGATACTTTGACTT TAACCGCTTCCACTGCCACTTCTCGCCAAGAGACT GGCAAAGGCTCATCAACAACCACTGGGGACTGCG CCCCAAAAGCATGCAAGTCCGCATCTTCAACATCC AAGTTAAGGAGGTCACGACGTCTAACGGGGAGAC GACCGTATCCAACAACCTCACCAGCACGGTCCAG	MSFVDHPPDWLESIGDGFREFLGLEAGPP KPKANQQKQDNARGLVLPGYKYLGPGN GLDKGDPVNFADEVAREHDLSYQKQLE AGDNPYLKYNHADAEFQEKLASDTSFG GNLGKAVFQAKKRILEPLGLVETPDKTA PAAKKRPLEQSPQEPDSSSGVGKKGKQP ARKRLNFDDEPGAGDGPPPEGPSSGAMS TETEMRAAAGGNGGDAGQGAEGVGNA SGDWHCDSTWSESHVTTTSTRTWVLPTY NNHLYLRLGSSNASDTFNGFSTPWGYFD FNRFHCHFSPRDWQRLINNHWGLRPKSM QVRIFNIQVKEVTTSNGETTVSNNLTSTV QIFADSTYELPYVMDAGQEGSLPPFPND VFMVPQYGYCGLVTGGSSQNQTDRNAF YCLEYFPSQMLRTGNNFEMVYKFENVPF HSMYAHSQSLDRLMNPLLDQYLWELQS TTSGGTLNQGNSATNFAKLTKTNFSGYR KNWLPGPMMKQQRFSKTASQNYKIPQG RNNSLLHYETRTTLDGRWSNFAPGTAM ATAANDATDFSQAQLIFAGPNITGNTTTD ANNLMFTSEDELRATNPRDTDLFGHLAT NQQNATTVPTVDDVDGVGVYPGMVWQ DRDIYYQGPIWAKIPHTDGHFHPSPLIGG FGLKSPPPQIFIKNTPVPANPATTFSPARIN SFITQYSTGQVAVKIEWEIQKERSKRWNP EVQFTSNYGAQDSLLWAPDNAGAYKEP RAIGSRYLTNHL (SEQ ID NO: 824)

	ATCTTTGCGGACAGCACGTACGAGCTCCCGTACGT GATGGATGCAGGTCAGGAGGGCAGCTTGCCTCCTT TCCCCAACGACGTGTTCATGGTGCCTCAGTACGGG TACTGCGGACTGGTAACCGGAGGCAGCTCTCAAA ACCAGACAGACAGAAATGCCTTCTACTGTCTGGAG TACTTTCCCAGCCAGATGCTGAGAACCGGAAACA ACTTTGAGATGGTGTACAAGTTTGAAAACGTGCCC TTCCACTCCATGTACGCTCACAGCCAGAGCCTGGA TAGGCTGATGAACCCGCTGCTGGACCAGTACCTGT GGGAGCTCCAGTCTACCACCTCTGGAGGAACTCTC AACCAGGGCAATTCAGCCACCAACTTTGCCAAGCT GACCAAAACAAACTTTTCTGGCTACCGCAAAAACT GGCTCCCGGGGCCCATGATGAAGCAGCAGAGATT CTCCAAGACTGCCAGTCAAAACTACAAGATTCCCC AGGGAAGAAACAACAGTCTGCTCCATTATGAGAC CAGAACTACCCTCGACGGAAGATGGAGCAATTTT GCCCCGGGAACGGCCATGGCAACCGCAGCCAACG ACGCCACCGACTTCTCTCAGGCCCAGCTCATCTTT GCGGGGCCCAACATCACCGGCAACACCACCACAG ATGCCAATAACCTGATGTTCACTTCAGAAGATGAA CTTAGGGCCACCAACCCCCGGGACACTGACCTGTT TGGCCACCTGGCAACCAACCAGCAAAACGCCACC ACCGTTCCTACCGTAGACGACGTGGACGGAGTCG GCGTGTACCCGGGAATGGTGTGGCAGGACAGAGA CATTTACTACCAAGGGCCCATTTGGGCCAAAATTC CACACACGGATGGACACTTTCACCCGTCTCCTCTC ATTGGCGGATTTGGACTGAAAAGCCCGCCTCCACA AATATTCATCAAAAACACTCCTGTACCCGCCAATC CCGCAACGACCTTCTCTCCGGCCAGAATCAACAGC TTCATCACCCAGTACAGCACCGGACAGGTGGCTGT CAAAATAGAATGGGAAATCCAGAAGGAGCGGTCC AAGAGATGGAACCCAGAGGTCCAGTTCACGTCCA ACTACGGAGCACAGGACTCGCTTCTCTGGGCTCCC GACAACGCCGGAGCCTACAAAGAGCCCAGGGCCA TTGGATCCCGATACCTCACCAACCACCTCTAG (SEQ ID NO: 823)	
MG2	ATGTCGTTCTTTGATTATCTAGGTAAAAAATACGC TCAAGGAGCCGCGGATTGGTGGGACGCGAAAGCA GGACCACCTCCTCCGAAAAAGGCCAAAGGGAAAG GCTTCGTCTTTCCCGGACACAACTACCTAGGTCCC GGCAACGACCTAGACAACGGTAAACCCGTTGACG CGGACGACGCTGCCGCCAAAAAGCACGACGAAGC TTACGACGAGCAGCTCAAACAGGGCGACAACCCG TACGTGAAATATAATCACGCGGACGCCCAGCTGC AAGAAGACCTCAAAGACGTCAATACCTTTGAGGC CAATCTCGCCAAAGGAGTCTTCCAGGCCAAGAAA CGGGTCCTCGAACCGTTTGGTCTGGTAGAAGAACT GGCCGCAAAGAAGCGACCTCACTCTGAGATCGAA TCGCCCGATTCCAGCTCGGGAATCGGCAAGAAGG GAGACCAACCCGCACGACGTCGGTTGGACTTTGAC CAGGAAGACAGACAACAAGCCGCTCAAGACGGAG CCACCGGTACCGGAAGTCTCGAGCCGTCCTCAAAT TTGGGATCTGATACGATGTCTACAGGCGGTGGCGC TCCAATGGCAGGAGATAACCAAGGCGCCGATGGA	MSFFDYLGKKYAQGAADWWDAKAGPP PPKKAKGKGFVFPGHNYLGPGNDLDNG KPVDADDAAAKKHDEAYDEQLKQGDN PYVKYNHADAQLQEDLKDVNTFEANLA KGVFQAKKRVLEPFGLVEELAAKKRPHS EIESPDSSSGIGKKGDQPARRRLDFDQED RQQAAQDGATGTGSLEPSSNLGSDTMST GGGAPMAGDNQGADGVGNASGNWHCD TTWMGDRVVTRSTRTWLLPSYNNHVYR RINAGSNSNAFFGFSTPWGYFDYNRFHC HFSPRDWQRLVNNNWGIRPKHLKFKLFN IQVKEYTVQEGGGETVANNLTSTVQVFT DSDYQLPYVLGSAQEGAMPPFPADVFML PQYAYCTRQNGTNPTPRSAFYCLEYFPS QMLRTGNNFDFSYEFENVPFHSMWAHN QTLDRLMNPLVDQYLWRLQNVANVSGT KTNQFSQNVAGDLQLFTRNWLPGPAYR TQSWSTNTAQNATQTNIPNQENVYFLEG

	GTGGGTAATGCCTCAGGAAATTGGCATTGCGATAC CACATGGATGGGAGACCGAGTGGTCACCCGCTCC ACTCGCACCTGGTTGCTGCCCAGCTACAACAACCA CGTGTACAGAAGAATCAACGCCGGATCAAACAGC AACGCCTTCTTCGGCTTCAGCACCCCGTGGGGGTA CTTTGACTACAACAGATTCCACTGCCACTTTTCTCC ACGAGACTGGCAACGACTCGTCAACAACAACTGG GGCATTCGACCCAAGCACCTCAAGTTCAAACTCTT CAACATACAAGTTAAAGAATACACGGTCCAGGAG GGCGGAGGCGAGACCGTCGCCAATAACCTCACAA GCACCGTGCAGGTCTTTACGGACTCGGACTATCAG CTACCGTACGTGCTAGGTTCGGCGCAAGAGGGTGC CATGCCACCGTTCCCCGCCGACGTGTTTATGTTGC CTCAATACGCGTACTGCACTCGGCAGAACGGAAC CAACCCGACGCCCAGAAGCGCCTTTTACTGTCTGG AATACTTTCCCAGTCAGATGCTGAGAACCGGCAAC AACTTTGACTTCTCGTACGAGTTCGAGAACGTGCC CTTCCACAGTATGTGGGCTCACAACCAGACTCTTG ATCGTCTGATGAATCCTCTGGTGGATCAATACTTG TGGAGACTACAAAATGTAGCTAACGTAAGCGGAA CAAAGACCAATCAGTTCAGCCAGAACGTGGCCGG CGACCTACAACTGTTTACTCGGAACTGGCTACCCG GGCCCGCTTACAGAACGCAGTCTTGGTCGACCAAC ACGGCTCAGAACGCGACTCAGACCAACATCCCGA ATCAGGAAAACGTCTACTTTCTGGAAGGAAGAAC TAACACCATCGCACCCGGCATGGCCATGGGGCCC ACAACCGCTACACACTTTCAATCCAACGGAGTGAT ATTCGCCAAGAACGAATCGGGTACTCCGGCAACT GGTGCCAACATTAACCAGTTGAACATCACACGGG AATCAGAAGTCAACTCGACCAATCCTCTGCTGTCA AGACCACCCCTGGCTATTACCAACAACGAGCAGA CCGCAGCTGCCACGCCAGACAGAGCCAACTGGAA TCTGGCTCCCGTCATGCCCGGACAAGTCTGGATGG ACAAAGACATCTACCTGCAAGGCCAGATCTGGGC CAAGATCCCACACACGGACGGACACTTTCATCCGT CTCCTCTCATGGGAGGATTCGGACTGAAACACCCT CCACCTCAAATCTTCATCAAGAACACTCCCGTACC CGGATCACCCGCTCCCAACTTTACCGTGGACAAGG TCAACAAGTTCATCAACCAATACTCTACCGGTCAA GTAACCGTGGAGATTGAATGGGAACTGCAGAAGG AATCGTCCAAGAGATGGAATCCAGAGATCCAATA CACGGACGACGCATCCGCCACCGACGGATCTATCC TGCACTTTGCTCCCAACGACGAAGGATTGTACACC CAGTTCCGATCTATCGGGACTCGGTACCTCACCCG CCCTCTGTAA (SEQ ID NO: 825)	RTNTIAPGMAMGPTTATHFQSNGVIFAK NESGTPATGANINQLNITRESEVNSTNPL LSRPPLAITNNEQTAAATPDRANWNLAP VMPGQVWMDKDIYLQGQIWAKIPHTDG HFHPSPLMGGFGLKHPPPQIFIKNTPVPGS PAPNFTVDKVNKFINQYSTGQVTVEIEW ELQKESSKRWNPEIQYTDDASATDGSILH FAPNDEGLYTQFRSIGTRYLTRPL (SEQ ID NO: 826)

ZC	ATGTCTTCTTTGTTTAAAGAGTACCTGCAGAGTAC CGGACTGGTGGGTATCCAGTCGGGAGCGCCGAAG CCAAAGGCCGGCCAGCAGAAGCAGGATACAGGGT CCTTTGAGTGGAAAAAGAAGGAGGACACAGCCAG AGGGCTGGTCCTTCCCGGTTACAAGTACCTCGGGC CCTTCAACGGTCTAGAGCGGGGCGAGCCTGTGAAT GCCGCGGACGCCGCGGCCCAGCGACACGACCGAC AGTACGATCGTATTCTACAACAAGGGGGTAATCCA TACCTCACGTACAACCACGCCGACCGAGAGTTCCA GGAGGAGCTCCAGTCTGACGAGTCTTTTGGGGGA AATCTTGGCAAGGCTGTGTTTCAGGCCAAGAAGCG AATCCTAGAGCCCCTCGGTCTGGTTGAAGAGGATC CCTCTCAGCCAGATTCGTCCTCGGACAACACGGCG CCACCCGTGAAAAAGTCCCGTCTCGAAGAAGCCC AGCCCATTCAAAGTCCAGACGTTTCCAGCAGCACT GGCGGAGGAATTGCCGACGTCATGTCTGGAGATG CTGAAATGGCTGCAGTGGGCGGGGGAGCACCGGG CGTCGACGGCCAGGGTGCCGAGGGAGTGGGTACT TCCTCGGGTAATTGGCATTGCGATTCCCAGTGGTC AGAAGGACACGTCAGAACCACCAGCACCAGAACC TGGGTGTTGCCCAGCTACAACAACCACCTGTATAA ACGGCTTGGAAGTAGCGCACAATCCAATACCTAC AACGGATTCTCCACCCCCTGGGGATACCTCGACTT CAATAGATGGCACTGTCACTTCAGTCCTCGGAACT GGCAACGTCTCATCAACAACAACTGGGGCATCAG ACCAAAAAGACTTAATGTTAAATTGTTCAACATAC AAGTCAAAGAGGTCACGACGGAAGGGGGGACGAC GACCGTCGCCAATAACCTTACCAGCACGATTCAGG TGTTTGCGGACAACGCGTACGAACTCCCGTATGTT GTCGACGCGGGTCACGAGGGGGCATTGCCGCCGT TCCCAAACGACGTGTTTATGATTCCCCAATACGGA TACTGCGGGCTGGTGTCTGGTCAAAGTCAGGCTCA GTCGGACTTGTGTTCGTTTTATTGCCTGGAGTATTT TCCATCCCAGATGCTGAGAACAGGAAACAATTTTG AAATGAATTTCAGGTTTGAAGACGTTCCATTTCAC TCCATGTACGCCCACAGCCAGTCTCTAGACAGGCT CATGAATCCTTTAATTGATCAGTATCTGTGGAGTC TAAAACAAACAGGAAACCCGACTGGATCAACAAC GAGAGACTTAAAATTTATCAAGAACAAAGTTCCC AATTTTGCACATTATGGAAAAAATTGGCTTCCTGG ACCTTTTATTAGACAACAGGGGTGGACAACACAA AATATTAATAATAGTGTTGTTAATTTTAATGACAT GCTGGGAAAAAATTCGACATTTACTTTGGACACTA GATGGAGTTCATTAGCGCCTGGTCCGTGTATGGGG GATGACGGACGAACTCCATCCACCACCAAGTTCTC AAATGCTCAGCTCATGTTTGGATCTGGAACACAAC CCACCGAAGGCGGTGAAGATGCTGTACATATTAC ATCCGAGTCGGAGGTCAAGGCAACCAACCCAACT GCAATCGATGAATACGGACGAGTGGCCGATAATA CGCAAAATGCAACAACCGCCCCAACCACAGTGGG AAATGCTGCAATGGGGGCCATGCCTGGGATGGTG TGGCAAGATAGGGATATCTATCTTCAAGGACCCAT CTGGGGAAAAATACCCCATACAGACGGACATTTTC ATCCGTCTCCTCTCATGGGGGGGTTTGGATACAGA	MSSLFKEYLQSTGLVGIQSGAPKPKAGQ QKQDTGSFEWKKKEDTARGLVLPGYKY LGPFNGLERGEPVNAADAAAQRHDRQY DRILQQGGNPYLTYNHADREFQEELQSD ESFGGNLGKAVFQAKKRILEPLGLVEEDP SQPDSSSDNTAPPVKKSRLEEAQPIQSPD VSSSTGGGIADVMSGDAEMAAVGGGAP GVDGQGAEGVGTSSGNWHCDSQWSEG HVRTTSTRTWVLPSYNNHLYKRLGSSAQ SNTYNGFSTPWGYLDFNRWHCHFSPRN WQRLINNNWGIRPKRLNVKLFNIQVKEV TTEGGTTTVANNLTSTIQVFADNAYELP YVVDAGHEGALPPFPNDVFMIPQYGYCG LVSGQSQAQSDLCSFYCLEYFPSQMLRT GNNFEMNFRFEDVPFHSMYAHSQSLDRL MNPLIDQYLWSLKQTGNPTGSTTRDLKFI KNKVPNFAHYGKNWLPGPFIRQQGWTT QNINNSVVNFNDMLGKNSTFTLDTRWSS LAPGPCMGDDGRTPSTTKFSNAQLMFGS GTQPTEGGEDAVHITSESEVKATNPTAID EYGRVADNTQNATTAPTTVGNAAMGA MPGMVWQDRDIYLQGPIWGKIPHTDGH FHPSPLMGGFGYRKPPPQIFIKNTPVPGNP ATTFSPNRINNFITQYSTGQVTVTIDWEL QKENSKRWNPEVQFTSNFGTVDSLNWA PDNAGNYKEPRVVGSRFLTHIL (SEQ ID NO: 828)

	AAACCCCCTCCGCAAATTTTTATTAAAAACACCCC CGTTCCTGGAAATCCGGCAACTACATTTTCTCCAA ATAGAATAAACAATTTCATTACTCAATACTCAACG GGACAAGTGACCGTTACTATTGACTGGGAGCTGCA AAAGGAAAACTCAAAGAGATGGAACCCAGAAGTA CAATTTACATCAAATTTTGGCACGGTCGATTCACT AAACTGGGCACCGGACAACGCGGGAAACTACAAA GAACCAAGGGTGGTGGGGTCTCGATTCCTTACGCA TATACTATAA (SEQ ID NO: 827)	
MA1	ATGTCTTTTTTAGAAGCGTTTGAGGACTGGTACGA AAAAGCAGCGGCTACTTGGCGACATCTCGAACCT GGCGCCCCCCGTCCTAAAGCCAACCAGCAGCATC AGGAACAGCATCATCCGGTTCCTGCTGAAGAGCCT AAAGATCAAGCTCGAGGCCTTGTGCTACCTGGCTA CAAGTACCTCGGCCCCTTTAACGGCCTTGATAAAG GTGATCCGGTTAATCAAGCGGACGCTGCGGCGCTT GAACACGATAAAGCGTACAACCACCTTCTCGAAG AAGGACAAAACCCGTACCTCGTGTATAACGAGGC TGATAAGCATTTCCAGGAAACTCTACAGGGGGAT ACGAGCTTCGGCGGTAACCTCGGCAAAGCAGTCTT CCAAGGCAAGAAACGAGTCCTAGAGCCCTTTGGA CTCGTTGACGAGGAGCCTCACGAACTGGCTCCACC TCCAAAAAAGAAGACAAAACAGGAATACAAAGAC ACTCTAGAGGCTATTCCTCCTCCTAGCAGAGAGGA ACAGACTCCTCAGAAGGGATCTTCAGCATCCAAA AACGGAGCCTCATCTTCAGCAGCAGCAGCAGCCC CCAGTAATTTGGGATCTGGTATCATGGCAGAAGGC GGTGGCGCACCAATGGGCGACAATCAACAGGGTG CCGATGGAGTGGGTAATGCCTCGGGAAATTGGCA TTGCGATTCCCACTGGATGGGCGACAAAGTCGTCA CCCGCACCACCAGAACCTGGGTCCTGCCCACCTAC AACAACCACCTCTACCAAGCCATCGACGAGAATA CTACTCTCGGATCCGCCAACAACTTCTACGGATTC AGCACCCCATGGGGGTACTTTGACTTTAACCGCTT CCACTGCCACTTCTCCCCACGTGACTGGCAAAGAC TCATCAACAACAACTGGGGCATCCGACCCAGAAG CCTCAGATTCAAGCTGTTCAACATCCAAGTCAAGG AGGTCACGACGACGTCGGGCGAAACGACCATCGC CAATAACCTTACCAGCACGGTTCAGGTTTTTGCGG ATGAAGAAATGCACCTCCCGTACGTCTTAGGGAGT GCCCACGACGGGTGCTTTCCTCCATTTCCACCAGA CATTTACATGCTGCCCCAATACGGGTACTGCACTC TGGACTATCAGGGACTCACCACAGACAGGAGCGC CTTCTACTGTCTGGAGTACTTCCCTTCTCAGATGCT GAGAACGGGCAACAACTTTGAGCTTACTTACAACT TTGAGAAGCTGCCTTTCCACAGCATGTGGATGCAC AACCAGTCCCTGGACAGGATCATGAATCCTCTCAT CGATCAGTACCTGTATCGCTTCGCCTCCAAGACTG GCAACACATGGACATACCAAAAAGGCAGCAAAGA CGACAAGACCGCTCAAGCCAGAAACTGGCTTCCT	MSFLEAFEDWYEKAAATWRHLEPGAPR PKANQQHQEQHHPVPAEEPKDQARGLV LPGYKYLGPFNGLDKGDPVNQADAAAL EHDKAYNHLLEEGQNPYLVYNEADKHF QETLQGDTSFGGNLGKAVFQGKKRVLEP FGLVDEEPHELAPPPKKKTKQEYKDTLE AIPPPSREEQTPQKGSSASKNGASSSAAA AAPSNLGSGIMAEGGGAPMGDNQQGAD GVGNASGNWHCDSHWMGDKVVTRTTR TWVLPTYNNHLYQAIDENTTLGSANNFY GFSTPWGYFDFNRFHCHFSPRDWQRLIN NNWGIRPRSLRFKLFNIQVKEVTTTSGET TIANNLTSTVQVFADEEMHLPYVLGSAH DGCFPPFPPDIYMLPQYGYCTLDYQGLTT DRSAFYCLEYFPSQMLRTGNNFELTYNF EKLPFHSMWMHNQSLDRIMNPLIDQYLY RFASKTGNTWTYQKGSKDDKTAQARN WLPGPSIRNQPLYSSGNQTNLSGYDIAPK TEVNEKQMTVFPGLSMATENERSGTIKE QPTSNLLIFANKADAATSQTTSSVGSTAT LLVTNENEVKTVNPSATEAWGSAAANK QTSATASQSVAIHAQGVVPGMVWQDRD IYFQGPIWAKIPHTDGHFHPSPLMGGFGL KHPPPQIFIKNTPVPSNPATTYTTAKYNSF ITQYSTGQVTVEIDWELQKETSKKWNPEI QFTSNYNGTDTLGFAPDAAGDYTEPRAI GTRYLTRPL (SEQ ID NO: 830)

	GGACCCTCCATCAGAAATCAGCCTCTGTACTCGAG TGGTAATCAAACCAACCTCTCGGGATACGATATCG CTCCAAAAACAGAAGTCAACGAGAAACAGATGAC TGTGTTTCCTGGACTTTCTATGGCCACCGAAAACG AGAGGAGCGGAACCATCAAAGAGCAGCCAACGTC CAATCTGCTCATCTTTGCCAACAAAGCGGATGCGG CTACCTCTCAGACCACTTCTAGCGTGGGCAGTACA GCGACACTCCTGGTCACCAACGAAAACGAAGTCA AGACTGTCAATCCCTCCGCCACCGAAGCGTGGGGT TCGGCTGCGGCCAACAAACAGACATCTGCTACCGC CTCTCAAAGCGTGGCTATTCACGCTCAAGGCGTGG TGCCAGGCATGGTCTGGCAAGACAGAGACATCTA CTTTCAGGGCCCCATCTGGGCCAAGATCCCGCATA CAGATGGACACTTTCACCCATCTCCTCTCATGGGT GGATTCGGACTCAAACATCCTCCTCCACAAATTTT CATTAAAAACACTCCTGTCCCCTCCAATCCTGCTA CCACGTACACCACCGCCAAGTACAACTCCTTCATC ACCCAGTACTCCACCGGTCAAGTGACTGTGGAGAT TGACTGGGAACTCCAGAAGGAGACCTCCAAAAAG TGGAACCCCGAGATCCAGTTCACCTCCAATTACAA CGGGACAGACACGCTCGGATTCGCTCCAGATGCC GCTGGAGACTACACCGAACCACGTGCCATCGGAA CCCGATACCTTACCCGACCCCTTTAA (SEQ ID NO: 829)	
CH1	ATGTCTTTTGTTGATCACCCTCCAGATTGGTTGGA AGAAGTTGGTGAAGGTCTTCGCGAGTTTTTGGGCC TTGAAGCGGGCCCACCGAAACCAAAACCCAATCA GCAGCATCAAGATCAAGCCCGTGGTCTTGTGCTGC CTGGTTATAACTATCTCGGACCCGGAAACGGTCTC GATCGAGGAGAGCCTGTCAACAGGGCAGACGAGG TCGCGCGAGAGCACGACATCTCGTACAACGAGCA GCTTGAGGCGGGAGACAACCCCTACCTCAAGTAC AACCACGCGGACGCCGAGTTTCAGGAGAAGCTCG CCGACGACACATCCTTCGGGGGAAACCTCGGAAA GGCAGTCTTTCAGGCCAAGAAAAGGGTTCTCGAA CCTTTTGGCCTGGTTGAAGAGGGTGCTAAGACGGC CCCTACCGGAAAGCGGATAGACGACCACTTTCCA AAAAGAAAGAAGGCTCGGACCGAAGAGGACTCCA AGCCTTCCACCTCGTCAGACGCCGAAGCTGGACCC AGCGGATCCCAGCAGCTGCAAATCCCAGCACAAC CAGCCTCAAGTTTGGGAGCTGATACAATGTCTGCG GGAGGTGGCGGCCCATTGGGCGACAATAACCAAG GTGCCGATGGAGTGGGCAATGCCTCGGGAGATTG GCATTGCGATTCCACGTGGATGGGGGACAGAGTC GTCACCAAGTCCACCCGCACCTGGGTGCTGCCCAG CTACAACAACCACCAGTACCGAGAGATCAAAAGC GGCTCCGTCGACGGAAGCAACGCCAACGCCTACTT TGGATACAGCACCCCCTGGGGGTACTTTGACTTTA ACCGCTTCCACAGCCACTGGAGCCCCCGAGACTGG CAAAGACTCATCAACAACTATTGGGGCTTCAGACC CCGGTCTCTCAGAGTCAAAATCTTCAACATCCAAG TCAAAGAGGTCACGGTGCAGGACTCCACCACCAC CATCGCCAACAACCTCACCTCCACCGTCCAAGTGT TTACGGACGACGACTACCAACTCCCGTACGTCGTC	MSFVDHPPDWLEEVGEGLREFLGLEAGP PKPKPNQQHQDQARGLVLPGYNYLGPG NGLDRGEPVNRADEVAREHDISYNEQLE AGDNPYLKYNHADAEFQEKLADDTSFG GNLGKAVFQAKKRVLEPFGLVEEGAKT APTGKRIDDHFPKRKKARTEEDSKPSTSS DAEAGPSGSQQLQIPAQPASSLGADTMS AGGGGPLGDNNQGADGVGNASGDWHC DSTWMGDRVVTKSTRTWVLPSYNNHQY REIKSGSVDGSNANAYFGYSTPWGYFDF NRFHSHWSPRDWQRLINNYWGFRPRSLR VKIFNIQVKEVTVQDSTTTIANNLTSTVQ VFTDDDYQLPYVVGNGTEGCLPAFPPQV FTLPQYGYATLNRDNGDNPTERSSFFCLE YFPSKMLRTGNNFEFTYSFEEVPFHCSFA PSQNLFKLANPLVDQYLYRFVSTSATGAI QFQKNLAGRYANTYKNWFPGPMGRTQG WNTSSGSSTNRVSVNNFSVSNRMNLEGA SYQVNPQPNGMTNTLQGSNRYALENTMI FNAQNATPGTTSVYPEDNLLLTSESETQP VNRVAYNTGGQMATNAQNATTAPTVGT YNLQEVLPGSVWMERDVYLQGPIWAKIP ETGAHFHPSPAMGGFGLKHPPPMMLIKN TPVPGNITSFSDVPVSSFITQYSTGQVTVE MEWELKKENSKRWNPEIQYTNNYNDPQ FVDFAPDGSGEYRTTRAIGTRYLTRPL (SEQ ID NO: 832)

	GGCAACGGGACCGAGGGATGCCTGCCGGCCTTCC CCCCGCAGGTCTTTACGCTGCCGCAGTACGGCTAC GCGACGCTGAACCGAGACAACGGAGACAACCCGA CAGAGCGGAGCAGCTTCTTTTGCCTAGAGTACTTT CCCAGCAAGATGCTGAGGACGGGCAACAACTTTG AGTTTACCTACAGCTTTGAAGAGGTGCCCTTCCAC TGCAGCTTCGCCCCGAGCCAGAACCTCTTTAAGCT GGCCAACCCGCTGGTGGACCAGTACCTGTACCGCT TCGTGAGCACCTCGGCCACGGGCGCCATCCAGTTC CAAAAGAACCTGGCGGGCAGATACGCCAACACCT ACAAAAACTGGTTCCCGGGGCCCATGGGCCGAAC CCAGGGCTGGAACACGAGCTCTGGCAGCAGCACC AACAGAGTCAGCGTCAACAACTTTTCCGTCTCAAA CCGGATGAACCTGGAGGGGGCCAGCTACCAAGTG AACCCCCAGCCCAACGGGATGACAAACACGCTCC AAGGCAGCAACCGCTACGCGCTGGAAAACACCAT GATCTTCAACGCTCAAAACGCCACGCCGGGAACT ACCTCGGTGTACCCAGAGGACAATCTACTGCTGAC CAGCGAGAGCGAGACTCAGCCCGTCAACCGGGTG GCTTACAACACGGGCGGTCAGATGGCCACCAACG CCCAGAACGCCACCACGGCTCCCACGGTCGGGAC CTACAACCTCCAGGAAGTGCTTCCTGGCAGCGTAT GGATGGAGAGGGACGTGTACCTCCAAGGACCCAT CTGGGCCAAGATCCCAGAGACGGGGGCGCACTTT CACCCCTCTCCGGCCATGGGCGGATTCGGACTCAA ACACCCGCCGCCCATGATGCTCATCAAAAACACGC CGGTGCCCGGCAACATCACCAGCTTCTCGGACGTG CCCGTCAGCAGCTTCATCACCCAGTACAGCACCGG GCAGGTCACCGTGGAGATGGAATGGGAGCTCAAA AAGGAAAACTCCAAGAGGTGGAACCCAGAGATCC AGTACACCAACAACTACAACGACCCCCAGTTTGTG GACTTTGCTCCAGACGGCTCCGGCGAATACAGAAC CACCAGAGCCATCGGAACCCGATACCTCACCCGA CCCCTTTAA (SEQ ID NO: 831)	
SD1	ATGTCGTTTGTTGATCACCCTCCAGATTGGCTTGA GGAGATTGGTGAGGGTCTAAAGGAGTTTTTGGGA CTCGAACCTGGCCCACCCAAACCGAAGCCCAACC AGCAGAAGCAAGACGACGCCCGTGGTCTTGTACT GCCTGGATATAATTACCTGGGACCCGGAAACGGTC TCGACCGCGGAGAACCTGTCAACCGGGCTGACGA GGTCGCGCGAGAGCACGACATCTCGTACAACGAG CAGCTCCAGGCGGGAGACAACCCCTACCTCAAGT ACAACCACGCGGACGCCGAGTTTCAGGAGAAGCT CGCGGACGACACCTCCTTCGGGGGCAACCTCGGC AAGGCAGTCTTTCAGGCCAAAAAAAGGGTTCTCG AACCTTTTGGCCTGGTTGAGGAGCCTGTTAAGACG GCTGCTAAAGGCGAGCGGATAGACGACCACTATC CCAAAAAGAAGAAGGCTCGGATCGAAGAGACCGA AGCTGGAACCAGCGGAGCCCAGCAGCTGCAGATC CCAGCCCAACCAGCCTCAAGTTTGGGAGCTGATAC AATGTCTGCGGGAGGTGGCAGCCCACTGGGCGAC AATAACCAAGGCGCCGATGGAGTGGGCAATGCCT CGGGAGATTGGCATTGCGATTCCACGTGGATGGG GGACCGAGTCATCACCAAGTCCACCCGAACCTGG	MSFVDHPPDWLEEIGEGLKEFLGLEPGPP KPKPNQQKQDDARGLVLPGYNYLGPGN GLDRGEPVNRADEVAREHDISYNEQLQA GDNPYLKYNHADAEFQEKLADDTSFGG NLGKAVFQAKKRVLEPFGLVEEPVKTAA KGERIDDHYPKKKKARIEETEAGTSGAQ QLQIPAQPASSLGADTMSAGGGSPLGDN NQGADGVGNASGDWHCDSTWMGDRVI TKSTRTWVLPSYNNHQYLEIHSGSVDGS NANAYFGYSTPWGYFDFNRFHSHWSPR DWQRLVNNYWGFRPRSLKVKIFNIQVKE VTTQDGTTTIANNLTSTVQVFTDNDYQL PYVIGNGTEGCLPAFPPQVFTLPQYGYAT LNRNNTDDPTERSSFFCLEYFPSKMLRTG NNFEFTYSFEEVPFHCSFAPSQNLFKLAN PLVDQYLYRFVSTDTSGNLQFQKNLKAR YANTYKNWFPGPMCRTQGWYTSAGTY NNKGVANFDTSNKMELEGASYQVNPQP NGMTNTLQDSNKYALENTMIFNAQNAP PGTTSLYQENNLLITSESETQPVNRLAYN

	GTGCTGCCCAGCTACAACAACCATCAGTACCTTGA GATCCACAGCGGTTCCGTCGACGGAAGCAACGCT AACGCTTATTTTGGATACAGCACCCCCTGGGGGTA CTTTGACTTCAACCGCTTCCACAGCCACTGGAGCC CCCGAGACTGGCAGCGACTCGTCAACAACTACTG GGGATTCAGACCCCGGTCCCTCAAGGTCAAGATCT TTAACATCCAAGTCAAGGAAGTCACGACGCAGGA CGGCACCACCACCATCGCCAACAACCTCACCTCCA CCGTCCAAGTGTTTACGGACAACGACTACCAGCTA CCGTACGTCATCGGCAACGGAACGGAGGGGTGCC TGCCGGCCTTCCCTCCGCAGGTCTTTACGCTGCCG CAGTACGGCTACGCGACACTGAACCGTAACAACA CCGACGATCCCACCGAGCGGAGCAGTTTCTTCTGC CTGGAATACTTTCCCAGCAAGATGCTGCGGACGGG CAACAACTTTGAATTCACCTACAGCTTCGAGGAGG TGCCCTTCCACTGCAGCTTCGCTCCCAGCCAGAAC CTCTTCAAGCTGGCCAATCCGCTGGTGGACCAGTA CCTGTACCGCTTTGTGAGCACCGACACTTCCGGTA ACCTACAGTTCCAAAAGAACTTGAAGGCCAGATA TGCCAACACTTACAAGAATTGGTTTCCGGGGCCCA TGTGCCGGACCCAGGGCTGGTACACAAGCGCGGG CACATATAACAACAAAGGCGTTGCCAACTTTGATA CTTCAAACAAGATGGAACTGGAGGGGGCTAGTTA CCAAGTAAACCCTCAACCAAATGGAATGACAAAC ACGCTTCAGGATAGTAACAAATACGCGCTTGAAA ACACCATGATCTTCAACGCACAGAACGCCCCTCCG GGAACGACCTCTCTGTACCAGGAGAACAATCTTTT GATAACCAGCGAGAGCGAGACGCAGCCTGTGAAC CGATTGGCCTACAACACCGGTGGTCAGGTATCAAA CAACAACCAGAATTCAAATACACATCCTACGGTCG GAGTATACAATCACCAGGAAGTGTTGCCTGGTAGC GTGTGGATGGACAGAGACGTATACCTTCAGGGCC CCATCTGGGCCAAAATCCCGGAGACAGGGGCACA CTTTCATCCTTCTCCGGCTATGGGCGGATTCGGAC TCAAACACCCACCGCCCATGATGCTCATCAAGAAC ACACCGGTACCTAGCAACGTCGCTGCCTTCTCTGA CGTGCCCGTTAAAAGCTTCATCACCCAGTACAGCA CCGGACAGGTCACGGTGGAGATTGAATGGGAGCT CAAGAAAGAAAACTCCAAGAGGTGGAATCCCGAG ATACAGTACACCAACAACTACAACAACCCTACATT CGTGGACTTTGCTCCAGACACCTCCGGCGAGTACA GGACTACGAGGGCTATTGGAACCCGTTACCTTACC CGACCCCTGTAA (SEQ ID NO: 833)	TGGQVSNNNQNSNTHPTVGVYNHQEVL PGSVWMDRDVYLQGPIWAKIPETGAHF HPSPAMGGFGLKHPPPMMLIKNTPVPSN VAAFSDVPVKSFITQYSTGQVTVEIEWEL KKENSKRWNPEIQYTNNYNNPTFVDFAP DTSGEYRTTRAIGTRYLTRPL (SEQ ID NO: 834)

SD2	ATGTCGTTTGTTGATCACCCTCCAGATTGGCTTGA GGAGATTGGTGAGGGTCTAAAGGAGTTTTTGGGA CTCGAACCTGGCCCACCCAAACCGAAGCCCAACC AGCAGAAGCAAGACAACGCCCGTGGTCTTGTACT GCCTGGATATAATTACCTGGGACCCGGCAACGGTC TCGACCGCGGAGAACCTGTCAACCGGGCCGACGA GGTCGCGCGAGAGCACGACATCTCGTACAACGAG CAGCTCCAGGCGGGAGACAACCCCTACCTCAAGT ACAACCACGCGGACGCCGAGTTTCAAGAGAAGCT CGCGGACGACACCTCCTTCGGGGGCAACCTCGGC AAGGCAGTCTTTCAGGCCAAGAAAAGGGTTCTCG AACCTTTTGGCCTGGTTGAGGATCCTGTTAAGACG GCTGCTAAAGGCGAGCGGATAGACGACCACTATC CCAAGAAGAAGAAGGCTCGGGTTGAAGAAACCGA AGCTGGAACCAGCGGAGCCCAGCAGCTGCAGATC CCAGCCCAACCAGCCTCAAGTTTGGGAGCTGATAC AATGTCTGCGGGAGGTGGCAGCCCACTGGGCGAC AATAACCAAGGCGCCGATGGAGTGGGCAATGCCT CGGGAGATTGGCATTGCGATTCCACATGGATGGG GGACAGAGTCATCACCAAGTCCACCCGAACCTGG GTGCTGCCCAGCTACAATAACCACCTGTACAAAGA GATCCACAACAACGGGGTCGACGGCAGCACCGCC AACGCTTACTTTGGATACAGTACCCCCTGGGGGTA CTTTGACTTCAACCGCTTCCACAGCCACTGGAGCC CTCGAGACTGGCAGCGACTCATCAACAACTACTGG GGCTTCAGACCCCGGTCCCTCAAGGTCAAGATTTT CAACATCCAAGTCAAAGAGGTCACGGTGCAGGAC AACACCACCACCATCGCCAACAACCTCACCTCCAC CGTCCAAGTGTTTACGGACAACGACTACCAGCTGC CGTACGTCATCGGCAACGGGACCGAGGGGTGTCT ACCGGCCTTCCCTCCGCAGGTCTTTACGCTGCCGC AGTATGGCTACGCGACGCTGAATCGCAACAATAG CGAAAATCCCACCGAGCGAAGCAGCTTCTTCTGTC TGGAGTACTTTCCCAGCAAGATGCTGCGGACGGGC AACAACTTTGAGTTCACATACAGCTTCGAGGAGGT ACCCTTCCACTGCAGCTTCGCGCCCAGCCAAAACC TCTTCAAACTGGCTAACCCGTTGGTGGATCAGTAC CTTTACCGCTTCGTGAGCACGGACACCTCCGGTGC CGTCCAGTTCAAAAAGAACCTGGCGGGCAGATAC GCCAACACCTACAAGAACTGGTTCCCAGGACCCAT GTGCCGAACCCAGGGCTGGTACACAGGAACGGGT ATATATAACAACAAAGGCGCTACCAGCTTTAACAC CTCAAACAGAATGGACCTAGAGGGAGCCAGTTAT CAAGTGCCTCCCCAGCCCAACGGGATGACAAACA CGGTTCAGGACAGCAACCTTTACGCGCTGGAAAA CACCATGATTTTTAACGCACAAAACGCCACCCCGG GAACGAATACAACGTATCCGGAGGAGAACCTTTT GATAACCAGTGAGAGCGAGACTCAACCCGTGAAC AGAGTGGCTTACAGCTCCGGAGGACAAATAGCCA ACAACAATCAGAATACCAACACGGCTCCTACTGC AGGAACCTACAACCACCAGGAAATACTACCTGGC AGCGTGTGGATGGACAGGGACGTGTACCTCCAGG GTCCCATCTGGGCCAAGATCCCAGAGACAGGGGC ACACTTTCATCCTTCTCCAGCCATGGGCGGGTTCG	MSFVDHPPDWLEEIGEGLKEFLGLEPGPP KPKPNQQKQDNARGLVLPGYNYLGPGN GLDRGEPVNRADEVAREHDISYNEQLQA GDNPYLKYNHADAEFQEKLADDTSFGG NLGKAVFQAKKRVLEPFGLVEDPVKTA AKGERIDDHYPKKKKARVEETEAGTSGA QQLQIPAQPASSLGADTMSAGGGSPLGD NNQGADGVGNASGDWHCDSTWMGDR VITKSTRTWVLPSYNNHLYKEIHNNGVD GSTANAYFGYSTPWGYFDFNRFHSHWSP RDWQRLINNYWGFRPRSLKVKIFNIQVK EVTVQDNTTTIANNLTSTVQVFTDNDYQ LPYVIGNGTEGCLPAFPPQVFTLPQYGYA TLNRNNSENPTERSSFFCLEYFPSKMLRT GNNFEFTYSFEEVPFHCSFAPSQNLFKLA NPLVDQYLYRFVSTDTSGAVQFKKNLAG RYANTYKNWFPGPMCRTQGWYTGTGIY NNKGATSFNTSNRMDLEGASYQVPPQPN GMTNTVQDSNLYALENTMIFNAQNATP GTNTTYPEENLLITSESETQPVNRVAYSS GGQIANNNQNTNTAPTAGTYNHQEILPG SVWMDRDVYLQGPIWAKIPETGAHFHPS PAMGGFGLKYPPPMMLIKNTPVPGNITTF SDVPVQSFITQYSTGQVTVEMEWELKKE NSKRWNPEIQYTNNYNNPTFVDFAPDTE GEYRTTRAIGTRYLTRPL (SEQ ID NO: 836)

	GACTCAAATACCCGCCTCCCATGATGCTAATCAAG AACACGCCAGTGCCCGGAAACATCACCACCTTCTC GGACGTGCCCGTCCAAAGTTTTATTACCCAGTACA GCACCGGACAAGTCACCGTGGAGATGGAGTGGGA GCTCAAGAAGGAAAACTCTAAGAGGTGGAACCCC GAGATACAGTACACCAACAACTACAACAACCCTA CGTTCGTGGACTTTGCTCCAGACACAGAAGGAGA ATACAGGACCACTAGGGCTATTGGAACCCGCTACC TTACCCGACCTCTGTAA (SEQ ID NO: 835)	
MR1	ATGTCGTTTTTTGATTACATAGGTAAAAAATACGC TCAAGGAGCGGCCGATTGGTGGAACGCAGAAGCG GGACCACCTCCTCCAAAAAAAGCCAAAGGGAAAG GATTCGTCTTTCCCGGTCACAACTATCTAGGTCCG GGAAACGACCTAGATAACGGGAATCCCGTCGACG AAGACGACGCTGCCGCCAAAAAGCACGACGAGGC TTACGACGAGCAGCTCAAGCAGGGCGACAATCCG TACGTGAAGTATAATCACGCGGACGCTCAGCTTCA GGAAGACCTCAAGGACGTCAATACCTTCGAGGCC AATCTCGCTAAAGGAATCTTTCAGGCCAAAAAAC GGGTCCTAGAACCGTTTGGCCTGGTGGAAGAACTG GCCGGGAAAAAGAGGCCTCATACTGAACAAGAGT CCCCCGATTCCAGCGCAGGAGTCGGAAAGAAAGG CGAACAGCCAGCTAAGCGTCGGTTGGATTTCTCCG ACGCAGAAGCGCCTCAAGACGGAGCAGCAGCAGC GCCCTCTGGTGTGGAAGATATTACAATGTCTGGCG GAGGAGGCGCTCCGATGGACGGAAATCAACAAGG TGCCGATGGAGTGGGCAATGCCTCGGGAAATTGG CATTGCGATTCCCAATGGATGGGGAACAGAGTCGT TACCCATTCCACCAGAACCTGGGTCCTGCCCAGCT ACAACGACCACATCTACCGACAAATCAGCAGCGG AGCCCAGACCGACCACTCCAAGCAATACTTTGGAT ACAGCACCCCCTGGGGGTACTTTGACTTTAACCGA TTCCACTGCCACTTTTCACCGCGAGACTGGCAGCG ATTGGTCAACAACCACTGGGGATTCAGACCCAAG AAGCTGCACGTCAAGCTCTTTAACATCCAAGTCAA GGAGTGGGTCACTCAAGATGGGACTAACGTCGTA TCCAACAACCTCACGAGCACCGTACAAGTGTTCAC GGACGACGAGTACCAGCTCCCCTACGTGCTCGGCA ACGCCACAGAAGGATGTCTCCCCGCCTTCCCTCCG ACGGTCTTTACGCTGCCGCAGTACGGCTACGCGAC GCTGAACGACCCGGGGGACGACGCTCCGACCAAC AGGAGCTCGTTCTTTTGCCTGGAGTACTTTCCCAG TCAGATGCTGAGAACCGGCAACAACTTTGAGTTCA GCTACTCGTTCGAAGACGTGCCCTTTCACACGGGA TGGGCTCCCAACCAGCATCTGTTTAAGACGGCGAA TCCTCTGGTGGACCAGTACCTGTATCAACTGGTCG	MSFFDYIGKKYAQGAADWWNAEAGPPP PKKAKGKGFVFPGHNYLGPGNDLDNGN PVDEDDAAAKKHDEAYDEQLKQGDNPY VKYNHADAQLQEDLKDVNTFEANLAKG IFQAKKRVLEPFGLVEELAGKKRPHTEQE SPDSSAGVGKKGEQPAKRRLDFSDAEAP QDGAAAAPSGVEDITMSGGGGAPMDGN QQGADGVGNASGNWHCDSQWMGNRV VTHSTRTWVLPSYNDHIYRQISSGAQTD HSKQYFGYSTPWGYFDFNRFHCHFSPRD WQRLVNNHWGFRPKKLHVKLFNIQVKE WVTQDGTNVVSNNLTSTVQVFTDDEYQ LPYVLGNATEGCLPAFPPTVFTLPQYGY ATLNDPGDDAPTNRSSFFCLEYFPSQMLR TGNNFEFSYSFEDVPFHTGWAPNQHLFK TANPLVDQYLYQLVGTNNTGDVQFEKV LAGNYGDAFKNWITGANRRTNSWLLGG GNRANVTALNPNYVQIEGDFYKSQPEPA GMANNLSGSDAVALGNSLIFNPQQTAPG TTTRLPAQNVLVTREDETQPVNTYASNS TGYVATNQQSTTSAPQNGQYHFQGIMPG AVWMDRDVYLHGPIWAKIPNTGAHFHP SPMMGGFGLKAPPPMILMKNTPVPGVTQ AFSEVVSNNFINQYSTGQVTVSIEWELQK ENSKRWNPEIQYTNNYNDPQFVDFAPDS TGEYRTTRPIGTRYLTRPL (SEQ ID NO: 838)

	GAACCAACAACACCGGAGACGTTCAGTTTGAGAA AGTGCTGGCCGGTAACTACGGAGACGCGTTCAAG AACTGGATCACGGGAGCCAACAGGCGTACCAACT CTTGGCTGCTGGGCGGAGGAAACAGAGCCAACGT CACCGCTCTGAATCCAAACTACGTCCAGATCGAAG GAGACTTTTACAAGTCGCAGCCGGAACCCGCCGG CATGGCCAACAACCTCAGCGGATCGGACGCGGTT GCCCTGGGCAACAGTCTGATATTCAACCCTCAGCA GACGGCGCCAGGAACTACCACCAGACTGCCGGCG CAGAACGTACTGGTGACCAGAGAAGACGAGACGC AGCCCGTCAACACGTACGCGTCCAACTCTACCGGA TACGTGGCGACCAATCAGCAGAGTACTACGTCGG CTCCTCAGAACGGACAGTACCACTTTCAAGGCATC ATGCCAGGGGCCGTGTGGATGGACAGAGACGTGT ACCTGCACGGACCCATCTGGGCCAAGATTCCGAAT ACGGGAGCGCACTTTCATCCTTCACCCATGATGGG CGGATTCGGATTGAAGGCTCCTCCTCCCATGATCC TGATGAAGAATACGCCCGTTCCCGGAGTGACGCA GGCGTTCAGCGAGGTCGTGTCCAACAACTTCATCA ACCAGTATTCTACGGGTCAGGTGACAGTGTCCATC GAATGGGAGCTGCAGAAGGAAAACAGCAAGCGGT GGAATCCGGAGATTCAGTACACCAACAACTACAA CGACCCCCAGTTTGTGGACTTTGCCCCGGACAGCA CCGGGGAATACAGAACCACCAGACCTATCGGAAC CCGATACCTTACCCGACCCCTTTAA (SEQ ID NO: 837)	
RN1	ATGTCTTTTTTCGATTGGATAGGTCGTAAGTACGC CAACGGCGCCGCTGAGTTCTGGGATCTCGAGCCGG GTCCTCCGAAGCCCAAAAAGGCCAGAGTCGACGA CTCGGCCGGCTTCAACTTCCCCGGGCACAAGTACT TGGGCCCCGGCAACGGTCTCGACCGAGGGGAGCC CGTGGATGCCGACGACGCGGCGGCCCAGAAACAC GACCAGGAATATCAGGCTCTGCTTGAGAGCGGTG AGAATCCTTACCTCACTTTTAATCACGCTGATCGG CAGCTGCAGAAGGACCTCGCGGAGGACCAATCTT TCGAGGGGAACCTCGCCCGGGGTCTGTTCGAAGCC AAGAAGCTCGCGGCACAGGTCGTAGGCGTCGAGG AGCCGGAGCTGGCTCCTCCAGTCAAGCGTCCACAC TCGCCCGAGAAAACCCCAGAGAACCAGAAGGGTC AGCCGCGACCGGATCCCCGGACGCCGGCCAAGAA GAGGCTCGAGTTCTCCGATCAGCCTGGATCTTCAG CAGACTTACCTGCATCCTCACAACAGTCCCAGCCT CCCGCAGGGGTTCCCGGTGTGGTTCCTGGTACGAT GTCTGCAGGAGGAGGCGCTCCAGTGGACGATGCT CAACAAGGTGCCGACGGAGTGGGCAATGCCTCGG GAGATTGGCATTGCGATTCCAAATGGCTGGGCAAC CGAGTTCTCACCCGATCCACCCGGACCTGGGTGCT GCCCAGCTACAACAACCACCTGTACAAGCAGATCT CAGACGCCTCCGGCGTGCACAGCCTCCCCGGGAG CCGATACTTTGGCTACAGCACCCCCTGGGGGTACT TCGACTTCAATCGCTTCCACTGCCACTTCTCGCCCA GAGACTGGCAGCGCCTCGTCAATAACCACTGGGG CTTCCGACCCAAGAGACTGCGAGTCAAACTCTTCA ACATCCAGGTCAAGGAGGTCACGACTACTGATTCG	MSFFDWIGRKYANGAAEFWDLEPGPPKP KKARVDDSAGFNFPGHKYLGPGNGLDR GEPVDADDAAAQKHDQEYQALLESGEN PYLTFNHADRQLQKDLAEDQSFEGNLAR GLFEAKKLAAQVVGVEEPELAPPVKRPH SPEKTPENQKGQPRPDPRTPAKKRLEFSD QPGSSADLPASSQQSQPPAGVPGVVPGT MSAGGGAPVDDAQQGADGVGNASGDW HCDSKWLGNRVLTRSTRTWVLPSYNNH LYKQISDASGVHSLPGSRYFGYSTPWGY FDFNRFHCHFSPRDWQRLVNNHWGFRP KRLRVKLFNIQVKEVTTTDSTTTVSNNLT STVQVFTDDEYQLPYVCGNATEGCLPPF PPDVFTLPQYGYATLNDSNNGNPTNRSS FFCLEYFPSKMLRTGNNFEFTFEFEDVPF HTGFAPCQNLFKLSNPLVDQYLYRFQGT DNSSSSTPGVVKFEKCVAHNYAQQYKN WLTGAFQRNQDYNYVSGTSNYKGVVGS NQNNLQRIENVQFAIAPDVPSMCNHLEG TNMIALDNSLIFRDVSTAPGDTTQYNINQ VIVTSEAETQSVNAYSGDTCGRIVNNSQ NSGTNAGTTGINYKGTMPSSVWMDRDV YLQGPIWAKIPHTGAHFHPSPMMGGFGL RNPPPMMLIKNTPVPGNVTTFTEVKVNQ FITQYSTGQITVDVEWELQKENSKRWNP EIQYTNNYSNNTFVDFAPNANGDYQMT RPIGTRYPTRPP (SEQ ID NO: 840)

	ACGACCACGGTCTCCAACAACCTCACGAGCACGG TCCAGGTCTTCACAGACGACGAGTACCAGCTGCCG TACGTCTGCGGCAACGCCACCGAGGGATGCCTGCC GCCGTTCCCCCCGGACGTCTTCACGCTGCCGCAGT ACGGCTACGCGACGCTGAACGACAGCAACAATGG CAACCCCACCAACAGATCGAGCTTCTTCTGTCTGG AATACTTTCCCAGCAAGATGCTGAGAACCGGGAA CAACTTTGAATTTACCTTCGAGTTTGAAGACGTGC CCTTCCACACGGGCTTCGCCCCCTGTCAGAACCTG TTCAAGCTGAGCAATCCCCTGGTGGACCAGTACCT GTACCGCTTCCAGGGTACTGACAACTCCTCCAGCA GCACGCCCGGGGTGGTCAAGTTCGAGAAATGCGT GGCCCACAACTACGCCCAGCAGTACAAGAACTGG CTGACCGGTGCTTTCCAGAGGAACCAGGACTATAA CTACGTGTCTGGTACTAGCAACTACAAGGGCGTGG TTGGATCTAATCAGAATAACCTGCAGCGCATAGAG AATGTGCAGTTCGCCATTGCTCCGGACGTGCCCAG CATGTGTAACCATCTGGAGGGCACCAATATGATCG CTCTGGACAACAGCCTGATCTTCCGGGATGTCTCT ACCGCCCCTGGTGACACGACTCAGTATAACATCAA CCAGGTGATCGTGACCAGCGAAGCCGAGACTCAA TCTGTCAATGCGTACTCGGGCGACACCTGTGGCCG CATTGTGAATAACAGCCAGAATAGCGGAACCAAC GCTGGGACCACCGGCATCAACTACAAAGGTACCA TGCCGTCCTCGGTCTGGATGGACCGTGACGTCTAC CTGCAGGGCCCGATCTGGGCCAAGATTCCCCACAC GGGGGCGCACTTTCACCCGTCTCCCATGATGGGGG GCTTCGGCCTGAGGAACCCTCCTCCGATGATGCTG ATCAAGAACACGCCCGTTCCTGGCAACGTGACCAC CTTCACCGAGGTCAAGGTTAACCAGTTCATCACCC AGTACTCGACCGGGCAGATCACCGTGGACGTGGA ATGGGAGCTGCAGAAGGAGAACTCCAAACGATGG AACCCCGAGATCCAGTACACAAATAACTACAGTA ATAATACTTTTGTGGACTTTGCCCCCAATGCTAAT GGAGATTACCAGATGACCCGCCCCATTGGAACCC GCTACCCGACCCGCCCCCCGTAA (SEQ ID NO: 839)	
RN2	ATGTCTTTTTTCGATTGGATAGGTCGTAAGTACGC CAACGGCGCCGCTGAGTTCTGGGACCTCGAGCCG GGTCCTCCGAAGCCCAAAAAGGCCAGAGTCGACG ACTCGGCCGGCTTCAACTTCCCCGGGCACAAGTAC CTGGGCCCCGGCAACGGTCTCGACCGAGGCGAGC CCGTGGATGCCGACGACGCGGCGGCCCAGAAGCA CGACCAGGAATATCAGGCTCTGCTTGAGAGCGGC GAGAATCCTTACCTCACTTTTAATCACGCTGATCG GCAGCTGCAGAAGGACCTCGCGGAGGACCAATCT TTCGAGGGGAACCTCGCCCGGGGCCTGTTCGAAGC CAAGAAGCTCGCGGCGCAGGTCGTAGGCGTCGAG GAGCCTGAGCTGGCTCCTCCAGTCAAGCGTCCACA CTCTCCCGAGAAGACCCCGGAAGACCAGAAGGGT CAGCCGCGACCGGATCCCCGGACGCCGGCCAAGA AGCGGCTCGAGTTCTCCGATCAGCCTGGATCTGCA GCAGACTTACGTACATCCTCACAACAGTCCCAGCC TCCCGCAGGGGTTCCCGGTGTGGTTTCTGGTACGA TGTCTGCAGGAGGAGGCGCTCCAGTGGACGATGC	MSFFDWIGRKYANGAAEFWDLEPGPPKP KKARVDDSAGFNFPGHKYLGPGNGLDR GEPVDADDAAAQKHDQEYQALLESGEN PYLTFNHADRQLQKDLAEDQSFEGNLAR GLFEAKKLAAQVVGVEEPELAPPVKRPH SPEKTPEDQKGQPRPDPRTPAKKRLEFSD QPGSAADLRTSSQQSQPPAGVPGVVSGT MSAGGGAPVDDAQQGADGVGNASGDW HCDSKWLGNRVLTRSTRTWVLPSYNNH LYKQISDASGVHGLPGSRYFGYSTPWGY FDFNRFHCHFSPRDWQRLVNNHWGFRP KRLRVKLFNIQVKEVTTTDSTTTVSNNLT STVQVFTDDEYQLPYVCGNATEGCLPPF PPDVFTLPQYGYATLNDSNNGNPTNRSS FFCLEYFPSKMLRTGNNFEFTFEFEDVPF HTGFAPCQNLFKLSNPLVDQYLYRFQGT DASTSSTPGIVKFEKCVAHNYAQQYKN WLTGAFQRNQDYNYVSGTTNYKGIVGS

	TCAACAAGGTGCCGACGGAGTGGGCAATGCCTCG GGAGATTGGCATTGCGATTCCAAATGGCTGGGCA ACCGAGTTCTCACCCGATCCACCCGGACCTGGGTG CTGCCCAGCTACAACAACCACCTGTACAAGCAGAT CTCCGACGCCTCCGGCGTGCACGGCCTCCCCGGGA GTCGATACTTTGGCTACAGCACCCCCTGGGGGTAC TTCGACTTCAATCGCTTCCACTGCCACTTCTCGCCC AGAGACTGGCAGCGCCTCGTCAACAACCACTGGG GCTTCCGACCCAAGCGACTGCGAGTCAAGCTCTTC AACATCCAGGTCAAGGAGGTCACGACTACTGATTC GACGACCACGGTCTCCAACAACCTCACGAGCACG GTCCAGGTCTTCACAGACGACGAGTACCAGCTGCC GTACGTCTGCGGCAACGCCACCGAGGGATGCCTG CCGCCGTTCCCTCCGGACGTCTTCACGCTGCCGCA GTACGGCTACGCGACGCTGAACGACAGCAACAAC GGCAACCCCACCAACAGATCGAGCTTTTTCTGTCT GGAATACTTTCCCAGCAAGATGCTGAGGACCGGG AACAACTTTGAATTCACCTTCGAGTTCGAGGACGT GCCTTTCCACACGGGCTTCGCCCCCTGTCAGAACC TGTTCAAGCTGAGTAACCCCCTGGTGGACCAGTAC CTGTATCGCTTCCAAGGAACAGATGCCTCCACTAG CAGTACTCCCGGAATCGTCAAGTTCGAGAAATGCG TAGCTCATAATTACGCCCAGCAGTACAAGAACTGG CTGACTGGAGCTTTCCAGAGGAATCAGGACTACA ACTATGTGTCTGGTACCACTAATTATAAAGGAATC GTGGGATCCAACCAGAATAACTTGCAGCGCATAG AGAACGTTCAGTTTGCCATCGCTCCGGACGTGCCC AGCATGTGTAACCATCTCGAAGGCACCAATATGAT CGCTCTGGATAATAGCCTTATCTTCAGAGATGTGT CTACCGCCCCCGGCGACACGACTCAGTATAACATC AACCAGGTGATCGTGACCAGCGAGGCCGAGACTC AACCTGTCAATTCGTACTCGGGCGACAACTGTGGC CGCATTGTGAATAACAGCCAGAATAACGGGACCA ACGCCGGGACTACCGGCATCAACTACAAAGGTAC CATGCCTTCCTCGGTCTGGATGGACCGTGACGTCT ACCTGCAGGGCCCGATCTGGGCCAAGATTCCCCAC ACGGGGGCGCACTTTCACCCGTCTCCCATGATGGG GGGCTTCGGCCTGAAGAACCCTCCCCCGATGATGC TGATCAAGAACACGCCCGTTCCTGGCAACGTGACC ACCTTCACCGAGGTCAAGGTTAACCAGTTCATCAC CCAGTACTCGACCGGGCAGATCACCGTGGACGTG GAATGGGAGCTGCAGAAGGAGAACTCCAAGAGAT GGAACCCCGAGATCCAGTACACAAACAACTACAG CAATAACACTTTTGTGGACTTTGCCCCCAATGCCA ATGGAGATTACCAGATGACCCGCCCTATCGGAACC CGATACCTTACCCGACCCCTTTAA (SEQ ID NO: 841)	NQNNLQRIENVQFAIAPDVPSMCNHLEG TNMIALDNSLIFRDVSTAPGDTTQYNINQ VIVTSEAETQPVNSYSGDNCGRIVNNSQN NGTNAGTTGINYKGTMPSSVWMDRDVY LQGPIWAKIPHTGAHFHPSPMMGGFGLK NPPPMMLIKNTPVPGNVTTFTEVKVNQFI TQYSTGQITVDVEWELQKENSKRWNPEI QYTNNYSNNTFVDFAPNANGDYQMTRPI GTRYLTRPL (SEQ ID NO: 842)

RN3	ATGTCTTTCTTCGATTGGATAGGTCGTAAGTACGC CAACGGCGCCGCTGAGTTCTGGGACCTCGAGCCG GGTCCTCCGAAGCCCAAAAAGGCCCGAGTCGACG ACTCGGCCGGCTTCAACTTCCCCGGGCACAAGTAC CTGGGCCCCGGCAACGGTCTCGACCGCGGGGAGC CCGTGGATGCCGACGACGCGGCGGCCCAGAAGCA CGACCAGGAATATCAGGCTCTGCTTGAGAGCGGC GAGAATCCTTACCTCACTTTTAATCACGCTGATCG GCAGCTGCAGAAGGACCTCGCGGAGGACCAATCT TTCGAGGGGAACCTCGCCCGGGGCCTGTTCGAAGC CAAGAAGCTCGCGGCGCAGGTCGTAGGCGTCGAG GAGCCGGAGCTGGCTCCTCCAGTCAAGCGTCCACA CTCGCCCGAGAAGACCCCGGAGGACCAGAAGGGT CAGCCGCGACCGGATCCCCGGACGCCGGCAAAGA AGCGGCTCGAGTTCTCCGATCAGCCTGGATCTTCA GCAGACTTACCTGCATCCTCACAACAGTCCCAGCC TCCCGCAGGGGTTCCCGGTGTGGTTCCTGGTACGA TGTCTGCAGGAGGAGGCGCTCCAGTGGACGATGC TCAACAAGGTGCCGACGGAGTGGGCAATGCCTCG GGAGATTGGCATTGCGATTCCAAATGGCTGGGCA ACCGAGTTCTCACCCGATCCACCCGGACCTGGGTG CTGCCCAGCTACAACAACCACCTGTACAAGCAGAT CTCCGACGCCTCCGGCGTGCACGGCCTCCCCGGGA GCCGATACTTTGGCTACAGCACCCCCTGGGGGTAC TTCGACTTCAATCGCTTCCACTGCCACTTCTCGCCC AGAGACTGGCAGCGACTCGTCAACAACCACTGGG GCTTCCGACCCAAGCGTCTCCGAGTCAAGCTCTTC AACATCCAGGTCAAGGAGGTCACGACTACTGATTC GACGACCACGGTCTCCAACAACCTCACGAGCACG GTCCAGGTCTTCACAGACGACGAGTACCAGCTGCC GTACGTCTGCGGCAACGCCACCGAGGGATGCCTG CCGCCGTTCCCTCCGGACGTCTTCACGCTGCCGCA GTACGGCTACGCGACGCTGAACGACAGCAACAAC GGCAACCCCACCAACAGATCGAGCTTCTTCTGTCT GGAATACTTTCCCAGCAAGATGCTGAGGACCGGG AACAACTTTGAATTCACCTACGAGTTCGAGGAAGT GCCCTTCCACACGGGCTTCGCCCCCTGTCAGAACC TGTTCAAGCTGAGTAACCCCCTGGTGGACCAGTAC CTGTATCGCTTCCAAGGAACAGATGCCTCAACTAG CAGCACTCCCGGAATCGTCAAGTTCGAGAAATGC GTAGCCCATAATTACGCCCAGCAGTACAAGAACT GGCTGACTGGAGCTTTCCAGAGGAACCAGGACTA CAACTATGTGTCTGGCACCACTAATTATAAAGGCA TCGTGGGATCTAATCAGAATAACCTGCAGCGCATA GAGAATGTGCAGTTCGCCATTGCTCCGGACGTGCC CAGCATGTGTAACCATCTGGAGGGCACCAACATG ATTGCCCTGGACAACAGCCTGATCTTCCGAGATGT CTCTACCGCCCCCGGTGACACGACTCAGTATAACA TCAACCAGGTGATCGTGACCAGCGAGGCCGAGAC TCAACCCGTCAATGCGTACTCGGGCGATAACTGTG GCCGCATTGTGAATAATAGCCAGAATAACGGGAC CAACGCCGGGACCACCGGCATCAACTACAAAGGT ACCATGCCTTCCTCGGTCTGGATGGACCGTGACGT CTACCTGCAGGGCCCGATCTGGGCCAAGATTCCCC	MSFFDWIGRKYANGAAEFWDLEPGPPKP KKARVDDSAGFNFPGHKYLGPGNGLDR GEPVDADDAAAQKHDQEYQALLESGEN PYLTFNHADRQLQKDLAEDQSFEGNLAR GLFEAKKLAAQVVGVEEPELAPPVKRPH SPEKTPEDQKGQPRPDPRTPAKKRLEFSD QPGSSADLPASSQQSQPPAGVPGVVPGT MSAGGGAPVDDAQQGADGVGNASGDW HCDSKWLGNRVLTRSTRTWVLPSYNNH LYKQISDASGVHGLPGSRYFGYSTPWGY FDFNRFHCHFSPRDWQRLVNNHWGFRP KRLRVKLFNIQVKEVTTTDSTTTVSNNLT STVQVFTDDEYQLPYVCGNATEGCLPPF PPDVFTLPQYGYATLNDSNNGNPTNRSS FFCLEYFPSKMLRTGNNFEFTYEFEEVPF HTGFAPCQNLFKLSNPLVDQYLYRFQGT DASTSSTPGIVKFEKCVAHNYAQQYKN WLTGAFQRNQDYNYVSGTTNYKGIVGS NQNNLQRIENVQFAIAPDVPSMCNHLEG TNMIALDNSLIFRDVSTAPGDTTQYNINQ VIVTSEAETQPVNAYSGDNCGRIVNNSQ NNGTNAGTTGINYKGTMPSSVWMDRDV YLQGPIWAKIPHTGAHFHPSPMMGGFGL KNPPPMMLIKNTPVPGNVITFTEVKVNQF ITQYSTGQITVDVEWELQKENSKRWNPEI QYTNNYSNNTFVDFAPNANGDYQMTRPI GTRYLTRPL (SEQ ID NO: 844)

	ACACGGGGGCGCACTTTCACCCGTCTCCCATGATG GGGGGCTTCGGCCTGAAGAACCCTCCTCCGATGAT GCTGATCAAGAACACGCCCGTTCCTGGCAACGTGA TCACCTTCACCGAGGTCAAGGTCAACCAGTTCATC ACCCAGTACTCGACCGGGCAGATCACCGTGGACG TGGAATGGGAGCTGCAGAAGGAGAACTCCAAGAG ATGGAACCCCGAGATCCAGTACACAAACAACTAC AGCAACAACACTTTTGTGGACTTTGCCCCCAATGC CAATGGAGATTACCAGATGACCCGCCCTATCGGA ACCCGATACCTTACCCGACCCCTTTAA (SEQ ID NO: 843)	
MM1	ATGTCTTTCTTTGATTGGGTAGGTAAAAAGTACGC TCAAGGAGCTGCTGAATTCTGGGATCTGAAGTCCG GTCCTCCTGCACCAAAAAAGGCGCGAAAGGACGG ATCTGCCGGATTCAATTTCCCCGGGCATAAATATC TCGGACCCGGTAACTCTCTCGATCGAGGAGATCCG GTTGACGCCGACGACGCCGCTGCTCAAAAGCACG ATCAAGCTTATCAAGAGCAGCTTGAAGCGGGAGA TAATCCCTATCTGAAATACAATCACGCCGATCGCG AGTTCCAAGAGGCGTTGAAAGACGATACCTCGTTT GAAGGAAATCTCGCGAGAGGACTCTTTGAGGCCA AGAAGCTCGTGGCCGAGCCTCTAGGTCTCGTTGAA CCGGAACTGGCTCCGGGGAAGAAACGCCCAGCGG CTTCCCCCGTTCAATCGCCAGACTCGTCTTCCGGT ATCGGCAAGAAGGGCGCGCAACCTTCCCGCCGAA GATTGGACTTTGACCAGGCCGATCGAGAAACCGCT GCCGAGGCTGGAGCCACAGGCACCGGAAGTCTAC CACCCGCCTCTAATTTGGGACCTGTTACAATGTCT GCTGGGGGAGGCGCACCGATGGACGACAATCAAC AAGGTGCCGATGGAGTGGGCAATGCCTCGGGAAA TTGGCATTGCGATTCCCAATGGCTGGGCGACAGAG TCATTACAAAGTCCACCCGAACCTGGAGCCTGCCC ACCTACAACAACCACTTGTACAAGAAGATCTCCAG CAACGAAGCTTCAGGAGACAAGACTTGGTTTGGA TTCAGTACCCCTTGGGGATACTTTGACTTCAACCG CTTCCACTGTCACTTCTCCCCCCGAGACTGGCAGC GTCTCATCAACAACCACTGGGGCATCCGCCCGAAG AAGATGCACGTCAAGCTCTTCAATATTCAAGTTAA GGAGGTCACGACGCAAGACTCGACGACGACCATC GCTAATAACCTTACCAGTACGGTTCAGGTCTTTGC GGACTCGGAATATCAGCTGCCGTACGTAGCCGGG AATGCGCACGAAGGATGCTTGCCACCTTTTCCAGC TGATGTCTTTATGCTGCCGCAGTACGGCTACTGCA CTCTCCAGAACAGCGCTTCCAACGATCCAGTGGCC CAGAGTGCCTTCTACTGTCTGGAATATTTCCCCAG TCAGATGCTGAGAACGGGAAATAACTTTGTCTTCA GCTACGAGTTCGAAGACGTCCCGTTCCACAGTATG	MSFFDWVGKKYAQGAAEFWDLKSGPPA PKKARKDGSAGFNFPGHKYLGPGNSLDR GDPVDADDAAAQKHDQAYQEQLEAGD NPYLKYNHADREFQEALKDDTSFEGNLA RGLFEAKKLVAEPLGLVEPELAPGKKRP AASPVQSPDSSSGIGKKGAQPSRRRLDFD QADRETAAEAGATGTGSLPPASNLGPVT MSAGGGAPMDDNQQGADGVGNASGN WHCDSQWLGDRVITKSTRTWSLPTYNN HLYKKISSNEASGDKTWFGFSTPWGYFD FNRFHCHFSPRDWQRLINNHWGIRPKKM HVKLFNIQVKEVTTQDSTTTIANNLTSTV QVFADSEYQLPYVAGNAHEGCLPPFPAD VFMLPQYGYCTLQNSASNDPVAQSAFYC LEYFPSQMLRTGNNFVFSYEFEDVPFHS MYMHNQSLDRLMNPLIDQYLYYLSATN APANTNKFTKSNKGDLPSQLRNWIPGPSF QQAAWTLDGQSQTQTNNTWDYANKYQ LDARKNKLAPGPVNAPREGQFQANTLIF GKDPTVSSTLAEQDKVNITRDNETRTVN PILGTAQIFDIINNDTPRANGSISKINMNIC PVMPGSVWQDRDIYLQGPIWAKVPRTD GKFHPSPLMGGFGLRCPPPQIFIKNTPVPG NPPTTFEPTKVASFITQYSTGQVTVEMEW ELEKENSKRWNPEIQFTNNFSSSDGNMV HFAVDSTGLYSEPRPVGTRYLTRSL (SEQ ID NO: 846)

	TACATGCACAACCAATCGTTGGACCGCTTGATGAA CCCGCTGATTGATCAGTACTTGTACTATCTCTCCGC CACCAATGCTCCCGCCAATACAAATAAATTTACCA AGTCAAATAAAGGTGATCTACCGTCTCAGCTCAGG AACTGGATTCCTGGACCTTCATTTCAACAAGCAGC TTGGACTCTAGACGGACAGAGTCAAACCCAAACG AACAACACGTGGGATTATGCCAATAAATACCAATT GGATGCACGAAAGAACAAGTTGGCTCCCGGTCCA GTTAATGCACCTCGAGAAGGACAGTTTCAAGCAA ATACGCTGATATTCGGCAAAGATCCAACCGTCTCT TCTACCTTGGCTGAACAAGACAAAGTTAACATTAC ACGAGACAATGAAACTCGGACAGTAAACCCAATT CTGGGAACGGCTCAGATATTTGACATCATCAACAA CGACACTCCGCGTGCGAACGGCTCAATAAGCAAA ATCAACATGAACATCTGTCCTGTCATGCCTGGAAG CGTCTGGCAAGATCGAGATATCTACCTACAAGGAC CAATCTGGGCCAAGGTTCCTCGCACAGATGGAAA GTTCCATCCCTCTCCTCTCATGGGTGGATTTGGACT TAGATGTCCACCTCCACAGATTTTCATTAAGAATA CTCCGGTTCCTGGCAACCCTCCAACAACATTTGAA CCAACCAAAGTGGCCAGCTTCATCACGCAGTACTC GACCGGACAAGTCACCGTCGAAATGGAATGGGAG CTTGAAAAGGAAAACTCCAAACGCTGGAACCCGG AAATACAGTTCACCAACAACTTCTCCTCTTCGGAC GGAAACATGGTCCACTTTGCTGTGGACTCTACCGG ACTCTACTCAGAACCCCGTCCCGTGGGCACCCGTT ACCTTACCCGCTCTTTGTAA (SEQ ID NO: 845)	
MG1	ATGTCTTTTGTCGATCACCCTCCAGATTGGTTGGA AGACGTTCTCGACGGCTTCAACGAATTTCTGGGTC TCAAATCGGGACCACCTCAGCCAAAGGCCGGGCA GTACGATAAGCCCAAAGACAACACCAGAGGCCTA GTCCTTCCTGGCTACAAATACTTGGGTCCGGGCAA CGGACTCGATAAAGGAGAGCCCGTCAACGAGGCT GACGCCGCTGCTCTCGAGCACGACAAGGCCTACG ACGAGCAGCTTAAAGCGGGAGACAACCCGTACGT CAAATACAATCACGCAGACGCCGAATTCCAAGAG CGTCTGCAGGGAGATACGTCTTTTGGCGGGAACCT TGCTCGTGGTGTATTCCAGGCCAAGAAGCGGGTCC TAGAGCCGCTCGGTCTGGTCGAAGGAGACGAGAC GGGTCCCATCAAGAGACCCCATACGCAAGAGAGT CCCGAGTCCAGTTCCGGGATCGGAAAGAAGGGCC ATCAGCCTGCCCGCAAGAGGCTCAACTTTGATCAA GCTGGAGAATCGAGCTCTGGTGCCGCAGGACAGT CTAACGGAAGCATGTCTGCAGACGGTACTGAGCG TGCAGCAGCTGGCGGAGCTGCGGTCGATGCAGCT CAAGGTGCCGATGGAGTGGGTAATTCCTCCGGTAA TTGGCATTGCGATTCCACGTGGGGAGATAATCACG TCACCACCACCAGCACCAGAACCTGGGTGCTGCCC AACTACAACAACCACCTGTACCGAAGACTCGGCG TCAGCAACACCTCCAACACCTACAACGGATTCAGC ACCCCCTGGGGGTACTTTGACTTTAACCGATTCCA CTGCCACTTCTCTCCACGAGACTGGCAGCGACTCA TCAATAACCATTGGGGAATCAGACCCAAACGCAT CAGCGTCAAAATCTTCAACATTCAAGTCAAGGAG	MSFVDHPPDWLEDVLDGFNEFLGLKSGP PQPKAGQYDKPKDNTRGLVLPGYKYLG PGNGLDKGEPVNEADAAALEHDKAYDE QLKAGDNPYVKYNHADAEFQERLQGDT SFGGNLARGVFQAKKRVLEPLGLVEGDE TGPIKRPHTQESPESSSGIGKKGHQPARK RLNFDQAGESSSGAAGQSNGSMSADGTE RAAAGGAAVDAAQGADGVGNSSGNWH CDSTWGDNHVTTTSTRTWVLPNYNNHL YRRLGVSNTSNTYNGFSTPWGYFDFNRF HCHFSPRDWQRLINNHWGIRPKRISVKIF NIQVKEVTVQDGTTTVANNLTSTIQVFA DKAYDLPYVMDAGQEGSLPPFPNDVFMI PQYGYMGLVTGGSSQNQTDRNAFYCLE YFPSQMLRTGNNFEMSFDFEDVPFHSMF AHSQSLDRLMNPLLDQYLWSLHSTTGG AYGTGTKLNTFIKNQTTDFSSYRRNWLP GPFIKQQRWNKGNTNYSVPVSGNDSML HHESHTTLNGRWSKLAPGPAMATPENT ATSFTQGQLMFAGPGQTTTSTTNSNNIF TSEDELIATNPTDYDYWGAVTDGQQTST VAPSERDVDAMPTMPGMVWQDRDIYY QGPIWAKIPHTDGRFHPSPLIGGFGLHKP PPQIFIKNTPVPANPATTFSATRINTFLNQ YSTGQVAVQIEWELRKEKSKRWNPEIQF TSNYGTQNSMLWAPDAAGNYEEPRAIG TRYLTHCL (SEQ ID NO: 848)

	GTCACGGTGCAGGACGGCACCACGACCGTCGCCA ACAATCTCACCAGCACCATACAGGTCTTTGCGGAC AAGGCGTACGACCTGCCGTACGTGATGGACGCCG GACAGGAGGGCAGTCTGCCTCCGTTCCCCAACGAC GTCTTTATGATACCGCAGTACGGCTACATGGGACT GGTGACAGGCGGCAGCAGCCAGAACCAGACGGAC AGAAACGCCTTCTACTGTCTAGAGTACTTTCCCTC TCAGATGCTGAGAACGGGAAACAACTTTGAAATG AGCTTCGACTTTGAAGACGTGCCCTTCCACTCCAT GTTCGCTCACAGCCAGAGTTTGGACAGACTGATGA ATCCTCTGCTGGATCAGTACCTGTGGAGTCTGCAC TCGACCACAGGGGGAGCCTATGGCACCGGCACCA AGCTCAATACGTTCATCAAGAACCAGACCACAGA CTTTTCGTCTTATAGACGCAACTGGCTGCCTGGTC CTTTCATCAAACAGCAGAGATGGAACAAGGGCAA TACCAACTACTCGGTGCCCGTCTCTGGAAACGACT CGATGCTCCACCACGAGTCTCACACGACTCTGAAC GGTCGATGGAGCAAACTGGCTCCCGGACCGGCCA TGGCAACTCCAGAAAACACGGCTACCAGCTTTACA CAAGGCCAGCTCATGTTTGCCGGGCCCGGGCAGA CCACTACCAGTACCACCAACTCCAACAACATCATC TTCACCTCGGAGGACGAGCTGATTGCCACTAATCC CACCGACTACGATTACTGGGGAGCCGTCACAGAC GGACAGCAGACCTCTACCGTGGCACCCAGCGAAC GAGACGTGGACGCGATGCCGACTATGCCCGGCAT GGTCTGGCAAGACAGAGACATTTACTACCAGGGC CCCATCTGGGCCAAGATTCCTCATACCGACGGACG CTTCCATCCGTCTCCTTTGATTGGAGGCTTCGGTCT GCACAAGCCTCCTCCTCAAATCTTCATCAAAAACA CGCCGGTACCTGCCAATCCAGCAACCACCTTCTCG GCTACTCGCATCAACACCTTCCTGAACCAGTACAG CACCGGACAAGTCGCCGTCCAGATCGAATGGGAG CTGCGCAAAGAAAAGTCCAAGCGATGGAACCCGG AAATCCAGTTCACTTCCAACTACGGTACCCAGAAT TCCATGCTCTGGGCGCCAGATGCCGCTGGAAACTA CGAAGAGCCCAGAGCCATCGGTACCCGCTACCTC ACCCATTGCCTGTAA (SEQ ID NO: 847)	
MM2	ATGTCTTTCTTTGATTGGTTAGGTAAACAGTACGC TCAAGGAGCTGCTGAATTCTGGGATTTGAAGTCCG GTCCTCCTGCACCAAAAAAGGCGCGAAAGGACGG ATCAGCCGGATTCAATTTCCCTGGGCACAAATACC TGGGTCCTGGCAATTCATTGGATCGCGGAGATCCC GTGGACGCTGACGACGCCGCTGCTCAAAAGCACG ACCAGTCGTACCAAGAGCAGCTTGAGGCGGGAGA CAATCCCTACCTCAAGTACAACCACGCCGACCGCG AGTTCCAGGAGGCGTTGAAGGACGACACCTCCTTT GAAGGAAATCTCGCGAGAGGACTCTTTGAGGCCA AGAAGCTCGTGGCAGAGCCTCTTGGTCTCGTGGAA CCAGAACTGGCGCCACCCAGTGGCCGTAAACGAC CGGTGCAATCGAGTCAAGAGTCTGGTTACTCGAGT AGCCAAGACAAGCGGCCCAACCTCGACGTAGACG AGGAGGACCGTGAGTTCGCTGCCGCTGCAGCGGA GACCGAAACTGGAAGCGCTCCCCCCACCGGCAAT TTGGGACCTGGTACGATGGCTGGAGGCGGTAGCG	MSFFDWLGKQYAQGAAEFWDLKSGPPA PKKARKDGSAGFNFPGHKYLGPGNSLDR GDPVDADDAAAQKHDQSYQEQLEAGD NPYLKYNHADREFQEALKDDTSFEGNLA RGLFEAKKLVAEPLGLVEPELAPPSGRKR PVQSSQESGYSSSQDKRPNLDVDEEDREF AAAAAETETGSAPPTGNLGPGTMAGGGS APIDDGSYGADGVGNASGDWHCDSTWL DNCVITRTTRTWNLPTYNNHIYKRLNGT TSGDQSYFGFSTPWGYFDFNRFHCHFSPR DWQRLINNNWGLRPKSLRFKIFNIQVKE VTTQDSTKIISNNLTSTVQVFADTEYQLP YVIGSAHEGCLPPFPADVFMLPQYGYCT RQDGNSNNPTPRSAFYCLEYFPSKMLRT GNSFEFTYNFEKVPFHSMWAHNQSLDRL MNPLIDQYLYYLDVTSSTGFTYQKGVHT NLPEQERNWLPGPGIRNQAWFNSATGNN

	CGCCAATCGACGACGGCTCGTATGGTGCCGATGG AGTGGGCAATGCCTCGGGAGATTGGCATTGCGATT CCACATGGCTGGACAACTGTGTCATCACCCGAACC ACTCGGACCTGGAATCTGCCAACCTACAACAACCA CATCTACAAACGACTCAACGGAACGACCTCCGGA GACCAAAGCTACTTCGGATTCAGCACCCCCTGGGG ATACTTTGACTTCAACCGCTTCCACTGTCATTTCTC CCCTCGAGACTGGCAAAGACTCATCAACAATAACT GGGGACTCCGACCAAAGAGCCTACGGTTCAAAAT CTTTAACATTCAAGTTAAAGAAGTCACGACGCAAG ACTCAACGAAGATCATCTCCAATAACCTTACCAGC ACGGTTCAGGTATTTGCGGACACGGAGTACCAGCT CCCGTACGTGATTGGATCGGCTCACGAAGGATGTC TGCCTCCTTTCCCGGCTGACGTGTTCATGCTGCCGC AGTATGGATACTGTACTCGACAAGACGGAAACAG CAACAATCCGACCCCGAGAAGCGCCTTCTATTGTT TGGAGTACTTTCCCAGCAAGATGCTAAGAACTGGG AACAGTTTTGAATTTACATATAACTTTGAGAAGGT GCCGTTCCACAGCATGTGGGCTCACAACCAGAGTC TGGATCGATTGATGAATCCATTGATTGATCAGTAC CTGTACTACCTTGATGTCACTTCGAGTACCGGGTT TACCTATCAAAAAGGAGTTCACACGAACTTGCCCG AACAAGAGCGCAACTGGCTGCCGGGACCAGGAAT TCGAAATCAAGCTTGGTTTAATTCTGCAACTGGAA ACAACCCACTAACTGGTACCTGGCAATATTCCAAC AAATACGTACTTGAAAACCGTGCTTCGAAGATTGC TCCTGGACCTGCCATGGGAATTGAATCAACAAAAT TCGACGGGAACGGAATCATCTTTTCTAAAGAATAC ATCACCAATGTAAATACAGCCAATCCGAATCAAGT AAACATCACACGCGAGACCGAAATAAACTCAACG AACCCCTTGGCTGGAGGATCTCTCGGAGCCCACGC TAATAATTCACAAAACACAACAACGGCACCGACG CTAGACCACACCAACGTCATGGGTGTGTTTCCGGG TAGTGTCTGGCAAGACAGAGACATTTACCTTCAAG GACAAATCTGGGCCAAGATTCCCCACACAGACGG ACATTTCCACCCTTCTCCTCTCATGGGAGGATTTG GACTGAAGAACCCGCCTCCTCAAATTCTGATCAAA AACACACCTGTTCCGGCTGACCCACCAACTGAATT CAATGCGAACAAAATCTCTTCTTTCATCACTCAAT ACTCAACCGGACAAGTTACAGTGGAAATGGAATG GGAACTTCAGAAAGAAACCTCCAAAAGATGGAAT CCAGAAATCCAGTACAGCGACGACTCGTCTTCGAC GTCTGGCTCCATTCTGCACTTTGCTCCGGATGATGT TGGAAACTACAAAGAGTTCCGCTCTATCGGAACTC GTTACCTTACCCGTCCTCTGTAA (SEQ ID NO: 849)	PLTGTWQYSNKYVLENRASKIAPGPAMG IESTKFDGNGIIFSKEYITNVNTANPNQVN ITRETEINSTNPLAGGSLGAHANNSQNTT TAPTLDHTNVMGVFPGSVWQDRDIYLQ GQIWAKIPHTDGHFHPSPLMGGFGLKNP PPQILIKNTPVPADPPTEFNANKISSFITQY STGQVTVEMEWELQKETSKRWNPEIQYS DDSSSTSGSILHFAPDDVGNYKEFRSIGT RYLTRPL (SEQ ID NO: 850)

MM3	ATGTCTTTCTTTGATTGGATAGGTAAACAGTACGC TCAAGGAGCTGCTGAATTCTGGGATTTGAAGTCCG GTCCTCCTGCACCAAAAAAGGCGCGAAAGGACGG ATCCGCCGGATTCAATTTCCCTGGGCACAAGTACC TGGGTCCTGGCAATTCATTGGATCGCGGAGATCCC GTGGACGCCGACGACGCTGCTGCTCAAAAGCACG ACCAGTCGTATCAAGAACAGCTTGAGGCGGGAGA CAATCCCTACCTCAAGTACAATCACGCGGATCGCG AGTTCCAGGAGGCGTTGAAAGGAGACACCTCCTTC GAAGGAAATCTCGCGAGAGGACTCTTTGAGGCCA AGAAGCTCGTTGCAGAGCCTCTCGGCCTTGTTGAA CCGGAACTGGCGCCACCTAGTGGCCGCAAGCGAC CGGTGCAATCGAGTCAAGAATCTGGTTACTCGAGT AGCCAAGACAAGCGGCCCAACCTCGACTTAGACG AGGAGGATCGTGAGCTCGCTGCCGCTGCAGCGGA GACCGAAACTGGAAGCGCTCCCCCCACCGGCAAT TTGGGACCTGGTACGATGGCTGGAGGCGGTAGCG CGCCAATCGACGACGGCTCGTATGGTGCCGATGG AGTGGGCAATGCCTCGGGAGATTGGCATTGCGATT CCATATGGATGGACGACCGTGTCGTCACCCGCACC ACTCGGACCTGGAACCTGCCCACCTACAATAACCA CGTCTACAAACGACTCAACGGAAACACGGAACAG ACCACCGGAGATGGAAGCTACTTCGGATTCAGCA CCCCGTGGGGATACTTCGACTTCAACCGCTTCCAC TGTCATTTCTCCCCCAGAGACTGGCAAAGACTCAT CAATAACAATTGGGGACTCAGACCAAAAAGTCTA CGGTTCAAAATCTTTAACATTCAGGTTAAAGAGGT CACGGTGCAAGACCAAACGACGATCATCTCCAAT AACCTTACCAGCACGGTTCAAGTGTTTGCGGACAC GGACTATCAATTGCCGTACGTGATTGGTTCTGCTC ACGAAGGTTGTCTGCCTCCTTTCCCAGCCGATGTG TTTATGCTGCCGCAGTATGCATACTGCACTCGACA AGACACAAACAGCAGTAATCCCACTCCCAGAAGC GCTTTCTATTGTTTGGAATACTTTCCCAGCAAGAT GTTGAGAACTGGTAACAACTTTGAATTTACTTACA ATTTTGAGAAGGTGCCCTTCCACAGCATGTGGGCT CATAACCAGTCTTTAGATCGGCTGATGAACCCTCT CATCGATCAGTATTTGTACTACCTCAATCAAACTA CAAATAGTGCTCTATCCAGCCGATCTAAGACTGGT TACGAGCCTAGCCACAGAAGCGTGTTCGTGCTGCC CCAGAGCAAGTACAATAGTGCTCTATCCTACAACA AATCAACAAAATCAAATCTACCTGAGCAAGAACG CAACTGGCTACCTGGACCTGGCATCAGAAATCAA ACGTGGTTCAGCAGTGGAACTGGCAACATGGCTGT AGACGCCATGTGGAACTACTCGAATAAGTATGTGC TGGAGAACAGAGCTTCCAAGATTGCTCCTGGTCCA GCTATGGGGATTGAGTCAACAAAAGTAGATGGCA ATGGACTAATCTTCTCTAAAGAGTATATAACTAAT TCAACCACCGCGAATGCTGATCAAGTGAACGTCAC TAGAGAAACTGAAATCAACTCAACCAATCCCTTGG CGGGCAAATCACTTGGAGCTCATGCGAACAATAT ACAGAACCAGACAACACTGCCAACACTTGCTAAT ACTGACATCATGCCGGTACTACCTGGCAGCGTTTG GCAGAATAGAGATATCTACCTTCAAGGCCAAATCT	MSFFDWIGKQYAQGAAEFWDLKSGPPA PKKARKDGSAGFNFPGHKYLGPGNSLDR GDPVDADDAAAQKHDQSYQEQLEAGD NPYLKYNHADREFQEALKGDTSFEGNLA RGLFEAKKLVAEPLGLVEPELAPPSGRKR PVQSSQESGYSSSQDKRPNLDLDEEDREL AAAAAETETGSAPPTGNLGPGTMAGGGS APIDDGSYGADGVGNASGDWHCDSIWM DDRVVTRTTRTWNLPTYNNHVYKRLNG NTEQTTGDGSYFGFSTPWGYFDFNRFHC HFSPRDWQRLINNNWGLRPKSLRFKIFNI QVKEVTVQDQTTIISNNLTSTVQVFADTD YQLPYVIGSAHEGCLPPFPADVFMLPQY AYCTRQDTNSSNPTPRSAFYCLEYFPSK MLRTGNNFEFTYNFEKVPFHSMWAHNQ SLDRLMNPLIDQYLYYLNQTTNSALSSRS KTGYEPSHRSVFVLPQSKYNSALSYNKS TKSNLPEQERNWLPGPGIRNQTWFSSGT GNMAVDAMWNYSNKYVLENRASKIAP GPAMGIESTKVDGNGLIFSKEYITNSTTA NADQVNVTRETEINSTNPLAGKSLGAHA NNIQNQTTLPTLANTDIMPVLPGSVWQN RDIYLQGQIWAKIPHTDGHFHPSPLMGGF GLKNPPPQILIKNTPVPANPPVEFTAGKIN SFITQYSTGQVTVEMEWELQKETSKRWN PEIQYSDDSSATNQSVLHFAPNETGEYKE FRSIGSRYLTRPL (SEQ ID NO: 852)

	GGGCGAAGATTCCTCACACAGATGGACACTTTCAT CCTTCACCGCTAATGGGAGGATTCGGTCTAAAGAA TCCTCCTCCTCAAATACTGATCAAGAACACACCAG TGCCAGCAAATCCTCCGGTAGAATTTACAGCTGGC AAGATTAACTCTTTCATCACTCAGTATTCAACTGG TCAAGTCACAGTCGAAATGGAATGGGAACTTCAG AAAGAAACATCCAAAAGATGGAATCCAGAGATCC AATACAGCGACGACTCATCAGCCACCAACCAATC GGTGCTGCACTTTGCTCCAAACGAAACAGGAGAG TACAAGGAATTCCGTTCAATTGGATCCCGTTATCT CACTCGTCCTCTGTAA (SEQ ID NO: 851)	
SD3	ATGTCGTTTGTTGATCACCCTCCAGATTGGTTAGA AGAGGTTGGTGAAGGCCTTCACGAGTTTTTGGAGC TCGAAGCTGGCCCACCCAAACCGAAGCCCAATCA GCAAAAACAAGACAACGCCCGTGGTCTTGTACTG CCTGGATATAATTATCTGGGACCCTTCAACGGACT CGACAAGGGAGAGCCCGTCAACCGAGCGGACGCT GTTGCGCGAGAGCACGATATCTCGTACAACGAGC AGCTCCAGGCGGGAGACAACCCCTACCTCAAGTA CAACCACGCGGACGCCGAGTTTCAGGAGAAGCTC GCCGACGACACCTCCTTTGGGGGCAACCTCGGAA AGGCAATCTTTCAGGCCAAGAAGAGGGTTCTCGA ACCTTTTGGCCTGGTTGAAGAGCCTGTTAAGACGG CTCCGGCAAAGAAGAGGCCTGTAGAGCAGTCTCC TGCGGAACCGGGCCCATCCTCTGGCATCGGAAAG GCGGGCGATCAGCCTGCGAGAAAGCGACTAAACT TTGGTCAGACTGGAGACGCCGACTCCGAACCCGA CCCCCAGCCTCTCGGAGAAGCACCAGCAGCCAAT CCCAATTTGGGTTCTGCTACAATGTCTGGAGGCGG AGGCGCGCCGATGGCAGATGATCAACAGGGTGCC GATGGAGTGGGTACTTCCTCAGGAAATCGGCATTG CGATTCCACATGGCTGGACAAGCGAATCATTACCA AGTCCACCCGAACCTGGAGCCTGCCCACCTACAAC AACCATCTCTACAACAAGATCATGTCCACCAACGC CAATGGAGACAGCGCTTGGTTTGGCTTCAGCACTC CCTGGGGATACTTTGACTTCAACCGCTTCCACTGC CACTTCTCGCCCCGAGACTGGCAACGACTCATCAA CAACAACTGGGGCATCCGGCCCAAGAAAATGCAC GTCAACCTCTTCAATATCCAAGTCAAGGAAGTCAC GACGCAAGACTCGACGACGACCACCACCAATAAC CTTACCAGCACGATTCAGGTGTTTGCGGACACGGA GTACCAGCTGCCGTACGTAGCCAGCAACGCCCAC GAGGGCTGTCTCCCGCCCTTTCCGGCGGACGTCTT CATGCTTCCGCAGTACGGCTACTGCACGCTGCAAC GCGAGAACTCCAACGATCCTGTGGCCCAGAGCTC GTTCTACTGCCTGGAGTACTTCCCCAGCCAGATGC TGAGGACCGGAAACAACTACTCCATCAGCTACAC	MSFVDHPPDWLEEVGEGLHEFLELEAGP PKPKPNQQKQDNARGLVLPGYNYLGPF NGLDKGEPVNRADAVAREHDISYNEQL QAGDNPYLKYNHADAEFQEKLADDTSF GGNLGKAIFQAKKRVLEPFGLVEEPVKT APAKKRPVEQSPAEPGPSSGIGKAGDQPA RKRLNFGQTGDADSEPDPQPLGEAPAAN PNLGSATMSGGGGAPMADDQQGADGV GTSSGNRHCDSTWLDKRIITKSTRTWSLP TYNNHLYNKIMSTNANGDSAWFGFSTP WGYFDFNRFHCHFSPRDWQRLINNNWGI RPKKMHVNLFNIQVKEVTTQDSTTTTTN NLTSTIQVFADTEYQLPYVASNAHEGCLP PFPADVFMLPQYGYCTLQRENSNDPVAQ SSFYCLEYFPSQMLRTGNNYSISYTFEDV PFHSMYMHNQSLDRLMNPLIDQYLWYL NTTVSTNNTNTFSKSTKGDLARQNRNWL PGPAFQTVAFNLNGQGNFGSTNHWLYL NKYTMNGKSVAIGPGPVVAPKSTTFQAD GMIFAKSADTATSAAPQNEVNITSEAETS TVNPMMGTNPLIINSSNTPSTSAPTMSNV SVNPVMPGSVWQDRDIYLQGPIWAKIPR TDGTFHPSPLMGGFGLRRPPPQIFIKNTPV PSNPPTTFNPAKINSFITQYSTGQVTVEME WELEKENSKRWNPEIQFTNNFQVSDGN MVPFCVDSRGVYSEPRPIGTRYLTRTL (SEQ ID NO: 854)

	GTTCGAAGACGTGCCCTTCCACAGCATGTACATGC ACAACCAGAGCCTGGACCGGCTGATGAACCCGCT CATCGACCAGTACTTGTGGTACCTGAATACCACCG TATCGACCAACAACACAAACACCTTCTCCAAGTCG ACCAAAGGCGACCTGGCACGCCAGAACCGAAACT GGCTGCCTGGACCCGCCTTCCAGACGGTGGCCTTC AACCTGAACGGACAAGGCAACTTTGGATCGACCA ATCACTGGCTTTACTTGAACAAGTACACAATGAAT GGAAAAAGCGTGGCTATCGGCCCCGGCCCCGTCG TGGCTCCCAAGAGTACAACCTTCCAGGCCGACGGC ATGATCTTCGCCAAGAGCGCAGATACGGCGACCTC TGCTGCCCCTCAAAATGAGGTCAACATCACCAGCG AGGCGGAGACGTCCACCGTCAATCCCATGATGGG CACCAATCCTCTCATTATCAATTCGAGCAATACTC CATCTACCTCCGCTCCCACGATGTCAAATGTATCC GTTAATCCCGTCATGCCTGGCTCCGTTTGGCAAGA CAGAGACATCTACCTTCAAGGCCCGATCTGGGCCA AGATTCCGCGTACCGACGGAACCTTCCACCCCTCG CCTCTGATGGGAGGATTCGGGCTGCGTCGTCCTCC TCCTCAAATCTTCATCAAGAACACGCCCGTTCCGT CCAATCCTCCAACCACGTTCAATCCCGCCAAGATC AACTCCTTCATCACCCAATACTCGACCGGACAGGT CACCGTTGAGATGGAATGGGAGCTCGAGAAGGAA AACAGCAAGAGGTGGAACCCGGAAATACAGTTCA CCAACAATTTCCAGGTTTCGGACGGGAACATGGTT CCCTTCTGCGTGGACTCACGTGGTGTCTACTCTGA ACCCCGCCCCATCGGAACCCGCTACCTTACCCGAA CCCTCTAA (SEQ ID NO: 853)	
MK1	ATGTCGTTTGTCGATCACCCCCCAGACTGGCTCGA GGAGATCGGAGAGGGGCTCTCCGAGTTCATCGGA CTAGAGGCGGGACCGCCGAAGCCTAAGCCGGGCT ACCAGGACCGCGCGCGCGGCCTCGTCGTCCCCGGC TACAAGTACCTCGGGCCCTTCAACGGACTCGACAG GGGCGAGCCCGTCAACGCGGCCGACGCGGCGGCC AAGAAGCACGACGAGGAGTACGACCGGCTCCTCA AGGCGGGGGACAACCCGTACCTCGCCTACAACCA CGCGGACGCCGAGTTCCAGAAGGACCTCAGCGGC GACTCCAGCCTAGCCGGCAACGCCGCGAACGCGC TCTTCCAGGCCAAGAAGAGGGTCCTCGAGCCCTTC GGACTGGTCGAGGGAGAGCCGGAGCCCAAGAAGA CGCCTTCCGTCAAGAGGCCGCACGCATCGCCGGAC TCGTCGAGCGGCGTCGGGAAGAAGGGCGACCAGC CCGCGCGGAAGAGGCTCGACTTTGGGACGGAGCC CGCGGGTCAAGACGGAGCGGGACGAGCGGCCCAG GCAGCCGGAGATATGGCATCTGCTGAGGTGGCTG CGGGTGGTGGCGGACCAGTGGGCGACGATGCACA AGGTGCCGATGGAGTGGGTAATGCCTCAGGAAAT TGGCATTACGATTCCGTTTGGATGGACGGCGCTGT CATCACCAAGTCCACCCGAACCTGGAGCCTGCCCG CCTACAACAACCACCTCTACCGCCAGATCCAGTCC AGCGGCACCGGAGACGGCACGTACTTTGGTTACA GCACGCCTTGGGGATACTTCGATTTCAATCGATTC CACTGCCACTTTTCTCCCAGGGACTGGCAGCGGCT CATCAACAACCACTGGGGCATCCGCCCCAAGCGG	MSFVDHPPDWLEEIGEGLSEFIGLEAGPP KPKPGYQDRARGLVVPGYKYLGPFNGL DRGEPVNAADAAAKKHDEEYDRLLKAG DNPYLAYNHADAEFQKDLSGDSSLAGN AANALFQAKKRVLEPFGLVEGEPEPKKT PSVKRPHASPDSSSGVGKKGDQPARKRL DFGTEPAGQDGAGRAAQAAGDMASAEV AAGGGGPVGDDAQGADGVGNASGNWH YDSVWMDGAVITKSTRTWSLPAYNNHL YRQIQSSGTGDGTYFGYSTPWGYFDFNR FHCHFSPRDWQRLINNHWGIRPKRLHFK LFNIQVKEVTTTDGTTTIANNLTSTIQVFA DTEYQLPYVLGNAHEGCLPPFPADVFML PQYAYLTLNANPTNQGAAALSLPQSAFY CLEYFPSQMLRTGNNFSFSYEFEKLPFHS MFMHSQSLDRLMNPLIDQYLWYLNATT GNNLSFNKAGAKNFPEYFRNWLPGPGQ RVQQWSTIGTQNNAQTGTWASANKWIL MGRSSKMAPGLAQPVRNAQTVTNGSQLI FNNETIKGSTATASTVHSGLLVTNESETA PTNPNSATKWGVMTDNQQTTSTTPTVSD DLEAHVFPGMVWQDRDIYLQGPIWAKIP ETDGHFHPSPLMGGFGLKNPPPQILVKNT PVPAVPPTTFTPQKVNSFITQYSTGQVTV EIEWELRKEKSKRWNPEIQYTSNFENSAN VQFSVNGDGAYIEPRPIGTRYLTHNL

	CTCCACTTTAAGCTCTTCAACATCCAGGTCAAGGA GGTCACGACGACCGACGGCACCACGACCATCGCC AACAACCTCACCAGCACCATACAGGTCTTTGCGGA CACGGAGTACCAGCTCCCGTACGTGCTCGGCAACG CCCACGAGGGCTGCCTGCCGCCGTTCCCGGCGGAC GTCTTCATGCTGCCGCAGTACGCGTACCTGACGCT CAACGCCAACCCCACGAACCAGGGCGCGGCCGCG CTCAGCCTGCCGCAGAGCGCCTTCTACTGCCTGGA GTACTTTCCGAGCCAGATGCTGAGGACCGGGAAC AACTTCTCGTTCAGCTACGAGTTCGAGAAGCTGCC CTTCCACTCGATGTTCATGCACAGCCAGAGCCTGG ACCGGCTCATGAACCCGCTCATCGACCAGTACCTC TGGTACCTCAACGCCACGACCGGCAACAACCTCTC CTTCAACAAGGCCGGCGCCAAGAACTTCCCCGAGT ACTTCCGCAACTGGCTGCCCGGGCCCGGCCAGCGC GTGCAGCAGTGGAGCACGATCGGAACCCAGAACA ACGCGCAGACCGGCACGTGGGCCAGCGCCAACAA GTGGATCCTCATGGGCAGGTCCAGCAAGATGGCC CCCGGGCTGGCCCAGCCGGTCAGGAACGCCCAGA CCGTCACCAACGGCTCGCAGCTCATCTTCAACAAC GAGACCATCAAGGGCTCCACCGCGACGGCCTCCA CCGTACACTCGGGGCTGCTCGTCACCAACGAGTCG GAGACCGCCCCGACCAACCCCAACTCGGCGACAA AGTGGGGCGTCATGACGGACAACCAGCAGACCAC CTCCACTACGCCGACCGTCAGCGACGACCTCGAGG CGCACGTCTTCCCGGGCATGGTCTGGCAGGACCGC GACATCTACCTGCAGGGCCCGATATGGGCCAAGA TCCCGGAGACGGACGGCCACTTCCACCCGTCCCCG CTCATGGGCGGCTTCGGGCTCAAGAACCCGCCGCC GCAGATCCTCGTCAAGAACACGCCCGTGCCCGCCG TGCCGCCCACGACCTTCACACCGCAGAAGGTCAAC TCCTTCATCACGCAGTACTCCACCGGCCAGGTCAC CGTCGAGATCGAGTGGGAGCTGCGCAAGGAGAAG AGCAAGCGCTGGAACCCCGAGATCCAGTACACCT CCAACTTCGAGAACTCGGCCAACGTACAGTTCTCC GTCAACGGCGACGGCGCGTACATCGAGCCGCGAC CGATCGGAACCCGGTACCTCACCCACAACCTGTAA (SEQ ID NO: 855)	(SEQ ID NO: 856)

CT1	ATGTCGTTTGTCGATCACCCCCCAGACTGGCTCGA GGAGATCGGAGAGGGGCTCTCCGAGTTTATCGGA CTAGAGGCGGGACCGCCGAAGCCTAAGCCGGGCT ACCAGGACCGCGCGCGCGGCCTCGTCGTCCCCGGC TACAAGTACCTCGGGCCCTTCAACGGACTCGACAG GGGCGAGCCCGTCAACGCGGCCGACGCGGCGGCC AAGAAGCACGACGAGGAGTACGACCGGCTCCTCA AGGAGGGGGACAACCCGTACCTCACCTACAACCA CGCCGACGCCCAGTTCCAGCAGGACCTCCGGGGC GACGCCAGCCTAGCCGGCAACGCCGCGAACGCGC TCTTCCAGGCCAAGAAGAGGGTCCTCGAGCCCTTC GGACTGGTCGAGGGAGAGCCGGAGCCGAAGAAGA CGCCTTCCGTCAAGAGGCCGCACGCATCGCCGGAC TCGTCGAGCGGCGTCGGGAAGAAGGGCGACCAGC CCGCGAGGAAGAGGCTCGACTTTGGAACGGAGCC CGCGAGTCAAGACGGAGCCGGACGAGCGGCGCAG GCAACCGGAGATATGGCATCTGCTGAAGTGGCTG CGGGTGGTGGCGGACCAGTGGGCGACGATGCACA AGGTGCCGATGGAGTGGGTAGCGCCTCGGGAAAT TGGCATTGCGATTCCGTTTGGATGGACGGCGCTGT CATCACCAAGTCCACCCGAACCTGGAGCCTGCCCG CCTACAACAACCACCTCTACCGCCAGATCCAGTCC AGCGGCACCGGAGACGGCACGTACTTTGGTTACA GCACGCCTTGGGGATACTTCGATTTCAATCGATTC CACTGCCACTTTTCTCCCAGGGACTGGCAGCGGCT CATCAACAACCACTGGGGCATCCGCCCCAAGCGG CTCCACTTCAAGCTCTTCAACATCCAGGTCAAGGA GGTCACGACGACCGACGGCACCACGACCATCGCC AACAACCTCACCAGCACCATACAGGTCTTTGCGGA CACGGAGTACCAGCTCCCGTACGTGCTCGGCAACG CCCACGAGGGCTGCCTGCCGCCGTTCCCGGCGGAC GTCTTCATGCTGCCGCAGTACGCGTACCTGACGCT CAACGCCAACCCGACGTCGCAGAGCGCGGCCGCG CTCAGCCTGCCGCAGAGCGCCTTCTACTGCCTCGA GTACTTCCCGAGCCAGATGCTGAGGACCGGGAAC AACTTCTCGTTCAGCTACGAGTTCGAGAAGCTGCC CTTCCACTCGATGTTCATGCACAGCCAGAGCCTGG ACCGGCTCATGAACCCGCTCATCGACCAGTACCTC TGGTACCTCAACGCCACGACCGGCAACAACCTCTC CTTCAACAAGGCCGGCGCCAAGAACTTCCCCGAGT ACTTCCGCAACTGGCTGCCCGGGCCCGGCCAGCGC GTGCAGCAGTGGAGCACGATCGGGACCCAGAACA ACGCGCAGACCGGCACGTGGGCCAGCGCCAACAA GTGGATCCTCATGGGCAGGTCCAGCAAGATGGCC CCCGGGCTGGCCCAGCCGGTCCGGAACGCCCAGA CCGTCACCAACGGCTCGCAGCTCATCTTCAACAAC GAGACCATCAAGGGCTCCACCGCGACGGCCTCCA CCGTGCACTCGGGGCTGCTCGTCACCAACGAGTCG GAGACCGCCCCGACCAACCCCAACTCGGCGACAA AGTGGGGCGTCATGACGGACAATCAGCAGACCAC CTCCACTACGCCGACCGTCAGCGACGACCTCGAGG CGCACGTCTTCCCGGGCATGGTCTGGCAGGACCGC GACATCTACCTGCAGGGCCCGATCTGGGCCAAGAT CCCGGAGACGGACGGCCACTTCCACCCGTCCCCGC	MSFVDHPPDWLEEIGEGLSEFIGLEAGPP KPKPGYQDRARGLVVPGYKYLGPFNGL DRGEPVNAADAAAKKHDEEYDRLLKEG DNPYLTYNHADAQFQQDLRGDASLAGN AANALFQAKKRVLEPFGLVEGEPEPKKT PSVKRPHASPDSSSGVGKKGDQPARKRL DFGTEPASQDGAGRAAQATGDMASAEV AAGGGGPVGDDAQGADGVGSASGNWH CDSVWMDGAVITKSTRTWSLPAYNNHL YRQIQSSGTGDGTYFGYSTPWGYFDFNR FHCHFSPRDWQRLINNHWGIRPKRLHFK LFNIQVKEVTTTDGTTTIANNLTSTIQVFA DTEYQLPYVLGNAHEGCLPPFPADVFML PQYAYLTLNANPTSQSAAALSLPQSAFY CLEYFPSQMLRTGNNFSFSYEFEKLPFHS MFMHSQSLDRLMNPLIDQYLWYLNATT GNNLSFNKAGAKNFPEYFRNWLPGPGQ RVQQWSTIGTQNNAQTGTWASANKWIL MGRSSKMAPGLAQPVRNAQTVTNGSQLI FNNETIKGSTATASTVHSGLLVTNESETA PTNPNSATKWGVMTDNQQTTSTTPTVSD DLEAHVFPGMVWQDRDIYLQGPIWAKIP ETDGHFHPSPLMGGFGLKNPPPQILVKNT PVPAVPPTTFTPQKVNSFITQYSTGQVTV EIEWELRKEKSKRWNPEIQYTSNFENSAN VQFSVNGDGAYIEPRPIGTRYLTHTL (SEQ ID NO: 858)

	TCATGGGCGGCTTCGGGCTCAAGAACCCGCCGCCG CAGATCCTCGTCAAGAACACGCCCGTGCCCGCCGT GCCGCCCACGACCTTCACGCCGCAGAAGGTCAACT CCTTCATCACGCAGTACTCCACCGGCCAGGTCACC GTCGAGATCGAGTGGGAGCTGCGCAAGGAGAAGA GCAAGCGCTGGAACCCCGAGATCCAGTACACGTC CAACTTCGAGAACTCGGCCAACGTTCAGTTCTCCG TCAACGGCGACGGCGCGTACATCGAGCCCCGCCC GATCGGAACCCGATACCTCACCCACACCCTGTAA (SEQ ID NO: 857)	
CT2	ATGTCGTTTGTCGATCACCCCCCAGACTGGCTCGA GGAGATCGGAGAGGGGCTCTCCGAGTTTATCGGA CTAGAGGCGGGACCGCCGAAGCCTAAGCCGGGCT ACCAGGACCGCGCGCGCGGCCTCGTCGTCCCCGGC TACAAGTACCTCGGGCCCTTCAACGGACTCGACAG GGGCGAGCCCGTCAACGCGGCCGACGCGGCGGCC AAGAAGCACGACGAGGAGTACGACCGGCTCCTCA AGGCGGGGGACAACCCGTACCTCACCTACAACCA CGCGGACGCCGAGTTCCAGAAGGACCTCCGGGGC GACGCCAGCCTAGCCGGCAACGCCGCGAACGCGC TCTTCCAGGCCAAGAAGAGGGTCCTCGAGCCCTTC GGACTGGTCGAGGGAGAGCCGGAGCCGAAGAAGA CGCCTTCCGTCAAGAGGCCGCACGCATCGCCGGAC TCGTCGAGCGGAATCGGGAAGAGGGGCGACCAGC CCGCGAGGAAGAGGCTCGACTTTGGAACGGAGCC CGCGAGTCAAGACGGAGCGGGAAGAGCGGCCAAC GCAGCCGGAGATATGGCATCTGCTGAGGTGGCTG CGGGTGGTGGCGGACCAGTGGGCGACGATGCACA AGGTGCCGATGGAGTGGGTCAGTCCTCGGGAAAT TGGCATTGCGATTCCGTTTGGATGGGCGACCGAGT GCTCACCAAGTCCACCCGGACCTGGAGCCTGCCCA CCTACAACAACCACCTGTACAAGCAGATCAACGG CAGCGGAACCGGGGACGCGGTGTACTTTGGTTAC AGCACGCCTTGGGGATACTTCGATTTCAATCGATT CCACTGCCACTTTTCTCCCCGAGACTGGCAGCGCC TCGTCAACAACCACTGGGGCATCCGCCCCCGAAG GCTCAACTTCAAGCTCTTCAACATCCAGGTCAAGG AGGTCACGACGACGGACGGCACCAAGACCATCGC CAATAACCTCACGTCCACCGTCCAGGTGTTTGCGG ACACGGAGCACCAGCTCCCGTACATCCTAGGCTCC GCGCACGAGGGCTGCATGCCGCCCTTCCCCGCGGA CGTCTTCATGCTGCCGCAGTACGGCTACCTCACGC TCAACGGGCCCGGCAGCAACAACAACAACCTCAG CACGCCCTCGAGCGCCTTCTACTGCCTCGAGTACT TCCCGAGCCAGATGCTGCGCACGGGCAACAACTTC GTCTTCACCTACGAGTTCGAGAAGGTGCCCTTCCA	MSFVDHPPDWLEEIGEGLSEFIGLEAGPP KPKPGYQDRARGLVVPGYKYLGPFNGL DRGEPVNAADAAAKKHDEEYDRLLKAG DNPYLTYNHADAEFQKDLRGDASLAGN AANALFQAKKRVLEPFGLVEGEPEPKKT PSVKRPHASPDSSSGIGKRGDQPARKRLD FGTEPASQDGAGRAANAAGDMASAEVA AGGGGPVGDDAQGADGVGQSSGNWHC DSVWMGDRVLTKSTRTWSLPTYNNHLY KQINGSGTGDAVYFGYSTPWGYFDFNRF HCHFSPRDWQRLVNNHWGIRPRRLNFKL FNIQVKEVTTTDGTKTIANNLTSTVQVFA DTEHQLPYILGSAHEGCMPPFPADVFML PQYGYLTLNGPGSNNNNLSTPSSAFYCLE YFPSQMLRTGNNFVFTYEFEKVPFHSMF MHNQALDRLMNPLVDQYLWYLDATSG NNLTFRKAGAKNFPEYFRNWIPGPGCRN QQWNKVGTKNNPQTGTWASANKWRLQ GRLNKYAPGQPNAPAEGFLTNAGDLAFA NAKATGATTAAGTVPADILLTSESETTTT NMMSNNGWGAIASNNQNASVAPTVQYE DSAHVLPGMVWQDRDIYLQGPIWAKIPE TDGHFHPSPLMGGFGLKNPPPQILIKNTP VPADPPTQFSSQKINSFITQYSTGQMTVEI EWELRKENSKRWNPEIQYTANFNNSAN AQFSVNNNGLYIEDRTIGTRYLTHTL (SEQ ID NO: 860)

	CTCGATGTTCATGCACAACCAGGCGCTGGACCGCC TCATGAACCCGCTCGTCGACCAGTACCTCTGGTAC CTCGACGCCACCTCCGGCAACAACCTCACCTTCCG CAAGGCCGGCGCCAAGAACTTCCCCGAGTACTTCC GCAACTGGATCCCAGGCCCGGGCTGCCGCAACCA GCAGTGGAACAAGGTCGGCACCAAGAACAACCCG CAGACCGGCACGTGGGCGTCCGCCAACAAGTGGC GGCTGCAGGGCCGCCTCAACAAGTACGCGCCCGG ACAGCCCAACGCCCCCGCGGAGGGCTTCCTCACCA ACGCCGGCGACCTGGCCTTCGCCAACGCCAAGGC CACCGGCGCCACCACGGCCGCCGGAACCGTGCCC GCCGACATCCTGCTCACCAGCGAGAGCGAGACCA CCACCACCAACATGATGTCCAACAACGGCTGGGG CGCGATCGCCTCCAACAACCAGAACGCGTCCGTCG CGCCCACCGTGCAGTACGAGGACTCGGCGCACGT GCTGCCCGGCATGGTCTGGCAGGACCGCGACATCT ACCTGCAGGGCCCGATCTGGGCCAAGATCCCGGA GACGGACGGCCACTTCCACCCCTCGCCGCTCATGG GCGGCTTCGGACTCAAGAACCCGCCCCCGCAGATC CTCATCAAGAACACGCCCGTGCCCGCCGACCCCCC GACGCAGTTCTCCTCGCAGAAGATCAACTCCTTCA TCACGCAGTACTCCACCGGCCAGATGACCGTCGAG ATCGAGTGGGAGCTGCGCAAGGAGAACTCCAAGC GCTGGAACCCCGAGATCCAGTACACGGCCAACTTC AACAACAGCGCCAACGCGCAGTTCTCCGTCAACA ACAACGGCCTCTACATCGAGGACCGCACCATCGG AACCCGATACCTCACCCACACCCTGTAA (SEQ ID NO: 859)	
RP1	ATGATTCGAGTAGTCATGTCGTTTGTTGACTATCCT CCAGATTGGCTGAAAGAGGTTGCGGGTGGACTCA AAGAGTTCATTGACCTCAAACCGGGTGCGCCGAA GCCCAAAGCGAATCAACAAAAACAAGATCAACGC GGTCTTGTCCTCCCGGGTTACAAATACCTGGGACC CTTCAACGGCCTTGATAAAGGCCAACCGGTGAAC AAAGCCGACGAAGTCGCGCGGGAGCACGACATCG AATACGACAAACTCCTGAAAGAAGGAGACAACCC GTACGTGAAGTACAACCACGCCGATCAAGAGTTTC AAGAGAAACTCGAGGGCGACGAATCTTTTGGCGG TAACCTCGGACGAGCGGTCTTCCAAGCCAAGAAG CGGTTACTGGAACCGTTTGGTTTATCGGAAGAGAC ACCTCCGAGAAAGAGACCCGCTTCCGAGAACAAC CAGTCAGACGTCCCGGCTAAGAGACGTCTGATCCT CGATCCCGAAGACAACGGAGCCGCCGAAGGAGAA CCAATCGAACAACAACAACCAGCCAACGATATGG CATCTGCTGAAATGGCTGCAGGTGGTGGCGGGCC AGTGGGCGACGATGCACAGGGTGCCGATGGAGTG GGTAGCGCCTCAGGAAATTGGCATTGCGATTCCAT TTGGATGGGTGACCGAGTCATCACCAAATCCACTC GCACCTGGAGCCTGCCCACCTACAACAACCACCTC TACAAGCAAATCAACTCTAACACCACCGATGCCA AAGAAACCTACTTTGGCTACAGCACTCCTTGGGGG TATTTTGACTTCAACCGGTTCCACTGCCACTTCTCG CCGAGGGACTGGCAACGACTCATCAACAACCATT GGGGCATGCGACCAAAGAGACTCAACTTTAAACT	MIRVVMSFVDYPPDWLKEVAGGLKEFID LKPGAPKPKANQQKQDQRGLVLPGYKY LGPFNGLDKGQPVNKADEVAREHDIEYD KLLKEGDNPYVKYNHADQEFQEKLEGD ESFGGNLGRAVFQAKKRLLEPFGLSEETP PRKRPASENNQSDVPAKRRLILDPEDNG AAEGEPIEQQQPANDMASAEMAAGGGG PVGDDAQGADGVGSASGNWHCDSIWM GDRVITKSTRTWSLPTYNNHLYKQINSN TTDAKETYFGYSTPWGYFDFNRFHCHFS PRDWQRLINNHWGMRPKRLNFKLFNIQ VKEVTTNDGTKTIANNLTSTVQVFADSE YQLPYVLGNAHEGCFPPFPADVFMLPQY GYLTLNSNDNKAMPQSSFYCLEYFPSQM LRTGNNFTFSYEFESVPFHSMFMHNQSL DRLMNPLIDQYLWYLHSTNSLSLNFKKV KRDHLYKYYRNWVPGPGQRRQQWSES GGANKQLTDLWPYNNKYNLMGRYSSLA PGMPQAAADGFITADGQLIFSNVLQESEE ASGGKPSPNLLITNESEVRPTNPAINAPW GDVALNQQTSTVTPSVAEDNEAHIFPGM VWQDRDIYLQGPIWAKIPDTDGSFHPSPL MGGFGLKHPPPQILIKNTPVPSDPPLEFN GKKIQSFITQYSTGQVTVEIEWELKKEHS KRWNPEIQYTSNFDNTDRVQFSVNGAGN YVEDRPIGTRYLTHNL (SEQ ID NO: 862)

	CTTTAACATCCAAGTCAAAGAGGTCACGACGAAC GACGGTACGAAAACGATCGCGAACAATCTTACCT CGACGGTGCAAGTCTTCGCGGACTCAGAGTACCA ACTCCCCTACGTGCTAGGCAACGCGCACGAGGGA TGCTTTCCACCGTTTCCGGCCGACGTCTTCATGCTA CCGCAGTATGGATACCTGACCTTGAACTCAAACGA CAACAAAGCCATGCCACAGAGCTCGTTCTACTGTC TGGAATATTTTCCGAGCCAGATGCTGAGAACCGGC AACAACTTCACCTTCAGCTATGAGTTCGAGAGCGT TCCATTCCACAGCATGTTTATGCATAACCAATCGC TGGACAGATTGATGAATCCATTGATTGACCAGTAC TTATGGTACCTACACTCTACCAACAGTCTATCTCT AAATTTTAAAAAAGTGAAGAGAGATCATCTGTAC AAATACTACAGAAATTGGGTACCCGGACCCGGTC AAAGAAGACAACAATGGAGCGAAAGTGGTGGAGC AAACAAACAATTGACAGACTTGTGGCCATACAAC AACAAATACAATCTGATGGGAAGATATTCTTCTTT GGCACCGGGAATGCCACAAGCGGCGGCAGATGGA TTCATCACAGCCGATGGACAATTAATATTTTCTAA CGTGCTACAAGAATCAGAAGAAGCCAGCGGTGGA AAACCATCTCCAAATCTATTAATTACAAACGAGTC AGAAGTCAGACCAACAAATCCAGCTATCAACGCA CCTTGGGGAGATGTCGCATTAAATCAACAAACGTC AACAGTAACACCATCAGTAGCTGAAGACAACGAA GCCCACATTTTTCCTGGCATGGTCTGGCAAGACCG AGACATCTACCTGCAAGGACCCATCTGGGCCAAG ATACCAGACACCGACGGCAGCTTCCACCCGTCTCC GCTCATGGGTGGATTTGGACTCAAACACCCTCCGC CGCAGATCCTGATCAAGAACACGCCGGTACCGTC GGATCCTCCGCTGGAATTCAACGGCAAGAAGATC CAGTCCTTCATCACGCAGTACTCGACCGGTCAGGT GACGGTGGAGATCGAGTGGGAGCTGAAGAAGGAG CACAGCAAGCGGTGGAATCCGGAGATCCAGTACA CCTCCAACTTCGACAACACCGACCGGGTGCAGTTC TCGGTCAACGGGGCGGGCAACTATGTCGAAGATC GACCAATCGGGACCCGTTATCTCACCCACAATCTG TAA (SEQ ID NO: 861)	
RF1	ATGAGTCGAGTAGTCATGTCGTTTGTTGATCACCC TCCAGATTGGCTGAAGCGTGTGGCGAGTGGCTTGA CTGAGTTCATCGACCTCGAACCCGGACCCCCAAAA CCCAAACCCGGCTATCAGGACCAGCAGCGCGGTC TCGTGGTACCCGGATACAAATACCTCGGTCCCTTT AACGGTCTCGATAAAGGAGAACCCGTTAACAAGG CTGACGAGGTTGCTCGAGAGCACGACCTCGAGTA CAACAAGCTACTTGAGCAAGGAGATAATCCTTATC TCAAGTACAACCACGCCGATAAGGAGTTTCAAGA GAAACTCTCTGACGACACTTCCTGGGGCGGCAACC TCGGGAAGGCGGTTTTTCAAGCCAAGAAGCGTATC CTCGAGCCCTTTGGCCTAGCCGAAGAGCTCGGCGG TGAAAAGACGGCTCCTGGAAAAAAGCGTCCCAGG CCTCCGGAACTTCCACCTTCGGCTCCGCATTCGGA CGACGAAGAACTGATTGCTCGACCGGCAGTCAGA GCTCGGCTCGAAGGTCAAAACGGAGCAGCAAACG CAGGACCAGCAGCAGCACCAGCCGCAGATATGGC	MSRVVMSFVDHPPDWLKRVASGLTEFID LEPGPPKPKPGYQDQQRGLVVPGYKYLG PFNGLDKGEPVNKADEVAREHDLEYNK LLEQGDNPYLKYNHADKEFQEKLSDDTS WGGNLGKAVFQAKKRILEPFGLAEELGG EKTAPGKKRPRPPELPPSAPHSDDEELIA RPAVRARLEGQNGAANAGPAAAPAADM ASTEVASGGGGPMGDNNQGSDGVGSSS GNWHCDSIWMDGRVITKSTRTWSLPTY NNHLYKQINSNAEVKDEKYFGYSTPWG YFDFNRFHCHFSPRDWQRLINNHWGMR PKRLNFKLFNIQVKEVTTNDGTTTIANNL TSTVQVFADTEYQLPYVLGNAHEGCLPP FPADVFMLPQYGYLTLNDGNQVIPQSSF YCLEYFPSQMLRTGNNFTFSYEFENVPFH SMFMHNQSLDRLMNPLVDQYLWYLSAT NGLNLHFRKSKKGAFASQYRNWMPGPN

	ATCTACTGAAGTGGCTTCAGGCGGTGGCGGCCCAA TGGGCGACAATAACCAAGGTTCCGATGGAGTGGG TAGCTCCTCCGGAAATTGGCATTGCGATTCCATTT GGATGGACGGCAGAGTCATCACCAAGTCTACCAG AACCTGGAGCCTGCCCACCTACAACAACCACCTCT ACAAGCAAATCAACTCAAATGCCGAAGTCAAAGA CGAAAAGTACTTTGGATATAGTACTCCCTGGGGAT ACTTTGATTTCAATCGATTCCACTGCCACTTTTCTC CCAGAGACTGGCAGCGACTCATCAACAACCACTG GGGAATGCGACCCAAAAGACTCAACTTTAAGCTG TTCAACATCCAGGTCAAGGAGGTCACGACGAATG ACGGAACGACAACGATCGCCAATAACCTCACGTC GACCGTCCAGGTGTTTGCAGACACAGAGTACCAGT TGCCGTACGTGCTCGGGAATGCGCACGAGGGCTG CCTCCCGCCGTTCCCGGCAGACGTGTTTATGCTGC CGCAGTACGGATACTTGACTCTAAACGACGGCAA CCAGGTCATTCCTCAGAGCTCTTTTTACTGCTTGGA GTACTTTCCCAGTCAGATGCTGAGAACCGGAAACA ACTTTACGTTCAGCTACGAGTTTGAGAACGTGCCT TTCCACAGCATGTTTATGCATAATCAGTCTCTGGA CCGACTGATGAATCCTTTAGTCGATCAGTACCTGT GGTACCTCTCGGCTACCAACGGATTAAATTTACAT TTTAGAAAATCCAAGAAGGGAGCGTTTGCCAGTC AATACAGAAACTGGATGCCCGGTCCTAACCAGCG ACGTCAGCAGTGGAGTAAAGGAGCTGGAAACAAC AACGCCTTGACCAATCTCTGGGAATACAAAAACA AGTTCCACATCATGAACAGAGACTCGGTACTGGCT CCAGGACTCGCACAGGCTGCGGCGGATGGATATTT CACAGCCGACGGACAATTAATTTTTGCTAACGCTA CACAACAATCGGACGAGGCTTCCGGTGGAAATCC ACCGAGCAGCATGTTAATTACCGGTGAATCCGAA ACTCGACCCACTAATCCCGGAGCAAACACCAAGT GGGGTGAAATGGCCGACAACCAACAGTCTACAAC GGCAAAGCCGACCATGGTTATCGACAACGAGGCT CACATCTTTCCCGGCATGGTCTGGCAGGATCGGGA CATTTACCTGCAAGGGCCCATCTGGGCCAAGATTC CAGAGACGGACGGACACTTTCATCCTTCTCCGCTG ATGGGCGGATTCGGTCTCAAGAATCCTCCTCCTCA GATACTCATCAAGAATACACCCGTTCCGGCTAATC CTCCGGAAGCGTTCACCGCCAACAAGATAAACTCC TTCATCACGCAGTACTCGACCGGACAAGTAACGGT GGAAATTGAATGGGAGCTCAGGAAGGAGCACAGC AAGCGCTGGAACCCAGAGGTGCAGTACACGTCCA ACTTTGACTCTACAGACAGTCTGCAGTTTACCGTT AACAACGACGGACTCTACATCGAGACTCGACCAA TCGGAACCCGTTTCCTTTCCCACAATATTTAA (SEQ ID NO: 863)	QRRQQWSKGAGNNNALTNLWEYKNKF HIMNRDSVLAPGLAQAAADGYFTADGQ LIFANATQQSDEASGGNPPSSMLITGESE TRPTNPGANTKWGEMADNQQSTTAKPT MVIDNEAHIFPGMVWQDRDIYLQGPIWA KIPETDGHFHPSPLMGGFGLKNPPPQILIK NTPVPANPPEAFTANKINSFITQYSTGQV TVEIEWELRKEHSKRWNPEVQYTSNFDS TDSLQFTVNNDGLYIETRPIGTRFLSHNI (SEQ ID NO: 864)

RS1	ATGAGTCGAGTAGTCATGTCGTTTGTCGATCACCC TCCAGATTGGCTGAAAGAGGCCGCGGGCGGACTA CGCGAGTTTCTCGATCTCGAGCCGGGAGCGCCTAA ACCCAAACCTAATCAACAAAAACAGGACCAGCGA GGTCTGGTTCTGCCGGGATACAAATATCTCGGACC TTTCAACGGCCTTGATAAAGGCCAGCCCGTCAATA AGGCGGACGAGGTCGCGCGAGAGCACGATCTCGA GTATAACAAGCTACTTGAGGCGGGCGACAATCCCT ACCTCAAGTACAACCACGCAGACCAAGAGTTTCA GGAGAAACTCGAGGGGGACACTTCCTTCGGAGGT AACCTCGGTCGCGCGGTATTCCAGGCCAAGAAGA GGGTGCTAGAGCCCTTTGGATTGGCCGAAGAAGCT CCGGCGAAGACGGCTCCTGGGAAGAAACGTCCCA GGCCCATCGAACCCCCTCCTTCCTCGGAGAACTCG GTTGACGAGGACTTGGTCGAGCGCCAAGCTCAGC GTCAACGTCTCGGATTTCAAGACGGACCCGGAAA CAGCGCTAACCAAGTCCAGCAACAACCAGCCGCA AATATGGCATCTGCTGAAGTGGCTGCGGGCAGTG GCGGCCCAATGGGCGACAATGCACAAGGTGCCGA TGGAGTGGGTAGCTCCTCCGGAAATTGGCATTGCG ATTCCCTTTGGATGGGCGATAGGGTGATGCGATTC CCTTGGATGGGAGACCGAGTCATCACCAAGTCCAC CCGAACCTGGAGCCTGCCCACCTACAACAACCATC TCTACAAGCAAATCAATTCCAACTCGGATGTCACC GACGAAAAGTACTTTGGATACAGTACCCCGTGGG GATACTTTGACTTTAACCGCTTCCACTGCCACTTTT CACCCCGAGACTGGCAGCGCCTCATCAACAACCAT TGGGGTATGCGCCCGAAGAGGCTCAACTTTAAGCT GTTCAACATCCAGGTTAAGGAGGTCACGACGAAC GACGGCACCACGACCATCGCCAACAACCTCACGT CTACCGTCCAGATTTTCGCCGACAGCGAGTACCAG TTGCCGTACGTGCTCGGGAATGCGCACGAGGGCTG CCTCCCGCCGTTCCCGGCAGACGTGTTCATGCTGC CGCAGTACGGGTACCTCACGTTGAACGACGGAAA TCAAGTGATTCCACAAAGTTCCTTCTACTGTCTGG AGTATTTTCCCAGCCAGATGCTGAGAACCGGCAAC AACTTTTCCTTCAGCTACGAATTTGAAAACGTGCC TTTCCACAGCATGTTTATGCACAACCAGTCTTTGG ATCGGTTAATGAACCCCCTCATGGACCAGTACCTG TGGTACCTGACGGCCACCAACGGCGTTCAGTTGCA CTTTAAAAAGTCCAAGAAGGGGGCGTTTGCCAGC CAGTTTAGGAACTGGCTGCCGGGTCCCAACCAGCG GCGTCAGCAGTGGAGCAAAGGGGCCGGTAACAAC AACGCTCAGACCAACCTGTGGAGTTACAAAAACA AGTACCGGATCATGAACCGGGACAGCGTGGCGGC TCCTGGACTGGCTCAAGCGGCGGCCGATGGCTTCA TCACGGCCGACGGACAGCTCATCTTCTCCAACGCC GGACAACAGTCGGATGAAGCGGCTGGAGCTAAAC CACCGACCACGTTGCTGATTACAGACGAGAGTGA AACTCGTCCTACCAACCCGGGTGCCAACACCAAAT GGGGAGAAGTGGCAGACAACCAACAGACATCTTC CACGGCCCCCACCATGGCCATCGACAACGAGGCC CACATCTTTCCCGGAATGGTGTGGCAGGATCGAGA CATTTACCTGCAGGGTCCCATCTGGGCCAAGATTC	MSRVVMSFVDHPPDWLKEAAGGLREFL DLEPGAPKPKPNQQKQDQRGLVLPGYK YLGPFNGLDKGQPVNKADEVAREHDLE YNKLLEAGDNPYLKYNHADQEFQEKLE GDTSFGGNLGRAVFQAKKRVLEPFGLAE EAPAKTAPGKKRPRPIEPPPSSENSVDED LVERQAQRQRLGFQDGPGNSANQVQQQ PAANMASAEVAAGSGGPMGDNAQGAD GVGSSSGNWHCDSLWMGDRVMRFPWM GDRVITKSTRTWSLPTYNNHLYKQINSNS DVTDEKYFGYSTPWGYFDFNRFHCHFSP RDWQRLINNHWGMRPKRLNFKLFNIQV KEVTTNDGTTTIANNLTSTVQIFADSEYQ LPYVLGNAHEGCLPPFPADVFMLPQYGY LTLNDGNQVIPQSSFYCLEYFPSQMLRTG NNFSFSYEFENVPFHSMFMHNQSLDRLM NPLMDQYLWYLTATNGVQLHFKKSKKG AFASQFRNWLPGPNQRRQQWSKGAGNN NAQTNLWSYKNKYRIMNRDSVAAPGLA QAAADGFITADGQLIFSNAGQQSDEAAG AKPPTTLLITDESETRPTNPGANTKWGEV ADNQQTSSTAPTMAIDNEAHIFPGMVWQ DRDIYLQGPIWAKIPETDGHFHPSPLMGG FGLRNPPPQILIKNTPVPANPPETFNANKI NSFITQYSTGQVTVEIEWELKKEHSKRW NPEVQYTSNFASADSLQFTVNGEGVYIE DRPIGTRYLSHNI (SEQ ID NO: 866)

	CAGAGACGGACGGTCACTTTCACCCGTCTCCTCTG ATGGGAGGATTCGGACTGCGAAATCCTCCTCCTCA GATTCTCATCAAAAACACACCGGTACCGGCCAATC CACCAGAGACCTTCAACGCCAACAAGATCAACTC ATTCATCACCCAGTACTCAACCGGACAGGTTACAG TCGAAATAGAGTGGGAGCTGAAGAAGGAGCACAG CAAACGCTGGAACCCGGAGGTGCAGTACACGTCC AATTTTGCCTCGGCGGACAGCCTCCAATTCACCGT CAACGGAGAAGGCGTGTACATCGAAGACAGACCA ATCGGAACCCGCTACCTCTCCCACAACATTTAA (SEQ ID NO: 865)	

[0095] The disclosure shows that one of the discovered variants (AAV9.DKK1) exhibit deprecated transduction when the cognate receptor for the ligand from which the peptide was derived is knocked out **(Fig. 5a)**. Furthermore, due to the systematic tiling nature of the peptide library, it was possible to generate full-length transduction maps across the entire residue space of ligands utilized in this study **(Fig. 5a,c)**. This not only increases confidence in the screening hits, in that homologous peptides behave similarly, but could also provide insight into critical structural domains mediating virus-receptor interactions and more broadly, provide a platform for further expanding the known protein interactome especially when the dataset is integrated with the cell surfaceome.

[0096] Additionally, in recent years, developing predictive models of AAV infectivity has garnered significant interest from multiple research groups. The application of machine learning to AAV engineering parallels major advances in machine learning across multiple areas of protein science such as structure prediction, enzyme activity forecasting, and antibody binding optimization. While deep learning and similar blackbox methodologies have rapidly become mature technologies, applying these methodologies to AAV engineering is still severely limited by the lack of available training data. The AAV screening data is an ideal training dataset for several reasons: (1) the experimental design features a large,

defined library of variants (**Fig. 1**), meaning that every variant has a reliable quantification of package infectivity, (2) each variant was screened across a panel of 9 major organs to map the infectivity across diverse tissue types, (3) the library was inserted across 2 AAV serotypes and 2 surface loops, providing important functional information on the capsid context in which the peptide resides, and (4) a large cohort of variants were validated to demonstrate the screening data is trustworthy (**Fig. 4**) and (5) to illustrate the utility of the dataset as training data, the packaging efficiency of AAV variants could be accurately predicted from the biophysical characteristics of the inserted peptides (**Fig. 2e**), and also peptide amino acid sequence is directly predictive of tissue-tropism across multiple capsids and insertion sites (**Fig. 3d**). As such, the screening data will have great utility for the machine-learning and computational biology community.

[0097] While the variants identified via the pooled-screen have tissue transduction exceeding AAV9 in many organs **(Fig 5a,c)** and several variants exhibit drastic liver de-targeting **(Fig 4d, 6c)**, further engineering could be performed to enhance potency and specificity. Accordingly, peptides with at least 85% identity to any one of the peptide sequences provided herein are also contemplated. This may be relevant in particular for certain peptides derived from ligands that engage receptors expressed on multiple cell types or which have promiscuous binding activity. In the validation experiments, a standard promoter (CMV) was used to drive expression of the mCherry transgene. Alternatively, tissue-specific promoters could be used to increase the specificity and magnitude of transgene expression in the organ of interest. Furthermore, the hit capsids identified here could be further engineered for increased activity. Existing hits could serve as a scaffold for further rounds of targeted mutagenesis and screening, or peptides could be inserted on both loop1 and loop2 of the AAV capsid to increase the valency of the displayed ligands. Additionally, recently developed direct chemical engineering or peptide display strategies and/or alternative peptide linkers could be utilized to enhance peptide-mediated transduction. Also, scRNAseq could be used to screen hit

variants towards more specific cell-types within the organ of interest.

[0098] While the screening strategy addresses the gene therapy hurdle of organ targeting, the issue of pre-existing AAV immunity also can limit clinical translation. Therefore, experiments computationally identified and fully characterized a set of 4 immune-orthogonal AAV capsids **(Fig. 7a-d)**, and showed that insertion of a screen identified re-targeting peptide can alter the *in vivo* tropism of one of these identified AAVs **(Fig. 7e)**. This not only increases the number of known serotypes with interesting immunological properties to be investigated alongside the current AAV set, but also lends credence that the peptide insertions identified through the screening methodology can be grafted onto diverse biological scaffolds to achieve unique targeting across broad delivery contexts.

[0099] Taken together, the disclosure presents a massive functional screen of engineered AAV variants, spanning over one million total variants derived from two capsids and multiple sites of insertional mutagenesis. Using this screening data, validation of 21 AAV variants, identifying AAV capsids with increased organ transduction across multiple organs (heart, muscle, and lung for AAV9.PDGFC, **Fig. 4c**), as well as highly specific AAV capsids (AAV5.APOA1, AAV9.DKK1, **Fig. 5a**). Improved broad targeting AAV variants have massive potential for genetic diseases such as hemophilia A, where total factor VIII expression levels are most critical. At the same time, highly specific AAV capsids such as AAV5.APOA1 (which has less than 1% the liver infectivity of WT AAV9) would have great utility for neurodegenerative disorders, where maximizing transgene expression in the brain is essential. In addition to the novel variants identified herein, the bulk screening data itself is high value. Given the scale, reliability, and translational relevance of the screening dataset, the data set can serve as a foundation for future computational engineering of designer AAV capsids.

[00100] The following examples are intended to illustrate but not limit the disclosure. While they are typical of those that might be

used, other procedures known to those skilled in the art may alternatively be used.

EXAMPLES

[00101] Mice

[00102] All animal care and experimental methods were performed in accordance with the University of California Institutional Animal Care and Use Committee. 6-8 week old male C57Bl/6J (JAX, #000664) and Balb/cJ (JAX, #000651) mice were purchased from the Jackson Laboratories and systemic injections were administered retro-orbitally with either AAV or PBS.

[00103] Cell lines and culture conditions

[00104] All cells were cultured in a 37°C 5% CO2 humidified incubator. HEK293T cells were cultured in DMEM medium supplemented with 10% FBS, GlutaMAX (1x) (GIBCO), and Penicillin-Streptomycin (100 U/mL) (GIBCO).

[00105] Visualizing AAV capsid structures

[00106] To obtain structural representations of AAV capsids, AAV5 and AAV9 structure files were downloaded from the Protein Data Bank. These were visualized using the PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC with the *ramp_new* function utilized for coloring the surface capsid representation.

[00107] Design of displayed peptide library

[00108] Each AAV library consisted of 275,298 peptides, derived from 6,465 proteins. These protein sources were mined from a variety of protein families, including all protein ligands cataloged in the Guide to Pharmacology database, toxins, nuclear localization signals (NLS), viral receptor binding domains, albumin and Fc binding domains, transmembrane domains, histones, granzymes, and predicted cell penetrating motifs. In addition to peptides coding for functional biomolecules, 444 control peptides coding for FLAG-tags with premature stop codons were included. For all human proteins, the cDNA coding for each protein was fragmented *in silico* to generate DNA coding for all possible 20mer peptides. For viral proteins, cell penetrating motifs, and FLAG stop codon controls, the protein sequence was back-translated to DNA using the most abundant human codon for each amino acid.

[00109] Oligonucleotide array synthesis, amplification, and cloning

[00110] Oligonucleotide libraries were synthesized by GenScript as three 91,766 element pools. Each oligonucleotide library was amplified using KAPA Hifi Hotstart Readymix, and the manufacturer recommended cycling conditions with an annealing temperature of 60 °C and an extension time of 30 seconds. The number of PCR cycles was optimized to avoid over-amplification of the peptide libraries. After amplifying each oligonucleotide pool and confirming amplicon size on an agarose gel, the amplified sub-libraries were pooled to yield the total 275,298 element peptide library.

[00111] These pools were cloned into loop 1 or loop 2 sites to yield pAAV5L1_Screen, pAAV5L2_Screen, pAAV9L1_Screen, and pAAV9L2_Screen. The AAV *rep* and *cap* were flanked by AAV inverted terminal repeat (ITR) sequences to facilitate packaging of *cap* genes into a recombinant AAV particle.

[00112] Recombinant AAV production

[00113] Utilizing the library plasmid pools described above (AAV5-Loop1, AAV5-Loop2, AAV9-Loop1, and AAV9-Loop2), each AAV capsid library was produced by transfecting HEK293T cells in 40 15 cm dishes with the plasmid library pool (diluted 1:100 with pUC19 filler DNA to prevent capsid cross-packaging) and an adenoviral helper plasmid (pHelper). Titers were determined via qPCR using the iTaq Universal SYBR green supermix and primers binding to the AAV ITR region. To prepare the capsid particles as templates for qPCR, 2 µL of virus was added to 50 µL of alkaline digestion buffer (25mM NaOH, 0.2 mM EDTA) and boiled for 8 minutes. Following this, 50 µL of neutralization buffer (40mM Tris-HCl, .05% Tween-20, pH 5) was added to each sample.

[00114] In vivo evaluation of AAV display libraries

[00115] Each AAV capsid library was retro-orbitally administered to mice in duplicate at a dose of 2E12 vg/mouse for the AAV9-based libraries or 1E12 vg/mouse for the AAV5-based libraries. Two weeks after injection, the heart, lung, liver, intestine, spleen, pancreas, kidneys, brain, and gastrocnemius muscle were harvested and placed in RNAlater storage solution. Total DNA was extracted from all mouse tissues using TRIzol reagent and the TNES-6U back

extraction method. The resulting precipitated DNA was centrifuged for 15 minutes at 18,000G, and the supernatant discarded. The DNA pellet was then washed three times with 70% ethanol, and finally resuspended in 300 µL of EB after allowing the pellet to air dry.

[00116] Preparation of plasmid and capsid DNA for next generation sequencing

[00117] To sequence the plasmid libraries (AAV5/9 and loop1/2 peptide insertions), 50 ng of plasmid was used as template for a 50 µL KAPA Hifi Hotstart Readymix PCR reaction with an annealing temperature of 60 °C and an extension time of 30 seconds. The primers were designed to amplify the peptide coding region from each sub-library. The number of cycles was optimized to avoid overamplification. The PCR reactions were purified using a QIAquick PCR Purification Kit according to the manufacturer's protocol. Following this, 50 ng of the PCR amplicon was used as template for a secondary 50 µL KAPA Hifi Hotstart Readymix PCR reaction to add illumina compatible adapters and indices (NEBNext Cat# E7600S). The PCR reaction was performed with an annealing temperature of 60 °C, an extension time of 30 seconds. To sequence the capsid libraries, a similar protocol was performed, with a modified template amount in the step-1 PCR. To prepare the capsid particles as templates for PCR, 2 µL of virus was added to 50 µL of alkaline digestion buffer (25mM NaOH, 0.2 mM EDTA) and boiled for 8 minutes. Following this, 50 µL of neutralization buffer (40mM Tris-HCl, .05% Tween-20, pH 5) was added to each sample. 1 µL of this digested capsid mix was then used as a template for a 50 µL PCR reaction. For each sample, the number of cycles was optimized to avoid overamplification, and a secondary PCR was subsequently performed to add illumina compatible adapters and indices. After generating illumina compatible libraries, the plasmid and capsid samples were sequenced on a NovaSeq 6000 with an S4 flowcell generating 100bp paired end reads.

[00118] Preparation of tissue DNA for next generation sequencing

[00119] To sequence the AAV *cap* genes from each tissue for the pooled screen, as with the plasmid/capsid libraries, a two-step PCR based library prep method was used. For each organ and replicate, a 300 µL PCR reaction was performed with 120 uL of genomic DNA used as a template. For each tissue, the number of cycles was optimized via

an initial qPCR to avoid overamplification of the library. All other parameters such as primers, and melting temperatures were identical to the PCRs for the plasmid libraries. Following this initial PCR, a secondary PCR was performed as above to add illumina compatible adapters and indices. The libraries were then sequenced on a NovaSeq 6000 with an S4 flowcell generating 100bp paired end reads.

[00120] In vivo validation of AAV variants

[00121] Either saline or the AAV-variant-mCherry, AAV9-mCherry, or AAV5-mCherry capsids were systematically administered to mice in duplicate at a dose of 5E11 vg/mouse. Three weeks after injection, the lungs were inflated with a PBS/OCT solution and the lungs, heart, liver, intestine, spleen, pancreas, kidneys, brain, and gastrocnemius muscle were harvested. Each organ was split with one portion placed in RNAlater and the other embedded in OCT blocks and flash frozen in a dry-ice/ethanol slurry. Total RNA was then isolated from all mouse tissues using TRIzol reagent and RNA Isolation kits with on-column DNase treatment (Zymo Cat# R2072). cDNA synthesis was performed with random primers from the Protoscript cDNA synthesis kit (NEB Cat#E6560S). Transgene expression was then quantified via qPCR using the iTaq Universal SYBR green supermix and primers binding to the mCherry transcript. mCherry transgene expression was normalized to that of an internal GAPDH control, using GAPDH specific primers. For histological examination, OCT frozen blocks were cryosectioned at approximately 10 µm thickness and tissue slides were then imaged on an Olympus SlideScanner S200. Exposure times between 5-1000 ms were used, with identical exposure times used for all samples of a given tissue type. mCherry expression from histological sections was then quantified using the Olympus OlyVIA software to calculate the mean pixel intensity across the entire organ section.

[00122] Quantifying AAV variant abundance from NGS data

[00123] Starting with FASTQ sequencing files, the MAGeCK (*94*) 'count' function was used to generate count matrices describing AAV abundance in each sample (plasmids/capsids/tissues). Following this, the count matrices were normalized (via multiplication with a constant size-factor) for each sample to account for non-identical read depth. The sequencing counts were then transformed by taking

the log base 2 of the raw counts, after addition of a pseudocount. Variants with no counts across all of the experimental samples were excluded from analysis.

[00124] Biophysical analysis of AAV capsids

[00125] The biophysical characteristics of the inserted peptides was calculated using the "ProteinAnalysis" module within the Biopython Python package. A variant was considered a successful packager if it had higher abundance in the capsid particles compared to the plasmid pool. Support vector machine training and visualization was accomplished via the "svm" module within the sklearn Python package(*96*). UMAP projection of peptide biophysical characteristics was accomplished via the "plot" functionality within the UMAP Python package. All default parameters were used for the visualization. Boxplots and hexbin plots were generated using the matplotlib and seaborn Python packages.

[00126] Identifying significantly enriched variants in each tissue

[00127] To identify variants which successfully transduce each tissue, for each variant a one sample T-test was applied, comparing the abundance in the capsid particles to the abundance in the tissue. Resulting p-values were adjusted for multiple hypothesis testing via the Benjamini-Hochberg procedure. A variant was considered a significant transducer of an organ if it had an FDR adjusted p-value < .05, and a $Log_2FC > 1$ in both replicates. When choosing variants for validation experiments, variants were prioritized that had inserted peptides which were identified as hits in multiple capsid/loop contexts, and variants for which similar inserted peptides infecting the same organ were identified.

[00128] Visualizing tissue transduction from pooled screen

[00129] Heatmaps for visualizing AAV transduction were generated using the 'clustermap' function within the seaborn Python package. Rows and columns were ordered via the scipy 'optimal_leaf_ordering' function to minimize the euclidean distance between adjacent leaves of the dendrogram. UMAP projections visualizing AAV tissue specificity were generated by embedding the tissue level log_2 fold change into two dimensions via the "plot" functionality within the UMAP Python package. All default parameters were used for generating

the embedding. The variants were colored by the organ in which they had the max $\log_2$ fold change.

[00130] Assessing accuracy of predicted AAV variant tropism

[00131] For each variant which was individually validated, the accuracy of both positive and negative predictions of tissue infectivity were assessed. For variants predicted to target a specific organ, a prediction was considered accurate if the individual validations showed greater than 50% of wild-type AAV9 infectivity in that organ. For variants predicted not to target a specific organ, a prediction was considered accurate accurate if the individual validations showed less than 50% of wild-type AAV9 activity.

[00132] Peptide Distance Projections

[00133] To calculate the Levenshtein distance between inserted peptides, the "levenshtein" function from the Python package "rapidfuzz" was used with default parameters(*100*). After building the pairwise distance matrix between all significantly enriched peptides, the matrix was projected into two dimensions via UMAP with metric="precomputed" , n_neighbors=1500, and min_dist=.1. Clusters of peptides with similar functions were then hand annotated onto the resulting plot.

[00134] Convolutional Neural Networks

[00135] To train convolutional neural networks (CNNs) to predict the tissue specificity of AAVs, the AA sequences of the inserted peptides were converted to a one-hot encoding via the "get_dummies" function from the pandas Python package. Among the significantly enriched variants, a variant was considered a transducer of a given organ if the $\log_2$FC relative to the capsid in both replicates was greater than 0. The data was then randomly split into training (⅔) and validation (⅓) datasets. For each variant, the one-hot encoding was reshaped to a 20x20 matrix with rows indicating residue positions, and columns indicating the presence or absence of a particular amino acid. The model architecture was instantiated via a Keras sequential model(*102*). In brief, a convolutional layer (Conv1D) with 32 filters, a kernel size of 3 and "relu" activation was fed into a max pooling layer (MaxPool1D) with pool size of 2. These layers were followed with another set of convolutional and max

pooling layers, this time with 64 filters in the convolutional layer. These layers were followed with a dense layer with units=20. Finally a dropout layer was added with the dropout rate=.5. A flattening layer and final dense layer (with sigmoid activation) was then used to output resulting class probabilities. A separate independent model was trained for each organ. When training the models, the classes (infective versus non-infective variants) were weighted proportionally to the inverse of the number of class examples. When calling 'model.fit()' to train the CNN, a dictionary describing the class weights was passed via the 'class_weight' parameter. Model performance was evaluated via accuracy, area under the receiver operator characteristic curve (AUROC), F1-score, and Matthews Correlation Coefficient (MCC). Metrics were calculated via builtin Keras functions, and plotted via matplotlib. To assess how model accuracy changes as a function of the edit distance from the training data, the levenshtein distances between peptides in the testing and training datasets were calculated using the 'rapidfuzz' python package as above.

[00136] Transmission electron microscopy (TEM)

[00137] To obtain transmission electron microscopy images of select AAV variants, 20µL of the AAV solution was applied to formvar/carbon coated EM grids. Upon washing 3 times with water, the sample-containing grid was then negative-stained for 1 minute using a solution of 2 % of Uranyl acetate in water and blot-dried. The EM grids with each AAV variant were then imaged at 68,000X magnification using the FEI Tecnai Spirit G2 BioTWIN transmission electron microscope operated at 80keV.

[00138] Mechanism knockout experiments

[00139] sgRNA sequences targeting LRP6 or non-targeting controls were identified using CRISPick and cloned into the lentiCRISPR v2 plasmid backbone. Lentivirus was then produced as described in (*33*). Briefly, HEK293T cells were seeded at ~40% confluency the day before transfection. The day of transfection, Optimem serum reduced media was mixed with Lipofectamine 2000 (Thermo Fisher), 3 µg of pMD2.G plasmid, 12 µg of pCMV deltaR8.2 plasmid, and 9 µg of the respective lentiCRISPRv2 plasmid and added drop-wise onto the HEK293T cells following a 30 minute incubation period. 48 hours after

transfection, the media was collected and replaced with fresh DMEM with 10% FBS. At 72 hours post-transfection, the supernatant containing viral particles were collected again, pooled, and concentrated to 1mL using Amicon-15 centrifugal filters with a 100,000 NMWL cutoff (EMD Millipore). For lentiviral transduction, HEK293T cells were seeded in a 12-well plate at ~20% confluency the day before transduction. The day of transduction, lentiviral containing DMEM with 8µg/mL polybrene was added to the cells. The media was then replaced 24 hours later and then changed into puromycin (2µg/mL) containing DMEM 28 hours post-transfection. Post selection, once the cells reached confluency, they were passaged into a 24-well plate at ~40% confluency. 24 hours later, DMEM containing either AAV9 ($4x10^9$ viral genomes) or AAV9.DKK1 ($1x10^9$ viral genomes) was added to the cells. The cells were then collected 24 hours later and flow cytometry was performed to quantify mCherry transgene expression.

[00140] Identifying immune orthogonal AAV capsids

[00141] Potential AAV orthologs were curated by first identifying sequences which exhibited similarity to the AAV2 *cap* gene using the National Center for Biotechnology Information (NCBI) basic local alignment search tool (BLAST). Next, incomplete, heavily truncated, and highly homologous sequences to one another were filtered out of the selection criteria. Viruses within the human AAV clade and those from non-mammalian hosts were then filtered out. Finally, sequences with high similarity to previously identified human serotypes were removed, resulting in the 23 potential AAV orthologues assessed.

[00142] Immune orthogonal AAV production

[00143] To clone computationally identified AAV orthologs, the capsid sequences were codon-optimized and cloned downstream of the AAV2 *rep* gene using Gibson Assembly. Immune orthogonal AAV capsids were produced as described for the AAV variant validation capsids and AAV production titer was measured via qPCR with primers binding to the AAV ITR region. Any ortholog with a production titer within a power of 10 of AAV5 was considered to have successfully packaged.

[00144] Immune orthogonal AAV in vivo transduction

[00145] Immune orthogonal AAV capsids with sufficient packaging titer were then injected retro-orbitally into C57BL/6 mice at a dose

of $1x10^{12}$ viral genomes/mouse. Livers were harvested 3 weeks post-injection and total RNA was isolated as described above. cDNA was then generated and transgene expression was quantified via qPCR using the iTaq Universal SYBR green supermix and primers binding to the mCherry transcript. mCherry transgene expression was then normalized to GAPDH and the relative expression was compared to AAV5.

[00146] Assessing immune cross-reactivity

[00147] The assay to assess immune antibody cross-reactivity of the identified immune orthogonal was performed as previously described. Prior to injection, serum was collected via tail snip procedure and then the mice were injected with $1x10^{12}$ viral genomes/mouse of AAV or PBS in triplicate. 3 weeks later, serum was collected from each of the mice and the antibody cross-reactivity ELISA was performed. For this, $1x10^{9}$ viral genomes of AAV8, AAV MM2, AAV MG1, AAV MG2, or AAV CH1 were diluted in a 1x coating buffer and incubated overnight in each well of 96-well Nunc MaxiSorp plates. Plates were washed three times for 5 mins with 1x wash buffer (Bethyl) and blocked with 1x BSA blocking buffer (Bethyl) for 2 hours at room temperature. The wells were then washed again and serum samples were added at a 1:40 dilutions. Plates were incubated for 5 hours at 4℃ with shaking. Wells were 3x washed and 100 µL of HRP-labeled goat anti-mouse IgG1 (Bethyl; diluted 1:100,000 in 1% BSA) was added to each well. Secondary antibody was incubated for 1 hour at room temperature, wells were washed 3 times, and 100 µL of TMB substrate was added to each well. Optical density at 450 nm was measured using a microplate absorbance reader (BioRad iMark).

[00148] Engineering immune orthogonal AAV variants

[00149] To engineer the immune orthogonal 'AAV MG2' capsid, the AAV9 capsid amino acid sequence was aligned to the MG2 sequence using Clustal Omega (*108*). The PDGFC peptide coding sequence was then ligated into the 'AAV MG2' vector at the appropriate Loop 1 location with an identical protocol as was done for cloning peptides into AAV9. The engineered 'AAV MG2.PDGFC' vector capsid was then assayed *in vivo* identically to the above validation experiments using AAV9.

[00150] Ethical Compliance

[00151] All experiments involving live vertebrates performed at UCSD were done in compliance with ethical regulations approved by the UCSD IACUC committee.

[00152] **Statistics**

[00153] Unless otherwise noted, differences in means were calculated via an unpaired t test. When necessary, p values were adjusted for multiple hypothesis testing via the benjamini-hochberg procedure(*99*). For all Figures: unless otherwise noted *p<0.05, **p<0.01, ***p<0.001, ****p<0.0001.

[00154] **A systematic library of AAV variants displaying fragmented proteins**

[00155] To create the libraries AAV5 and AAV9 were chosen as the starting serotypes. This was due to their established clinical utility as well as two key characteristics: one, AAV5 is more evolutionarily distant to other AAV serotypes in clinical use, and has previously been shown to be immune orthogonal (to other prevalent AAV serotypes thereby enabling their sequential redosing); and two, AAV9 has been used extensively for clinical trials and has been shown to cross the blood-brain barrier outperforming other AAV serotypes in most tissues. To generate the library of diverse AAV variants, a DNA oligonucleotide pool of 275,298 gene fragments was generated (**Fig. 1a-b**). Each gene fragment coded for a 20 amino acid peptide derived from the coding sequence of ligands with known extracellular receptors, or a gene predicted to have cell-penetrating or internalizing properties (**Fig. 1a-b**). Protein ligands were sourced from the Guide to Pharmacology database, an expertly curated list of pharmacological targets and their associated ligands, and cell-penetrating/internalizing functionality was inferred through text mining of UniProt entries. Examples of protein classes identified as having potential internalizing function included toxins, histones, granzymes, viral receptor binding domains, and nuclear localization signal domains (NLS). Mouse and human genomes share 80% of their protein coding genes, with 85% amino acid sequence identity between orthologs. These genomic similarities extend to the receptors for ligands in the libraries (e.g., 86% of mouse GPCR proteins have human orthologs), implying many human ligands will be similarly functional across human and

mice contexts. After synthesizing the pool of single-stranded oligonucleotides coding for these gene fragments, it was amplified to double-stranded DNA via PCR, digested, and ligated into two distinct locations on the AAV5 and AAV9 *cap* genes. Seamless cloning was enabled by type IIS restriction enzymes cutting outside their recognition sequence. PaqCI sites engineered to yield compatible overhangs (coding for glycine-serine 2 amino acid linkers) were inserted at the ends of the peptide coding DNA library and on the AAV5/AAV9 *cap* plasmid DNA at sites coding for two distinct surface loops, hereon referred to as loop 1 and loop 2. Surface loop 1 (AA 443 and 456 on AAV5 and AAV9, respectively) and loop 2 (AA 576 and 587 on AAV5 and AAV9, respectively) were chosen as peptide insertion locations due to their distance from the viral particle core facilitating potential receptor engagement (**Fig. 1b**). While many previous AAV engineering efforts have focused on inserting peptides at surface loop 2, loop 1 sites have been shown to accommodate large insertions including even a full length fluorescent protein. Collectively, four libraries of variants were produced, spanning two AAV capsids, with two loop insertion sites each. In addition to protein coding gene fragments, 444 stop codon containing gene fragments were included as negative controls. This defined library synthesis methodology was adopted to enable quantitative inference of variant packaging and transduction efficiencies. The starting plasmid libraries were sequenced to establish initial relative variant abundances, and packaging efficiencies were quantified via comparison to this initial baseline.

[00156] Biophysical drivers of AAV capsid formation

[00157] To quantify how well different AAV *cap* variants package into functional capsids, recombinant AAV particles were generated with the engineered AAV5 and AAV9 *cap* plasmid libraries via transient triple transfection of HEK293T cells (**Fig. 2a**). These viral particles were treated with benzonase to degrade residual plasmid DNA, and then subjected to next generation sequencing (NGS) to quantify relative variant abundance. Packaging efficiency was quantified by ranking AAV variants by the $\log_2$ fold change ($\log_2$FC) of their relative capsid abundance compared to their count in the plasmid pool (**Fig. 2b**). Utilizing this method, over 250,000 AAV

variants were identified which package efficiently into functional AAV capsids with a $log_2FC > 0$. To validate this packaging metric quantified from screening data, 25 AAV capsids were produced including 23 identified as successful packagers ($log_2FC > 0$) and 2 identified as non-packagers ($log_2FC < 0$). In these individual validations, the 2 non-packagers yielded >10-fold lower titer than those identified as packagers, thus providing confidence in the AAV packaging metric. Consistent with their disruption of the AAV capsid structure there was also a depletion of non-functional stop codon control AAV variants in the capsid pool, and importantly this confirmed lack of library cross-packaging during AAV production(**Fig. 2C**).

[00158] To better understand what features drove successful capsid formation, the biophysical characteristics of the inserted peptides that yielded AAV variants that packaged successfully were analyzed. It was found that peptide charge, flexibility, alpha helical content, and hydrophobicity were all significantly different in packaging AAV variants versus AAV variants unable to package (**Fig. 2d**). The set of successfully packaged variants had a narrower charge distribution than the variants unable to package, suggesting peptides with extreme charge densities have a negative impact on capsid formation. Successfully packaging AAV variants also had inserted peptides with higher flexibility, lower alpha-helical content, and lower hydrophobicity than the variants unable to package. The observed depletion of hydrophobic peptide displaying variants is consistent with the solvent exposed nature of the AAV surface loops.

[00159] To build an integrated model predicting if AAV variants will package based on the biophysical features of the inserted peptides, a support vector machine classifier was trained using the charge, flexibility, alpha helical content, and hydrophobicity of the peptides in the dataset (**Fig. 2e**). While all of these biophysical features were significantly different when comparing packaging versus non-packaging AAV variants, the magnitude of this difference was relatively modest for each individual feature **(Fig. 2d)**. However, collectively these features were sufficient to train a model which could differentiate between packaging and non-packaging

AAV variants (area under the receiver operating characteristic curve = 0.89). For each AAV variant, embedding the inserted peptide's charge, flexibility, alpha helix content, and hydrophobicity into two dimensions using uniform manifold approximation and projections (UMAP) enabled visualization of this class separability (**Fig. 2f**). The resulting embedding laid out AAV variants into distinct clusters, indicating that while each underlying biophysical feature is continuous, there are separable groups of AAV variants with similar biophysical features. AAV variants which package tended to cluster with other packaging variants in this unsupervised embedding, further supporting the predictive power of these four biophysical features. This thorough quantification and analytical framework for assessing AAV packaging, which is enabled by the diversity and length of the inserted peptide library, is of critical translational importance for identified AAV variants as AAV production costs are directly related to their packaging capability.

[00160] High-throughput mapping of engineered AAV tissue tropism

[00161] Having produced libraries of recombinant AAV particles packaging their own *cap* genes, next four viral pools were injected into C57BL/6 mice in duplicate (**Fig. 3a**). After two weeks, mice were sacrificed and the peptide-containing region of the AAV *cap* gene was amplified from DNA isolated from mouse liver, kidney, spleen, brain, lung, heart, skeletal muscle, intestine, and pancreas. The relative abundance of each AAV variant via NGS was quantified, and the $\log_2$ enrichment calculated for each variant relative to its abundance in the capsid pool, observing good correlation between replicates (**Fig. 3b**). Using this data, over 15,000 variants were identified which efficiently infected at least one mouse tissue (**Fig. 3c**). The spleen and liver were the most frequent tissue targets of the infectious variants, consistent with the established wild-type (WT) tropism of AAV5 and AAV9 towards the liver, as well as more recent research showing AAV5 and AAV9 readily transduce the spleen. The fewest AAV variants targeting the skeletal muscle and brain were identified, in line with the high therapeutic AAV doses needed to achieve clinical efficacy for muscle targeting gene therapies, and the challenge of delivery across the blood-brain barrier. A substantial number of the AAV variants were comprised of the same

peptide inserted across different AAV capsids and insertion sites, giving credence to the hypothesis that tropism reprogramming is, at least partially, peptide-specific (**Fig. 3c**).

[00162] Towards mapping and understanding the tissue-transduction patterns mediated by the inserted peptides, the feasibility of training predictive models linking inserted peptide sequence to tissue tropism were analyzed. Inspired by contemporary work using convolutional neural networks (CNN) to predict antibody specificity, a CNN multi-label classifier was trained to predict AAV tissue tropism using one-hot encoded inserted peptide sequences as input features (**Fig. 3d**). To evaluate performance, the model was trained using a random selection of ⅔ of the significantly enriched AAV variants, and evaluated on the ⅓ testing dataset. This CNN model architecture had good performance across all organs, with a minimum accuracy of 72% in the kidney (**Fig. 3d**). The highest F1 scores and Matthews Correlation Coefficients (MCC) were observed for the liver and spleen, likely due to the high number of liver and spleen targeting variants identified in the pooled screen (**Fig. 3c**). It was also observed that for the liver and spleen, the CNN had only modest reduction in F1 scores when evaluated on peptides very different (greater than 10 edits away) form any training examples, implying the models had learned generalizable features relevant to AAV transduction. For other organs, the models were much less accurate when predicting the tissue tropism of peptides very different from the training data, likely due to the reduced amount of positive training examples. Collectively, the ability to predict AAV tropism supports the conclusion that the inserted peptides are mediating retargeting of tissue-tropism, and that there is a learnable relationship between peptide sequence and tissue-tropism.

[00163] **Engineered AAV variants with clinically relevant tissue tropism**

[00164] Like the WT scaffolds from which they were derived, transduction of multiple organs is a near ubiquitous phenotype among the infectious AAV variants identified (**Fig. 4a**). Significant transduction of the liver and spleen was observed for the majority of infectious variants regardless of which other organs were co-transduced. This is true even for variants with insertions in

surface loops known to be involved in WT capsid receptor binding. While liver and spleen targeting were near ubiquitous, allowing for identifying variants which specifically target the liver/spleen plus one other organ, as well as variants which transduced all tissues at high levels (**Fig. 4a**). When variants were hierarchically clustered based on their tissue detection levels, variants derived from the same sub-library tended to cluster together, suggesting that the tissue specificity of the wild-type scaffold was at least partially a determinant of engineered variant tropism. Hierarchical clustering of the organ samples resulted in replicates clustering together, giving further confidence to the reliability of the screen results. To visualize the overall screen results, the tissue detection levels for each variant were embedded into two dimensions using UMAP, coloring the variants by the organ they most readily transduce (**Fig. 4b**). In this reduced dimensional space, organ specific clusters can be readily identified, with the liver and spleen targeting variants especially prominent.

[00165] To confirm the tissue tropism of the novel AAV variants that were identified via the pooled screen, 21 variants were individually produced and individually validated by quantifying their ability to package and deliver an mCherry transgene in vivo (**Fig. 4c**). All 21 variants were significantly enriched in at least one organ, and variants for validation which were internally consistent within the screening data were prioritized: consistent AAV capsids being defined as hits where other variants with similar inserted peptides enriching in the same organ were identified. Variants were initially characterized by *in vivo* quantification of tissue tropism at the mRNA level (via RT-qPCR quantification of mCherry transgene expression) (**Fig. 4c**). The tissue tropism of the variants largely recapitulated the screen predictions (**Fig. 4c**), with 74.3% of the tissue tropism predictions matching expectations. AAV variants which specifically targeted hard to infect organs such as the muscle, lung, and brain, while simultaneously de-targeting away from the liver were identified (**Fig. 4c**). Notably, protein level quantification of mCherry delivery to the liver, quantified via fluorescent microscopy, confirmed excellent concordance between mRNA and protein measures of tissue transduction (R^2=0.97)(**Fig. 4d**).

Altogether, across the variants individually tested, 9/21 variants were found to exceed AAV9 infectivity in at least one organ. Variants were identified which exceed AAV9 infectivity in all organs except the liver and pancreas, which had max relative transduction of 98.8% of WT AAV9 and 82.2% of WT AAV9, respectively. Additionally, 18/21 variants had less than half the liver transduction of WT AAV9, with three variants below 5% AAV9 liver transduction levels (**Fig. 4d**) indicating clinically important liver de-targeting. To increase confidence in the fidelity of the individually validated variants in C67BL/6 mice, 3 of the above 21 variants were individually validated in BALB/c mice. Notably, their relative tropism was consistent across the two mouse strains.

[00166] Mechanistic insights into AAV reprogramming

[00167] Having confirmed the efficacy of the AAV variants, experiments were performed to explore the mechanisms underlying their reprogrammed tropism, and in particular, investigate whether the hypothesis that AAV variants displaying peptides derived from distinct ligand domains could indeed drive their transduction patterns. First, experiments were performed to interrogate how peptides derived from specific ligand protein regions alter AAV tropism. This was motivated by observation that tiled peptide enrichment patterns in a screen can yield insight into functional domains of the proteins from which they were derived. Analysis of how the peptides were derived was performed. To do this, experiments were performed to interrogate tissue-specific AAV variants identified in the screen, focusing first on a lung-specific variant (termed AAV9.DKK1) that was derived via display of a peptide fragment form the DKK gene on the AAV9-Loop2 scaffold. Specifically, as shown in **Fig. 5a (top left)**, it was observed that significantly enhanced lung transduction from peptides derived from a specific domain in the N-terminus of the protein. Upon confirming that this variant is packaged into a functional capsid via electron microscopy **(Fig. 5a, top right)**, next experiments quantified its transduction profile *in vivo*. This confirmed that AAV9.DKK1 was indeed an efficient lung transducing variant with greater than 2-3 fold higher expression than AAV9, and with consistent de-targeting across all

other organs compared to AAV9 when quantified at both the RNA and protein level **(Fig. 5a, bottom left and right).**

[00168] Next, experiments were conducted on the DKK1 displayed peptide to determine insight into the mechanism of transduction. Interestingly, the N-termini region of lung enrichment **(Fig. 5a, top left)** was centered on an evolutionarily conserved, linear peptide motif mediating binding to the low density lipoprotein receptor-related proteins 5 and 6 (LRP5/6). As HEK293T cells robustly express LRP6, thus lentiviral-mediated CRISPR-Cas9 was utilized to disrupt LRP6 expression in these cells to investigate the relative transducability of the AAV9.DKK1 variant in this system **(Fig. 5b).** Specifically, cells transduced with either non-targeting control (NTC) guides, or *LRP6* targeting guides were subsequently transduced with either AAV9.DKK1 or wild-type AAV9. Using flow cytometry to quantify AAV delivery of the mCherry transgene, it was observed that a significant reduction in the infectivity of AAV9.DKK1 in the LRP6 knockout population, with no corresponding reduction in the infectivity of wild-type AAV9 **(Fig. 5b).** The LRP6 dependent *in vitro* infectivity of AAV9.DKK1, combined with the known interaction between LRP6 and the N-termini of DKK1 (see crystal structure, **Fig. 5b),** and its known robust expression in lung alveolar cells, to together gives strong support to the hypothesis that LRP6 mediates, at least in part, the reprogrammed tropism of AAV9.DKK1.

[00169] Building on these observations, and focusing on the brain as an exemplar, experiments were performed to examine more broadly the extent to which the inserted peptides were mediating re-targeting. First, a distance matrix quantifying the similarities between all peptide hits identified as significantly enriched in at least one organ was generated (**Fig. 6a**). By projecting this distance matrix into two dimensions, allowed for visualization of distinct clusters of homologous inserted peptides which infect the brain. It was found that families of similar inserted peptides yielded similar transduction rates (**Fig. 6a**). This effect was not limited to a particular capsid or insertion site, insofar as inserted peptides were often functional across all tested capsids and insertion sites (**Fig. 6a**). The observation that inserted peptides yielded consistent phenotypes across multiple capsids suggests that re-targeting is

directly due to a peptide mediated mechanism. This result also highlights the power of peptide tiling library designs, insofar as having multiple overlapping peptides with similar sequences can function as internal controls, imparting confidence in the identification of particular AAV variant hits.

[00170] Experiments were performed to specifically interrogated two brain targeting variants (AAV5.APOA1 and AAV9.APOA1) which contained the same inserted apolipoprotein-A 1 (APOA1) derived peptide. Both variants were confirmed to package into functional capsids by electron microscopy (**Fig. 6b**). Across the entire APOA1 protein, functional APOA1-derived peptides were primarily from a tandem repeat region at the C-terminus of the protein (**Fig. 6b).** Highlighting the peptide-specific effect of altered AAV tropism, it was observed that strong detargeting away from the liver at less than 2% of AAV9 levels for both variants, with unperturbed brain transduction relative to wild-type AAV9 (**Fig. 6c**). This was especially remarkable for AAV5.APOA1 as the engineered variant was able to overcome the limited transduction of its parental AAV serotype across the blood-brain barrier. AAV5.APOA1's improved specificity was subsequently shown to be strain independent as it retained both its brain targeting and liver detargeting functionality in BALB/c mice (FIG. 6d).

[00171] **Engineering re-targeted immune orthogonal AAV vectors**

[00172] To generate additional variants with clinically relevant features, the AAV engineering approach was expanded to include novel capsids beyond AAV5 and AAV9. In particular, the potential of engineering tropism of highly divergent AAV orthologs was explored. This was motivated by the fact that one of the biggest challenges with AAV therapy is the associated immune responses: either, pre-existing immunity that currently excludes a significant fraction of the human population from being eligible for any AAV therapy; or induced immunity upon AAV injection which prevents subsequent AAV re-dosing. This makes it extremely difficult to create effective therapies since it only allows a single opportunity to treat a patient, thus often mandating extremely high titers for a one-time therapy which in turn can lead to dangerous levels of toxicity. Recently, the concept of immune orthogonality to address this

challenge has been performed. Specifically, it has been suggested that an orthologue, given sufficient sequence divergence, will not cross-react with the immune response generated by exposure to other orthologues, thereby allowing re-dosing that avoids neutralization by existing antibodies or clearance of treated cells by activated cytotoxic T cells. While in that study experiments focused on commonly used AAV orthologs, experiments were performed to explore highly divergent natural orthologs from the full spectrum of mammalian species, including those which have not yet been thoroughly tested for use as *in vivo* vectors. Towards this a computational experimental pipeline was established to identify novel immune-orthogonal AAV serotypes. First, a basic local alignment search tool (BLAST) was used to identify 687 capsids with sequence homology to the AAV2 *cap* gene. This initial list was then filtered to exclude truncated genomes, redundant samples, human and non-mammalian serotypes, as well as close orthologs, resulting in 23 AAV capsid sequences for experimental investigation **(Fig. 7a)**. Analyzing the evolutionary distance of these potential AAV orthologs, it was confirmed that they exhibit high dissimilarity from all wild-type AAV capsids (except AAV5 which had previously been shown to be immune-orthogonal), with many exhibiting less than 60% sequence similarity to AAV2 **(Fig. 7b)**. To investigate if these computationally predicted AAV orthologs were functional, their ability to package into capsids was examined. Comparing titers to AAV5, 11 of the 23 AAV capsids were identified with packaging titers within 10-fold of AAV5 **(Fig. 7c)**. Next, to evaluate *in vivo* transduction potential, these 11 AAV capsids carrying a mCherry transgene were administered individually to C57BL/6 mice. Given that most AAV capsids have at least a degree of liver transducing potential, the livers of these mice were harvested 3 weeks post infection. Of the 11 AAV capsids analyzed, 4 of them had detectable levels of mCherry in the liver with 2 of them generating 3-fold higher mCherry expression than AAV5 **(Fig 7c)**.

[00173] Having identified 4 functional AAV capsids, their immune-orthogonality was characterized by injecting C57BL/6 mice with each of these and collecting serum at day 0 and day 21 timepoints **(Fig. 7d)**. AAV8 was selected for comparison in this study due to its

propensity for transducing the liver and its similar serological profile to other wild-type serotypes. Performing an ELISA at the 3-week timepoint confirmed that antibodies developed against the 4 studied immune-orthogonal AAVs indeed exhibited no immune cross-reactivity with AAV8 **(Fig. 7d)**. This result highlights the utility in combining computational and experimental methods to identify and characterize novel AAV serotypes which could be repurposed to enable gene therapy redosing.

[00174] Next experiments were performed to assess whether the tropism of the mined AAV immune-orthogonals could be re-engineered through insertion of one of the screen identified hit peptides. The muscle targeting AAV variants had broader tissue-tropism than the brain and lung targeting variants, insofar as they also readily infected the heart, lung, intestine, and spleen at level comparable to, or exceeding, AAV9, it was hypothesized that grafting the peptide from one of these variants onto an AAV ortholog would result in enhanced muscle targeting with minimal effects on parental AAV infectivity. To assess this AAV9.PDGFC was selected as a candidate variant, and confirmed its ability to form functional capsids (FIG. 6d) and enable efficient targeting of muscular tissue (with >10 fold expression above AAV9 as quantified by RT-qPCR)(Fig. 6d). The skeletal muscle and cardiac muscle (heart) transduction enhancement relative to AAV9 was also confirmed by protein level visualization (Fig. 6d). Grafts where then performed of the 20-mer PDGFC derived peptide onto one of the surface loops of the immune orthogonal AAV MGS (Fig. 7e), and notably observed similar enhanced muscle transduction of the engineered variant above its parental AAV (FIG. 7e). Taken together, the result highlights the fidelity of the ligand tiling display approach and further lends credence to the exciting possibility that displayed peptides identified in the initial screening approach could be utilized to re-engineer a diverse clade of AAV serotypes via peptide transfer.

[00175] It will be understood that various modifications may be made without departing from the spirit and scope of this disclosure. Accordingly, other embodiments are within the scope of the following claims.

WHAT IS CLAIMED IS:

1. A method of improving tropism of a virus or other delivery agent, the method comprising identifying ligand protein sequences derived from all known receptor-interacting ligands; systematically tile the ligand peptides into 5-50 or 10-20 or 20 amino acid peptides which are inserted into surface-exposed loops of AAV capsids; assessing the engineered capsids for their packaging capacity, in vivo tropism, and enhanced protein interactions.

2. The method of claim 1, wherein the virus is an adeno-associated virus (AAV).

3. The method of claim 2, wherein the AAV is selected from the group consisting of AAV1, AAV2, AAV5, AAV6, AAV7, AAV8, and AAV9.

4. The method of claim 3, wherein the AAV is AAV5 or AAV9.

5. The method of claim 3, wherein peptide sequences were generated via pooled oligonucleotide synthesis and inserted into 4 distinct loop regions: AAV5-Loop1 (N443), AAV5-Loop2 (S576), AAV9-Loop1 (Q456), and AAV9-Loop2 (A587) to generate over 1 million AAV variants.

6. The method of claim 3, wherein a 20'mer ligand peptide is inserted into one or both of 2 surface-exposed loops of AAV5 (SEQ ID NO: 2) or AAV9 (SEQ ID NO: 4).

7. A recombinant vector comprising capsid proteins containing one or two peptides inserted into one or both of two surface-exposed loops in the capsid protein, wherein the vector has a desired tropism or immune-orthogonality and wherein the peptides are independently selected from the group consisting of SEQ ID NOs:5 to 820 and 870-911.

8. The recombinant vector of claim 7, wherein the vector is an adeno-associated virus (AAV).

9. The recombinant vector of claim 8, wherein the AAV vector is an AAV5 serotype.

10. The recombinant vector of claim 9, wherein the AAV5 comprises a capsid protein having a sequence as set forth in SEQ ID NO:2.

11. The recombinant vector of claim 8, wherein the AAV vector is an AAV9 serotype.

12. The recombinant vector of claim 11, wherein the AAV9 comprises a capsid protein having a sequence as set forth in SEQ ID NO: 4.

13. The recombinant vector of claim 7, wherein the vector has tropism to pancreas, heart, brain, lung, liver, kidney, muscle, spleen or intestine.

14. The recombinant vector of claim 10 or 12, wherein a peptide of SEQ ID NOs:530-820, 870-910 or 911 is inserted into loop 1 and/or loop 2 of an AAV5 capsid or an AAV9 capsid.

15. The recombinant vector of claim 7, wherein the vector is immune orthogonal.

16. The recombinant vector of claim 15, wherein the vector is an adeno-associated virus (AAV).

17. The recombinant vector of claim 16, wherein the AAV comprises a capsid protein of any one of SEQ ID NOs:822, 824, 826, 828, 830, 832, 834, 836, 838, 840, 842, 844, 846, 848, 850, 852, 854, 856, 858, 860, 862, 864, or 866, or a sequence that it at least 85% to 99% identical to any of the foregoing sequences.

18. A method of making a delivery vehicle or vector with a desired tropism, the method comprising selecting a peptide sequence from any one of SEQ ID NOs: 5-820 or 870-911 and (i) cloning a nucleic acid sequence encoding the peptide into a coding sequence for a capsid

protein at an exposed loop site to obtain a recombinant capsid coding sequence and producing the vector using the recombinant capsid coding sequence or (ii) inserting the peptide into an exposed surface of the delivery vehicle.

19. The method of claim 18, wherein the vector is an adeno-associated virus (AAV).

20. The method of claim 19, wherein the AAV is selected from the group consisting of AAV1, AAV2, AAV5, AAV6, AAV7, AAV8, and AAV9

21. The method of claim 20, wherein the AAV is AAV5 or AAV9.

22. The method of claim 21, wherein the AAV5 capsid coding sequence comprises SEQ ID NO:1.

23. The method of claim 21, wherein the AAV9 capsid coding sequence comprises SEQ ID NO:2.

24. An AAV vector wherein the capsid protein has been modified according to the method of claim 18.

25. An AAV vector comprising a capsid protein wherein the capsid protein expresses a peptide of any one of SEQ ID NO: 5-820 or 870-911 in surface exposed loop 1 and/or loop 2 of the capsid protein.

26. The AAV vector of claim 25, wherein the wild-type capsid protein sequence comprises SEQ ID NO:2 or 4.

27. The AAV vector of claim 25, wherein the AAV vector has a AAV5 or AAV9 serotype.

28. The AAV vector of claim 25, wherein the vector has a desired tropism.

29. The AAV vector of claim 28, wherein the AAV vector has a tropism to pancreas, heart, brain, lung, liver, kidney, muscle, spleen or intestine.

30. A viral vector having a capsid protein comprising a heterologous targeting peptide in a range of 10-30 amino acids in length inserted into a surface exposed portion of a capsid protein and wherein the targeting peptide is set forth in any one of SEQ ID NOs: 5-820 or 870-911.

31. The viral vector of claim 30, wherein the heterologous targeting peptide is about 15-25 amino acids in length.

32. The viral vector of claim 31, wherein the heterologous targeting peptide is about 20 amino acids in length.

33. The viral vector of claim 30, wherein the viral vector is an adeno-associated virus (AAV).

34. The viral vector of claim 30, wherein the viral vector is a lentiviral vector.

35. The viral vector of claim 30, wherein the capsid protein is a VP1 capsid protein.

36. The viral vector of claim 30, wherein the capsid protein is a VP2 capsid protein.

37. The viral vector of claim 30, wherein the capsid protein is a VP3 capsid protein.

38. The viral vector of claim 30, wherein the heterologous targeting peptide is inserted into an AAV capsid protein at loop 1 and/or loop 2.

39. The viral vector of claim 30, wherein the viral vector is an AAV5.

40. The viral vector of claim 30, wherein the viral vector is an AAV9.

41. The viral vector of claim 39, wherein the heterologous targeting peptide is flanked by a linker peptide at the N-terminal and C-terminal ends of the heterologous targeting peptide.

42. The viral vector of claim 40, wherein the heterologous targeting peptide is flanked by a linker peptide at the N-terminal and C-terminal ends of the heterologous targeting peptide.

43. The viral vector of claim 30, wherein the heterologous targeting peptide targets the viral vector to hepatocytes or liver tissue.

44. The viral vector of claim 30, wherein the heterologous targeting peptide targets the viral vector to neuronal cells or brain tissue.

45. The viral vector of claim 30, wherein the heterologous targeting peptide targets the viral vector to pancreatic cells or pancreas tissue.

46. The viral vector of claim 30, wherein the heterologous targeting peptide targets the viral vector to cardiac cells or heart tissue.

47. The viral vector of claim 30, wherein the heterologous targeting peptide targets the viral vector to lung tissue.

48. The viral vector of claim 30, wherein the heterologous targeting peptide targets the viral vector to intestinal tissue.

49. The viral vector of claim 30, wherein the heterologous targeting peptide targets the viral vector to spleen tissue.

50. The viral vector of claim 30, wherein the heterologous targeting peptide targets the viral vector to renal cells or kidney tissue.

51. The viral vector of claim 30, wherein the heterologous targeting peptide targets the viral vector to muscle cells or tissue.

52. An adeno-associated virus (AAV) capsid protein comprising a heterologous targeting peptide cloned into loop 1 and/or loop 2 of the capsid protein, wherein the heterologous targeting peptide is about 10-30 amino acids in length and is contained or comprises any one of the peptides of SEQ ID NO:5-820 or 870-911.

53. The AAV capsid protein of claim 52, wherein the capsid protein is a VP1 capsid protein.

54. The AAV capsid protein of claim 52, wherein the capsid protein is a VP2 capsid protein.

55. The AAV capsid protein of claim 52, wherein the capsid protein is a VP3 capsid protein.

56. The AAV capsid protein of claim 52, wherein the heterologous targeting peptide is about 15-25 amino acids in length.

57. The AAV capsid protein of claim 52, wherein the heterologous targeting peptide is about 20 amino acids in length.

58. The AAV capsid protein of claim 52, wherein the heterologous targeting peptide is flanked by a linker peptide at the N-terminal and C-terminal ends of the heterologous targeting peptide.

59. The AAV capsid protein of claim 52, wherein the heterologous targeting peptide targets hepatocytes or liver tissue.

60. The AAV capsid protein of claim 52, wherein the heterologous targeting peptide targets neuronal cells or brain tissue.

61. The AAV capsid protein of claim 52, wherein the heterologous targeting peptide targets pancreatic cells or pancreas tissue.

62. The AAV capsid protein of claim 52, wherein the heterologous targeting peptide targets cardiac cells or heart tissue.

63. The AAV capsid protein of claim 52, wherein the heterologous targeting peptide targets lung tissue.

64. The AAV capsid protein of claim 52, wherein the heterologous targeting peptide targets intestinal tissue.

65. The AAV capsid protein of claim 52, wherein the heterologous targeting peptide targets spleen tissue.

66. The AAV capsid protein of claim 52, wherein the heterologous targeting peptide targets renal cells or kidney tissue.

67. The AAV capsid protein of claim 52, wherein the heterologous targeting peptide targets muscle cells or tissue.

68. A recombinant AAV (rAAV) comprising a capsid protein of any one of claims 52-67.

69. A recombinant AAV (rAAV) comprising a capsid protein having a targeting peptide in loop 1 and/or loop 2 wherein the targeting peptide is independently selected from SEQ ID Nos: 5-820 or 870-911.

70. The recombinant AAV of claim 69, wherein the recombinant AAV further comprises a heterologous polynucleotide for gene delivery.

71. The recombinant AAV of claim 70, wherein the heterologous polynucleotide is a therapeutic gene.

72. A composition comprising the recombinant rAAV of any one of claims 69-71.

73. The composition of claim 72 further comprising a pharmaceutically acceptable carrier.

73. A method for delivering a transgene to a subject comprising: administering a recombinant AAV (rAAV) to a subject, wherein the rAAV comprises:

(i) a capsid protein of any one of claims 52-67, and

(ii) at least one transgene, and wherein the rAAV infects cells of a target tissue of the subject.

74. The method of claim 73, wherein the at least one transgene encodes a protein.

75. The method of claim 74, wherein the protein is an immunoglobulin heavy chain or light chain or fragment thereof.

76. The method of claim 73, wherein the at least one transgene encodes a small interfering nucleic acid.

77. The method of claim 76, wherein the small interfering nucleic acid is a miRNA.

78. The method of claim 76, wherein the small interfering nucleic acid is a miRNA sponge or TuD RNA that inhibits the activity of at least one miRNA in the subject or animal.

79. The method of claim 77, wherein the miRNA is expressed in a cell of the target tissue.

80. The method of claim of claim 73, wherein the target tissue is skeletal muscle, heart, liver, pancreas, brain or lung.

81. The method of claim 73, wherein the transgene expresses a transcript that comprises at least one binding site for a miRNA,

wherein the miRNA inhibits activity of the transgene, in a tissue other than the target tissue, by hybridizing to the binding site.

82. The method of claim 73, wherein the at least one transgene encodes a gene product that mediates genome editing.

83. The method of claim 73, wherein the transgene comprises a tissue specific promoter or inducible promoter.

84. The method of claim 83, wherein the tissue specific promoter is a liver-specific thyroxin binding globulin (TBG) promoter, an insulin promoter, a glucagon promoter, a somatostatin promoter, a pancreatic polypeptide (PPY) promoter, a synapsin-1 (Syn) promoter, a creatine kinase (MCK) promoter, a mammalian desmin (DES) promoter, a a-myosin heavy chain (a-MHC) promoter, or a cardiac Troponin T (cTnT) promoter.

85. The method of claim 73, wherein the rAAV is administered intravenously, intravascularly, transdermally, intraocularly, intrathecally, orally, intramuscularly, subcutaneously, intranasally, or by inhalation.

86. The method of claim 73, wherein the subject is selected from a mouse, a rat, a rabbit, a dog, a cat, a sheep, a pig, and a non-human primate.

59. The method of claim 73, wherein the subject is a human.

87. An isolated nucleic acid encoding an AAV capsid protein containing an amino acid sequence selected from the group consisting of SEQ ID No: 5-820 and 870-911.

88. A delivery vehicle for delivery of a small molecule drug or biological agent having a desired tropism, wherein the delivery vehicle comprises a peptide or peptide fragment of at least 10-20 amino acids of any one of SEQ ID NOs:5-820 or 870-911.

89. The delivery vehicle of claim 88, wherein the delivery vehicle is selected from the group consisting of a liposome, a nanoparticle, a bacteria, a bacteriophage, a virus-like particle (VLP), a erythrocyte ghost, and an exosome.

90. The delivery vehicle of claim 88, wherein the biological agent comprises an siRNA, an antisense molecule, a protein or polypeptide, insulin, a vaccine, or an antibody.

91. The delivery vehicle of claim 88, wherein the small molecule drug comprises a chemotherapeutic agent, an anti-inflammatory, a steroid, and an antibiotic.

92. A biological agent having a desired tropism, the biological agent linked to a peptide or peptide fragment of at least 10-20 amino acids of any one of SEQ ID NOs:5-820 or 870-911.

93. The biological agent of claim 92, wherein the biological agent is a nucleic acid, a protein, a polypeptide, a peptide, an antibody, an antibody fragment, a non-immunoglobulin binding agent, or an enzyme.

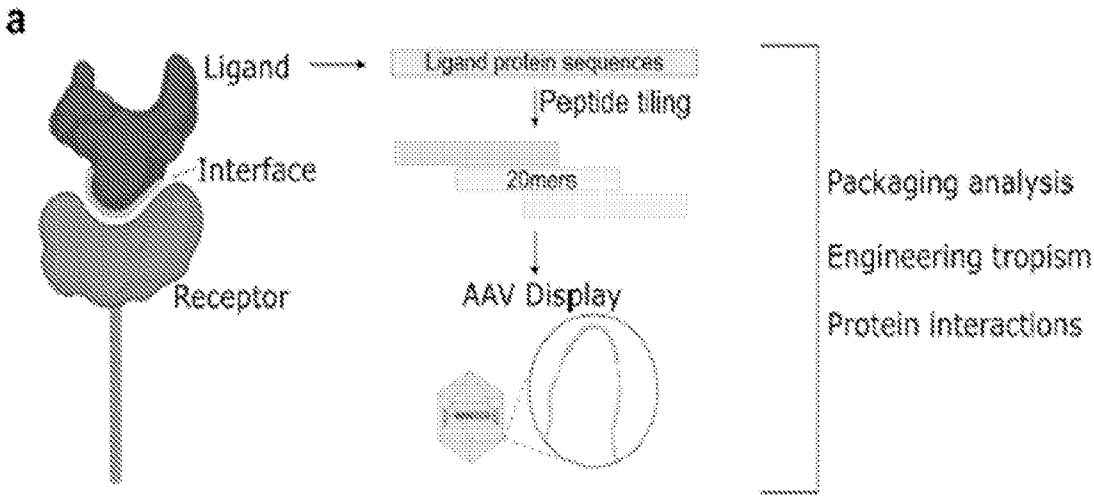
a
Ligand
Ligand protein sequences
Peptide tiling
Interface
20mers
Receptor
AAV Display
Packaging analysis
Engineering tropism
Protein interactions

FIG. 1A

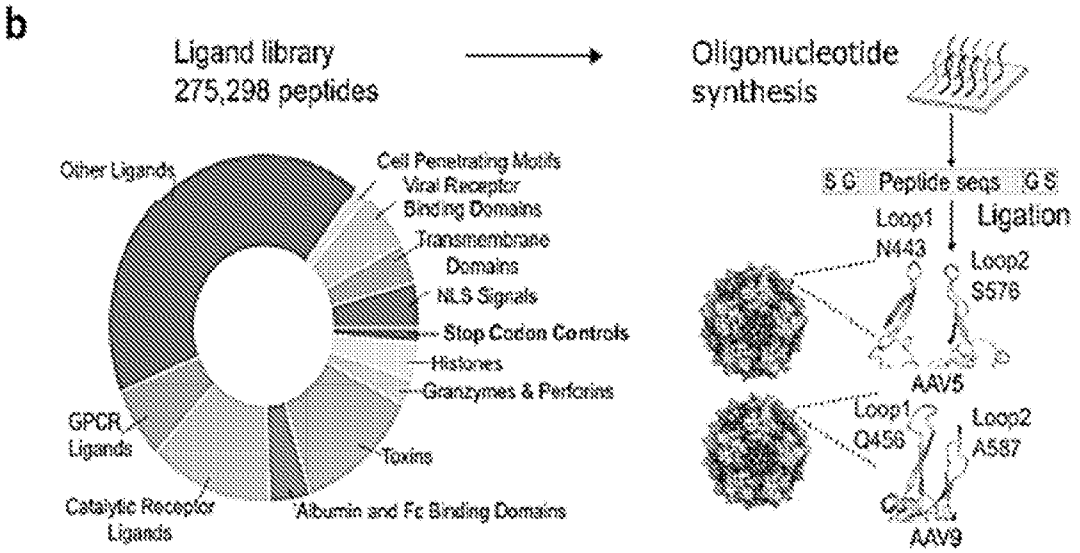
b
Ligand library
275,298 peptides
Oligonucleotide
synthesis
Other Ligands
Cell Penetrating Motifs
Viral Receptor
Binding Domains
Transmembrane
Domains
NLS Signals
Stop Codon Controls
Histones
Granzymes & Perforins
Toxins
Albumin and Fc Binding Domains
Catalytic Receptor
Ligands
GPCR
Ligands
S G Peptide seqs G S
Loop1
N443
Ligation
Loop2
S576
AAV5
Loop1
Q456
Loop2
A587
AAV9

FIG. 1B

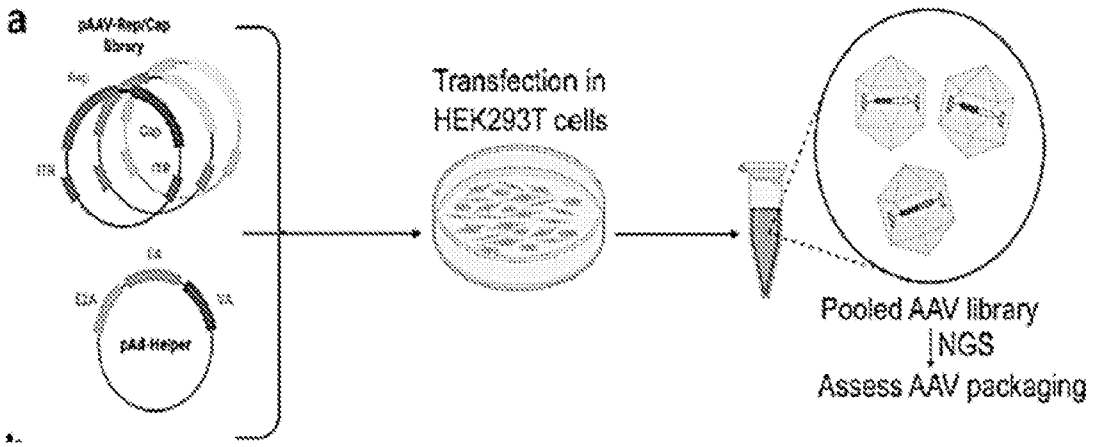

FIG. 2A

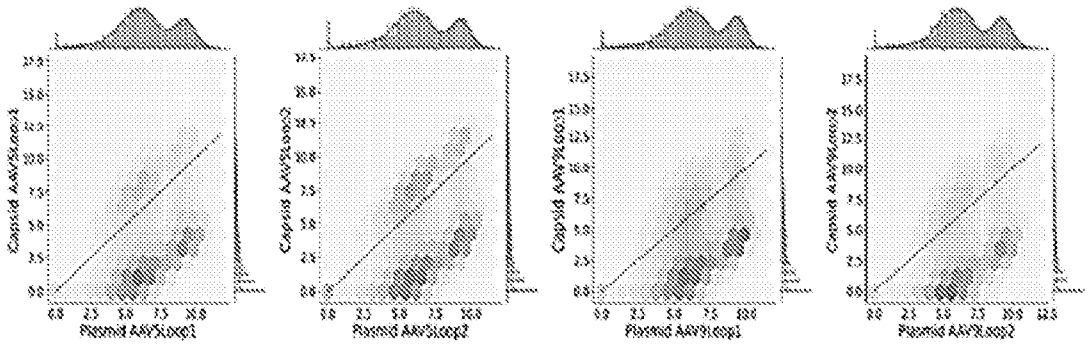

FIG. 2B

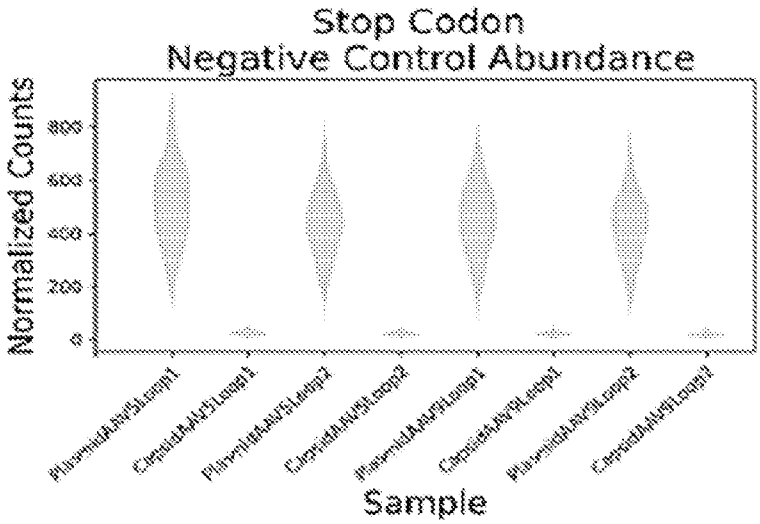

FIG. 2C

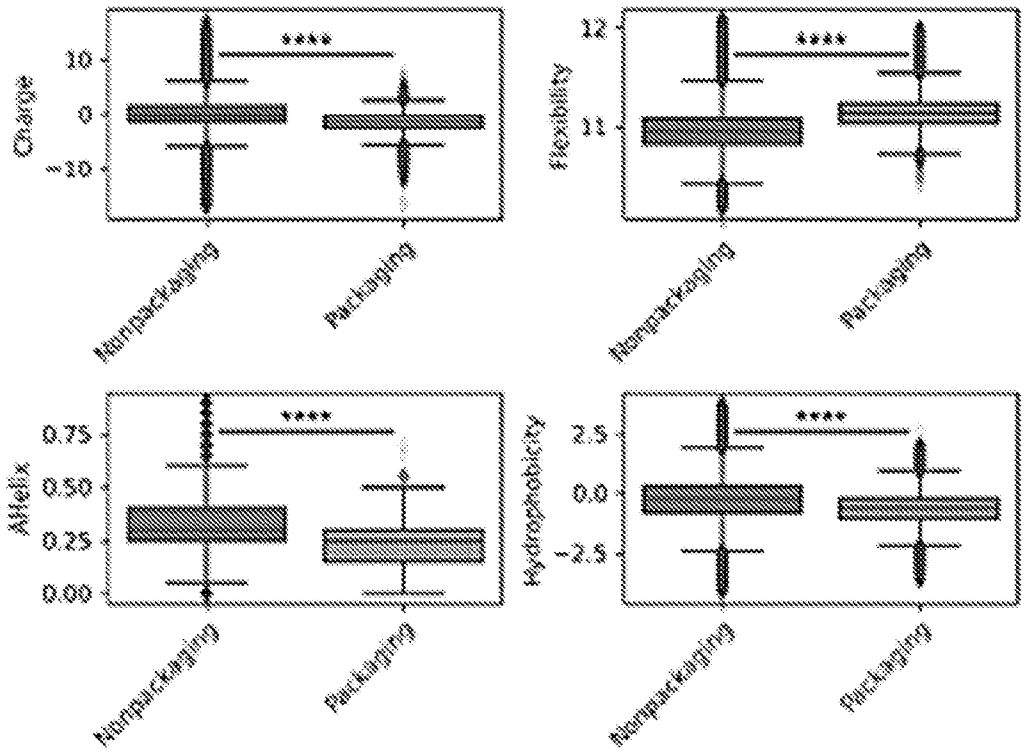
Charge
10
0
−10
****
Nonpackaging
Packaging
flexibility
12
11
****
Nonpackaging
Packaging
Alfa-helix
0.75
0.50
0.25
0.00
****
Nonpackaging
Packaging
Hydrophobicity
2.5
0.0
−2.5
****
Nonpackaging
Packaging

FIG. 2D

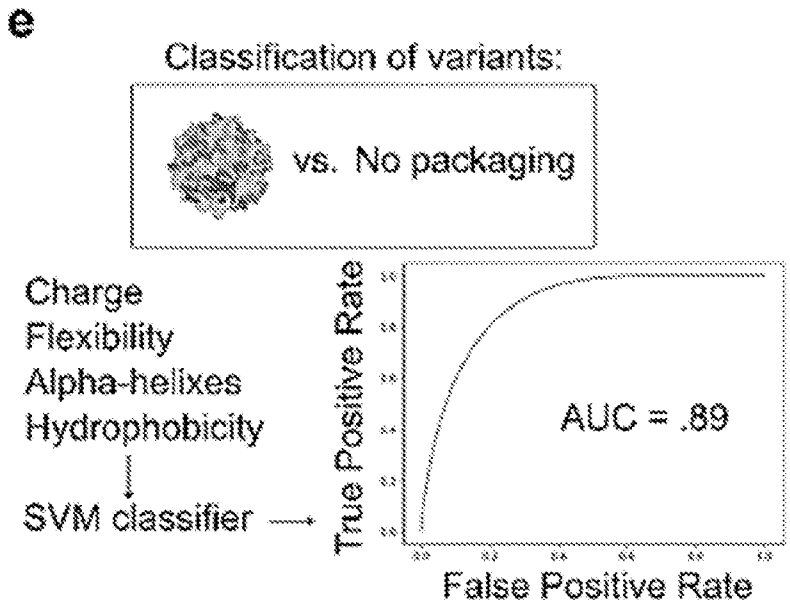
e
Classification of variants:
vs. No packaging
Charge
Flexibility
Alpha-helixes
Hydrophobicity
SVM classifier
True Positive Rate
False Positive Rate
AUC = .89

FIG. 2E

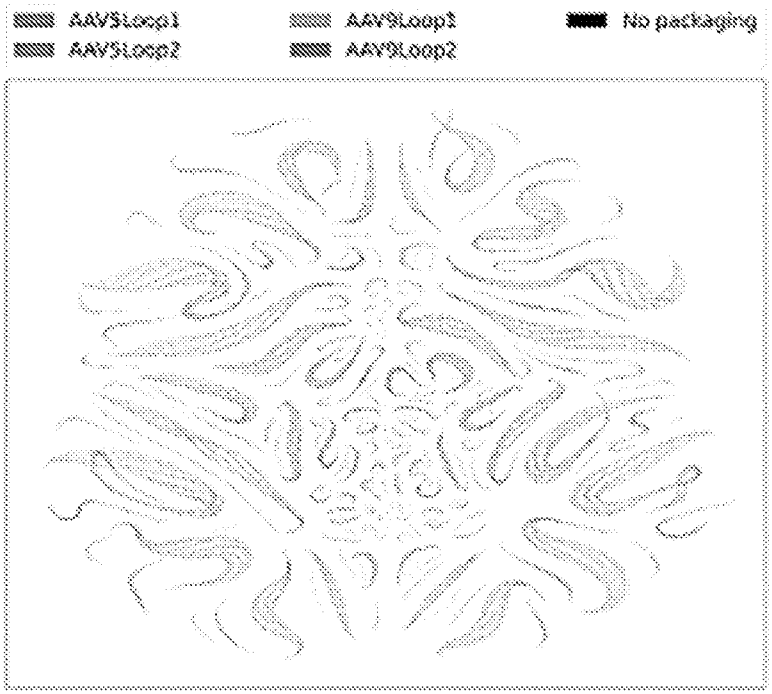
AAV9Loop1
AAV9Loop2
AAV9Loop1
AAV9Loop2
No packaging

FIG. 2F

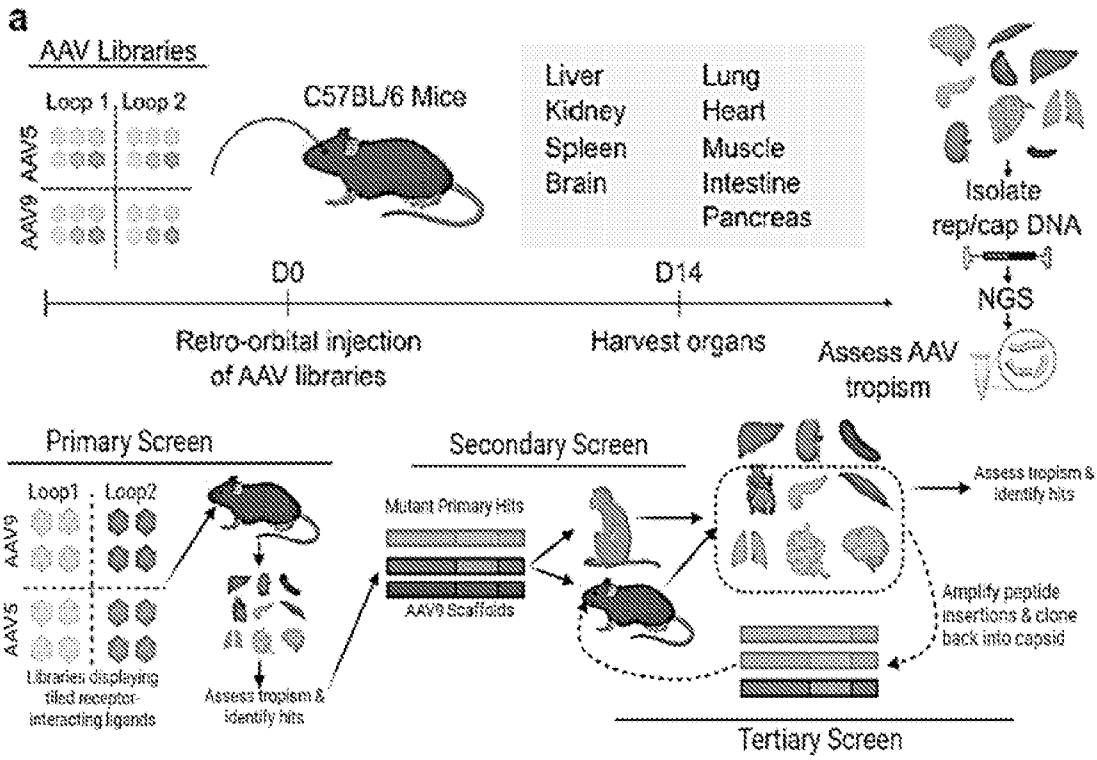
a
AAV Libraries
Loop 1
Loop 2
AAV5
AAV9
C57BL/6 Mice
Liver
Kidney
Spleen
Brain
Lung
Heart
Muscle
Intestine
Pancreas
Isolate rep/cap DNA
NGS
D0
D14
Retro-orbital injection of AAV libraries
Harvest organs
Assess AAV tropism
Primary Screen
Loop1
Loop2
AAV9
AAV5
Libraries displaying tiled receptor-interacting ligands
Assess tropism & identify hits
Secondary Screen
Mutant Primary Hits
AAV9 Scaffolds
Assess tropism & identify hits
Amplify peptide insertions & clone back into capsid
Tertiary Screen

FIG. 3A

b

Replicate Pearson Correlation

	AAV5-Loop1	AAV5-Loop2	AAV9-Loop1	AAV9-Loop2
Liver	0.78	0.71	0.82	0.83
Brain	0.52	0.72	0.59	0.57
Pancreas	0.60	0.61	0.48	0.61
Heart	0.61	0.61	0.55	0.59
Lung	0.50	0.78	0.46	0.56
Intestine	0.59	0.65	0.58	0.33
Spleen	0.82	0.83	0.89	0.87
Kidney	0.59	0.68	0.66	0.69
Muscle	0.63	0.60	0.45	0.66

FIG. 3B

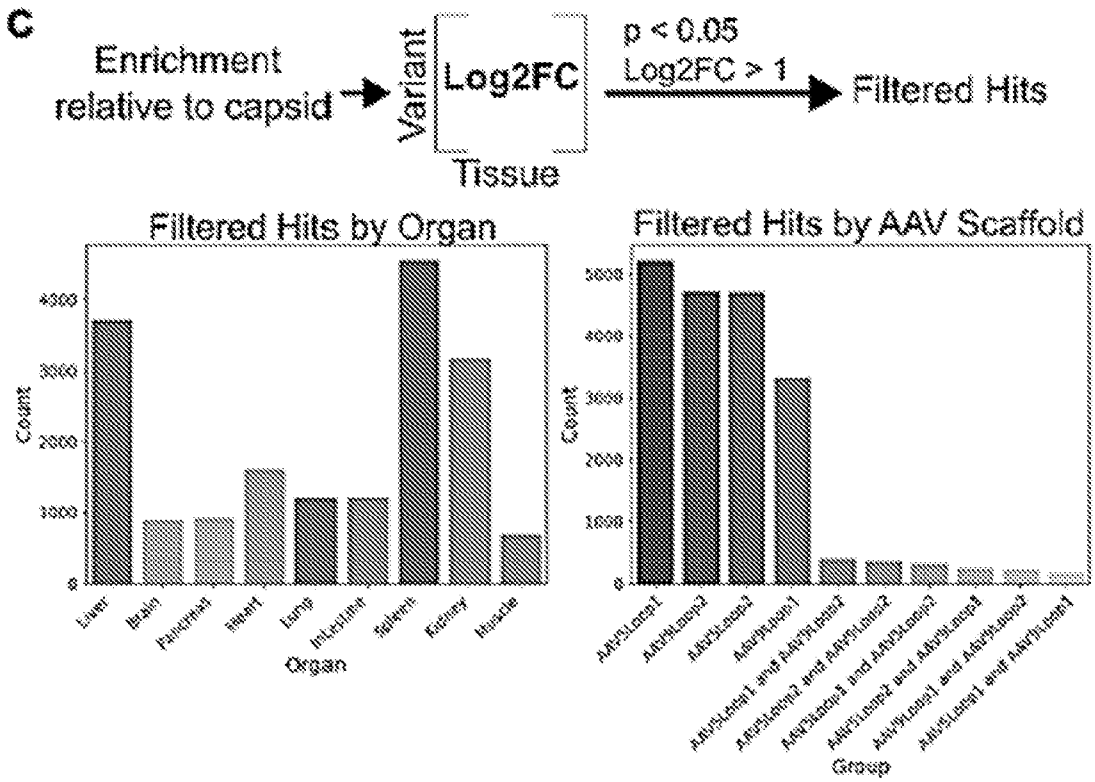

FIG. 3C

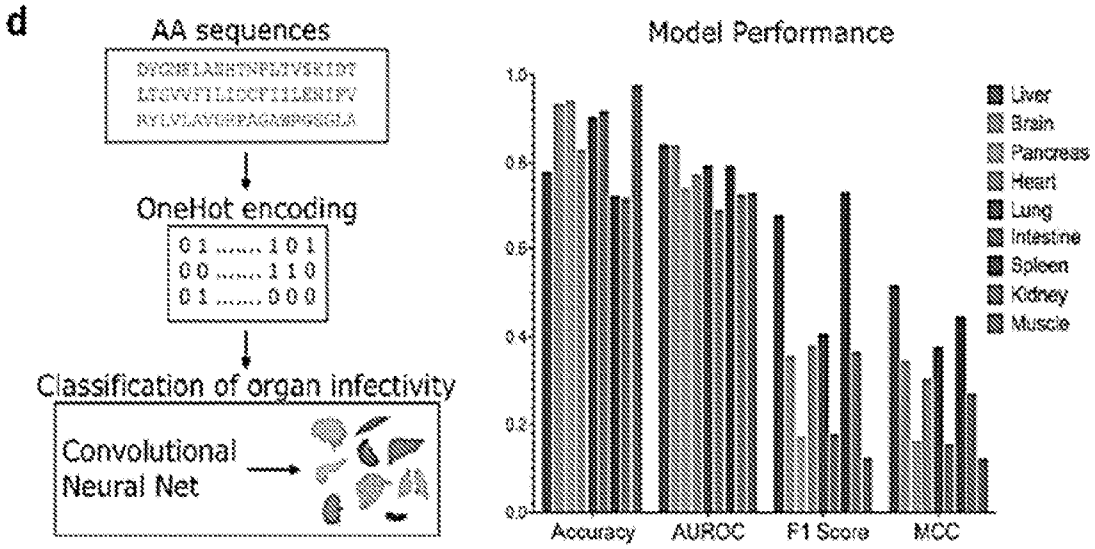
d
AA sequences
OneHot encoding
Classification of organ infectivity
Convolutional Neural Net
Model Performance
Liver
Brain
Pancreas
Heart
Lung
Intestine
Spleen
Kidney
Muscle
Accuracy
AUROC
F1 Score
MCC

FIG. 3D

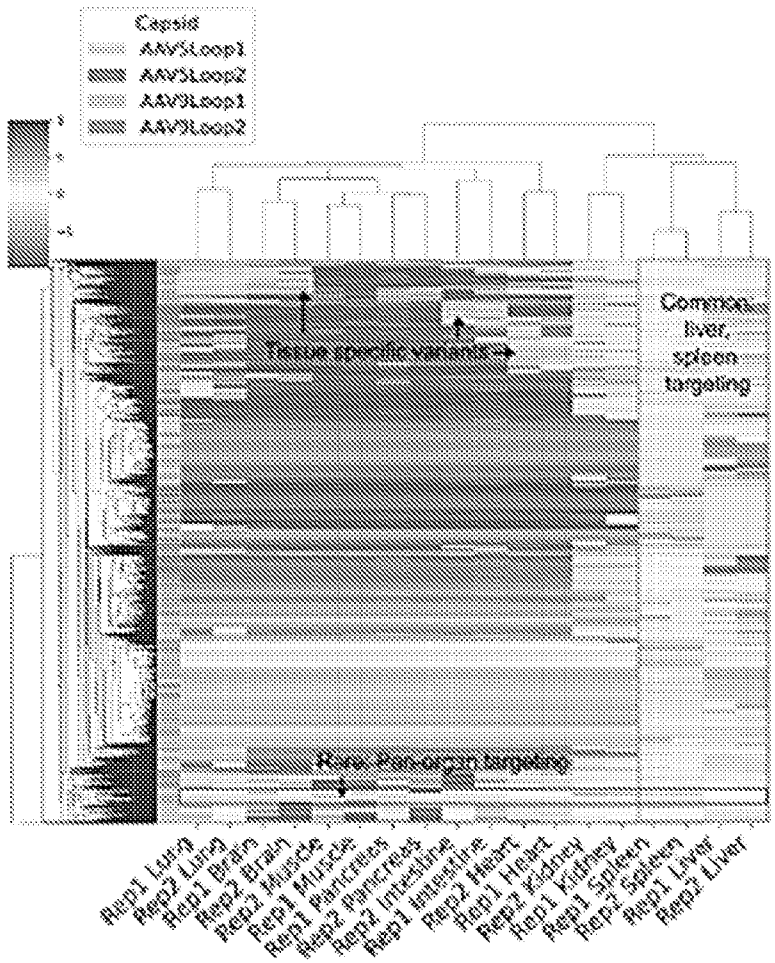
Capsid
AAV3Loop1
AAV3Loop2
AAV9Loop1
AAV9Loop2
Common liver, spleen targeting
Tissue specific variants
Rep1 Lung
Rep2 Lung
Rep1 Brain
Rep2 Brain
Rep1 Muscle
Rep2 Muscle
Rep1 Pancreas
Rep2 Pancreas
Rep1 Intestine
Rep2 Intestine
Rep1 Heart
Rep2 Heart
Rep1 Kidney
Rep2 Kidney
Rep1 Spleen
Rep2 Spleen
Rep1 Liver
Rep2 Liver

FIG. 4A

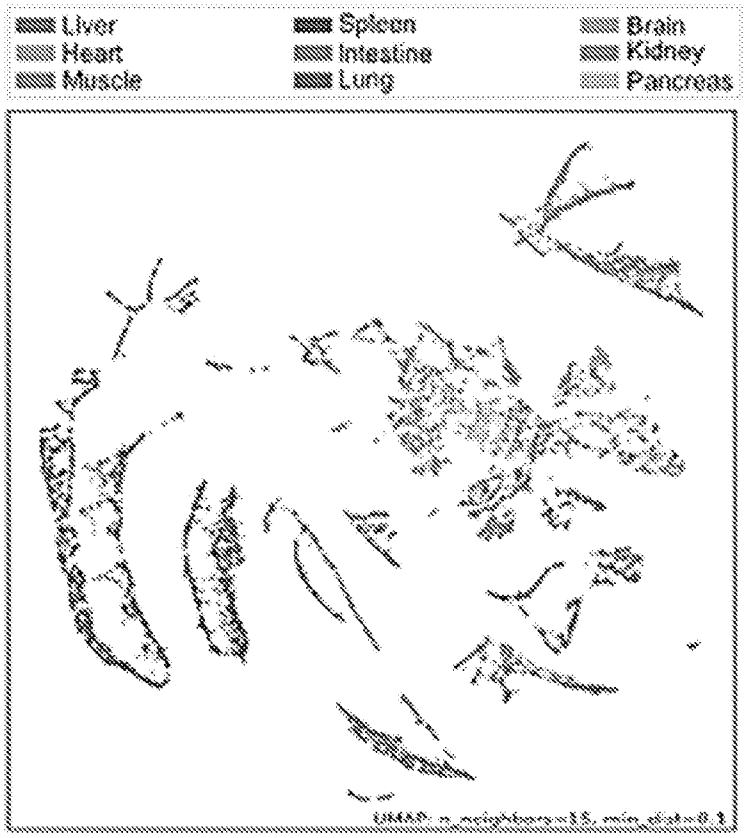
Liver
Heart
Muscle
Spleen
Intestine
Lung
Brain
Kidney
Pancreas

FIG. 4B

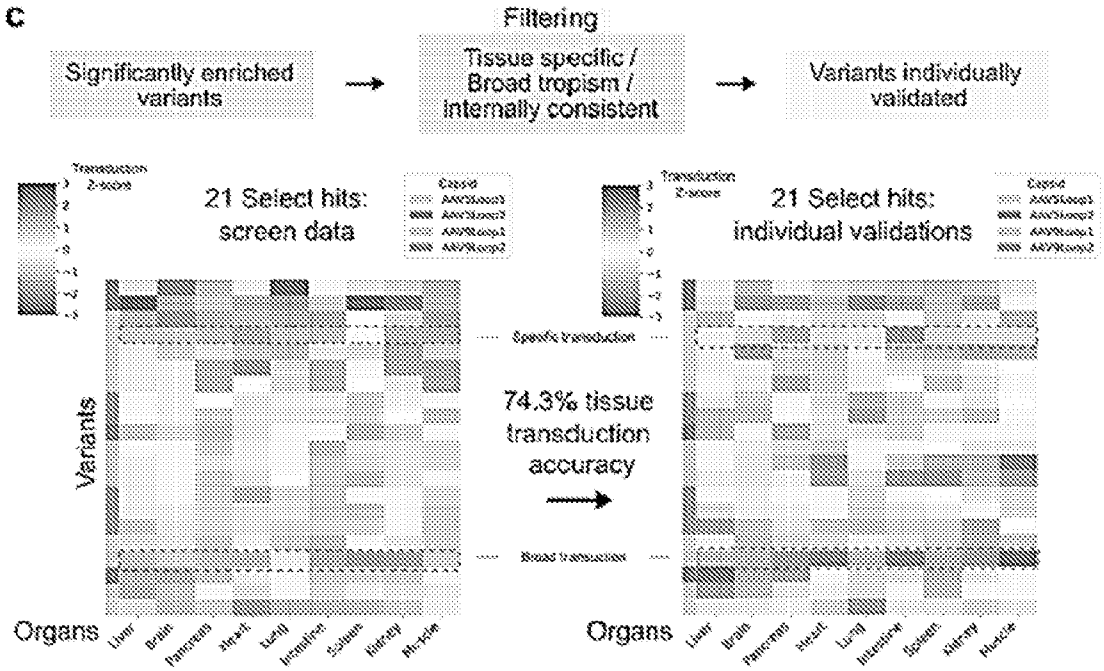
c
Filtering
Significantly enriched variants
Tissue specific / Broad tropism / internally consistent
Variants individually validated
21 Select hits: screen data
21 Select hits: individual validations
74.3% tissue transduction accuracy
Variants
Organs
Liver
Brain
Pancreas
Heart
Lung
Intestine
Spleen
Kidney
Muscle

FIG. 4C

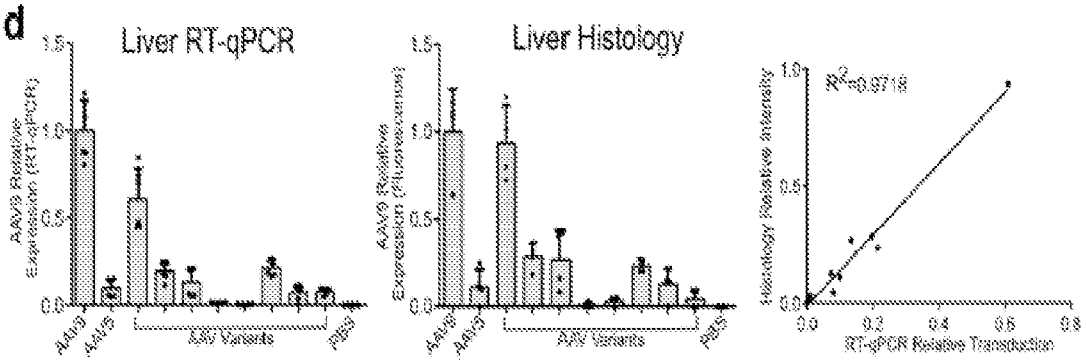

FIG. 4D

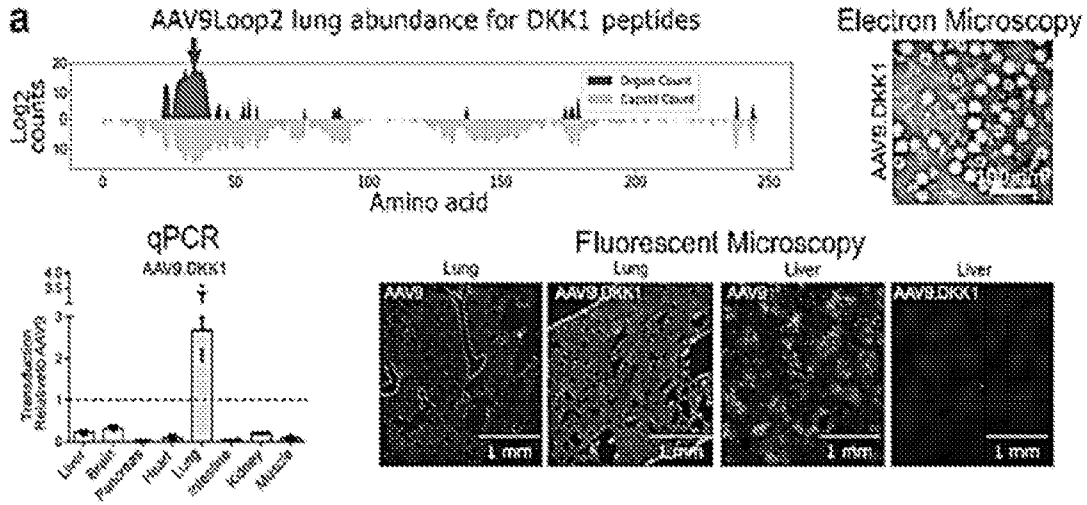

FIG. 5A

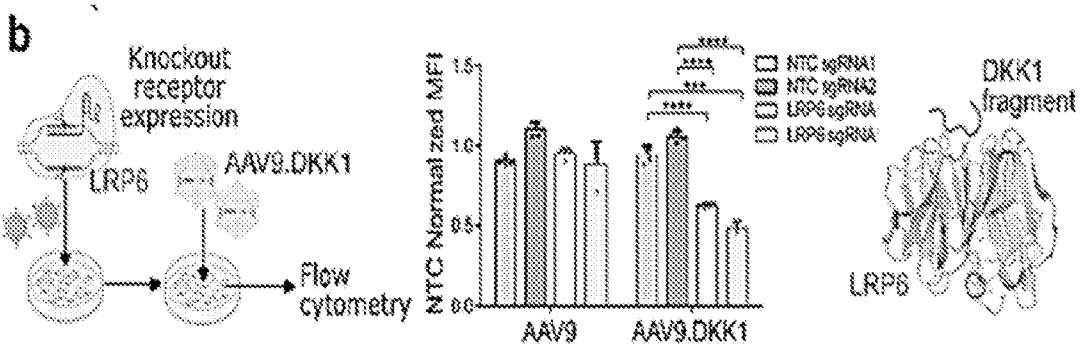

FIG. 5B

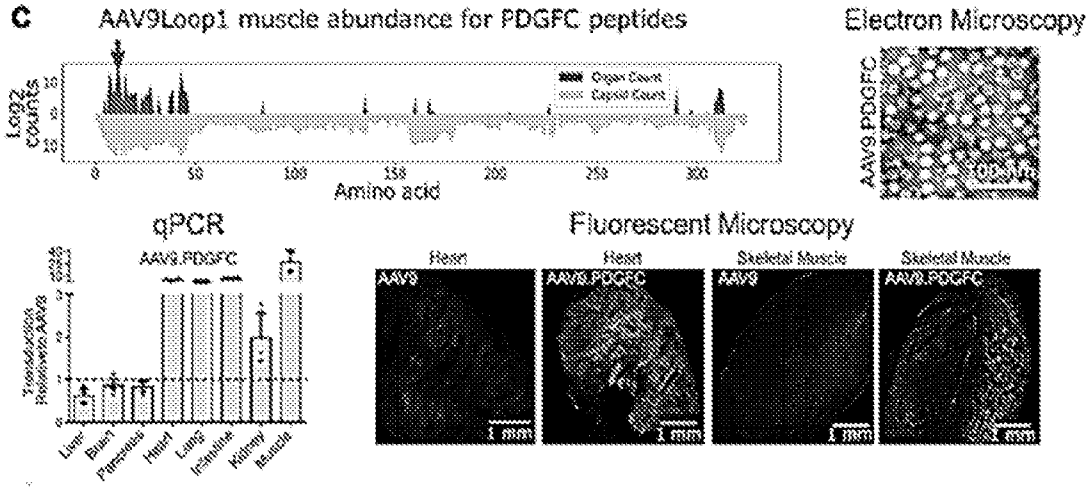

FIG. 5C

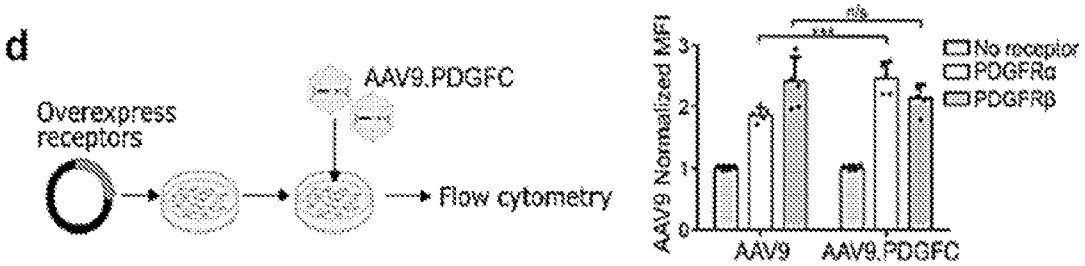

FIG. 5D

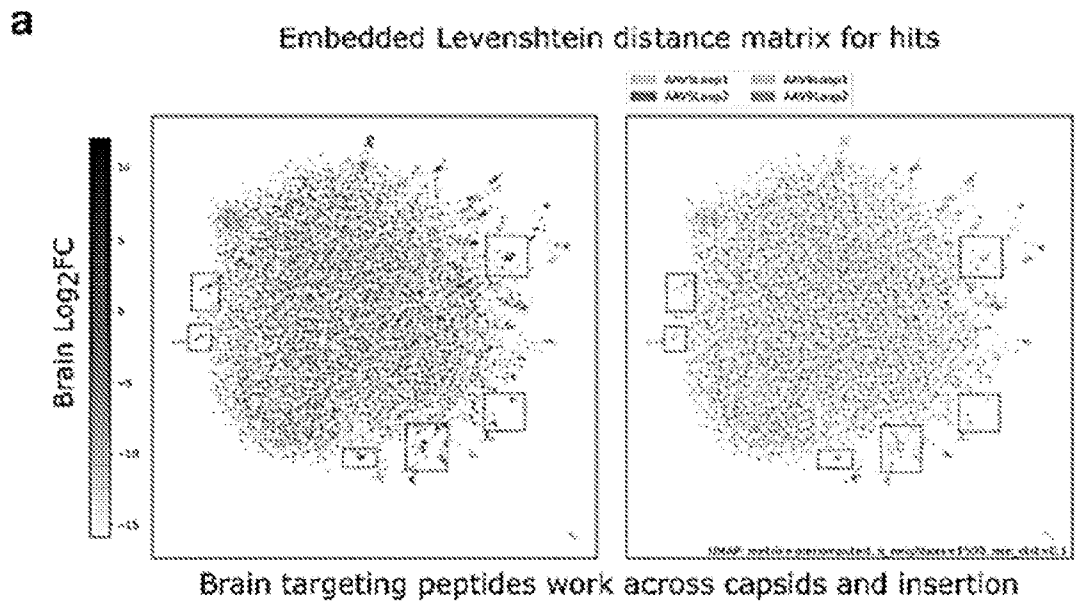

FIG. 6A

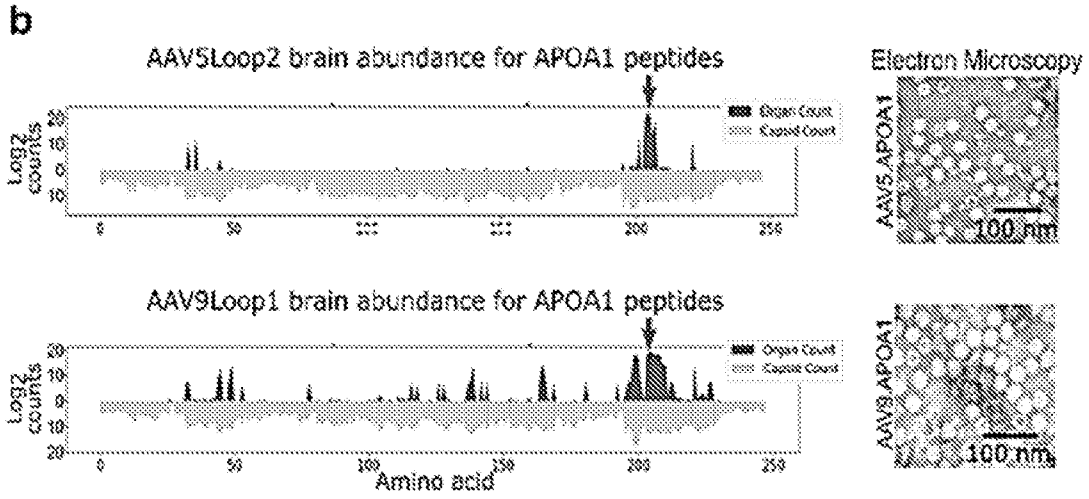
b
AAV5Loop2 brain abundance for APOA1 peptides
Electron Microscopy
AAV5.APOA1
100 nm
AAV9Loop1 brain abundance for APOA1 peptides
AAV9.APOA1
100 nm
Amino acid

FIG. 6B

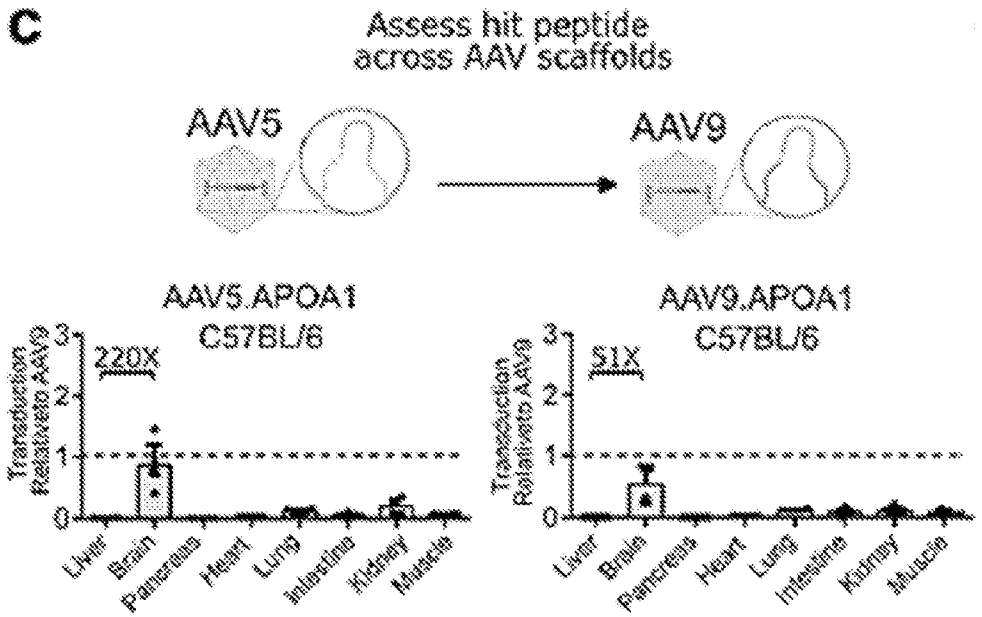
C
Assess hit peptide
across AAV scaffolds
AAV5
AAV9
AAV5.APOA1
C57BL/6
220X
AAV9.APOA1
C57BL/6
51X
Liver
Brain
Pancreas
Heart
Lung
Intestine
Kidney
Muscle

FIG. 6C

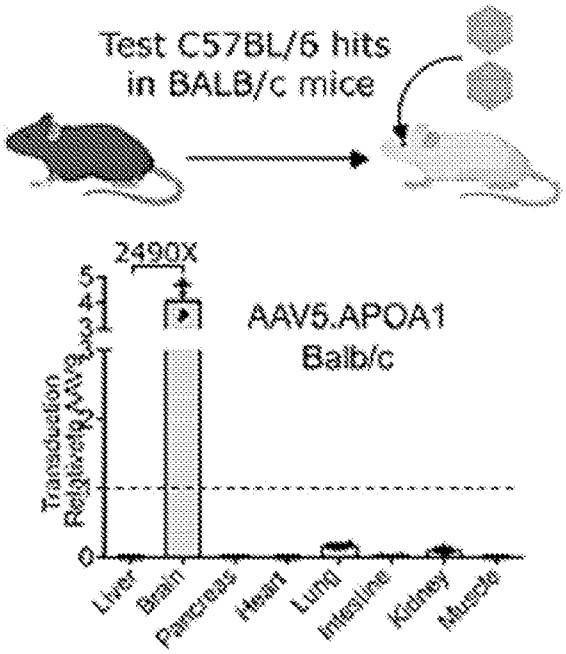
Test C57BL/6 hits
in BALB/c mice
2490X
AAV5.APOA1
Balb/c
Liver
Brain
Pancreas
Heart
Lung
Intestine
Kidney
Muscle

FIG. 6D

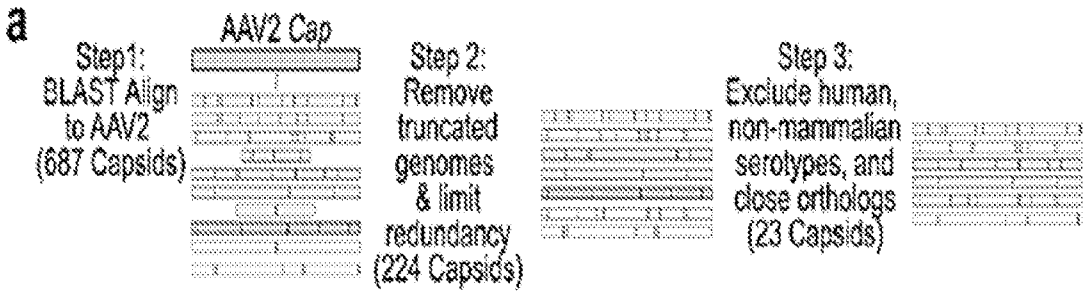
a
AAV2 Cap
Step1:
BLAST Align
to AAV2
(687 Capsids)
Step 2:
Remove
truncated
genomes
& limit
redundancy
(224 Capsids)
Step 3:
Exclude human,
non-mammalian
serotypes, and
close orthologs
(23 Capsids)

FIG. 7A

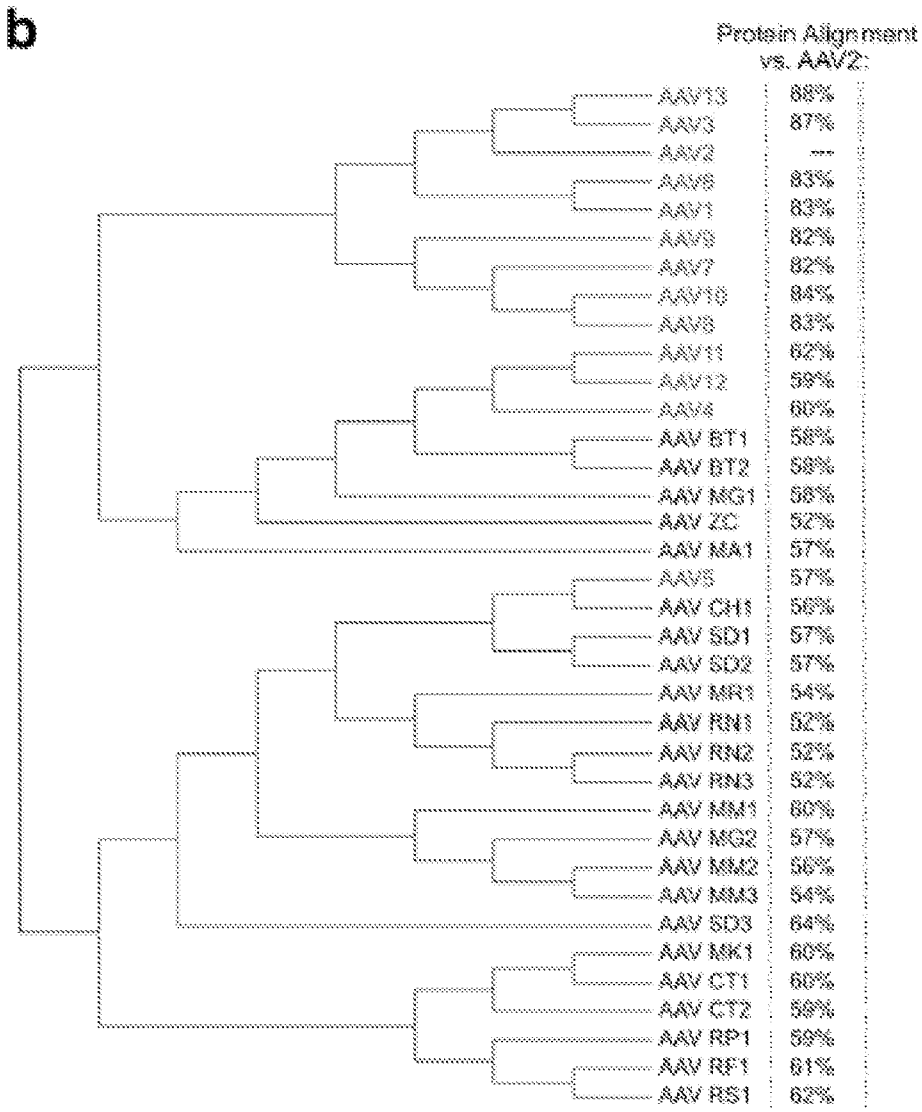
b
Protein Alignment vs. AAV2:
AAV13 88%
AAV3 87%
AAV2 ---
AAV9 83%
AAV1 83%
AAV5 82%
AAV7 82%
AAV10 84%
AAV8 83%
AAV11 62%
AAV12 59%
AAV4 60%
AAV BT1 58%
AAV BT2 58%
AAV NG1 58%
AAV ZC 52%
AAV MA1 57%
AAV6 57%
AAV CH1 58%
AAV SD1 57%
AAV SD2 57%
AAV MR1 54%
AAV RN1 52%
AAV RN2 52%
AAV RN3 52%
AAV MM1 60%
AAV MG2 57%
AAV MM2 56%
AAV MM3 54%
AAV SD3 64%
AAV MK1 60%
AAV CT1 60%
AAV CT2 59%
AAV RP1 59%
AAV RF1 61%
AAV RS1 62%

FIG. 7B

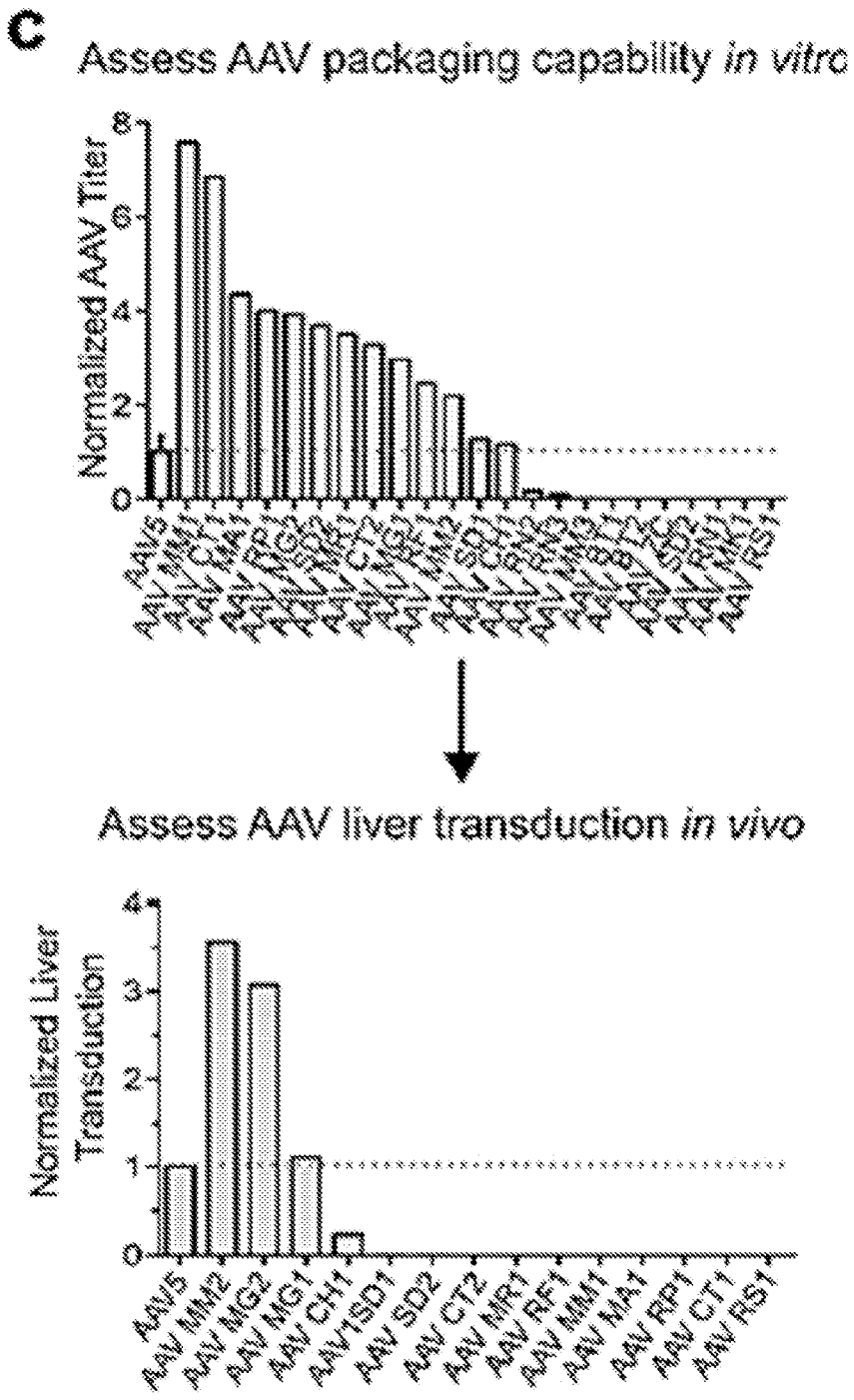
C
Assess AAV packaging capability in vitro
Normalized AAV Titer
0
2
4
6
8
Assess AAV liver transduction in vivo
Normalized Liver Transduction
0
1
2
3
4
AAV5
AAV MM2
AAV MG2
AAV MG1
AAV CH1
AAV1SD1
AAV SD2
AAV CT2
AAV MR1
AAV RF1
AAV MM1
AAV MA1
AAV RP1
AAV CT1
AAV RS1

FIG. 7C

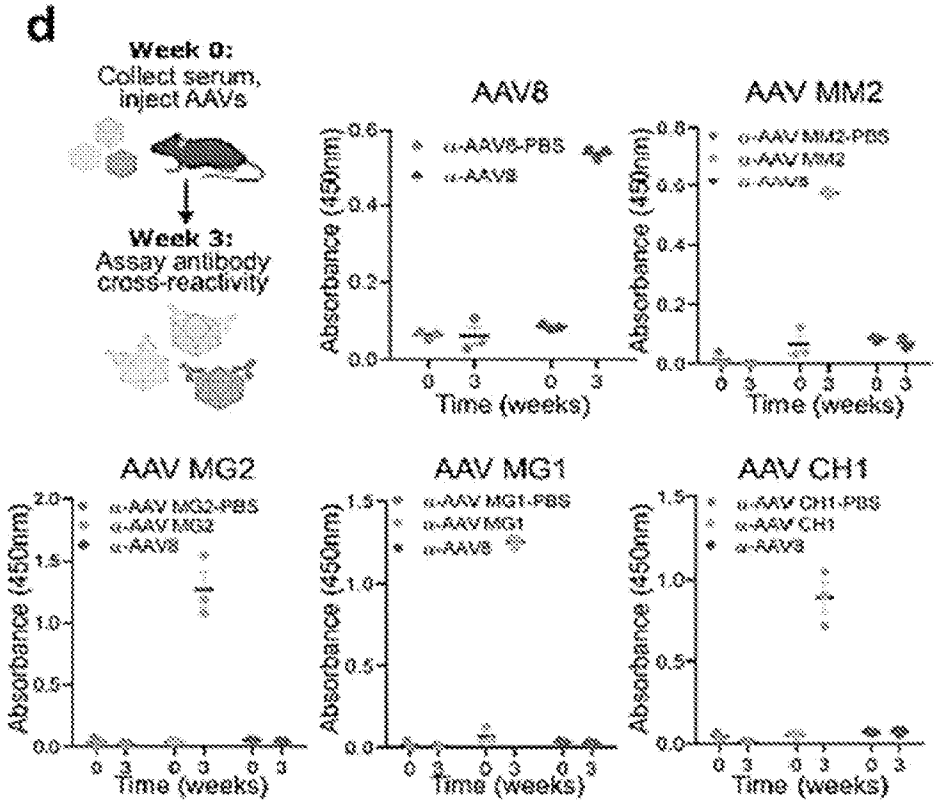
d
Week 0:
Collect serum,
inject AAVs
Week 3:
Assay antibody
cross-reactivity
AAV8
AAV MM2
AAV MG2
AAV MG1
AAV CH1
Absorbance (450nm)
Time (weeks)

FIG. 7D

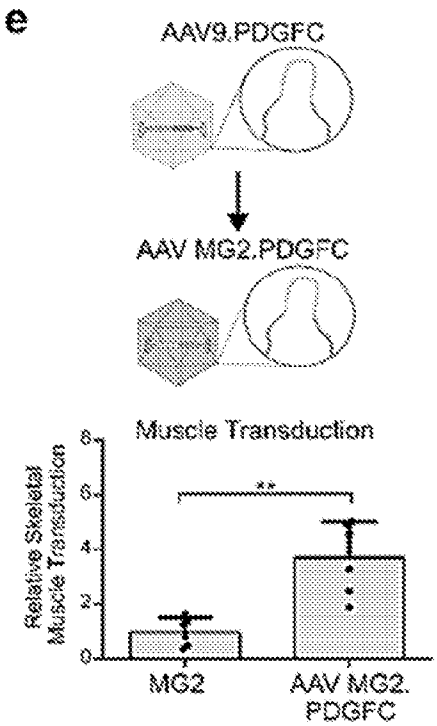
e
AAV9.PDGFC
AAV MG2.PDGFC
Muscle Transduction
Relative Skeletal Muscle Transduction
MG2
AAV MG2.PDGFC

FIG. 7E