

**AAV CAPSID MODIFICATIONS THAT ENABLE IMPROVED CNS-WIDE GENE
DELIVERY THROUGH INTERACTIONS WITH THE TRANSFERRIN RECEPTOR**

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application No. 63/621,839, filed January 17, 2024. The entire content of the above-identified application is hereby incorporated herein by reference in its entirety.

STATEMENT REGARDING FEDERALLY SPONSORED RESEARCH

[0002] This invention was made with government support under Grant Nos. NS111689 and MH120096 awarded by the National Institutes of Health (NIH). The government has certain rights in the invention.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

[0003] Reference is made to the electronic sequence listing ("BROD-5975WP_ST26.xml"; Size is 2,318,176 bytes, created on January 16, 2025), which is herein incorporated by reference in its entirety.

REFERENCE TO ELECTRONIC TABLES

[0004] Reference is also made to Tables 1-2, filed on January 17, 2024, in the United States Patent and Trademark Office and assigned Serial No. 63/621,839. The Tables are herein incorporated by reference in their entirety.

TECHNICAL FIELD

[0005] The subject matter disclosed herein relates generally to enhancing the transduction of an engineered AAV capsid into cells of the central nervous system (CNS) through interaction with the transferrin receptor expressed on the surface of the cells. In particular, in examples described herein, at least one of the capsid's proteins is modified to include an n-mer motif. Further examples relate to a vector system having one or more vectors for the expression of the modified AAV capsid proteins in a cell, the packaging of a recombinant nucleic acid cargo into the engineered AAV capsids, and a method of delivering the cargo to the CNS by either administering the AAV capsids directly to a host organism *in vivo* or transducing cells with the AAV capsids *ex vivo* and transplanting the cells back into the host organism.

BACKGROUND

[0006] The development of gene therapies for neurodevelopmental and neurological disorders has been constrained by the inability to deliver transgenes throughout the CNS. Several studies have reported engineered AAV9 capsids, most notably the AAV-PHP.B family, capable of highly effective gene transfer throughout the CNS after intravenous administration in adult mice. However, none of the engineered AAV capsids that cross the blood-brain barrier (BBB) and transduce the mouse brain with high efficiency exhibit enhanced CNS tropism in primates. There is an unmet need for AAVs with enhanced transduction efficiency.

SUMMARY

[0007] In an aspects, the techniques described herein relate to an engineered adeno associated virus (AAV) capsid polypeptide, including a transferrin receptor (TfR1) binding modification sequence defined by the formula X1-X2-X3-[7-mer]-X4-X5-X6-X7, wherein the 7-mer is one of the amino acid sequences YSRIGPN (SEQ ID NO: 1390), YSRNSDN (SEQ ID NO: 1391), LHRLGPN (SEQ ID NO: 1392), FRSTNGV (SEQ ID NO: 1393), FVSTNGV (SEQ ID NO: 1394), FZ1STNGZ2 (SEQ ID NO: 1395), FRSTNGZ3 (SEQ ID NO: 1396), or VESTNGR (SEQ ID NO: 1397), and wherein the 7-mer is inserted between amino acids 588 and 589 of an AAV9 capsid polypeptide (e.g., SEQ ID NO: 1), or in an analogous position of a capsid polypeptide of another AAV serotype, and wherein X1, X2, X3, X4, X5, X6, X7 indicate an amino acid substitution at one or more of these positions in the capsid polypeptide flanking the inserted 7-mer (i.e., residues 586, 587, 588, 589, 590, 591 and 592 of AAV9 or corresponding positions of any other AAV serotypes).

[0008] In an embodiment, the 7-mer includes one of YSRIGPN (SEQ ID NO: 1390), YSRNSDN (SEQ ID NO: 1391), or LHRLGPN (SEQ ID NO: 1392), and wherein X1 is A, G, E, L, N, Q, S, W, or M; X2 includes A, F, I, L, M, N, Q, P, T, V, or Y; X3 includes Q; X4 includes any amino acid; X5 includes D, F, G, I, L, M, Q, P, S, T, or V; X6 includes A, C, F, G, H, I, P, S, T, V, W, or Y; X7 includes D, E, Q, or T; or any combination thereof where there is at least one amino acid substitution relative to the wild type sequence. In an embodiment, the TfR1 binding modification sequence includes or consists of the amino acid sequence of one of SEQ ID Nos: 60-1210.

[0009] In an embodiment, the n-mer includes the amino acid sequence YSRIGPN (SEQ ID NO: 1390), and wherein X1 is A, G, E, L, N, Q, S, W, or M; X2 is A, F, I, L, M, N, Q, P, T, V, or Y; X3 is Q; X4 is A, E, F, H, I, L, M, N, P, Q, V, Y, D, or G; X5 is D, F, G, I, L, M, Q, P, S, T, or V; X6 is A, C, F, G, H, I, P, S, T, V, W, or Y; X7 is D, E, Q, or T; or any combination thereof where there is at least one amino acid substitution relative to the wild type sequence. In an embodiment, the TfR1 binding modification sequence includes or consists of the amino acid sequence of one of SEQ ID Nos: 60-1136.

[0010] In an embodiment, the n-mer is YSRNSDN (SEQ ID NO: 1391), and X1 is S or W; X2 is V, I, or F; X3 is Q; X4 is any amino acid; X5 is Q or T; X6 is A; X7 is Q; or any combination thereof. In an embodiment, the TfR1 binding modification includes or consists of the amino acid sequence of one of SEQ ID Nos: 1152-1157.

[0011] In an embodiment, the n-mer is LHRLGPN (SEQ ID NO: 1392), and X1 is S, A, L, or M; X2 is A or P; X3 is Q; X4 is A, E, F, H, I, L, M, N, P, Q, V, or Y; X5 is Q; X6 is A, P, S, or T; X7 is D, E, Q, or T; or any combination thereof. In an embodiment, the TfR1 binding modification sequence includes or consists of the amino acid sequence of one of SEQ ID Nos: 1158-1210.

[0012] In an embodiment, the 7-mer is FRSTNGV (SEQ ID NO: 1393), FVSTNGV (SEQ ID NO: 1394), FZ1STNGZ2 (SEQ ID NO: 1395), FRSTNGZ3 (SEQ ID NO: 1396), or VESTNGR (SEQ ID NO: 1397) and wherein X1 is S; X2 is A, S, M, or D; X3 is F, H, I, L, M, N, Q, R, Y, D, or E; X4 is A, S, or M; X5 is Q or P; X6 is A, F, H, Q, or S; X7 is A, D, E, F, Q, S, or T; or any combination thereof. In an embodiment, the TfR1 binding modification sequence includes or consists of the amino acid sequence of one of SEQ ID Nos: 1211-1389.

[0013] In an embodiment, the n-mer is FRSTNGV (SEQ ID NO: 1393), and X1 is S; X2 is A or S; X3 is D; X4 is A or S; X5 is Q or P; X6 is A, F, H, Q, or S; X7 is D, E, Q, or T; or any combination thereof. In an embodiment, the TfR1 binding modification sequence includes or consists of the amino acid sequence of one of SEQ ID Nos: 1211-1266.

[0014] In an embodiment, the n-mer is FVSTNGV (SEQ ID NO: 1394), and X1 is S; X2 is A, S, or M; X3 is Q, E, or D; X4 is A or M; X5 is Q or P; X6 is A; X7 is E; or any combination thereof. In an embodiment, the TfR1 binding modification sequence includes or consists of the amino acid sequence of one of SEQ ID Nos: 1267-1298.

[0015] In an embodiment, the n-mer is FZ1STNGZ2 (SEQ ID NO: 1395) or FRSTNGZ3 (SEQ ID NO: 1396), and X1 is S; X2 is A or S; X3 is Q or D; X4 is A; X5 is Q; X6 is A; X7 is E; or any

combination thereof. In an embodiment, the TfR1 binding modification sequence includes or consists of the amino acid sequence of one of SEQ ID Nos: 1299-1319.

[0016] In an embodiment, the 7-mer is VESTNGR (SEQ ID NO: 1397), and X1 is S; X2 is S or D; X3 is F, H, I, L, M, N, Q, R, or Y; X4 is A; X5 is Q or P; X6 is A; X7 is A, D, E, F, Q, S, or T; or any combination thereof. In an embodiment, the TfR1 binding modification sequence comprises or consists of the amino acid sequence of one of SEQ ID Nos: 1320-1388.

[0017] In an embodiment, Z1 is selected from the group consisting of A, D, H, N, Q, and S; Z2 is K or R; and Z3 is selected from the group consisting of L, M, and R.

[0018] In an embodiment, the TfR1 binding modification comprises or consists of the amino acid sequence of EAQYSRIGPNNQAQ (SEQ ID NO: 170); SADFRSTNGVAQAE (SEQ ID NO: 1215); or GAQYSRIGPNPQPE (SEQ ID NO: 243).

[0019] In an embodiment, the viral capsid protein is VP1, VP2, VP3, or a combination thereof. In an embodiment, the other AAV serotype is AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAVrh8, AAV rh.74, or AAV rh. 10.

[0020] In some aspects, the techniques described herein relate to an engineered AAV particle, including any engineered AAV capsid polypeptide as described herein, and further include a recombinant AAV genome encoding a transgene.

[0021] In an embodiment, the transgene encodes a therapeutic polypeptide, an antibody or fragment thereof, a siRNA, a CRISPR-Cas system, Transcription Activator-like Effector (TALE)- or Zinc Finger Protein (ZFP)-based transcriptional activator; repressor; or epigenomic silencer, a RNA encoding a partial gene fragment designed for transplacing into an endogenous RNA, one or more transfer RNAs, or a component thereof, or an OMEGA system or any component thereof. In an embodiment, the transgene is operably linked to a regulatory sequence that promotes expression in the CNS.

[0022] In one aspect, embodiments are directed to a pharmaceutical composition, including any recombinant engineered AAV particle described herein and an acceptable carrier.

[0023] In another aspect, embodiments described herein provide a method of delivering a polypeptide or polynucleotide to the CNS of a subject comprising administering the pharmaceutical composition of any of those described herein to the subject. In an embodiment, the pharmaceutical composition is delivered at a dosage between 0.1×10^{12} vg/kg to 1×10^{14} vg/kg. In an embodiment, the dosage is between 0.1×10^{12} vg/kg to 100×10^{12} vg/kg. In an embodiment,

the dosage is between 1×10^{12} vg/kg to 10×10^{12} vg/kg. In an embodiment, the dosage is 5×10^{12} vg/kg. In an embodiment the dosage is 2×10^{12} vg/kg. In an embodiment, the pharmaceutical composition is administered systemically or directly to the CNS.

[0024] In one embodiment, a method of manufacturing a recombinant engineered AAV comprises culturing mammalian cells comprising (1) a polynucleotide encoding the engineered AAV capsid polypeptide of any of those as described herein, (2) a polynucleotide encoding a recombinant AAV genome including a transgene operably linked to a regulatory sequence and flanked by AAV ITR sequences, and optionally (3) a polynucleotide encoding adenoviral helper genes, under conditions sufficient for the production of recombinant engineered AAV particles; and recovering the recombinant engineered AAV particles from said culture.

[0025] In another aspect, embodiments described herein relate to a host cell for the production of recombinant engineered AAV particles, which comprises a polynucleotide encoding the engineered AAV capsid polypeptide of any of those described herein, a polynucleotide encoding a recombinant AAV genome including a transgene operably linked to a regulatory sequence and flanked by AAV ITR sequences, and optionally a polynucleotide encoding adenoviral helper genes.

[0026] In another aspect, embodiments described herein relate to an AAV library including a population of variant engineered recombinant AAV particles wherein each member of the population of variant engineered recombinant AAV particles includes a variant recombinant AAV capsid polypeptide having a modification defined by the formula X1-X2-X3-[7-mer]-X4-X5-X6-X7, and wherein the 7-mer is inserted between amino acids 588 and 589 of an AAV9 capsid polypeptide, or in an analogous position of a capsid polypeptide of another AAV serotype, and wherein X1, X2, X3, X4, X5, X6, X7 indicate one or more modifications at amino acid positions in the capsid polypeptide flanking the inserted 7-mer, wherein the modification has been selected for binding of TfR1 and/or increased tropism for the CNS relevant to a reference AAV particle without the modification. In an embodiment, the another AAV serotype is AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAVrh8, AAV rh.74, or AAV rh. 10.

[0027] In another aspect, embodiments described herein relate to a method of screening an AAV library for a recombinant AAV particle that binds TfR1 and/or has increased tropism for the CNS. Said method comprises assaying any AAV library as described herein for increased binding

of Tfr1 and/or tropism for the CNS relative to an AAV vector with a reference capsid and selecting those recombinant AAV vectors that have increased binding of Tfr1 and/or tropism for the CNS.

[0028] In another aspect, embodiments described herein relate to a method for training a machine learning algorithm, including (a) receiving, by at least one computing device, a plurality of AAV capsid polypeptide sequences, including a modification for binding to a Tfr1; (b) training, by at least one computing device, with the plurality of AAV capsid polypeptide sequences including a modification for binding to a Tfr1, a transferrin receptor (Tfr1) targeting machine learning model; and (c) deploying, by at least one computing device, the Tfr1 targeting machine learning algorithm.

[0029] In an embodiment, the Tfr1 targeting machine learning model is trained to identify one or more sequences from the plurality of sequences with increased binding of Tfr1.

[0030] In an embodiment, the Tfr1 targeting machine learning model is trained to identify one or more sequences from the plurality of sequences that decrease the transduction of off-target tissues.

[0031] In an embodiment, the Tfr1 targeting machine learning model is trained to identify one or more sequences from the plurality of sequences that bind to an extracellular domain of TFRC.

[0032] In an embodiment, the Tfr1 targeting machine learning model is trained to identify one or more sequences from the plurality of sequences that bind to an apical domain.

[0033] In an embodiment, the modification is defined by the formula X1-X2-X3-[7-mer]-X4-X5-X6-X7, wherein X1, X2, X3, X4, X5, X6, and X7 indicate modifications at one or more amino acid positions in the capsid polypeptide flanking the inserted 7-mer (that is, at each of X1, X2, X3, X4, X5, X6, and X7, the amino acid may be the wild-type amino acid at that position or an amino acid substitution).

[0034] In an embodiment, the training includes unsupervised, supervised, semi-supervised, reinforcement, transfer, incremental, and curriculum. The transferrin receptor (Tfr1) targeting machine learning model comprises of linear classifiers, logistic classifiers, random forest, artificial neural networks, matrix factorization, support vector machines, K-means clustering, or K-nearest neighbor.

[0035] In an embodiment, the transferrin receptor (Tfr1)- targeting machine learning model comprises Boltzmann machines, Bayesian networks, autoregressive models, variational

autoencoders (VAEs), diffusion models, energy-based models, flow-based models, generative adversarial networks (GANs), mixture models, hidden Markov models, or large language models (LLMs).

[0036] In an embodiment, the transferrin receptor (TfR1) targeting machine learning model comprises convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory models (LSTMs), gated recurrent units (GRUs), capsule networks, attention mechanisms, or transformer networks.

[0037] In an embodiment, the transferrin receptor (TfR1) targeting machine learning model is pre-trained and further trained to predict transferrin receptor (TfR1) targeting by a plurality of sequences.

[0038] These and other aspects, objects, features, and advantages of the example embodiments will become apparent to those with ordinary skill in the art upon considering the following detailed description of example embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0039] An understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention may be utilized, and the accompanying drawings of which:

[0040] **FIG. 1A-1F – AAV9 can be programmed to bind human TfR1.** (1A) An AAV9-based NNK capsid library of variants with random 7-mer insertions between VP1 residues 588–589 was screened for selective binding to human TfR1 in pull-down or cell binding assays. Individually produced human TfR1-binding variants carrying a single-stranded AAVCAG-NLS-mScarlet-P2A-Luciferase-WPRE-SV40pA genome exhibited enhanced species-specific (1B) association with and (1C) transduction (Luciferase activity) of CHO cells stably expressing *TFRC*. Values reported are normalized to AAV9 in each cell line. Two-way ANOVA test to determine significant differences in transduction among five AAV variants on five CHO cell lines with AAV9 as the main comparison group for each cell line and using a Bonferroni multiple comparison correction: **** and *** indicate $p \leq 0.0001$ and ≤ 0.001 , respectively; $n = 3$ replicates, error bars indicate $\pm$ SEM. The TfR1-binding variants exhibited enhanced (1D) association with and (1E) transduction (Luciferase activity) of human brain microvascular endothelial cells (hBMVEC) and hCMEC/D3 cells. Values are normalized to AAV9 in each cell line. One-way ANOVA test to

determine significant differences among five AAV variants with AAV9 as the comparison group for each cell line and using a Dunnett's multiple comparison correction: ****, ***, and * indicate $p \leq 0.0001$, ≤ 0.001 , and ≤ 0.05 , respectively; $n = 4$ replicates, error bars indicate $\pm$ SEM. **(1F)** The binding kinetics between each capsid and full-length human TfR1 were assessed by BLI. AAVX probes were loaded with capsid, and human TfR1 was used as an analyte. Sensorgram curve fits (black dash-dot lines) were generated by applying global 2:1 exponential association and decay models.

[0041] FIG. 2A-2D – Human TfR1-targeted capsids transduce human brain endothelial cells via interactions with the apical domain of human TfR1. (2A) The plots show transduction (Luciferase, normalized RLU) of hCME/D3 cells incubated with 3×10^8 vg/mL of the indicated AAV and the specified OKT9 or AF2474 antibody concentrations. Two-way ANOVA test to determine significant differences in transduction with AAV9 or BI-hTFR1 and one of two antibodies at different concentrations with the no antibody control for each condition as the main comparison group and using a Bonferroni multiple comparison correction: **** indicates $p \leq 0.0001$; $n = 3$ replicates, error bars indicate $\pm$ SEM. **(2B)** The effect of Tf on BI-hTFR1 binding to full-length human TfR1 was assessed by BLI. AAVX probes were loaded with AAV9 or BI-hTFR1. Human TfR1 that had or had not been pre-incubated with 300 nM holo-Tf was used as an analyte. 2:1 binding model curve fits (black dash-dot lines) are shown. **(2C)** Biotinylated holo-Tf was immobilized on streptavidin-coated BLI probes (SA probe), introduced first into human TfR1, and then into a buffer with or without the OKT9 antibody, and finally into BI-hTFR1 virus particles. Segments shaded in gray highlight the indicated association step. **(2D)** BI-hTFR1 or AAV2 was incubated with hCMEC/D3 cells at 50,000 vg/cell for one hour at 4°C, with or without OKT9 (1 μ g/mL) or Tf (1 μ g/mL), and immunostained for AAV and TfR1. Scale bars = 15 μ m. AAV9 binding to hCMEC/D3 cells was rarely detected; therefore, AAV2 was used as a control (Fig. 9).

[0042] FIG. 3A-3C – BI-hTFR1 is efficiently endocytosed and actively transported across human brain endothelial cells. (3A) Schematic shows the pooled transwell BBB model experimental design. **(3B)** The vector genomes in the bottom chamber were quantified by qPCR. Two-way ANOVA with Bonferroni multiple comparison correction: **** and *** indicate $p \leq 0.0001$ and ≤ 0.001 , respectively; $n = 3$ transwell replicates, error bars indicate $\pm$ SEM. **(3C)** BI-

hTFR1 or AAV2 were incubated with hCMEC/D3 cells at 25,000 vg/cell for one hour at 37°C and stained for endosomal markers Rab5 and Rab7 and the AAVs. Scale bar = 15 μ m.

[0043] FIG. 4A-4D – BI-hTFR1 efficiently delivers genes to the CNS of *TFRC* KI mice.

(4A) In *TFRC* KI mice, mouse *Tfrc* exons 4–19 encoding the extracellular region of TfR1 have been replaced by those of human *TFRC*. **(4B)** BI-hTFR1 or AAV9 encoding CAG-NLS-mScarlet-P2A-Luciferase-WPRE-SV40pA were intravenously injected into adult female C57BL/6J or *TFRC* KI mice at 5×10^{11} vg/mouse. AAV9 in C57BL/6J and BI-hTFR1 in *TFRC* KI mice had $n = 4$ mice per group. AAV9 in *TFRC* KI and BI-hTFR1 in C57BL/6J had $n = 3$ mice per group. The **(4C)** biodistribution reported as vector genomes per mouse genome and **(4D)** Luciferase activity within different organs are shown three weeks post-injection. Two-way ANOVA test to determine significant differences within organs among the four groups of mice with BI-hTFR1 in *TFRC* KI mice as the main comparison group and using a Bonferroni multiple comparison correction: ****, ***, **, and * indicate $p \leq 0.0001$, ≤ 0.001 , ≤ 0.01 , and ≤ 0.05 , respectively; each data point represents an individual mouse, error bars indicate $\pm$ SEM.

[0044] FIG. 5A-5F – BI-hTFR1 efficiently transduces neurons and astrocytes throughout the CNS.

Representative **(5A)** whole brain and **(5B)** spinal cord images from each group of mice at three weeks post-injection are shown. Representative images show cells transduced by BI-hTFR1 overlaid with **(5C)** NeuN⁺ or **(5E)** SOX9⁺ stained cells in the cortex, thalamus, and striatum of *TFRC* KI mice. The percentages of **(5D)** NeuN⁺ neurons or **(5F)** SOX9⁺ astrocytes that expressed mScarlet in the cortex, striatum, and thalamus are shown. Two-way ANOVA test to determine significant differences in transduction within specific brain regions among the four groups of mice with BI-hTFR1 in *TFRC* KI mice as the main comparison group and using a Bonferroni multiple comparison correction: **** indicates $p \leq 0.0001$; each data point represents an individual mouse, error bars indicate $\pm$ SEM.

[0045] FIG. 6A-6D – BI-hTFR1 efficiently delivered *GBA1* and increased GCase activity in the brains of *TFRC* KI mice.

(6A) Schematic of the experiment shows the ssDNA AAV genome expressing human glucocerebrosidase that was packaged into AAV9 or BI-hTFR1 and administered to *TFRC* KI transgenic mice at either 1×10^{14} vg/kg or 5×10^{12} vg/kg. **(6B)** The biodistribution of AAV genomes found in brain and liver tissue relative to AAV9 ($n = 4$ mice per condition) is shown. One-way ANOVA with Sidak's multiple comparison post-hoc test using AAV9 as the control group: **** and *** indicate $p \leq 0.0001$ and ≤ 0.001 ; each data point

represents an individual mouse, and error bars indicate $\pm$ SEM. **(6C)** Sagittal brain sections (top) from the mice in each group show GBA-HA (magenta) in whole brain sagittal sections. Scale bar = 1 mm. Images of immunostaining (bottom) show neurons (NeuN, light gray) and GBA-HA (gray) in the substantia nigra pars compacta. Scale bar = 25 μ m. **(6D)** GCase enzyme activity levels in brain and liver tissue homogenate, CSF, and serum are shown. One-way ANOVA with Sidak's multiple comparison post-hoc tests with AAV9 as the control group: **** indicates $p \leq 0.0001$; each data point represents an individual mouse, and error bars indicate $\pm$ SEM.

[0046] FIG. 7A-7B – Validation assays confirm the expression of TfR1 from the indicated species in stable CHO cell lines. **(7A)** Species-specific primers (table S1) were used to assess *TFRC* or *Tfrc* mRNA expression levels by RT-qPCR. Each species primer set is shown with the indicated color (legend, top). The graph shows qPCR cycle threshold (Ct) values showing selective early cycle number amplification of the target species sequence. The bar indicates the mean Ct value ($n = 3$ replicates per cell line). **(7B)** Images show immunofluorescence for TfR1. Each cell line was fixed, permeabilized, and stained with the indicated anti-TfR1 antibodies to assess the expression of exogenous TfR1.

[0047] FIG. 8A-8D – The binding between BI-hTFR1 and human TfR1 is inhibited by the apical domain binding antibody OKT9 but not by holo-Tf. **(8A)** To assess binding between each AAV capsid and full-length human TfR1, AAVx probes were loaded with capsid and TfR1 was used as an analyte. Average kinetic constants and standard errors were computed from triplicate trials; * indicates a capsid for which values were computed from only two trials. **(8B)** The sensorgram shows the association of BI-hTFR1 to the Fc-hTfR1-loaded HFCII probe and dissociation in the presence of the indicated concentration of OKT9. The gray curve shows OKT9 binding to Fc-hTfR1 in the absence of BI-hTFR1. **(8C)** The binding kinetics between full-length human TfR1 and human holo-Tf were assessed by BLI. Mouse Fc (MFC) probes were loaded with an anti-transferrin antibody, 12A6. Holo-Tf was loaded onto the probe and associated with serial dilutions of TfR1 analyte. **(8D)** The effect of Tf on BI-hTFR1 binding to full-length human TfR1 was assessed by BLI. AAVx probes were loaded with AAV9 or BI-hTFR1. Human TfR1 that either had or had not been pre-incubated with 300 nM holo-Tf was used as an analyte. Average kinetic constants and standard errors were computed from triplicate trials.

[0048] FIG. 9A-9C – AAV colocalizes with TfR1 and Tf in brain endothelial cells. **(9A)** The indicated AAVs were incubated with hCMEC/D3 cells at 25,000 vg/cell for one hour at 37°C.

Cells were then fixed, permeabilized, and stained for AAV (light gray) and nuclei (dark gray). AAV2 was chosen as a control for subsequent experiments because AAV9 immunostaining was nearly undetectable. **(9B)** BI-hTFR1 or AAV2 was incubated with hCMEC/D3 cells at 30,000 vg/cell for one hour at 37°C. Cells were fixed, permeabilized, and stained for AAV (light gray) and TfR1 (dark grey). **(9C)** BI-hTFR1 or AAV2 was incubated with hCMEC/D3 cells at 25,000 vg/cell in prechilled media containing Tf-647 (1 µg/mL) for one hour at 4°C. Cells were then fixed and stained for AAV, Tf, and nuclei. All scale bars = 15 µm.

[0049] FIG. 10 – AAV colocalization with subcellular compartment markers. BI-hTFR1 or AAV2 was incubated with hCMEC/D3 cells at 25,000 vg/cell for one hour at 37°C. Cells were then fixed, permeabilized, and stained for AAV, endosomes (Rab5 and Rab7), cis-Golgi (RCAS1), trans-Golgi network (TGN46) or endoplasmic reticulum (KDEL (SEQ ID NO: 1787)), and nuclei. All scale bars = 15 µm.

[0050] FIG. 11A-11B – TfR1 levels in *TFRC* KI mice match those in WT mice. **(11A)** RT-qPCR using primers spanning exon 1 and exon 2 of the mouse *Tfrc* gene (table S1) was performed to assess mRNA levels in *TFRC* KI mice or C57BL/6J controls treated with either AAV9 or BI-hTFR1 ($n = 3$ or 4 animals per group; error bars indicate $\pm$ SEM). No significant difference in TfR1 mRNA between the C57BL/6J and *TFRC* KI mice was detected in the cortex or spinal cord (two-way ANOVA). **(11B)** Western blots of protein from the brain tissues of C57BL/6J and *TFRC* KI homozygous mice injected with AAV9 or BI-hTFR1 and purified full-length human or mouse TfR1 Fc-fusion proteins with N-terminal flag tag were stained using a polyclonal anti-human TfR1 antibody, AF2474 (R&D systems), or a monoclonal antibody, H68.4 (Thermofisher, Catalog #13-6890) that recognizes mouse and human TfR1.

[0051] FIG. 12A-12C – BI-hTFR1 exhibited a CNS-specific enhanced tropism in *TFRC* KI mice and similar transduction to AAV9 in the liver and dorsal root ganglion. BI-hTFR1 or AAV9:CAG-NLS-mScarlet-P2A-Luciferase-WPRE-SV40pA was intravenously injected into adult female C57BL/6J or *TFRC* KI mice at a dose of 5×10^{11} vg/mouse. The **(12A)** mScarlet transcript levels relative to AAV9 in C57BL/6J mice and native fluorescence in the **(12B)** liver and **(12C)** dorsal root ganglion at three weeks post-injection are shown. For **(12A)**, two-way ANOVA to determine significant differences within organs among the four groups of mice with BI-hTFR1 in *TFRC* KI mice as the main comparison group and using a Bonferroni multiple

comparison correction: ****, ***, **, and * indicate $p \leq 0.0001$, ≤ 0.001 , ≤ 0.01 , and ≤ 0.05 , respectively; each data point represents an individual mouse, error bars indicate $\pm$ SEM.

[0052] FIG. 13A-13C – GBA-HA-expression in mice's liver and brain regions injected with AAV9 or BI-hTFR1:GBA1. (13A) Images show NeuN (light gray) and HA (dark gray) immunostaining in the deep cerebellar nuclei and brainstem in *TFRC* KI mice with the indicated treatment conditions. Scale bar = 25 μ m. (13B) The cell counts of transduced NeuN+ cells expressing GBA-HA in the cortex, striatum, and thalamus of *TFRC* KI mice injected with 1×10^{14} vg/kg BI-hTFR1:GBA1 are shown. Each data point represents a mouse ($n = 4$). Error bars indicate $\pm$ SEM. (13C) Representative images of transduced liver hepatocytes stained for HA in the indicated conditions are shown. Scale bar = 50 μ m.

[0053] FIG. 14A-14C – CNS transduction by example Tfr1 binding AAVs. 5×10^{10} vg/animal of ssAAV-CAG-NLS-mScarlet-2A-Luciferase-pA was administered into the retro-orbital sinus of adult human *TFRC* knock-in mice and transduction was assessed 3 weeks post-injection. (14A) Images show native mScarlet fluorescence in sagittal brain sections from mice injected with Tfr1 binding capsid, BI-hTFR1 (top), a variant of BI294 (middle), or the BI295 (bottom). (14B, C) The graphs show the fraction of neurons (14B) or astrocytes (14C) with detectable mScarlet expression ($n=3$ per group).

[0054] FIG. 15A-15B – Peripheral organ transduction by example Tfr1 binding AAVs. 5×10^{10} vg/animal of ssAAV-CAG-NLS-mScarlet-2A-Luciferase-pA was administered into the retro-orbital sinus of adult *TFRC* knock-in mice and transduction was assessed 3 weeks post injection. (15A) Images show native mScarlet fluorescence in liver sections from mice injected with Tfr1 binding capsid, BI-hTFR1 (left), variant of BI294 (middle), or the BI295 (right). (15B) The graph shows the biodistribution assessment for AAV specific sequences ($n=3$ per group).

[0055] FIG. 16A-16D – Amino acid swaps between human and macaque Tfr1 identify specific human residues that are required for BI294 and BI295 binding to Tfr1. (16A) (SEQ ID NO: 1398-1400) An alignment of the human, macaque and marmoset amino acid sequences for the region containing the apical domain of Tfr1 (AA 200 to 380). AAs that differ from the human Tfr1 sequence are highlighted in blue. Four groupings of one to three surface exposed amino acids that differ between human and macaque are highlighted in outlined boxes in A and on the predicted structural models in 16B. (16C-D) CHO cells were transiently transfected with plasmids carrying a cDNA encoding human Tfr1, macaque Tfr1, human Tfr1 variants with

indicated amino acids replaced with the corresponding macaque amino acid sequence (left panels), or with macaque *TFRC* variants with the indicated amino acids replaced with the corresponding human amino acid (right panels). BI294:ssAAV-CAG-NLS-mScarlet-2A-Luciferase-pA (**16C**) or BI295:ssAAV-CAG-NLS-mScarlet-2A-Luciferase-pA (**16D**) was added to CHO cells expressing the indicated *TFRC* expression plasmid and transduction was assessed by measuring luciferase activity.

[0056] **FIG. 17 – AAV9 capsid residues 539-605 aligned to other previously described capsids.** (SEQ ID NO: 1401-1420) The 7-mer insertion site between AAV9 residue 588 and 589 is shown. The black bars above the alignment highlight surrounding residues that were modified in this study. Corresponding residues in other example capsid sequences are outlined and residues that differ from AAV9 are shown in gray. Sequences were aligned using MUSCLE (SnapGene).

[0057] **FIG. 18 – The SEQ ID 34 and surrounding AAV9 sequence (SEQ ID 45) imparts TfR1 binding into multiple AAV capsids.** Control CHO cells or stable CHO cells expressing human *TFRC* (hTFRC) were transduced with SEQ ID#21 (AAV9 w/SEQ ID#34), or modified AAV7 capsids where the SEQ ID#30 was inserted between AA589 and 590 (SEQ ID 28; AAV7 w/TfR1 targeting 7-mer) or where AA587-AA593 were replaced with SEQ ID#40, which comprises SEQ ID 30 and flanking amino acids from AAV9 (SEQ ID 29; AAV7 w/TfR1 targeting 7-mer and surrounding sequence). Bar plot shows the mean transduction (relative light units) per condition. Individual data points represent transduction replicates (n=4).

[0058] **FIG. 19 – Transferrin Receptor binding modification (SEQ ID NO: 1421-1433) – example insertions between residues 558 and 559 AAV9 VP1.**

[0059] **FIG. 20 – Example serotype sequence alignment (SEQ ID NO: 1, 3-5, 7-11, 15-18).**

[0060] **FIG. 21 – Transferrin Receptor binding moiety insertion site in AAV9 VP1 capsid (SEQ ID NO: 1, 1434-1436).**

[0061] **FIG. 22A-22B – Example TfR1-binding capsids exhibit increased biodistribution to the brain and spinal cord with a decreased biodistribution to the liver relative to AAV9 and BI-hTfR1 following IV administration into *TFRC* KI mice.** ssAAV-CAG-NLS-mScarlet-P2A-ffLuciferase-WPRE-pA was packaged into the indicated capsid and administered to 10 to 16-week-old humanized *TFRC* KI mice. AAV genome (WPRE) biodistribution was assessed 21 days later. (**22A**). The graph shows AAV genome biodistribution normalized relative to the house keeping gene GAPDH. (**22B**). The graph reports the data from **22A** after normalization to AAV9.

The viral genomes detected from BI-hTFR1, BI-hTFR1v2, BI-hTFR1v3 and BI-hTFR2v2 are 64-fold, 107-fold, 105-fold, and 56-fold greater than AAV9, respectively, in the spinal cord. n=6 animals per capsid (3 female and 3 males). Error bars indicate $\pm$ SEM.

[0062] FIG. 23A-23B – Example Tfr1-binding capsids more efficiently transduce the brain and spinal cord relative to AAV9 IV administration in *TFRC* KI mice. Transduction in the indicated organ was assessed by measuring Luciferase activity (relative luminescence units (RLU) per microgram of protein lysate). (23A). The graph shows luciferase RLU/ug protein in lysates from the indicated organs by the indicated capsids (legend). (23B). The graph reports the data from A expressed as the fold change in RLU per microgram of protein lysate relative to AAV9. n=6 animals per capsid (3 female and 3 males) of 10-16 weeks of age at time of injection. Error bars indicate $\pm$ SEM.

[0063] FIG. 24 – The TFR1 binding capsids mediate efficient transduction across the CNS. ssAAV-CAG-NLS-mScarlet-P2A-ffLuciferase-WPRE-pA was packaged into the indicated capsid and administered to 10 to 16-week-old humanized *TFRC* KI mice. mScarlet expression was assessed 21 days later. Images show native mScarlet (white) fluorescence from representative sagittal sections of B-hTFR1 mouse brains.

[0064] FIG. 25 – Example Capsid efficiently transduces neurons in the cortex of B-hTFR1 mice. Adult (10- to 16-week-old) humanized *TFRC* KI mice (B-hTFR1) were dosed with 2×10^{12} vg/kg of ssCAG-NLS-mScarlet-P2A-fLuciferase-WPRE-pA packaged into AAV9, AAV-PHP.eB, BI-hTFR1, or BI-hTFR1v2. Representative images of native mScarlet fluorescence (top) and the immunofluorescence for the neuronal marker NeuN (bottom) in the cerebral cortex 21 days post administration are shown.

[0065] FIG. 26 – Example capsid demonstrates increased transduction of neurons than BI-hTFR1 after IV delivery to humanized *TFRC* KI mice. Adult (10- to 16-week-old) humanized *TFRC* KI mice (B-hTFR1) were dosed with 2×10^{12} vg/kg of ssCAG-NLS-mScarlet-P2A-fLuciferase-WPRE-pA packaged into AAV-PHP.eB, BI-hTFR1, or BI-hTFR1v2 and the transduction of NeuN positive neurons in the cortex or striatum was assessed 21 days later by measuring the fraction of the NeuN+ cells that expressed mScarlet (native fluorescence). n=6 animals per capsid (3 female and 3 males). Data points represent values from individual animals. Error bars indicate $\pm$ SEM.

[0066] The figures herein are for illustrative purposes only and are not necessarily drawn to scale.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

Definitions

[0067] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.); Antibodies, A Laboratory Manual (1988) (Harlow and Lane, eds.); Antibodies A Laboratory Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471185710); Singleton *et al.*, Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (2011).

[0068] Titles or subtitles may be used in the specification for the sole convenience of the reader but are not intended to influence the scope of the present disclosure or to limit any aspect of the disclosure to any subsection, subtitle, or paragraph.

[0069] As used herein, the singular forms “a”, “an”, and “the” include both singular and plural referents unless the context clearly dictates otherwise. The term “optional” or “optionally” means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[0070] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges and the recited endpoints. For example, “1-5 ng” or a range of “1 ng to 5 ng” is intended to encompass 1 ng, 2 ng, 3 ng, 4 ng, 5 ng, 1-2 ng, 1-3 ng, 1-4 ng, 1-5 ng, 2-3 ng, 2-4 ng, 2-5 ng, 3-4 ng, 3-5 ng, and 4-5 ng.

[0071] The terms “about” or “approximately” as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, +/-5% or less, +/-1% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier “about” or “approximately” refers is itself also specifically and preferably, disclosed.

[0072] As used herein, a “biological sample” may contain whole cells and/or live cells and/or cell debris. The biological sample may contain (or be derived from) a “bodily fluid”. The present invention encompasses embodiments wherein the bodily fluid is selected from amniotic fluid, aqueous humour, vitreous humour, bile, blood serum, breast milk, cerebrospinal fluid, cerumen (earwax), chyle, chyme, endolymph, perilymph, exudates, feces, female ejaculate, gastric acid, gastric juice, lymph, mucus (including nasal drainage and phlegm), pericardial fluid, peritoneal fluid, pleural fluid, pus, rheum, saliva, sebum (skin oil), semen, sputum, synovial fluid, sweat, tears, urine, vaginal secretion, vomit and mixtures of one or more thereof. Biological samples include cell cultures, bodily fluids, cell cultures from bodily fluids. Bodily fluids may be obtained from a mammal organism, for example by puncture, or other collecting or sampling procedures.

[0073] The terms “subject,” “individual,” and “patient” are used interchangeably herein to refer to a vertebrate, preferably a mammal, more preferably a human. Mammals include, but are not limited to, murines, simians, humans, farm animals, sport animals, lab animals and pets. Tissues, cells and their progeny of a biological entity obtained *in vivo* or cultured *in vitro* are also encompassed.

[0074] As used herein, the phrase “at least one,” in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements, and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase “at least one”

refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, “at least one of A and B” (or, equivalently, “at least one of A or B,” or, equivalently “at least one of A and/or B”) can refer, in one aspect, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another aspect, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another aspect, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

[0075] As used herein, the terms “treat,” “treatment,” “treating” refer to therapeutic treatments, wherein the object is to reverse, alleviate, ameliorate, inhibit, slow down or stop the progression or severity of a disorder. The term “treating” includes reducing or alleviating at least one adverse effect or symptom of a condition, disease, or disorder. Treatment is generally “effective” if one or more symptoms or clinical markers are reduced. Alternatively, treatment is “effective” if the progression of a disorder is reduced or halted. That is, “treatment” includes not just the improvement of symptoms or markers, but also a cessation of, or at least slowing of, progress or worsening of symptoms compared to what would be expected in the absence of treatment. Beneficial or desired clinical results include, but are not limited to, alleviation of one or more symptom(s), diminishment of extent of disease, stabilized (i.e., not worsening) state of disease, delay or slowing of disease progression, amelioration or palliation of the disease state, remission (whether partial or total), and/or decreased mortality, whether detectable or undetectable. The term “treatment” of a disease also includes providing relief from the symptoms or side-effects of the disease (including palliative treatment).

[0076] The term “expression vector”, as used herein, refers to a nucleic acid construct comprising nucleic acid elements sufficient for the expression of a transgene product in a cell or in an in vitro assay. For example, a transgene expression vector, disclosed herein, comprises a promoter operatively linked to a transgene transcription unit comprising a transcription initiation site, a 5' untranslated region (UTR), a transgene nucleotide sequence and a 3' untranslated region (UTR) comprising one or more post-transcriptional regulatory elements, e.g., a polyadenylation sequence.

[0077] The term "operably linked" refers to a functional relationship between two or more polynucleotide (e.g., DNA) segments. Typically, the term refers to the functional relationship of a

transcriptional regulatory sequence and a transgene to be transcribed. For example, a promoter or enhancer sequence is operably linked to a transgene if it, e.g., stimulates or modulates the transgene transcription in an appropriate host cell or other expression system. Generally, promoter transcriptional regulatory' sequences that are operably linked to a sequence are contiguous to that sequence or are separated by short spacer sequences, i.e., they are cis-acting. However, some transcriptional regulatory sequences, such as enhancers, need not be physically contiguous or located in close proximity to the coding sequences whose transcription they enhance.

[0078] As used herein, "recombinant" refers to nucleic acids or polypeptides that are genetically engineered.

[0079] A "transgene" is a polynucleotide sequence that may encode an RNA (mRNA) that is translated into protein. For example, a transgene may comprise a cDNA sequence. In other embodiments, a transgene may encode on "non-coding" RNA that is not translated into protein (e.g. guide RNAs, ribozymes, aptamers, antisense RNAs, piwi-interacting RNAs (piRNAs), short interfering RNAs (siRNAs), microRNAs (miRNAs), shRNAs or recombinant U RNAs). In some aspects, the transgene nucleotide sequence may comprise one or more introns. In other aspects, the transgene can be polycistronic (e.g., two coding regions separated by internal ribosome entry site (IRES)). In some aspects, a transgene may encode more than one protein. In one aspect, a transgene comprises a "protein coding sequence" or a sequence that encodes a particular protein or polypeptide, i.e., a nucleic acid sequence that is capable of being transcribed into mRNA and translated into a polypeptide *in vitro* or *in vivo* when placed under the control of appropriate regulatory sequences. The boundaries of the coding sequence may be determined by a start codon at the 5' terminus (N-terminus) and a translation stop nonsense codon at the 3' terminus (C-terminus). A coding sequence can include, but is not limited to, cDNA from prokaryotic or eukaryotic mRNA, genomic DNA sequences from prokaryotic or eukaryotic DNA, and synthetic nucleic acids. A transcription termination sequence will usually be located 3' to the coding sequence. A "serotype" is traditionally defined on the basis of a lack of cross-reactivity between antibodies to one virus as compared to another virus. Such cross-reactivity differences are usually due to differences in capsid protein sequences/antigenic determinants (e.g., due to VP1, VP2, and/or VP3 sequence differences of AAV serotypes). Under the traditional definition, a serotype means that the virus of interest has been tested against serum specific for all existing and characterized serotypes for neutralizing activity and no antibodies have been found that neutralize

the virus of interest. As more naturally occurring virus isolates are discovered and capsid mutants generated, there may or may not be serological differences with any of the currently existing serotypes. Thus, in cases where the new AAV has no serological difference, this new AAV would be a subgroup or variant of the corresponding serotype. In many cases, serology testing for neutralizing activity has yet to be performed on mutant viruses with capsid sequence modifications to determine if they are of another serotype according to the traditional definition of serotype. Accordingly, for the sake of convenience and to avoid repetition, the term “serotype” broadly refers to both serologically distinct viruses (e.g., AAV) as well as viruses (e.g., AAV) that are not serologically distinct that may be within a subgroup or a variant of a given serotype.

[0080] “Transduction” refers to the transfer of a transgene into a recipient host cell by a viral vector. Transduction of a target cell by an rAAV virion of the invention leads to transfer of the transgene contained in that rAAV virion into the transduced cell. “Host cell” or “target cell” refers to the cell into which the DNA delivery takes place, such as the cells of the CNS or HEK293T cells in case of the *in vitro* transduction assay. AAV vectors are able to transduce both dividing and non-dividing cells. In a cell comprising a gene product of interest, such as for example GFP, the gene product of interest has been introduced/transferred/transduced by rAAV “transduction” of the cell. A cell into which the transgene has been introduced is referred to as a “transduced” cell.

[0081] Various embodiments are described hereinafter. It should be noted that the specific embodiments are not intended as an exhaustive description or as a limitation to the broader aspects discussed herein. One aspect described in conjunction with a particular embodiment is not necessarily limited to that embodiment and can be practiced with any other embodiment(s). Reference throughout this specification to “one embodiment”, “an embodiment”, “an example embodiment,” means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases “in one embodiment,” “in an embodiment,” or “an example embodiment” in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of

different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

OVERVIEW

[0082] Embodiments disclosed herein provide transferrin receptor binding modifications which promote transduction into the central nervous system (CNS) through its interaction with the cell surface transferrin receptor. These transferrin receptor binding modifications may be incorporated into particles, such as viral capsid delivery particles, including AAV9 particles and AAV particles of other serotypes, to confer tropism on the delivery particles and promote transduction of CNS. Exemplary CNS tissues include brain and spinal cord tissue. Exemplary CNS cell types include neurons, ependymal cells, and glial cells, e.g., microglia, astrocytes, oligodendrocytes, and NG2-glia progenitors, pericytes, as well as endothelial cells. Further embodiments disclosed herein provide for a vector system comprising one or more vectors encoding AAV capsids according to embodiments described herein. Accordingly, embodiments disclosed herein provide compositions capable of delivering cargos with enhanced selectivity and efficiency to the CNS vasculature. Embodiments disclosed herein also provide vector systems for the generation and loading of such delivery particles with cargo. Likewise, embodiments disclosed herein provide methods for use of such compositions to target CNS endothelial cells, *in vitro* and *in vivo*, with implications for both therapeutic and research purposes.

[0083] Additional features and advantages of the aforementioned embodiments are further described below.

Transferrin Receptor Binding Modifications and Compositions Thereof

[0084] In an embodiment, compositions are provided herein of AAV capsids comprising capsid proteins having a transferrin receptor binding modification sequence conferring on the capsid an enhanced tropism for endothelial cells of the CNS. A transferrin receptor binding modification with an enhanced tropism for endothelial cells of the CNS promotes, increases, or otherwise improves binding to, and in some cases, transduction of the CNS as compared to a natural or wild-type target moiety. This transferrin receptor binding modification may be coupled directly to a cargo to be delivered, such as an oligonucleotide or polypeptide. Alternatively, the targeting molecule may be incorporated into a delivery particle, such as an AAV particle (for example, being incorporated into an AAV capsid protein) to confer tropism for endothelial cells of the CNS on the delivery particle. A non-limiting example of a delivery particle is a viral capsid

particle. In such embodiments, the transferrin receptor binding modification may be incorporated into a viral capsid polypeptide such that the transferrin receptor binding modification is incorporated into the assembled viral capsid. However, other particle delivery systems where the transferrin receptor binding modification may be incorporated or attached, for example on exosomes, liposomes, lipid nanoparticles, virus-like particles, ribonucleoproteins, nanobodies, antibodies, or antibody fragments are also envisioned and encompassed as alternative embodiments herein.

[0085] In an embodiment, provided herein is a composition comprising a transferrin receptor binding modification effective to increase transduction of CNS tissues via binding to a transferrin receptor, optionally further comprising a cargo coupled to or otherwise associated with the transferrin receptor binding modification. A transferrin receptor binding modification with increased transduction promotes, enhances, or otherwise improves binding to and transduction of the CNS as compared to a natural or wild-type capsid without the transferrin receptor binding modifications described herein. In an example embodiment, the transferrin receptor binding modification binds to the extracellular domain of transferrin receptor (TFR). In an example embodiment, the transferrin receptor binding modification binds to one or more of the apical, helical, and/or protease-like domains of the extracellular domain. In an example embodiment, the transferrin receptor binding modification binds to the apical domain. In an example embodiment, the n-mer is an amino acid sequence of length n. The length of the n-mer may be any necessary length to transduce the CNS. In an embodiment, the n-mer is 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, or 12 amino acids in length. In an example embodiment, the n-mer motif has a length of at least 7 amino acids. In an example embodiment, a composition comprising a targeting moiety effective to increase transduction of CNS tissues comprises a n-mer motif, the n-mer motif comprising or consisting of an amino acid sequence of $Z_1-X_1-Z_2-X_2-X_3-X_4-X_5$ wherein Z_1 is Y, F, or L, Z_2 is S, R, or K, X_1-X_5 are independently selected amino acids. In an example embodiment, X_1 optionally comprises of A, S, or H; X_2 optionally comprises of S, T, L, or I; X_3 optionally comprises N or G; X_4 optionally comprises G; X_5 optionally comprises of N, D, I, V, or R. In an example embodiment, the targeting moiety comprises a n-mer motif, the n-mer motif comprising or consisting of an amino acid sequence of is $X_1-H-X_2-L-X_3-X_4-X_5$ wherein X_1-X_5 are independently selected amino acids.

[0086] In an embodiment, the n-mer can be used to increase transduction in target cells i.e., CNS cells and tissues. The increase in transduction efficiency (which may correspond to the tropism efficiency) of the n-mer to a cell may be compared to a composition that does not contain the transferrin receptor binding modification for example inclusion of one or more transferrin receptor binding modifications in a composition can result in an increase in transduction and or transduction efficiency by 10%, 20%, 30%, 40%, 50%, 60% 70% 80% 90% a 100% or more. In an exemplary embodiment the increase in transduction and or transduction efficiency is one and a half fold, two-fold, three-fold, four-fold, five-fold, six-fold, seven-fold, eight-fold, nine-fold, ten-fold, 20-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold or more relative to a composition lacking the n-mer. In one embodiment the transduction and/or transduction efficiency is increased or enhanced in endothelial cells, in one embodiment increase in endothelial cells of the vasculature, for example, the central nervous system vasculature. In an embodiment, the transduction and /or transduction efficiency is increased or enhanced in cells of the central nervous system. In an embodiment, the transduction and /or transduction efficiency is increased or enhanced in neurons and glial cells In an embodiment, the composition comprising a n-mer is selective to a target cell as compared to other cell types and/or other virus particles. As used herein, 'selective' and 'cell-selective' refers to preferential targeting for cells as compared to other cell types. Preferably, the transferrin receptor binding modification is selective for a desired target (e.g., cell, organ, system e.g., CNS tissues) or set of targets by at least 2:1, 3:1, 4:1, 5:1, 6:1 7:1, 8:1, 9:1. 10:1 or more; or 10%, 20%, 30%, 40%, 50%, 60%, 70%, 75% 80%, 85%, 90% or more, relative to other targets or cells (e.g., CNS). In an example embodiment, the composition comprising a Transferrin receptor binding modification described herein can have an increased uptake, delivery rate, transduction rate, efficiency, amount, or a combination thereof in a target cell (e.g., endothelial cells across the CNS e.g., brain endothelium, arterio-venous axis in brain, retina, and spinal cord vasculature) as compared to other cells types (e.g. hepatocytes) and/or other virus particles (e.g., AAVs not containing the transferrin receptor binding modification) and other compositions that do not contain the cell-selective n-mer motif of the present invention.

[0087] In an example embodiment, the n-mer motif is selected from the group consisting of YSRIGPN (SEQ ID NO: 1390), YSRLNMN (SEQ ID NO: 1437), YSRLNKD (SEQ ID NO: 1438), and YHRLSNN (SEQ ID NO: 1439). In an example embodiment, the n-mer is VHRLQDK (SEQ ID NO: 1440) or LHALSHN (SEQ ID NO: 1441). In an example embodiment, the n-mer is

or comprises PSATNGV (SEQ ID NO: 1442), QVSTNGI (SEQ ID NO: 1443), SYSSNGV (SEQ ID NO: 1444), HQSSNGV (SEQ ID NO: 1445), VGSINGI (SEQ ID NO: 1446), AMSTNGR (SEQ ID NO: 1447), SASTNGV (SEQ ID NO: 1448), YMSTNGV (SEQ ID NO: 1449), YYSSNGV (SEQ ID NO: 1450), VHSTNGI (SEQ ID NO: 1451), PLSTNGV (SEQ ID NO: 1452), VYSTNGI (SEQ ID NO: 1453), IISTNGV (SEQ ID NO: 1454), RSVSSNGV (SEQ ID NO: 1455), YKSSNGV (SEQ ID NO: 1456), FRSTNGV (SEQ ID NO: 1393), and/or FVSTNGV (SEQ ID NO: 1394). In an example embodiment, the n-mer is selected from any of the amino acid sequences in Tables 1-2 or any combination thereof. In an example embodiment, the transferrin receptor binding modification is part of (e.g., inserted between consecutive amino acids of) a viral capsid protein, including an AAV capsid protein. In an embodiment, the TfR1 binding modification comprises or consists of the amino acid sequence of EAQYSRIGPNNQAQ (SEQ ID NO: 170); SADFRSTNGVAQAE (SEQ ID NO: 1215); or GAQYSRIGPNPQPE (SEQ ID NO: 243).

Transferrin Receptor (TfR) Binding

[0088] The role of the transferrin-transferrin receptor interaction is to transport iron into the vicinity of the cell surface, thereby increasing iron uptake. Transferrin is a blood plasma protein for iron ion delivery. Transferrin is a glycoprotein, which binds iron very tightly but reversibly. Although iron bound to transferrin is less than 0.1% (4 mg) of the total body iron, dynamically it represents the most important iron pool, with the highest rate of turnover (25 mg/24 h). Transferrin has a molecular weight of around 80 kDa and contains 2 specific high affinity Fe binding sites. The affinity of transferrin for Fe is extremely high (10^{23} M^{-1} at pH 7.4) but decreases progressively with decreasing pH below neutrality.

[0089] Transferrin-bound iron is delivered to a cell through a transferrin receptor-mediated endocytosis. In vertebrates, there are two transferrin receptors. Transferrin receptor protein 1 (TfR1), also known as Cluster of Differentiation 71, (CD71), P90, CD71, TRFR, Trfr, TFR, or IMD46, is a protein that in humans is encoded by the TFRC gene (Human TfR1 mRNA: Accession No. X01060.1). Human TFRC (HGNC:11763; Gene ID: ENSG00000072274), located on chromosome 3, is transcribed into 38 transcripts (splice variants), 284 orthologues, 5 paralogues and is associated with 47 phenotypes.

[0090] Human TfR1 mediates cellular uptake of both transferrin and H-ferritin, an iron storage protein. TfR1 is a transmembrane homodimer that consists of two identical monomers with a molecular mass of approximately 90 kDa (760 aa); each monomer is joined by two disulfide bonds

at Cys89 and Cys98. TfR1 further comprises a short, NH₂-terminal cytoplasmic region (residues 1 to 67), a single transmembrane pass (residues 68 to 88), and a large extracellular portion (ectodomain, residues 89 to 760), which is soluble and bears a trypsin-sensitive site and contains a binding site for transferrin. Synthesized in the endoplasmic reticulum, TfR1 is post-translationally modified with both phosphate and fatty acyl groups. TfR1's extracellular domain contains three N-linked glycosylation sites at Asn251, Asn317, and Asn727 and one O-linked glycosylation site at Thr104. TfR1 is typically found as a dimer linked by disulfide bonds on the cell surface, see FIG. 4.

[0091] Human transferrin receptor 2 (TfR2), located on chromosome 7, is a protein that in humans is encoded by the TFR2 gene (Gene Identifier: ENSG00000106327). TFR2 is less commonly expressed than TFRC and is predominantly expressed in hepatocytes.

[0092] The TFR1 contains an extracellular C-terminal domain (around 671 amino acids) and comprises of the TF binding site. The extracellular C-terminal domain comprises of three subdomains: apical, helical, and protease-like domain, see FIG. 4. Furthermore, the extracellular C-terminal domain comprises of three N-linked glycosylation sites at asparagine residues 251, 317, and 727 and one O-linked glycosylation site at threonine 104, which contribute to adequate function of the receptor. TFR1 additionally comprises of a transmembrane domain (around 29 amino acids), and an intracellular N-terminal domain (around 61 amino acids). Alternative to the delivery of iron through TF with TFR1, uptake of iron can also occur through H-ferritin binding to the apical domain. See also Candelaria, P. V.; *et al.* Antibodies Targeting the Transferrin Receptor 1 (TfR1) as Direct Anti-Cancer Agents. *Frontiers in Immunology*, 2021, 12, hereby incorporated by reference.

[0093] In an example embodiment, the transferrin receptor binding modification binds to the extracellular domain of TFR1. In an example embodiment, the transferrin receptor binding modification binds to one or more of the helical (residues 606-760), the protease-like (residues 121-183, 384-605) and the apical (residues 184-383) domain (Lawrence et al., Crystal structure of the ectodomain of human transferrin receptor. *Science*. 1999;286(5440):779-82). In an embodiment, the transferrin receptor binding modification binds to the apical domain. As used herein, the term “a transferrin receptor” refers to TfR1 or TfR2 or any one of the alternatively spliced variants, orthologues, paralogues. In an embodiment, the term “a transferrin receptor” refers to TfR1 or any one of the alternatively spliced variants, orthologues, paralogues.

Table A. Human Transferrin Receptor 1 cDNA sequence.

Location	HUMAN TRANSFERRIN RECEPTOR 1 cDNA SEQUENCE (AMINO ACID SEQ ID NO: 1459) (NUCLEOTIDE SEQ ID NO: 1460)	Location	DOMAINS
1	M M D Q A R S A F S N L F G G E P L S Y	20	NH2-TERMINAL CYTOPLASMIC REGION (aa 1-67)
1	ATGATGGATCAAGCTAGATCAGCATCTCTAACTTGTGGTGAGAACCATTTGTCATAT	60	
21	T R F S L A R Q V D G D N S H V E M K L	40	
61	ACCCGGTTTCAGCCTGGCTCGGCAAGTAGATGGCGATAACAGTCATGTGGAGATGAAACCTT	120	
41	A V D E E N A D N N T K A N V T K P K	60	
12	GCTGTAGATGAAGAAGAAAAATGCTGACAATAACACAAAAGGCCAATGTCACAAAAACCAAAA	180	
61	R C S G S I C (SEQ ID NO: 1461)		
181	AGGTGTAGTGAAGTATCTGC (SEQ ID NO: 1462)		
	Y G T I A V I V F F L I G	80	TRANSMEMBRANE SEGMENT (aa 68-88)
	TATGGGACTATTGCTGTGATCGTCTTTTCTTGATTGGA	240	
81	F M I G Y L G Y (SEQ ID NO: 1463)		
241	TTTATGATTGGCTACTTGGGCTAT (SEQ ID NO: 1464)		
	C K G V E P K T E C E R	100	EXTRACELLULAR DOMAIN (aa 89-760) (AMINO ACID SEQ ID NO: 1457) (NUCLEOTIDE SEQ ID NO: 1458)
	TGTAAGGGGTAGAACCAAAAACTGAGTGTGAGAGA	300	
101	L A G T E S P V R E E P G E D F P A A R	120	
301	CTGGCAGGAACCGAGTCTCCAGTGAGGGAGGAGCCAGAGAGGACTTCCCTGCAGCACGT	360	
121	R L Y W D D L K R K L S E K L D S T D F	140	
361	CGCTTATATTGGGATGACCTGAAGAGAAAAAGTTGTCGGAGAAACTGGACAGCACAGACTTC	420	
141	T S T I K L L N E N S Y V P R E A G S Q	160	
421	ACCAGCACCATCAAGCTGCTGAATGAAAAATTCATATGTCCTCGTGAGGCTGGATCTCAA	480	
161	K D E N L A L Y V E N Q F R E F K L S K	180	
481	AAAGATGAAAAATCTTGCGTTGTATGTTGAAAAATCAATTTCTGTGAATTTAAACTCAGCAAA	540	
181	V W R (SEQ ID NO: 1465)		PROTEASE -LIKE DOMAIN I (aa 121-183)
541	GTCTGGCGT (SEQ ID NO: 1466)		
	D Q H F V K I Q V K D S A Q N S V	200	APICAL DOMAIN (aa 184-383)
	GATCAACATTTTGTAAAGATTCAGGTCAAAGACAGCGCTCAAAACTCGGTG	600	
201	I I V D K N G R L V Y L V E N P G G Y V	220	
601	ATCATAGTTTGATAAGAACGGTAGACTTGTTTTACCTGGTGGAGAAATCCTGGGGGTTATGTG	660	

Location	HUMAN TRANSFERRIN RECEPTOR 1 cDNA SEQUENCE (AMINO ACID SEQ ID NO: 1459) (NUCLEOTIDE SEQ ID NO: 1460)																Location	DOMAINS				
221	A	Y	S	K	A	A	T	V	T	G	K	L	V	H	A	N	F	G	T	K	240	EXTRACELLULAR DOMAIN (aa 89-760) (AMINO ACID SEQ ID NO: 1457) (NUCLEOTIDE SEQ ID NO: 1458)
661	G	C	G	T	A	T	A	G	G	C	T	G	C	A	A	C	A	G	T	T	720	
241	K	D	F	E	D	L	Y	T	P	V	N	G	S	I	V	I	V	R	A	G	260	
721	A	A	G	A	T	T	T	G	A	G	A	T	T	A	C	A	C	T	C	T	780	
261	K	I	T	F	A	E	K	V	A	N	A	E	S	L	N	A	I	G	V	L	280	
781	A	A	A	T	C	A	C	T	T	T	G	C	A	G	A	A	A	G	T	T	840	
281	I	Y	M	D	Q	T	K	F	P	I	V	N	A	E	L	S	F	F	G	H	300	
841	A	T	A	T	A	C	A	T	G	G	A	C	A	C	T	A	A	A	T	T	900	
301	A	H	L	G	T	G	D	P	Y	T	P	G	F	P	S	F	N	H	T	Q	320	
901	G	C	T	C	A	T	C	T	G	G	G	A	C	A	G	T	T	C	C	T	960	
321	F	P	P	S	R	S	S	G	L	P	N	I	P	V	Q	T	I	S	R	A	340	
961	T	T	T	C	A	C	C	A	T	C	T	G	G	T	C	A	T	G	C	C	1020	
341	A	A	E	K	L	F	G	N	M	E	G	D	C	P	S	D	W	K	T	D	360	
1021	G	C	T	G	C	A	G	A	A	A	G	C	T	G	T	T	G	G	A	A	1080	
361	S	T	C	R	M	V	T	S	E	S	K	N	V	K	L	T	V	S	N	V	380	
1081	T	C	T	A	C	A	T	G	T	A	A	C	C	T	C	A	G	A	A	T	1140	
381	L	K	E	(SEQ ID NO: 1467)																		
1141	C	T	G	A	A	A	G	A	G	(SEQ ID NO: 1468)												
		I	K	I	L	N	I	F	G	V	I	K	G	F	V	E	P	D				
401	H	Y	V	V	G	A	Q	R	D	A	W	G	P	G	A	A	K	S	G			
1201	C	A	T	A	T	G	T	T	G	G	G	C	C	A	G	A	G	A	T	G	C	420
421	V	G	T	A	L	L	L	K	L	A	Q	M	F	S	D	M	V	L	K	D	1260	
1261	G	T	A	G	G	C	A	G	A	C	T	T	G	C	C	A	G	A	T	G	T	440
441	G	F	Q	P	S	R	S	I	I	F	A	S	W	S	A	G	D	F	G	S	1320	
1321	G	G	G	T	T	C	A	G	C	C	A	G	A	G	A	T	T	T	C	G	T	460
461	V	G	A	T	E	W	L	E	G	Y	L	S	S	L	H	L	K	A	F	T	1380	
1381	G	T	T	G	T	G	C	C	A	C	T	T	T	C	G	T	C	C	T	G	C	480
481	Y	I	N	L	D	K	A	V	L	G	T	S	N	F	K	V	S	A	S	P	1440	
1441	T	A	T	A	T	A	T	C	T	G	G	A	T	A	A	G	C	C	A	T	T	500
																					1500	

Location	HUMAN TRANSFERRIN RECEPTOR 1 cDNA SEQUENCE (AMINO ACID SEQ ID NO: 1459) (NUCLEOTIDE SEQ ID NO: 1460)																				Location	DOMAINS
501	L	L	Y	T	L	I	E	K	T	M	Q	N	V	K	H	P	V	T	G	Q	520	HELICAL DOMAIN (aa 606-760)
1501	CTGTTGTATACGCTTATTGAGAAAACAATGCAAAATGTGAAGCATCCGGTTACTGGGCAA																				1560	
521	F	L	Y	Q	D	S	N	W	A	S	K	V	E	K	L	T	L	D	N	A	540	
1561	TTTCTATATCAGGACAGCAAACTGGGCCAGCAAAGTTGAGAAACTCACTTTAGACAAATGCT																				1620	
541	A	F	P	F	L	A	Y	S	G	I	P	A	V	S	F	C	F	C	E	D	560	
1621	GCTTCCCTTTCCCTTGCATATTCTGGAATCCCAGCAGTTTCTTCTGTGTTTTTTCGAGGAC																				1680	
561	T	D	Y	P	Y	L	G	T	T	M	D	T	Y	K	E	L	I	E	R	I	580	
1681	ACAGATTATCCTTATTTTGGGTACCACCATGGACACCTATAAGGAACTGATTGAGAGGATT																				1740	
581	P	E	L	N	K	V	A	R	A	A	A	E	V	A	G	Q	F	V	I	K	600	
1741	CCTGAGTTGAACAAAGTGGCACGAGCAGCTGCAGAGGTGCTGGTCAGTTCGTGATTAAA																				1800	
601	(SEQ ID NO: 1469)																					
1801	CTAACCCATGATGTT (SEQ ID NO: 1470)																					
	E L N L D Y E R Y N S Q L L S					GAATTGAACCTGGACTATGAGAGGTACAACAGCCAACTGCTTTCA															620	
621	F	V	R	D	L	N	Q	Y	R	A	D	I	K	E	M	G	L	S	L	Q	1860	
1861	TTTGTGAGGGATCTGAACCAATACAGAGCAGACATAAAGGAAATGGGCTGAGTTTACAG																				1920	
641	W	L	Y	S	A	R	G	D	F	F	R	A	T	S	R	L	T	T	D	F	660	
1921	TGGCTGTATTCTGCTCGTGGAGACTTCTTCCGTGCTACTTCCAGACTAACCAACAGATTTC																				1980	
661	G	N	A	E	K	T	D	R	F	V	M	K	K	L	N	D	R	V	M	R	680	
1981	GGAAATGCTGAGAAAAACAGACAGATTGTGTCATGAAGAAACTCAATGATCGTGTCAATGAGA																				2040	
681	V	E	Y	H	F	L	S	P	Y	V	S	P	K	E	S	P	F	R	H	V	700	
2041	GTGGAGTATCACTTCTCTCTCCCTACGTATCTCCAAAAGAGTCTCCTTTCCGACATGTC																				2100	
701	F	W	G	S	G	S	H	T	L	P	A	L	L	E	N	L	K	L	R	K	720	
2101	TTCTGGGGCTCCGGCTCTCACACGCTGCCAGCTTTACTGGAGAACTTGAAAACGCGTAAA																				2160	
721	Q	N	N	G	A	F	N	E	T	L	F	R	N	Q	L	A	L	A	T	W	740	
216	CAAAATAACGGTGCTTTTAATGAAAACGCTGTTCAGAAAACCAAGTTGGCTCTAGCTACTTGG																				2220	
741	T	I	Q	G	A	A	N	A	L	S	G	D	V	W	D	I	D	N	E	F	760	
2221	ACTATTCAGGGAGCTGCAAAATGCCCTCTCTGGTGACGTTTGGGACATTGACAAATGAGTTT																				2280	
761	* (SEQ ID NO: 1471)																				761	
2281	TAA (SEQ ID NO: 1472)																				2283	
EXTRACELLULAR DOMAIN (aa 89-760) (AMINO ACID SEQ ID NO: 1457) (NUCLEOTIDE SEQ ID NO: 1458)																						

Engineered Viral Capsids

[0094] Described herein are various embodiments of engineered viral capsids, such as adeno-associated virus (AAV) capsids, that can be engineered to confer cell-selective tropism, such as CNS tissue- and cell-specific tropism, to an engineered viral particle. Engineered viral capsids can be adenoviral or AAV capsids. The engineered capsids can be included in an engineered virus particle (e.g., an engineered adenoviral or AAV virus particle), and can confer cell-selective tropism to the engineered viral particle. The engineered viral capsids described herein can include one or more engineered viral capsid proteins described herein. The engineered viral capsids described herein can include one or more engineered viral capsid proteins described herein that can contain one or more transferrin receptor binding modification sequence as described elsewhere herein.

[0095] The engineered viral capsids can be variants of a wild-type viral capsid. For example, in an embodiment, the engineered AAV capsids can be variants of wild-type AAV capsids. In an embodiment, the wild-type AAV capsids can be composed of VP1, VP2, VP3 capsid proteins or a combination thereof. In other words, the engineered AAV capsids can include one or more variants of a wild-type VP1, wild-type VP2, and/or wild-type VP3 capsid proteins.

[0096] In an embodiment, the serotype of the reference wild-type AAV capsid can be AAV9, AAV9 K449R (or K449R AAV9), AAV1, AAVrh10, AAV-DJ, AAV-DJ8, AAV5, AAV-PHP.B (PHP.B), AAV-PHP.A (PHP.A), AAVG2B-26, AAVG2B-13, AAVTH1.1-32, AAVTH1.1-35, AAV-PHP.B2 (PHP.B2), AAV-PHP.B3 (PHP.B3), AAV-PHP.N/PHP.B-DGT, AAV-PHP.B-EST, AAV-PHP.B-GGT, AAV-PHP.B-ATP, AAV-PHP.B-ATT-T, AAV-PHP.B-DGT-T, AAV-PHP.B-GGT-T, AAV-PHP.B-SGS, AAV-PHP.B-AQP, AAV-PHP.B-QQP, AAV-PHP.B-SNP(3), AAV-PHP.B-SNP, AAV-PHP.B-QGT, AAV-PHP.B-NQT, AAV-PHP.B-EGS, AAV-PHP.B-SGN, AAV-PHP.B-EGT, AAV-PHP.B-DST, AAV-PHP.B-DST, AAV-PHP.B-STP, AAV-PHP.B-PQP, AAV-PHP.B-SQP, AAV-PHP.B-QLP, AAV-PHP.B-TMP, AAV-PHP.B-TTP, AAV-PHP.S/G2A12, AAVG2A15/G2A3 (G2A3), AAVG2B4 (G2B4), AAVG2B5 (G2B5), PHP.S, AAV2, AAV2G9, AAV3, AAV3a, AAV3b, AAV3-3, AAV4, AAV4-4, AAV6, AAV6.1, AAV6.2, AAV6.1.2, AAV7, AAV7.2, AAV8, AAV9.11, AAV9.13, AAV9.16, AAV9.24, AAV9.45, AAV9.47, AAV9.61, AAV9.68, AAV9.84, AAV9.9, AAV10, AAV11, AAV12, AAV16.3, AAV24.1, AAV27.3, AAV42.12, AAV42-1b, AAV42-2, AAV42-3a, AAV42-3b, AAV42-4, AAV42-5a, AAV42-5b, AAV42-6b, AAV42-8, AAV42-10, AAV42-11,

AAV42-12, AAV42-13, AAV42-15, AAV42-aa, AAV43-1, AAV43-12, AAV43-20, AAV43-21, AAV43-23, AAV43-25, AAV43-5, AAV44.1, AAV44.2, AAV44.5, AAV223.1, AAV223.2, AAV223.4, AAV223.5, AAV223.6, AAV223.7, AAV1-7/rh.48, AAV1-8/rh.49, AAV2-15/rh.62, AAV2-3/rh.61, AAV2-4/rh.50, AAV2-5/rh.51, AAV3.1/hu.6, AAV3.1/hu.9, AAV3-9/rh.52, AAV3-11/rh.53, AAV4-8/rh.1.64, AAV4-9/rh.54, AAV4-19/rh.55, AAV5-3/rh.57, AAV5-22/rh.58, AAV7.3/hu.7, AAV16.8/hu.10, AAV16.12/hu.11, AAV29.3/bb.1, AAV29.5/bb.2, AAV106.1/hu.37, AAV114.3/hu.40, AAV127.2/hu.41, AAV127.5/hu.42, AAV128.3/hu.44, AAV130.4/hu.48, AAV145.1/hu.53, AAV145.5/hu.54, AAV145.6/hu.55, AAV161.10/hu.60, AAV161.6/hu.61, AAV33.12/hu.17, AAV33.4/hu.15, AAV33.8/hu.16, AAV52/hu.19, AAV52.1/hu.20, AAV58.2/hu.25, AAVA3.3, AAVA3.4, AAVA3.5, AAVA3.7, AAVC1, AAVC2, AAVC5, AAVF3, AAVF5, AAVH2, AAVrh.72, AAVhu.8, AAVrh.68, AAVrh.70, AAVpi.1, AAVpi.3, AAVpi.2, AAVrh.60, AAVrh.44, AAVrh.65, AAVrh.55, AAVrh.47, AAVrh.69, AAVrh.45, AAVrh.59, AAVhu.12, AAVH6, AAVH-1/hu.1, AAVH-5/hu.3, AAVLG-10/rh.40, AAVLG-4/rh.38, AAVLG-9/hu.39, AAVN721-8/rh.43, AAVCh.5, AAVCh.5R1, AAVcy.2, AAVcy.3, AAVcy.4, AAVcy.5, AAVCy.5R1, AAVCy.5R2, AAVCy.5R3, AAVCy.5R4, AAVcy.6, AAVhu.1, AAVhu.2, AAVhu.3, AAVhu.4, AAVhu.5, AAVhu.6, AAVhu.7, AAVhu.9, AAVhu.10, AAVhu.11, AAVhu.13, AAVhu.15, AAVhu.16, AAVhu.17, AAVhu.18, AAVhu.20, AAVhu.21, AAVhu.22, AAVhu.23.2, AAVhu.24, AAVhu.25, AAVhu.27, AAVhu.28, AAVhu.29, AAVhu.29R, AAVhu.31, AAVhu.32, AAVhu.34, AAVhu.35, AAVhu.37, AAVhu.39, AAVhu.40, AAVhu.41, AAVhu.42, AAVhu.43, AAVhu.44, AAVhu.44R1, AAVhu.44R2, AAVhu.44R3, AAVhu.45, AAVhu.46, AAVhu.47, AAVhu.48, AAVhu.48R1, AAVhu.48R2, AAVhu.48R3, AAVhu.49, AAVhu.51, AAVhu.52, AAVhu.54, AAVhu.55, AAVhu.56, AAVhu.57, AAVhu.58, AAVhu.60, AAVhu.61, AAVhu.63, AAVhu.64, AAVhu.66, AAVhu.67, AAVhu.14/9, AAVhu.19, AAVrh.2, AAVrh.2R, AAVrh.8, AAVrh.8R, AAVrh.10, AAVrh.12, AAVrh.13, AAVrh.13R, AAVrh.14, AAVrh.17, AAVrh.18, AAVrh.19, AAVrh.20, AAVrh.21, AAVrh.22, AAVrh.23, AAVrh.24, AAVrh.25, AAVrh.31, AAVrh.32, AAVrh.33, AAVrh.34, AAVrh.35, AAVrh.36, AAVrh.37, AAVrh.37R2, AAVrh.38, AAVrh.39, AAVrh.40, AAVrh.46, AAVrh.48, AAVrh.48.1, AAVrh.48.1.2, AAVrh.48.2, AAVrh.49, AAVA.51, AAVrh.52, AAVrh.53, AAVrh.54, AAVrh.56, AAVrh.57, AAVrh.58, AAVA.61, AAVrh.64, AAVA.64R1, AAVA.64R2, AAVrh.67, AAVrh.73, AAVrh.74, AAVA8R, AAVA8R A586R mutant, AAVA8R R533A mutant, AAV, BAAV, caprine AAV, bovine AAV, AAVhEl.1,

AAVhErl .5, AAVhErl .14, AAVhErl .8, AAVhErl .16, AAVhErl .18, AAVhErl .35, AAVhErl .7, AAVhErl .36, AAVhEr2.29, AAVhEr2.4, AAVhEr2.16, AAVhEr2.30, AAVhEr2.31, AAVhEr2.36, AAVhErl .23, AAVhEr3.1, AAV2.5T, AAV-PAEC, AAV-LK01, AAV-LK02, AAV-LK03, AAV-LK04, AAV-LK05, AAV-LK06, AAV-LK07, AAV-LK08, AAV-LK09, AAV-LK10, AAV-LK11, AAV-LK12, AAV-LK13, AAV-LK14, AAV-LK15, AAV-LK16, AAV-LK17, AAV-LK18, AAV-LK19, AAV-PAEC2, AAV-PAEC4, AAV-PAEC6, AAV-PAEC7, AAV-PAEC8, AAV-PAEC11, AAV-PAEC12, AAV-2-pre-miRNA-lol, AAV-8h, AAV-8b, AAV-h, AAV-b, AAV SM 10-2, AAV Shuffle 100-1, AAV Shuffle 100-3, AAV Shuffle 100-7, AAV Shuffle 10-2, AAV Shuffle 10-6, AAV Shuffle 10-8, AAV Shuffle 100-2, AAV SM 10-1, AAV SM 10-8, AAV SM 100-3, AAV SM 100-10, BNP61 AAV, BNP62 AAV, BNP63 AAV, AAVrh.50, AAVrh.43, AAVrh.62, AAVrh.48, AAVhu.19, AAVhu.11, AAVhu.53, AAV4-8/rh.64, AAVLG-9/hu.39, AAV54.5/hu.23, AAV54.2/hu.22, AAV54.7/hu.24, AAV54.1/hu.21, AAV54.4R/hu.27, AAV46.2/hu.28, AAV46.6/hu.29, AAV128.1/hu.43, true type AAV (ttAAV), UPEN AAV 10, Japanese AAV 10 serotypes, AAV CBr-7.1, AAV CBr-7.10, AAV CBr-7.2, AAV CBr-7.3, AAV CBr-7.4, AAV CBr-7.5, AAV CBr-7.7, AAV CBr-7.8, AAV CBr-B7.3, AAV CBr-B7.4, AAV CBr-E1, AAV CBr-E2, AAV CBr-E3, AAV CBr-E4, AAV CBr-E5, AAV CBr-e5, AAV CBr-E6, AAV CBr-E7, AAV CBr-E8, AAV CHt-1, AAV CHt-2, AAV CHt-3, AAV CHt-6.1, AAV CHt-6.10, AAV CHt-6.5, AAV CHt-6.6, AAV CHt-6.7, AAV CHt-6.8, AAV CHt-P1, AAV CHt-P2, AAV CHt-P5, AAV CHt-P6, AAV CHt-P8, AAV CHt-P9, AAV CKd-1, AAV CKd-10, AAV CKd-2, AAV CKd-3, AAV CKd-4, AAV CKd-6, AAV CKd-7, AAV CKd-8, AAV CKd-B1, AAV CKd-B2, AAV CKd-B3, AAV CKd-B4, AAV CKd-B5, AAV CKd-B6, AAV CKd-B7, AAV CKd-B8, AAV CKd-H1, AAV CKd-H2, AAV CKd-H3, AAV CKd-H4, AAV CKd-H5, AAV CKd-H6, AAV CKd-N3, AAV CKd-N4, AAV CKd-N9, AAV CLg-F1, AAV CLg-F2, AAV CLg-F3, AAV CLg-F4, AAV CLg-F5, AAV CLg-F6, AAV CLg-F7, AAV CLg-F8, AAV CLv-1, AAV CLvl-1, AAV Clvl-10, AAV CLvl-2, AAV CLv-12, AAV CLvl-3, AAV CLv-13, AAV CLvl-4, AAV Clvl-7, AAV Clvl-8, AAV Clvl-9, AAV CLv-2, AAV CLv-3, AAV CLv-4, AAV CLv-6, AAV CLv-8, AAV CLv-D1, AAV CLv-D2, AAV CLv-D3, AAV CLv-D4, AAV CLv-D5, AAV CLv-D6, AAV CLv-D7, AAV CLv-D8, AAV CLv-E1, AAV CLv-K1, AAV CLv-K3, AAV CLv-K6, AAV CLv-L4, AAV CLv-L5, AAV CLv-L6, AAV CLv-M1, AAV CLv-M11, AAV CLv-M2, AAV CLv-M5, AAV CLv-M6, AAV CLv-M7, AAV CLv-M8, AAV CLv-M9, AAV CLv-R1, AAV CLv-R2, AAV CLv-R3, AAV CLv-R4, AAV CLv-R5, AAV

CLv-R6, AAV CLv-R7, AAV CLv-R8, AAV CLv-R9, AAV CSp-I, AAV CSp-IO, AAV CSp-I 1, AAV CSp-2, AAV CSp-3, AAV CSp-4, AAV CSp-6, AAV CSp-7, AAV CSp-8, AAV CSp-8.10, AAV CSp-8.2, AAV CSp-8.4, AAV CSp-8.5, AAV CSp-8.6, AAV CSp-8.7, AAV CSp-8.8, AAV CSp-8.9, AAV CSp-9, AAV.hu.48R3, AAV.VR-355, AAV3B, AAV4, AAV5, AAVF1/HSC1, AAVF11/HSC11, AAVF12/HSC12, AAVF13/HSC13, AAVF14/HSC14, AAVF15/HSC15, AAVF16/HSC16, AAVF17/HSC17, AAVF2/HSC2, AAVF3/HSC3, AAVF4/HSC4, AAVF5/HSC5, AAVF6/HSC6, AAVF7/HSC7, AAVF8/HSC8, and/or AAVF9/HSC9 and variants thereof. See e.g., US20210380969, the content of which is incorporated by reference herein in its entirety. In an embodiment, the serotype of the wild-type AAV capsid can be AAV9. The engineered AAV capsids can have a different tropism than that of the reference wild-type AAV capsid.

[0097] In an embodiment, the transferrin receptor binding modifications comprises of modifying the AAV capsid polypeptide. In an embodiment, an engineered adeno associated virus (AAV) capsid polypeptide comprising a transferrin receptor binding modification defined by the formula $X_1-X_2-X_3-[7\text{-mer}]-X_4-X_5-X_6-X_7$, wherein $X_1, X_2, X_3, X_4, X_5, X_6, X_7$ indicate modifications at one or more amino acid positions in the capsid polypeptide flanking the inserted 7-mer (that is at each of $X_1, X_2, X_3, X_4, X_5, X_6, X_7$, the amino acid may be the wild type amino acid at that position or an amino acid substitutions).

[0098] The core of each wild-type AAV viral protein contains an eight-stranded beta-barrel motif (betaB to betaI) and an alpha-helix (alphaA) that are conserved in autonomous parvovirus capsids (see e.g., DiMattia et al. 2012. J. Virol. 86(12):6947-6958). Structural variable regions (VRs), also referred to as “loops”, occur in the surface loops that connect the beta-strands, which cluster to produce local variations in the capsid surface. AAVs have 12 variable regions (also referred to as hypervariable regions) (see e.g., Weitzman and Linden. 2011. “Adeno-Associated Virus Biology.” In Snyder, R.O., Moullier, P. (eds.) Totowa, NJ: Humana Press). In an embodiment, $X_1, X_2, X_3, X_4, X_5, X_6$, and X_7 modify amino acids in one or more of the 12 variable regions in the wild-type AAV capsid proteins. In an embodiment, the $X_1, X_2, X_3, X_4, X_5, X_6$, and X_7 modify amino acids in VR-I, VR-II, VR-III, VR-IV, VR-V, VR-VI, VR-VII, VR-III, VR-IX, VR-X, VR-XI, VR-XII, or a combination thereof. In an example embodiment, the transferrin receptor binding modification is inserted or substituted in loop IV and/or loop VIII.

[0099] In an embodiment, the transferrin receptor binding modification comprises of amino acids 586-588 and 589-592 of a capsid protein of AAV9 (including insertion of the n-mer, such as a 7-mer, between positions 588-589), or in an analogous position of a capsid protein from AAV9, AAV9 K449R (or K449R AAV9), AAV1, AAVrh10, AAV-DJ, AAV-DJ8, AAV5, AAV-PHP.B (PHP.B), AAV-PHP.A (PHP.A), AAVG2B-26, AAVG2B-13, AAVTH1.1-32, AAVTH1.1-35, AAV-PHP.B2 (PHP.B2), AAV-PHP.B3 (PHP.B3), AAV-PHP.N/PHP.B-DGT, AAV-PHP.B-EST, AAV-PHP.B-GGT, AAV-PHP.B-ATP, AAV-PHP.B-ATT-T, AAV-PHP.B-DGT-T, AAV-PHP.B-GGT-T, AAV-PHP.B-SGS, AAV-PHP.B-AQP, AAV-PHP.B-QQP, AAV-PHP.B-SNP(3), AAV-PHP.B-SNP, AAV-PHP.B-QGT, AAV-PHP.B-NQT, AAV-PHP.B-EGS, AAV-PHP.B-SGN, AAV-PHP.B-EGT, AAV-PHP.B-DST, AAV-PHP.B-DST, AAV-PHP.B-STP, AAV-PHP.B-PQP, AAV-PHP.B-SQP, AAV-PHP.B-QLP, AAV-PHP.B-TMP, AAV-PHP.B-TTP, AAV-PHP.S/G2A12, AAVG2A15/G2A3 (G2A3), AAVG2B4 (G2B4), AAVG2B5 (G2B5), PHP.S, AAV2, AAV2G9, AAV3, AAV3a, AAV3b, AAV3-3, AAV4, AAV4-4, AAV6, AAV6.1, AAV6.2, AAV6.1.2, AAV7, AAV7.2, AAV8, AAV9.11, AAV9.13, AAV9.16, AAV9.24, AAV9.45, AAV9.47, AAV9.61, AAV9.68, AAV9.84, AAV9.9, AAV10, AAV11, AAV12, AAV16.3, AAV24.1, AAV27.3, AAV42.12, AAV42-1b, AAV42-2, AAV42-3a, AAV42-3b, AAV42-4, AAV42-5a, AAV42-5b, AAV42-6b, AAV42-8, AAV42-10, AAV42-11, AAV42-12, AAV42-13, AAV42-15, AAV42-aa, AAV43-1, AAV43-12, AAV43-20, AAV43-21, AAV43-23, AAV43-25, AAV43-5, AAV44.1, AAV44.2, AAV44.5, AAV223.1, AAV223.2, AAV223.4, AAV223.5, AAV223.6, AAV223.7, AAV1-7/rh.48, AAV1-8/rh.49, AAV2-15/rh.62, AAV2-3/rh.61, AAV2-4/rh.50, AAV2-5/rh.51, AAV3.1/hu.6, AAV3.1/hu.9, AAV3-9/rh.52, AAV3-11/rh.53, AAV4-8/rh.1.64, AAV4-9/rh.54, AAV4-19/rh.55, AAV5-3/rh.57, AAV5-22/rh.58, AAV7.3/hu.7, AAV16.8/hu.10, AAV16.12/hu.11, AAV29.3/bb.1, AAV29.5/bb.2, AAV106.1/hu.37, AAV114.3/hu.40, AAV127.2/hu.41, AAV127.5/hu.42, AAV128.3/hu.44, AAV130.4/hu.48, AAV145.1/hu.53, AAV145.5/hu.54, AAV145.6/hu.55, AAV161.10/hu.60, AAV161.6/hu.61, AAV33.12/hu.17, AAV33.4/hu.15, AAV33.8/hu.16, AAV52/hu.19, AAV52.1/hu.20, AAV58.2/hu.25, AAVA3.3, AAVA3.4, AAVA3.5, AAVA3.7, AAVC1, AAVC2, AAVC5, AAVF3, AAVF5, AAVH2, AAVrh.72, AAVhu.8, AAVrh.68, AAVrh.70, AAVpi.1, AAVpi.3, AAVpi.2, AAVrh.60, AAVrh.44, AAVrh.65, AAVrh.55, AAVrh.47, AAVrh.69, AAVrh.45, AAVrh.59, AAVhu.12, AAVH6, AAVH-1/hu.1, AAVH-5/hu.3, AAVLG-10/rh.40, AAVLG-4/rh.38, AAVLG-9/hu.39, AAVN721-8/rh.43, AAVCh.5, AAVCh.5RI, AAVcy.2,

AAVcy.3, AAVcy.4, AAVcy.5, AAVCy.5R1, AAVCy.5R2, AAVCy.5R3, AAVCy.5R4, AAVcy.6, AAVhu.1, AAVhu.2, AAVhu.3, AAVhu.4, AAVhu.5, AAVhu.6, AAVhu.7, AAVhu.9, AAVhu.10, AAVhu.11, AAVhu.13, AAVhu.15, AAVhu.16, AAVhu.17, AAVhu.18, AAVhu.20, AAVhu.21, AAVhu.22, AAVhu.23.2, AAVhu.24, AAVhu.25, AAVhu.27, AAVhu.28, AAVhu.29, AAVhu.29R, AAVhu.31, AAVhu.32, AAVhu.34, AAVhu.35, AAVhu.37, AAVhu.39, AAVhu.40, AAVhu.41, AAVhu.42, AAVhu.43, AAVhu.44, AAVhu.44R1, AAVhu.44R2, AAVhu.44R3, AAVhu.45, AAVhu.46, AAVhu.47, AAVhu.48, AAVhu.48R1, AAVhu.48R2, AAVhu.48R3, AAVhu.49, AAVhu.51, AAVhu.52, AAVhu.54, AAVhu.55, AAVhu.56, AAVhu.57, AAVhu.58, AAVhu.60, AAVhu.61, AAVhu.63, AAVhu.64, AAVhu.66, AAVhu.67, AAVhu.14/9, AAVhu.t19, AAVrh.2, AAVrh.2R, AAVrh.8, AAVrh.8R, AAVrh.10, AAVrh.12, AAVrh.13, AAVrh.13R, AAVrh.14, AAVrh.17, AAVrh.18, AAVrh.19, AAVrh.20, AAVrh.21, AAVrh.22, AAVrh.23, AAVrh.24, AAVrh.25, AAVrh.31, AAVrh.32, AAVrh.33, AAVrh.34, AAVrh.35, AAVrh.36, AAVrh.37, AAVrh.37R2, AAVrh.38, AAVrh.39, AAVrh.40, AAVrh.46, AAVrh.48, AAVrh.48.1, AAVrh.48.L2, AAVrh.48.2, AAVrh.49, AAVA.51, AAVrh.52, AAVrh.53, AAVrh.54, AAVrh.56, AAVrh.57, AAVrh.58, AAVA.61, AAVrh.64, AAVA.64R1, AAVA.64R2, AAVrh.67, AAVrh.73, AAVrh.74, AAVA8R, AAVA8R A586R mutant, AAVA8R R533A mutant, AAV, BAAV, caprine AAV, bovine AAV, AAVhEl.1, AAVhEr1.5, AAVhEr1.14, AAVhEr1.8, AAVhEr1.16, AAVhEr1.18, AAVhEr1.35, AAVhEr1.7, AAVhEr1.36, AAVhEr2.29, AAVhEr2.4, AAVhEr2.16, AAVhEr2.30, AAVhEr2.31, AAVhEr2.36, AAVhEr1.23, AAVhEr3.1, AAV2.5T, AAV-PAEC, AAV-LK01, AAV-LK02, AAV-LK03, AAV-LK04, AAV-LK05, AAV-LK06, AAV-LK07, AAV-LK08, AAV-LK09, AAV-LK10, AAV-LK11, AAV-LK12, AAV-LK13, AAV-LK14, AAV-LK15, AAV-LK16, AAV-LK17, AAV-LK18, AAV-LK19, AAV-PAEC2, AAV-PAEC4, AAV-PAEC6, AAV-PAEC7, AAV-PAEC8, AAV-PAEC11, AAV-PAEC12, AAV-2-pre-miRNA-101, AAV-8h, AAV-8b, AAV-h, AAV-b, AAV SM 10-2, AAV Shuffle 100-1, AAV Shuffle 100-3, AAV Shuffle 100-7, AAV Shuffle 10-2, AAV Shuffle 10-6, AAV Shuffle 10-8, AAV Shuffle 100-2, AAV SM 10-1, AAV SM 10-8, AAV SM 100-3, AAV SM 100-10, BNP61 AAV, BNP62 AAV, BNP63 AAV, AAVrh.50, AAVrh.43, AAVrh.62, AAVrh.48, AAVhu.19, AAVhu.11, AAVhu.53, AAV4-8/rh.64, AAVLG-9/hu.39, AAV54.5/hu.23, AAV54.2/hu.22, AAV54.7/hu.24, AAV54.1/hu.21, AAV54.4R/hu.27, AAV46.2/hu.28, AAV46.6/hu.29, AAV128.1/hu.43, true type AAV (ttAAV), UPEN AAV 10, Japanese AAV 10 serotypes, AAV CBr-7.1, AAV CBr-7.10, AAV

CBBr-7.2, AAV CBr-7.3, AAV CBr-7.4, AAV CBr-7.5, AAV CBr-7.7, AAV CBr-7.8, AAV CBr-B7.3, AAV CBr-B7.4, AAV CBr-E1, AAV CBr-E2, AAV CBr-E3, AAV CBr-E4, AAV CBr-E5, AAV CBr-e5, AAV CBr-E6, AAV CBr-E7, AAV CBr-E8, AAV CHt-1, AAV CHt-2, AAV CHt-3, AAV CHt-6.1, AAV CHt-6.10, AAV CHt-6.5, AAV CHt-6.6, AAV CHt-6.7, AAV CHt-6.8, AAV CHt-P1, AAV CHt-P2, AAV CHt-P5, AAV CHt-P6, AAV CHt-P8, AAV CHt-P9, AAV CKd-1, AAV CKd-10, AAV CKd-2, AAV CKd-3, AAV CKd-4, AAV CKd-6, AAV CKd-7, AAV CKd-8, AAV CKd-B1, AAV CKd-B2, AAV CKd-B3, AAV CKd-B4, AAV CKd-B5, AAV CKd-B6, AAV CKd-B7, AAV CKd-B8, AAV CKd-H1, AAV CKd-H2, AAV CKd-H3, AAV CKd-H4, AAV CKd-H5, AAV CKd-H6, AAV CKd-N3, AAV CKd-N4, AAV CKd-N9, AAV CLg-F1, AAV CLg-F2, AAV CLg-F3, AAV CLg-F4, AAV CLg-F5, AAV CLg-F6, AAV CLg-F7, AAV CLg-F8, AAV CLv-1, AAV CLvl-1, AAV Clvl-10, AAV CLvl-2, AAV CLv-12, AAV CLvl-3, AAV CLv-13, AAV CLvl-4, AAV Clvl-7, AAV Clvl-8, AAV Clvl-9, AAV CLv-2, AAV CLv-3, AAV CLv-4, AAV CLv-6, AAV CLv-8, AAV CLv-D1, AAV CLv-D2, AAV CLv-D3, AAV CLv-D4, AAV CLv-D5, AAV CLv-D6, AAV CLv-D7, AAV CLv-D8, AAV CLv-E1, AAV CLv-K1, AAV CLv-K3, AAV CLv-K6, AAV CLv-L4, AAV CLv-L5, AAV CLv-L6, AAV CLv-M1, AAV CLv-M11, AAV CLv-M2, AAV CLv-M5, AAV CLv-M6, AAV CLv-M7, AAV CLv-M8, AAV CLv-M9, AAV CLv-R1, AAV CLv-R2, AAV CLv-R3, AAV CLv-R4, AAV CLv-R5, AAV CLv-R6, AAV CLv-R7, AAV CLv-R8, AAV CLv-R9, AAV CSp-1, AAV CSp-10, AAV CSp-11, AAV CSp-2, AAV CSp-3, AAV CSp-4, AAV CSp-6, AAV CSp-7, AAV CSp-8, AAV CSp-8.10, AAV CSp-8.2, AAV CSp-8.4, AAV CSp-8.5, AAV CSp-8.6, AAV CSp-8.7, AAV CSp-8.8, AAV CSp-8.9, AAV CSp-9, AAV.hu.48R3, AAV.VR-355, AAV3B, AAV4, AAV5, AAVF1/HSC1, AAVF11/HSC11, AAVF12/HSC12, AAVF13/HSC13, AAVF14/HSC14, AAVF15/HSC15, AAVF16/HSC16, AAVF17/HSC17, AAVF2/HSC2, AAVF3/HSC3, AAVF4/HSC4, AAVF5/HSC5, AAVF6/HSC6, AAVF7/HSC7, AAVF8/HSC8, and/or AAVF9/HSC9 and variants thereof. In an embodiment, X1-X2-X3 comprises of any sequential amino acids 449-459 and X4-X5-X6-X7 comprises of any amino sequential amino acids 452-463 of a capsid protein of AAV9, or in an analogous position of a capsid protein from AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAVrh8, AAV rh.74, or AAVrh.10.

[0100] In an embodiment, the transferrin receptor binding modification is incorporated into a viral protein, such as a capsid protein, including but not limited to adenoviral or AAV proteins. In an embodiment, the n-mer is located between two amino acids of the viral protein such that the

transferrin receptor binding modification is external (i.e., is presented on the surface of) to a viral capsid. In an example embodiment, the n-mer disclosed herein can be inserted between two consecutive amino acids in the wild-type viral protein (VP) (or capsid protein), including in regions that are surface exposed when incorporated into a viral capsid. In an embodiment, the n-mer can be inserted between two consecutive amino acids in a variable amino acid region in a viral capsid protein.

[0101] In an embodiment, the n-mer can be inserted between two consecutive amino acids in a variable amino acid region in an AAV capsid protein. In an embodiment, one or more n-mer can be inserted between two amino acids in one or more of the 12 variable regions in the wild-type AAV capsid proteins. In an embodiment, the one or more n-mers can be each be inserted between two amino acids in VR-I, VR-II, VR-III, VR-IV, VR-V, VR-VI, VR-VII, VR-III, VR-IX, VR-X, VR-XI, VR-XII, or a combination thereof. In an example embodiment, the transferrin receptor binding modification is inserted or substituted in loop IV and/or loop VIII. In an example embodiment, the n-mer or 7-mer is YSRIGPN (SEQ ID NO: 1390), YSRNSDN (SEQ ID NO: 1391), LHRLGPN (SEQ ID NO: 1392), FRSTNGV (SEQ ID NO: 1393), FVSTNGV (SEQ ID NO: 1394), FZ₁STNGZ₂ (SEQ ID NO: 1395), FRSTNGZ₃ (SEQ ID NO: 1396), and VESTNGR (SEQ ID NO: 1397). In an embodiment, Z₁ is selected from the group consisting of A, D, H, N, Q, and S; Z₂ is selected from the group consisting of K and R; and Z₃ is selected from the group consisting of L, M, and R.

[0102] In an embodiment, the engineered capsid is a modified AAV1 capsid and can have a n-mer motif inserted after or a neighbor of amino acid 590 (i.e., between amino acid 590 and 591). In an embodiment, the engineered capsid is a modified AAV3 capsid and can have a n-mer motif inserted after or a neighbor of amino acid 586. In an embodiment, the engineered capsid is a modified AAV4 capsid and can have a n-mer motif inserted after or a neighbor of amino acid 586. In an embodiment, the engineered capsid is a modified AAV5 capsid and can have a n-mer motif inserted after or a neighbor of amino acid 575. In an embodiment, the engineered capsid is a modified AAV6 capsid and can have a n-mer inserted at or a neighbor of amino acid 585 and optionally Y705-731, T492V, K531E. In an embodiment, the engineered capsid is a modified AAV8 capsid and can have a n-mer inserted after or a neighbor of amino acid 585 and 590. In an embodiment, the engineered capsid is a modified AAV9 capsid and can have a n-mer inserted in between amino acid 588 and 589. (Büning, H.; Srivastava, A. Capsid Modifications for Targeting

and Improving the Efficacy of AAV Vectors. *Molecular Therapy - Methods & Clinical Development* 2019, 12, 248–265). In an embodiment, the engineered capsid can have a 7-mer motif inserted between amino acids 588 and 589 of an AAV9 viral protein. SEQ ID NO: 1 is a reference AAV9 capsid sequence for at least referencing the insertion sites discussed above. In an embodiment, the engineered capsid can have a 7-mer motif inserted between two consecutive amino acids within amino acids 451-460 of a capsid protein of AAV9 viral protein. SEQ ID NO: 1 is a reference AAV9 capsid sequence for at least referencing the insertion sites discussed above.

[0103] It will be appreciated that n-mers can be inserted in analogous positions in AAV viral proteins of other wild-type serotypes or engineered capsid variants, such as but not limited to, AAV9, AAV9 K449R (or K449R AAV9), AAV1, AAVrh10, AAV-DJ, AAV-DJ8, AAV5, AAV-PHP.B (PHP.B), AAV-PHP.A (PHP.A), AAVG2B-26, AAVG2B-13, AAVTH1.1-32, AAVTH1.1-35, AAV-PHP.B2 (PHP.B2), AAV-PHP.B3 (PHP.B3), AAV-PHP.N/PHP.B-DGT, AAV-PHP.B-EST, AAV-PHP.B-GGT, AAV-PHP.B-ATP, AAV-PHP.B-ATT-T, AAV-PHP.B-DGT-T, AAV-PHP.B-GGT-T, AAV-PHP.B-SGS, AAV-PHP.B-AQP, AAV-PHP.B-QQP, AAV-PHP.B-SNP(3), AAV-PHP.B-SNP, AAV-PHP.B-QGT, AAV-PHP.B-NQT, AAV-PHP.B-EGS, AAV-PHP.B-SGN, AAV-PHP.B-EGT, AAV-PHP.B-DST, AAV-PHP.B-DST, AAV-PHP.B-STP, AAV-PHP.B-PQP, AAV-PHP.B-SQP, AAV-PHP.B-QLP, AAV-PHP.B-TMP, AAV-PHP.B-TTP, AAV-PHP.S/G2A12, AAVG2A15/G2A3 (G2A3), AAVG2B4 (G2B4), AAVG2B5 (G2B5), PHP.S, AAV2, AAV2G9, AAV3, AAV3a, AAV3b, AAV3-3, AAV4, AAV4-4, AAV6, AAV6.1, AAV6.2, AAV6.1.2, AAV7, AAV7.2, AAV8, AAV9.11, AAV9.13, AAV9.16, AAV9.24, AAV9.45, AAV9.47, AAV9.61, AAV9.68, AAV9.84, AAV9.9, AAV10, AAV11, AAV12, AAV16.3, AAV24.1, AAV27.3, AAV42.12, AAV42-1b, AAV42-2, AAV42-3a, AAV42-3b, AAV42-4, AAV42-5a, AAV42-5b, AAV42-6b, AAV42-8, AAV42-10, AAV42-11, AAV42-12, AAV42-13, AAV42-15, AAV42-aa, AAV43-1, AAV43-12, AAV43-20, AAV43-21, AAV43-23, AAV43-25, AAV43-5, AAV44.1, AAV44.2, AAV44.5, AAV223.1, AAV223.2, AAV223.4, AAV223.5, AAV223.6, AAV223.7, AAV1-7/rh.48, AAV1-8/rh.49, AAV2-15/rh.62, AAV2-3/rh.61, AAV2-4/rh.50, AAV2-5/rh.51, AAV3.1/hu.6, AAV3.1/hu.9, AAV3-9/rh.52, AAV3-11/rh.53, AAV4-8/rh.1.64, AAV4-9/rh.54, AAV4-19/rh.55, AAV5-3/rh.57, AAV5-22/rh.58, AAV7.3/hu.7, AAV16.8/hu.10, AAV16.12/hu.11, AAV29.3/bb.1, AAV29.5/bb.2, AAV106.1/hu.37, AAV114.3/hu.40, AAV127.2/hu.41, AAV127.5/hu.42, AAV128.3/hu.44, AAV130.4/hu.48, AAV145.1/hu.53, AAV145.5/hu.54, AAV145.6/hu.55, AAV161.10/hu.60,

AAV161.6/hu.61, AAV33.12/hu.17, AAV33.4/hu.15, AAV33.8/hu.16, AAV52/hu.19, AAV52.1/hu.20, AAV58.2/hu.25, AAVA3.3, AAVA3.4, AAVA3.5, AAVA3.7, AAVC1, AAVC2, AAVC5, AAVF3, AAVF5, AAVH2, AAVrh.72, AAVhu.8, AAVrh.68, AAVrh.70, AAVpi.1, AAVpi.3, AAVpi.2, AAVrh.60, AAVrh.44, AAVrh.65, AAVrh.55, AAVrh.47, AAVrh.69, AAVrh.45, AAVrh.59, AAVhu.12, AAVH6, AAVH-1/hu.1, AAVH-5/hu.3, AAVLG-10/rh.40, AAVLG-4/rh.38, AAVLG-9/hu.39, AAVN721-8/rh.43, AAVCh.5, AAVCh.5R1, AAVcy.2, AAVcy.3, AAVcy.4, AAVcy.5, AAVCy.5R1, AAVCy.5R2, AAVCy.5R3, AAVCy.5R4, AAVcy.6, AAVhu.1, AAVhu.2, AAVhu.3, AAVhu.4, AAVhu.5, AAVhu.6, AAVhu.7, AAVhu.9, AAVhu.10, AAVhu.11, AAVhu.13, AAVhu.15, AAVhu.16, AAVhu.17, AAVhu.18, AAVhu.20, AAVhu.21, AAVhu.22, AAVhu.23.2, AAVhu.24, AAVhu.25, AAVhu.27, AAVhu.28, AAVhu.29, AAVhu.29R, AAVhu.31, AAVhu.32, AAVhu.34, AAVhu.35, AAVhu.37, AAVhu.39, AAVhu.40, AAVhu.41, AAVhu.42, AAVhu.43, AAVhu.44, AAVhu.44R1, AAVhu.44R2, AAVhu.44R3, AAVhu.45, AAVhu.46, AAVhu.47, AAVhu.48, AAVhu.48R1, AAVhu.48R2, AAVhu.48R3, AAVhu.49, AAVhu.51, AAVhu.52, AAVhu.54, AAVhu.55, AAVhu.56, AAVhu.57, AAVhu.58, AAVhu.60, AAVhu.61, AAVhu.63, AAVhu.64, AAVhu.66, AAVhu.67, AAVhu.14/9, AAVhu.19, AAVrh.2, AAVrh.2R, AAVrh.8, AAVrh.8R, AAVrh.10, AAVrh.12, AAVrh.13, AAVrh.13R, AAVrh.14, AAVrh.17, AAVrh.18, AAVrh.19, AAVrh.20, AAVrh.21, AAVrh.22, AAVrh.23, AAVrh.24, AAVrh.25, AAVrh.31, AAVrh.32, AAVrh.33, AAVrh.34, AAVrh.35, AAVrh.36, AAVrh.37, AAVrh.37R2, AAVrh.38, AAVrh.39, AAVrh.40, AAVrh.46, AAVrh.48, AAVrh.48.1, AAVrh.48.1.2, AAVrh.48.2, AAVrh.49, AAVA.51, AAVrh.52, AAVrh.53, AAVrh.54, AAVrh.56, AAVrh.57, AAVrh.58, AAVA.61, AAVrh.64, AAVA.64R1, AAVA.64R2, AAVrh.67, AAVrh.73, AAVrh.74, AAVA8R, AAVA8R A586R mutant, AAVA8R R533A mutant, AAV, BAAV, caprine AAV, bovine AAV, AAVhEl.1, AAVhEl.5, AAVhEl.14, AAVhEl.8, AAVhEl.16, AAVhEl.18, AAVhEl.35, AAVhEl.7, AAVhEl.36, AAVhEr2.29, AAVhEr2.4, AAVhEr2.16, AAVhEr2.30, AAVhEr2.31, AAVhEr2.36, AAVhEl.23, AAVhEr3.1, AAV2.5T, AAV-PAEC, AAV-LK01, AAV-LK02, AAV-LK03, AAV-LK04, AAV-LK05, AAV-LK06, AAV-LK07, AAV-LK08, AAV-LK09, AAV-LK10, AAV-LK11, AAV-LK12, AAV-LK13, AAV-LK14, AAV-LK15, AAV-LK16, AAV-LK17, AAV-LK18, AAV-LK19, AAV-PAEC2, AAV-PAEC4, AAV-PAEC6, AAV-PAEC7, AAV-PAEC8, AAV-PAEC11, AAV-PAEC12, AAV-2-pre-miRNA-101, AAV-8h, AAV-8b, AAV-h, AAV-b, AAV SM 10-2, AAV Shuffle 100-1, AAV Shuffle 100-3, AAV

Shuffle 100-7, AAV Shuffle 10-2, AAV Shuffle 10-6, AAV Shuffle 10-8, AAV Shuffle 100- 2, AAV SM 10-1, AAV SM 10-8 , AAV SM 100-3, AAV SM 100-10, BNP61 AAV, BNP62 AAV, BNP63 AAV, AAVrh.50, AAVrh.43, AAVrh.62, AAVrh.48, AAVhu. 19, AAVhu. 11, AAVhu.53, AAV4-8/rh.64, AAVLG-9/hu.39, AAV54.5/hu.23, AAV54.2/hu.22, AAV54.7/hu.24, AAV54.1/hu.21, AAV54.4R/hu.27, AAV46.2/hu.28, AAV46.6/hu.29, AAV128. 1/hu.43, true type AAV (ttAAV), UPEN AAV 10, Japanese AAV 10 serotypes, AAV CBr-7. 1, AAV CBr-7. 10, AAV CBr-7.2, AAV CBr-7.3, AAV CBr-7.4, AAV CBr-7.5, AAV CBr-7.7, AAV CBr-7.8, AAV CBr-B7.3, AAV CBr-B7.4, AAV CBr-E1, AAV CBr- E2, AAV CBr-E3, AAV CBr-E4, AAV CBr-E5, AAV CBr-e5, AAV CBr-E6, AAV CBr-E7, AAV CBr-E8, AAV CHt-1, AAV CHt-2, AAV CHt-3, AAV CHt-6. 1, AAV CHt-6. 10, AAV CHt-6.5, AAV CHt-6.6, AAV CHt-6.7, AAV CHt-6.8, AAV CHt-P1, AAV CHt-P2, AAV CHt-P5, AAV CHt-P6, AAV CHt-P8, AAV CHt-P9, AAV CKd-1, AAV CKd-10, AAV CKd-2, AAV CKd-3, AAV CKd-4, AAV CKd-6, AAV CKd-7, AAV CKd-8, AAV CKd-B1, AAV CKd-B2, AAV CKd-B3, AAV CKd-B4, AAV CKd-B5, AAV CKd-B6, AAV CKd-B7, AAV CKd-B8, AAV CKd-H1, AAV CKd-H2, AAV CKd-H3, AAV CKd-H4, AAV CKd- H5, AAV CKd-H6, AAV CKd-N3, AAV CKd-N4, AAV CKd-N9, AAV CLg-F1, AAV CLg-F2, AAV CLg-F3, AAV CLg-F4, AAV CLg-F5, AAV CLg-F6, AAV CLg-F7, AAV CLg-F8, AAV CLv-1, AAV CLvl-1, AAV Clvl-10, AAV CLvl-2, AAV CLv-12, AAV CLvl-3, AAV CLv-1 3, AAV CLvl-4, AAV Clvl-7, AAV Clvl-8, AAV Clvl-9, AAV CLv- 2, AAV CLv-3, AAV CLv-4, AAV CLv-6, AAV CLv-8, AAV CLv-D1, AAV CLv-D2, AAV CLv-D3, AAV CLv-D4, AAV CLv-D5, AAV CLv-D6, AAV CLv-D7, AAV CLv-D8, AAV CLv-E1, AAV CLv-K1, AAV CLv-K3, AAV CLv-K6, AAV CLv-L4, AAV CLv-L5, AAV CLv-L6, AAV CLv-M1, AAV CLv-M1 1, AAV CLv-M2, AAV CLv-M5, AAV CLv- M6, AAV CLv-M7, AAV CLv-M8, AAV CLv-M9, AAV CLv-R1, AAV CLv-R2, AAV CLv-R3, AAV CLv-R4, AAV CLv-R5, AAV CLv-R6, AAV CLv-R7, AAV CLv-R8, AAV CLv-R9, AAV CSp-1, AAV CSp-10, AAV CSp-1 1, AAV CSp-2, AAV CSp-3, AAV CSp-4, AAV CSp-6, AAV CSp-7, AAV CSp-8, AAV CSp-8. 10, AAV CSp-8.2, AAV CSp-8.4, AAV CSp-8.5, AAV CSp-8.6, AAV CSp-8.7, AAV CSp-8.8, AAV CSp-8.9, AAV CSp-9, AAV.hu.48R3, AAV.VR-355, AAV3B, AAV4, AAV5, AAVF1/HSC1, AAVF11/HSC11, AAVF12/HSC12, AAVF13/HSC13, AAVF14/HSC14, AAVF15/HSC15, AAVF16/HSC16, AAVF17/HSC17, AAVF2/HSC2, AAVF3/HSC3, AAVF4/HSC4, AAVF5/HSC5, AAVF6/HSC6, AAVF7/HSC7, AAVF8/HSC8, and/or AAVF9/HSC9 and variants thereof. See e.g., US20210380969, the content of which is incorporated by reference

herein in its entirety. In an embodiment as previously discussed, the n-mer(s) can be inserted between any two contiguous amino acids within the AAV viral protein and in an embodiment the insertion is made in a variable region.

[0104] In an embodiment, the first 1, 2, 3, or 4 amino acids of a transferrin receptor binding modification can replace 1, 2, 3, or 4 amino acids of a polypeptide into which it is inserted and preceding the insertion site. Using an AAV as another non-limiting example, one or more of the n-mers can be inserted into e.g., an AAV9 capsid polypeptide between amino acids 588 and 589 and the insert can replace amino acids 586, 587, and 588 such that the amino acid immediately preceding the transferrin receptor binding modification after insertion is residue 585. It will be appreciated that this principle can apply in any other insertion context and is not necessarily limited to insertion between residues 588 and 589 of an AAV9 capsid or equivalent position in another AAV capsid. In an example embodiment, the AAV capsid protein is selected from SEQ ID NO: 1.

Table B. Example AAV VP1 amino acid sequences

Serotype	SEQ ID NO:	Accession number	VP1 AMINO ACID SEQUENCE
AAV1	3	NP_049542.1	MAADGYLPDWLEDNLSGEGIREWWDLKPGAPKQKQDDGRGLVLPGYKYLGPFGNGLDKG EPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEEFERLQEDTSFGGNLGRAVFQAKKR VLEPLGLVEEAKTAPGKKRPVEQSPQEPDSSSGIGKTQQQPAKKRLNFGQTGDSSEVPDPQ PLGEPAPATPAAVGPTTMASSGGAPMADNNEGADVGNASGNWHCDSTWLGDRVITTTSTRTWA LPT'YNNHLYKQISSASTGASNDNHFGYSTPWGYFDFNRPHCHFSPRDWQRQLINNNWGFPRPK RLNFKLFNIQVKEVT'TNDGVTTIANNLSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFM I PQGYLTLLNNGSQAVGRSSFYCLEYFPFSQMLRTGNFTFSYTFEEVPFHHSSYAHSSQLDRL MNP LI DQYLYYLNR TQNGSGSAQNKDLLFSRGS P AGMSVQPKNWLP GP CYRQQRVSKTKTDN NNSNFTWTGASKYNLNGRESIINPGTAMASHKDDDEKFFPMSGVMI FGKESAGASNTALDNV MITDEEEIKATNPVATERFRTVAVNFQSSSTDPATGDVHAMGALPGMWQDRDVYLQGP IWA KI PHTDGHFHPSP LMGGFGLKNPPQILIKNTPVPANPAEFSATKEASFITQYSTGQVSVE IEWELQKENS KRWNPEVQYTSNYAKSANVDFTVDNNGLYTEPRPIGTRYLTRNL
			MAADGYLPDWLEDTLSEGI RQWVKLKPGPPPKPAERHKDDSRGLVLPGYKYLGPFGNGLDKG EPVNEADAAALEHDKAYDRQLDSGDNPYLYKNHADAEEFERLKEDTSFGGNLGRAVFQAKKR VLEPLGLVEEPVKTAPGKKRPVEHSPVEPDSSSGTGKAGQQPARKRLNFGQTGDADSVDPDQ PLGQPPAAPSGLGNTMATGSGAPMADNNEGADVGNSSGNWHCDSTWMDRVI TTTSTRTWA LPT'YNNHLYKQISSQSGASNDNHFGYSTPWGYFDFNRPHCHFSPRDWQRQLINNNWGFPRKR LNFKLFNIQVKEVTQNDGTTTIANNLSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMV PQGYLTLLNNGSQAVGRSSFYCLEYFPFSQMLRTGNFTFSYTFEDVPFHSSYAHSSQLDRLM NPLIDQYLYLSRTNTPSGTTTQSR LQFSQAGASDIRQSRNWLPGPCYRQQRVSKTSADNN NSEYSWTGATKYHLNGRDSLVP GPAMASHKDDDEKFFPQSGVLI FGKQSGSEKTNVDIEKVM ITDEEEIRTTNPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMVWQDRDVYLQGP IWA KI PHTDGHFHPSP LMGGFGLKHP PPQILIKNTPVPANPSTTFSAAKPFASFITQYSTGQVSVEI IEWELQKENS KRWNPEIQYTSNYKSNVNVDTVDNTNGVYSEPRPIGTRYLTRNL
AAV3	5	NP_043941.1	MAADGYLPDWLEDNLSGEGIREWWDLKPGVPQPKANQHQDNRRGLVLPGYKYLGPFGNGLDKG EPVNEADAAALEHDKAYDQQLKAGDNPYLRYNHADAEEFERLQEDTSFGGNLGRAVFQAKKR I LEP LGLVEEAAKTAPGKKGAVDQSPQEPDSSSGVKGSKQKQPAKRRLNFGQTGDSSEVPDPQ PLGEPAPATPSLGSNTWASGGGAPMADNNEGADVGNSSGNWHCDSQWLGDRI VITTTSTRTWA LPT'YNNHLYKQISSQSGASNDNHFGYSTPWGYFDFNRPHCHFSPRDWQRQLINNNWGFPRPK LSFKLFNIQVRGVTONDGT'TTIANNLSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMV PQGYLTLLNNGSQAVGRSSFYCLEYFPFSQMLRTGNNFQFSYTFEDVPFHSSYAHSSQLDRLM NPLIDQYLYYLNR TQGTTSRLLFSQAGPQMSLQARNWLP GP CYRQQRVSKTKTANDN NNSNFPWTAASKYHLNGRDSLVP GPAMASHKDDDEKFFPMHGNLIFGKEGTTASNAELDNV MITDEEEIRTTNPVATEQYGTVANNLQSSNTAPTGTGVNHQGALPGMWQDRDVYLQGP IWA KI PHTDGHFHPSP LMGGFGLKHP PPQIMIKNTPVPANPPTTFSAPKPFASFITQYSTGQVSVE IEWELQKENS KRWNPEIQYTSNYKSNVNVDTVDNTNGVYSEPRPIGTRYLTRNL
			MAADGYLPDWLEDNLSGEGIREWWDLKPGVPOPKANQHQDNRRGLVLPGYKYLGPFGNGLDKG EPVNEADAAALEHDKAYDQQLKAGDNPYLRYNHADAEEFERLQEDTSFGGNLGRAVFQAKKR
AAV3B	6	3KIC_A	MAADGYLPDWLEDNLSGEGIREWWDLKPGVPOPKANQHQDNRRGLVLPGYKYLGPFGNGLDKG EPVNEADAAALEHDKAYDQQLKAGDNPYLRYNHADAEEFERLQEDTSFGGNLGRAVFQAKKR

Serotype	SEQ ID NO:	Accession number	VP1 AMINO ACID SEQUENCE
AAV4	7	NP_044927.1	ILEPLGLVEEAAKTAPGKRPRVDQSPQEPDSSGVGSKQPKARKRLNFGQTGDSSEVPDPQ PLGEPAAPTSLGNSNTMASGGGAPNADNNEGADGVGNSGNWHCDSQWLGDRLVITTTSTRTWA LPTYNHLYKQISSQSGASNDNHYGYSTPWGYFDENFRHCHFSRDPDQRLINNNWGFRPKK LSFKLFNIQKREVTQNDGTTTIANNLTSTVQVTDSEQLPYVLGSAHQGCLPPFPADVFMV PQGYLTNLNGSQAAGRSSFYCLEYFSPQMLRTGNFQFSYTFEDVPFHSSYAHSQSLDRLM NPLIDQLYLYLNRQTGTTSGTTNQSRLLFSQAQPQSMISQARNWLPGPCYRQORLSKTANDN NNSNFPWTAASKYHLNLRDSLVPNPGFAMASHKDEEKFFPMHGNLIFGKEGTASNAELDNV MITDEEEIRITNPNVATEQYGTVANNLQSSNTAPTTRTVNDQALPGMVWQDRDVLQGIWA KI PHTDGHFHPSPLMGGFGLKHPPIQIMIKNTVPANPTTTFSPAKFASFITQYSTGQVSVE IEWELQKENSKRWNPEIQYTSNYSKSVNVDFTVDTNGVYSEPRPIGTRYLTRNL
			MTDGYLPDWLEDNLSEGVREWWALQPGAPKPKANQQHQDNARGLVLPGYKILGPGNGLDKGE PVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADAEFQORLQGDTSFGGNLGRAVFQAKKRV LEPLGLVEQAGETAPGKRRLIESPQQPDSSTGIGKKQKQPAKKLIVFEDETGAAGDGPPEG TSGAMSDDESEMPAAAGAAVEGGQADGVGNASGDWHCDSTWSEGHVTTTSTRTWVLPYNN HLYKRLGESLQSNTYNGFSTPWGYFDENFRHCHFSRDPDQRLINNNWMPKAMRVKIFNIQ VKEVTSNGETVANNLSTVQIFADSSYELPYVMDAQEGSLPPFPNDVFMVPQYGYCGLV TGNTSQQQTDRNAFYCLEYFSPQMLRTGNFEITYSFEKVPFHSSYAHSQSLDRLMPLIDQ YLWGLQSTTTGTLNAGTATTNFKLRPTNFSNFKKNWLPGPSIKQQGFSKTANQNYKIPAT GSDSLIKYETHSTLDGRWSALTPGPPMATAGPADSKFSNSQLIFAGPKQNGTATVPGLIF TSEELAAATNATDTDMGNLPGGQDSNSNLPTVDRLTALGAVPGMVWQNRDIYYQGIWAKI PHTDGHFHPSPLMGGFGLKHPPIQIFIKNTVPANPATFSSTPVNSFITQYSTGQVSQID WEIQKERSKRWNPVQFTSNYGOQNSLLWAPDAAGKYTEPRRAIGTRYLTHHL
AAV5	8	YP_068409.1	MSFYDHPDWLEEVGEGLEFLGLEAGPPKPKPNQQHQDQARGLVLPGYNYLGPNGLDRGE PVNRADDEVAREHDI SYNEQLEAGDNPYLKYNHADAEFQKLAADDTSFGGNLGAVFQAKKRV LEPGLVEEGAKTAPTGRKIDHFPKRRKARTEEDSKFTSSDAEAGPSGQQLQIPAPAS SLGADTMSAGGGGGLGDNQAGADGVGNASGDWHCDSTWMDRVVTKSTRTWLPSYNNHOYR EIKSGSVDGSNANAYFGYSTPWGYFDENFRHSHWSRDPDQRLINNNWGFRRPSLRVKIFNIQ VKEVTVDSTTTIANNLSTVQVFTDDDYQLPYVVGNGTEGCLPAPPQVFTLPQYGYATLN RDNTENPTERSFFCLEYFPSKMLRTGNFEITYNFEYVPHSSFAPSQNLFKIANPLVDQY LYRFVSTNNTGGVQFNKMLAGRYANTYKNWFFGPMGRQGWNLGSGVNRASVSATATNRME LEGASYQVPPQPNGMTNNLQGSNTYALENTMIFNSQPANPGTTATYLEGNMLITSESETQPV NRVAYNVGGQMATNNQSTTAPATGTINLQEIIVPGSVMMERDVYLQGGPIWAKIPETGAHFHP SPAMGGFGLKHPPPMMLIKNTVPNGNITSFSDVPVSSFITQYSTGQVTEMWEWELKKENSKR WNPEIQYTNNNYNDPQFVDFAPDSTGEYRTRTTRIGTRYLTRPL
			MAADGYLPDWLEDNLSEGIREWWDILKPGAPKPKANQQDQDGRGLVLPGYKILGPFNGLDKG EPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTSFGGNLGRAVFQAKK VLEPGLVEEGAKTAPGKRPRVEQSPQEPDSSGGIGKQOQPAKKRLNFGQTGDSSEVPDPQ PLGEPATPAAVGPTTMASSGGGAPNADNNEGADGVGNASGNWHCDSTWLGDRLVITTTSTRTWA LPTYNHLYKQISSASTGASNDNHYFGYSTPWGYFDENFRHCHFSRDPDQRLINNNWGFRPK
AAV6	9	AAB95450.1	

Serotype	SEQ ID NO:	Accession number	VP1 AMINO ACID SEQUENCE
AAV7	10	YP_077178.1	RLNFKLFNIQVKEVTTNDGVTTIANNLSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFM IPQGYLTLLNNGSQAVGRSSFYCLEYFPQSOMLRTGNNFTFSYTFEDVPFHSSYAHSQSLDRL MNPILIDQYLYYLNRTOQSGSAQNKDLLFSRCSPAGMSVQPKMWLPGPCYRQQRVSKTKTDN NNSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDFPMSGVMIFGKESAGASNTALDNV MITDEEEIKATNPVATERFGTVAVNLQSSSTDPAATGDVHVMGALPGMWQDRDVYLQGPWA KIPTHDGHFHPSPLMGGFGLKHPPQILIKNTPVPANPAEFSATKFASTFIQYSTGQVSVE IEWELQKENSKRWNPEVQYTSNYSANVDFTDNNGLTPEPRIGTRYLIRPL
			MAADGYLPDWLEDNLSEGIREWDLKPGAPKPKANQQKQDNQGRGLVLPGYKYLGPENGDKG EPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEEQERLQEDTSFGGNLGRAVFOAKKR VLEPLGLVEEAKTAPGKRPVEPSPQRSPTSSTIGIKGQQPARKRLNFGQTGDSVDPDP QPLGEPPAAPSSVGSITVAAGGAPMADNNEGADVGNASGNWHCHDSTWLGDRTVITSTRTW ALPTYNNHLYKQISSSETAGSTNDNTYFGYSTPWGYDFNRFCHCFSPRDWQRLINNNWGFRP KKLRFKLFIQVKEVTNDGVTTIANNLSTTIQVFSDEYQLPYVLGSAHQGCLPPFPADVF MIPQGYLTLLNNGSQSVGRSSFYCLEYFPQSOMLRTGNNFEFSYFEDVPFHSSYAHSQSLDR LMNPLIDQYLYYLARTQSNPGGTAGNRELQFYQGGPSTMAEQAKNWLPGPCFRQQRVSKTLD QNNNSFAWTGATKYHLNGRNSLVNPGVAMATHKDDDEDFPSSGVLIIFGKTGATNKTTLEN VMTNEEEIRPTNPVATEEYGISSNLQAANTAAQTQVNNQGALPGMWQDRDVYLQGPWA AKIPTHDGNFHPSPLMGGFGLKHPPQILIKNTPVPANPEVFTPAKFASFTIQTSTGQVSV EIEWELQKENSKRWNPEIQTSTNFEKQTVDFAVDSQVGYSEPRPIGTRYLTRNL
AAV8	11	YP_077180.1	MAADGYLPDWLEDNLSEGIREWDLKPGAPKPKANQQKQDDGRGLVLPGYKYLGPENGDKG EPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEEQERLQEDTSFGGNLGRAVFOAKKR VLEPLGLVEEAKTAPGKRPVEPSPQRSPTSSTIGIKGQQPARKRLNFGQTGDSVDPDP QPLGEPPAAPSGVGPNTMAAGGAPMADNNEGADVGNASGNWHCHDSTWLGDRTVITSTRTW ALPTYNNHLYKQISSNGTSGGATNDNTYFGYSTPWGYDFNRFCHCFSPRDWQRLINNNWGFR PKRISFKLFNIQVKEVTQNEGKTTIANNLSTIQVFTDSEYQLPYVLGSAHQGCLPPFPADV FMIPOGYLTLLNNGSQAVGRSSFYCLEYFPQSOMLRTGNNFQFTYTFEDVPFHSSYAHSQSLD RLMNPILIDQYLYYLSRTQTTGGTANTQTLGFSQGGPNTMANQAKNWLPGPCYRQQRVSTTG QNNNSFAWTGATKYHLNGRNSLANPGIAMATHKDDDEEFFPSNGLIFGKQNAARDNADYS DVMLTSEEEKTNPVATEEYGIIVADNLQOQNTAPQIGTVNSQGALPGMWQDRDVYLQGPWA WAKIPTHDGNFHPSPIMGGFGLKHPPQILIKNTPVPADPPTTFNQSKLNSFTIQTSTGQVS VEIEWELQKENSKRWNPEIQTSTNYSYKSTSVDFAVNTEGYSEPRPIGTRYLTRNL
			MAADGYLPDWLEDNLSEGIREWDLKPGAPQPKANQQKQDNARGVLVLPGYKYLGPENGDKG EPVNAADAAALEHDKAYDQQLKAGDNPYLYKNHADAEEQERLQEDTSFGGNLGRAVFOAKKR LLEPLGLVEEAAKTA PGKRPVEQSPQEPDSSAGIGKSAQPAKKRLNFGQTGDTESVDPDPQ PIGEPAPAAPSGVGLTMAAGGAPVADNNEGADVGNASGNWHCHDSTWLGDRTVITSTRTW LPTVNNHLYKQISNSTSGSSNDNAYFGYSTPWGYDFNRFCHCFSPRDWQRLINNNWGFRP KRLNFKLFNIQVKEVTDNNGVKTTIANNLSTVQVFTDSDYQLPYVLGSAHEGCLPPFPADVF MIPQGYLTLLNDGSQAVGRSSFYCLEYFPQSOMLRTGNNFQFSYEFENVPFHSSYAHSQSLDR LMNPLIDQYLYYLSKTINGSQNQOTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQN
AAV9	1	AAS99264.1	

Serotype	SEQ ID NO:	Accession number	VP1 AMINO ACID SEQUENCE
AAV9 K449R	2		NNSFFAWPGASSWALNGRNSLMNPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDADKV MITNEEEIKTTNPVATESYGQVATNHHQSAQAQQTGWQNGQILPGMVWQDRDVLQGP KIPIHTDGNFHPSPLMGGCMKHPPIILIKNTVPADPTAFNKKDLNSFTIQYSTGQVSVE IEWELQKENSKRWNPEIQYTSNYKSNVFEAVNTEGVSEPRPIGTRYLTRNL*
			MAADGYLPDWLEDNLSEGIREWWDLPKGPAPKQKQDQDGRGLVLPGYKYLGPFGNGLDKG EPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEEQERLQEDTSFGGNLGRAVFOAKKR LLEPLGLVEEAAKTA PGKKRPVEPQSPDSSTGIGKKQQAQKRLNFGQTGESESVDP PIGEPPAAPSGVGSLTMAAGGAPVADNNEGADGVGSSGNWHCDSQWLGDRLVITTTSTRTWA LPTYNHLYKQISNSTSGGSNDNAYFGYSTPWGYFDFNRFRCHFSFPRDWQRLNNNWGRFP KRLNFKLFENIQVKEVTDNNGVKTIANNLSTVQVFTDSYQLPYVLGSAHEGCLPPFPADV MIPQYGYLTLDGSAVGRSSFYCLEYFSPQMLRTGNNFQFSYEFENVFPHSSYAHSQSLDR LMNPLIDQYLYLSPTINGSGQNQOTLKFSVAGSPSNMAVQGRNYIPGPSYRQQRVSTVTQ NNNSEFAWPGASSWALNGRNSLMNPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDADK VMITNEEEIKTTNPVATESYGQVATNHHQSAQAQQTGWQNGQILPGMVWQDRDVLQGP AKIPIHTDGNFHPSPLMGGCMKHPPIILIKNTVPADPTAFNKKDLNSFTIQYSTGQVS EIEWELQKENSKRWNPEIQYTSNYKSNVFEAVNTEGVSEPRPIGTRYLTRNL*
AAV10	15	AAT46337.1	MAADGYLPDWLEDNLSEGIREWWDLPKGPAPKQKQDQDGRGLVLPGYKYLGPFGNGLDKG EPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEEQERLQEDTSFGGNLGRAVFOAKKR VLEPLGLVEEAAKTA PGKKRPVEPQSPDSSTGIGKKQQAQKRLNFGQTGESESVDP QPIGEPPAGSGLTMAAGGAPVADNNEGADGVGSSGNWHCDSTWLGDRLVITTTSTRTW ALPTYNHLYKQISNSTSGGSNDNAYFGYSTPWGYFDFNRFRCHFSFPRDWQRLNNNWGRFP PKRLSFKLFENIQVKEVTDNNGVKTIANNLSTVQVFTDSYQLPYVLGSAHEGCLPPFPADV FMIPIQYGYLTLDGSAVGRSSFYCLEYFSPQMLRTGNNFQFSYEFENVFPHSSYAHSQSLDR RLMNPLIDQYLYLSRTSTGQTQQLLFSQAQGANMSAQAKNWLPGPCYPQQRVSTTSL QNNNSFAWTGATKYHLNGRDSLVPVAMATHKDDERFFPSSGVLMMFGKQAGRDNDVYS SVMLTSEEEIKTTNPVATESYGQVADNLQQAQNTGPIVGNVNSQALPGMVWQDRDVLQGP WAKIPIHTDGNFHPSPLMGGFGLKHPPIILIKNTVPADPTTFSQAKLASFTIQYSTGQVS VEIEWELQKENSKRWNPEIQYTSNYKSNVFEAVNTEGVSEPRPIGTRYLTRNL
			MAADGYLPDWLEDNLSEGIREWWDLPKGPAPKQKQDQDGRGLVLPGYKYLGPFGNGLDKG EPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEEQERLQEDTSFGGNLGRAVFOAKKR VLEPLGLVEEAAKTA PGKKRPVEPQSPDSSTGIGKKQQAQKRLNFEEDTGAGDGPPEGS DTSAMSSDIEMRAAPGGNAVDAGQSDGVGNASGDWHCDSTWSEGVTTTSTRTWVLTNN HLYRLGTSSSNTYNGFTPWGYFDFNRFRCHFSFPRDWQRLNNNWGLRPKAMRVKIFNIQ VKEVTTNGETTVANNLTSTVQIFADSSYELPYVMDAGQSGSLPPFPNDVFMVFPQYGCIV TGENQQTDRNAYFQYFSPQMLRTGNNFEMAYNFKVPFHSMTAHQSQSLRLMNPFLDQY LWHLQSTTSGETLNQNAATTFGRTSGDFAFYRKNWLPGPCVQQRFSKTASQNYKIPASG GNALLKYDTHYTLNNRWSNIAPGPPMATAGPSDGDFSNAQLIFPGPSVTGNTTTSANLLFT SEEEIAATNPRDMDMFGQIADNNQNATTAPITGNVTAMGVLPGMVWQDRDVLQGP
AAV11	16	AAT46339.1	

Serotype	SEQ ID NO:	Accession number	VP1 AMINO ACID SEQUENCE
			HADGHFHPSPILIGFGGLKHPPPQIFIKNTVPANPATTFTAARDVSFITQYSTGQVAVQIEW EIEKERSKRWNPEVQFTSNYGNQSSMLWAPDITGKYTERVIGSRYLTHHL
AAV12	17	ABI16639.1	MAADGYLPDWLEDNLSEGIREWALKPGAPQKANKQHQDNGRGLVLPGYKYLGPFGNGLDKG EPVNEADAAALEHDKAYDKQLEQGDNPYLKYNHADAEFQORLATDTSFGGNLGRAVFOAKKR ILEPLGLVEEGVKTAPGKKRPLEKTPNRPDPSGAKPAKKQKDGEPADSAARTLDFEDSG AGDGPPEGSSGEMSHDAEMRAAPGNAVEAGQAGADGVGNASGDWHCDSTWSEGRVTTSTR TWLPTYNHLYLRIGTTANSTYNGFSTPWGYFDENRFHCHFSRPDQWRLINNNWGLRPKS MRVKIFNIQVKEVTTNGETTVANNLTSTVQIPADSTVELPYMDAQEGEGSPFPNDVFMV PQYGYCGVVTGKNQNTDRNAFYCLEYFPSQMLRTGNNEFVSQFEKVPFHSMYAHQSILDR MMNPLLDQYLVHLSQTTTNSLNQGTATTYGIKTTDFAYYRKNWLPGACIKQKFKSKNAN QNYKIPASGGDALLKYDTHTTNLGRWSNMAPGPMTATAGADSDFSNSQLIFAGPNPSGNTT TSSNNLLFTSEEEIATTNPRDITMFGQIADNNQNAATTAPHIANLDAMGIVPGMWQNNDIYY QGPIWAKVPHTDGHFHPSPLMGGFGLKHPPPQIFIKNTVPANPNTTFSAARINSFLTQYST GOVAVQIDWEIQKEHSKRWNPEVQFTSNYGNQSSMLWAPDITGKYTERVIGSRYLTHHL
	18	ABZ10812.1	MTDGYLPDWLEDNLSEGIREWALKPGAPKPKANQHQDNGRGLVLPGYKYLGPFGNGLDKGE PVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADAEFQORLQEDTSFGGNLGRAVFOAKKR LEPLGLVEEAAKTAPGKKRPEQSPAEPDSSSGIKSGSQQPAKKRINFQGTGDTESVDPQP LGQPPAAPSGVSTTMASSGGAPMADNNEGADGVGNSGWNHCDSQWLGDRTTSTRTVAL PTYNNHLYKQISSQSGATNDNHYFGYSTPWGYFDENRFHCHFSRPDQWRLINNNWGRPKRL NFKLFNIQVKEVTQNDGTTIANNLSTVQVTFDSEYQLPYVLGSAHQGCLPEFPADVEMVP QYGYLTNNQSQAVGRSSFYCLEYFPSQMLRTGNNFQSYTFEDVPFHSSYAHQSILDRIMN PLIDQYLYLNRPTQTSQTQQRLLFSQAGPTMSLQAKNWLPGCYRQRLSKQANDNNNS NFPWTGATKYHLNGRDSLVPNPGPAMASHKDDKEKFFPMHGTLI FGKEGTNANNADLENVMIT DEEIRTTNPVATEQYGTVSNNNLQNSAGPTTGTVNHQALPGMWQDQDRDVLQGPIMAKIP HTDGHFHPSPLMGGFGLKHPPPQIMIKNTVPANPNTTFSAAKFSFITQYSTGQVSVLEIWE ELQKENSKRWNPEIQYTSNYSKSNVDFTDITNGVSEPRPIGTRYLTRNL
AAVrh. 8	12	AAO88183.1	MAADGYLPDWLEDNLSEGIREWDLKPGAPKPKANQHQDNGRGLVLPGYKYLGPFGNGLDKG EPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTSFGGNLGRAVFOAKKR VLEPLGLVEGAKTAPGKKRPEQSPAEPDSSSGIKGTQQPAKKRINFQGTGDSSEVPDPQ PLGEPAPASGLGPNMTMASGGGAPMADNNEGADGVGNSGWNHCDSTWLGDRVITTTSTRWA LPTYNNHLYKQISNGTSGSTNDNTYFGYSTPWGYFDENRFHCHFSRPDQWRLINNNWGRFP KRLNFKLFNIQVKEVTNTEGKTTIANNLSTVQVTFDSEYQLPYVLGSAHQGCLPPFPADV MVPQGYLTNNQSQALGRSSFYCLEYFPSQMLRTGNNFQSYTFEDVPFHSSYAHQSILDR LMNPLIDQYLYLNRPTQTSQTQQRLLFSQAGPTMSLQAKNWLPGCYRQRLSKQANDNNNS NNSNFAWTGAAKFKLNGRDSLMPNPGVAMASHKDDDDFFPSSGVLI FGKQAGNDGDVYSQV LITDEEEIKATNPVATEEYGAVANNQAAQTGLVHNQGVIPGMVWQNRDVLQGPIMAKIP KIPIHTDGNFHPSPLMGGFGLKHPPPQILIKNTVPADPPLTFNQAKLNSFITQYSTGQVSVLE IEWELQKENSKRWNPEIQYTSNYSKSNVDFAVNTEGVSEPRPIGTRYLTRNL

Serotype	SEQ ID NO:	Accession number	VP1 AMINO ACID SEQUENCE
AAVrh. 10	13	AAO88201.1	MAADGYLPDWLEDNLSSEGTREWMDLKPAPKPKANQQKQDDGRGLVLPGYKYLGPFNGLDKG EPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAERLQEDTSFGGNLGRAVFQAKKR VLEPLGLVEEAKTAPGKKRPVEPSPQSDSTGIGKKQQQPAKRLNFGQTGDSSEVPDP QPIGEPPAGPSGLSGTMAAGGAPMADNNEGADGVSGNSGNWHCHDSTWLGDRVITTTSTRTW ALPT'YNNHLYKQISNGTSGGSTNDNTYFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFR PKRLNFKLFNIQVKEVTQNEGKTIANNLTSTIQVFTDSEYQLPYVLGSAHQGCLPPEFADV FMI PQGYLTLLNNGSQAVERSSFYCLEYFQSMLRTGNNEFFSYQFEDVPFHSYAHQSOLD RLMNP LI DQYLYLSRTQSTGGTAGTQQLLFSQAQPNMMSAQAKNWLPGPCYRQQRVSTTLS QNNSNFAWTGATKYHLNGRDSLVPNGVAMATHKDDERFFPSSGVLMFGKQAGAKDNVDYS SVMLTSEEEIKTTNPVATEQYGVVADNLQQNAAPIVGAVNSQGALPGMVWQNDRDYLQGP I WAKI PHTDGNFHPSPLMGGFGLKHPPQIILIKNTVPADPPTTFSQAKLASFITQYSTGQVS VEIEWELQKENSKRWNPEIQYTSNYKSTNVDFAVNTDGTYSERPRIGTRYLTRNL
	14	SEQ ID NO: 2 of US20220370639	MAADGYLPDWLEDNLSSEGTREWMDLKPAPKPKANQQKQDNGRGLVLPGYKYLGPFNGLDKG EPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAERLQEDTSFGGNLGRAVFQAKKR VLEPLGLVESPVKTAPGKKRPVEPSPQSDSTGIGKKQQQPAKRLNFGQTGDSSEVPDP QPIGEPPAGPSGLSGTMAAGGAPMADNNEGADGVSGNSGNWHCHDSTWLGDRV ITTTSTRTWALPT'YNNHLYKQISNGTSGGSTNDNTYFGYSTPWGYFDENRFCHFSPRDWQRL INNNWGFRPKRLNFKLFNIQVKEVTQNEGKTIANNLTSTIQVFTDSEYQLPYVLGSAHQG LPFPADVFMIPQGYLTLLNNGSQAVERSSFYCLEYFQSMLRTGNNEFFSYNFEFVDPFHSS YAHQSOLDRLMNP LI DQYLYLSRTQSTGGTAGTQQLLFSQAQPNMMSAQAKNW LPGPCYRQQRVSTTLSQNNSNFAWTGATKYHLNGRDSLVPNGVAMATHKDDERFFPSSGV LMFGKQAGAKDNVDYSSVMLTSEELKTTNPVATEQYGVVADNLQQNAAPIVGAVNSQGAL PGMVWQNDRDYLQGP I WAKI PHTDGNFHPSPLMGGFGLKHPPQIILIKNTVPADPPTTFNQ AKLASFITQYSTGQVSVEIEWELQKENSKRWNPEIQYTSNYKSTNVDFAVNTE GTYSERPRIGTRYLTRNL
AAV- LK19	19	AGT20780.1	MAADGYLPDWLEDNLSSEGTREWWALQPGAPKPKANQQHODNARGLVLPGYKYLGPGNGLDKG EPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADAERLQEDTSFGGNLGRAVFQAKKR LLEPLGLVEEAAKTAPGKKRPVDQSPQEPDSSGVGSKGQPAKRLNFGQTGDSSEVPDPQ PLGPPPAAPTSLGNTMASSGGAPMADNNEGADGVSGNSGNWHCHDSTWLGDRVITTTSTRTWA LPT'YNNHLYKQISSQSGASNDNHYFGYSTPWGYFDENRFCHFSPRDWQRLINNNWGFRPKK LSFKLFNIQVKEVTQNDGTTTIANNLSTVQVTDSEYQLPYVLGSAHQGCLPFPADVFMV PQGYLTLLNNGSQAVERSSFYCLEYFQSMLRTGNNFQFSYTFEDVPFHSYAHQSOLDRLM NP LI DQYLYLNRTQGTSGTTNQSRLLFSQAQPSQMSLQARNWLPGPCYRQQRLSKTANDN NNSNFPWTAASKYHLNGRDSLVPNGFAMASHKDEKFFPMHGNLIFGKEGTASNAELDNV MITDEEEIRTTNPVATEQYGTVANNLQSSNTAPTTTRTNDQGALPGMVWQDRDYLQGP I WA KIPHTDGHFHPSPLMGGFGLKHPPQIIMIKNTVPANPTTFSQAKLASFITQYSTGQVSVE IEWELQKENSKRWNPEIQYTSNYKSVNVDFVDTNGVYSEPRIGTRYLTRPL
	20	3J1Q_A_	MAADGYLPDWLEDTLSEGTIRQWWKLPKPPPKPAERHKDDSRGLVLPGYKYLGPFNGLDKG EPVNEADAAALEHDKAYDRQLDSGDNPNYLKYNHADAERLQEDTSFGGNLGRAVFQAKKR

Serotype	SEQ ID NO:	Accession number	VP1 AMINO ACID SEQUENCE
			LLEPLGLVEEAAKTA PGKKRPVEHSPVEPDSSSGTGKAGQQPARKRLNFGQTGDADSVDPDQ PIGEPPAAPSGVGS LITMAAGGAPNADNNEGADGVNSGNGHCDSTWMDRVITTTSTRTWA LPTYNHLYKQISNSTSGSSNDNAYFGYSTFWGYFDFNRFCHCFSPRDWQRLINNNWGFRP KRLSFKLENIQVKEVTQNEGTKTIANNLTSTIQVFTDSEYQLPYVLGSAHQGCLPPFPADV MIPOGYLTLNNGSQAVGRSSFYCLEYFPQSQMLRTGNNFQFTYTFEDVFPFHSYAHSQSLDR LMNPLIDQYLYLSRTQTGGTTNTQTGFSQGGPNTMANQAKWILPGPCYRQQRVSKTSAD NNSEYSWTGATKYHLNGRDSLVPFAMASHKDDDEKFFPQSGVLI FGKQGSEKTNVDIEK VMITDEEEIRTNPVATEQYGSVSTNLQRGNRQAATAVNTQGVLPGMVWQDRDVIYLGPIW AKIPTHDTGHHFSPPLMGGFGLKHPPPPQILIKNTPVPADPTTFENQSKINSFITQYSTGQVSV EIEWELQKENSKRWNPEIQYTSNYKSTSVDFAVNTEGYVSEPRPIGTRYLTRNL

[0105] In an embodiment, the 7-mer comprises YSRIGPN (SEQ ID NO: 1390), YSRNSDN (SEQ ID NO: 1391), or LHRLGPN (SEQ ID NO: 1392), and wherein X₁ comprises A, G, E, L, N, Q, S, W, or M; X₂ comprises A, F, I, L, M, N, Q, P, T, V, or Y; X₃ comprises Q; X₄ comprises any amino acid; X₅ comprises D, F, G, I, L, M, Q, P, S, T, or V; X₆ comprises A, C, F, G, H, I, P, S, T, V, W, or Y; X₇ comprises D, E, Q, or T; or any combination thereof. In an embodiment, the transferrin receptor binding modification comprises of the amino acid sequence of one of SEQ ID Nos: 60-1210.

[0106] In an embodiment, the n-mer comprises of YSRIGPN (SEQ ID NO: 1390), and wherein X₁ comprises of A, G, E, L, N, Q, S, W, or M; X₂ comprises of A, F, I, L, M, N, Q, P, T, V, or Y; X₃ comprises of Q; X₄ comprises of A, E, F, H, I, L, M, N, P, Q, V, Y, D, or G; X₅ comprises of D, F, G, I, L, M, Q, P, S, T, or V; X₆ comprises of A, C, F, G, H, I, P, S, T, V, W, or Y; X₇ comprises of D, E, Q, or T; or any combination thereof. In an embodiment, the transferrin receptor binding modification comprises of the amino acid sequence of one of SEQ ID Nos: 60-1136.

[0107] In an embodiment, the n-mer comprises of YSRNSDN (SEQ ID NO: 1391) and X₁ comprises of S, or W; X₂ comprises of V, I, or F; X₃ comprises of Q; X₄ comprises of any amino acid; X₅ comprises of Q or T; X₆ comprises of A; X₇ comprises of Q; or any combination thereof. In an embodiment, the transferrin receptor binding modification comprises of the amino acid sequence of one of SEQ ID Nos: 1152-1157.

[0108] In an embodiment, the n-mer comprises of LHRLGPN (SEQ ID NO: 1392) and X₁ comprises of S, A, L, or M; X₂ comprises of A or P; X₃ comprises of Q; X₄ comprises of A, E, F, H, I, L, M, N, P, Q, V, or Y; X₅ comprises of Q; X₆ comprises of A, P, S, or T; X₇ comprises of D, E, Q, or T; or any combination thereof. In an embodiment, the transferrin receptor binding modification comprises of the amino acid sequence of one of SEQ ID Nos: 1158-1210.

[0109] In an embodiment, the 7-mer comprises of FRSTNGV (SEQ ID NO: 1393), FVSTNGV (SEQ ID NO: 1394), FZ₁STNGZ₂ (SEQ ID NO: 1395), FRSTNGZ₃ (SEQ ID NO: 1396), or VESTNGR (SEQ ID NO: 1397) and wherein X₁ comprises of S; X₂ comprises of A, S, M, or D; X₃ comprises of F, H, I, L, M, N, Q, R, Y, D, or E; X₄ comprises of A, S, or M; X₅ comprises of Q or P; X₆ comprises of A, F, H, Q, or S; X₇ comprises of A, D, E, F, Q, S, or T; or any combination thereof. In an embodiment, wherein the transferrin receptor binding modification comprises of the amino acid sequence of one of SEQ ID Nos: 1211-1389.

[0110] In an embodiment, the n-mer comprises of FRSTNGV (SEQ ID NO: 1393), and X₁ comprises of S; X₂ comprises of A or S; X₃ comprises of D; X₄ comprises of A, or S; X₅ comprises of Q or P; X₆ comprises of A, F, H, Q, or S; X₇ comprises of D, E, Q, or T; or any combination thereof. In an embodiment, the transferrin receptor binding modification comprises of the amino acid sequence of one of SEQ ID Nos: 1211-1266.

[0111] In an embodiment, the n-mer comprises of FVSTNGV (SEQ ID NO: 1394), and X₁ comprises of S; X₂ comprises of A, S, or M; X₃ comprises of Q, E, or D; X₄ comprises of A, or M; X₅ comprises of Q or P; X₆ comprises of A; X₇ comprises of E; or any combination thereof. In an embodiment, the transferrin receptor binding modification comprises of the amino acid sequence of one of SEQ ID Nos: 1267-1298.

[0112] In an embodiment, the n-mer comprises of FZ₁STNGZ₂ (SEQ ID NO: 1395) or FRSTNGZ₃ (SEQ ID NO: 1396), and X₁ comprises of S; X₂ comprises of A or S; X₃ comprises of Q or D; X₄ comprises of A; X₅ comprises of Q; X₆ comprises of A; X₇ comprises of E; or any combination thereof. In an embodiment, the transferrin receptor binding modification comprises of the amino acid sequence of one of SEQ ID Nos: 1299-1319.

[0113] In an embodiment, the 7-mer comprises of VESTNGR (SEQ ID NO: 1397), and X₁ comprises of S; X₂ comprises of S or D; X₃ comprises of F, H, I, L, M, N, Q, R, or Y; X₄ comprises of A; X₅ comprises of Q or P; X₆ comprises of A; X₇ comprises of A, D, E, F, Q, S, or T; or any combination thereof. In an embodiment, the transferrin receptor binding modification comprises of the amino acid sequence of one of SEQ ID Nos: 1320-1388.

Table C. Example location of transferrin receptor (TfR) binding modifications with example AAVs

	UPSTREAM SEQUENCE	7-mer INSERTION (X1-X2-X3-[7-mer]-X4-X5-X6-X7)	DOWNSTREAM SEQUENCE
AAV1	VAVNFQ (SEQ ID NO: 1473)	7-mer: YSRIGPN (SEQ ID NO: 1390), YSRNSDN (SEQ ID NO: 1391), LHRLGPN (SEQ ID NO: 1392), FZISTNGZ2 (SEQ ID NO: 1395), FVSTNGV (SEQ ID NO: 1394), FZISTNGZ2 (SEQ ID NO: 1395), FRSTNGZ3 (SEQ ID NO: 1396), or VESTNGR (SEQ ID NO: 1397) X1, X2, X3, X4, X5, X6, X7 indicate one or more modifications at amino acid positions in the capsid polypeptide flanking the inserted 7-mer. The transferrin receptor binding modification can include any one of the amino acid sequences of SEQ ID Nos: 60-1210.	TGDVHAMG (SEQ ID NO: 1486)
AAV2	VSTNLQ (SEQ ID NO: 1474)		TADVNTQG (SEQ ID NO: 1487)
AAV3	VANNLQ (SEQ ID NO: 1475)		TGTVNHQG (SEQ ID NO: 1488)
AAV4	PGGDQ (SEQ ID NO: 1476)		VDRLTALG (SEQ ID NO: 1489)
AAV5	MATNNQ (SEQ ID NO: 1477)		TGTYNLQE (SEQ ID NO: 1490)
AAV6	VAVNLQ (SEQ ID NO: 1478)		TGDVHVMG (SEQ ID NO: 1491)
AAV7	VSSNLQ (SEQ ID NO: 1479)		TQVVNNQG (SEQ ID NO: 1492)
AAV8	VADNLQ (SEQ ID NO: 1480)		IGTVNSQG (SEQ ID NO: 1493)
AAV9	VATNHQ (SEQ ID NO: 1481)		TGWVQNQG (SEQ ID NO: 1494)
AAV10	VADNLQ (SEQ ID NO: 1482)		VGNVNSQG (SEQ ID NO: 1495)
AAV11	IADNNQ (SEQ ID NO: 1483)		TGNVTAMG (SEQ ID NO: 1496)
AAV12	IADNNQ (SEQ ID NO: 1484)		IANLDAMG (SEQ ID NO: 1497)
AAV13	VSNLQ (SEQ ID NO: 1485)		TGTVNHQG (SEQ ID NO: 1498)

Table D. Example sequences including transferrin receptor (TFR1) binding modifications with example AAVs

SEROTYPE	UPSTREAM SEQUENCE	7-mer INSERTION	DOWNSTREAM SEQUENCE	UPSTREAM SEQUENCE + 7-mer INSERTION (SEQ ID NO.)	7-mer INSERTION+ DOWNSTREAM SEQUENCE (SEQ ID NO.)
AAV1	VAVIFQSSS (SEQ ID NO: 1499)	YSRIGPN (SEQ ID NO: 1390)	DPATGDVHA (SEQ ID NO: 1512)	VAVIFQSSSYSRIGPN (SEQ ID NO: 1525)	YSRIGPNDPATGDVHA (SEQ ID NO: 1538)
AAV2	VSINLQRGN (SEQ ID NO: 1500)	YSRIGPN (SEQ ID NO: 1390)	GQAATADVNT (SEQ ID NO: 1513)	VSINLQRGNYSRIGPN (SEQ ID NO: 1526)	YSRIGPNGQAATADVNT (SEQ ID NO: 1539)
AAV3	VANNLQSSN (SEQ ID NO: 1501)	YSRIGPN (SEQ ID NO: 1390)	SAPTTGTVNH (SEQ ID NO: 1514)	VANNLQSSNYSRIGPN (SEQ ID NO: 1527)	YSRIGPNSAPTTGTVNH (SEQ ID NO: 1540)
AAV4	LYGGNSNS (SEQ ID NO: 1502)	YSRIGPN (SEQ ID NO: 1390)	NLPTVDRLTALG (SEQ ID NO: 1515)	LYGGNSNSYSRIGPN (SEQ ID NO: 1528)	YSRIGPNNLPTVDRLTALG (SEQ ID NO: 1541)
AAV5	MATNNQSST (SEQ ID NO: 1503)	YSRIGPN (SEQ ID NO: 1390)	TAPATGTYNLQE (SEQ ID NO: 1516)	MATNNQSSTYSRIGPN (SEQ ID NO: 1529)	YSRIGPNTAPATGTYNLQE (SEQ ID NO: 1542)
AAV6	VAVNLMSSS (SEQ ID NO: 1504)	YSRIGPN (SEQ ID NO: 1390)	TDPATGDDVHVMG (SEQ ID NO: 1517)	VAVNLMSSSYSRIGPN (SEQ ID NO: 1530)	YSRIGPNTDPATGDDVHVMG (SEQ ID NO: 1543)
AAV7	VSSNLQAAN (SEQ ID NO: 1505)	YSRIGPN (SEQ ID NO: 1390)	TAAQTQVVNNQG (SEQ ID NO: 1518)	VSSNLQAANYSRIGPN (SEQ ID NO: 1531)	YSRIGPNTAAQTQVVNNQG (SEQ ID NO: 1544)
AAV8	VADNLQQQN (SEQ ID NO: 1506)	YSRIGPN (SEQ ID NO: 1390)	TAPQIGTVNSQG (SEQ ID NO: 1519)	VADNLQQQNYSRIGPN (SEQ ID NO: 1532)	YSRIGPNTAPQIGTVNSQG (SEQ ID NO: 1545)
AAV9	VATNHQSAQ (SEQ ID NO: 1507)	YSRIGPN (SEQ ID NO: 1390)	AQAQTGWVQNQG (SEQ ID NO: 1520)	VATNHQSAQYSRIGPN (SEQ ID NO: 1533)	YSRIGPNAQAQTGWVQNQG (SEQ ID NO: 1546)
AAV10	VADNLQQAN (SEQ ID NO: 1508)	YSRIGPN (SEQ ID NO: 1390)	TGPIVGNVNSQG (SEQ ID NO: 1521)	VADNLQQANYSRIGPN (SEQ ID NO: 1534)	YSRIGPNTGPIVGNVNSQG (SEQ ID NO: 1547)
AAV11	IADNNQNAT (SEQ ID NO: 1509)	YSRIGPN (SEQ ID NO: 1390)	TAPITGNVTAMG (SEQ ID NO: 1522)	IADNNQNATYSRIGPN (SEQ ID NO: 1535)	YSRIGPNTAPITGNVTAMG (SEQ ID NO: 1548)
AAV12	IADNNQNAT (SEQ ID NO: 1510)	YSRIGPN (SEQ ID NO: 1390)	TAPHIANLIDAMG (SEQ ID NO: 1523)	IADNNQNATYSRIGPN (SEQ ID NO: 1536)	YSRIGPNTAPHIANLIDAMG (SEQ ID NO: 1549)
AAV13	VSNLQNSN (SEQ ID NO: 1511)	YSRIGPN (SEQ ID NO: 1390)	AGPTTGTVNHQHG (SEQ ID NO: 1524)	VSNLQNSNYSRIGPN (SEQ ID NO: 1537)	YSRIGPNAGPTTGTVNHQHG (SEQ ID NO: 1550)

SEROTYPE	UPSTREAM SEQUENCE	7-mer INSERTION	DOWNSTREAM SEQUENCE	UPSTREAM SEQUENCE + 7-mer INSERTION (SEQ ID NO.)	7-mer INSERTION+ DOWNSTREAM SEQUENCE (SEQ ID NO.)
AAV1	VAVIFQSSS (SEQ ID NO: 1499)	YSRNSDN (SEQ ID NO: 1391)	DPATGDVHA (SEQ ID NO: 1512)	VAVIFQSSSYSRNSDN (SEQ ID NO: 1551)	YSRNSDNDPATGDVHA (SEQ ID NO: 1564)
AAV2	VSINLQRGN (SEQ ID NO: 1500)	YSRNSDN (SEQ ID NO: 1391)	GQAATADVNT (SEQ ID NO: 1513)	VSINLQRGNYSRNSDN (SEQ ID NO: 1552)	YSRNSDNGQAATADVNT (SEQ ID NO: 1565)
AAV3	VANNLQSSN (SEQ ID NO: 1501)	YSRNSDN (SEQ ID NO: 1391)	SAPTTGTVNH (SEQ ID NO: 1514)	VANNLQSSNSYSRNSDN (SEQ ID NO: 1553)	YSRNSDNSAPTTGTVNH (SEQ ID NO: 1566)
AAV4	LYGGNSNS (SEQ ID NO: 1502)	YSRNSDN (SEQ ID NO: 1391)	NLPTVDRLTALG (SEQ ID NO: 1515)	LYGGNSNSYSRNSDN (SEQ ID NO: 1554)	YSRNSDNNLPTVDRLTALG (SEQ ID NO: 1567)
AAV5	MATNNQSST (SEQ ID NO: 1503)	YSRNSDN (SEQ ID NO: 1391)	TAPATGTYNLQE (SEQ ID NO: 1516)	MATNNQSSTYSRNSDN (SEQ ID NO: 1555)	YSRNSDNTAPATGTYNLQE (SEQ ID NO: 1568)
AAV6	VAVNLMSSS (SEQ ID NO: 1504)	YSRNSDN (SEQ ID NO: 1391)	TDPATGDVHVVMG (SEQ ID NO: 1517)	VAVNLMSSSYSRNSDN (SEQ ID NO: 1556)	YSRNSDNTDPATGDVHVVMG (SEQ ID NO: 1569)
AAV7	VSSNLQAAN (SEQ ID NO: 1505)	YSRNSDN (SEQ ID NO: 1391)	TAAQTQVVNNQG (SEQ ID NO: 1518)	VSSNLQAANYSRNSDN (SEQ ID NO: 1557)	YSRNSDNTAAQTQVVNNQG (SEQ ID NO: 1570)
AAV8	VADNLQQQN (SEQ ID NO: 1506)	YSRNSDN (SEQ ID NO: 1391)	TAPQIGTVNSQG (SEQ ID NO: 1519)	VADNLQQQNYSRNSDN (SEQ ID NO: 1558)	YSRNSDNTAPQIGTVNSQG (SEQ ID NO: 1571)
AAV9	VATNHQSAQ (SEQ ID NO: 1507)	YSRNSDN (SEQ ID NO: 1391)	AQAQTGWVQNQG (SEQ ID NO: 1520)	VATNHQSAQYSRNSDN (SEQ ID NO: 1559)	YSRNSDNAQAQTGWVQNQG (SEQ ID NO: 1572)
AAV10	VADNLQQAN (SEQ ID NO: 1508)	YSRNSDN (SEQ ID NO: 1391)	TGPVGNVNSQG (SEQ ID NO: 1521)	VADNLQQANYSRNSDN (SEQ ID NO: 1560)	YSRNSDNTGPVGNVNSQG (SEQ ID NO: 1573)
AAV11	IADNNQNAT (SEQ ID NO: 1509)	YSRNSDN (SEQ ID NO: 1391)	TAPITGNVTAMG (SEQ ID NO: 1522)	IADNNQNATYSRNSDN (SEQ ID NO: 1561)	YSRNSDNTAPITGNVTAMG (SEQ ID NO: 1574)
AAV12	IADNNQNAT (SEQ ID NO: 1510)	YSRNSDN (SEQ ID NO: 1391)	TAPHIANLDAMG (SEQ ID NO: 1523)	IADNNQNATYSRNSDN (SEQ ID NO: 1562)	YSRNSDNTAPHIANLDAMG (SEQ ID NO: 1575)
AAV13	VSNNLQNSN (SEQ ID NO: 1511)	YSRNSDN (SEQ ID NO: 1391)	AGPTTGTVNHQ (SEQ ID NO: 1524)	VSNNLQNSNYSRNSDN (SEQ ID NO: 1563)	YSRNSDNAGPTTGTVNHQ (SEQ ID NO: 1576)

SEROTYPE	UPSTREAM SEQUENCE	7-mer INSERTION	DOWNSTREAM SEQUENCE	UPSTREAM SEQUENCE + 7-mer INSERTION (SEQ ID NO.)	7-mer INSERTION+ DOWNSTREAM SEQUENCE (SEQ ID NO.)
AAV1	VAVIFQSSS (SEQ ID NO: 1499)	LHRLGPN (SEQ ID NO: 1392)	DPATGDVHA (SEQ ID NO: 1512)	VAVIFQSSSLHRLGPN (SEQ ID NO: 1577)	LHRLGPNDPATGDVHA (SEQ ID NO: 1590)
AAV2	VSINLQRGN (SEQ ID NO: 1500)	LHRLGPN (SEQ ID NO: 1392)	GQAATADVNT (SEQ ID NO: 1513)	VSINLQRGNLHRLGPN (SEQ ID NO: 1578)	LHRLGPNGQAATADVNT (SEQ ID NO: 1591)
AAV3	VANNLQSSN (SEQ ID NO: 1501)	LHRLGPN (SEQ ID NO: 1392)	SAPTTGTVNH (SEQ ID NO: 1514)	VANNLQSSNLHRLGPN (SEQ ID NO: 1579)	LHRLGPN SAPTTGTVNH (SEQ ID NO: 1592)
AAV4	LYGGNSNS (SEQ ID NO: 1502)	LHRLGPN (SEQ ID NO: 1392)	NLPTVDRLTALG (SEQ ID NO: 1515)	LYGGNSNSLHRLGPN (SEQ ID NO: 1580)	LHRLGPNNLPTVDRLTALG (SEQ ID NO: 1593)
AAV5	MATNNQSST (SEQ ID NO: 1503)	LHRLGPN (SEQ ID NO: 1392)	TAPATGTYNLQE (SEQ ID NO: 1516)	MATNNQSSTLHRLGPN (SEQ ID NO: 1581)	LHRLGPN TAPATGTYNLQE (SEQ ID NO: 1594)
AAV6	VAVNLMSSS (SEQ ID NO: 1504)	LHRLGPN (SEQ ID NO: 1392)	TDPATGDVHVVMG (SEQ ID NO: 1517)	VAVNLMSSSLHRLGPN (SEQ ID NO: 1582)	LHRLGPN TDPATGDVHVVMG (SEQ ID NO: 1595)
AAV7	VSSNLQAAAN (SEQ ID NO: 1505)	LHRLGPN (SEQ ID NO: 1392)	TAAQTQVVNNQG (SEQ ID NO: 1518)	VSSNLQAAANLHRLGPN (SEQ ID NO: 1583)	LHRLGPN TAAQTQVVNNQG (SEQ ID NO: 1596)
AAV8	VADNLQQQN (SEQ ID NO: 1506)	LHRLGPN (SEQ ID NO: 1392)	TAPQIGTVNSQG (SEQ ID NO: 1519)	VADNLQQQNLHRLGPN (SEQ ID NO: 1584)	LHRLGPN TAPQIGTVNSQG (SEQ ID NO: 1597)
AAV9	VATNHQSAQ (SEQ ID NO: 1507)	LHRLGPN (SEQ ID NO: 1392)	AQAQTGWVQNQG (SEQ ID NO: 1520)	VATNHQSAQLHRLGPN (SEQ ID NO: 1585)	LHRLGPN AQAQTGWVQNQG (SEQ ID NO: 1598)
AAV10	VADNLQQAN (SEQ ID NO: 1508)	LHRLGPN (SEQ ID NO: 1392)	TGPVGNVNSQG (SEQ ID NO: 1521)	VADNLQQANLHRLGPN (SEQ ID NO: 1586)	LHRLGPN TGPVGNVNSQG (SEQ ID NO: 1599)
AAV11	IADNNQNAT (SEQ ID NO: 1509)	LHRLGPN (SEQ ID NO: 1392)	TAPITGNVTAMG (SEQ ID NO: 1522)	IADNNQNATLHRLGPN (SEQ ID NO: 1587)	LHRLGPN TAPITGNVTAMG (SEQ ID NO: 1600)
AAV12	IADNNQNAT (SEQ ID NO: 1510)	LHRLGPN (SEQ ID NO: 1392)	TAPHIANLDAMG (SEQ ID NO: 1523)	IADNNQNATLHRLGPN (SEQ ID NO: 1588)	LHRLGPN TAPHIANLDAMG (SEQ ID NO: 1601)
AAV13	VSNNLQNSN (SEQ ID NO: 1511)	LHRLGPN (SEQ ID NO: 1392)	AGPTTGTVNHQ (SEQ ID NO: 1524)	VSNNLQNSNLHRLGPN (SEQ ID NO: 1589)	LHRLGPN AGPTTGTVNHQ (SEQ ID NO: 1602)

SEROTYPE	UPSTREAM SEQUENCE	7-mer INSERTION	DOWNSTREAM SEQUENCE	UPSTREAM SEQUENCE + 7-mer INSERTION (SEQ ID NO.)	7-mer INSERTION+ DOWNSTREAM SEQUENCE (SEQ ID NO.)
AAV1	VAVIFQSSS (SEQ ID NO: 1499)	FRSTNGV (SEQ ID NO: 1393)	DPATGDVHA (SEQ ID NO: 1512)	VAVIFQSSSFRSTNGV (SEQ ID NO: 1603)	FRSTNGVDPATGDVHA (SEQ ID NO: 1616)
AAV2	VSINLQRGN (SEQ ID NO: 1500)	FRSTNGV (SEQ ID NO: 1393)	GQAATADVNT (SEQ ID NO: 1513)	VSINLQRGNFRSTNGV (SEQ ID NO: 1604)	FRSTNGVGQAATADVNT (SEQ ID NO: 1617)
AAV3	VANNLQSSN (SEQ ID NO: 1501)	FRSTNGV (SEQ ID NO: 1393)	SAPTTGTVNH (SEQ ID NO: 1514)	VANNLQSSNFRSTNGV (SEQ ID NO: 1605)	FRSTNGVSAPTTGTVNH (SEQ ID NO: 1618)
AAV4	LYGGNSNS (SEQ ID NO: 1502)	FRSTNGV (SEQ ID NO: 1393)	NLPTVDRLTALG (SEQ ID NO: 1515)	LYGGNSNSFRSTNGV (SEQ ID NO: 1606)	FRSTNGVNLPTVDRLTALG (SEQ ID NO: 1619)
AAV5	MATNNQSST (SEQ ID NO: 1503)	FRSTNGV (SEQ ID NO: 1393)	TAPATGTYNLQE (SEQ ID NO: 1516)	MATNNQSSTFRSTNGV (SEQ ID NO: 1607)	FRSTNGVTAPATGTYNLQE (SEQ ID NO: 1620)
AAV6	VAVNLMSSS (SEQ ID NO: 1504)	FRSTNGV (SEQ ID NO: 1393)	TDPATGDVHVMG (SEQ ID NO: 1517)	VAVNLMSSSFRSTNGV (SEQ ID NO: 1608)	FRSTNGVTDPATGDVHVMG (SEQ ID NO: 1621)
AAV7	VSSNLQAAAN (SEQ ID NO: 1505)	FRSTNGV (SEQ ID NO: 1393)	TAAQTQVVNNQG (SEQ ID NO: 1518)	VSSNLQAAANFRSTNGV (SEQ ID NO: 1609)	FRSTNGVTAQTQVVNNQG (SEQ ID NO: 1622)
AAV8	VADNLQQQN (SEQ ID NO: 1506)	FRSTNGV (SEQ ID NO: 1393)	TAPQIGTVNSQG (SEQ ID NO: 1519)	VADNLQQQNFRSTNGV (SEQ ID NO: 1610)	FRSTNGVTAPQIGTVNSQG (SEQ ID NO: 1623)
AAV9	VATNHQSAQ (SEQ ID NO: 1507)	FRSTNGV (SEQ ID NO: 1393)	AQAQTGWVQNQG (SEQ ID NO: 1520)	VATNHQSAQFRSTNGV (SEQ ID NO: 1611)	FRSTNGVAQAQTGWVQNQG (SEQ ID NO: 1624)
AAV10	VADNLQQAN (SEQ ID NO: 1508)	FRSTNGV (SEQ ID NO: 1393)	TGPVGNVNSQG (SEQ ID NO: 1521)	VADNLQQANFRSTNGV (SEQ ID NO: 1612)	FRSTNGVTGPVGNVNSQG (SEQ ID NO: 1625)
AAV11	IADNNQNAT (SEQ ID NO: 1509)	FRSTNGV (SEQ ID NO: 1393)	TAPITGNVTAMG (SEQ ID NO: 1522)	IADNNQNATFRSTNGV (SEQ ID NO: 1613)	FRSTNGVTAPITGNVTAMG (SEQ ID NO: 1626)
AAV12	IADNNQNAT (SEQ ID NO: 1510)	FRSTNGV (SEQ ID NO: 1393)	TAPHIANLDAMG (SEQ ID NO: 1523)	IADNNQNATFRSTNGV (SEQ ID NO: 1614)	FRSTNGVTAPHIANLDAMG (SEQ ID NO: 1627)
AAV13	VSNNLQNSN (SEQ ID NO: 1511)	FRSTNGV (SEQ ID NO: 1393)	AGPTTGTVNHQ (SEQ ID NO: 1524)	VSNNLQNSNFRSTNGV (SEQ ID NO: 1615)	FRSTNGVAGPTTGTVNHQ (SEQ ID NO: 1628)

SEROTYPE	UPSTREAM SEQUENCE	7-mer INSERTION ID NO: (SEQ ID NO: 1394)	DOWNSTREAM SEQUENCE	UPSTREAM SEQUENCE + 7-mer INSERTION (SEQ ID NO:)	7-mer INSERTION+ DOWNSTREAM SEQUENCE (SEQ ID NO:)
AAV1	VAVIFQSSS (SEQ ID NO: 1499)	FVSTNGV (SEQ ID NO: 1394)	DPATGDVHA (SEQ ID NO: 1512)	VAVIFQSSSFVSTNGV (SEQ ID NO: 1629)	FVSTNGVDPATGDVHA (SEQ ID NO: 1642)
AAV2	VSINLQRGN (SEQ ID NO: 1500)	FVSTNGV (SEQ ID NO: 1394)	GQAATADVNT (SEQ ID NO: 1513)	VSINLQRGNFVSTNGV (SEQ ID NO: 1630)	FVSTNGVGQAATADVNT (SEQ ID NO: 1643)
AAV3	VANNLQSSN (SEQ ID NO: 1501)	FVSTNGV (SEQ ID NO: 1394)	SAPTTGTVNH (SEQ ID NO: 1514)	VANNLQSSNFVSTNGV (SEQ ID NO: 1631)	FVSTNGVSAPTTGTVNH (SEQ ID NO: 1644)
AAV4	LYGGNSNS (SEQ ID NO: 1502)	FVSTNGV (SEQ ID NO: 1394)	NLPTVDRLTALG (SEQ ID NO: 1515)	LYGGNSNSFVSTNGV (SEQ ID NO: 1632)	FVSTNGVNLPTVDRLTALG (SEQ ID NO: 1645)
AAV5	MATNNQSSST (SEQ ID NO: 1503)	FVSTNGV (SEQ ID NO: 1394)	TAPATGTYNLQE (SEQ ID NO: 1516)	MATNNQSSSFVSTNGV (SEQ ID NO: 1633)	FVSTNGVTAPATGTYNLQE (SEQ ID NO: 1646)
AAV6	VAVNLMSSS (SEQ ID NO: 1504)	FVSTNGV (SEQ ID NO: 1394)	TDPATGDVHVVMG (SEQ ID NO: 1517)	VAVNLMSSSFVSTNGV (SEQ ID NO: 1634)	FVSTNGVTDPATGDVHVVMG (SEQ ID NO: 1647)
AAV7	VSSNLQAAAN (SEQ ID NO: 1505)	FVSTNGV (SEQ ID NO: 1394)	TAAQTQVVNNQG (SEQ ID NO: 1518)	VSSNLQAAANFVSTNGV (SEQ ID NO: 1635)	FVSTNGVTAAQTQVVNNQG (SEQ ID NO: 1648)
AAV8	VADNLQQQN (SEQ ID NO: 1506)	FVSTNGV (SEQ ID NO: 1394)	TAPQIGTVNSQG (SEQ ID NO: 1519)	VADNLQQQNFVSTNGV (SEQ ID NO: 1636)	FVSTNGVTAPQIGTVNSQG (SEQ ID NO: 1649)
AAV9	VATNHQSAQ (SEQ ID NO: 1507)	FVSTNGV (SEQ ID NO: 1394)	AQAQTGWVQNQG (SEQ ID NO: 1520)	VATNHQSAQFVSTNGV (SEQ ID NO: 1637)	FVSTNGVAQAQTGWVQNQG (SEQ ID NO: 1650)
AAV10	VADNLQQAN (SEQ ID NO: 1508)	FVSTNGV (SEQ ID NO: 1394)	TGPVGNVNSQG (SEQ ID NO: 1521)	VADNLQQANFVSTNGV (SEQ ID NO: 1638)	FVSTNGVTGPVGNVNSQG (SEQ ID NO: 1651)
AAV11	IADNNQNAT (SEQ ID NO: 1509)	FVSTNGV (SEQ ID NO: 1394)	TAPITGNVTAMG (SEQ ID NO: 1522)	IADNNQNATFVSTNGV (SEQ ID NO: 1639)	FVSTNGVTAPITGNVTAMG (SEQ ID NO: 1652)
AAV12	IADNNQNAT (SEQ ID NO: 1510)	FVSTNGV (SEQ ID NO: 1394)	TAPHIANLDAMG (SEQ ID NO: 1523)	IADNNQNATFVSTNGV (SEQ ID NO: 1640)	FVSTNGVTAPHIANLDAMG (SEQ ID NO: 1653)
AAV13	VSNNLQNSN (SEQ ID NO: 1511)	FVSTNGV (SEQ ID NO: 1394)	AGPTTGTVNHQ (SEQ ID NO: 1524)	VSNNLQNSNFVSTNGV (SEQ ID NO: 1641)	FVSTNGVAGPTTGTVNHQ (SEQ ID NO: 1654)

SEROTYPE	UPSTREAM SEQUENCE	7-mer INSERTION	DOWNSTREAM SEQUENCE	UPSTREAM SEQUENCE + 7-mer INSERTION (SEQ ID NO.)	7-mer INSERTION+ DOWNSTREAM SEQUENCE (SEQ ID NO.)
AAV1	VAVIFQSSS (SEQ ID NO: 1499)	FZ1STNGZ2 (SEQ ID NO: 1395)	DPATGDVHA (SEQ ID NO: 1512)	VAVIFQSSSFZ1STNGZ2 (SEQ ID NO: 1655)	FZ1STNGZ2DPATGDVHA (SEQ ID NO: 1668)
AAV2	VSINLQRGN (SEQ ID NO: 1500)	FZ1STNGZ2 (SEQ ID NO: 1395)	GQAATADVNT (SEQ ID NO: 1513)	VSINLQRGNFZ1STNGZ2 (SEQ ID NO: 1656)	FZ1STNGZ2GQAATADVNT (SEQ ID NO: 1669)
AAV3	VANNLQSSN (SEQ ID NO: 1501)	FZ1STNGZ2 (SEQ ID NO: 1395)	SAPTTGTVNH (SEQ ID NO: 1514)	VANNLQSSNFZ1STNGZ2 (SEQ ID NO: 1657)	FZ1STNGZ2SAPTTGTVNH (SEQ ID NO: 1670)
AAV4	LYGGNSNS (SEQ ID NO: 1502)	FZ1STNGZ2 (SEQ ID NO: 1395)	NLPTVDRLTALG (SEQ ID NO: 1515)	LYGGNSNSFZ1STNGZ2 (SEQ ID NO: 1658)	FZ1STNGZ2NLPTVDRLTALG (SEQ ID NO: 1671)
AAV5	MATNNQSS (SEQ ID NO: 1503)	FZ1STNGZ2 (SEQ ID NO: 1395)	TAPATGTYNLQE (SEQ ID NO: 1516)	MATNNQSSFZ1STNGZ2 (SEQ ID NO: 1659)	FZ1STNGZ2TAPATGTYNLQE (SEQ ID NO: 1672)
AAV6	VAVNLMSSS (SEQ ID NO: 1504)	FZ1STNGZ2 (SEQ ID NO: 1395)	TDPATGDVHVMG (SEQ ID NO: 1517)	VAVNLMSSSFZ1STNGZ2 (SEQ ID NO: 1660)	FZ1STNGZ2TDPATGDVHVMG (SEQ ID NO: 1673)
AAV7	VSSNLQAAAN (SEQ ID NO: 1505)	FZ1STNGZ2 (SEQ ID NO: 1395)	TAAQTQVVNNQG (SEQ ID NO: 1518)	VSSNLQAAANFZ1STNGZ2 (SEQ ID NO: 1661)	FZ1STNGZ2TAAQTQVVNNQG (SEQ ID NO: 1674)
AAV8	VADNLQQQN (SEQ ID NO: 1506)	FZ1STNGZ2 (SEQ ID NO: 1395)	TAPQIGTVNSQG (SEQ ID NO: 1519)	VADNLQQQNFZ1STNGZ2 (SEQ ID NO: 1662)	FZ1STNGZ2TAPQIGTVNSQG (SEQ ID NO: 1675)
AAV9	VATNHQSAQ (SEQ ID NO: 1507)	FZ1STNGZ2 (SEQ ID NO: 1395)	AQAQTGWVQNQG (SEQ ID NO: 1520)	VATNHQSAQFZ1STNGZ2 (SEQ ID NO: 1663)	FZ1STNGZ2AQAQTGWVQNQG (SEQ ID NO: 1676)
AAV10	VADNLQQAN (SEQ ID NO: 1508)	FZ1STNGZ2 (SEQ ID NO: 1395)	TGPIVGNVNSQG (SEQ ID NO: 1521)	VADNLQQANFZ1STNGZ2 (SEQ ID NO: 1664)	FZ1STNGZ2TGPIVGNVNSQG (SEQ ID NO: 1677)
AAV11	IADNNQNAT (SEQ ID NO: 1509)	FZ1STNGZ2 (SEQ ID NO: 1395)	TAPITGNVTAMG (SEQ ID NO: 1522)	IADNNQNATFZ1STNGZ2 (SEQ ID NO: 1665)	FZ1STNGZ2TAPITGNVTAMG (SEQ ID NO: 1678)
AAV12	IADNNQNAT (SEQ ID NO: 1510)	FZ1STNGZ2 (SEQ ID NO: 1395)	TAPHIANLDAMG (SEQ ID NO: 1523)	IADNNQNATFZ1STNGZ2 (SEQ ID NO: 1666)	FZ1STNGZ2TAPHIANLDAMG (SEQ ID NO: 1679)
AAV13	VSNNLQNSN (SEQ ID NO: 1511)	FZ1STNGZ2 (SEQ ID NO: 1395)	AGPTTGTVNHQ (SEQ ID NO: 1524)	VSNNLQNSNFZ1STNGZ2 (SEQ ID NO: 1667)	FZ1STNGZ2AGPTTGTVNHQ (SEQ ID NO: 1680)

Z1 is selected from the group consisting of A, D, H, N, Q, and S; Z2 is selected from the group consisting of K and R.

SEROTYPE	UPSTREAM SEQUENCE	7-mer INSERTION	DOWNSTREAM SEQUENCE	UPSTREAM SEQUENCE + 7-mer INSERTION (SEQ ID NO.)	7-mer INSERTION+ DOWNSTREAM SEQUENCE (SEQ ID NO.)
AAV1	VAVIFQSSS (SEQ ID NO: 1499)	FRSTNGZ3 (SEQ ID NO: 1396)	DPATGDVHA (SEQ ID NO: 1512)	VAVIFQSSSFRSTNGZ3 (SEQ ID NO: 1681)	FRSTNGZ3DPATGDVHA (SEQ ID NO: 1694)
AAV2	VSINLQRGN (SEQ ID NO: 1500)	FRSTNGZ3 (SEQ ID NO: 1396)	GQAAATADVNT (SEQ ID NO: 1513)	VSINLQRGNFRSTNGZ3 (SEQ ID NO: 1682)	FRSTNGZ3GQAAATADVNT (SEQ ID NO: 1695)
AAV3	VANNLQSSN (SEQ ID NO: 1501)	FRSTNGZ3 (SEQ ID NO: 1396)	SAPTTGTVNH (SEQ ID NO: 1514)	VANNLQSSNFRSTNGZ3 (SEQ ID NO: 1683)	FRSTNGZ3SAPTTGTVNH (SEQ ID NO: 1696)
AAV4	LYGGNSNS (SEQ ID NO: 1502)	FRSTNGZ3 (SEQ ID NO: 1396)	NLPTVDRLTALG (SEQ ID NO: 1515)	LYGGNSNSFRSTNGZ3 (SEQ ID NO: 1684)	FRSTNGZ3NLPTVDRLTALG (SEQ ID NO: 1697)
AAV5	MATNNQSST (SEQ ID NO: 1503)	FRSTNGZ3 (SEQ ID NO: 1396)	TAPATGTYNLQE (SEQ ID NO: 1516)	MATNNQSSTFRSTNGZ3 (SEQ ID NO: 1685)	FRSTNGZ3TAPATGTYNLQE (SEQ ID NO: 1698)
AAV6	VAVNLMSSS (SEQ ID NO: 1504)	FRSTNGZ3 (SEQ ID NO: 1396)	TDPATGDVHVMG (SEQ ID NO: 1517)	VAVNLMSSSFRSTNGZ3 (SEQ ID NO: 1686)	FRSTNGZ3TDPATGDVHVMG (SEQ ID NO: 1699)
AAV7	VSSNLQAAAN (SEQ ID NO: 1505)	FRSTNGZ3 (SEQ ID NO: 1396)	TAAQTQVVNNQG (SEQ ID NO: 1518)	VSSNLQAAANFRSTNGZ3 (SEQ ID NO: 1687)	FRSTNGZ3TAAQTQVVNNQG (SEQ ID NO: 1700)
AAV8	VADNLQQQN (SEQ ID NO: 1506)	FRSTNGZ3 (SEQ ID NO: 1396)	TAPQIGTVNSQG (SEQ ID NO: 1519)	VADNLQQQNFRSTNGZ3 (SEQ ID NO: 1688)	FRSTNGZ3TAPQIGTVNSQG (SEQ ID NO: 1701)
AAV9	VATNHQSAQ (SEQ ID NO: 1507)	FRSTNGZ3 (SEQ ID NO: 1396)	AQAQTGWVQNQG (SEQ ID NO: 1520)	VATNHQSAQFRSTNGZ3 (SEQ ID NO: 1689)	FRSTNGZ3AQAQTGWVQNQG (SEQ ID NO: 1702)
AAV10	VADNLQQAN (SEQ ID NO: 1508)	FRSTNGZ3 (SEQ ID NO: 1396)	TGPVGNVNSQG (SEQ ID NO: 1521)	VADNLQQANFRSTNGZ3 (SEQ ID NO: 1690)	FRSTNGZ3TGPVGNVNSQG (SEQ ID NO: 1703)
AAV11	IADNNQNAT (SEQ ID NO: 1509)	FRSTNGZ3 (SEQ ID NO: 1396)	TAPITGNVTAMG (SEQ ID NO: 1522)	IADNNQNATFRSTNGZ3 (SEQ ID NO: 1691)	FRSTNGZ3TAPITGNVTAMG (SEQ ID NO: 1704)
AAV12	IADNNQNAT (SEQ ID NO: 1510)	FRSTNGZ3 (SEQ ID NO: 1396)	TAPHIANLIDAMG (SEQ ID NO: 1523)	IADNNQNATFRSTNGZ3 (SEQ ID NO: 1692)	FRSTNGZ3TAPHIANLIDAMG (SEQ ID NO: 1705)
AAV13	VSNNLQNSN (SEQ ID NO: 1511)	FRSTNGZ3 (SEQ ID NO: 1396)	AGPTTGTVNHQ (SEQ ID NO: 1524)	VSNNLQNSNFRSTNGZ3 (SEQ ID NO: 1693)	FRSTNGZ3AGPTTGTVNHQ (SEQ ID NO: 1706)

Z3 is selected from the group consisting of L, M, and R.

SEROTYPE	UPSTREAM SEQUENCE	7-mer INSERTION	DOWNSTREAM SEQUENCE	UPSTREAM SEQUENCE + 7-mer INSERTION (SEQ ID NO.)	7-mer INSERTION+ DOWNSTREAM SEQUENCE (SEQ ID NO.)
AAV1	VAVIFQSSS (SEQ ID NO: 1499)	VESTNGR (SEQ ID NO: 1397)	DPATGDVHA (SEQ ID NO: 1512)	VAVIFQSSSVESTNGR (SEQ ID NO: 1707)	VESTNGRDPATGDVHA (SEQ ID NO: 1720)
AAV2	VSINLQRGN (SEQ ID NO: 1500)	VESTNGR (SEQ ID NO: 1397)	GQAATADVNT (SEQ ID NO: 1513)	VSINLQRGNVESTNGR (SEQ ID NO: 1708)	VESTNGRGQAATADVNT (SEQ ID NO: 1721)
AAV3	VANNLQSSN (SEQ ID NO: 1501)	VESTNGR (SEQ ID NO: 1397)	SAPTTGTVNH (SEQ ID NO: 1514)	VANNLQSSNVESTNGR (SEQ ID NO: 1709)	VESTNGRSAPTTGTVNH (SEQ ID NO: 1722)
AAV4	LYGGNSNS (SEQ ID NO: 1502)	VESTNGR (SEQ ID NO: 1397)	NLPTVDRLTALG (SEQ ID NO: 1515)	LYGGNSNSVESTNGR (SEQ ID NO: 1710)	VESTNGRNLPTVDRLTALG (SEQ ID NO: 1723)
AAV5	MATNNQSST (SEQ ID NO: 1503)	VESTNGR (SEQ ID NO: 1397)	TAPATGTYNLQE (SEQ ID NO: 1516)	MATNNQSSTVESTNGR (SEQ ID NO: 1711)	VESTNGRTAPATGTYNLQE (SEQ ID NO: 1724)
AAV6	VAVNLMSSS (SEQ ID NO: 1504)	VESTNGR (SEQ ID NO: 1397)	TDPATGDVHVVMG (SEQ ID NO: 1517)	VAVNLMSSSVESTNGR (SEQ ID NO: 1712)	VESTNGRTDPATGDVHVVMG (SEQ ID NO: 1725)
AAV7	VSSNLQAAAN (SEQ ID NO: 1505)	VESTNGR (SEQ ID NO: 1397)	TAAQTQVVNNQG (SEQ ID NO: 1518)	VSSNLQAAANVESTNGR (SEQ ID NO: 1713)	VESTNGRTAAQTQVVNNQG (SEQ ID NO: 1726)
AAV8	VADNLQQQN (SEQ ID NO: 1506)	VESTNGR (SEQ ID NO: 1397)	TAPQIGTVNSQG (SEQ ID NO: 1519)	VADNLQQQNVESTNGR (SEQ ID NO: 1714)	VESTNGRTAPQIGTVNSQG (SEQ ID NO: 1727)
AAV9	VATNHQSAQ (SEQ ID NO: 1507)	VESTNGR (SEQ ID NO: 1397)	AQAQTGWVQNQG (SEQ ID NO: 1520)	VATNHQSAQVESTNGR (SEQ ID NO: 1715)	VESTNGRAQAQTGWVQNQG (SEQ ID NO: 1728)
AAV10	VADNLQQAN (SEQ ID NO: 1508)	VESTNGR (SEQ ID NO: 1397)	TGPVGNVNSQG (SEQ ID NO: 1521)	VADNLQQANVESTNGR (SEQ ID NO: 1716)	VESTNGRTGPVGNVNSQG (SEQ ID NO: 1729)
AAV11	IADNNQNAT (SEQ ID NO: 1509)	VESTNGR (SEQ ID NO: 1397)	TAPITGNVTAMG (SEQ ID NO: 1522)	IADNNQNATVESTNGR (SEQ ID NO: 1717)	VESTNGRTAPITGNVTAMG (SEQ ID NO: 1730)
AAV12	IADNNQNAT (SEQ ID NO: 1510)	VESTNGR (SEQ ID NO: 1397)	TAPHIANLDAMG (SEQ ID NO: 1523)	IADNNQNATVESTNGR (SEQ ID NO: 1718)	VESTNGRTAPHIANLDAMG (SEQ ID NO: 1731)
AAV13	VSNNLQNSN (SEQ ID NO: 1511)	VESTNGR (SEQ ID NO: 1397)	AGPTTGTVNHQ (SEQ ID NO: 1524)	VSNNLQNSNVESTNGR (SEQ ID NO: 1719)	VESTNGRAGPTTGTVNHQ (SEQ ID NO: 1732)

[0114] In an embodiment, in addition to the n-mer motif(s), the transferrin receptor binding modification can include a polypeptide, a polynucleotide, a lipid, a polymer, a sugar, or a combination thereof.

[0115] The engineered viral capsid and/or capsid proteins can be encoded by one or more engineered viral capsid polynucleotides. In an embodiment, the engineered viral capsid polynucleotide is an engineered AAV capsid polynucleotide or engineered adenovirus capsid polynucleotide. In an embodiment, an engineered viral capsid polynucleotide (e.g., an engineered AAV capsid polynucleotide or engineered adenovirus capsid polynucleotide) can include a 3' polyadenylation signal. The polyadenylation signal can be an SV40 polyadenylation signal.

[0116] In an embodiment, the engineered polynucleotide can be included in a polynucleotide that is configured to express the engineered capsid in a host cell system for production of viral particles. The host cell system may also include a construct that expresses a recombinant viral genome that comprises a transgene encoding a polypeptide or nucleic acid operably linked to one or more regulatory sequences that promote expression of the transgene in a target cell, including a recombinant AAV genome where the transgene and regulatory sequences are flanked by AAV ITR sequences.

[0117] In an embodiment, the engineered AAV capsid encoding polynucleotide can be included in a polynucleotide that is configured to express the engineered capsid in a host cell system for production of AAV viral particles. The host cell system may also include a construct that expresses a recombinant AAV viral genome that comprises a transgene encoding a polypeptide or nucleic acid operably linked to one or more regulatory sequences that promote expression of the transgene in a target cell, including a recombinant AAV genome where the transgene and regulatory sequences are flanked by AAV ITR sequences. In an embodiment, the engineered AAV capsid encoding polynucleotide can be operably coupled to a polyadenylation tail. In an embodiment, the polyadenylation tail can be an SV40 polyadenylation tail. In an embodiment, the AAV capsid encoding polynucleotide can be operably coupled to a promoter. In an embodiment, the regulatory sequence that regulates the expression of the transgene is a promoter and can be a tissue- or cell type-specific promoter. In an embodiment, the tissue-specific promoter is specific for muscle (e.g., cardiac, skeletal, and/or smooth muscle), neurons or other nervous system cells (e.g., astrocytes, glial cells, Schwann cells, ependymal cells, pericyte, oligodendrocyte, oligodendrocyte progenitor), specific neuronal subtype (e.g., dopaminergic

neuron; Purkinje Cell; Parvalbumin, somatostatin, VIP inhibitory neuron; medium spiny neuron, Pyramidal neuron, motor neuron, etc.), endothelial cell, fat, spleen, liver, kidney, immune cells, synovial fluid cells, skin cells, cartilage, tendons, connective tissue, bone, pancreas, adrenal gland, blood cell, bone marrow cells, placenta, endothelial cells, and combinations thereof. In an embodiment, the promoter can be a constitutive promoter. Suitable tissue specific promoters and constitutive promoters are discussed elsewhere herein and are generally known in the art and can be commercially available. Suitable neuronal tissue/cell specific promoters include, but are not limited to, GFAP promoter (astrocytes), SYN1 promoter (neurons), and NSE/RU5' (mature neurons). In an embodiment, the regulatory sequence that regulates the expression of the transgene is a promoter and can be a cell state regulating promoter or drug inducible promoter.

[0118] A neuron-specific promoter refers to a promoter that, when administered e.g., peripherally, directly into the central nervous system (CNS), or delivered to neuronal cells, including in vitro, ex vivo, or in vivo, preferentially drives or regulates expression of an operatively-linked transgene in neurons as compared to expression in non-neuronal cells. Non-limiting example of tissue-specific expression elements for neurons include neuron-specific enolase (NSE) (see, e.g., EMBL HSEN02, X51956); an aromatic amino acid decarboxylase (AADC) promoter; a neurofilament promoter (see, e.g., GenBank HUMNFL, L04147); a synapsin promoter (see, e.g., GenBank HUMSYNIB, M55301); a thy-1 promoter (see, e.g., Chen et al., (1987) *Cell*, 51 :7-19; Llewellyn et al. (2010) *Nat. Med.*, 16(10):1161-1 166); a serotonin receptor promoter (see, e.g., GenBank S62283); a tyrosine hydroxylase promoter (TH) (see, e.g., Oh et al., (2009) *Gene Ther.*, 16:437; Sasaoka et al., (1992) *Mol. Brain Res.*, 16:274; Boundy et al., (1998) *J. Neurosci.*, 18:9989; and Kaneda et al., (1991) *Neuron*, 6:583-594); a methyl-CpG binding protein 2 (MeCP2) promoter, an optimized methyl- CpG binding protein 2 (MeCP2) promoter (the published International Patent Application No. WO2020180928, the content of which is incorporated by reference herein in its entirety), a Ca²⁺-calmodulin-dependent protein kinase II- α (CaMKII α) promoter (see, e.g., Mayford et al., (1996) *Proc. Natl. Acad. Sci. USA*, 93:13250; and Casanova et al., (2001) *Genesis*, 31 :37); a GnRH promoter (see, e.g., Radovick et al., (1991) *Proc. Natl. Acad. Sci. USA*, 88:3402- 3406); an L7 promoter (see, e.g., Oberdick et al., (1990) *Science*, 248:223-226); a DNMT promoter (see, e.g., Badge et al., (1988) *Proc. Natl. Acad. Sci. USA*, 85:3648-3652); an enkephalin promoter (see, e.g., Comb et al., (1988) *EMBO J.*, 17:3793-3805); a myelin basic protein (MBP) promoter; a CMV enhancer/platelet-derived growth factor-p

promoter (see, e.g., Liu et al., (2004) *Gene Ther.*, 11 :52-60); and the like. In some aspects, a portion of or all the minimal human synapsin 1 promoter (SYN) can be used (Kugler et al., (2003) *Gene Ther.*, 10(4): 337-47; Thiel et al, (1991) *Proc. Natl. Acad. Sci. USA*, 88(8) 3431 -5; Castle et al., (2016) *Methods Mol. Biol.*, 1382: 133-49; McLean et al., (2014) *Neurosci. Lett.*, 576: 73-78; Kugler et al., (2003) *Virology*, 311 (1): 89-95). In other aspects, the neural-specific promoter can be mGluR2, NFL, NFH, nß2, PPE, Enk and EAAT2 promoters. A non-limiting example of a tissue-specific expression elements for astrocytes include the glial fibrillary acidic protein (GFAP) and EAAT2 promoters. A non-limiting example of a tissue-specific expression element for oligodendrocytes include the myelin basic protein (MBP) promoter. In certain aspects, a neuronal promoter can include a neuronal enhancer to direct expression to specific regions of the brain (see, for example, published U.S. Patent Application No.2019/0247516, the content of which is incorporated by reference herein in its entirety). In one aspect, the promoter can be a fugu SST (somatostatin) promoter (Nathanson, et al. *Frontiers in Neural Circuits* 3: 19). Examples of retinal-specific promoters include, but are not limited to, NA65p (RPE cells), Nefh (ganglion cells), hGRK1 (rod and cone photoreceptor cells), hRLBP1 (Müller glial cells and RPE cells), human RHO (rhodopsin), human rhodopsin kinase (RHOK/GRK1) (an exemplary list of retina cell-specific promoters can be found in Buck et al. (2020) *International Journal of Molecular Sciences* 21 (12), the content of which is incorporated by reference in its entirety). Non-limiting examples of liver promoters include hAAT and TBG. Non-limiting examples of skeletal muscle promoters include Desmin, MCK and C5-12. Additional exemplary tissue-specific promoters can be found in the TiProD (Tissue specific promoter database webpage tiprod.bioinf.med.uni-goettingen.de).

[0119] In other aspects, a promoter can be an inducible promoter (i.e., a promoter whose activity is controlled by an external stimulus, e.g., the presence of a particular temperature, compound, or protein). In some aspects, a promoter may be a temporally restricted promoter that drives expression depending on the temporal context in which the promoter is found. For example, a temporally restricted promoter may drive expression only during specific stages of a biological process. Prokaryotic (Gossen et al. *TIBS* 18: 471475, 1993) and insect regulatory systems (No et al. *Proc. Natl. Acad. Sci. USA* 93: 3346-3351, 1996) have been adapted to construct gene switches that function in mammalian cells. Since inducer molecules are not expected to have targets in mammalian cells, the possibility of interference with cellular processes is reduced. Of the prokaryotic proteins, the repressors from the lac operon (Brown, M., et al. *Cell* 49: 603- 612, 1987;

and Hu, M. C. -T. and N. Davidson Cell 48: 555-566, 1987), the tet operon (e.g., U.S. Patent No. 7,541,446, the content of which is incorporated by reference herein in its entirety) and the cumate operon (e.g., U.S. Patent No. 7,745,592, the content of which is incorporated by reference herein in its entirety) have been shown to function in mammalian cells. Many have been incorporated in eukaryotic inducible expression systems using different strategies to control activation and repression of expression. Activation of expression is mediated by a chimeric transactivator protein formed by the fusion of the bacterial repressor with an activation domain (Gossen, M. and H. Bujard, Proc. Natl. acad. sci. USA 89: 5547- 5551, 1992, and Gossen, M., et al. Science 268: 1766-1769, 1995; U.S. Patent No. 7,745,592, the contents of which are incorporated by reference herein in their entireties). The transactivator can activate transcription when bound to its DNA recognition sequence placed upstream of the minimal promoter. The ability of the activator to bind DNA is dependent on the presence/absence of the inducer molecule (e.g., doxycycline or cumate depending on the inducible system being used). Repression of expression is mediated by the repressor bound to operator sites placed downstream of the minimal promoter in the absence of inducer and repression is relieved on the addition of the inducer (Brown, M., et al. Cell 49: 603-612, 1987). In one aspect, the promoter may be a promoter which is less than 1 kb. The promoter may have a length of 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800 or more than 800. The promoter may have a length between 200-300, 200-400, 200-500, 200-600, 200-700, 200-800, 300-400, 300-500, 300-600, 300-700, 300-800, 400-500, 400-600, 400-700, 400-800, 500-600, 500-700, 500-800, 600-700, 600-800 or 700-800 nucleotides. In one aspect, the promoter can be a pol III-dependent promoter, e.g., a U6 snRNA or H1-RNA promoter, for the expression of non-coding RNAs including, but not limited to, U snRNAs or miRNAs. In another aspect, the promoter can be a polymerase II U snRNA-dependent promoter, e.g., a human U1 snRNA gene and of its promoter and terminator regions (see, for example, published U.S. Patent No. 7,947,823, the content of which is incorporated by reference herein in its entirety).

[0120] Additional promoters can be found in Wang, E. T.-S. & Poukalov, K. K. Methods and compositions to confer regulation to gene therapy cargoes by heterologous use of alternative splicing cassettes. *World Patent* (2022); Boyne, A. R., Danos, O. F., Volles, M. J. & Guo, X.

Regulation of gene expression by aptamer-mediated modulation of alternative splicing. *US Patent* (2022); Ranum, P. T., Monteys, A. M., Hundley, A. A. & Davidson, B. L. Compositions and methods for inducible alternative splicing regulation of gene expression. *World Patent* (2021); Monteys, A. M. *et al.* Regulated control of gene therapies by drug-induced splicing. *Nature* 1–5 (2021); Doshi, A., Sadeghi, F., Varadarajan, N. & Cirino, P. C. Small-molecule inducible transcriptional control in mammalian cells. *Crit. Rev. Biotechnol.* **40**, 1131–1150 (2020); Monteys, A. M. *et al.* Regulated control of gene therapies with a drug induced switch. 2020.02.21.956664 (2020) doi:10.1101/2020.02.21.956664.; Davidson, B. L., Monteys, A. M., Al HUNDLEY, A. & Ranum, P. T. ALTERNATIVE SPLICING REGULATION OF GENE EXPRESSION AND THERAPEUTIC METHODS. *PCT* (2020); Boyne, A. R., Olivier DANOS, F., Michael VOLLES, J. & Xuecui, G. U. O. Regulation of gene expression by aptamer-mediated modulation of alternative splicing. *World Patent* (2016) ; REGULATABLE EXPRESSION SYSTEMS. *World Patent*; BERGLUND, John, Andrew DELGADO, Elizabeth JENQUIN, Jana, Rose WANG, Eric, Tzy-Shi. GENE THERAPY VECTORS. *World Patent*, incorporated herein by reference in their entirety.

Further Capsid Modifications

[0121] In an embodiment, the viral capsid protein may comprise one or more mutations relative to wild type. In an example embodiment, the one or more mutations comprise a K449R substitution in a capsid polypeptide of AAV 9 20507, or a substitution in an analogous position of a capsid polypeptide from AAV9, AAV9 K449R (or K449R AAV9), AAV1, AAVrh10, AAV-DJ, AAV-DJ8, AAV5, AAV-PHP.B (PHP.B), AAV-PHP.A (PHP.A), AAVG2B-26, AAVG2B-13, AAVTH1.1-32, AAVTH1.1- 35, AAV-PHP.B2 (PHP.B2), AAV-PHP.B3 (PHP.B3), AAV-PHP.N/PHP.B-DGT, AAV-PHP.B-EST, AAV-PHP.B-GGT, AAV-PHP.B-ATP, AAV-PHP.B-ATT-T, AAV-PHP.B- DGT-T, AAV-PHP.B-GGT-T, AAV-PHP.B-SGS, AAV-PHP.B-AQP, AAV-PHP.B-QQP, AAV-PHP.B-SNP(3), AAV-PHP.B-SNP, AAV-PHP.B-QGT, AAV-PHP.B-NQT, AAV-PHP.B- EGS, AAV-PHP.B-SGN, AAV-PHP.B-EGT, AAV-PHP.B-DST, AAV-PHP.B-DST, AAV-PHP.B-STP, AAV-PHP.B-PQP, AAV-PHP.B-SQP, AAV-PHP.B-QLP, AAV-PHP.B- TMP, AAV-PHP.B-TTP, AAV-PHP.S/G2A12, AA VG2A 15/G2A3 (G2A3), AAVG2B4 (G2B4), AAVG2B5 (G2B5), PHP.S, AAV2, AAV2G9, AAV3, AAV3a, AAV3b, AAV3-3, AAV4, AAV4-4, AAV6, AAV6.1, AAV6.2, AAV6.1.2, AAV7, AAV7.2, AAV8, AAV9.11, AAV9.13, AAV9.16, AAV9.24, AAV9.45, AAV9.47, AAV9.61, AAV9.68, AAV9.84,

AAV9.9, AAV 10, AAV11, AAV 12, AAV16.3, AAV24.1, AAV27.3, AAV42.12, AAV42- 1b, AAV42-2, AAV42-3a, AAV42-3b, AAV42-4, AAV42-5a, AAV42-5b, AAV42-6b, AAV42-8, AAV42-10, AAV42-11, AAV42-12, AAV42-13, AAV42-15, AAV42-aa, AAV43-1, AAV43-12, AAV43-20, AAV43-21, AAV43-23, AAV43-25, AAV43-5, AAV44.1, AAV44.2, AAV44.5, AAV223.1, AAV223.2, AAV223.4, AAV223.5, AAV223.6, AAV223.7, AAV1-7/rh.48, AAV1-8/rh.49, AAV2-15/rh.62, AAV2-3/rh.61, AAV2-4/rh.50, AAV2-5/rh.5 1, AAV3. 1/hu.6, AAV3.1/hu.9, AAV3-9/rh.52, AAV3-1 1/rh.53, AAV4- 8/rl 1.64, AAV4-9/rh.54, AAV4-19/rh.55, AAV5-3/rh.57, AAV5-22/rh.58, AAV7.3/hu.7, AAV16.8/hu.10, AAV16. 12/hu.1 1, AAV29.3/bb.1, AAV29.5/bb.2, AAV106.1/hu.37, AAV1 14.3/hu.40, AAV127.2/hu.41, AAV127.5/hu.42, AAV128.3/hu.44, AAV130.4/hu.48, AAV145. 1/hu.53, AAV145.5/hu.54, AAV145.6/hu.55, AAV161.10/hu.60, AAV161.6/hu.61, AAV33.12/hu. 17, AAV33.4/hu.15, AAV33.8/hu. 16, AAV52/hu.19, AAV52. 1/hu.20, AAV58.2/hu.25, AAVA3.3, AAVA3.4, AAVA3.5, AAVA3.7, AAVC1, AAVC2, AAVC5, AAVF3, AAVF5, AAVH2, AAVrh.72, AAVhu.8, AAVrh.68, AAVrh.70, AAVpi. 1, AAVpi.3, AAVpi.2, AAVrh.60, AAVrh.44, AAVrh.65, AAVrh.55, AAVrh.47, AAVrh.69, AAVrh.45, AAVrh.59, AAVhu.12, AAVH6, AAVH-1/hu.1, AAVH-5/hu.3, AAVLG- 10/rh.40, AAVLG-4/rh.38, AAVLG-9/hu.39, AAVN721-8/rh.43, AAVCh.5, AAVCh.5R1, AAVcy.2, AAVcy.3, AAVcy.4, AAVcy.5, AAVCy.5R1, AAVCy.5R2, AAVCy.5R3, AAVCy.5R4, AAVcy.6, AAVhu. 1, AAVhu.2, AAVhu.3, AAVhu.4, AAVhu.5, AAVhu.6, AAVhu.7, AAVhu.9, AAVhu.10, AAVhu.11, AAVhu.13, AAVhu.15, AAVhu.16, AAVhu.1 7, AAVhu.18, AAVhu.20, AAVhu.21, AAVhu.22, AAVhu.23.2, AAVhu.24, AAVhu.25, AAVhu.27, AAVhu.28, AAVhu.29, AAVhu.29R, AAVhu.31, AAVhu.32, AAVhu.34, AAVhu.35, AAVhu.37, AAVhu.39, AAVhu.40, AAVhu.41, AAVhu.42, AAVhu.43, AAVhu.44, AAVhu.44R1, AAVhu.44R2, AAVhu.44R3, AAVhu.45, AAVhu.46, AAVhu.47, AAVhu.48, AAVhu.48R1, AAVhu.48R2, AAVhu.48R3, AAVhu.49, AAVhu.51, AAVhu.52, AAVhu.54, AAVhu.55, AAVhu.56, AAVhu.57, AAVhu.58, AAVhu.60, AAVhu.61, AAVhu.63, AAVhu.64, AAVhu.66, AAVhu.67, AAVhu.14/9, AAVhu. t 19, AAVrh.2, AAVrh.2R, AAVrh.8, AAVrh.8R, AAVrh.10, AAVrh.12, AAVrh.13, AAVrh. 13R, AAVrh.14, AAVrh.17, AAVrh.18, AAVrh.19, AAVrh.20, AAVrh.21, AAVrh.22, AAVrh.23, AAVrh.24, AAVrh.25, AAVrh.31, AAVrh.32, AAVrh.33, AAVrh.34, AAVrh.35, AAVrh.36, AAVrh.37, AAVrh.37R2, AAVrh.38, AAVrh.39, AAVrh.40, AAVrh.46, AAVrh.48, AAVrh.48.1, AAVrh.48. L2, AAVrh.48.2, AAVrh.49, AAVA.51, AAVrh.52, AAVrh.53, AAVrh.54, AAVrh.56, AAVrh.57, AAVrh.58,

AAVA.61, AAVrh.64, AAVA.64R1, AAVA.64R2, AAVrh.67, AAVrh.73, AAVrh.74, AAVA8R, AAVA8R A586R mutant, AAVA8R R533A mutant, AAV, BAAV, caprine AAV, bovine AAV, AAVhEl .1, AAVhErI .5, AAVhERl .14, AAVhErI .8, AAVhErI .16, AAVhErI .18, AAVhErI .35, AAVhErI .7, AAVhErI .36, AAVhEr2.29, AAVhEr2.4, AAVhEr2.16, AAVhEr2.30, AAVhEr2.31, AAVhEr2.36, AAVhERl .23, AAVhEr3.1, AAV2.5T, AAV-PAEC, AAV-LK01, AAV-LK02, AAV-LK03, AAV-LK04, AAV-LK05, AAV-LK06, AAV-LK07, AAV-LK08, AAV-LK09, AAV-LK10, AAV-LK11, AAV-LK12, AAV-LK13, AAV-LK14, AAV-LK15, AAV-LK16, AAV-LK17, AAV-LK18, AAV-LK19, AAV-PAEC2, AAV-PAEC4, AAV-PAEC6, AAV-PAEC7, AAV-PAEC8, AAV-PAEC 11, AAV-PAEC 12, AAV-2-pre-miRNA-IOL, AAV-8h, AAV-8b, AAV-h, AAV-b, AAV SM 10-2, AAV Shuffle 100-1, AAV Shuffle 100-3, AAV Shuffle 100-7, AAV Shuffle 10-2, AAV Shuffle 10-6, AAV Shuffle 10-8, AAV Shuffle 100-2, AAV SM 10-1, AAV SM 10-8, AAV SM 100-3, AAV SM 100-10, BNP61 AAV, BNP62 AAV, BNP63 AAV, AAVrh.50, AAVrh.43, AAVrh.62, AAVrh.48, AAVhu.19, AAVhu.11, AAVhu.53, AAV4-8/rh.64, AAVLG-9/hu.39, AAV54.5/hu.23, AAV54.2/hu.22, AAV54.7/hu.24, AAV54.1/hu.21, AAV54.4R/hu.27, AAV46.2/hu.28, AAV46.6/hu.29, AAVI28.1/hu.43, true type AAV (ttAAV), UPEN AAV 10, Japanese AAV 10 serotypes, AAV CBr-7.1, AAV CBr-7.10, AAV CBr-7.2, AAV CBr-7.3, AAV CBr-7.4, AAV CBr-7.5, AAV CBr-7.7, AAV CBr-7.8, AAV CBr-B7.3, AAV CBr-B7.4, AAV CBr-El, AAV CBr-E2, AAV CBr-E3, AAV CBr-E4, AAV CBr-E5, AAV CBr-e5, AAV CBr-E6, AAV CBr-E7, AAV CBr-E8, AAV CHt-1, AAV CHt-2, AAV CHt-3, AAV CHt-6.1, AAV CHt-6.10, AAV CHt-6.5, AAV CHt-6.6, AAV CHt-6.7, AAV CHt-6.8, AAV CHt-Pl, AAV CHt-P2, AAV CHt-P5, AAV CHt-P6, AAV CHt-P8, AAV CHt-P9, AAV CKd-1, AAV CKd-10, AAV CKd-2, AAV CKd-3, AAV CKd-4, AAV CKd-6, AAV CKd-7, AAV CKd-8, AAV CKd-B1, AAV CKd-B2, AAV CKd-B3, AAV CKd-B4, AAV CKd-B5, AAV CKd-B6, AAV CKd-B7, AAV CKd-B8, AAV CKd-H1, AAV CKd-H2, AAV CKd-H3, AAV CKd-H4, AAV CKd-H5, AAV CKd-H6, AAV CKd-N3, AAV CKd-N4, AAV CKd-N9, AAV CLg-F1, AAV CLg-F2, AAV CLg-F3, AAV CLg-F4, AAV CLg-F5, AAV CLg-F6, AAV CLg-F7, AAV CLg-F8, AAV CLv-1, AAV CLvl-1, AAV Clvl-10, AAV CLvl-2, AAV CLv-12, AAV CLvl-3, AAV CLv-13, AAV CLvl-4, AAV Clvl-7, AAV Clvl-8, AAV Clvl-9, AAV CLv-2, AAV CLv-3, AAV CLv-4, AAV CLv-6, AAV CLv-8, AAV CLv-D1, AAV CLv-D2, AAV CLv-D3, AAV CLv-D4, AAV CLv-D5, AAV CLv-D6, AAV CLv-D7, AAV CLv-D8, AAV CLv-El, AAV CLv-K1, AAV CLv-K3, AAV CLv-K6, AAV CLv-L4, AAV CLv-

L5, AAV CLv-L6, AAV CLv-M1, AAV CLv-M1 1, AAV CLv-M2, AAV CLv-M5, AAV CLv-M6, AAV CLv-M7, AAV CLv-M8, AAV CLv-M9, AAV CLv-R1, AAV CLv-R2, AAV CLv-R3, AAV CLv-R4, AAV CLv-R5, AAV CLv-R6, AAV CLv-R7, AAV CLv-R8, AAV CLv-R9, AAV CSp-1, AAV CSp-10, AAV CSp-1 1, AAV CSp-2, AAV CSp-3, AAV CSp-4, AAV CSp-6, AAV CSp-7, AAV CSp-8, AAV CSp-8. 10, AAV CSp-8.2, AAV CSp-8.4, AAV CSp-8.5, AAV CSp-8.6, AAV CSp-8.7, AAV CSp-8.8, AAV CSp-8.9, AAV CSp-9, AAV.hu.48R3, AAV.VR-355, AAV3B, AAV4, AAV5, AAVF1/HSC1, AAVF11/HSC11, AAVF12/HSC12, AAVF13/HSC13, AAVF14/HSC14, AAVF15/HSC15, AAVF16/HSC16, AAVF17/HSC17, AAVF2/HSC2, AAVF3/HSC3, AAVF4/HSC4, AAVF5/HSC5, AAVF6/HSC6, AAVF7/HSC7, AAVF8/HSC8, and/or AAVF9/HSC9 and variants thereof. In an example embodiment, the K449R substituted AAV capsid is selected from SEQ ID NO: 20507.

[0122] In an embodiment, the viral capsid protein may comprise additional targeting motifs that are in addition to the transferrin receptor binding modification of the present disclosure. Without being bound by theory the additional targeting moieties can be antibodies or fragments thereof. In an embodiment, the additional targeting moiety can be any molecule or composition capable of recognizing, binding, attaching to, or otherwise interacting with a binding partner that can be present on the surface of a target cell. Binding partners include, but are not limited to, nucleic acids, proteins, peptides, sugars, fats, or any combination thereof or any other molecule or molecules that are present on the surface of a target cell. In an embodiment, the binding partner is unique to a cell type or cell state or a to a group of related cell types or cell states. In an embodiment, the binding partner is a receptor, channel, or other complex present on the surface of a target cell. These additional targeting moieties can be used to target, e.g., specific cell types or cell states within those the set of target cells targeted by the transferrin receptor binding modification. As used herein, “cell state” is used to describe transient elements of a cell’s identity. Cell state can be thought of as the transient characteristic profile or phenotype of a cell. Cell states arise transiently during time-dependent processes, either in a temporal progression that is unidirectional (e.g., during differentiation, or following an environmental stimulus) or in a state vacillation that is not necessarily unidirectional and in which the cell may return to the origin state. Vacillating processes can be oscillatory (e.g., cell-cycle or circadian rhythm) or can transition between states with no predefined order (e.g., due to stochastic, or environmentally controlled, molecular events). These time-dependent processes may occur transiently within a stable cell type

(as in a transient environmental response), or may lead to a new, distinct type (as in differentiation). See e.g., Wagner et al., 2016. Nat Biotechnol. 34(11): 1145-1160.

[0123] In an embodiment, the additional targeting moiety is or includes a peptide or a polypeptide. In an embodiment, the additional targeting moiety is or includes an antibody or fragment thereof. Exemplary antibodies and fragments thereof are described in greater detail elsewhere herein, see e.g., discussion on exemplary cargos. In an embodiment, the additional targeting moiety is or includes an aptamer. In an embodiment, the additional targeting moiety is or includes a small molecule. In an embodiment, the additional targeting moiety is or includes a nucleic acid (e.g., DNA or RNA). In an embodiment, the additional targeting moiety is or includes a receptor. In an embodiment, the additional targeting moiety is or includes a receptor ligand. In an embodiment, the additional targeting moiety is or includes a carbohydrate (e.g., a sugar). In an embodiment, the additional targeting moiety is or includes a lipid. In an embodiment, the additional targeting moiety is an engineered protein scaffold. In an embodiment, the additional targeting moiety is an affibody. In an embodiment, the additional targeting moiety is an antibody mimetic. In an embodiment, the additional targeting moiety is an engineered binding protein, such as a designed ankyrin repeat proteins (DARPs) (see e.g., Plückthun et al., Annu. Rev. Pharmacol. Toxicol. (2015) 55(1): 489-511), avimers (Silverman et al., Nat. Biotechnol. (2005) 23 (12): 1556–1561 and Jeong et al. Nat. Biotechnol. (2005) 23(12): 1493-1494), or affibodies (see e.g., Nord et al., Nat. Biotechnol. (1997) 15(8):772-777). In an embodiment, the additional targeting moiety is a receptor ligand or binding protein. In an embodiment, the additional targeting moiety is attached or otherwise coupled to the capsid surface. In an embodiment, the additional targeting moiety is encoded by a vector that produces a capsid of the present invention described herein.

[0124] In an embodiment, the capsid polypeptide can be covalently modified by the covalent coupling of at least one compound comprising a lactam moiety (e.g., β -lactam) to at least one amino group of an amino acid residue of the capsid of the AAV vectors (see, for example, the published International PCT application No. PCT/EP2021/080832 and U.S. Patent No. US11382988, the contents of which are incorporated by reference herein in their entirety). In other embodiments, a ligand, e.g., an 7-mer covalently linked to a primary amino group of a capsid polypeptide via a CSNH- bond, (see, e.g., U.S. Patent No. 11,648,319, the content of which is incorporated by reference herein in its entirety).

[0125] In an embodiment, the transferrin receptor binding modification can be bound to an AAV capsid polypeptide through a specific protein: protein binding pair that forms a covalent, e.g., isopeptide, bond. For example, the SpyTag/SpyCatcher bioconjugation technology can be used to append a targeting ligand, e.g., the transferrin receptor binding modification, to the surface of a capsid protein, where it can specifically bind a transferrin receptor, expressed on the cell of interest (see, for example, the published U.S. Patent Application No. 2020/0140492, the content of which is incorporated by reference herein in its entirety). According to this embodiment, the transferrin receptor binding modification can be fused in frame to the 13 amino acid SpyTag peptide. Advantages of this approach include binding of the SpyTag - transferrin receptor binding modification to a fully assembled AAV capsid and the relative ease of testing different 7-mer transferrin receptor binding modifications using the same AAV preparation. The SpyCatcher moiety can be bound to the AAV capsid polypeptide either covalently or non-covalently, for example, using a protein binding domain-specific for one or more epitopes on the surface of the capsid polypeptide.

Engineered Vectors and Vector Systems

[0126] Also provided herein are vectors and vector systems that can encode one or more of the engineered polypeptides described herein that includes one or more of the transferrin receptor binding modifications of the present invention, including but not limited to engineered viral polynucleotides (e.g., polynucleotides encoding engineered AAV capsid proteins). In a preferred embodiment, provided herein is a vector system comprising one or more vectors encoding a transferrin receptor binding modification effective to increase transduction of central nervous system tissues (CNS), optionally further comprising a vector encoding a recombinant viral genome comprising a transgene. In preferred embodiment, the transferrin receptor binding modification encoded in the vector system binds to Transferrin Receptor (TFR1). As used in this context, engineered viral capsid polynucleotides refers to any one or more of the polynucleotides described herein encoding an engineered viral capsid as described elsewhere herein and/or polynucleotide(s) encoding one or more engineered viral capsid proteins described elsewhere herein. Further, where the vector includes an engineered viral capsid polynucleotide described herein, the vector can also be referred to and considered an engineered vector or system thereof although not specifically noted as such. In embodiments, the vector can contain one or more polynucleotides encoding one or more elements of an engineered viral capsid described herein. The vectors and systems thereof

can be useful in producing bacterial, fungal, yeast, plant cells, animal cells, and transgenic animals that can express one or more components of the engineered viral capsid, particle, or other compositions described herein. Within the scope of this disclosure are vectors containing one or more of the polynucleotide sequences described herein. One or more of the polynucleotides that are part of the engineered viral capsid and system thereof described herein can be included in a vector or vector system.

[0127] In an embodiment, a vector used in the production of the rAAVs disclosed herein comprises a rep gene and cap gene). The rep gene typically encodes Rep78, Rep68, Rep52 and Rep40 from a single ORF. These replication factors aid AAV genome replication and virion assembly. The cap gene typically encodes the three capsid proteins (i.e., virion protein 1 (VP1), VP2 and VP3) from a single ORF as well. In addition, the three capsid proteins are regulated by transcription from a start codon (ACG) and alternative splicing. The cap gene also encodes, from an in-frameshifted ORF, an assembly-activating protein (AAP). The AAP is essential for capsid assembly.

[0128] In an embodiment, the vector can include an engineered viral (e.g., AAV) capsid polynucleotide having a 3' polyadenylation signal. In an embodiment, the 3' polyadenylation is an SV40 polyadenylation signal. In an embodiment the vector does not have splice regulatory elements. In an embodiment, the vector includes one or more minimal splice regulatory elements. In an embodiment, the vector can further include a modified splice regulatory element, wherein the modification inactivates the splice regulatory element. In an embodiment, the modified splice regulatory element is a polynucleotide sequence sufficient to induce splicing, between a rep protein polynucleotide and the engineered viral (e.g., AAV) capsid protein variant polynucleotide. In an embodiment, the polynucleotide sequence can be sufficient to induce splicing is a splice acceptor or a splice donor. In an embodiment, the viral (e.g., AAV) capsid polynucleotide is an engineered viral (e.g., AAV) capsid polynucleotide as described elsewhere herein. In some embodiments, the vector does not include one or more minimal splice regulatory elements, modified splice regulatory agent, splice acceptor, and/or splice donor.

[0129] The vectors and/or vector systems can be used, for example, to express one or more of the engineered viral (e.g., AAV) capsid and/or other polynucleotides in a cell, such as a producer cell, to produce engineered viral (e.g., AAV) particles and/or other compositions (e.g., polypeptides, particles, etc.) containing an engineered viral (e.g., AAV) capsid or other

composition containing an n-mer motif of the present invention described elsewhere herein. Other uses for the vectors and vector systems described herein are also within the scope of this disclosure. In general, and throughout this specification, the term is a tool that allows or facilitates the transfer of an entity from one environment to another. In some contexts which will be appreciated by those of ordinary skill in the art, “vector” can be a term of art to refer to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. A vector can be a replicon, such as a plasmid, phage, or cosmid, into which another DNA segment may be inserted so as to bring about the replication of the inserted segment. Generally, a vector is capable of replication when associated with the proper control elements.

[0130] Vectors include, but are not limited to, nucleic acid molecules that are single-stranded, double-stranded, or partially double-stranded; nucleic acid molecules that comprise one or more free ends, no free ends (e.g., circular); nucleic acid molecules that comprise DNA, RNA, or both; and other varieties of polynucleotides known in the art. One type of vector is a “plasmid,” which refers to a circular double stranded DNA loop into which additional DNA segments can be inserted, such as by standard molecular cloning techniques. Another type of vector is a viral vector, wherein virally derived DNA or RNA sequences are present in the vector for packaging into a virus (e.g., retroviruses, replication defective retroviruses, adenoviruses, replication defective adenoviruses, and adeno-associated viruses (AAVs)). Viral vectors also include polynucleotides carried by a virus for transfection into a host cell. 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.” Common expression vectors of utility in recombinant DNA techniques are often in the form of plasmids.

[0131] Recombinant expression vectors can be composed of a nucleic acid (e.g., a polynucleotide) of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory elements, which can be selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector,

“operably linked” and “operatively-linked” are used interchangeably herein and further defined elsewhere herein. In the context of a vector, the term “operably linked” is intended to mean that the nucleotide sequence of interest is linked to the regulatory element(s) in a manner that allows for expression of the nucleotide sequence (e.g., in an in vitro transcription/translation system or in a host cell when the vector is introduced into the host cell).

[0132] In an embodiment, the vector can be a bicistronic vector. In an embodiment, a bicistronic vector can be used for one or more elements of the engineered viral (e.g., AAV) capsid system described herein. In an embodiment, expression of elements of the engineered viral (e.g., AAV) capsid system described herein can be driven by a suitable constitutive or tissue specific promoter. Where the element of the engineered viral (e.g., AAV) capsid system is an RNA, its expression can be driven by a Pol III promoter, such as a U6 promoter. In an embodiment, the two are combined.

Cell-based Vector Amplification and Expression

[0133] Vectors can be designed for expression of one or more elements of the engineered viral (e.g., AAV) capsid system or other compositions containing a transferrin receptor binding modification of the present disclosure described herein (e.g., nucleic acid transcripts, proteins, enzymes, and combinations thereof) in a suitable host cell.

[0134] In an embodiment, the suitable host cell is a prokaryotic cell. Suitable host cells include, but are not limited to, bacterial cells, yeast cells, insect cells, and mammalian cells. The vectors can be viral-based or non-viral based. In an embodiment, the suitable host cell is a eukaryotic cell. In an embodiment, the suitable host cell is a suitable bacterial cell. Suitable bacterial cells include, but are not limited to, bacterial cells from the bacteria of the species *Escherichia coli*. Many suitable strains of *E. coli* are known in the art for expression of vectors. These include, but are not limited to Pir1, Stb12, Stb13, Stb14, TOP10, XL1 Blue, and XL10 Gold. In an embodiment, the host cell is a suitable insect cell. Suitable insect cells include those from *Spodoptera frugiperda*. Suitable strains of *S. frugiperda* cells include, but are not limited to, Sf9 and Sf21. In an embodiment, the host cell is a suitable yeast cell. In an embodiment, the yeast cell can be from *Saccharomyces cerevisiae*. In an embodiment, the host cell is a suitable mammalian cell. Many types of mammalian cells have been developed to express vectors. Suitable mammalian cells include, but are not limited to, HEK293, Chinese Hamster Ovary Cells (CHOs), mouse myeloma cells, HeLa, U2OS, A549, HT1080, CAD, P19, NIH 3T3, L929, N2a, MCF-7, Y79, SO-

Rb50, HepG 2, DIKX-X11, J558L, Baby hamster kidney cells (BHK), and chicken embryo fibroblasts (CEFs). Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990).

[0135] In an embodiment, the vector can be a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec1 (Baldari, et al., 1987. EMBO J. 6: 229-234), pMFa (Kuijan and Herskowitz, 1982. Cell 30: 933-943), pJRY88 (Schultz et al., 1987. Gene 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Calif.). As used herein, a "yeast expression vector" refers to a nucleic acid that contains one or more sequences encoding an RNA and/or polypeptide and may further contain any desired elements that control the expression of the nucleic acid(s), as well as any elements that enable the replication and maintenance of the expression vector inside the yeast cell. Many suitable yeast expression vectors and features thereof are known in the art; for example, various vectors and techniques are illustrated in Yeast Protocols, 2nd edition, Xiao, W., ed. (Humana Press, New York, 2007) and Buckholz, R.G. and Gleeson, M.A. (1991) Biotechnology (NY) 9(11): 1067-72. Yeast vectors can contain, without limitation, a centromeric (CEN) sequence, an autonomous replication sequence (ARS), a promoter, such as an RNA Polymerase III promoter, operably linked to a sequence or gene of interest, a terminator such as an RNA polymerase III terminator, an origin of replication, and a marker gene (e.g., auxotrophic, antibiotic, or other selectable markers). Examples of expression vectors for use in yeast may include plasmids, yeast artificial chromosomes, 2 μ plasmids, yeast integrative plasmids, yeast replicative plasmids, shuttle vectors, and episomal plasmids.

[0136] In an embodiment, the vector is a baculovirus vector or expression vector and can be suitable for expression of polynucleotides and/or proteins in insect cells. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., SF9 cells) include the pAc series (Smith, et al., 1983. Mol. Cell. Biol. 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. Virology 170: 31-39). rAAV (recombinant Adeno-associated viral) vectors are preferably produced in insect cells, e.g., *Spodoptera frugiperda* Sf9 insect cells, grown in serum-free suspension culture. Serum-free insect cells can be purchased from commercial vendors, e.g., Sigma Aldrich (EX-CELL 405).

[0137] In an embodiment, the vector is a mammalian expression vector. In an embodiment, the mammalian expression vector is capable of expressing one or more polynucleotides and/or polypeptides in a mammalian cell. Examples of mammalian expression vectors include, but are not limited to, pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, et al., 1987. *EMBO J.* 6: 187-195). The mammalian expression vector can include one or more suitable regulatory elements capable of controlling expression of the one or more polynucleotides and/or proteins in the mammalian cell. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, simian virus 40, and others disclosed herein and known in the art. More detail on suitable regulatory elements is described elsewhere herein.

[0138] For other suitable expression vectors and vector systems for both prokaryotic and eukaryotic cells see, e.g., Chapters 16 and 17 of 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.

[0139] In an embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (e.g., tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, et al., 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Baneiji, et al., 1983. *Cell* 33: 729-740; Queen and Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (e.g., the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, et al., 1985. *Science* 230: 912-916), and mammary gland-specific promoters (e.g., milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also encompassed, e.g., the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546). With regards to these prokaryotic and eukaryotic vectors, mention is made of U.S. Patent 6,750,059, the contents of which are incorporated by reference herein in their entirety. Other embodiments can utilize viral vectors, with regards to which mention is made of U.S. Patent application 13/092,085, the contents of which are incorporated by reference herein in their entirety. Tissue-specific regulatory elements

are known in the art and in this regard, mention is made of U.S. Patent 7,776,321, the contents of which are incorporated by reference herein in their entirety. In an embodiment, a regulatory element can be operably linked to a transgene in a recombinant genome packaged by the engineered AAV capsid system so as to drive expression of the one or more elements of the transgene delivered by the viral vector as described herein in a tissue specific manner.

[0140] Vectors may be introduced and propagated in a prokaryote or prokaryotic cell. In an embodiment, a prokaryote is used to amplify copies of a vector to be introduced into a eukaryotic cell or as an intermediate vector in the production of a vector to be introduced into a eukaryotic cell (e.g., amplifying a plasmid as part of a viral vector packaging system). In an embodiment, a prokaryote is used to amplify copies of a vector and express one or more nucleic acids, such as to provide a source of one or more proteins for delivery to a host cell or host organism.

[0141] In an embodiment, the vector can be a fusion vector or fusion expression vector. In an embodiment, fusion vectors add a number of amino acids to a protein encoded therein, such as to the amino terminus, carboxy terminus, or both of a recombinant protein. Such fusion vectors can serve one or more purposes, such as: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. In an embodiment, expression of polynucleotides (such as non-coding polynucleotides) and proteins in prokaryotes can be carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion polynucleotides and/or proteins. In an embodiment, the fusion expression vector can include a proteolytic cleavage site, which can be introduced at the junction of the fusion vector backbone or other fusion moiety and the recombinant polynucleotide or protein to enable separation of the recombinant polynucleotide or protein from the fusion vector backbone or other fusion moiety subsequent to purification of the fusion polynucleotide or protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Example fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann et al., (1988) *Gene* 69:301-

315) and pET 11d (Studier et al., GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

[0142] In an embodiment, one or more vectors driving expression of one or more elements of an engineered viral (e.g., AAV) capsid system or other composition containing a transferrin receptor binding modification described herein are introduced into a host cell such that expression of the elements of the engineered delivery system described herein direct formation of an engineered viral (e.g., AAV) capsid system or other composition containing a transferrin receptor binding modification described herein (including but not limited to an engineered gene transfer agent particle, which is described in greater detail elsewhere herein). For example, different elements of the engineered viral (e.g., AAV) capsid system or other composition containing a transferrin receptor binding modification described herein can each be operably linked to separate regulatory elements on separate vectors. RNA(s) of different elements of the engineered delivery system described herein can be delivered to an animal or mammal or cell thereof to produce an animal or mammal or cell thereof that constitutively or inducibly or conditionally expresses different elements of the engineered viral (e.g., AAV) capsid system or other composition containing a transferrin receptor binding modification described herein that incorporates one or more elements of the engineered viral (e.g., AAV) capsid system or other composition containing a transferrin receptor binding modification described herein or contains one or more cells that incorporates and/or expresses one or more elements of the engineered viral (e.g., AAV) capsid system or other composition containing a transferrin receptor binding modification described herein.

[0143] In an embodiment, two or more of the elements expressed from the same or different regulatory element(s) can be combined in a single vector, with one or more additional vectors providing any components of the system not included in the first vector. Engineered polynucleotides of the present invention that are combined in a single vector may be arranged in any suitable orientation, such as one element located 5' with respect to ("upstream" of) or 3' with respect to ("downstream" of) a second element. The coding sequence of one element may be located on the same or opposite strand of the coding sequence of a second element, and oriented in the same or opposite direction. In an embodiment, a single promoter drives expression of a transcript encoding one or more engineered viral (e.g., AAV) capsid proteins or other composition containing a transferrin receptor binding modification described herein, embedded within one or

more intron sequences (e.g., each in a different intron, two or more in at least one intron, or all in a single intron). In an embodiment, the engineered polynucleotides of the present invention (including but not limited to engineered viral polynucleotides) can be operably linked to and expressed from the same promoter.

Vector Features

[0144] The vectors can include additional features that can confer one or more functionalities to the vector, the polynucleotide to be delivered, a virus particle produced there from, or polypeptide expressed thereof. Such features include, but are not limited to, regulatory elements, selectable markers, molecular identifiers (e.g., molecular barcodes), stabilizing elements, and the like. It will be appreciated by those skilled in the art that the design of the expression vector and additional features included can depend on such factors as the choice of the host cell to be transformed, the level of expression desired, etc.

Regulatory Elements

[0145] In embodiments, the polynucleotides and/or vectors thereof described herein (including, but not limited to, the engineered AAV capsid polynucleotides of the present disclosure) can include one or more regulatory elements that can be operatively linked to the polynucleotide. The term “regulatory element” is intended to include promoters, enhancers, internal ribosomal entry sites (IRES), and other expression control elements (e.g., transcription termination signals, such as polyadenylation signals and poly-U sequences). Such regulatory elements are described, for example, in Goeddel, *GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY* 185, Academic Press, San Diego, Calif. (1990). Regulatory elements include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (e.g., tissue-specific regulatory sequences). A tissue-specific promoter can direct expression primarily in a desired tissue of interest, such as muscle, neuron, bone, skin, blood, specific organs (e.g., liver, brain), or particular cell types (e.g., lymphocytes). Regulatory elements may also direct expression in a temporal-dependent manner, such as in a cell-cycle dependent or developmental stage-dependent manner, which may or may not also be tissue or cell-type specific. In an embodiment, a vector comprises one or more pol III promoter (e.g., 1, 2, 3, 4, 5, or more pol III promoters), one or more pol II promoters (e.g., 1, 2, 3, 4, 5, or more pol II promoters), one or more pol I promoters (e.g., 1, 2, 3, 4, 5, or more pol I promoters), or combinations thereof. Examples

of pol III promoters include, but are not limited to, U6 and H1 promoters. Examples of pol II promoters include, but are not limited to, the retroviral Rous sarcoma virus (RSV) LTR promoter (optionally with the RSV enhancer), the cytomegalovirus (CMV) promoter (optionally with the CMV enhancer) (see, e.g., Boshart et al, Cell, 41:521-530 (1985)), the SV40 promoter, the dihydrofolate reductase promoter, the β -actin promoter, the phosphoglycerol kinase (PGK) promoter, and the EF1 α promoter. Also encompassed by the term “regulatory element” are enhancer elements, such as WPRE; CMV enhancers; the R-U5' segment in LTR of HTLV-I (Mol. Cell. Biol., Vol. 8(1), p. 466-472, 1988); SV40 enhancer; and the intron sequence between exons 2 and 3 of rabbit β -globin (Proc. Natl. Acad. Sci. USA., Vol. 78(3), p. 1527-31, 1981).

[0146] In an embodiment, the regulatory sequence can be a regulatory sequence described in U.S. Pat. No. 7,776,321, U.S. Pat. Pub. No. 2011/0027239, and PCT publication WO 2011/028929, the contents of which are incorporated by reference herein in their entirety. In an embodiment, the vector can contain a minimal promoter. In an embodiment, the minimal promoter is the Mecp2 promoter, tRNA promoter, or U6. In a further embodiment, the minimal promoter is tissue specific. In an embodiment, the length of the vector polynucleotide the minimal promoters and polynucleotide sequences is less than 4.4Kb.

[0147] To express a polynucleotide, the vector can include one or more transcriptional and/or translational initiation regulatory sequences, e.g., promoters, that direct the transcription of the gene and/or translation of the encoded protein in a cell. In an embodiment a constitutive promoter may be employed. Suitable constitutive promoters for mammalian cells are generally known in the art and include, but are not limited to SV40, CAG, CMV, EF-1 α , β -actin, RSV, and PGK. Suitable constitutive promoters for bacterial cells, yeast cells, and fungal cells are generally known in the art, such as a T-7 promoter for bacterial expression and an alcohol dehydrogenase promoter for expression in yeast.

[0148] In an embodiment, the regulatory element can be a regulated promoter. “Regulated promoter” refers to promoters that direct gene expression not constitutively, but in a temporally- and/or spatially-regulated manner, and includes tissue-specific, tissue-preferred and inducible promoters. In an embodiment, the regulated promoter is a tissue specific promoter as previously discussed elsewhere herein. Regulated promoters include conditional promoters and inducible promoters. In an embodiment, conditional promoters can be employed to direct expression of a polynucleotide in a specific cell type, under certain environmental conditions, and/or during a

specific state of development. Suitable tissue specific promoters can include, but are not limited to, liver specific promoters (e.g., APOA2, SERPIN A1 (hAAT), CYP3A4, and MIR122), pancreatic cell promoters (e.g., INS, IRS2, Pdx1, Alx3, Ppy), cardiac specific promoters (e.g. Myh6 (alpha MHC), MYL2 (MLC-2v), TNI3 (cTnl), NPPA (ANF), Slc8a1 (Ncx1)), central nervous system cell promoters (SYN1, GFAP, INA, NES, MOBP, MBP, TH, FOXA2 (HNF3 beta)), skin cell specific promoters (e.g., FLG, K14, TGM3), immune cell specific promoters, (e.g. ITGAM, CD43 promoter, CD14 promoter, CD45 promoter, CD68 promoter), urogenital cell specific promoters (e.g., Pbsn, Upk2, Sbp, Fer114), endothelial cell specific promoters (e.g., ENG), pluripotent and embryonic germ layer cell specific promoters (e.g. Oct4, NANOG, Synthetic Oct4, T brachyury, NES, SOX17, FOXA2, MIR122), and muscle cell specific promoter (e.g., Desmin). Other tissue and/or cell specific promoters are discussed elsewhere herein and can be generally known in the art and are within the scope of this disclosure.

[0149] Inducible/conditional promoters can be positively inducible/conditional promoters (e.g., a promoter that activates transcription of the polynucleotide upon appropriate interaction with an activated activator, or an inducer (compound, environmental condition, or other stimulus) or a negative/conditional inducible promoter (e.g., a promoter that is repressed (e.g., bound by a repressor) until the repressor condition of the promoter is removed (e.g., inducer binds a repressor bound to the promoter stimulating release of the promoter by the repressor or removal of a chemical repressor from the promoter environment). The inducer can be a compound, environmental condition, or other stimulus. Thus, inducible/conditional promoters can be responsive to any suitable stimuli such as chemical, biological, or other molecular agents, temperature, light, and/or pH. Suitable inducible/conditional promoters include, but are not limited to, Tet-On, Tet-Off, Lac promoter, pBad, AlcA, LexA, Hsp70 promoter, Hsp90 promoter, pDawn, XVE/OlexA, GVG, and pOp/LhGR.

[0150] In an embodiment, the vector or system thereof can include one or more elements capable of translocating and/or expressing an engineered polynucleotide of the present disclosure (e.g., an engineered viral (e.g., AAV) capsid polynucleotide) to/in a specific cell component or organelle. Such organelles can include, but are not limited to, nucleus, ribosome, endoplasmic reticulum, Golgi apparatus, chloroplast, mitochondria, vacuole, lysosome, cytoskeleton, plasma membrane, cell wall, peroxisome, centrioles, etc.

CpG content

[0151] In one aspect, a rAAV vector, including an rAAV vector genome as described herein, comprises at least one synthetic AAV ITR, wherein one or more CpG islands (a cytosine base followed immediately by a guanine base (a CpG) in which the cytosines in such arrangement tend to be methylated) that typically occur at, or near the transcription start site in an ITR are deleted and/or substituted. In one aspect, deletion, or reduction in the number of CpG islands can reduce the immunogenicity of the rAAV vector. This results from a reduction or complete inhibition in TLR-9 binding to the rAAV vector DNA sequence, which occurs at CpG islands. It is also well known that methylation of CpG motifs results in transcriptional silencing. Removal of CpG motifs in the ITR is expected to result in decreased TLR-9 recognition and/or decreased methylation and therefore decreased transgene silencing. In some aspects, it is the minimal functional ITR in which one or more CpG islands are deleted and/or substituted. In one aspect, AAV ITR2 is known to contain 16 CpG islands of which one or more, or all 16 can be deleted.

[0152] In some aspects, at least 1 CpG motif is deleted and/or substituted, e.g., at least 4 or more or 8 or more CpG motifs, e.g., at least 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, or 16 CpG motifs. The phrase “deleted and/or substituted” as used herein means that one or both nucleotides in the CpG motif is deleted, substituted with a different nucleotide, or any combination of deletions and substitutions.

[0153] In certain aspects, the transgene nucleic acid sequence can also be codon optimized to enhance expression in vivo and/or to reduce the number of CpG islands and avoid an innate immune response to the vector. An example of CpG depletion can be found in the published International Application No. PCT/US2023/067901, the content of which is incorporated by reference herein in its entirety.

Dual AAV vectors

[0154] Various strategies have been investigated to overcome the limitation of AAV cargo capacity. Several groups have attempted to “force” large genes into one of the many AAV capsids available by developing the so-called oversize vectors. Although administration of oversize AAV vectors can achieve therapeutically relevant levels of transgene expression in rodent and canine models of human inherited diseases, including the retina of the *Abca4*^{-/-} and shaker 1 (*shl*) mouse models of STGD and USH1B, the mechanism underlying oversize AAV-mediated transduction remains elusive. Oversize AAV vectors do not contain a pure population of intact large size

genomes but rather a heterogeneous mixture of mostly truncated genomes <5 kb in length. Following infection, reassembly of these truncated genomes in the target cell nucleus has been proposed as a mechanism for oversize AAV vector transduction. Independent of transduction mechanism and *in vivo* efficacy, the heterogeneity in oversize AAV genome sizes is a major limitation for their application in human gene therapy.

[0155] Alternatively, the inherent ability of AAV genomes to undergo intermolecular concatemerization can be exploited to transfer large genes *in vivo* by splitting a large gene expression cassette into halves (<5 kb in size), each contained in one of two separate (dual) AAV vectors. In the dual AAV trans-splicing strategy, a splice donor (SD) signal is placed at the 3' end of the 5' - half vector and a splice acceptor (SA) signal is placed at the 5' end of the 3 -half vector. Upon coinfection of the same cell by the dual AAV vectors and inverted terminal repeat (ITR)-mediated head-to-tail concatemerization of the two halves, trans-splicing results in the production of a mature mRNA and full-size protein. Trans-splicing has been successfully used to express large genes in muscle and retina.

[0156] Alternatively, the two halves of a large transgene expression cassette contained in dual AAV vectors may contain homologous overlapping sequences (at the 3' end of the 5 -half vector and at the 5' end of the 3 -half vector, dual AAV overlapping), which will mediate reconstitution of a single large genome by homologous recombination. This strategy depends on the recombinogenic properties of the transgene overlapping sequences.

[0157] A third dual AAV strategy (hybrid) is based on adding a highly recombinogenic region from an exogenous gene [i.e. alkaline phosphatase, AP] to the trans-splicing vector. The added region is placed downstream of the SD signal in the 5'-half vector and upstream of the SA signal in the 3 -half vector in order to increase recombination between the dual AAVs. The published US patent application No. 2010/003218 and US Patent No. 10,494,645, both of which are incorporated by reference herein in their entireties, provide additional examples of dual vector systems.

Selectable Markers and Tags

[0158] One or more of the engineered polynucleotides of the present disclosure (e.g., an engineered viral (e.g., AAV) capsid polynucleotide) can be operably linked, fused to, or otherwise modified to include a polynucleotide that encodes or is a selectable marker or tag, which can be a polynucleotide or polypeptide. In an embodiment, the polypeptide encoding a polypeptide selectable marker can be incorporated in the engineered polynucleotide of the present disclosure

(e.g., an engineered viral (e.g., AAV) capsid polynucleotide) such that the selectable marker polypeptide, when translated, is inserted between two amino acids between the N- and C- terminus of an engineered polypeptide (e.g., the engineered AAV capsid polypeptide) or at the N- and/or C-terminus of the engineered polypeptide (e.g., an engineered AAV capsid polypeptide). In an embodiment, the selectable marker or tag is a polynucleotide barcode or unique molecular identifier (UMI).

[0159] It will be appreciated that the polynucleotide encoding such selectable markers or tags can be incorporated into a polynucleotide encoding one or more components of the engineered AAV capsid system described herein in an appropriate manner to allow expression of the selectable marker or tag. Such techniques and methods are described elsewhere herein and will be instantly appreciated by one of ordinary skill in the art in view of this disclosure. Many such selectable markers and tags are generally known in the art and are intended to be within the scope of this disclosure.

[0160] Suitable selectable markers and tags include, but are not limited to, affinity tags, such as chitin binding protein (CBP), maltose binding protein (MBP), glutathione-S-transferase (GST), poly(His) tag; solubilization tags such as thioredoxin (TRX) and poly(NANP), MBP, and GST; chromatography tags such as those consisting of polyanionic amino acids, such as FLAG-tag; epitope tags such as V5-tag, Myc-tag, HA-tag and NE-tag; protein tags that can allow specific enzymatic modification (such as biotinylation by biotin ligase) or chemical modification (such as reaction with FLAsH-EDT2 for fluorescence imaging), DNA and/or RNA segments that contain restriction enzyme or other enzyme cleavage sites; DNA segments that encode products that provide resistance against otherwise toxic compounds including antibiotics, such as, spectinomycin, ampicillin, kanamycin, tetracycline, Basta, neomycin phosphotransferase II (NEO), hygromycin phosphotransferase (HPT) and the like; DNA and/or RNA segments that encode products that are otherwise lacking in the recipient cell (e.g., tRNA genes, auxotrophic markers); DNA and/or RNA segments that encode products which can be readily identified (e.g., phenotypic markers such as β -galactosidase, GUS; fluorescent proteins such as green fluorescent protein (GFP), cyan (CFP), yellow (YFP), red (RFP), luciferase, and cell surface proteins); polynucleotides that can generate one or more new primer sites for PCR (e.g., the juxtaposition of two DNA sequences not previously juxtaposed), DNA sequences not acted upon or acted upon by a restriction endonuclease or other DNA modifying enzyme, chemical, etc.; epitope tags (e.g. GFP,

FLAG- and His-tags), and, DNA sequences that make a molecular barcode or unique molecular identifier (UMI), DNA sequences required for a specific modification (e.g., methylation) that allows its identification. Other suitable markers will be appreciated by those of skill in the art.

[0161] Selectable markers and tags can be operably linked to one or more components of the engineered AAV capsid system or other compositions and/or systems described herein via suitable linker, such as a glycine or glycine serine linkers as short as GS or GG up to (GGGGG)₃ (SEQ ID NO: 1733) or (GGGGS)₃ (SEQ ID NO: 1734). Other suitable linkers are described elsewhere herein.

[0162] The vector or vector system can include one or more polynucleotides encoding one or more transferrin receptor binding modifications. In an embodiment, the transferrin receptor binding modification encoding polynucleotides can be included in the vector or vector system, such as a viral vector system, such that they are expressed within and/or on the virus particle(s) produced such that the virus particles can be targeted to specific cells, tissues, organs, etc. In an embodiment, the transferrin receptor binding modification encoding polynucleotides can be included in the vector or vector system such that the engineered polynucleotide(s) of the present disclosure (e.g., an engineered viral (e.g., AAV) capsid polynucleotide(s)) and/or products expressed therefrom include the transferrin receptor binding modification and can be targeted to specific cells, tissues, organs, etc. In an embodiment, such as non-viral carriers, the transferrin receptor binding modification can be attached to the carrier (e.g., polymer, lipid, inorganic molecule etc.) and can be capable of targeting the carrier and any attached or associated engineered polynucleotide(s) of the present disclosure, the engineered polypeptides, or other compositions of the present disclosure described herein, to specific cells, tissues, organs, etc. In an embodiment, the specific cells are CNS cells.

Cell-free Vector and Polynucleotide Expression

[0163] In an embodiment, the polynucleotide(s) encoding a transferrin receptor binding modification of the present disclosure can be expressed from a vector or suitable polynucleotide in a cell-free in vitro system. In an embodiment, the polynucleotide encoding one or more features of the engineered AAV capsid system can be expressed from a vector or suitable polynucleotide in a cell-free in vitro system. In other words, the polynucleotide can be transcribed and optionally translated in vitro. In vitro transcription/translation systems and appropriate vectors are generally known in the art and commercially available. Generally, in vitro transcription and in vitro

translation systems replicate the processes of RNA and protein synthesis, respectively, outside of the cellular environment. Vectors and suitable polynucleotides for *in vitro* transcription can include T7, SP6, T3, promoter regulatory sequences that can be recognized and acted upon by an appropriate polymerase to transcribe the polynucleotide or vector.

[0164] *In vitro* translation can be stand-alone (e.g., translation of a purified polyribonucleotide) or linked/coupled to transcription. In an embodiment, the cell-free (or *in vitro*) translation system can include extracts from rabbit reticulocytes, wheat germ, and/or *E. coli*. The extracts can include various macromolecular components that are needed for translation of exogenous RNA (e.g., 70S or 80S ribosomes, tRNAs, aminoacyl-tRNA, synthetases, initiation, elongation factors, termination factors, etc.). Other components can be included or added during the translation reaction, including but not limited to, amino acids, energy sources (ATP, GTP), energy regenerating systems (creatine phosphate and creatine phosphokinase (eukaryotic systems)) (phosphoenol pyruvate and pyruvate kinase for bacterial systems), and other co-factors (Mg²⁺, K⁺, etc.). As previously mentioned, *in vitro* translation can be based on RNA or DNA starting material. Some translation systems can utilize an RNA template as starting material (e.g., reticulocyte lysates and wheat germ extracts). Some translation systems can utilize a DNA template as a starting material (e.g., *E. coli*-based systems). In these systems transcription and translation are coupled and DNA is first transcribed into RNA, which is subsequently translated. Suitable standard and coupled cell-free translation systems are generally known in the art and are commercially available.

Codon Optimization of Vector Polynucleotides

[0165] As described elsewhere herein, the polynucleotide encoding a transferrin receptor binding modification of the present disclosure and/or other polynucleotides described herein, or the transgene contained within the recombinant AAV genome can be codon optimized. In an embodiment, polynucleotides of the engineered AAV capsid system described herein can be codon optimized. In an embodiment, one or more polynucleotides contained in a vector (“vector polynucleotides”) described herein that are in addition to an optionally codon optimized polynucleotide encoding a transferrin receptor binding modification, including but not limited to, embodiments of the engineered AAV capsid system described herein, can be codon optimized. In general, codon optimization refers to a process of modifying a nucleic acid sequence for enhanced expression in the host cells of interest by replacing at least one codon (e.g., about or more than

about 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more codons) of the native sequence with codons that are more frequently or most frequently used in the genes of that host cell while maintaining the native amino acid sequence. Various species exhibit particular bias for certain codons of a particular amino acid. Codon bias (differences in codon usage between organisms) often correlates with the efficiency of translation of messenger RNA (mRNA), which is in turn believed to be dependent on, among other things, the properties of the codons being translated and the availability of particular transfer RNA (tRNA) molecules. The predominance of selected tRNAs in a cell is generally a reflection of the codons used most frequently in peptide synthesis. Accordingly, genes can be tailored for optimal gene expression in a given organism based on codon optimization. Codon usage tables are readily available, for example, at the “Codon Usage Database” available at www.kazusa.or.jp/codon/ and these tables can be adapted in a number of ways. See Nakamura, Y., et al. “Codon usage tabulated from the international DNA sequence databases: status for the year 2000” *Nucl. Acids Res.* 28:292 (2000). Computer algorithms for codon optimizing a particular sequence for expression in a particular host cell are also available, such as Gene Forge (Aptagen; Jacobus, PA), are also available. In an embodiment, one or more codons (e.g., 1, 2, 3, 4, 5, 10, 15, 20, 25, 50, or more, or all codons) in a sequence encoding a DNA/RNA-targeting Cas protein corresponds to the most frequently used codon for a particular amino acid. As to codon usage in yeast, reference is made to the online Yeast Genome database available at www.yeastgenome.org/community/codon_usage.shtml, or *Codon selection in yeast*, Bennetzen and Hall, *J Biol Chem.* 1982 Mar 25;257(6):3026-31. As to codon usage in plants including algae, reference is made to *Codon usage in higher plants, green algae, and cyanobacteria*, Campbell and Gowri, *Plant Physiol.* 1990 Jan; 92(1): 1–11.; as well as *Codon usage in plant genes*, Murray et al, *Nucleic Acids Res.* 1989 Jan 25;17(2):477-98; or *Selection on the codon bias of chloroplast and cyanelle genes in different plant and algal lineages*, Morton BR, *J Mol Evol.* 1998 Apr;46(4):449-59.

[0166] The vector polynucleotide can be codon optimized for expression in a specific cell-type, tissue type, organ type, and/or subject type. In an embodiment, a codon optimized sequence is a sequence optimized for expression in a eukaryote, e.g., humans (i.e., being optimized for expression in a human or human cell), or for another eukaryote, such as another animal (e.g., a mammal or avian) as is described elsewhere herein. Such codon optimized sequences are within the ambit of the ordinary skilled artisan in view of the description herein. In an embodiment, the

polynucleotide is codon optimized for a specific cell type. Such cell types can include, but are not limited to, epithelial cells (including skin cells, cells lining the gastrointestinal tract, cells lining other hollow organs), nerve cells (nerves, brain cells, spinal column cells, nerve support cells (e.g., astrocytes, glial cells, Schwann cells etc.)), muscle cells (e.g., cardiac muscle, smooth muscle cells, and skeletal muscle cells), connective tissue cells (fat and other soft tissue padding cells, bone cells, tendon cells, cartilage cells), blood cells, stem cells and other progenitor cells, immune system cells, germ cells, and combinations thereof. Such codon optimized sequences are within the ambit of the ordinary skilled artisan in view of the description herein. In an embodiment, the polynucleotide is codon optimized for a specific tissue type. Such tissue types can include, but are not limited to, muscle tissue, connective tissue, nervous tissue, and epithelial tissue. Such codon optimized sequences are within the ambit of the ordinary skilled artisan in view of the description herein. In an embodiment, the polynucleotide is codon optimized for a specific organ. Such organs include, but are not limited to, muscles, skin, intestines, liver, spleen, brain, lungs, stomach, heart, kidneys, gallbladder, pancreas, bladder, thyroid, bone, blood vessels, blood, and combinations thereof. Such codon optimized sequences are within the ambit of the ordinary skilled artisan in view of the description herein.

[0167] In an embodiment, a vector polynucleotide is codon optimized for expression in particular cells, such as prokaryotic or eukaryotic cells. The eukaryotic cells may be those of or derived from a particular organism, such as a plant or a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as discussed herein, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate.

Non-Viral Vectors and Carriers

[0168] In an embodiment, the vector is a non-viral vector or carrier. In an embodiment, non-viral vectors can have the advantage(s) of reduced toxicity and/or immunogenicity and/or increased bio-safety as compared to viral vectors. The terms of art “Non-viral vectors and carriers” and as used herein in this context refers to molecules and/or compositions that are not based on one or more component of a virus or virus genome (excluding any nucleotide to be delivered and/or expressed by the non-viral vector) that can be capable of attaching to, incorporating, coupling, and/or otherwise interacting with an engineered capsid polynucleotide (e.g., an engineered AAV capsid polynucleotide) or other composition of the present disclosure described herein and can be capable of ferrying the polynucleotide to a cell and/or expressing the polynucleotide. It will be

appreciated that this does not exclude the inclusion of a virus-based polynucleotide that is to be delivered. For example, if a gRNA to be delivered is directed against a virus component and it is inserted or otherwise coupled to an otherwise non-viral vector or carrier, this would not make said vector a “viral vector”. Non-viral vectors and carriers include naked polynucleotides, chemical-based carriers, polynucleotide (non-viral) based vectors, and particle-based carriers. It will be appreciated that the term “vector” as used in the context of non-viral vectors and carriers refers to polynucleotide vectors and “carriers” used in this context refers to a non-nucleic acid or polynucleotide molecule or composition that be attached to or otherwise interact with a polynucleotide to be delivered, such as an engineered AAV capsid polynucleotide of the present disclosure.

Naked Polynucleotides

[0169] In an embodiment one or more engineered AAV capsid polynucleotides or other polynucleotides of the present disclosure described elsewhere herein can be included in a naked polynucleotide. The term of art “naked polynucleotide” as used herein refers to polynucleotides that are not associated with another molecule (e.g., proteins, lipids, and/or other molecules) that can often help protect it from environmental factors and/or degradation. As used herein, associated with includes, but is not limited to, linked to, adhered to, adsorbed to, enclosed in, enclosed in or within, mixed with, and the like. Naked polynucleotides that include one or more of the engineered AAV capsid polynucleotides or other polynucleotides of the present disclosure described herein can be delivered directly to a host cell and optionally expressed therein. The naked polynucleotides can have any suitable two- and three-dimensional configurations. By way of non-limiting examples, naked polynucleotides can be single-stranded molecules, double stranded molecules, circular molecules (e.g., plasmids and artificial chromosomes), molecules that contain portions that are single stranded and portions that are double stranded (e.g., ribozymes), and the like. In an embodiment, the naked polynucleotide contains only the engineered AAV capsid polynucleotide(s) or other polynucleotides of the present disclosure. In an embodiment, the naked polynucleotide can contain other nucleic acids and/or polynucleotides in addition to the engineered AAV capsid polynucleotide(s) or other polynucleotides of the present disclosure described elsewhere herein. The naked polynucleotides can include one or more elements of a transposon system. Transposons and system thereof are described in greater detail elsewhere herein.

Non-Viral Polynucleotide Vectors

[0170] In an embodiment, one or more of the engineered AAV capsid polynucleotides or other polynucleotides of the present disclosure can be included in a non-viral polynucleotide vector. Suitable non-viral polynucleotide vectors include, but are not limited to, transposon vectors and vector systems, plasmids, bacterial artificial chromosomes, yeast artificial chromosomes, AR(antibiotic resistance)-free plasmids and miniplasmids, circular covalently closed vectors (e.g., minicircles, minivectors, miniknots.), linear covalently closed vectors (“dumbbell shaped”), MIDGE (minimalistic immunologically defined gene expression) vectors, MiLV (micro-linear vector) vectors, Ministrings, mini-intronic plasmids, PSK systems (post-segregationally killing systems), ORT (operator repressor titration) plasmids, and the like. See e.g., Hardee et al. 2017. *Genes*. 8(2):65.

[0171] In an embodiment, the non-viral polynucleotide vector can have a conditional origin of replication. In an embodiment, the non-viral polynucleotide vector can be an ORT plasmid. In an embodiment, the non-viral polynucleotide vector can have a minimalistic immunologically defined gene expression. In an embodiment, the non-viral polynucleotide vector can have one or more post-segregationally killing system genes. In an embodiment, the non-viral polynucleotide vector is AR-free. In an embodiment, the non-viral polynucleotide vector is a minivector. In an embodiment, the non-viral polynucleotide vector includes a nuclear localization signal. In an embodiment, the non-viral polynucleotide vector can include one or more CpG motifs. In an embodiment, the non-viral polynucleotide vectors can include one or more scaffold/matrix attachment regions (S/MARs). See e.g., Mirkovitch et al. 1984. *Cell*. 39:223-232, Wong et al. 2015. *Adv. Genet.* 89:113-152, whose techniques and vectors can be adapted for use in the present disclosure. S/MARs are AT-rich sequences that play a role in the spatial organization of chromosomes through DNA loop base attachment to the nuclear matrix. S/MARs are often found close to regulatory elements such as promoters, enhancers, and origins of DNA replication. Inclusion of one or S/MARs can facilitate a once-per-cell-cycle replication to maintain the non-viral polynucleotide vector as an episome in daughter cells. In an embodiment, the S/MAR sequence is located downstream of an actively transcribed polynucleotide (e.g., one or more engineered AAV capsid polynucleotides or other polynucleotides or molecules of the present disclosure) included in the non-viral polynucleotide vector. In an embodiment, the S/MAR can be a S/MAR from the beta-interferon gene cluster. See e.g., Verghese et al. 2014. *Nucleic Acid Res.*

42:e53; Xu et al. 2016. *Sci. China Life Sci.* 59:1024-1033; Jin et al. 2016. 8:702-711; Koirala et al. 2014. *Adv. Exp. Med. Biol.* 801:703-709; and Nehlsen et al. 2006. *Gene Ther. Mol. Biol.* 10:233-244, whose techniques and vectors can be adapted for use in the present disclosure.

[0172] In an embodiment, the non-viral vector is a transposon vector or system thereof. As used herein, “transposon” (also referred to as transposable element) refers to a polynucleotide sequence that is capable of moving from location in a genome to another. There are several classes of transposons. Transposons include retrotransposons and DNA transposons. Retrotransposons require the transcription of the polynucleotide that is moved (or transposed) in order to transpose the polynucleotide to a new genome or polynucleotide. DNA transposons are those that do not require reverse transcription of the polynucleotide that is moved (or transposed) in order to transpose the polynucleotide to a new genome or polynucleotide. In an embodiment, the non-viral polynucleotide vector can be a retrotransposon vector. In an embodiment, the retrotransposon vector includes long terminal repeats. In an embodiment, the retrotransposon vector does not include long terminal repeats. In an embodiment, the non-viral polynucleotide vector can be a DNA transposon vector. DNA transposon vectors can include a polynucleotide sequence encoding a transposase. In an embodiment, the transposon vector is configured as a non-autonomous transposon vector, meaning that the transposition does not occur spontaneously on its own. In some of these embodiments, the transposon vector lacks one or more polynucleotide sequences encoding proteins required for transposition. In an embodiment, the non-autonomous transposon vectors lack one or more Ac elements.

[0173] In an embodiment, a non-viral polynucleotide transposon vector system can include a first polynucleotide vector that contains the engineered AAV capsid polynucleotide(s) or other polynucleotides, or molecules of the present disclosure described herein flanked on the 5' and 3' ends by transposon terminal inverted repeats (TIRs) and a second polynucleotide vector that includes a polynucleotide capable of encoding a transposase coupled to a promoter to drive expression of the transposase. When both are expressed in the same cell the transposase can be expressed from the second vector and can transpose the material between the TIRs on the first vector (e.g., the engineered AAV capsid polynucleotide(s) or other polynucleotides or molecules of the present disclosure) and integrate it into one or more positions in the host cell's genome. In an embodiment the transposon vector or system thereof can be configured as a gene trap. In an embodiment, the TIRs can be configured to flank a strong splice acceptor site followed by a

reporter and/or other gene (e.g., one or more of the engineered AAV capsid polynucleotide(s) or other polynucleotides or molecules of the present disclosure) and a strong poly A tail. When transposition occurs while using this vector or system thereof, the transposon can insert into an intron of a gene and the inserted reporter or other gene can provoke a mis-splicing process and as a result it in activates the trapped gene.

[0174] Any suitable transposon system can be used. Suitable transposon and systems thereof can include Sleeping Beauty transposon system (Tc1/mariner superfamily) (see e.g., Ivics et al. 1997. Cell. 91(4): 501-510), piggyBac (piggyBac superfamily) (see e.g., Li et al. 2013 110(25): E2279-E2287 and Yusa et al. 2011. PNAS. 108(4): 1531-1536), Tol2 (superfamily hAT), Frog Prince (Tc1/mariner superfamily) (see e.g., Miskey et al. 2003 Nucleic Acid Res. 31(23):6873-6881) and variants thereof.

Chemical Carriers

[0175] In an embodiment, the engineered AAV capsid polynucleotide(s) or other polynucleotides or other molecules of the present disclosure described herein can be coupled to a chemical carrier. Chemical carriers that can be suitable for delivery of polynucleotides can be broadly classified into the following classes: (i) inorganic particles, (ii) lipid-based, (iii) polymer-based, and (iv) peptide based. They can be categorized as (1) those that can form condensed complexes with a polynucleotide (such as the engineered AAV capsid polynucleotide(s) of the present disclosure), (2) those capable of targeting specific cells, (3) those capable of increasing delivery of the polynucleotide or other molecules (such as the engineered AAV capsid polynucleotide(s) of the present disclosure to the nucleus or cytosol of a host cell, (4) those capable of disintegrating from DNA/RNA in the cytosol of a host cell, and (5) those capable of sustained or controlled release. It will be appreciated that any one given chemical carrier can include features from multiple categories. The term “particle” as used herein, refers to any suitable sized particles for delivery of the compositions (including particles, polypeptides, polynucleotides, and other compositions described herein) present disclosure described herein. Suitable sizes include macro-, micro-, and nano-sized particles.

[0176] In an embodiment, the non-viral carrier can be an inorganic particle. In an embodiment, the inorganic particle, can be a nanoparticle. The inorganic particles can be configured and optimized by varying size, shape, and/or porosity. In an embodiment, the inorganic particles are optimized to escape from the reticulo endothelial system. In an embodiment, the inorganic particles

can be optimized to protect an entrapped molecule from degradation. The suitable inorganic particles that can be used as non-viral carriers in this context can include, but are not limited to, calcium phosphate, silica, metals (e.g., gold, platinum, silver, palladium, rhodium, osmium, iridium, ruthenium, mercury, copper, rhenium, titanium, niobium, tantalum, and combinations thereof), magnetic compounds, particles, and materials, (e.g., supermagnetic iron oxide and magnetite), quantum dots, fullerenes (e.g., carbon nanoparticles, nanotubes, nanostrings, and the like), and combinations thereof. Other suitable inorganic non-viral carriers are discussed elsewhere herein.

[0177] In an embodiment, the non-viral carrier can be lipid-based. Suitable lipid-based carriers are also described in greater detail herein. In an embodiment, the lipid-based carrier includes a cationic lipid or an amphiphilic lipid that is capable of binding or otherwise interacting with a negative charge on the polynucleotide to be delivered (e.g., such as an engineered AAV capsid polynucleotide of the present disclosure). In an embodiment, chemical non-viral carrier systems can include a polynucleotide (such as the engineered AAV capsid polynucleotide(s)) or other composition or molecule of the present disclosure) and a lipid (such as a cationic lipid). These are also referred to in the art as lipoplexes. Other embodiments of lipoplexes are described elsewhere herein. In an embodiment, the non-viral lipid-based carrier can be a lipid nano emulsion. Lipid nano emulsions can be formed by the dispersion of an immiscible liquid in another stabilized emulsifying agent and can have particles of about 200 nm that are composed of the lipid, water, and surfactant that can contain the polynucleotide to be delivered (e.g., the engineered AAV capsid polynucleotide(s) of the present disclosure). In an embodiment, the lipid-based non-viral carrier can be a solid lipid particle or nanoparticle.

[0178] In an embodiment, the non-viral carrier can be peptide-based. In an embodiment, the peptide-based non-viral carrier can include one or more cationic amino acids. In an embodiment, 35 to 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, 95, 99 or 100 % of the amino acids are cationic. In an embodiment, peptide carriers can be used in conjunction with other types of carriers (e.g., polymer-based carriers and lipid-based carriers to functionalize these carriers). In an embodiment, the functionalization is targeting a host cell. Suitable polymers that can be included in the polymer-based non-viral carrier can include, but are not limited to, polyethylenimine (PEI), chitosan, poly (DL-lactide) (PLA), poly (DL-Lactide-co-glycoside) (PLGA), dendrimers (see e.g., US Pat. Pub. 2017/0079916 whose techniques and compositions can be adapted for use with the engineered

AAV capsid polynucleotides of the present disclosure), polymethacrylate, and combinations thereof.

[0179] In an embodiment, the non-viral carrier can be configured to release an engineered delivery system polynucleotide that is associated with or attached to the non-viral carrier in response to an external stimulus, such as pH, temperature, osmolarity, concentration of a specific molecule or composition (e.g., calcium, NaCl, and the like), pressure and the like. In an embodiment, the non-viral carrier can be a particle that is configured includes one or more of the engineered AAV capsid polynucleotides or other compositions of the present disclosure describe herein and an environmental triggering agent response element, and optionally a triggering agent. In an embodiment, the particle can include a polymer that can be selected from the group of polymethacrylates and polyacrylates. In an embodiment, the non-viral particle can include one or more embodiments of the compositions microparticles described in US Pat. Pubs. 20150232883 and 20050123596, whose techniques and compositions can be adapted for use in the present disclosure.

[0180] In an embodiment, the non-viral carrier can be a polymer-based carrier. In an embodiment, the polymer is cationic or is predominantly cationic such that it can interact in a charge-dependent manner with the negatively charged polynucleotide to be delivered (such as the engineered AAV capsid polynucleotide(s) of the present disclosure). Polymer-based systems are described in greater detail elsewhere herein.

Viral Vectors

[0181] In an embodiment, the vector is a viral vector. The term of art “viral vector” and as used herein in this context refers to polynucleotide based vectors that contain one or more elements from or based upon one or more elements of a virus that can be capable of expressing and packaging a polynucleotide, such as an engineered AAV capsid polynucleotide, cargo, or other composition or molecule of the present disclosure, into a virus particle and producing said virus particle when used alone or with one or more other viral vectors (such as in a viral vector system). Viral vectors and systems thereof can be used for producing viral particles for delivery of and/or expression and/or generation of one or more compositions of the present disclosure described herein (including, but not limited to, any viral particle and associated cargo). The viral vector can be part of a viral vector system involving multiple vectors. In an embodiment, systems incorporating multiple viral vectors can increase the safety of these systems. Suitable viral vectors

can include adenoviral-based vectors, adeno associated vectors, helper-dependent adenoviral (HdAd) vectors, hybrid adenoviral vectors, and the like. Other embodiments of viral vectors and viral particles produced therefrom are described elsewhere herein. In an embodiment, the viral vectors are configured to produce replication incompetent viral particles for improved safety of these systems.

Adenoviral vectors, Helper-dependent Adenoviral vectors, and Hybrid Adenoviral Vectors

[0182] In an embodiment, the vector can be an adenoviral vector. In an embodiment, the adenoviral vector can include elements such that the virus particle produced using the vector or system thereof can be serotype 2, 5, or 9. In an embodiment, the polynucleotide to be delivered via the adenoviral particle can be up to about 8 kb. Thus, in an embodiment, an adenoviral vector can include a DNA polynucleotide to be delivered that can range in size from about 0.001 kb to about 8 kb. Adenoviral vectors have been used successfully in several contexts (see e.g., Teramoto et al. 2000. *Lancet*. 355:1911-1912; Lai et al. 2002. *DNA Cell. Biol.* 21:895-913; Flotte et al., 1996. *Hum. Gene. Ther.* 7:1145-1159; and Kay et al. 2000. *Nat. Genet.* 24:257-261. The vector can encode the engineered AAV capsids, said capsids forming adenoviral particles.

[0183] In an embodiment the vector can be a helper-dependent adenoviral vector or system thereof. These are also referred to in the field as “gutless” or “guttled” vectors and are a modified generation of adenoviral vectors (see e.g., Thrasher et al. 2006. *Nature*. 443:E5-7). In an embodiment of the helper-dependent adenoviral vector system one vector (the helper) can contain all the viral genes required for replication but contains a conditional gene defect in the packaging domain. The second vector of the system can contain only the ends of the viral genome, one or more engineered AAV capsid polynucleotides, and the native packaging recognition signal, which can allow selective packaged release from the cells (see e.g., Cideciyan et al. 2009. *N Engl J Med*. 361:725-727). Helper-dependent Adenoviral vector systems have been successful for gene delivery in several contexts (see e.g., Simonelli et al. 2010. *J Am Soc Gene Ther.* 18:643-650; Cideciyan et al. 2009. *N Engl J Med*. 361:725-727; Crane et al. 2012. *Gene Ther.* 19(4):443-452; Alba et al. 2005. *Gene Ther.* 12:18-S27; Croyle et al. 2005. *Gene Ther.* 12:579-587; Amalfitano et al. 1998. *J. Virol.* 72:926-933; and Morral et al. 1999. *PNAS*. 96:12816-12821). The techniques and vectors described in these publications can be adapted for inclusion and delivery of the engineered AAV capsid polynucleotides described herein. In an embodiment, the polynucleotide to be delivered via the viral particle produced from a helper-dependent adenoviral vector or system

thereof can be up to about 38 kb. Thus, in an embodiment, an adenoviral vector can include a DNA polynucleotide to be delivered that can range in size from about 0.001 kb to about 37 kb (see e.g., Rosewell et al. 2011. J. Genet. Syndr. Gene Ther. Suppl. 5:001).

[0184] In an embodiment, the vector is a hybrid-adenoviral vector or system thereof. Hybrid adenoviral vectors are composed of the high transduction efficiency of a gene-deleted adenoviral vector and the long-term genome-integrating potential of adeno-associated, and transposon based-gene transfer. In an embodiment, such hybrid vector systems can result in stable transduction and limited integration site. See e.g., Balague et al. 2000. Blood. 95:820-828; Morral et al. 1998. Hum. Gene Ther. 9:2709-2716; Kubo and Mitani. 2003. J. Virol. 77(5): 2964-2971; Zhang et al. 2013. PloS One. 8(10) e76771; and Cooney et al. 2015. Mol. Ther. 23(4):667-674), whose techniques and vectors described therein can be modified and adapted for use in the engineered AAV capsid system of the present disclosure. In an embodiment, a hybrid-adenoviral vector can include one or more features of a retrovirus and/or an adeno-associated virus. In an embodiment the hybrid-adenoviral vector can include one or more features of a spuma retrovirus or foamy virus (FV). See e.g., Ehrhardt et al. 2007. Mol. Ther. 15:146-156 and Liu et al. 2007. Mol. Ther. 15:1834-1841, whose techniques and vectors described therein can be modified and adapted for use in the engineered AAV capsid system of the present disclosure. Advantages of using one or more features from the FVs in the hybrid-adenoviral vector or system thereof can include the ability of the viral particles produced therefrom to infect a broad range of cells, a large packaging capacity as compared to other retroviruses, and the ability to persist in quiescent (non-dividing) cells. See also e.g., Ehrhardt et al. 2007. Mol. Ther. 15:146-156 and Shuji et al. 2011. Mol. Ther. 19:76-82, whose techniques and vectors described therein can be modified and adapted for use in the engineered AAV capsid system of the present disclosure.

Adeno Associated Vectors

[0185] In an embodiment, the engineered vector or system thereof can be an adeno-associated vector (AAV). See, e.g., West et al., Virology 160:38-47 (1987); U.S. Pat. No. 4,797,368; WO 93/24641; Kotin, Human Gene Therapy 5:793-801 (1994); and Muzyczka, J. Clin. Invest. 94:1351 (1994). Although similar to adenoviral vectors in some of their features, AAVs have some deficiency in their replication and/or pathogenicity and thus can be safer than adenoviral vectors. In an embodiment, the AAV can integrate into a specific site on chromosome 19 of a human cell with no observable side effects. In an embodiment, the capacity of the AAV vector, system thereof,

and/or AAV particles can be up to about 4.7 kb. The AAV vector or system thereof can include one or more engineered capsid polynucleotides described herein.

[0186] The AAV vector or system thereof can be operably linked to a regulatory sequence, said regulatory sequence encoding one or more regulatory molecules. In an embodiment the regulatory molecules can be promoters, enhancers, repressors and the like, which are described in greater detail elsewhere herein. In an embodiment, the AAV vector or system thereof can include one or more polynucleotides that can encode one or more regulatory proteins. In an embodiment, the promoter can be a tissue specific promoter as previously discussed. In an embodiment, the tissue specific promoter can drive expression of an engineered capsid AAV capsid polynucleotide described herein.

[0187] The AAV vector or system thereof can include one or more polynucleotides that can encode one or more capsid proteins, such as the engineered AAV capsid proteins described elsewhere herein. The engineered capsid proteins can be capable of assembling into a protein shell (an engineered capsid) of the AAV virus particle. The engineered capsid can have a cell-, tissue-, and/or organ-specific tropism.

[0188] In an embodiment, the AAV vector or system thereof can include one or more adenovirus helper factors or polynucleotides that can encode one or more adenovirus helper factors. Such adenovirus helper factors can include, but are not limited, E1A, E1B, E2A, E4ORF6, and VA RNAs. In an embodiment, a producing host cell line expresses one or more of the adenovirus helper factors.

[0189] The AAV vector or system thereof can be configured to produce AAV particles having a specific serotype. In an embodiment, the serotype can be AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV8, AAV9 or any combinations thereof. In an embodiment, the AAV can be AAV1, AAV2, AAV5, AAV9 or any combination thereof. One can select the AAV of the AAV with regard to the cells to be targeted; e.g., one can select AAV serotypes 1, 2, 5, 9 or a hybrid capsid AAV1, AAV2, AAV5, AAV9 or any combination thereof for targeting brain and/or neuronal cells; and one can select AAV4 for targeting cardiac tissue; and one can select AAV8 for delivery to the liver. Thus, in an embodiment, an AAV vector or system thereof capable of producing AAV particles capable of targeting the brain and/or neuronal cells can be configured to generate AAV particles having serotypes 1, 2, 5 or a hybrid capsid AAV1, AAV2, AAV5 or any combination thereof. In an embodiment, an AAV vector or system thereof capable of producing AAV particles

capable of targeting cardiac tissue can be configured to generate an AAV particle having an AAV4 serotype. In an embodiment, an AAV vector or system thereof capable of producing AAV particles capable of targeting the liver can be configured to generate an AAV having an AAV8 serotype. See also Srivastava. 2017. *Curr. Opin. Virol.* 21:75-80.

[0190] The nucleotide sequences of the genomes of the AAV serotypes are known in the art. For example, the complete genome of AAV1 is provided in GenBank Accession No. NC 002077; the complete genome of AAV2 is provided in GenBank Accession No. NC 001401 and Srivastava et al., *J. Virol.*, 45: 555-564 (1983); the complete genome of AAV3 is provided in GenBank Accession No. NC 1829; the complete genome of AAV4 is provided in GenBank Accession No. NC 001829; the AAV5 genome is provided in GenBank Accession No. AF085716; the complete genome of AAV6 is provided in GenBank Accession No. NC 00 1862; at least portions of AAV 7 and AAV8 genomes are provided in GenBank Accession Nos. AX753246 and AX753249, respectively; the AAV9 genome is provided in Gao et al., *J. Virol.*, 78: 6381-6388 (2004); the AAV10 genome is provided in *Mol. Ther.*, 13(1): 67-76 (2006); the AAV11 genome is provided in *Virology*, 330(2): 375-383 (2004); AAV PHP.B is described by Deverman et al., *Nature Biotech.* 34(2), 204-209 and its sequence deposited under GenBank Accession No. KU056473.1. Exemplary reviews of AAV serotypes may be found in Choi et al (2005) *Curr Gene Ther* 5(3): 299-310 and Wu et al (2006) *Molecular Therapy* 14(3), 316-327.

[0191] It will be appreciated that while the different serotypes can provide some level of cell, tissue, and/or organ specificity, each serotype still is multi-tropic and thus can result in tissue-toxicity if using that serotype to target a tissue that the serotype is less efficient in transducing. Thus, in addition to achieving some tissue targeting capacity via selecting an AAV of a particular serotype, it will be appreciated that the tropism of the AAV serotype can be modified by an engineered AAV capsid described herein. As described elsewhere herein, variants of wild-type AAV of any serotype can be generated via a method described herein and determined to have a particular cell-specific tropism, which can be the same or different as that of the reference wild-type AAV serotype. In an embodiment, the cell, tissue, and/or specificity of the wild-type serotype can be enhanced (e.g., made more selective or specific for a particular cell type that the serotype is already biased towards). For example, wild-type AAV9 is biased towards muscle and brain in humans (see e.g., Srivastava. 2017. *Curr. Opin. Virol.* 21:75-80.) By including an engineered AAV capsid and/or capsid protein variant of wild-type AAV9 as described herein, the bias for e.g., brain

can be reduced or eliminated and/or the septicity increased such that the brain specificity appears reduced in comparison, thus enhancing the specificity for the muscle as compared to the wild-type AAV9. As previously mentioned, inclusion of an engineered capsid and/or capsid protein variant of a wild-type AAV serotype can have a different tropism than the wild-type reference AAV serotype. For example, an engineered AAV capsid and/or capsid protein variant of AAV9 can have specificity for a tissue other than muscle or brain in humans.

[0192] In an embodiment, the AAV vector is a hybrid AAV vector or system thereof. Hybrid AAVs are AAVs that include genomes with elements from one serotype that are packaged into a capsid derived from at least one different serotype. For example, if it is the rAAV2/5 that is to be produced, and if the production method is based on the helper-free, transient transfection method discussed below, the 1st plasmid and the 3rd plasmid (the adeno helper plasmid) will be the same as discussed for rAAV2 production. However, the 2nd plasmid, the pRepCap will be different. In this plasmid, called pRep2/Cap5, the Rep gene is still derived from AAV2, while the Cap gene is derived from AAV5. The production scheme is the same as the above-mentioned approach for AAV2 production. The resulting rAAV is called rAAV2/5, in which the genome is based on recombinant AAV2, while the capsid is based on AAV5. It is assumed the cell or tissue-tropism displayed by this AAV2/5 hybrid virus should be the same as that of AAV5. It will be appreciated that wild-type hybrid AAV particles suffer the same specificity issues as with the non-hybrid wild-type serotypes previously discussed.

[0193] Advantages achieved by the wild-type based hybrid AAV systems can be combined with the increased and customizable cell-specificity that can be achieved with the engineered AAV capsids can be combined by generating a hybrid AAV that can include an engineered AAV capsid described elsewhere herein. It will be appreciated that hybrid AAVs can contain an engineered AAV capsid containing a genome with elements from a different serotype than the reference wild-type serotype that the engineered AAV capsid is a variant of. For example, a hybrid AAV can be produced that includes an engineered AAV capsid that is a variant of an AAV9 serotype that is used to package a genome that contains components (e.g., AAV2 ITRs) from an AAV2 serotype. As with wild-type based hybrid AAVs previously discussed, the tropism of the resulting AAV particle will be that of the engineered AAV capsid.

[0194] A tabulation of certain wild-type AAV serotypes as to these cells can be found in Grimm, D. et al, J. Virol. 82: 5887-5911 (2008) reproduced below as **Table E**. Further tropism details can be found in Srivastava. 2017. Curr. Opin. Virol. 21:75-80 as previously discussed.

Table E								
Cell Line	AAV1	AAV2	AAV3	AAV4	AAV5	AAV6	AAV8	AAV9
Huh-7	13	100	2.5	0.0	0.1	10	0.7	0.0
HEK293	25	100	2.5	0.1	0.1	5	0.7	0.1
HeLa	3	100	2.0	0.1	6.7	1	0.2	0.1
HepG2	3	100	16.7	0.3	1.7	5	0.3	ND
Hep1A	20	100	0.2	1.0	0.1	1	0.2	0.0
911	17	100	11	0.2	0.1	17	0.1	ND
CHO	100	100	14	1.4	333	50	10	1.0
COS	33	100	33	3.3	5.0	14	2.0	0.5
MeWo	10	100	20	0.3	6.7	10	1.0	0.2
NIH3T3	10	100	2.9	2.9	0.3	10	0.3	ND
A549	14	100	20	ND	0.5	10	0.5	0.1
HT1180	20	100	10	0.1	0.3	33	0.5	0.1
Monocytes	1111	100	ND	ND	125	1429	ND	ND
Immature DC	2500	100	ND	ND	222	2857	ND	ND
Mature DC	2222	100	ND	ND	333	3333	ND	ND

[0195] In an embodiment, the AAV vector or system thereof is AAV rh.74 or AAV rh.10.

[0196] In an embodiment, the AAV vector or system thereof is configured as a “gutless” vector, similar to that described in connection with a retroviral vector. In an embodiment, the “gutless” AAV vector or system thereof can have the cis-acting viral DNA elements involved in genome amplification and packaging in linkage with the heterologous sequences of interest (e.g., a transgene encoding a therapeutic protein or nucleic acid of interest)).

Table F: Example virus sequences – see Figure 17 for an example alignment of the example virus sequences.

SEQ ID	Sequence	Name
1	MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYL GPGNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLKED TSFGGNLGRAVFQAKKRLLLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSG AQPAKKRLNFGQTGDTESVPDPQPIGEPPAAPSGVGSLTMASGGGAPVADNNEGAD GVGSSSGNWHCDSQWLGDRTVTTSTRTWALPTYNNHLYKQISNSTSGGSSNDNAYF GYSTPWGYFDNRFHCHFSRWDQRLINNNWGFRPKRLNFKLFNIQVKEVTDNNG VKTIANNLTSTVQVFSDSYQLPYVLGSAHEGCLPPFPADVFMIPQYGYLTLNDGSQ AVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVPFHSSYAHSQSLDRLMNPLIDQYL YYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQNNNSE FAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDA DKVMITNEEEIKTTNPVATESYQVATNHQSAQAQAQTGWVQNGILPGMVWQD RDVYLQGPWAKIPHTDGNFHPSPMLMGFGMKHPPQILIKNTVPADPPTAFNKDK LNSFITQYSTGQVSVEIEWELQKENS KRWNPEIQYTSNYYKSNNVEFAVNTEGVYSE PRPIGTRYLTRNL*	AAV9
2	MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNARGLVLPGYKYL GPGNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLKED TSFGGNLGRAVFQAKKRLLLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSG AQPAKKRLNFGQTGDTESVPDPQPIGEPPAAPSGVGSLTMASGGGAPVADNNEGAD GVGSSSGNWHCDSQWLGDRTVTTSTRTWALPTYNNHLYKQISNSTSGGSSNDNAYF GYSTPWGYFDNRFHCHFSRWDQRLINNNWGFRPKRLNFKLFNIQVKEVTDNNG VKTIANNLTSTVQVFSDSYQLPYVLGSAHEGCLPPFPADVFMIPQYGYLTLNDGSQ AVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVPFHSSYAHSQSLDRLMNPLIDQYL YYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNYIPGPSYRQQRVSTTVTQNNNSE FAWPGASSWALNGRNSLMNPGPAMASHKEGEDRFFPLSGSLIFGKQGTGRDNVDA DKVMITNEEEIKTTNPVATESYQVATNHQSAQAQAQTGWVQNGILPGMVWQD RDVYLQGPWAKIPHTDGNFHPSPMLMGFGMKHPPQILIKNTVPADPPTAFNKDK LNSFITQYSTGQVSVEIEWELQKENS KRWNPEIQYTSNYYKSNNVEFAVNTEGVYSE PRPIGTRYLTRNL*	AAV9 K449R
3	MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYL GPFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADADEFQERLQED TSFGGNLGRAVFQAKKRVLLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSSGIGKTG QQPAKKRLNFGQTGDSESVDPQPLGEPPATPAAVGPTTMASGGGAPMADNNEGA DVGNGASGNWHCDSTWLGDRVTTSTRTWALPTYNNHLYKQISSASTGASNDNH FGYSTPWGYFDNRFHCHFSRWDQRLINNNWGFRPKRLNFKLFNIQVKEVTTNDG VTIANNLTSTVQVFSDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQ AVGRSSFYCLEYFPSQMLRTGNNFTFSYTFEEVPFHSSYAHSQSLDRLMNPLIDQYL YYLNRTQNQSGSAQNKDLLFSRGSAPGMSVQPKNWLPGPCYRQQRVSKTKTDNN NSNFTWTGASKYNLNGRESIINPGTAMASHKDDKFFPMMSGVMIFGKESAGASNT ALDNVMITDEEEIKATNPVATERFGTVAVNFQSSSTDPATGDVHAMGALPGMVWQ DRDVYLQGPWAKIPHTDGHFHPSPMLMGFGGLKNPPQILIKNTVPANPPAEFSATK FASFITQYSTGQVSVEIEWELQKENS KRWNPEVQYTSNYAKSANVDFTVDNNGLYT EPRPIGTRYLTRPL	AAV1
4	MAADGYLPDWLEDTLSEGIRQWWKLKPGPPPKPAERHKDDSRGLVLPGYKYLGP FNGLDKGEPVNEADAAALEHDKAYDRQLDSGDNPYLKYNHADADEFQERLKEDTSF GGNLGRAVFQAKKRVLLEPLGLVEEPVKTAPGKKRPVEHSPVEPDSSSGTGKAGQQP ARKRLNFGQTGDADSVPDPQPLGQPPAAPSGLGTNTMATGSGAPMADNNEGADG VGNSSGNWHCDSTWMGDRVTTSTRTWALPTYNNHLYKQISSQSGASNDNHFGY STPWGYFDNRFHCHFSRWDQRLINNNWGFRPKRLNFKLFNIQVKEVTQNDGTTT	AAV2

	IANNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVMPQYGYLTLNNGSQAV GRSSFYCLEYFPSQMLRTGNNFTFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLY LSRTNTPSGTTTQSRLLQFSQAGASDIRDQSRNWLPGPCYRQQRVSKTSADNNNSEY SWTGATKYHLNGRDSL VNP GPAMASHKDDEEKFFPQSGVLIFGKQGSEKTNVDIEK VMITDEEEIRTTNPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMVWQDRDV YLQGPWIAKIPHTDGHFHPSPLMGGFGLKHPPQILIKNTPVPANPSTTFSAAKFASFI TQYSTGQVSVEIEWELQKENS KRWNPEIQYTSNYNKS VNVDFTVD TNGVYSEPRPI GTRYLTRNL	
5	MAADGYLPDWLEDNLSEGIREW WALKPGVPQPKANQQHQDNRRGLVLPGYKYLG PGNGLDKGEPVNEADAAALEHDKAYDQQLKAGDNPYLKYNHADA EFQERLQEDT SFGGNLGRAVFQAKKRILEPLGLVEEAAKTAPGKKGAVDQSPQEPDSSSGVGKSGK QPARKRLNFGQTGDSESVPDPQPLGEPPAAPTSLGSNTMASGGGAPMADNNEGAD GVGNSSGNWHCDSQWLGD RVITTTSTRTWALPTYNNHLYKQISSQSGASNDNHYFG YSTPWGYFDNRFHCHFSRDPWQRLINNNWGRPKKLSFKLFNIQVRGVTQNDGTT TIANNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVMPQYGYLTLNNGSQAV VGRSSFYCLEYFPSQMLRTGNNFQFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLY YLNRQTGTTSGTTNQSRLIFSQAQPSMSLQARNWLPGPCYRQQRLSKTANDNNN SNFPWTAASKYHLNGRDSL VNP GPAMASHKDDEEKFFPMHGNLIFGKEGTTASNA ELDNVMITDEEEIRTTNPVATEQYGTVANNLQSSNTAPTGT VNHQGALPGMVWQ DRDVYLQGPWIAKIPHTDGHFHPSPLMGGFGLKHPPQIMIKNTPVPANPPTTFSPA KFASFITQYSTGQVSVEIEWELQKENS KRWNPEIQYTSNYNKS VNVDFTVD TNGVY SEPRPIGTRYLTRNL	AAV3
6	MAADGYLPDWLEDNLSEGIREW WALKPGVPQPKANQQHQDNRRGLVLPGYKYLG PGNGLDKGEPVNEADAAALEHDKAYDQQLKAGDNPYLKYNHADA EFQERLQEDT SFGGNLGRAVFQAKKRILEPLGLVEEAAKTAPGKKRPVDQSPQEPDSSSGVGKSGK QPARKRLNFGQTGDSESVPDPQPLGEPPAAPTSLGSNTMASGGGAPMADNNEGAD GVGNSSGNWHCDSQWLGD RVITTTSTRTWALPTYNNHLYKQISSQSGASNDNHYFG YSTPWGYFDNRFHCHFSRDPWQRLINNNWGRPKKLSFKLFNIQVKEVTQNDGTT TIANNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVMPQYGYLTLNNGSQAV VGRSSFYCLEYFPSQMLRTGNNFQFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLY YLNRQTGTTSGTTNQSRLIFSQAQPSMSLQARNWLPGPCYRQQRLSKTANDNNN SNFPWTAASKYHLNGRDSL VNP GPAMASHKDDEEKFFPMHGNLIFGKEGTTASNA ELDNVMITDEEEIRTTNPVATEQYGTVANNLQSSNTAPTTRTVNDQGALPGMVWQ DRDVYLQGPWIAKIPHTDGHFHPSPLMGGFGLKHPPQIMIKNTPVPANPPTTFSPA KFASFITQYSTGQVSVEIEWELQKENS KRWNPEIQYTSNYNKS VNVDFTVD TNGVY SEPRPIGTRYLTRNL	AAV3B
7	MTDGYLPDWLEDNLSEGVREWWALQPGAPKPKANQQHQDNARGLVLPGYKYLG PGNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADA EFQQRQLQGD TSFGGNLGRAVFQAKKRILEPLGLVEQAGETAPGKKRPLIESPQQPDSSTGIGKKGK QPAKKKL VFEDET GAGDGPPEGSTSGAMSDDSEMRAAAGGAAVEGGQGADGVGN ASGDWHCDSTWSEGHVTTTSTRTWVLPPTYNNHLYKRLGESLQSN TYNGFSTPWGY FDNRFHCHFSRDPWQRLINNNWGM RPKAMRVKIFNIQVKEVTTSNGETTVANNL TSTVQIFADSSYELPYVMDAQEGSLPPFPNDVFMVMPQYGYCGLVTGNTSQQQTD NAFYCLEYFPSQMLRTGNNFEITYSFEKVPFHSMYAHSQSLDRLMNPLIDQYWLGL QSTTTGTTLNAGTATTNFTKL RPTNFSNFKKNWLPGPSIKQQGFSKTANQNYKIPAT GSDSLIKYETHSTLDGRWSALT PGPMPATAGPADSKFSNSQLIFAGPKQNGNTATVP GTLIFTSEEELAATNATD TDMWGNLPGGDQSN SNLPTVDRLTALGAVPGMVWQNR DIYYQGPWIAKIPHTDGHFHPSP LIGGFGLKHPPQIFIKNTPVPANPATTFSTPVNSF ITQYSTGQVSVIDWEIQKERSKRWNPEVQFTSNYQQQNSLLWAPDAAGKYTEPRA IGTRYLTHHL	AAV4
8	MSFVDHPPDWLEEVGEGLEFLGLEAGPPKPKPNQQHQDQARGLVLPGYNYLGP NGLDRGEPVNRAD E VAREHDISYNEQLEAGDNPYLKYNHADA EFQEKLAADDTSFG GNLGKAVFQAKKRILEPFG LVEEGAKTAPTGRIDDHFPKRKKARTEEDSKPSTSS DAEAGPSGQQLQIPAQPASSLGADTMSAGGGGPGDNNQAGDGVGNASGDWHC DSTWMGDRVVT KSTRTWVLP SYNNHQYREIKSGSVDGSNANAYFGYSTPWGYFD FNRFHSHWSPRDWQRLINNYWGRPRSLRVKIFNIQVKEVTQDSTTTIANNLTSTV	AAV5

	QVFTDDDDYQLPYVVGNGTEGCLPAFPQVFTLPQYGYATLNRDNTENPTERSSFFC LEYFPSKMLRTGNNFEFTYNFEEVVPFHSSFAPSQNLFKLANPLVDQYLYRFVSTNNT GGVQFNKNLAGRYANTYKNWFGPMGRTQGWNLGSGVNRASVSFAFATTNRMEL EGASYQVPPQPNGMTNNLQGSNTYALENTMIFNSQPANPGTTATYLEGNMLITSES ETQPVNRVAYNVGGQMATNNQSSTTAPATGTYNLQEIVPGSVWMERDVYLQGP WAKIPETGAHFHPSAMGGFGLKHPPPMMLIKNTPVPGNITSFSDVPVSSFITQYSTG QVTVEMEWELKKENSKRWNPFIQYTNNYNDPQFVDFAPDSTGEYRTTRPIGTRYLT RPL	
9	MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYL GPFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADADEFQERLQED TSFGGNLGRAVFQAKKRVLEPFLVEEGAKTAPGKKRPVEQSPQEPDSSSGIGKTG QQPAKKRLNFGQTGDSESVDPQPLGEPPATPAAVGPTTMAASGGGAPMADNNEGA DGVGNASGNWHCDSTWLGDRVITSTRTWALPTYNNHLYKQISSASTGASNDNH FGYSTPWGYFDNRFHCHFSRDPWQRLINNNWGFRPKRLNFKLFNIQVKEVTTNDG VTIANNNLTSTVQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGS AVGRSSFYCLEYFPSQMLRTGNNFTFSYTFEDVPFHSSY AHSQSLDRLMNPLIDQYL YYLNRQTQNSGSAQNKDLLFSRGSAPGMSVQPKNWLPGPCYRQQRVSKTKTDNN NSNFTWTGASKYNLNGRESIINPGTAMASHKDDKDKFFPMMSGVMIFGKESAGASNT ALDNVMTDEEEIKATNPVATERFTVAVNLQSSSTDPATGDVHVMGALPGMVWQ DRDVYLQGPWAKIPHDTGDFHPSPLMGGFGLKHPPQILIKNTPVPANPPAEFSATK FASFITQYSTGQVSVEIEWELQKENSkrwnpevqytsnyaksanvdfTVDNNGLYT EPRPIGTRYLTRPL	AAV6
10	MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDNGRGLVLPGYKYL GPFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADADEFQERLQED TSFGGNLGRAVFQAKKRVLEPLGLVEEGAKTAPAKKRPVEPSQSPDSSTGIGKK GQPAKKRLNFGQTGDSESVDPQPLGEPPAAPSSVSGSTVAAGGGAPMADNNEGA ADGVGNASGNWHCDSTWLGDRVITSTRTWALPTYNNHLYKQISSETAGSTNDNT YFGYSTPWGYFDNRFHCHFSRDPWQRLINNNWGFRPKRLNFKLFNIQVKEVTTND GVTIANNNLTSTIQVFSDEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGS QSVGRSSFYCLEYFPSQMLRTGNNFEFSYFEDVPFHSSY AHSQSLDRLMNPLIDQY LYYLARTQSNPGGTAGNRELQFYQGGPSTMAEQAKNWLPGPCFRQQRVSKTLNQ NNSNFAWTGATKYHLNGRNSLVNPGVAMATHKDDERFFPSSGVLIFGKTGATNK TTLENVLMTNEEEIRPTNPVATEEYGIVSSNLQAANTAAQTQVNNQ GALPGMVW QNRDVYLQGPWAKIPHDTGDFHPSPLMGGFGLKHPPQILIKNTPVPANPPEVFTPA KFASFITQYSTGQVSVEIEWELQKENSkrwnpeiqytsnfekqTGVDFAVDSQGVYS EPRPIGTRYLTRNL	AAV7
11	MAADGYLPDWLEDNLSEGIREWWALKPGAPKPKANQQKQDDGRGLVLPGYKYL GPFNGLDKGEPVNAADAAALEHDKAYDQQLQAGDNPYLRYNHADADEFQERLQED TSFGGNLGRAVFQAKKRVLEPLGLVEEGAKTAPGKKRPVEPSQSPDSSTGIGKK GQPAKKRLNFGQTGDSESVDPQPLGEPPAAPSGVGPNTMAAGGGAPMADNNEGA ADGVGSSSGNWHCDSTWLGDRVITSTRTWALPTYNNHLYKQISNGTSGGATNDN TYFGYSTPWGYFDNRFHCHFSRDPWQRLINNNWGFRPKRLNFKLFNIQVKEVTQN EGTKTIANNNLTSTIQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNG SQAVGRSSFYCLEYFPSQMLRTGNNFQFTYTFEDVPFHSSY AHSQSLDRLMNPLIDQ YLYLRSRTQTTGGTANTQTLGFSQGGPNTMANQAKNWLPGPCYRQQRVSTTTGQN NNSNFAWTAGTKYHLNGRNSLANPGIAMATHKDDERFFPSNGILIFGKQNAARDN ADYSDVMLTSEEEIKTTNPVATEEYGIVADNLQQQNTAPQIGTVNSQGALPGMVW QNRDVYLQGPWAKIPHDTGDFHPSPLMGGFGLKHPPQILIKNTPVPADPPTTFNQ KLNSFITQYSTGQVSVEIEWELQKENSkrwnpeiqytsnyyKSTSVDFAVNTEGVYS EPRPIGTRYLTRNL	AAV8
12	MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYL GPFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADADEFQERLQED TSFGGNLGRAVFQAKKRVLEPLGLVEEGAKTAPGKKRPVEQSPQEPDSSSGIGKTG QQPAKKRLNFGQTGDSESVDPQPLGEPPAAPSGLGPNTMASGGGAPMADNNEGA DGVGNSSSGNWHCDSTWLGDRVITSTRTWALPTYNNHLYKQISNGTSGGSTNDNT YFGYSTPWGYFDNRFHCHFSRDPWQRLINNNWGFRPKRLNFKLFNIQVKEVTTNE	rh.8

	GTKTIANNLTSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVMPQYGYLTLNNG SQALGRSSFYCLEYFPSQMLRTGNNFQFSYTFEDVPFHSSYAHSQSLDRLMNPLIDQ YLYYLVRTQTGTGGTQTLAFSQAGPSSMANQARNWVPGPCYRQQRVSTTTNQN NSNFAWTGA AKFKLNGRDSLMPGVAMASHKDDDDRRFFPSSGVLIFGKQGAGND GVDYSQVLITDEEEKATNPVATEEYGAVAINNQAANTQAQTGLVHNQGVIPGMV WQNRDVYLGQPIWAKIPHTDGNFHPSPLMGGFGLKHPPPQILIKNTVPADPPLTFN QAKLNSFITQYSTGQVSVEIEWELQKENS KRWNPEIQYTSNYYKSTNVDFAVNTEG VYSEPRPIGTRYLTRNL	
13	MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYL GPFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADADEFQERLQED TSFGGNLGRAVFQAKKRVLEPLGLVEEGA K TAPGKKRPVEPSPQRSPDSSTGIGKK GQQPAKKRLNFGQTGDSESVDPQPIGEPPAGPSGLSGTMAAGGGAPMADNNEG ADGVGSSSGNWHCDSTWLGDRTVTTSTRTWALPTYNNHLYKQISNGTSGGSTNDN TYFGYSTPWGYFDNRFHCHFSRDPWQRLINNNWGRPKRLNFKLFNIQVKEVTQN EGTKTIANNLTSTIQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNG SQAVGRSSFYCLEYFPSQMLRTGNNFESYQFEDVPFHSSYAHSQSLDRLMNPLIDQ YLYYLRTSTGGTAGTQQLFSQAGPNMMSAQAKNWLPGPCYRQQRVSTTSLQN NNSNFAWTGATKYHLNGRDSLVPNGVAMATHKDDEERFFPSSGVL MFGKQGAGK DNVDYSSVMLTSEEEIKTTNPVATEQYGVVADNLQQNAAPIVGAVNSQGALPGM VWQNRDVYLGQPIWAKIPHTDGNFHPSPLMGGFGLKHPPPQILIKNTVPADPPTTF SQAKLASFITQYSTGQVSVEIEWELQKENS KRWNPEIQYTSNYYKSTNVDFAVNTD GTYSEPRPIGTRYLTRNL	rh.10
14	MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYL GPFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADADEFQERLQED TSFGGNLGRAVFQAKKRVLEPLGLVEEGA K TAPGKKRPLESPQEPDSSSGIGKKGK QPAKKRLNFEEDTGAGDGPPEGSDTSAMSSDIEMRAAPGGNAVDAGQSDGVGNA SGDWHCDSTWSEGKVTSTRTWVLPPTYNNHLYLRLGTTSSNTYNGFSTPWGYF DFNRFHCHFSRDPWQRLINNNWGLRPAKMRVKIFNIQVKEVTTSNGETTVANNLTS TVQIFADSSYELPYVMDAGQEGSLPPFPNDVFMVMPQYGYCGIVTGENQNQTDNRNFA YCLEYFPSQMLRTGNNFEMAYNFEKVPFHSMYAHSQSLDRLMNPLLDQYLWHLQS TTSGETLNQGNAATTFGKIRSGDFAFYRKNWLPGPCVKQQRFSKTASQNYKIPASG GNALLKYDTHYTLNNRWSNIAPGPPMATAGPSDGDFSNAQLIFFGPSVTGNTTTSA NNLLFTSEEEIAATNPRD TDMFGQIADNNQNATTAPITGNVTAMGVLPGMVWQNR DIYYQGPIWAKIPHADGHFHSPLIGGFGLKHPPPQIFIKNTVPANPATTTFAARVDS FITQYSTGQVAVQIEWEIEKERSKRWNPEVQFTSNYGNQSSMLWAPD TTGKYTEPR VIGSRYLTNHL	rh.74
15	MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYL GPFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADADEFQERLQED TSFGGNLGRAVFQAKKRVLEPLGLVEEAAKTAPGKKRPVEPSPQRSPDSSTGIGKK GQQPAKKRLNFGQTGESESVPDPQPIGEPPAGPSGLSGTMAAGGGAPMADNNEG ADGVGSSSGNWHCDSTWLGDRTVTTSTRTWALPTYNNHLYKQISNGTSGGSTNDN TYFGYSTPWGYFDNRFHCHFSRDPWQRLINNNWGRPKRLSFKLFNIQVKEVTQN EGTKTIANNLTSTIQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNG SQAVGRSSFYCLEYFPSQMLRTGNNFESYTFEDVPFHSSYAHSQSLDRLMNPLIDQ YLYYLRTSTGGTQGTQQLFSQAGPANMSAQAKNWLPGPCYRQQRVSTTSLQN NNSNFAWTGATKYHLNGRDSLVPNGVAMATHKDDEERFFPSSGVL MFGKQGAGR DNVDYSSVMLTSEEEIKTTNPVATEQYGVVADNLQQANTGPVGNVNSQGALPGM VWQNRDVYLGQPIWAKIPHTDGNFHPSPLMGGFGLKHPPPQILIKNTVPADPPTTF SQAKLASFITQYSTGQVSVEIEWELQKENS KRWNPEIQYTSNYYKSTNVDFAVNTE GTYSEPRPIGTRYLTRNL	AAV10
16	MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYL GPFNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADADEFQERLQED TSFGGNLGRAVFQAKKRVLEPLGLVEEGA K TAPGKKRPLESPQEPDSSSGIGKKGK QPAKKRLNFEEDTGAGDGPPEGSDTSAMSSDIEMRAAPGGNAVDAGQSDGVGNA SGDWHCDSTWSEGKVTSTRTWVLPPTYNNHLYLRLGTTSSNTYNGFSTPWGYF DFNRFHCHFSRDPWQRLINNNWGLRPAKMRVKIFNIQVKEVTTSNGETTVANNLTS	AAV11

	TVQIFADSSYELPYVMDAGQEGSLPPFPNDVFMVPPQYGYCGIVTGENQNQTDNRNAY YCLEYFPSQMLRTGNNFEMAYNFEKVPFHSMY AHSQSLDRLMNPLLDQYLWHLQS TTSGETLNQGNAAATTFGKIRSGDFAFYRKNWLPGPCVKQQRFSKTASQNYKIPASG GNALLKYDTHYTLNNRWSNIAPGPPMATAGPSDGDGFSNAQLIFGPSVTGNTTTS NNLLFTSEEEIAATNPRDMDMFGQIADNNQNATTAPITGNVTAMGVLPGMVWQNR DIYYQGPIWAKIPHADGHFHPSPILIGGFGLKHPPPIQIFIKNTVPANPATTTAARVDS FITQYSTGQVAVQIEWEIEKERSKRWNPEVQFTSNYGNQSSMLWAPDITGKYTEPR VIGSRYLTNHL	
17	MAADGYLPDWLEDNLSEGIREWWALKPGAPQPKANQQHQDNNGRGLVLPGYKYL GPFNGLDKGEPVNEADAAALEHDKAYDKQLEQGDNPYLKYNHADADEFQQLATD TSFGGNLGRAVFQAKKRILEPLGLVEEGVKTAPGKKRPLEKTPNRPTNPDSGKAPA KKKQKDGEPADSARRTLDFEDSGAGDGPPEGSSSGEMSHDAEMRAAPGNAVEA GQGADGVGNASGDWHCDSTWSEGRVTTTSTRTWVLPYNNHLYLRIGTTANSNT YNGFSTPWGYDFNRFHCHFSRDPWQRLINNNWGLRPKSMRVKIFNIQVKEVTTSN GETTVANNLTSTVQIFADSTYELPYVMDAGQEGSFPPFPNDVFMVPPQYGYCGVVTG KNQNQTDNRNAYCLEYFPSQMLRTGNNFEVSQFEKVPFHSMY AHSQSLDRMMNP LLDQYLWHLQSTTGNSLNQGTATTTYGKITTGDFAYYRKNWLPACIKQKFKSK NANQNYKIPASGGDALLKYDTHTTLNNGRWSNMAPGPPMATAGAGDSDFSNSQLIF AGPNPSGNTTSSNNLLFTSEEEIATNPRDMDMFGQIADNNQNATTAPHIANLDAM GIVPGMVWQNRDIYYQGPIWAKVPHTDGHFHPSPILMGGFGLKHPPPIQIFIKNTVP ANPNTTFSARINSFLTQYSTGQVAVQIDWEIQKEHSKRWNPEVQFTSNYGTQNSM LWAPDNAGNYHELRAIGSRFLTHHL	AAV12
18	MTDGYLPDWLEDNLSEGIREWWALQPGAPKPKANQQHQDNARGLVLPGYKYL PGNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLQEDT SFGGNLGRAVFQAKKRILEPLGLVEEAAKTAPGKKRPVEQSPAEPDSSSGIGKSGQ PARKRLNFGQTDGTSVPDPQPLGQPPAAPSGVGSTTMASGGGAPMADNNEGADG VGNSSGNWHCDSQWLGDRTVTTSTRTWALPTYNNHLYKQISSQSGATNDNHYFGY STPWGYDFNRFHCHFSRDPWQRLINNNWGRPKRLNFKLFNIQVKEVTQNDGTTT IANNLSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPPQYGYLTLNNGSQAV GRSSFYCLEYFPSQMLRTGNNFQFSYTFEDVPFHSSY AHSQSLDRLMNPLIDQYL YLNRTQTAGSTQSRLLFSQAGPMSLSLQAKNWLPGPCYRQRLSKQANDNNNSNF PWTGATKYHLNGRDSLVPNPGPAMASHKDDKEKFFPMHGTILIFGKEGTNANNADLE NVMITDEEEIRTTPVATEQYGTVSNLQNSNAGPTTGTVNHQALPGMVWQDRD VYLQGPWAKIPHDTGDFHPSPLMGGFGLKHPPPIQIMIKNTVPANPPTNFSAAKFA SFITQYSTGQVSVEIEWELQKENS KRWNPEIQYTSNYSNYSVNVDFTVDTNGVYSEPR PIGTRYLTRNL	AAV13
19	MAADGYLPDWLEDNLSEGIREWWALQPGAPKPKANQQHQDNARGLVLPGYKYL GPGNGLDKGEPVNAADAAALEHDKAYDQQLKAGDNPYLKYNHADADEFQERLKED TSFGGNLGRAVFQAKKRILLEPLGLVEEAAKTAPGKKRPVDQSPQEPDSSSGVGKSG KQPARKRLNFGQTDGDSVDPDPQPLGEPPAAPTSLGSNTMASGGGAPMADNNEGA DGVGNSSGNWHCDSQWLGDRTVTTSTRTWALPTYNNHLYKQISSQSGASNDNHYF GYSTPWGYDFNRFHCHFSRDPWQRLINNNWGRPKLSFKLFNIQVKEVTQNDGT TTIANNLSTVQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMVPPQYGYLTLNNGSQ AVGRSSFYCLEYFPSQMLRTGNNFQFSYTFEDVPFHSSY AHSQSLDRLMNPLIDQYL YYLNRTQGTSGTTNQSRLFSQAGPMSLSLQARNWLPGPCYRQRLSKTANDNN NSNFPWTAASKYHLNGRDSLVPNPGPAMASHKDDDEEKFFPMHGNLIFGKEGTTASN AELDNVMITDEEEIRTTPVATEQYGTVANLQSSNTAPTTRTVNDQALPGMVW QDRDVYLQGPWAKIPHDTGDFHPSPLMGGFGLKHPPPIQIMIKNTVPANPPTTFS AKFASFITQYSTGQVSVEIEWELQKENS KRWNPEIQYTSNYSNYSVNVDFTVDTNGV YSEPRPIGTRYLTRPL	LK03
20	MAADGYLPDWLEDNLSEGIRQWWKLKPGPPPKPAERHKDDSRGLVLPGYKYLGP FNGLDKGEPVNEADAAALEHDKAYDRQLDSGDNPYLKYNHADADEFQERLKEDTSF GGNLGRAVFQAKKRILLEPLGLVEEAAKTAPGKKRPVEHSPVEPDSSSGTGKAGQ ARKRLNFGQTDGADSVDPDPQIGEPAPSGVGSLTMAAGGGAPMADNNEGADGV GNSSGNWHCDSTWMGDRVTTTSTRTWALPTYNNHLYKQISNSTSGGSSNDNAYFG YSTPWGYDFNRFHCHFSRDPWQRLINNNWGRPKRLSFKLFNIQVKEVTQNEGK	AAV-DJ

TIANNLTSTIQVFTDSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAV GRSSFYCLEYFPSQMLRTGNNFQFTYTFEDVPFHSSYAHSQSLDRLMNPLIDQYLYY LSRTQTTGGTTNTQTLGFSQGGPNTMANQAKNWLPGPCYRQQRVSKTSADNNNSE YSWGATKYHLNGRDSLVPNPGPAMASHKDDEEKFFPQSGVLIFGKQGSEKTNVDIE KVMITDEEEIRTTNPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMVWQDRD VYLQGPWAKIPHTDGHFHPSPLMGGFGLKHPPQILIKNTPVPADPPTTFNQSKLNS FITQYSTGQVSVEIEWELQKENSQRWNPEIQYTSNYYKSTSVDFAVNTEGVYSEPRP IGTRYLTRNL	
---	--

Vectors Encoding the Transgene

[0197] In an embodiment, the vector encoding the transgene (also referred to as “an artificial genome”) comprises the transgene to be delivered flanked on either side by AAV ITRs. Only ~145 bp AAV ITRs are required for recombinant AAV (rAAV) propagation because they participate in vector production, induce transgene expression, and ensure continual cell transduction. Accordingly, ~96% of the AAV genome can be removed for gene therapy. For example, the rep and cap genes can be substituted for the expression cassette containing a promoter (such as those described herein), a therapeutic transgene (for example, IDS) and a poly(A) tail forms the essence of all AAV vectors.

[0198] In an embodiment, additional modifications may be implemented to further increase the efficacy of the AAV. For example, the AAV ITRs may be modified to increase the expression of the rAAV vector upon transduction, which may allow the transgene to be expressed without second-strand DNA synthesis; the promoter may be modified to increase transcription; and the codons in the transgene may be engineered to modify mRNA production and/or translation.

[0199] In an embodiment, the ITRs are modified to overcome second-strand synthesis after infection. The AAV transduction rate is restricted by the synthesis of dsDNA from the single-stranded AAV genome. ITRs initiate second-strand synthesis. In an example embodiment, modified ITRs are no longer suitable substrates for the Rep68 and Rep78 proteins. As a result, the terminal resolution of replication is obviated, and specific self-complementary AAV (scAAV) replication intermediates are produced. The scAAV intermediates comprise plus and minus strands of DNA fused by the modified ITRs encapsulated into the virion shell. Wild-type AAVs package either a single plus-strand or minus-strand DNA. The modified scAAV intermediates are delivered to the nucleus, these plus and minus strands instantaneously anneal to form dsDNA.

[0200] In an embodiment, the cis-elements are optimized for targeted delivery. The cis-elements are optimized because the packaging capacity of AAVs is restricted. In an example

embodiment, small cis-elements replace long promoter sequences for the delivery of large therapeutic transgenes (e.g., 4.4–4.5 kbs).

[0201] In an embodiment, several strategies may be used to deliver transgenes using AAV vectors. Example approach 1 takes advantage of an AAV genome concatemerized via the homologous recombination of ITR sequences. In this approach, transgene cassettes may be split into two or more vectors, which are then delivered to the same cells. After the virus is uncoated, an intact transgene is formed by the homologous recombination between the two or more fragments.

[0202] In example approach 2, truncated transgene fragments of different lengths are packaged into different AAV virions at undefined locations on the vector genome. Either homologous recombination of the overlapping regions of the different AAV vector genomes or annealing of different AAV vector genomes at complementary regions via single-stranded templates produces the transgene cassette. In an embodiment, overlapping fragments may be added to the end of the individual AAV vectors to encourage homologous recombination.

[0203] In example approach 3, a hybrid dual-vector incorporates an overlapping region with intron splice sites in the split vector transgenes. Approach 3 uses concatemerization activity of AAV genomes to bring independent AAV vector genomes together. Recombination (for example, the starting vectors are segregated into two halves each carrying the 5' and 3' splicing elements, respectively), and splicing provide the appropriate transgene protein. This strategy may increase the expression of full functional protein.

[0204] In example approach 4, an AAV genome is cross-packaged into the capsids of other parvoviruses thus creating chimeric vectors. In example approach 5, intein-mediated protein trans-splicing is used. Intein catalyzes protein splicing thereby causing the ligation of two polypeptides via trans-splicing (this approach is similar to intron-mediated RNA splicing). Multiple AAV vectors are delivered to the same cells. Each of the AAV vectors encode one of the fragments of target proteins, the fragments are flanked by short split inteins. The full-length protein forms after protein trans-splicing. See e.g., Li, C., Samulski, R.J. Engineering adeno-associated virus vectors for gene therapy. *Nat Rev Genet* 21, 255–272 (2020), herein incorporated by reference.

Vector Construction

[0205] The vectors described herein can be constructed using any suitable process or technique. In an embodiment, one or more suitable recombination and/or cloning methods or

techniques can be used to the vector(s) described herein. Suitable recombination and/or cloning techniques and/or methods can include, but not limited to, those described in U.S. Application publication No. US 2004-0171156 A1. Other suitable methods and techniques are described elsewhere herein.

[0206] Construction of recombinant AAV vectors is described in a number of publications, including U.S. Pat. No. 5,173,414; Tratschin et al., *Mol. Cell. Biol.* 5:3251-3260 (1985); Tratschin, et al., *Mol. Cell. Biol.* 4:2072-2081 (1984); Hermonat & Muzyczka, *PNAS* 81:6466-6470 (1984); and Samulski et al., *J. Virol.* 63:03822-3828 (1989). Any of the techniques and/or methods can be used and/or adapted for constructing an AAV or other vector described herein. AAV vectors are discussed elsewhere herein.

[0207] In an embodiment, the vector can have one or more insertion sites, such as a restriction endonuclease recognition sequence (also referred to as a “cloning site”). In an embodiment, one or more insertion sites (e.g., about or more than about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, or more insertion sites) are located upstream and/or downstream of one or more sequence elements of one or more vectors.

[0208] Delivery vehicles, vectors, particles, nanoparticles, formulations and components thereof for expression of one or more elements of an engineered AAV capsid system described herein are as used in the foregoing documents, such as International Patent Application Publication WO 2014/093622 (PCT/US2013/074667) and are discussed in greater detail herein.

Virus Particle Production from Viral Vectors

AAV Particle Production

[0209] There are two main strategies for producing AAV particles from AAV vectors and systems thereof, such as those described herein, which depend on how the adenovirus helper factors are provided (helper v. helper free). In an embodiment, a method of producing AAV particles from AAV vectors and systems thereof can include adenovirus infection into cell lines that stably harbor AAV replication and capsid encoding polynucleotides along with AAV vector containing the polynucleotide to be packaged and delivered by the resulting AAV particle (e.g., the engineered AAV capsid polynucleotide(s)). In an embodiment, a method of producing AAV particles from AAV vectors and systems thereof can be a “helper free” method, which includes co-transfection of an appropriate producing cell line with three vectors (e.g., plasmid vectors): (1) an AAV vector that contains a polynucleotide of interest (e.g., a transgene encoding a therapeutic

protein or nucleic acid operably linked to a regulatory element that promotes expression in the target tissue) between 2 ITRs; (2) a vector that carries the AAV Rep-Cap encoding polynucleotide, including the engineered capsid protein described herein; and helper polynucleotides. One of skill in the art will appreciate various methods and variations thereof that are both helper and -helper free and as well as the different advantages of each system.

[0210] The engineered AAV vectors and systems thereof described herein can be produced by any of these methods.

Vector and Virus Particle Delivery

[0211] A vector (including non-viral carriers) described herein can be introduced into host cells to thereby produce transcripts, proteins, or peptides, including fusion proteins or peptides encoded by nucleic acids as described herein (e.g., engineered AAV capsid system transcripts, proteins, enzymes, mutant forms thereof, fusion proteins thereof, etc.), and virus particles (such as from viral vectors and systems thereof).

[0212] AAV capsids prepared from one or more engineered AAV capsid polynucleotides can be used to deliver a recombinant AAV genome encoding a therapeutic protein or nucleic acid of interest. Alternatively, adenovirus or other plasmid or viral vector types as previously described, can be used, in particular, using formulations and doses from, for example, US Patents Nos. 8,454,972 (formulations, doses for adenovirus), 8,404,658 (formulations, doses for AAV) and 5,846,946 (formulations, doses for DNA plasmids) and from clinical trials and publications regarding the clinical trials involving lentivirus, AAV and adenovirus. For examples, for AAV, the route of administration, formulation and dose can be as in US Patent No. 8,454,972 and as in clinical trials involving AAV. For Adenovirus, the route of administration, formulation and dose can be as in US Patent No. 8,404,658 and as in clinical trials involving adenovirus.

[0213] For plasmid delivery, the route of administration, formulation and dose can be as in US Patent No 5,846,946 and as in clinical studies involving plasmids. In an embodiment, doses can be based on or extrapolated to an average 70 kg individual (e.g., a male adult human), and can be adjusted for patients, subjects, mammals of different weight and species. Frequency of administration is within the ambit of the medical or veterinary practitioner (e.g., physician, veterinarian), depending on usual factors including the age, sex, general health, other conditions of the patient or subject and the particular condition or symptoms being addressed. The viral vectors can be injected into or otherwise delivered to the tissue or cell of interest.

[0214] In terms of *in vivo* delivery, AAV is advantageous over other viral vectors for a couple of reasons such as low toxicity (this may be due to the purification method not requiring ultra-centrifugation of cell particles that can activate the immune response) and a low probability of causing insertional mutagenesis because it does not integrate into the host genome.

[0215] The vector(s) and virus particles described herein can be delivered into a host cell *in vitro*, *in vivo*, and or *ex vivo*. Delivery can occur by any suitable method including, but not limited to, physical methods, chemical methods, and biological methods. Physical delivery methods are those methods that employ physical force to counteract the membrane barrier of the cells to facilitate intracellular delivery of the vector. Suitable physical methods include, but are not limited to, needles (e.g., injections), ballistic polynucleotides (e.g., particle bombardment, micro projectile gene transfer, and gene gun), electroporation, sonoporation, photoporation, magnetofection, hydroporation, and mechanical massage. Chemical methods are those methods that employ a chemical to elicit a change in the cells membrane permeability or other characteristic(s) to facilitate entry of the vector into the cell. For example, the environmental pH can be altered which can elicit a change in the permeability of the cell membrane. Biological methods are those that rely and capitalize on the host cell's biological processes or biological characteristics to facilitate transport of the vector (with or without a carrier) into a cell. For example, the vector and/or its carrier can stimulate an endocytosis or similar process in the cell to facilitate uptake of the vector into the cell.

[0216] Delivery of engineered AAV capsid system components (e.g., polynucleotides encoding engineered AAV capsid and/or capsid proteins) to cells via particles. The term "particle" as used herein, refers to any suitable sized particles for delivery of the engineered AAV capsid system components described herein. Suitable sizes include macro-, micro-, and nano-sized particles. In an embodiment, any of the of the engineered AAV capsid system components (e.g., polypeptides, polynucleotides, vectors, and combinations thereof described herein) can be attached to, coupled to, integrated with, otherwise associated with one or more particles or component thereof as described herein. The particles described herein can then be administered to a cell or organism by an appropriate route and/or technique. In an embodiment, particle delivery can be selected and be advantageous for delivery of the polynucleotide or vector components. It will be appreciated that in embodiments, particle delivery can also be advantageous for other engineered capsid system molecules and formulations described elsewhere herein.

Engineered Virus Particles Including an Engineered Viral (e.g., AAV) Capsid

[0217] Also described herein are engineered virus particles (also referred to here and elsewhere herein as “engineered viral particles” that can contain an engineered viral capsid (e.g., AAV capsid, referred to as “engineered AAV particles”) as described in detail elsewhere herein. It will be appreciated that the engineered AAV particles can be adenovirus-based particles, helper adenovirus-based particles, AAV-based particles, or hybrid adenovirus-based particles that contain at least one engineered AAV capsid proteins as previously described. An engineered AAV capsid is one that contains one or more engineered AAV capsid proteins as are described elsewhere herein. In an embodiment, the engineered AAV particles can include 1-60 engineered AAV capsid proteins described herein. In an embodiment, the engineered AAV particles can contain 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, or 60 engineered capsid proteins. In an embodiment, the engineered AAV particles can contain 0-59 wild-type AAV capsid proteins. In an embodiment, the engineered AAV particles can contain 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, or 59 wild-type AAV capsid proteins. The engineered AAV particles can thus include one or more transferrin receptor binding modifications as is previously described.

[0218] The engineered AAV particle can include one or more cargo polynucleotides. Cargo polynucleotides are discussed in greater detail elsewhere herein. Methods of making the engineered AAV particles from viral and non-viral vectors are described elsewhere herein. Formulations containing the engineered virus particles are described elsewhere herein.

Example Cargos

[0219] The transferrin receptor binding modifications can be coupled to or otherwise associated with a cargo. Cargos can include any molecule that is capable of being coupled to or associated with the transferrin receptor binding modifications described herein. Cargos can include, without limitation, nucleotides, oligonucleotides, polynucleotides, amino acids, peptides, polypeptides, riboproteins, lipids, sugars, pharmaceutically active agents (e.g., drugs, imaging and other diagnostic agents, and the like), chemical compounds, and combinations thereof. In an embodiment, the cargo is DNA, RNA, amino acids, peptides, polypeptides, antibodies, aptamers,

ribozymes, guide sequences for ribozymes that inhibit translation or transcription of essential tumor proteins and genes, hormones, immunomodulators, antipyretics, anxiolytics, antipsychotics, analgesics, antispasmodics, anti-inflammatories, antihistamines, anti-infectives, radiation sensitizers, chemotherapeutics, radioactive compounds, imaging agents, and combinations thereof. In an embodiment, the cargo is a recombinant AAV genome comprising a transgene, for example, encoding a therapeutic protein or nucleic acid, operably linked to regulatory sequences that direct expression of the therapeutic protein or nucleic acid in a target tissue, flanked by AAV ITR sequences.

[0220] In an embodiment, the cargo is capable of treating or preventing a neurological disease or disorder, details of which are described herein.

[0221] In an embodiment, the cargo is a morpholino, a peptide-linked morpholino, an antisense oligonucleotide, a PMO, a therapeutic transgene, a polynucleotide encoding a therapeutic polypeptide or peptide, a PPMO, one or more peptides, one or more polynucleotides encoding a CRISPR-Cas protein, a guide RNA, or both, a ribonucleoprotein, wherein the ribonucleoprotein comprises a CRISPR-Cas system molecule, a therapeutic transgene RNA, or other gene modifying or therapeutic RNA and/or protein, or any combination thereof.

[0222] In an embodiment, one or more transferrin receptor binding modifications described herein is directly attached to the cargo. In an embodiment, one or more transferrin receptor binding modifications described herein is indirectly coupled to the cargo, such as via a linker molecule. In an embodiment, one or more one or more transferrin receptor binding modifications described herein is coupled to associated with a polypeptide or other particle that is coupled to, attached to, encapsulates, and/or contains a cargo.

[0223] Exemplary particles include, without limitation, viral particles (e.g., viral capsids, which is inclusive of bacteriophage capsids), polysomes, liposomes, nanoparticles, microparticles, exosomes, micelles, and the like. The term “nanoparticle” as used herein includes a nanoscale deposit of a homogenous or heterogeneous material. Nanoparticles may be regular or irregular in shape and may be formed from a plurality of co-deposited particles that form a composite nanoscale particle. Nanoparticles may be generally spherical in shape or have a composite shape formed from a plurality of co-deposited generally spherical particles. Exemplary shapes for the nanoparticles include, but are not limited to, spherical, rod, elliptical, cylindrical, disc, and the like. In an embodiment, the nanoparticles have a substantially spherical shape.

Cargo Polynucleotides

[0224] Cargos are also described elsewhere herein. In an embodiment, the cargo is a cargo polynucleotide that can be packaged into an engineered viral particle and subsequently delivered to a cell. In an embodiment, delivery is cell selective, e.g., neurons and glial cells of the central nervous system. In an embodiment, the one or more cargo polynucleotides are part of the engineered viral (e.g., AAV) genome of the viral (e.g., AAV) system and packaged within the engineered capsid containing a transferrin receptor binding modification of the present disclosure. The cargo polynucleotides can be packaged into an engineered viral (e.g., AAV) particle, which can be delivered to, e.g., a cell. In an embodiment, the cargo polynucleotide can be capable of modifying a polynucleotide (e.g., gene or transcript) of a cell to which it is delivered. As used herein, “gene” can refer to a hereditary unit corresponding to a sequence of DNA that occupies a specific location on a chromosome and that contains the genetic instruction for a characteristic(s) or trait(s) in an organism. The term gene can refer to translated and/or untranslated regions of a genome. “Gene” can refer to the specific sequence of DNA that is transcribed into an RNA transcript that can be translated into a polypeptide or be a catalytic RNA molecule, including but not limited to, tRNA, siRNA, piRNA, miRNA, long-non-coding RNA and shRNA. Polynucleotide, gene, transcript, etc. modification includes all genetic engineering techniques including, but not limited to, gene editing as well as conventional recombinational gene modification techniques (e.g., whole or partial gene insertion, deletion, and mutagenesis (e.g., insertional and deletional mutagenesis) techniques.

[0225] In an embodiment, the cargo molecule is a polynucleotide that is or can encode a vaccine. In an embodiment, the cargo molecule is a polynucleotide encoding an antibody.

RNA Interference Agents

[0226] In an embodiment, the one or more polynucleotides may encode one or more RNA interference agents. RNA interference agents are RNA molecules capable of suppressing gene expressions. Examples of RNA interference agents include, but are not limited to, small interfering RNAs (siRNA), microRNAs (miRNA), and short hairpin RNAs (shRNA).

[0227] In an embodiment, the interference RNA may be a siRNAs. Small interfering RNA (siRNA) molecules are capable of inhibiting target gene expression by interfering RNA. siRNAs may be chemically synthesized, or may be obtained by in vitro transcription, or may be synthesized in vivo in target cell. siRNAs may comprise double-stranded RNA from 15 to 40 nucleotides in

length and can contain a protuberant region 3' and/or 5' from 1 to 6 nucleotides in length. Length of protuberant region is independent from total length of siRNA molecule. siRNAs may act by post-transcriptional degradation or silencing of target messenger. In some cases, the exogenous polynucleotides encode shRNAs. In shRNAs, the antiparallel strands that form siRNA are connected by a loop or hairpin region.

[0228] In an embodiment, the cargo polynucleotide is an RNAi molecule, antisense molecule, and/or a gene silencing oligonucleotide or a polynucleotide that encodes an RNAi molecule, antisense molecule, and/or gene silencing oligonucleotide.

[0229] As used herein, “gene silencing oligonucleotide” refers to any oligonucleotide that can alone or with other gene silencing oligonucleotides utilize a cell’s endogenous mechanisms, molecules, proteins, enzymes, and/or other cell machinery or exogenous molecule, agent, protein, enzyme, and/or polynucleotide to cause a global or specific reduction or elimination in gene expression, RNA level(s), RNA translation, RNA transcription, that can lead to a reduction or effective loss of a protein expression and/or function of a non-coding RNA as compared to wild-type or a suitable control. This is synonymous with the phrase “gene knockdown” Reduction in gene expression, RNA level(s), RNA translation, RNA transcription, and/or protein expression can range from about 100, 99, 98, 97, 96, 95, 94, 93, 92, 91, 90, 89, 88, 87, 86, 85, 84, 83, 82, 81, 80, 79, 78, 77, 76, 75, 74, 73, 72, 71, 70, 69, 68, 67, 66, 65, 64, 63, 62, 61, 60, 59, 58, 57, 56, 55, 54, 53, 52, 51, 50, 49, 48, 47, 46, 45, 44, 43, 42, 41, 40, 39, 38, 37, 36, 35, 34, 33, 32, 31, 30, 29, 28, 27, 26, 25, 24, 23, 22, 21, 20, 19, 18, 17, 16, 15, 14, 13, 12, 11, 10, 9, 8, 7, 6, 5, 4, 3, 2, to 1% or less reduction. “Gene silencing oligonucleotides” include, but are not limited to, any antisense oligonucleotide, ribozyme, any oligonucleotide (single or double stranded) used to stimulate the RNA interference (RNAi) pathway in a cell (collectively RNAi oligonucleotides), small interfering RNA (siRNA), microRNA, and short-hairpin RNA (shRNA). Commercially available programs and tools are available to design the nucleotide sequence of gene silencing oligonucleotides for a desired gene, based on the gene sequence and other information available to one of ordinary skill in the art.

[0230] In an embodiment, a cargo polynucleotide, such as an encoding polynucleotide, is flanked by at least a retroelement polypeptide encoding polynucleotide 3' UTR or portion thereof, such as the proximal region of about 500 base pairs of the 3' UTR. In an embodiment a cargo polynucleotide, such as an encoding polynucleotide, is flanked by a (e.g., endogenous or

engineered) retroelement polypeptide (such as a retroviral gag protein or gag homolog) 5' UTR. In an embodiment a cargo polynucleotide, such as an encoding polynucleotide, is flanked by an (e.g., endogenous or engineered) retroelement polypeptide encoding polynucleotide 5' and 3' UTR. In an embodiment, the flanking retroelement polypeptide encoding polynucleotide UTR(s) are from PNMA, Arc, PEG10 or other Sushi Class polypeptide. In an embodiment, the inclusion of the 3' UTR, the 5'UTR, or both can increase packaging and/or delivery of the cargo that they flank. These and other packaging elements are described in greater detail elsewhere herein.

Gene Modification Cargo Polynucleotides

[0231] In an embodiment, the cargo molecule can be a polynucleotide or polypeptide or polynucleotide encoding a polypeptide that can alone or when delivered as part of a system, whether or not delivered with other components of the system, operate to modify the genome, epigenome, and/or transcriptome of a cell to which it is delivered. Such systems include, but are not limited to, CRISPR-Cas systems. Other gene modification systems, e.g., Transcription Activator-like Effector (TALE)- or Zinc Finger Protein (ZFP)-based transcriptional activator; repressor; or epigenomic silencer, a RNA encoding a partial gene fragment designed for transplacing into an endogenous RNA, one or more transfer RNAs, or a component thereof, or an OMEGA system or any component thereof, Cre-Lox, morpholinos, etc. are other non-limiting examples of gene modification systems whose one or more components can be delivered by the engineered viral (e.g., AAV) particles described herein.

[0232] In an embodiment, the cargo molecule is or encodes a gene editing system or component thereof. In an embodiment, the cargo molecule is or encodes a CRISPR-Cas system molecule or a component thereof. In an embodiment, the cargo molecule is a polynucleotide that encodes one or more components of a gene modification system (such as a CRISPR-Cas system). In an embodiment the cargo molecule is or encodes a gRNA. CRISPR-Cas system as used herein is intended to encompass by Class 1 and Class 2 CRISPR-Cas systems and derivatives of CRISPR-Cas systems such as base editors, prime editors, and CRISPR-associated transposases (CAST) systems.

[0233] In an embodiment, the cargo molecule can be a polynucleotide or polypeptide or polynucleotide encoding a polypeptide that can alone or when delivered as part of a system, whether or not delivered with other components of the system, operate to modify the genome, epigenome, and/or transcriptome of a cell to which it is delivered, is such that it treats or prevents

a disease, a disorder, or a symptom thereof of a neurologic disease or disorder, and/or viruses (such as single stranded RNA viruses). In an embodiment, the cargo molecule, whether or not delivered with other components of the system, operates to modify the genome, epigenome, and/or transcriptome of a cell to which it is delivered, is such that it treats or prevents a neurological disease or disorder described further herein.

[0234] In an embodiment, the cargo molecule, whether or not delivered with other components of the system, operates to modify the genome, epigenome, and/or transcriptome of a cell to which it is delivered, is such that can modify the GAA gene, such as any of those described in US Pat. App. Pub. 20190284555, the contents of which are incorporated by reference as if expressed in their entirety herein and can be adapted for use with the present disclosure.

[0235] In an embodiment, the cargo molecule is or encodes an antisense oligomer or RNA molecule, such as those described in U.S. Pat. App. Pub. US20160251398, US20150267202, and US20180216111, the contents of which are incorporated by reference as if expressed in their entirety herein and can be adapted for use with the present disclosure.

[0236] In an embodiment, the cargo molecule can be a peptide-oligomer, conjugate as described in e.g., International Patent Application Publication WO2017106304A1, the contents of which are incorporated by reference as if expressed in their entirety herein and can be adapted for use with the present disclosure.

[0237] An embodiment of the disclosure encompasses methods of modifying a genomic locus of interest to change gene expression in a cell by introducing into the cell any of the compositions described herein.

[0238] An embodiment of the disclosure is that the above elements are comprised in a single composition or comprised in individual compositions. These compositions may advantageously be applied to a host to elicit a functional effect on the genomic level.

Polypeptides

[0239] In an embodiment, the cargo molecule may one or more polypeptides or may be a nucleic acid encoding a polypeptide. The polypeptide may be a full-length protein or a functional fragment or functional domain thereof, that is a fragment or domain that maintains the desired functionality of the full-length protein. As used within this section “protein” is meant to refer to full-length proteins and functional fragments and domains thereof. A wide array of polypeptides may be delivered using the engineered delivery vesicles described herein, including but not limited

to, secretory proteins, immunomodulatory proteins, anti-fibrotic proteins, proteins that promote tissue regeneration and/or transplant survival functions, hormones, anti-microbial proteins, anti-fibrillating polypeptides, and antibodies. The one or more polypeptides may also comprise combinations of the aforementioned example classes of polypeptides. It will be appreciated that any of the polypeptides described herein can also be delivered via the engineered delivery vesicles and systems described herein via delivery of the corresponding encoding polynucleotide.

Antibodies

[0240] In an embodiment, the one or more polypeptides may comprise one or more antibodies. The term "antibody" is used interchangeably with the term "immunoglobulin" herein, and includes intact antibodies, fragments of antibodies, e.g., Fab, F(ab')₂ fragments, scFv, and intact antibodies and fragments that have been mutated either in their constant and/or variable region (e.g., mutations to produce chimeric, partially humanized, or fully humanized antibodies, as well as to produce antibodies with a desired trait, e.g., enhanced binding and/or reduced FcR binding). The term "fragment" refers to a part or portion of an antibody or antibody chain comprising fewer amino acid residues than an intact or complete antibody or antibody chain. Fragments can be obtained via chemical or enzymatic treatment of an intact or complete antibody or antibody chain. Fragments can also be obtained by recombinant means. Exemplary fragments include Fab, Fab', F(ab')₂, Fabc, Fd, dAb, V_{HH} and scF and/or Fv fragments. As used herein, a preparation of antibody protein "having less than about 50% of non-antibody protein (also referred to herein as a "contaminating protein"), or of chemical precursors, is considered to be "substantially free." 40%, 30%, 20%, 10% and more preferably 5% (by dry weight) of non-antibody protein, or of chemical precursors is considered to be substantially free. When the antibody protein or biologically active portion thereof is recombinantly produced, it is also preferably substantially free of culture medium, i.e., culture medium represents less than about 30%, preferably less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume or mass of the protein preparation.

[0241] In an embodiment, the antibody is a fragment or portion thereof. In an example embodiment, the antibody is an epitope binding protein or portion thereof. The term "binding portion" of an antibody (or "antibody portion") includes one or more complete domains, e.g., a pair of complete domains, as well as fragments of an antibody that retain the ability to specifically bind to a target molecule. It has been shown that the binding function of an antibody can be

performed by fragments of a full-length antibody. Binding fragments are produced by recombinant DNA techniques, or by enzymatic or chemical cleavage of intact immunoglobulins. Binding fragments include Fab, Fab', F(ab')₂, Fabc, Fd, dAb, Fv, single chains, single-chain antibodies, e.g., scFv, and single domain antibodies.

[0242] In an embodiment, the cargo or antibody is an antibody fragment or portion. In an embodiment, the cargo is an epitope binding protein. Examples of portions of antibodies or epitope-binding proteins encompassed by the present definition include: (i) the Fab fragment, having V_L, C_L, V_H and C_{H1} domains; (ii) the Fab' fragment, which' is a Fab fragment having one or more cysteine residues at the C-terminus of the C_{H1} domain; (iii) the Fd fragment having V_H and C_{H1} domains; (iv) the Fd' fragment having V_H and C_{H1} domains and one or more cysteine residues at the C-terminus of the C_{H1} domain; (v) the Fv fragment having the V_L and V_H domains of a single arm of an antibody; (vi) the dAb fragment (Ward et al., 341 Nature 544 (1989)) which consists of a V_H domain or a V_L domain that binds antigen; (vii) isolated CDR regions or isolated CDR regions presented in a functional framework; (viii) F(ab')₂ fragments which are bivalent fragments including two Fab' fragments linked by a disulphide bridge at the hinge region; (ix) single chain antibody molecules (e.g., single chain Fv; scFv) (Bird et al., 242 Science 423 (1988); and Huston et al., 85 PNAS 5879 (1988)); (x) "diabodies" with two antigen binding sites, comprising a heavy chain variable domain (V_H) connected to a light chain variable domain (V_L) in the same polypeptide chain (see, e.g., EP 404,097; WO 93/11161; Hollinger et al., 90 PNAS 6444 (1993)); (xi) "linear antibodies" comprising a pair of tandem Fd segments (V_H-C_{H1}-V_H-C_{H1}) which, together with complementary light chain oligopeptides, form a pair of "antigen binding regions" (Zapata et al., Protein Eng. 8(10):1057-62 (1995); and U.S. Patent No. 5,641,870).

[0243] The term "antigen-binding fragment" refers to a polypeptide fragment of an immunoglobulin or antibody that binds antigen or competes with intact antibody (i.e., with the intact antibody from which they were derived) for antigen binding (i.e., specific binding). As such these antibodies or fragments thereof are included in the scope of the disclosure, provided that the antibody or fragment binds specifically to a target molecule.

[0244] In an embodiment, the antibody is a single-chain antibody (scFvs). The term "single-chain variable fragment", as used herein refers to a fusion protein containing the variable region(s) of the heavy (V_H) and light (V_L) of an immunoglobulin that are connected via a linker peptide. The linker peptide typically ranges from about 10 to about 25 amino acids. The linker can be flexible

and can contain one or more glycine residues for flexibility. The linker can contain one or more serine or threonine residues to increase or modify solubility. The V_H and light (V_L) can be linked via the linker in any order. In an embodiment, N terminus of the V_H and is coupled, via a linker, C terminus of the (V_L). In an embodiment, C terminus of the V_H and is coupled, via a linker, N terminus of the (V_L). In an embodiment, the scFV is a bivalent or trivalent scFvs. In an embodiment bivalent or trivalent scFvs are bi or trispecific, meaning that they can target 2 or 3, respectively, different epitopes. See also e.g., Hollinger, Philipp; Prospero, T; Winter, G (July 1993). "Diabodies": small bivalent and bispecific antibody fragments". *Proceedings of the National Academy of Sciences of the United States of America*. **90** (14): 6444–8; incq, S; Bosman, F; Buyse, MA; Degrieck, R; Celis, L; De Boer, M; Van Doorselaere, V; Sablon, E (2001). "Expression and purification of monospecific and bispecific recombinant antibody fragments derived from antibodies that block the CD80/CD86-CD28 costimulatory pathway". *Protein Expression and Purification*. **22** (1): 11–24. doi:10.1006/prep.2001.1417; Le Gall, F.; Kipriyanov, SM; Moldenhauer, G; Little, M (1999). "Di-, tri- and tetrameric single chain Fv antibody fragments against human CD19: effect of valency on cell binding". *FEBS Letters*. **453** (1): 164–168. doi:10.1016/S0014-5793(99)00713-9; Huston, J. S.; Levinson, D.; Mudgett-Hunter, M.; Tai, M. S.; Novotný, J.; Margolies, M. N.; Crea, R. (1988). "Protein engineering of antibody binding sites: recovery of specific activity in an anti-digoxin single-chain Fv analogue produced in *Escherichia coli*". *Proceedings of the National Academy of Sciences of the United States of America*. **85** (16): 5879–5883; de Graaf et al., *Methods Mol Biol*. 2002;178:379-87. doi: 10.1385/1-59259-240-6:379; Zhou, H.X., *J Mol Biol*. 2003 May 23;329(1):1-8. doi: 10.1016/s0022-2836(03)00372-3; Bird and Walker. *Trends Biotechnol*. 1991 Apr;9(4):132-7. doi: 10.1016/0167-7799(91)90044-I; Wörn et al., *J Mol Biol*. 2001 Feb 2;305(5):989-1010. doi: 10.1006/jmbi.2000.4265.

[0245] As used herein, "heavy chain antibody," "VHH" or "single-domain antibodies" (sdAbs) refers to an antibody which is composed only of two heavy chains and lacks the two light chains usually found in antibodies (see, e.g., Henry and MacKenzie, *Antigen recognition by single-domain antibodies: structural latitudes and constraints*. *MAbs*. 2018 Aug-Sep; 10(6): 815–826). VHH can refer to an antibody or VHH domain. Single-domain antibodies (sdAb) are also referred to as a "nanobody", which is defined herein as an antibody fragment composed of a single monomeric variable antibody domain. As used herein "VHH" is used interchangeably with

“nanobody.” The ~12–15 kDa variable domains of these antibodies (VHHs and VNARs) can be produced recombinantly and can recognize antigen in the absence of the remainder of the antibody heavy chain. In common antibodies, the antigen binding region consists of the variable domains of the heavy and light chains (VH and VL). Heavy-chain antibodies can bind antigens despite having only VH domains. In an embodiment, the heavy chain antibody is an antibody derived from cartilaginous fishes (immunoglobulin new antigen receptor (IgNAR)) or camelid ungulates. Non-limiting examples of camelids include dromedaries, camels, llamas and alpacas.

[0246] It is intended that the term “antibody” encompass any Ig class or any Ig subclass (e.g., the IgG1, IgG2, IgG3, and IgG4 subclasses of IgG) obtained from any source (e.g., humans and non-human primates, and in rodents, lagomorphs, caprines, bovines, equines, ovines, etc.).

[0247] The term “Ig class” or “immunoglobulin class”, as used herein, refers to the five classes of immunoglobulin that have been identified in humans and higher mammals, IgG, IgM, IgA, IgD, and IgE. The term “Ig subclass” refers to the two subclasses of IgM (H and L), three subclasses of IgA (IgA1, IgA2, and secretory IgA), and four subclasses of IgG (IgG1, IgG2, IgG3, and IgG4) that have been identified in humans and higher mammals. The antibodies can exist in monomeric or polymeric form; for example, IgM antibodies exist in pentameric form, and IgA antibodies exist in monomeric, dimeric or multimeric form.

[0248] The term “IgG subclass” refers to the four subclasses of immunoglobulin class IgG - IgG1, IgG2, IgG3, and IgG4 that have “been identified in humans and higher mammals by the heavy chains of the immunoglobulins, $VH1 - VH4$, respectively. The term “single-chain immunoglobulin” or “single-chain antibody” (used interchangeably herein) refers to a protein having a two-polypeptide chain structure consisting of a heavy and a light chain, said chains being stabilized, for example, by interchain peptide linkers, which has the ability to specifically bind antigen. The term “domain” refers to a globular region of a heavy or light chain polypeptide comprising peptide loops (e.g., comprising 3 or 4 peptide loops) stabilized, for example, by β pleated sheet and/or intrachain disulfide bond. Domains are further referred to herein as “constant” or “variable”, based on the relative lack of sequence variation within the domains of various class members in the case of “constant” domain, or the significant variation within the domains of various class members in the case of a “variable” domain. Antibody or polypeptide “domains” are often referred to interchangeably in the antibody or polypeptide “regions”. The “constant” domains of an antibody light chain are referred to interchangeably as “light chain constant regions”, “light

chain constant domains", "CL" regions or "CL" domains." The "constant" domains of an antibody heavy chain are referred to interchangeably as "heavy chain constant regions", "heavy chain constant domains", "CH" regions or "CH" domains." The "variable" domains of an antibody light chain are referred to interchangeably as "light chain variable regions", "light chain variable domains", "VL" regions or "VL" domains." The "variable" domains of an antibody heavy chain are referred to interchangeably as "heavy chain constant regions", "heavy chain constant domains", "VH" regions or "VH" domains.

[0249] The term "region" can also refer to a part or portion of an antibody chain or antibody chain domain (e.g., a part or portion of a heavy or light chain or a part or portion of a constant or variable domain, as defined herein), as well as more discrete parts or portions of said chains or domains. For example, light and heavy chains or light and heavy chain variable domains include "complementarity determining regions" or "CDRs" interspersed among "framework regions" or "FRs", as defined herein.

[0250] The term "conformation" refers to the tertiary structure of a protein or polypeptide (e.g., an antibody, antibody chain, domain or region thereof). For example, the phrase "light (or heavy) chain conformation" refers to the tertiary structure of a light (or heavy) chain variable region, and the phrase "antibody conformation" or "antibody fragment conformation" refers to the tertiary structure of an antibody or fragment thereof.

[0251] The term "antibody-like protein scaffolds" or "engineered protein scaffolds" broadly encompasses proteinaceous non-immunoglobulin specific-binding agents, typically obtained by combinatorial engineering (such as site-directed random mutagenesis in combination with phage display or other molecular selection techniques). Usually, such scaffolds are derived from robust and small soluble monomeric proteins (such as Kunitz inhibitors or lipocalins) or from a stably folded extra-membrane domain of a cell surface receptor (such as protein A, fibronectin or the ankyrin repeat).

[0252] Such scaffolds have been extensively reviewed in Binz et al. (Engineering novel binding proteins from non-immunoglobulin domains. *Nat Biotechnol* 2005, 23:1257-1268), Gebauer and Skerra (Engineered protein scaffolds as next-generation antibody therapeutics. *Curr Opin Chem Biol.* 2009, 13:245-55), Gill and Damle (Biopharmaceutical drug discovery using novel protein scaffolds. *Curr Opin Biotechnol* 2006, 17:653-658), Skerra (Engineered protein scaffolds for molecular recognition. *J Mol Recognit* 2000, 13:167-187), and Skerra (Alternative

non-antibody scaffolds for molecular recognition. *Curr Opin Biotechnol* 2007, 18:295-304), and include without limitation affibodies, based on the Z-domain of staphylococcal protein A, a three-helix bundle of 58 residues providing an interface on two of its alpha-helices (Nygren, *Alternative binding proteins: Affibody binding proteins developed from a small three-helix bundle scaffold. FEBS J* 2008, 275:2668-2676); engineered Kunitz domains based on small polypeptides of 58 residues) and robust, disulphide-crosslinked serine protease inhibitor, typically of human origin (e.g., LACI-D1), which can be engineered for different protease specificities (Nixon and Wood, *Engineered protein inhibitors of proteases. Curr Opin Drug Discov Dev* 2006, 9:261-268); monobodies or adnectins based on the 10th extracellular domain of human fibronectin III (10Fn3), which adopts an Ig-like beta-sandwich fold (94 residues) with 2–3 exposed loops but lacks the central disulphide bridge (Koide and Koide, *Monobodies: antibody mimics based on the scaffold of the fibronectin type III domain. Methods Mol Biol* 2007, 352:95-109); anticalins derived from the lipocalins, a diverse family of eight-stranded beta-barrel proteins (ca. 180 residues) that naturally form binding sites for small ligands by means of four structurally variable loops at the open end, which are abundant in humans, insects, and many other organisms (Skerra, *Alternative binding proteins: Anticalins—harnessing the structural plasticity of the lipocalin ligand pocket to engineer novel binding activities. FEBS J* 2008, 275:2677-2683); DARPins, designed ankyrin repeat domains (166 residues), which provide a rigid interface arising from typically three repeated beta-turns (Stumpp et al., *DARPins: a new generation of protein therapeutics. Drug Discov Today* 2008, 13:695-701); avimers (multimerized LDLR-A module) (Silverman et al., *Multivalent avimer proteins evolved by exon shuffling of a family of human receptor domains. Nat Biotechnol* 2005, 23:1556-1561); and cysteine-rich knottin peptides (Kolmar” *Alternative binding proteins: biological activity and therapeutic potential of cystine-knot miniproteins. FEBS J* 2008, 275:2684-2690).

[0253] "Specific binding" of an antibody means that the antibody exhibits appreciable affinity for a particular antigen or epitope and, generally, does not exhibit significant cross reactivity. "Appreciable" binding includes binding with an affinity of at least 25 μM . Antibodies with affinities greater than $1 \times 10^7 \text{ M}^{-1}$ (or a dissociation coefficient of $1 \mu\text{M}$ or less or a dissociation coefficient of 1nm or less) typically bind with correspondingly greater specificity. Values intermediate of those set forth herein are also intended to be within the scope of the present disclosure and antibodies of the disclosure bind with a range of affinities, for example, 100nM or

less, 75nM or less, 50nM or less, 25nM or less, for example 10nM or less, 5nM or less, 1nM or less, or in embodiments 500pM or less, 100pM or less, 50pM or less or 25pM or less. An antibody that "does not exhibit significant cross reactivity" is one that will not appreciably bind to an entity other than its target (e.g., a different epitope or a different molecule). For example, an antibody that specifically binds to a target molecule will appreciably bind the target molecule but will not significantly react with non-target molecules or peptides. An antibody specific for a particular epitope will, for example, not significantly cross react with remote epitopes on the same protein or peptide. Specific binding can be determined according to any art-recognized means for determining such binding. Preferably, specific binding is determined according to Scatchard analysis and/or competitive binding assays.

[0254] As used herein, the term "affinity" refers to the strength of the binding of a single antigen-combining site with an antigenic determinant. Affinity depends on the closeness of stereochemical fit between antibody combining sites and antigen determinants, on the size of the area of contact between them, on the distribution of charged and hydrophobic groups, etc. Antibody affinity can be measured by equilibrium dialysis or by the kinetic BIACORE™ method. The dissociation constant, K_d , and the association constant, K_a , are quantitative measures of affinity.

[0255] As used herein, the term "monoclonal antibody" refers to an antibody derived from a clonal population of antibody-producing cells (e.g., B lymphocytes or B cells) which is homogeneous in structure and antigen specificity. The term "polyclonal antibody" refers to a plurality of antibodies originating from different clonal populations of antibody-producing cells which are heterogeneous in their structure and epitope specificity, but which recognize a common antigen. Monoclonal and polyclonal antibodies may exist within bodily fluids, as crude preparations, or may be purified, as described herein. "Humanized" forms of non-human (e.g., murine) antibodies are chimeric antibodies that contain minimal sequence derived from non-human immunoglobulin. For the most part, humanized antibodies are human immunoglobulins (recipient antibody) in which residues from a hypervariable region of the recipient are replaced by residues from a hypervariable region of a non-human species (donor antibody) such as mouse, rat, rabbit, or nonhuman primate having the desired specificity, affinity, and capacity. In some instances, FR residues of the human immunoglobulin are replaced by corresponding non-human residues. Furthermore, humanized antibodies may comprise residues that are not found in the

recipient antibody or in the donor antibody. These modifications are made to further refine antibody performance. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or substantially all of the hypervariable regions correspond to those of a non-human immunoglobulin and all or substantially all of the FR regions are those of a human immunoglobulin sequence. The humanized antibody optionally also will comprise at least a portion of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin.

[0256] As used herein, a "blocking" antibody or an antibody "antagonist" is one which inhibits or reduces biological activity of the antigen(s) it binds. In an embodiment, the blocking antibodies or antagonist antibodies or portions thereof described herein completely inhibit the biological activity of the antigen(s).

[0257] Antibodies may act as agonists or antagonists of the recognized polypeptides. For example, the present disclosure includes antibodies which disrupt receptor/ligand interactions either partially or fully. The disclosure features both receptor-specific antibodies and ligand-specific antibodies. The disclosure also features receptor-specific antibodies which do not prevent ligand binding but prevent receptor activation. Receptor activation (i.e., signaling) may be determined by techniques described herein or otherwise known in the art. For example, receptor activation can be determined by detecting the phosphorylation (e.g., tyrosine or serine/threonine) of the receptor or of one of its down-stream substrates by immunoprecipitation followed by western blot analysis. In specific embodiments, antibodies are provided that inhibit ligand activity or receptor activity by at least 95%, at least 90%, at least 85%, at least 80%, at least 75%, at least 70%, at least 60%, or at least 50% of the activity in absence of the antibody.

[0258] The disclosure also features receptor-specific antibodies which both prevent ligand binding and receptor activation as well as antibodies that recognize the receptor-ligand complex. Likewise, encompassed by the disclosure are neutralizing antibodies which bind the ligand and prevent binding of the ligand to the receptor, as well as antibodies which bind the ligand, thereby preventing receptor activation, but do not prevent the ligand from binding the receptor. Further included in the disclosure are antibodies which activate the receptor. These antibodies may act as receptor agonists, i.e., potentiate or activate either all or a subset of the biological activities of the ligand-mediated receptor activation, for example, by inducing dimerization of the receptor. The antibodies may be specified as agonists, antagonists or inverse agonists for biological activities

comprising the specific biological activities of the peptides disclosed herein. The antibody agonists and antagonists can be made using methods known in the art. See, e.g., PCT publication WO 96/40281; U.S. Pat. No. 5,811,097; Deng et al., *Blood* 92 (6):1981-1988 (1998); Chen et al., *Cancer Res.* 58(16):3668-3678 (1998); Harrop et al., *J. Immunol.* 161(4):1786-1794 (1998); Zhu et al., *Cancer Res.* 58(15):3209-3214 (1998); Yoon et al., *J. Immunol.* 160(7):3170-3179 (1998); Prat et al., *J. Cell. Sci.* III (Pt2):237-247 (1998); Pitard et al., *J. Immunol. Methods* 205 (2):177-190 (1997); Liautard et al., *Cytokine* 9(4):233-241 (1997); Carlson et al., *J. Biol. Chem.* 272(17):11295-11301 (1997); Taryman et al., *Neuron* 14(4):755-762 (1995); Muller et al., *Structure* 6(9):1153-1167 (1998); Bartunek et al., *Cytokine* 8(1):14-20 (1996).

[0259] The antibodies as defined for the present disclosure include derivatives that are modified, i.e., by the covalent attachment of any type of molecule to the antibody such that covalent attachment does not prevent the antibody from generating an anti-idiotypic response. For example, but not by way of limitation, the antibody derivatives include antibodies that have been modified, e.g., by glycosylation, acetylation, pegylation, phosphorylation, amidation, derivatization by known protecting/blocking groups, proteolytic cleavage, linkage to a cellular ligand or other protein, etc. Any of numerous chemical modifications may be carried out by known techniques, including, but not limited to specific chemical cleavage, acetylation, formylation, metabolic synthesis of tunicamycin, etc. Additionally, the derivative may contain one or more non-classical amino acids.

Programmable Nuclease

[0260] As used herein, a programmable nuclease refers to a nuclease capable of forming a complex with said guide molecule and wherein the site of nuclease activity on a genomic loci is dictated by the guide molecule. A programmable nuclease may comprise a CRISPR-Cas system or a component thereof (e.g., a Cas nuclease) or an OMEGA system or a component thereof (e.g., an IscB nuclease, an IsrB nuclease, an IshB nuclease, a TnpB nuclease, a Fanzor, etc.) (see e.g., Altae-Tran, H.; *et al.* The Widespread IS200/IS605 Transposon Family Encodes Diverse Programmable RNA-Guided Endonucleases. *Science*, 2021, 374:57–65; Karvelis et al., *Nature*, 599: 692–696 (2021); Hirano *et al.*, *Nature*, 610:575–581 (2022); and Saito et al., *Nature*, 620:660–668 (2023); Jiang et al., *Science Advances*, 9(39), DOI: 10.1126/sciadv.adk0171 (2023)). The same programmable nuclease may be used with the entire set of guide molecules, or the set of

molecules may comprise one or more guides capable of complexing with different types of programmable nucleases.

CRISPR-Cas Systems

[0261] The programmable nuclease may be a Cas nuclease. CRISPR-Cas systems can generally fall into two classes based on their architectures of their effector molecules, which are each further subdivided by type and subtype. The two classes are Class 1 and Class 2. Class 1 CRISPR-Cas systems have effector modules composed of multiple Cas proteins, some of which form crRNA-binding complexes, while Class 2 CRISPR-Cas systems include a single, multi-domain crRNA-binding protein.

[0262] In some embodiments, the CRISPR-Cas system that can be used to modify a polynucleotide as described herein can be a Class 1 CRISPR-Cas system. Class 1 CRISPR-Cas systems are divided into types I, III, and IV. Makarova et al. 2020. Nat. Rev. 18: 67-83., particularly as described in Figure 1. Type I CRISPR-Cas systems are divided into 9 subtypes (I-A, I-B, I-C, I-D, I-E, I-F1, I-F2, I-F3, and IG). Makarova *et al.*, 2020. Class 1, Type I CRISPR-Cas systems can contain a Cas3 protein that can have helicase activity. Type III CRISPR-Cas systems are divided into 6 subtypes (III-A, III-B, III-C, III-D, III-E, and III-F). Type III CRISPR-Cas systems can contain a Cas10 that can include an RNA recognition motif called Palm and a cyclase domain that can cleave polynucleotides. Makarova *et al.*, 2020. Type IV CRISPR-Cas systems are divided into 3 subtypes. (IV-A, IV-B, and IV-C). Makarova *et al.*, 2020. Class 1 systems also include CRISPR-Cas variants, including Type I-A, I-B, I-E, I-F and I-U variants, which can include variants carried by transposons and plasmids, including versions of subtype I-F encoded by a large family of Tn7-like transposon and smaller groups of Tn7-like transposons that encode similarly degraded subtype I-B systems. Peters *et al.*, PNAS 114 (35) (2017); DOI: 10.1073/pnas.1709035114; *see also*, Makarova et al. 2018. The CRISPR Journal, v. 1, n5, Figure 5.

[0263] The compositions, systems, and methods described in greater detail elsewhere herein can be designed and adapted for use with Class 2 CRISPR-Cas systems. Thus, in some embodiments, the CRISPR-Cas system is a Class 2 CRISPR-Cas system. Class 2 systems are distinguished from Class 1 systems in that they have a single, large, multi-domain effector protein. In certain example embodiments, the Class 2 system can be a Type II, Type V, or Type VI system, which are described in Makarova et al. “Evolutionary classification of CRISPR-Cas systems: a

burst of class 2 and derived variants” Nature Reviews Microbiology, 18:67-81 (Feb 2020), incorporated herein by reference. Each type of Class 2 system is further divided into subtypes. *See* Markova et al. 2020, particularly at Figure. 2. Class 2, Type II systems can be divided into 4 subtypes: II-A, II-B, II-C1, and II-C2. Class 2, Type V systems can be divided into 17 subtypes: V-A, V-B1, V-B2, V-C, V-D, V-E, V-F1, V-F1(V-U3), V-F2, V-F3, V-G, V-H, V-I, V-K (V-U5), V-U1, V-U2, and V-U4. Class 2, Type VI systems can be divided into 5 subtypes: VI-A, VI-B1, VI-B2, VI-C, and VI-D.

OMEGA nucleases

[0264] The programmable nuclease may be an OMEGA nuclease. The OMEGA nuclease may be an IscB nuclease, an IsrB nuclease, a TnpB nuclease, or a Fanzor nuclease.

IscB nuclease

[0265] In one embodiment, IscB nucleases may comprise a split RuvC nuclease domain comprising RuvC-I, RuvC-II, and RuvC-III subdomains. Some IscB proteins may further comprise a HNH endonuclease domain. In one example embodiment, the RuvC endonuclease domain is split by the insertion of a bridge helix, a HNH domain, or both. However, unlike Cas9, IscB nucleases do not contain a Rec domain. In addition, IscB nucleases may further comprise a conserved N-terminal domain (also referred to herein as a PLMP domain), which is not present in Cas9 proteins. IscB proteins may also further comprise a conserved C-terminal domain. In one example embodiment, an IscB nuclease comprises, moving from the N- to C-terminus, a PLMP domain, a RuvC-I subdomain, a bridge helix, a RuvC-II subdomain, a HNH domain, a RuvC-III subdomain, and a C terminal domain. In one embodiment, IscB nucleic acid-guided nucleases may comprise CRISPR-associated IscB nucleases. In one embodiment, the IscB nucleases are CRISPR-associated proteins, e.g., the loci of the nucleases are associated with an CRISPR array. In one embodiment the IscBs may be referred to as Cas IscBs. The Cas IscB nucleic acid-guided nuclease may comprise one or more domains, e.g., one or more of a X domain (e.g., at N-terminus), a RuvC domain, a Bridge Helix domain, and a Y domain (e.g., at C-terminus). See International Application Publication No. WO 2022/087494 A1 incorporated herein by reference in its entirety.

IsrB nuclease

[0266] IsrBs are homologs of IscB nucleases. IsrB nucleases comprise the PLMP and RuvC domains but do not comprise a HNH domain. In one embodiment, the IsrB nuclease comprises a PLMP domain and a split RuvC but lacks the HNH domain present between the RuvC-II and III

subdomains in IscB nucleases. In one embodiment, the IsrB is an OMEAG RNA guided nickase. In one embodiment, the OMEAG RNA guided IsrB nicks a DNA target. In one embodiment, the DNA target is a dsDNA and the nicks occurs on the non-target strand of the dsDNA target. In one embodiment, the IsrB nicks the dsDNA in a guide and TAM specific manner. Accordingly, applications where a nickase is utilized can be used with the IsrB nucleases detailed herein in a manner functionally similar to an IscB that has been inactivated at the HNH domain.

IshB nuclease

[0267] As noted above IshBs are IscB homologs and may be referred to herein as an Insertion sequence HNH-like OrfB (IshB) nuclease. IshB nucleases are generally smaller than IsrB or IscB nucleases and contain only the PLMP and HNH domain, but no RuvC domain. In one embodiment, the IshB, or IscB homolog, comprises a PLMP domain and an HNH domain, but does not comprise a RuvC domain.

[0268] Some IshB nucleases may be part of the IS605 OrfB family of transposases. In an embodiment, the IshB nuclease is from *Actinoplanes lobatus* and has the Genbank accession number MBB4752409. In an embodiment, the RefSeq database accession number for the nuclease with accession number MBB4752409 is WP_188124268 and the INSDC number is GGN95087.

TnpB nuclease

[0269] In one aspect, embodiments disclosed herein are directed to compositions comprising a TnpB and an OMEGA RNA capable of forming a complex with the TnpB and directing site-specific binding of the TnpB to a target sequence on a target polynucleotide.

[0270] TnpB nucleases may comprise a RuvC-like domain. Exemplary TnpB sequences are shown in FIG. 1, Table 1A, Table 1B, Table 1C and Table 5 of International Patent Publication Application No. WO 2022/159892 A1, herein incorporated by reference in its entirety. The RuvC domain may be a split RuvC domain comprising RuvC-I, RuvC-II, and RuvC-III subdomains. The TnpB may further comprise one or more of a HTH domain, a bridge helix domain and a zinc finger domain. TnpB nucleases do not comprise an HNH domain. In one example embodiment, TnpB proteins comprise, starting at the N-terminus a HTH domain, a RuvC-I sub-domain, a bridge helix domain, a RuvC-II sub-domain, a zinc finger domain, and a RuvC-III sub-domain. In one example embodiment, the RuvC-III sub-domain forms the C-terminus of the TnpB nuclease.

[0271] In one embodiment, the TnpB nuclease is from *Epsilonproteobacteria bacterium*, or *Actinoplanes lobatus* strain DSM 43150, *Actinomadura cellulositica* strain DSM 45823,

Actinomadura namibiensis strain DSM 44197, *Alicyclobacillus macrosprangioides* strain DSM 17980, *Lipingzhangella halophila* strain DSM 102030, or *Ktedonobacter racemifer*. In one embodiment, the TnpB nuclease is from *Ktedonobacter racemifer*, or comprises a conserved RNA region with similarity to the 5' ITR of *K. racemifer* TnpB loci.

Fanzor nuclease

[0272] In one aspect, embodiments disclosed herein are directed to compositions comprising an engineered Fanzor and/or OMEGA RNA capable of forming a complex with the Fanzor and directing site-specific binding of the Fanzor to a target sequence on a target polypeptide.

[0273] Fanzor nucleases may comprise a RuvC-like domain. Exemplary Fanzor sequences are shown or encoded by those in Table 1, Table 6, Table 7, Table 8, Table 9, Table 10, Table 11, Table 12, Table 13, Table 14, and FIG. 20 of International Patent Application Publication No. WO 2023/114872 A3, herein incorporated by reference in its entirety. In some embodiments, the Fanzor nuclease is a nuclease as shown and described in relation with FIGS. 10C-10E, FIG. 35, and FIG. 56A-56D of International Patent Application Publication No. WO 2023/114872 A3. The RuvC domain may be a split RuvC domain comprising a RuvC-I, RuvC-II, and RuvC-III subdomains. The Fanzor may further comprise one or more of a HTH domain, a bridge helix domain, a REC domain, a zinc finger domain, or any combination thereof. Fanzor nucleases do not comprise an HNH domain. In one example embodiment, Fanzor proteins comprise, starting at the N-terminus a HTH domain, a RuvC-I sub-domain, a bridge helix domain, a RuvC-II sub-domain, a zinc finger domain, and a RuvC-III sub-domain. In one example embodiment, the RuvC-III sub-domain forms the C-terminus of the Fanzor nuclease.

Engineered Cells and Organisms Expressing said Engineered AAV capsids

[0274] Described herein are engineered cells that can include one or more of the engineered AAV capsid polynucleotides, polypeptides, vectors, and/or vector systems. In an embodiment, one or more of the engineered AAV capsid polynucleotides can be expressed in the engineered cells. In an embodiment, the engineered cells can be capable of producing engineered AAV capsid proteins and/or engineered AAV capsid particles that are described elsewhere herein. Also described herein are modified or engineered organisms that can include one or more engineered cells described herein. The engineered cells can be engineered to express a cargo molecule (e.g., a cargo polynucleotide) dependently or independently of an engineered AAV capsid polynucleotide

as described elsewhere herein, *e.g.*, packaged within an engineered AAV capsid as described herein.

[0275] A wide variety of animals, plants, algae, fungi, yeast, etc. and animal, plant, algae, fungus, yeast cell or tissue systems may be engineered to express one or more nucleic acid constructs of the engineered AAV capsid system described herein using various transformation methods mentioned elsewhere herein. This can produce organisms that can produce engineered AAV capsid particles, such as for production purposes, engineered AAV capsid design and/or generation, and/or model organisms. In an embodiment, the polynucleotide(s) encoding one or more components of the engineered AAV capsid system described herein can be stably or transiently incorporated into one or more cells of a plant, animal, algae, fungus, and/or yeast or tissue system. In an embodiment, one or more of engineered AAV capsid system polynucleotides are genomically incorporated into one or more cells of a plant, animal, algae, fungus, and/or yeast or tissue system. Further embodiments of the modified organisms and systems are described elsewhere herein. In an embodiment, one or more components of the engineered AAV capsid system described herein are expressed in one or more cells of the plant, animal, algae, fungus, yeast, or tissue systems.

Engineered Cells

[0276] Described herein are various embodiments of engineered cells that can include one or more of the engineered AAV capsid system polynucleotides, polypeptides, vectors, and/or vector systems described elsewhere herein. In an embodiment, the cells can express one or more of the engineered AAV capsid polynucleotides and can produce one or more engineered AAV capsid particles, which are described in greater detail herein. Such cells are also referred to herein as “producer cells”. It will be appreciated that these engineered cells are different from “modified cells” described elsewhere herein in that the modified cells are not necessarily producer cells unless they include one or more of the engineered AAV capsid polynucleotides, engineered AAV capsid vectors or other vectors described herein that render the cells capable of producing an engineered AAV capsid particle. Modified cells can be recipient cells of an engineered AAV capsid particles and can, in an embodiment, be modified by the engineered AAV capsid particle(s) and/or a cargo polynucleotide delivered to the recipient cell. Modified cells are discussed in greater detail elsewhere herein. The term modification can be used in connection with modification of a cell that

is not dependent on being a recipient cell. For example, isolated cells can be modified prior to receiving an engineered AAV capsid molecule.

[0277] In an embodiment, the disclosure provides a non-human eukaryotic organism; for example, a multicellular eukaryotic organism, including a eukaryotic host cell containing one or more components of an engineered delivery system described herein according to any of the described embodiments. In other embodiments, the disclosure provides a eukaryotic organism; preferably a multicellular eukaryotic organism, comprising a eukaryotic host cell containing one or more components of an engineered delivery system described herein according to any of the described embodiments. In an embodiment, the organism is a host of AAV.

[0278] In an embodiment, the plants, algae, fungi, yeast, etc., cells or parts obtained are transgenic plants, comprising an exogenous DNA sequence incorporated into the genome of all or part of the cells.

[0279] The engineered cell can be a prokaryotic cell. The prokaryotic cell can be bacterial cell. The prokaryotic cell can be an archaea cell. The bacterial cell can be any suitable bacterial cell. Suitable bacterial cells can be from the genus *Escherichia*, *Bacillus*, *Lactobacillus*, *Rhodococcus*, *Rodhobacter*, *Synechococcus*, *Synechoystis*, *Pseudomonas*, *Psedoaltermonas*, *Stenotrophamonas*, and *Streptomyces*. Suitable bacterial cells include, but are not limited to *Escherichia coli* cells, *Caulobacter crescentus* cells, *Rodhobacter sphaeroides* cells, *Psedoaltermonas haloplanktis* cells. Suitable strains of bacterial include, but are not limited to BL21(DE3), DL21(DE3)-pLysS, BL21 Star-pLysS, BL21-SI, BL21-AI, Tuner, Tuner pLysS, Origami, Origami B pLysS, Rosetta, Rosetta pLysS, Rosetta-gami-pLysS, BL21 CodonPlus, AD494, BL2trxB, HMS174, NovaBlue(DE3), BLR, C41(DE3), C43(DE3), Lemo21(DE3), Shuffle T7, ArcticExpress and ArticExpress (DE3).

[0280] The engineered cell can be a eukaryotic cell. The eukaryotic cells may be those of or derived from a particular organism, such as a plant or a mammal, including but not limited to human, or non-human eukaryote or animal or mammal as herein discussed, e.g., mouse, rat, rabbit, dog, livestock, or non-human mammal or primate. In an embodiment the engineered cell can be a cell line. Examples of cell lines include, but are not limited to, C8161, CCRF-CEM, MOLT, mIMCD-3, NHDF, HeLa-S3, Huh1, Huh4, Huh7, HUVEC, HASMC, HEK_n, HEK_a, MiaPaCell, Panc1, PC-3, TF1, CTLL-2, C1R, Rat6, CV1, RPTE, A10, T24, J82, A375, ARH-77, Calu1, SW480, SW620, SKOV3, SK-UT, CaCo2, P388D1, SEM-K2, WEHI-231, HB56, TIB55, Jurkat, J45.01, LRMB, Bcl-1, BC-3, IC21, DLD2, Raw264.7, NRK, NRK-52E, MRC5, MEF, Hep G2,

HeLa B, HeLa T4, COS, COS-1, COS-6, COS-M6A, BS-C-1 monkey kidney epithelial, BALB/3T3 mouse embryo fibroblast, 3T3 Swiss, 3T3-L1, 132-d5 human fetal fibroblasts; 10.1 mouse fibroblasts, 293-T, 3T3, 721, 9L, A2780, A2780ADR, A2780cis, A172, A20, A253, A431, A-549, ALC, B16, B35, BCP-1 cells, BEAS-2B, bEnd.3, BHK-21, BR 293, BxPC3, C3H-10T1/2, C6/36, Cal-27, CHO, CHO-7, CHO-IR, CHO-K1, CHO-K2, CHO-T, CHO Dhfr ^{-/-}, COR-L23, COR-L23/CPR, COR-L23/5010, COR-L23/R23, COS-7, COV-434, CML T1, CMT, CT26, D17, DH82, DU145, DuCaP, EL4, EM2, EM3, EMT6/AR1, EMT6/AR10.0, FM3, H1299, H69, HB54, HB55, HCA2, HEK-293, HeLa, Hepa1c1c7, HL-60, HMEC, HT-29, Jurkat, JY cells, K562 cells, Ku812, KCL22, KG1, KYO1, LNCap, Ma-Mel 1-48, MC-38, MCF-7, MCF-10A, MDA-MB-231, MDA-MB-468, MDA-MB-435, MDCK II, MDCK II, MOR/0.2R, MONO-MAC 6, MTD-1A, MyEnd, NCI-H69/CPR, NCI-H69/LX10, NCI-H69/LX20, NCI-H69/LX4, NIH-3T3, NALM-1, NW-145, OPCN / OPCT cell lines, Peer, PNT-1A / PNT 2, RenCa, RIN-5F, RMA/RMAS, Saos-2 cells, Sf-9, SkBr3, T2, T-47D, T84, THP1 cell line, U373, U87, U937, VCaP, Vero cells, WM39, WT-49, X63, YAC-1, YAR, and transgenic varieties thereof. Cell lines are available from a variety of sources known to those with skill in the art (see, e.g., the American Type Culture Collection (ATCC) (Manassas, Va.)).

[0281] In an embodiment, the engineered cell is a muscle cell (e.g. cardiac muscle, skeletal muscle, and/or smooth muscle), bone cell, blood cell, immune cell (including but not limited to B cells, macrophages, T-cells, CAR-T cells, and the like), kidney cells, bladder cells, lung cells, heart cells, liver cells, brain cells, neurons, skin cells, stomach cells, neuronal support cells, intestinal cells, epithelial cells, endothelial cells, stem or other progenitor cells, adrenal gland cells, cartilage cells, and combinations thereof.

[0282] In an embodiment, the engineered cell can be a fungus cell. As used herein, a "fungal cell" refers to any type of eukaryotic cell within the kingdom of fungi. Phyla within the kingdom of fungi include Ascomycota, Basidiomycota, Blastocladiomycota, Chytridiomycota, Glomeromycota, Microsporidia, and Neocallimastigomycota. Fungal cells may include yeasts, molds, and filamentous fungi. In an embodiment, the fungal cell is a yeast cell.

[0283] As used herein, the term "yeast cell" refers to any fungal cell within the phyla Ascomycota and Basidiomycota. Yeast cells may include budding yeast cells, fission yeast cells, and mold cells. Without being limited to these organisms, many types of yeast used in laboratory and industrial settings are part of the phylum Ascomycota. In an embodiment, the yeast cell is an

S. cerevisiae, *Kluyveromyces marxianus*, or *Issatchenkia orientalis* cell. Other yeast cells may include without limitation *Candida* spp. (e.g., *Candida albicans*), *Yarrowia* spp. (e.g., *Yarrowia lipolytica*), *Pichia* spp. (e.g., *Pichia pastoris*), *Kluyveromyces* spp. (e.g., *Kluyveromyces lactis* and *Kluyveromyces marxianus*), *Neurospora* spp. (e.g., *Neurospora crassa*), *Fusarium* spp. (e.g., *Fusarium oxysporum*), and *Issatchenkia* spp. (e.g., *Issatchenkia orientalis*, a.k.a. *Pichia kudriavzevii* and *Candida acidothermophilum*). In an embodiment, the fungal cell is a filamentous fungal cell. As used herein, the term "filamentous fungal cell" refers to any type of fungal cell that grows in filaments, i.e., hyphae or mycelia. Examples of filamentous fungal cells may include without limitation *Aspergillus* spp. (e.g., *Aspergillus niger*), *Trichoderma* spp. (e.g., *Trichoderma reesei*), *Rhizopus* spp. (e.g., *Rhizopus oryzae*), and *Mortierella* spp. (e.g., *Mortierella isabellina*).

[0284] In an embodiment, the fungal cell is an industrial strain. As used herein, "industrial strain" refers to any strain of fungal cell used in or isolated from an industrial process, e.g., production of a product on a commercial or industrial scale. Industrial strain may refer to a fungal species that is typically used in an industrial process, or it may refer to an isolate of a fungal species that may be also used for non-industrial purposes (e.g., laboratory research). Examples of industrial processes may include fermentation (e.g., in production of food or beverage products), distillation, biofuel production, production of a compound, and production of a polypeptide. Examples of industrial strains can include, without limitation, JAY270 and ATCC4124.

[0285] In an embodiment, the fungal cell is a polyploid cell. As used herein, a "polyploid" cell may refer to any cell whose genome is present in more than one copy. A polyploid cell may refer to a type of cell that is naturally found in a polyploid state, or it may refer to a cell that has been induced to exist in a polyploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). A polyploid cell may refer to a cell whose entire genome is polyploid, or it may refer to a cell that is polyploid in a particular genomic locus of interest.

[0286] In an embodiment, the fungal cell is a diploid cell. As used herein, a "diploid" cell may refer to any cell whose genome is present in two copies. A diploid cell may refer to a type of cell that is naturally found in a diploid state, or it may refer to a cell that has been induced to exist in a diploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the *S. cerevisiae* strain S228C may be maintained in a haploid or diploid state. A diploid cell may refer to a cell whose entire genome is

diploid, or it may refer to a cell that is diploid in a particular genomic locus of interest. In an embodiment, the fungal cell is a haploid cell. As used herein, a "haploid" cell may refer to any cell whose genome is present in one copy. A haploid cell may refer to a type of cell that is naturally found in a haploid state, or it may refer to a cell that has been induced to exist in a haploid state (e.g., through specific regulation, alteration, inactivation, activation, or modification of meiosis, cytokinesis, or DNA replication). For example, the *S. cerevisiae* strain S228C may be maintained in a haploid or diploid state. A haploid cell may refer to a cell whose entire genome is haploid, or it may refer to a cell that is haploid in a particular genomic locus of interest.

[0287] In an embodiment, the engineered cell is a cell obtained from a subject. In an embodiment, the subject is a healthy or non-diseased subject. In an embodiment, the subject is a subject with a desired physiological and/or biological characteristic such that when a engineered AAV capsid particle is produced it can package one or more cargo polynucleotides that can be related to the desired physiological and/or biological characteristic and/or capable of modifying the desired physiological and/or biological characteristic. Thus, the cargo polynucleotides of the produced engineered AAV capsid particle can be capable of transferring the desired characteristic to a recipient cell. In an embodiment, the cargo polynucleotides are capable of modifying a polynucleotide of the engineered cell such that the engineered cell has a desired physiological and/or biological characteristic.

[0288] In an embodiment, a cell transfected with one or more vectors described herein is used to establish a new cell line comprising one or more vector-derived sequences.

[0289] The engineered cells can be used to produce engineered viral (e.g., AAV) capsid polynucleotides, vectors, and/or particles. In an embodiment, the engineered viral (e.g., AAV) capsid polynucleotides, vectors, and/or particles are produced, harvested, and/or delivered to a subject in need thereof. In an embodiment, the engineered cells are delivered to a subject. Other uses for the engineered cells are described elsewhere herein. In an embodiment, the engineered cells can be included in formulations and/or kits described elsewhere herein.

[0290] The engineered cells can be stored short-term or long-term for use at a later time. Suitable storage methods are generally known in the art. Further, methods of restoring the stored cells for use (such as thawing, reconstitution, and otherwise stimulating metabolism in the engineered cell after storage) at a later time are also generally known in the art.

Formulations

[0291] The compositions, polynucleotides, polypeptides, particles, cells, vector systems and combinations thereof described herein can be contained in a formulation, such as a pharmaceutical formulation. In an embodiment, the formulations can be used to generate polypeptides and other particles that include one or more CNS-specific transferrin receptor binding modifications described herein. In an embodiment, the formulations can be delivered to a subject in need thereof. In an embodiment, component(s) of the engineered AAV capsid system, engineered cells, engineered AAV capsid particles, and/or combinations thereof described herein can be included in a formulation that can be delivered to a subject or a cell. In an embodiment, the formulation is a pharmaceutical formulation. One or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein can be provided to a subject in need thereof or a cell alone or as an active ingredient, such as in a pharmaceutical formulation. As such, also described herein are pharmaceutical formulations containing an amount of one or more of the polypeptides, polynucleotides, vectors, cells, or combinations thereof described herein. In an embodiment, the pharmaceutical formulation can contain an effective amount of the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein. The pharmaceutical formulations described herein can be administered to a subject in need thereof or a cell.

[0292] In an embodiment, the amount of the one or more of the polypeptides, polynucleotides, vectors, cells, virus particles, nanoparticles, other delivery particles, and combinations thereof described herein contained in the pharmaceutical formulation can range from about 1 pg/kg to about 10 mg/kg based upon the bodyweight of the subject in need thereof or average bodyweight of the specific patient population to which the pharmaceutical formulation can be administered. The amount of the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein in the pharmaceutical formulation can range from about 1 pg to about 10 g, from about 10 nL to about 10 mL. In embodiments where the pharmaceutical formulation contains one or more cells, the amount can range from about 1 cell to 1×10^2 , 1×10^3 , 1×10^4 , 1×10^5 , 1×10^6 , 1×10^7 , 1×10^8 , 1×10^9 , 1×10^{10} or more cells. In embodiments where the pharmaceutical formulation contains one or more cells, the amount can range from about 1 cell to 1×10^2 , 1×10^3 , 1×10^4 , 1×10^5 , 1×10^6 , 1×10^7 , 1×10^8 , 1×10^9 , 1×10^{10} or more cells per nL, μ L, mL, or L.

[0293] In an embodiment, were engineered AAV capsid particles are included in the formulation, the formulation can contain 1 to 1×10^1 , 1×10^2 , 1×10^3 , 1×10^4 , 1×10^5 , 1×10^6 , 1×10^7 , 1×10^8 , 1×10^9 , 1×10^{10} , 1×10^{11} , 1×10^{12} , 1×10^{13} , 1×10^{14} , 1×10^{15} , 1×10^{16} , 1×10^{17} , 1×10^{18} , 1×10^{19} , or 1×10^{20} transducing units (TU)/mL of the engineered AAV capsid particles. In an embodiment, the formulation can be 0.1 to 100 mL in volume and can contain 1 to 1×10^1 , 1×10^2 , 1×10^3 , 1×10^4 , 1×10^5 , 1×10^6 , 1×10^7 , 1×10^8 , 1×10^9 , 1×10^{10} , 1×10^{11} , 1×10^{12} , 1×10^{13} , 1×10^{14} , 1×10^{15} , 1×10^{16} , 1×10^{17} , 1×10^{18} , 1×10^{19} , or 1×10^{20} transducing units (TU)/mL of the engineered AAV capsid particles.

Pharmaceutically Acceptable Carriers and Auxiliary Ingredients and Agents

[0294] In an embodiment, the pharmaceutical formulation containing an amount of one or more of the polypeptides, polynucleotides, vectors, cells, virus particles, nanoparticles, other delivery particles, and combinations thereof described herein can further include a pharmaceutically acceptable carrier. Suitable pharmaceutically acceptable carriers include, but are not limited to, water, salt solutions, alcohols, gum arabic, vegetable oils, benzyl alcohols, polyethylene glycols, gelatin, carbohydrates such as lactose, amylose or starch, magnesium stearate, talc, silicic acid, viscous paraffin, perfume oil, fatty acid esters, hydroxy methylcellulose, and polyvinyl pyrrolidone, which do not deleteriously react with the active composition.

[0295] The pharmaceutical formulations can be sterilized, and if desired, mixed with auxiliary agents, such as lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, coloring, flavoring and/or aromatic substances, and the like which do not deleteriously react with the active composition.

[0296] In addition to an amount of one or more of the polypeptides, polynucleotides, vectors, cells, engineered AAV capsid particles, nanoparticles, other delivery particles, and combinations thereof described herein, the pharmaceutical formulation can also include an effective amount of an auxiliary active agent, including but not limited to, polynucleotides, amino acids, peptides, polypeptides, antibodies, aptamers, ribozymes, hormones, immunomodulators, antipyretics, anxiolytics, antipsychotics, analgesics, antispasmodics, anti-inflammatories, anti-histamines, anti-infectives, chemotherapeutics, and combinations thereof.

[0297] Suitable hormones include, but are not limited to, amino-acid derived hormones (e.g., melatonin and thyroxine), small peptide hormones and protein hormones (e.g., thyrotropin-releasing hormone, vasopressin, insulin, growth hormone, luteinizing hormone, follicle-

stimulating hormone, and thyroid-stimulating hormone), eicosanoids (e.g., arachidonic acid, lipoxins, and prostaglandins), and steroid hormones (e.g., estradiol, testosterone, tetrahydrocortisol). Suitable immunomodulators include, but are not limited to, prednisone, azathioprine, 6-MP, cyclosporine, tacrolimus, methotrexate, interleukins (e.g., IL-2, IL-7, and IL-12), cytokines (e.g., interferons (e.g., IFN- α , IFN- β , IFN- ϵ , IFN-K, IFN- ω , and IFN- γ), granulocyte colony-stimulating factor, and imiquimod), chemokines (e.g., CCL3, CCL26 and CXCL7), cytosine phosphate-guanosine, oligodeoxynucleotides, glucans, antibodies, and aptamers).

[0298] Suitable antipyretics include, but are not limited to, non-steroidal anti-inflammatories (e.g., ibuprofen, naproxen, ketoprofen, and nimesulide), aspirin and related salicylates (e.g., choline salicylate, magnesium salicylate, and sodium salicylate), paracetamol/acetaminophen, metamizole, nabumetone, phenazone, and quinine.

[0299] Suitable anxiolytics include, but are not limited to, benzodiazepines (e.g., alprazolam, bromazepam, chlordiazepoxide, clonazepam, clorazepate, diazepam, flurazepam, lorazepam, oxazepam, temazepam, triazolam, and tofisopam), serotonergic antidepressants (e.g., selective serotonin reuptake inhibitors, tricyclic antidepressants, and monoamine oxidase inhibitors), mebicar, fabomotizole, selank, bromantane, emoxypine, azapirones, barbiturates, hydroxyzine, pregabalin, validol, and beta blockers.

[0300] Suitable antipsychotics include, but are not limited to, benperidol, bromoperidol, droperidol, haloperidol, moperone, pipamperone, timiperone, fluspirilene, penfluridol, pimozide, acepromazine, chlorpromazine, cyamemazine, dixyrazine, fluphenazine, levomepromazine, mesoridazine, perazine, pericyazine, perphenazine, pipotiazine, prochlorperazine, promazine, promethazine, prothipendyl, thioproperazine, thioridazine, trifluoperazine, triflupromazine, chlorprothixene, clopenthixol, flupentixol, thiothixene, zuclopenthixol, clotiapine, loxapine, prothipendyl, caripramine, clocapramine, molindone, mosapramine, sulpiride, veralipride, amisulpride, amoxapine, aripiprazole, asenapine, clozapine, blonanserin, iloperidone, lurasidone, melperone, nemonapride, olanzapine, paliperidone, perospirone, quetiapine, remoxipride, risperidone, sertindole, trimipramine, ziprasidone, zotepine, alstonie, bifeprunox, bitopertin, brexpiprazole, cannabidiol, cariprazine, pimavanserin, pomaglumetad methionil, vabicaserin, xanomeline, and zicronapine.

[0301] Suitable analgesics include, but are not limited to, paracetamol/acetaminophen, nonsteroidal anti-inflammatories (e.g., ibuprofen, naproxen, ketoprofen, and nimesulide), COX-2 inhibitors (e.g., rofecoxib, celecoxib, and etoricoxib), opioids (e.g., morphine, codeine, oxycodone, hydrocodone, dihydromorphine, pethidine, buprenorphine), tramadol, norepinephrine, flupirtine, nefopam, orphenadrine, pregabalin, gabapentin, cyclobenzaprine, scopolamine, methadone, ketobemidone, piritramide, and aspirin and related salicylates (e.g., choline salicylate, magnesium salicylate, and sodium salicylate).

[0302] Suitable antispasmodics include, but are not limited to, mebeverine, papaverine, cyclobenzaprine, carisoprodol, orphenadrine, tizanidine, metaxalone, methocarbamol, chlorzoxazone, baclofen, dantrolene, baclofen, tizanidine, and dantrolene. Suitable anti-inflammatories include, but are not limited to, prednisone, non-steroidal anti-inflammatories (e.g., ibuprofen, naproxen, ketoprofen, and nimesulide), COX-2 inhibitors (e.g., rofecoxib, celecoxib, and etoricoxib), and immune selective anti-inflammatory derivatives (e.g., submandibular gland peptide-T and its derivatives)

[0303] Suitable anti-histamines include, but are not limited to, H1-receptor antagonists (e.g., acrivastine, azelastine, bilastine, brompheniramine, buclizine, bromodiphenhydramine, carbinoxamine, cetirizine, chlorpromazine, cyclizine, chlorpheniramine, clemastine, cyproheptadine, desloratadine, dexbrompheniramine, dexchlorpheniramine, dimenhydrinate, dimetindene, diphenhydramine, doxylamine, ebastine, emramine, fexofenadine, hydroxyzine, levocetirizine, loratadine, meclizine, mirtazapine, olopatadine, orphenadrine, phenindamine, pheniramine, phenyltoloxamine, promethazine, pyrilamine, quetiapine, rupatadine, tripeleminamine, and triprolidine), H2-receptor antagonists (e.g., cimetidine, famotidine, lafutidine, nizatidine, ranitidine, and roxatidine), tritoqualine, catechin, cromoglicate, nedocromil, and p2-adrenergic agonists.

[0304] Suitable anti-infectives include, but are not limited to, amebicides (e.g., nitazoxanide, paromomycin, metronidazole, tinidazole, chloroquine, miltefosine, amphotericin b, and iodoquinol), aminoglycosides (e.g., paromomycin, tobramycin, gentamicin, amikacin, kanamycin, and neomycin), anthelmintics (e.g., pyrantel, mebendazole, ivermectin, praziquantel, albendazole, thiabendazole, oxfamiquine), antifungals (e.g., azole antifungals (e.g., itraconazole, fluconazole, parconazole, ketoconazole, clotrimazole, miconazole, and voriconazole), echinocandins (e.g., caspofungin, anidulafungin, and micafungin), griseofulvin, terbinafine, flucytosine, and polyenes

(e.g., nystatin, and amphotericin b), antimalarial agents (e.g., pyrimethamine/sulfadoxine, artemether/lumefantrine, atovaquone/proguanil, quinine, hydroxychloroquine, mefloquine, chloroquine, doxycycline, pyrimethamine, and halofantrine), antituberculosis agents (e.g., aminosalicylates (e.g., aminosalicylic acid), isoniazid/rifampin, isoniazid/pyrazinamide/rifampin, bedaquiline, isoniazid, ethambutol, rifampin, rifabutin, rifapentine, capreomycin, and cycloserine), antivirals (e.g., amantadine, rimantadine, abacavir/lamivudine, emtricitabine/tenofovir, cobicistat/elvitegravir/emtricitabine/tenofovir, efavirenz/emtricitabine/tenofovir, abacavir/lamivudine/zidovudine, lamivudine/zidovudine, emtricitabine/tenofovir, emtricitabine/lopinavir/ritonavir/tenofovir, interferon alfa-2v/ribavirin, peginterferon alfa-2b, maraviroc, raltegravir, dolutegravir, enfuvirtide, foscarnet, fomivirsen, oseltamivir, zanamivir, nevirapine, efavirenz, etravirine, rilpivirine, delavirdine, nevirapine, entecavir, lamivudine, adefovir, sofosbuvir, didanosine, tenofovir, abacavir, zidovudine, stavudine, emtricitabine, zalcitabine, telbivudine, simeprevir, boceprevir, telaprevir, lopinavir/ritonavir, boceprevir, darunavir, ritonavir, tipranavir, atazanavir, nelfinavir, amprenavir, indinavir, saquinavir, ribavirin, valacyclovir, acyclovir, famciclovir, ganciclovir, and valganciclovir), carbapenems (e.g., doripenem, meropenem, ertapenem, and cilastatin/imipenem), cephalosporins (e.g., cefadroxil, cephadrine, cefazolin, cephalixin, cefepime, cefazoline, loracarbef, cefotetan, cefuroxime, cefprozil, loracarbef, cefoxitin, cefaclor, ceftibuten, ceftriaxone, cefotaxime, cefpodoxime, cefdinir, cefixime, ceftidoren, ceftizoxime, and ceftazidime), glycopeptide antibiotics (e.g., vancomycin, dalbavancin, oritavancin, and telavancin), glycyclcyclines (e.g., tigecycline), leprostatics (e.g., clofazimine and thalidomide), lincomycin and derivatives thereof (e.g., clindamycin and lincomycin), macrolides and derivatives thereof (e.g., telithromycin, fidaxomicin, erythromycin, azithromycin, clarithromycin, dirithromycin, and troleandomycin), linezolid, sulfamethoxazole/trimethoprim, rifaximin, chloramphenicol, Fosfomycin, metronidazole, aztreonam, bacitracin, penicillin (amoxicillin, ampicillin, bacampicillin, carbenicillin, piperacillin, ticarcillin, amoxicillin/clavulanate, ampicillin/sulbactam, piperacillin/tazobactam, clavulanate/ticarcillin, penicillin, procaine penicillin, oxacillin, dicloxacillin, and nafcillin), quinolones (e.g., lomefloxacin, norfloxacin, ofloxacin, gatifloxacin, moxifloxacin, ciprofloxacin, levofloxacin, gemifloxacin, moxifloxacin, cinoxacin, nalidixic acid, enoxacin, grepafloxacin, gatifloxacin, trovafloxacin, and sparfloxacin), sulfonamides (e.g., sulfamethoxazole/trimethoprim, sulfasalazine, and sulfisoxazole), tetracyclines (e.g., doxycycline,

demeclocycline, minocycline, doxycycline/salicylic acid, doxycycline/omega-3 polyunsaturated fatty acids, and tetracycline), and urinary anti-infectives (e.g., nitrofurantoin, methenamine, Fosfomycin, cinoxacin, nalidixic acid, trimethoprim, and methylene blue).

[0305] Suitable chemotherapeutics include, but are not limited to, paclitaxel, brentuximab vedotin, doxorubicin, 5-FU (fluorouracil), everolimus, pemetrexed, melphalan, pamidronate, anastrozole, exemestane, nelarabine, ofatumumab, bevacizumab, belinostat, tositumomab, carmustine, bleomycin, bosutinib, busulfan, alemtuzumab, irinotecan, vandetanib, bicalutamide, lomustine, daunorubicin, clofarabine, cabozantinib, dactinomycin, ramucirumab, cytarabine, Cytosan, cyclophosphamide, decitabine, dexamethasone, docetaxel, hydroxyurea, dacarbazine, leuprolide, epirubicin, oxaliplatin, asparaginase, estramustine, cetuximab, vismodegib, asparaginase *Erwinia chrysanthemi*, amifostine, etoposide, flutamide, toremifene, fulvestrant, letrozole, degarelix, pralatrexate, methotrexate, floxuridine, obinutuzumab, gemcitabine, afatinib, imatinib mesylate, carmustine, eribulin, trastuzumab, altretamine, topotecan, ponatinib, idarubicin, ifosfamide, ibrutinib, axitinib, interferon alfa-2a, gefitinib, romidepsin, ixabepilone, ruxolitinib, cabazitaxel, ado-trastuzumab emtansine, carfilzomib, chlorambucil, sargramostim, cladribine, mitotane, vincristine, procarbazine, megestrol, trametinib, mesna, strontium-89 chloride, mechlorethamine, mitomycin, busulfan, gemtuzumab ozogamicin, vinorelbine, filgrastim, pegfilgrastim, sorafenib, nilutamide, pentostatin, tamoxifen, mitoxantrone, pegaspargase, denileukin diftitox, alitretinoin, carboplatin, pertuzumab, cisplatin, pomalidomide, prednisone, aldesleukin, mercaptopurine, zoledronic acid, lenalidomide, rituximab, octreotide, dasatinib, regorafenib, histrelin, sunitinib, siltuximab, omacetaxine, thioguanine (tioguanine), dabrafenib, erlotinib, bexarotene, temozolomide, thiotepa, thalidomide, *Bacillus Calmette-Guerin* (BCG), temsirolimus, bendamustine hydrochloride, triptorelin, arsenic trioxide, lapatinib, valrubicin, panitumumab, vinblastine, bortezomib, tretinoin, azacitidine, pazopanib, teniposide, leucovorin, crizotinib, capecitabine, enzalutamide, ipilimumab, goserelin, vorinostat, idelalisib, ceritinib, abiraterone, epothilone, tafluposide, azathioprine, doxifluridine, vindesine, and all-trans retinoic acid.

[0306] In embodiments where there is an auxiliary active agent contained in the pharmaceutical formulation in addition to the one or more of the polypeptides, polynucleotides, CRISPR-Cas complexes, vectors, cells, virus particles, nanoparticles, other delivery particles, and combinations thereof described herein, amount, such as an effective amount, of the auxiliary active

agent will vary depending on the auxiliary active agent. In an embodiment, the amount of the auxiliary active agent ranges from 0.001 micrograms to about 1 milligram. In other embodiments, the amount of the auxiliary active agent ranges from about 0.01 IU to about 1000 IU. In further embodiments, the amount of the auxiliary active agent ranges from 0.001 mL to about 1 mL. In yet other embodiments, the amount of the auxiliary active agent ranges from about 1 % w/w to about 50% w/w of the total pharmaceutical formulation. In additional embodiments, the amount of the auxiliary active agent ranges from about 1 % v/v to about 50% v/v of the total pharmaceutical formulation. In still other embodiments, the amount of the auxiliary active agent ranges from about 1 % w/v to about 50% w/v of the total pharmaceutical formulation.

Dosage Forms

[0307] In an embodiment, the pharmaceutical formulations described herein may be in a dosage form. The dosage forms can be adapted for administration by any appropriate route. Appropriate routes include, but are not limited to, rectal, epidural, intracranial, intraocular, inhaled, intranasal, topical (including buccal, sublingual, or transdermal), vaginal, intraurethral, parenteral, intracranial, subcutaneous, intramuscular, intravenous, intraperitoneal, intradermal, intraosseous, intracardiac, intraarticular, intracavernous, intrathecal, intravitreal, intracerebral, gingival, subgingival, intracerebroventricular, and intradermal. Such formulations may be prepared by any method known in the art.

[0308] For delivery to the CNS, the AAV formulations described herein may be administered intra-parenchymally, intrathecally, intracerebroventricularly, intracisternally, intravenously, into the carotid artery, systemically or a combination of these. In some aspects, the AAV formulation is administered by intrathecally in equal portions to the cisterna magna and the lumbar spine.

[0309] In some aspects, an AAV pharmaceutical formulation can be administered as a single bolus injection of about 0.1 mL, 0.2 mL, 0.3 mL, 0.4 mL, 0.5 mL, 0.6 mL, 0.7 mL, 0.8 mL, 0.9 mL, 1 mL, 2 mL, 3 mL, 4 mL, or 5 mL. In other aspects, an AAV pharmaceutical formulation is delivered as an infusion at a rate of 0.001 mL/min to 1 mL/min, (e.g., 0.01 mL/min).

[0310] In certain aspects, more than one administration (e.g., two, three, four, five, six, seven, eight, nine, 10, etc., or more administrations) may be employed to achieve the desired level of gene expression over a period of various intervals, e.g., hourly, daily, weekly, monthly, yearly, etc. Dosing can be single dosage or cumulative (serial dosing), and can be readily determined by one skilled in the art. For instance, treatment of a disease or disorder may comprise a one-time

administration of an effective dose of a pharmaceutical composition virus vector disclosed herein. Alternatively, treatment of a disease or disorder may comprise multiple administrations of an effective dose of a virus vector carried out over a range of time periods, such as, e.g., once daily, twice daily, trice daily, once every few days, or once weekly. In certain aspects, rAAV particles may be administered to multiple locations, for example, 1, 2, 3, 4, or 5 locations simultaneously or staggered over time.

[0311] In an embodiment, the more than one administration may include immuno-suppression or immunomodulatory agents (e.g., steroids, anti-B cell antibodies, rapamycin or other mTOR inhibitors). In an embodiment, the immuno-suppression or immunomodulatory agent is an inhibitor, see e.g., WO2021067598A1, hereby incorporated by reference. In an embodiment, the immuno-suppression or immunomodulatory agent is an antibody, see e.g., EP3909602A1, hereby incorporated by reference. In an embodiment, the immuno-suppression or immunomodulatory agent is a protease or glycosidase, see e.g., WO2020016318A1, hereby incorporated by reference. In an embodiment, the immuno-suppression or immunomodulatory agent is a steroid, see e.g., WO2021163322, hereby incorporated by reference.

[0312] The timing of administration can vary from individual to individual, depending upon such factors as the severity of an individual's symptoms. For example, an effective dose of a virus vector disclosed herein can be administered to an individual once every six months for an indefinite period, or until the individual no longer requires therapy. A person of ordinary skill in the art will recognize that the condition of the individual can be monitored throughout the course of treatment and that the effective amount of a virus vector disclosed herein that is administered can be adjusted accordingly.

[0313] The pharmaceutical composition is typically in liquid form. Liquid pharmaceutical compositions generally include a liquid carrier such as water, petroleum, animal or vegetable oils, mineral oil or synthetic oil. Physiological saline solution, magnesium chloride, dextrose or other saccharide solution or glycols such as ethylene glycol, propylene glycol or polyethylene glycol may be included. In some cases, a surfactant, such as pluronic acid (PF68) 0.001% may be used.

[0314] For injection, the active ingredient will be in the form of an aqueous solution which is pyrogen-free and has suitable pH, isotonicity and stability. Those of relevant skill in the art are well able to prepare suitable solutions using, for example, isotonic vehicles such as Sodium

Chloride Injection, Ringer's Injection, Lactated Ringer's Injection. Preservatives, stabilizers, buffers, antioxidants and/or other additives may be included, as required.

[0315] Methods of delivering an AAV preparation to a specific region of the CNS by stereotactic injection can be found in U.S. Patent No. 10,898,585, the content of which is incorporated by reference herein in its entirety.

[0316] Dosage forms adapted for parenteral administration and/or adapted for any type of injection (e.g., intravenous, intraperitoneal, subcutaneous, intramuscular, intradermal, intraosseous, epidural, intracardiac, intraarticular, intracavernous, gingival, subgingival, intrathecal, intravitreal, intracerebral, and intracerebroventricular) can include aqueous and/or non-aqueous sterile injection solutions, which can contain anti-oxidants, buffers, bacteriostats, solutes that render the composition isotonic with the blood of the subject, and aqueous and non-aqueous sterile suspensions, which can include suspending agents and thickening agents. The dosage forms adapted for parenteral administration can be presented in a single-unit dose or multi-unit dose containers, including but not limited to sealed ampoules or vials. The doses can be lyophilized and resuspended in a sterile carrier to reconstitute the dose prior to administration. Extemporaneous injection solutions and suspensions can be prepared, in an embodiment, from sterile powders, granules, and tablets. See e.g., Glascock, J. J., *et al.* Delivery of Therapeutic Agents Through Intracerebroventricular (ICV) and Intravenous (IV) Injection in Mice. *J. Vis. Exp.* (56), e2968 and Foley CP, *et al.* Intra-arterial delivery of AAV vectors to the mouse brain after mannitol mediated blood brain barrier disruption. *J Control Release*. 2014 Dec 28;196:71-78.

[0317] The dosage form can also be prepared to prolong or sustain the release of any ingredient. In an embodiment, the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein can be the ingredient whose release is delayed. In other embodiments, the release of an optionally included auxiliary ingredient is delayed. Suitable methods for delaying the release of an ingredient include, but are not limited to, coating or embedding the ingredients in material in polymers, wax, gels, and the like. Delayed release dosage formulations can be prepared as described in standard references such as "Pharmaceutical dosage form tablets," eds. Liberman et. al. (New York, Marcel Dekker, Inc., 1989), "Remington - The science and practice of pharmacy", 20th ed., Lippincott Williams & Wilkins, Baltimore, MD, 2000, and "Pharmaceutical dosage forms and drug delivery systems", 6th Edition, Ansel et al., (Media, PA: Williams and Wilkins, 1995).

[0318] Dosage forms adapted for nasal or inhalation administration include aerosols, solutions, suspension drops, gels, or dry powders. In an embodiment, the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein is contained in a dosage form adapted for inhalation is in a particle-size-reduced form that is obtained or obtainable by micronization. In an embodiment, the particle size of the size reduced (e.g., micronized) compound or salt or solvate thereof, is defined by a D50 value of about 0.5 to about 10 microns as measured by an appropriate method known in the art. Dosage forms adapted for administration by inhalation also include particle dusts or mists. Suitable dosage forms wherein the carrier or excipient is a liquid for administration as a nasal spray or drops include aqueous or oil solutions/suspensions of an active ingredient (e.g., the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein and/or auxiliary active agent), which may be generated by various types of metered dose pressurized aerosols, nebulizers, or insufflators.

[0319] In an embodiment, the dosage forms can be aerosol formulations suitable for administration by inhalation. In some of these embodiments, the aerosol formulation can contain a solution or fine suspension of the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein and a pharmaceutically acceptable aqueous or non-aqueous solvent. Aerosol formulations can be presented in single or multi-dose quantities in sterile form in a sealed container. For some of these embodiments, the sealed container is a single dose or multi-dose nasal, or an aerosol dispenser fitted with a metering valve (e.g., metered dose inhaler), which is intended for disposal once the contents of the container have been exhausted.

[0320] Where the aerosol dosage form is contained in an aerosol dispenser, the dispenser contains a suitable propellant under pressure, such as compressed air, carbon dioxide, or an organic propellant, including but not limited to a hydrofluorocarbon. The aerosol formulation dosage forms in other embodiments are contained in a pump-atomizer. The pressurized aerosol formulation can also contain a solution or a suspension of one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein. In further embodiments, the aerosol formulation can also contain co-solvents and/or modifiers incorporated to improve, for example, the stability and/or taste and/or fine particle mass characteristics (amount and/or profile) of the formulation. Administration of the aerosol formulation can be once daily or several times daily, for example 2, 3, 4, or 8 times daily, in which 1, 2, or 3 doses are delivered each time.

[0321] For some dosage forms suitable and/or adapted for inhaled administration, the pharmaceutical formulation is a dry powder inhalable formulation. In addition to the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein, an auxiliary active ingredient, and/or pharmaceutically acceptable salt thereof, such a dosage form can contain a powder base such as lactose, glucose, trehalose, mannitol, and/or starch. In some of these embodiments, the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein is in a particle-size reduced form. In further embodiments, a performance modifier, such as L-leucine or another amino acid, cellobiose octaacetate, and/or metals salts of stearic acid, such as magnesium or calcium stearate.

[0322] In an embodiment, the aerosol dosage forms can be arranged so that each metered dose of aerosol contains a predetermined amount of an active ingredient, such as the one or more of the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein.

[0323] Dosage forms adapted for ocular administration can include aqueous and/or nonaqueous sterile solutions that can optionally be adapted for injection, and which can optionally contain antioxidants, buffers, bacteriostats, solutes that render the composition isotonic with the eye or fluid contained therein or around the eye of the subject, and aqueous and nonaqueous sterile suspensions, which can include suspending agents and thickening agents.

[0324] For some embodiments, the dosage form contains a predetermined amount of the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein per unit dose. In an embodiment, the predetermined amount of the Such unit doses may therefore be administered once or more than once a day. Such pharmaceutical formulations may be prepared by any of the methods well known in the art.

Effective Amounts

[0325] In an embodiment, the amount of the primary active agent and/or optional secondary agent can be an effective amount, least effective amount, and/or therapeutically effective amount. As used herein, “effective amount”, “effective concentration”, and/or the like refers to the amount, concentration, etc. of the primary and/or optional secondary agent included in the pharmaceutical formulation that achieve one or more therapeutic effects or desired effect. As used herein, “least effective”, “least effective concentration”, and/or the like amount refers to the lowest amount, concentration, etc. of the primary and/or optional secondary agent that achieves the one or more

therapeutic or other desired effects. As used herein, “therapeutically effective amount”, “therapeutically effective concentration” and/or the like refers to the amount, concentration, etc. of the primary and/or optional secondary agent included in the pharmaceutical formulation that achieves one or more therapeutic effects. In an embodiment, the one or more therapeutic effects comprise transducing the CNS.

[0326] In an embodiment, the amount or effective amount, particularly where an infective particle is being delivered (e.g., a virus particle having the primary or secondary agent as a cargo), the effective amount of virus particles can be expressed as a titer (plaque forming units per unit of volume) or as a MOI (multiplicity of infection). In an embodiment, the effective amount can be about 1×10^1 particles per pL, nL, μ L, mL, or L to 1×10^{20} particles per pL, nL, μ L, mL, or L or more, such as about 1×10^1 , 1×10^2 , 1×10^3 , 1×10^4 , 1×10^5 , 1×10^6 , 1×10^7 , 1×10^8 , 1×10^9 , 1×10^{10} , 1×10^{11} , 1×10^{12} , 1×10^{13} , 1×10^{14} , 1×10^{15} , 1×10^{16} , 1×10^{17} , 1×10^{18} , 1×10^{19} , to/or about 1×10^{20} particles per pL, nL, μ L, mL, or L. In an embodiment, the effective titer can be about 1×10^1 transforming units per pL, nL, μ L, mL, or L to 1×10^{20} transforming units per pL, nL, μ L, mL, or L or more, such as about 1×10^1 , 1×10^2 , 1×10^3 , 1×10^4 , 1×10^5 , 1×10^6 , 1×10^7 , 1×10^8 , 1×10^9 , 1×10^{10} , 1×10^{11} , 1×10^{12} , 1×10^{13} , 1×10^{14} , 1×10^{15} , 1×10^{16} , 1×10^{17} , 1×10^{18} , 1×10^{19} , to/or about 1×10^{20} transforming units per pL, nL, μ L, mL, or L or any numerical value or subrange within these ranges. In an embodiment, the MOI of the pharmaceutical formulation can range from about 0.1 to 10 or more, such as 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, 1.9, 2, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 3, 3.1, 3.2, 3.3, 3.4, 3.5, 3.6, 3.7, 3.8, 3.9, 4, 4.1, 4.2, 4.3, 4.4, 4.5, 4.6, 4.7, 4.8, 4.9, 5, 5.1, 5.2, 5.3, 5.4, 5.5, 5.6, 5.7, 5.8, 5.9, 6, 6.1, 6.2, 6.3, 6.4, 6.5, 6.6, 6.7, 6.8, 6.9, 7, 7.1, 7.2, 7.3, 7.4, 7.5, 7.6, 7.7, 7.8, 7.9, 8, 8.1, 8.2, 8.3, 8.4, 8.5, 8.6, 8.7, 8.8, 8.9, 9, 9.1, 9.2, 9.3, 9.4, 9.5, 9.6, 9.7, 9.8, 9.9, 10 or more or any numerical value or subrange within these ranges.

[0327] In an embodiment, the amount or effective amount, particularly where an infective particle is being delivered (e.g., a virus particle having the primary or secondary agent as a cargo), the effective amount of virus particles can be expressed as vector genomes per kilogram bodyweight (vg/kg). In an embodiment, the effective amount can be about 1×10^1 vg/kg or more, such as about 1×10^2 , 1×10^3 , 1×10^4 , 1×10^5 , 1×10^6 , 1×10^7 , 1×10^8 , 1×10^9 , 1×10^{10} , 1×10^{11} , 1×10^{12} , 1×10^{13} , 1×10^{14} , 1×10^{15} , 1×10^{16} , 1×10^{17} , 1×10^{18} , 1×10^{19} , to/or about 1×10^{20} such that the effective amount can be anything in between, for example, 0.1 to 10 (0.1×10^{12} to 10×10^{12}). In an embodiment, the effective amount can be about 1×10^{11} , 5×10^{11} , 1×10^{12} , 5×10^{12} , 1×10^{13} , 5×10^{13} ,

1×10^{14} , 5×10^{14} such that the effective amount can be anything in between. In an embodiment, the effective amount can be about 1×10^{12} , 2×10^{12} , 3×10^{12} , 4×10^{12} , 5×10^{12} , 6×10^{12} , 7×10^{12} , 8×10^{12} , 9×10^{12} , 10×10^{12} such that the effective amount can be anything in between. In an embodiment, the pharmaceutical composition is delivered at a dosage between 0.1×10^{12} vg/kg to 1×10^{14} vg/kg. In an embodiment, the dosage is between 0.1×10^{12} vg/kg to 100×10^{12} vg/kg. In an embodiment, the dosage is between 1×10^{12} vg/kg to 10×10^{12} vg/kg. In an embodiment, the dosage is 5×10^{12} vg/kg. In an embodiment the dosage is 2×10^{12} vg/kg.

[0328] In an embodiment, the amount of virus particles can be measured as vector genomes per cell (vg/cell). This unit of measurement describes the concentration of vector genomes in a cell. Vg/cell may be used to measure the effective amount of vector genomes in a cell after delivery of a particular dosage, such as those described above. For example, the vg/cell can be computed by taking the absolute DNA value derived from a standard curve and multiplying by the number of single-stranded plasmid copies in 1 femtogram of plasmid, e.g., $(2 \times ([6.022 \times 10^{23}] / [\text{plasmid size in bp} \times 1 \times 10^{15} \times 650]))$, then compute the diploid genome by taking the absolute DNA value derived from the standard curve and dividing by the number of human diploid genomes per nanogram of DNA (167 genomes/ng), finally divide the total number of vector genomes by the number of diploid genomes. See e.g., Bio Protoc. 2017 Jan 20;7(2):e2110. doi: 10.21769/BioProtoc.2110.

[0329] In an embodiment, the effective amount can be about 0.1 vg/cell or more such as about 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, to/or about 20 such that the effective amount can be anything in between, for example, 2.1-2.9 (e.g., 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9). In an embodiment, the dosage is 2 vg/cell or greater.

[0330] In an embodiment, the effective amount, least effective amount, and/or therapeutically effective amount can be an effective concentration, least effective concentration, and/or therapeutically effective concentration, which can each be any non-zero amount ranging from about 0 to 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 210, 220, 230, 240, 250, 260, 270, 280, 290, 300, 310, 320, 330, 340, 350, 360, 370, 380, 390, 400, 410, 420, 430, 440, 450, 460, 470, 480, 490, 500, 510, 520, 530, 540, 550, 560, 570, 580, 590, 600, 610, 620, 630, 640, 650, 660, 670, 680, 690, 700, 710, 720, 730, 740, 750, 760, 770, 780, 790, 800, 810, 820, 830, 840, 850, 860, 870, 880, 890, 900, 910, 920, 930, 940, 950, 960,

970, 980, 990, 1000 pM, nM, μ M, mM, or M or be any numerical value or subrange within any of these ranges.

[0331] In an embodiment, the primary and/or the optional secondary active agent present in the pharmaceutical formulation can be any non-zero amount ranging from about 0 to 0.001, 0.002, 0.003, 0.004, 0.005, 0.006, 0.007, 0.008, 0.009, 0.01, 0.02, 0.03, 0.04, 0.05, 0.06, 0.07, 0.08, 0.09, 0.1, 0.11, 0.12, 0.13, 0.14, 0.15, 0.16, 0.17, 0.18, 0.19, 0.2, 0.21, 0.22, 0.23, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.3, 0.31, 0.32, 0.33, 0.34, 0.35, 0.36, 0.37, 0.38, 0.39, 0.4, 0.41, 0.42, 0.43, 0.44, 0.45, 0.46, 0.47, 0.48, 0.49, 0.5, 0.51, 0.52, 0.53, 0.54, 0.55, 0.56, 0.57, 0.58, 0.59, 0.6, 0.61, 0.62, 0.63, 0.64, 0.65, 0.66, 0.67, 0.68, 0.69, 0.7, 0.71, 0.72, 0.73, 0.74, 0.75, 0.76, 0.77, 0.78, 0.79, 0.8, 0.81, 0.82, 0.83, 0.84, 0.85, 0.86, 0.87, 0.88, 0.89, 0.9, 0.91, 0.92, 0.93, 0.94, 0.95, 0.96, 0.97, 0.98, 0.9, to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9 % w/w, v/v, or w/v of the pharmaceutical formulation or be any numerical value or subrange within any of these ranges.

[0332] In an embodiment, the amount or effective amount of the one or more of the active agent(s) described herein contained in the pharmaceutical formulation can range from about 1 pg/kg to about 10 mg/kg based upon the bodyweight of the subject in need thereof or average body weight of the specific patient population to which the pharmaceutical formulation can be administered.

[0333] In an embodiment where there is a secondary agent contained in the pharmaceutical formulation, the effective amount of the secondary active agent will vary depending on the secondary agent, the primary agent, the administration route, subject age, disease, stage of disease, among other things, which will be one of ordinary skill in the art.

[0334] When optionally present in the pharmaceutical formulation, the secondary active agent can be included in the pharmaceutical formulation or can exist as a stand-alone compound or pharmaceutical formulation that can be administered contemporaneously or sequentially with the compound, derivative thereof, or pharmaceutical formulation thereof.

[0335] In an embodiment, the effective amount of the secondary active agent, when optionally present, is any non-zero amount ranging from about 0 to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14,

15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9 % w/w, v/v, or w/v of the total active agents present in the pharmaceutical formulation or any numerical value or subrange within these ranges. In additional embodiments, the effective amount of the secondary active agent is any non-zero amount ranging from about 0 to 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 99.1, 99.2, 99.3, 99.4, 99.5, 99.6, 99.7, 99.8, 99.9 % w/w, v/v, or w/v of the total pharmaceutical formulation or any numerical value or subrange within these ranges.

Manufacturing

[0336] In an embodiment, the techniques described herein relate a method of manufacturing a recombinant engineered AAV. In an embodiment, the method of manufacturing comprising: culturing mammalian cells comprising; (1) a polynucleotide encoding the engineered AAV capsid polypeptide of any one of claims 1 to 26, (2) a polynucleotide encoding a recombinant AAV genome comprising a transgene operably linked to a regulatory sequence and flanked by AAV ITR sequences, and optionally (3) polynucleotide encoding adenoviral helper genes, under conditions sufficient for production of recombinant engineered AAV particles; and recovering the recombinant engineered AAV particles from said culture. In an embodiment, recovering includes any known standard techniques in the art. In an embodiment, the known standard techniques include those used for recombinant proteins. In an embodiment, recovering may include precipitation, centrifugation, chromatography, and/or tangential flow filtration.

Kits

[0337] Also described herein are kits that contain one or more of the one or more of the compositions, polypeptides, polynucleotides, vectors, cells, or other components described herein and combinations thereof and pharmaceutical formulations described herein. In an embodiment, one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof described herein can be presented as a combination kit. As used herein, the terms "combination

kit" or "kit of parts" refers to the compounds, or formulations and additional components that are used to package, screen, test, sell, market, deliver, and/or administer the combination of elements or a single element, such as the active ingredient, contained therein. Such additional components include but are not limited to, packaging, syringes, blister packages, bottles, and the like. The combination kit can contain one or more of the components (e.g., one or more of the one or more of the polypeptides, polynucleotides, vectors, cells, and combinations thereof) or formulation thereof can be provided in a single formulation (e.g., a liquid, lyophilized powder, etc.), or in separate formulations. The separate components or formulations can be contained in a single package or in separate packages within the kit. The kit can also include instructions in a tangible medium of expression that can contain information and/or directions regarding the content of the components and/or formulations contained therein, safety information regarding the content of the components(s) and/or formulation(s) contained therein, information regarding the amounts, dosages, indications for use, screening methods, component design recommendations and/or information, recommended treatment regimen(s) for the components(s) and/or formulations contained therein. As used herein, "tangible medium of expression" refers to a medium that is physically tangible or accessible and is not a mere abstract thought or an unrecorded spoken word. "Tangible medium of expression" includes, but is not limited to, words on a cellulosic or plastic material, or data stored in a suitable computer readable memory form. The data can be stored on a unit device, such as a flash memory drive or CD-ROM or on a server that can be accessed by a user via, e.g., a web interface.

[0338] In one embodiment, the disclosure provides a kit comprising one or more of the components described herein. In an embodiment, the kit comprises a vector system and instructions for using the kit. In an embodiment, the vector system includes a regulatory element operably linked to one or more engineered polynucleotides, such as those containing a selective transferrin receptor binding modification, as described elsewhere herein and, optionally, a cargo molecule, which can optionally be operably linked to a regulatory element. The one or more engineered polynucleotides such as those containing a selective transferrin receptor binding modification, as described elsewhere herein and, can be included on the same or different vectors as the cargo molecule in embodiments containing a cargo molecule within the kit.

[0339] In an embodiment, the kit comprises a vector system and instructions for using the kit. In an embodiment, the vector system comprises (a) a first regulatory element operably linked to a

direct repeat sequence and one or more insertion sites for inserting one or more guide sequences up- or downstream (whichever applicable) of the direct repeat sequence, wherein when expressed, the guide sequence directs sequence-specific binding of a Cas9 CRISPR complex to a target sequence in a eukaryotic cell, wherein the Cas9 CRISPR complex comprises a Cas9 enzyme complexed with the guide sequence that is hybridized to the target sequence; and/or (b) a second regulatory element operably linked to an enzyme-coding sequence encoding said Cas9 enzyme comprising a nuclear localization sequence. Where applicable, a tracr sequence may also be provided. In an embodiment, the kit comprises components (a) and (b) located on the same or different vectors of the system. In an embodiment, component (a) further comprises two or more guide sequences operably linked to the first regulatory element, wherein when expressed, each of the two or more guide sequences direct sequence specific binding of a CRISPR complex to a different target sequence in a eukaryotic cell. In an embodiment, the Cas9 enzyme comprises one or more nuclear localization sequences of sufficient strength to drive accumulation of said CRISPR enzyme in a detectable amount in the nucleus of a eukaryotic cell. In an embodiment, the CRISPR enzyme is a type V or VI CRISPR system enzyme. In an embodiment, the CRISPR enzyme is a Cas9 enzyme. In an embodiment, the Cas9 enzyme is derived from *Francisella tularensis* 1, *Francisella tularensis* subsp. *novicida*, *Prevotella albensis*, *Lachnospiraceae* bacterium MC2017 1, *Butyrivibrio proteoclasticus*, *Peregrinibacteria* bacterium GW2011_GWA2_33_10, *Parcubacteria* bacterium GW2011_GWC2_44_17, *Smithella* sp. SCADC, *Acidaminococcus* sp. BV3L6, *Lachnospiraceae* bacterium MA2020, *Candidatus Methanoplasma termitum*, *Eubacterium eligens*, *Moraxella bovoculi* 237, *Leptospira inadai*, *Lachnospiraceae* bacterium ND2006, *Porphyromonas crevioricanis* 3, *Prevotella disiens*, or *Porphyromonas macacae* Cas9 (e.g., modified to have or be associated with at least one DD), and may include further alteration or mutation of the Cas9, and can be a chimeric Cas9. In an embodiment, the DD-CRISPR enzyme is codon-optimized for expression in a eukaryotic cell. In an embodiment, the DD-CRISPR enzyme directs cleavage of one or two strands at the location of the target sequence. In an embodiment, the DD-CRISPR enzyme lacks or substantially DNA strand cleavage activity (e.g., no more than 5% nuclease activity as compared with a wild-type enzyme or enzyme not having the mutation or alteration that decreases nuclease activity). In an embodiment, the first regulatory element is a polymerase III promoter. In an embodiment, the second regulatory element is a polymerase II promoter. In an

embodiment, the guide sequence is at least 16, 17, 18, 19, 20, 25 nucleotides, or between 16-30, or between 16-25, or between 16-20 nucleotides in length.

Methods of Use for Delivery of Cargo to CNS

[0340] The compositions including one or more of the cell-selective transferrin receptor binding modifications, engineered AAV capsid system polynucleotides, polypeptides, vector(s), engineered cells, engineered AAV capsid particles can be used generally to package and/or deliver one or more cargos to neurons, glial cells, and endothelial cells of the CNS. In an embodiment, delivery is done in cell-selective manner based upon the selectivity of the transferrin receptor binding modification. In an embodiment this is conferred by the tropism of the engineered AAV capsid, which can be influenced at least in part by the inclusion of one or transferrin receptor binding modifications described elsewhere herein. In an embodiment, compositions including one or more of the CNS transferrin receptor binding modifications, including engineered AAV capsid particles where the capsids incorporate the transferrin receptor binding modification, can be administered to a subject or a cell, tissue, and/or organ and facilitate the transfer and/or integration of the cargo to the recipient cell. In other embodiments, engineered cells capable of producing compositions, such as polypeptides and other particles (e.g., engineered AAV capsids and viral particles), containing one or more of the transferrin receptor binding modifications can be generated from the polynucleotides, vectors, and vector systems etc., described herein. This includes without limitation, the engineered AAV capsid system molecules (e.g., polynucleotides, vectors, and vector systems, etc.). In an embodiment, the polynucleotides, vectors, and vector systems etc., described herein capable of generating the compositions, such as polypeptides and other particles (e.g., engineered AAV capsids and viral particles), containing one or more of the transferrin receptor binding modifications can be delivered to a cell or tissue, in vivo, ex vivo, or in vitro. In an embodiment, when delivered to a subject, the composition can transform a subject's cell in vivo or ex vivo to produce an engineered cell that can be capable of making a composition described herein that contains one or more of the cell-selective transferrin receptor binding modifications described herein, including but not limited to the engineered AAV capsid particles, which can be released from the engineered cell and deliver cargo molecule(s) to a recipient cell in vivo or produce personalized engineered compositions (e.g., AAV capsid particles) for reintroduction into the subject from which the recipient cell was obtained.

[0341] In an embodiment, an engineered cell can be delivered to a subject, where it can release produced compositions of the present disclosure (including but not limited to engineered AAV capsid particles) such that they can then deliver a cargo (e.g., a cargo polynucleotide(s)) to a recipient cell. These general processes can be used in a variety of ways to treat and/or prevent disease or a symptom thereof in a subject, generate model cells, generate modified organisms, provide cell selection and screening assays, in bioproduction, and in other various applications.

[0342] In an embodiment, the compositions, such as polypeptides and other particles (e.g., engineered AAV capsids and viral particles), containing one or more of the transferrin receptor binding modifications can be delivered to neural cells of the CNS. In another example embodiment, the compositions containing one or more transferrin receptor binding modifications can be delivered to glial cells of the CNS.

[0343] In an embodiment, the engineered AAV capsid polynucleotides, vectors, and systems thereof can be used to generate engineered AAV capsid variant libraries that can be mined for variants with a desired cell-selectivity. The description provided herein as supported by the various Examples can demonstrate that one having a desired cell-selectivity in mind could utilize the present disclosure as described herein to obtain a capsid with the desired cell-selectivity.

Modified Tropism

[0344] The adeno-associated virus (AAV) has more than one hundred naturally occurring variants, and this library has been further expanded by various techniques for modifying the viral capsid. These AAV capsid variants have unique antigenic profiles and demonstrate distinct cellular tropisms driven by differences in receptor binding. Additionally, AAV capsids can be chemically modified to alter tropism, produced as hybrid vectors combining the properties of multiple serotypes, and carry peptide insertions that introduce novel receptor-binding activity. Furthermore, directed evolution of shuffled genome libraries can identify engineered variants with unique properties, and rational modification of the viral capsid can alter tropism, reduce blockage by neutralizing antibodies, or enhance transduction efficiency. (reviewed by Castle, M. J., Turunen, H. T., Vandenberghe, L. H. & Wolfe, J. H. Controlling AAV Tropism in the Nervous System with Natural and Engineered Capsids. *Methods Mol. Biol.* **1382**, 133–149 (2016)).

[0345] The transferrin receptor binding modifications of the capsid polypeptides disclosed herein enhance AAV transduction in the CNS.

[0346] In an embodiment, the modified AAV capsid facilitates AAV transduction to at least one region of the brain.

[0347] In an embodiment, the modified AAV capsid facilitates AAV transduction to the forebrain, the midbrain or the hindbrain.

[0348] In an embodiment, the modified AAV capsid facilitates AAV transduction to at least one of the following brain regions chosen from the occipital lobes, the temporal lobe, the parietal lobe, the frontal lobe, the cerebral cortex, the cerebellum, the hypothalamus, the thalamus, the pituitary gland, the pineal gland, the amygdala, the hippocampus and the mid-brain.

[0349] In an embodiment, the modified AAV capsid facilitates AAV transduction to at least two of the following brain regions chosen from the occipital lobes, the temporal lobe, the parietal lobe, the frontal lobe, the cerebral cortex, the cerebellum, the hypothalamus, the thalamus, the pituitary gland, the pineal gland, the amygdala, the hippocampus and the mid-brain.

[0350] In an embodiment, the modified AAV capsid facilitates AAV transduction to at least three of the following brain regions chosen from the occipital lobes, the temporal lobe, the parietal lobe, the frontal lobe, the cerebral cortex, the cerebellum, the hypothalamus, the thalamus, the pituitary gland, the pineal gland, the amygdala, the hippocampus and the mid-brain.

[0351] In an embodiment, the modified AAV capsid facilitates AAV transduction of the endothelial cells of the CNS vasculature.

Regulation of transgene expression by microRNAs

[0352] MicroRNAs (miRNAs) are indeed small, non-coding RNA molecules that play a crucial role in regulating gene expression at the post-transcriptional level. They bind to complementary sequences on messenger RNAs (mRNAs), typically resulting in gene silencing either by degrading the mRNA or by inhibiting its translation into proteins.

[0353] In an embodiment, AAV transgene expression in the various regions of the brain can be further fine-tuned by the incorporation of specific miRNA binding sites into the transgene's 5' or 3'-UTR. Thus, AAV transgene expression can be suppressed in those cells where the microRNA is expressed and the corresponding miRNA binding site is placed in the 5' or 3'-UTR. A summary of miRNA gene expression in specific cell types of the CNS can be found in the Table below, reproduced from Zolboot, N., Du, J. X., Zampa, F. & Lippi, G. MicroRNAs instruct and maintain cell type diversity in the nervous system. *Front. Mol. Neurosci.* 14, 646072 (2021).

[0354] In one embodiment, the transgene's 5' or 3'-UTR may contain one or more miR-122 binding sites to suppress spurious transgene expression and potential toxicity in the liver (see, for example, Qiao, C. et al. Liver-specific microRNA-122 target sequences incorporated in AAV vectors efficiently inhibits transgene expression in the liver. *Gene Ther.* 18, 403–410 (2011)). In one embodiment, the transgene's 5' or 3'-UTR may contain one or more miR-183 binding sites to suppress spurious transgene expression and potential toxicity in the liver (see, for example, Hordeaux, J.; Buza, E. L.; Jeffrey, B.; Song, C.; Jahan, T.; Yuan, Y.; Zhu, Y.; Bell, P.; Li, M.; Chichester, J. A.; Calcedo, R.; Wilson, J. M. MicroRNA-Mediated Inhibition of Transgene Expression Reduces Dorsal Root Ganglion Toxicity by AAV Vectors in Primates. *Science Translational Medicine*, 2020, 12).

Table F. Example miRNAs and known targets.

CELL type	Developmental processes	miRNA	Known targets	References
Neural progenitors	Proliferation	let-7, miR-125, miR-9, miR-137	lin-28	Fyfe et al., 2006
		let-7, miR-9, miR-137	Nr2e1	Zhao et al., 2009
	Proliferation, cell cycle	miR-2115	Ccr4	Nowotarski et al., 2016
		miR-302a-d		Pargament et al., 2015
	Proliferation, differentiation	miR-20a, miR-20b, miR-23a	CyclinD1	Chen et al., 2014
Intermediate progenitors		miR-934		Prodromidou et al., 2020
		miR-186a/b-5p		Lee et al., 2020
	Neuronal differentiation	miR-124, miR-9	REST, E2F complexes	Yoo et al., 2011; Lee et al., 2013
		miR-124	Ptbp1	Makay et al., 2007
		miR-92b	Tb2	Nowakowski et al., 2015
Retinal progenitors	Neurogenesis, differentiation	let-7, miR-125, miR-9	Prig, Lin28b	La Torre et al., 2013
Olfactory progenitor cells	Neurogenesis, survival	miR-200a-c, miR-429, miR-141	Mash1	Chen et al., 2008
Adult neurogenic progenitors	Proliferation, progenitor identity	miR-184, miR-34a	Numa	Shen et al., 2002
				Finocchietti et al., 2012
	Neuronal differentiation	miR-124	Sox9	Cheng et al., 2009
Cortical pyramidal neurons	Laminar identity	miR-125, let-7b, miR-9		Shu et al., 2019
	Migration	mi-129-3p/5p	Fmr1	Wu et al., 2018
		miR-386-3p, miR-498, miR-543	Cdh2	Pagès et al., 2014
	Dendritic outgrowth	miR-101	Scn12a2, Ank2, Klf1a	Lipini et al., 2018
		miR-9		Chen et al., 2014
Adult-born hippocampal neurons	Migration	miR-19	Rargad2	Han et al., 2018
	Dendritic outgrowth	miR-132, miR-212		Yu et al., 2009; Magill et al., 2010
		miR-19		Han et al., 2018
Casal-Retzius cell	Differentiation	miR-9	Foxg1	Shen et al., 2011
Corticospinal motor neurons	Differentiation, axon growth	miR-409-3p	LMN4	Chen et al., 2020
Dopamine neurons	Differentiation, survival	miR-133b	Pitc3	Kim et al., 2007
		miR-200c	Zeb2	Yang et al., 2018
		miR-212	Koht1	Trivedi et al., 2015
Spinal motor neurons	Differentiation, survival			Fecherstein et al., 2013
Dorsal root ganglion neurons	Axon growth	miR-132	Ras1	Han et al., 2014
Retinal ganglion cells	Axon growth	miR-182	Cf1	Brake et al., 2017
Retinal photoreceptors	Differentiation, morphology	miR-183/98/182 cluster		Stankovic et al., 2016
Olfactory interneurons	Migration, dendritic outgrowth	miR-125		Akerman et al., 2011
Olfactory dopamine neurons	Differentiation	miR-7a	Pax6	de Chevigny et al., 2019
Mechanosensory neurons	Subtype specification	miR-183/98/182 cluster	Shox2, Capn2d1/2	Perry et al., 2016
Microglia	Activation state	miR-128, miR-124	Cebpa	Yang et al., 2017
Oligodendrocytes	Differentiation	miR-218, miR-338	Sox6, Hes5, Zfp298	Dugas et al., 2010
				Zhao et al., 2010
		miR-7a	Pax6, NeuroD4	Zhao et al., 2012
		miR-28	UmnB1	Lin and Fu, 2009
	Survival	miR-17/92 cluster	Pten	Baron et al., 2010
Schwann cells	Differentiation, proliferation	miR-34a	Notch1, Cend1	Waller et al., 2011
		miR-143	Egr2	Waller et al., 2011
Astrocytes	Differentiation, proliferation	miR-31	lin-28	Waters et al., 2018
		miR-129b		Pogue et al., 2010

For each cell type, the developmental processes and the relevant miRNAs are listed. In addition, known targets of those miRNAs and the main references are indicated (e.g., in human neural progenitors, the timing and duration of the cell cycle is regulated by the great age-specific miR-2115 through the repression of the DNA replication regulator CCR4).

Methods of Treatment

[0355] Provided herein are methods for treating a disease or disorder, the method comprising administering to a subject in need thereof, a composition as disclosed herein to cells of the CNS. In an aspect, the compositions used in methods disclosed herein are capable of increasing

transduction of neurons and/or glial cells of the CNS, allowing for delivery of cargo and therapeutics directly to such cell types. In an embodiment, a method is disclosed wherein the cargo is one or more polypeptides.

Diseases or Disorders

[0356] In an embodiment, a method is disclosed wherein the disease or disorder is a cancer, neurological disorder, or infection.

[0357] In an embodiment, methods of treatment comprise administering a composition as detailed herein to a subject in need thereof. In an example embodiment, the cancer is a neuroepithelial cancer. In an embodiment, the cancer is a neuroepithelial tumor, for example, Astrocytic tumors, e.g., Diffuse Astrocytoma (fibrillary, protoplasmic, gemistocytic, mixed), Anaplastic (malignant) astrocytoma, Glioblastoma (giant cell, gliosarcoma variants), Pilocytic astrocytoma, Pleomorphic xanthoastrocytoma, or Subependymal giant cell astrocytoma; Oligodendroglial tumors, e.g., Oligodendroglioma, Anaplastic (malignant) Oligodendroglioma, Ependymal tumors, Ependymoma (cellular, papillary, clear cell, tancytic), Anaplastic (malignant) ependymoma, ependymoma, Subependymoma; Mixed tumors, e.g., Oligoastrocytoma or Anaplastic (malignant) oligoastrocytoma; Choroid plexus tumors, e.g., Choroid Plexus papilloma or Choroid Plexus carcinoma; Neuronal and mixed neuronal-glial tumors, e.g., Gangliocytoma, Ganglioglioma, Dysembryoplastic neuroepithelial tumor (DNET), Dysplastic gangliocytoma of the cerebellum (Lhermitte-Duclos), Desmoplastic infantile astrocytoma/ganglioglioma, Central neurocytoma, Anaplastic ganglioglioma, Cerebellar liponeurocytoma, Paraganglioma of the filum terminale; Pineal tumors, e.g., Pineocytoma, Pineoblastoma, Pineal parenchymal tumor of intermediate differentiation; Embryonal tumors, e.g., Medulloblastoma (desmoplastic, large cell, melanotic, medulloblastoma), Medulloepithelioma, Supratentorial primitive neuroectodermal tumors, PNETs such as Neuroblastoma, Ganglioneuroblastoma, Ependymblastoma, or Atypical teratoid/rhabdoid tumor; Neuroblastic tumors, e.g., Olfactory (esthesioneuroblastoma), Olfactory neuroepithelioma, Neuroblastomas of the adrenal gland and sympathetic nervous system; Glial tumors of uncertain etiology, e.g., Astroblastoma, Gliomatosis cerebri, Chordoid glioma of the third ventricle.

[0358] In an embodiment, the cancer is a primary cancer metastasized to brain or other region of the central nervous system.

[0359] In an example embodiment, the neurological disorder is caused by a neurodegenerative or a neurodevelopmental disease. Example neurodegenerative diseases include, but are not limited to, Alzheimer's disease and other memory disorders, amyotrophic lateral sclerosis (ALS), ataxia, Huntington's disease, Parkinson's disease, Friedreich's ataxia, motor neuron disease, multiple system atrophy, progressive supranuclear palsy. Examples of neurodevelopmental diseases include, but are not limited to, attention-deficit/hyperactivity disorder (ADHD), autism, learning disabilities, intellectual disability (also known as mental retardation), conduct disorders, cerebral palsy, speech and language disorders, Tourette syndrome, Schizophrenia, Rett syndrome, Angelman's syndrome, Fragile X syndrome, and impairments in vision and hearing.

[0360] In an embodiment, the disease is a lysosomal storage disease. The lysosomal storage disease may include Gaucher's or Parkinson's disease. In particular, Gaucher's disease may be Type II or Type III Gaucher's disease.

Non-Human Transgenic Animals

[0361] In one aspect, as provided herein a method of creating humanized transgenic non-human animals comprising: delivering to one or more cells of a non-human animal a vector system or a recombinant viral particle comprising a recombinant viral genome, wherein: the vector system or recombinant viral genome encodes a human transferrin polypeptide, wherein the encoded human transferrin polypeptide is under the control of a tissue-specific promoter or miRNA binding element that has selective activity in the desired cell, tissue, or organ. In an example embodiment, the one or more cells are endothelial cells. In an example embodiment, the one or more cells are CNS cells. In an example embodiment, the one or more cells are of the CNS vasculature, lungs, kidneys, liver, or any combination thereof. In an example embodiment, the endothelial cells are of the CNS vasculature. In an example embodiment, the recombinant viral particle, optionally an AAV viral particle, comprises a capsid polypeptide, optionally an AAV capsid polypeptide, wherein the capsid polypeptide comprises a CNS specific transferrin receptor binding modification.

[0362] In one aspect, provided herein a humanized transgenic non-human animal comprising: one or more cells expressing a human transferrin polypeptide, optionally wherein the one or more cells are CNS cells. In an example embodiment, the transgenic non-human animal is a rodent, optionally a mouse. In one aspect, provided herein a humanized transgenic non-human animal

produced by the methods described herein. In an example embodiment, the humanized non-human animal has a suppressed immune system.

[0363] It will be appreciated that in the present methods, where the non-human transgenic organism is multicellular, e.g., an animal or plant, the modification may occur *ex vivo* or *in vitro*, for instance in a cell culture and in some instances not *in vivo*. In other embodiments, it may occur *in vivo*. In an aspect, the disclosure provides a method of modifying an organism or a non-human organism by manipulation of a target sequence in a genomic locus of interest comprising: Delivering, e.g., via particle(s) or nanoparticle(s) or vector(s) (e.g., viral vector, e.g., AAV, adenovirus) a non-naturally occurring or engineered composition.

[0364] A single cell or a population of cells may preferably be modified *ex vivo* and then re-introduced, e.g., transplanted to make transgenic organisms that express TFRC in certain cells. The disclosure In an embodiment comprehends a method of modifying a eukaryote, such as a transgenic eukaryote comprising delivering, e.g., via vector(s) and/or particle(s) and/or nanoparticles a non-naturally occurring or engineered composition. The system may comprise one, two, three or four different vectors; and the system may comprise one, two, three or four different nanoparticle complex(es) delivering the component(s) of the system. Components I, II, III and IV may thus be located on one, two, three or four different vectors, and may be delivered by one, two, three or four different particle or nanoparticle complex(es) or AAVs or components I, II, III and IV can be located on same or different vector(s) / particle(s) / nanoparticle(s), with all combinations of locations envisaged. And complexes that target the CNS or CNS tissue or CNS cells are advantageous.

[0365] In an embodiment, vectors are delivered to the eukaryotic cell in a transgenic eukaryote. In an embodiment, said modifying takes place in said eukaryotic cell in a cell culture. In one aspect, the disclosure provides a method of generating a model eukaryotic cell or a model transgenic eukaryote comprising one or more human proteins.

[0366] In one aspect, the present disclosure provides a transgenic eukaryote, e.g., mouse. In one aspect, the present disclosure provides a constitutive transgenic eukaryote, e.g., mouse line obtained by crossing of the transgenic mouse with another mouse line. In an embodiment, progeny (or progenies) derived from said transgenic eukaryote, e.g., mouse line, may be successfully bred over at least five generations without exhibiting increased levels of genome instability or cellular toxicity. In one aspect, the present disclosure provides a method for simultaneously introducing

multiple mutations *ex vivo* in a tissue, organ or a cell line (of the CNS), or in vivo in the same. It can be appreciated that using the novel transferrin receptor binding modification tools disclosed herein, may produce a transgenic non-human eukaryote, e.g., animal model with multiple mutations in any number of loci can be envisioned and are within the scope of the present disclosure. It will be appreciated that such a transgenic non-human eukaryote, e.g., animal model provides a valuable tool for research purposes, e.g., viral transduction, and opens the door for developing and testing new therapeutic interventions for targeting specific tissue involving mutations at multiple loci. Such uses are within the scope of the present disclosure.

[0367] The eukaryotic cell can comprise a constitutive promoter, or a tissue specific promoter, or an inducible promoter; and, the eukaryotic cell can be part of a non-human transgenic eukaryote, e.g., a non-human mammal, primate, rodent, mouse, rat, rabbit, canine, dog, cow, bovine, sheep, ovine, goat, pig, fowl, poultry, chicken, fish, insect or arthropod; advantageously a mouse. The isolated eukaryotic cell or the non-human transgenic eukaryote can express an additional protein or enzyme, such as transferrin receptor ; and, the expression of transferrin receptor can be driven by coding therefor functionally or operatively linked to a constitutive promoter, or a tissue specific promoter, or an inducible promoter.

[0368] The eukaryotic cell can be a mammalian cell, e.g., a mouse cell, such as a mouse cell that is part of a transgenic mouse having cells that express transferrin receptor .

[0369] Transgenic non-human eukaryotic organisms, e.g., animals are also provided in an aspect of practice of the instant disclosure. Preferred examples include animals comprising transferrin receptor , in terms of polynucleotides encoding transferrin receptor or the protein itself. In certain aspects, the disclosure involves a constitutive or conditional or inducible transferrin receptor non-human eukaryotic organism, such as an animal, e.g., a primate, rodent, e.g., mouse, rat and rabbit, are preferred; and can include a canine or dog, livestock (cow / bovine, sheep / ovine, goat or pig), fish, fowl or poultry, e.g., chicken, and an insect or arthropod, with it mentioned that it is advantageous if the animal is a model as to a human or animal protein, cell, or tissue, as use of the non-human eukaryotic organisms in condition modeling, e.g., via inducing a plurality, are preferred. To generate transgenic mice with the constructs, as exemplified herein one may inject pure, linear DNA into the pronucleus of a zygote from a pseudo pregnant female, e.g. a CB56 female. Founders may then be identified, genotyped, and backcrossed to CB57 mice. The

constructs may then be cloned and optionally verified, for instance by Sanger sequencing. Knock ins are envisaged (alone or in combination).

[0370] Accordingly, the disclosure involves a non-human eukaryote, animal, mammal, primate, rodent, etc. or cell thereof or tissue thereof that may be used as a model. For example, a method of the disclosure may be used to create a non-human eukaryote, e.g., an animal, mammal, primate, rodent or cell that comprises a modification in one or more nucleic acid sequences associated or correlated with a cell or tissue of such (such as the CNS). The cell may be *in vivo* or *ex vivo* in the cases of multicellular organisms. In the instance where the cell is in cultured, a cell line may be established if appropriate culturing conditions are met and preferably if the cell is suitably adapted for this purpose (for instance a stem cell). Hence, cell lines are also envisaged.

[0371] In an aspect the disclosure can involve cells, e.g., non-human eukaryotic, e.g., animal, such as mammal, e.g., primate, rodent, mouse, rat, rabbit, etc., or even human cells, transformed to contain transferrin receptor, e.g., such cells as to which a vector that contains nucleic acid molecule(s) encoding a transferrin receptor, e.g., with nucleic acid(s) encoding a promoter and at least one NLS, advantageously two or more NLSs, or such cells that have had their genome altered, e.g., through the vector being an integrating virus or through such cells being stem cells or cells that give rise to a cell line or a living organism (but wherein such an organism is advantageously non-human), that contains and expresses nucleic acid molecule(s) encoding transferrin receptor. Such cells are then transplanted into or onto an animal suitable for expressing a cell or tissue. The cells proliferate on or in the non-human eukaryote, e.g., animal model. The non-human eukaryote, e.g., animal model, having the proliferated heterologous transplanted transferrin receptor -containing cells, is then administered RNA(s) or vector(s), e.g., AAV, adenovirus containing or providing RNA(s), e.g., under the control of a promoter such as a U6 promoter and/or particle(s) and/or nanoparticle(s). The non-human eukaryote, e.g., animal model can then be used for testing, e.g., as to potential therapy and/or putative treatment via a possibly pharmaceutically active compound. The administering can be at or to or for body delivery to the proliferated heterologous transplanted transferrin receptor -containing cells, e.g., direct injection at or near such proliferated heterologous transplanted transferrin receptor -containing cells, or injection or other administration in such a way that the RNA(s) are delivered into the proliferated heterologous transplanted transferrin receptor -containing cells, e.g., injection into the bloodstream whereby bodily functions transport to the proliferated heterologous transplanted transferrin receptor -

containing cells. In an aspect of the disclosure, barcoding techniques of WO/2013/138585 A1 can be adapted or integrated into the practice of the disclosure. WO/2013/138585 A1 provides methods for simultaneously determining the effect of a test condition on viability or proliferation of each of a plurality of genetically heterogeneous cell types. The number of living cells in the sample after exposure to the test condition as compared to the reference number of cells indicates the effect of the test condition on viability or proliferation of each cell type. WO/2013/138585 A1 also provides methods for simultaneously determining the effect of a test condition on viability or proliferation of each of a plurality of genetically heterogeneous cell types. Referenced patent applications are hereby incorporated by reference.

[0372] The practice of the present disclosure employs, unless otherwise indicated, conventional techniques for generation of genetically modified mice. See Marten H. Hofker and Jan van Deursen, TRANSGENIC MOUSE METHODS AND PROTOCOLS, 2nd edition (2011).

Screening and Cell Selection

[0373] In one aspect, as provided herein a method of screening an AAV library for a recombinant AAV particle which binds to a transferrin receptor and/or has increased tropism for the CNS, said method comprising assaying the AAV library of any of those described herein for increased binding of a transferrin receptor and/or tropism for the CNS relative to an AAV vector with a reference capsid and selecting those recombinant AAV vectors which have increased binding of transferrin receptor and/or tropism for the CNS. In an example embodiment, the composition is a viral particle comprising one or more capsid proteins each comprising the candidate transferrin receptor binding modification. In an example embodiment, at least one of the one or more compositions further comprises a cargo. In an example embodiment, the cargo is or encodes a therapeutic nucleic acid or polypeptide, a selectable marker, or a control polypeptide or nucleic acid.

[0374] In one aspect, a method of screening transferrin receptor binding modifications, including incorporated within an AAV, capable of transduction of central nervous system (CNS) tissues via binding to a transferrin receptor in humanized transgenic non-human animals comprising: (a) introducing a plurality of vector systems to one or more humanized transgenic non-human animal that express the human TFRC and (b) detecting the transferrin receptor binding modification that binds to a transferrin receptor.

[0375] The engineered AAV capsid system vectors, engineered cells, and/or engineered AAV capsid particles described herein can be used in a screening assay and/or cell selection assay. The engineered delivery system vectors, engineered cells, and/or engineered AAV capsid particles can be delivered to a subject and/or cell. In an embodiment, the cell is a eukaryotic cell. The cell can be in vitro, ex vivo, in situ, or in vivo. The engineered AAV capsid system molecules, vectors, engineered cells, and/or engineered AAV capsid particles described herein can introduce an exogenous molecule or compound to subject or cell to which they are delivered. The presence of an exogenous molecule or compound can be detected which can allow for identification of a cell and/or attribute thereof. In an embodiment, the delivered molecules or particles can impart a gene or other nucleotide modification (e.g., mutations, gene or polynucleotide insertion and/or deletion, etc.). In an embodiment the nucleotide modification can be detected in a cell by sequencing. In an embodiment, the nucleotide modification can result in a physiological and/or biological modification to the cell that results in a detectable phenotypic change in the cell, which can allow for detection, identification, and/or selection of the cell. In an embodiment, the phenotypic change can be cell death, such as embodiments where binding of a CRISPR complex to a target polynucleotide results in cell death. Embodiments of the disclosure allow for selection of specific cells without requiring a selection marker or a two-step process that may include a counter-selection system. The cell(s) may be prokaryotic or eukaryotic cells.

[0376] In an embodiment, the disclosure provides for a method of selecting one or more cell(s) by introducing one or more mutations in a gene in the one or more cell(s), the method comprising: introducing one or more vectors, which can include one or more engineered delivery system molecules or vectors described elsewhere herein, into the cell(s), wherein the one or more vectors can include a CRISPR enzyme and/or drive expression of one or more of: a guide sequence linked to a tracr mate sequence, a tracr sequence, and an editing template; or other polynucleotide to be inserted into the cell and/or genome thereof; wherein, for example that which is being expressed is within and expressed in vivo by the CRISPR enzyme and/or the editing template, when included, comprises the one or more mutations that abolish CRISPR enzyme cleavage; allowing homologous recombination of the editing template with the target polynucleotide in the cell(s) to be selected; allowing a CRISPR complex to bind to a target polynucleotide to effect cleavage of the target polynucleotide within said gene, wherein the CRISPR complex comprises the CRISPR enzyme complexed with (1) the guide sequence that is hybridized to the target sequence within the target

polynucleotide, and (2) the tracer sequence that is hybridized to the target sequence, wherein binding of the CRISPR complex to the target polynucleotide induces cell death, thereby allowing one or more cell(s) in which one or more mutations have been introduced to be selected. In a preferred embodiment, the CRISPR enzyme is a Cas protein. In another embodiment of the disclosure the cell to be selected may be a eukaryotic cell.

[0377] The screening methods involving the engineered AAV capsid system molecules, vectors, engineered cells, and/or engineered AAV capsid particles, including but not limited to those that deliver one or more CRISPR-Cas system molecules to cell, can be used in detection methods such as fluorescence in situ hybridization (FISH). In an embodiment, one or more components of an engineered CRISPR-Cas system that includes a catalytically inactive Cas protein, can be delivered by an engineered AAV capsid system molecule, engineered cell, and/or engineered AAV capsid particle described elsewhere herein to a cell and used in a FISH method. The CRISPR-Cas system can include an inactivated Cas protein (dCas) (e.g., a dCas9), which lacks the ability to produce DNA double-strand breaks may be fused with a marker, such as fluorescent protein, such as the enhanced green fluorescent protein (eGFP) and co-expressed with small guide RNAs to target pericentric, centric and telomeric repeats in vivo. The dCas system can be used to visualize both repetitive sequences and individual genes in the human genome. Such new applications of labelled dCas, dCas CRISPR-Cas systems, engineered AAV capsid system molecules, engineered cells, and/or engineered AAV capsid particles can be used in imaging cells and studying the functional nuclear architecture, especially in cases with a small nucleus volume or complex 3-D structures. (Chen B, Gilbert LA, Cimini BA, Schnitzbauer J, Zhang W, Li GW, Park J, Blackburn EH, Weissman JS, Qi LS, Huang B. 2013. Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155(7):1479-91. doi: 10.1016/j.cell.2013.12.001., the teachings of which can be applied and/or adapted to the CRISPR systems described herein. A similar approach involving a polynucleotide fused to a marker (e.g., a fluorescent marker) can be delivered to a cell via an engineered AAV capsid system molecule, vector, engineered cell, and/or engineered AAV capsid particle described herein and integrated into the genome of the cell and/or otherwise interact with a region of the genome of a cell for FISH analysis.

[0378] Similar approaches for studying other cell organelles and other cell structures can be accomplished by delivering to the cell (e.g., via an engineered delivery AAV capsid molecule,

engineered cell, and/or engineered AAV capsid particle described herein) one or more molecules fused to a marker (such as a fluorescent marker), wherein the molecules fused to the marker are capable of targeting one or more cell structures. By analyzing the presence of the markers, one can identify and/or image specific cell structures.

[0379] In an embodiment, the engineered AAV capsid system molecules and/or engineered AAV capsid particles can be used in a screening assay inside or outside of a cell. In an embodiment, the screening assay can include delivering a CRISPR-Cas cargo molecule(s) via an engineered AAV capsid particle.

[0380] Use of the present system in screening is also provided by the present disclosure, e.g., gain of function screens. Cells which are artificially forced to overexpress a gene are able to down regulate the gene over time (re-establishing equilibrium) e.g., by negative feedback loops. By the time the screen starts the unregulated gene might be reduced again. Other screening assays are discussed elsewhere herein.

[0381] In an embodiment, the disclosure provides a cell from or of an *in vitro* method of delivery, wherein the method comprises contacting the delivery system with a cell, optionally a eukaryotic cell, whereby there is delivery into the cell of constituents of the delivery system, and optionally obtaining data or results from the contacting, and transmitting the data or results.

[0382] In an embodiment, the disclosure provides a cell from or of an *in vitro* method of delivery, wherein the method comprises contacting the delivery system with a cell, optionally a eukaryotic cell, whereby there is delivery into the cell of constituents of the delivery system, and optionally obtaining data or results from the contacting, and transmitting the data or results; and wherein the cell product is altered compared to the cell not contacted with the delivery system, for example altered from that which would have been wild type of the cell but for the contacting. In an embodiment, the cell product is non-human or animal. In an embodiment, the cell product is human.

[0383] In an embodiment, a host cell is transiently or non-transiently transfected with one or more vectors described herein. In an embodiment, a cell is transfected as it naturally occurs in a subject optionally to be reintroduced therein. In an embodiment, a cell that is transfected is taken from a subject. In an embodiment, the cell obtained from or is derived from cells taken from a subject, such as a cell line. Delivery mechanisms and techniques of the engineered AAV capsid system, engineered AAV capsid particles are described elsewhere herein.

[0384] In an embodiment, it is envisaged to introduce the engineered AAV capsid system molecule(s) and/or engineered AAV capsid particle(s) directly to the host cell. For instance, the engineered AAV capsid system molecule(s) can be delivered together with one or more cargo molecules to be packaged into an engineered AAV capsid particle.

[0385] The disclosure is further described in the following examples, which do not limit the scope of the disclosure described in the claims.

Machine Learning

[0386] Machine learning is a field of study within artificial intelligence that allows computers to learn functional relationships between inputs and outputs without being explicitly programmed. Machine learning involves a module comprising algorithms that may learn from existing data by analyzing, categorizing, or identifying the data. Such machine-learning algorithms operate by first constructing a model from training data to make predictions or decisions expressed as outputs. In an embodiment, the training data includes data for one or more identified features and one or more outcomes, for example a transferrin receptor sequence and a plurality of AAV capsid polypeptide sequences comprising a modification for binding to a transferrin receptor. Although example embodiments are presented with respect to a few machine-learning algorithms, the principles presented herein may be applied to other machine-learning algorithms.

[0387] Data supplied to a machine learning algorithm can be considered a feature, which can be described as an individual measurable property of a phenomenon being observed. The concept of feature is related to that of an independent variable used in statistical techniques such as those used in linear regression. The performance of a machine learning algorithm in pattern recognition, classification and regression is highly dependent on choosing informative, discriminating, and independent features. Features may comprise numerical data, categorical data, time-series data, strings, graphs, or images. Features of the disclosure may further comprise a plurality of AAV capsid polypeptide sequences comprising a modification for binding to a transferrin receptor. These plurality of AAV capsid polypeptide sequences comprising a modification for binding to a transferrin receptor may include any one or more SEQ ID NO: 60-1388.

[0388] In general, there are two categories of machine learning problems: classification problems and regression problems. Classification problems, also referred to as categorization problems, aim at classifying items into discrete category values. Training data teaches the classifying algorithm how to classify. In an embodiment, features to be categorized may include a

plurality of AAV capsid polypeptide sequences comprising a modification for binding to a transferrin receptor, which can be provided to the classifying machine learning algorithm and then placed into categories of, for example, one or more novel transferrin receptor binding modifications. Regression algorithms aim at quantifying and correlating one or more features. Training data teaches the regression algorithm how to correlate the one or more features into a quantifiable value. In an embodiment, features such as a plurality of AAV capsid polypeptide sequences comprising a modification for binding to a transferrin receptor can be provided to the regression machine learning algorithm resulting in one or more continuous values, for example, of one or more novel transferrin receptor binding modifications.

Generative Machine Learning Models

[0389] In an embodiment, the machine learning network comprises a generative model. A generative model typically refers to a model that can generate new data instances. In general, once a generative model is trained to learn a pattern from the training data, the generative model can then generate new data that resembles (e.g., shares similar characteristics) the pattern from the training data set. For example, a generative model can take in an input X, for example a transferrin receptor sequence, and generate an output Y, for example one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1, based on computed probabilities of sequential outcomes.

[0390] Example generative models include, Boltzmann machines, Bayesian networks, autoregressive models, variational auto encoders (VAEs), diffusion models, energy-based models, flow-based models, generative adversarial networks (GANs), mixture models, hidden Markov models, or large language models (LLMs), convolutional neural network (CNN), recurrent neural networks (RNNs), long short-term memory models (LSTMs), gated recurrent units (GRUs), capsule networks, attention mechanisms, or transformer networks. These models are known in the art but are further described herein. For example, a machine learning model can be trained with the AAV capsid polypeptide sequences comprising a modification for binding to a TfR1 as described herein to generate a transferrin receptor (TfR1) targeting machine learning model, see e.g., Shin, J.E., Riesselman, A.J., Kollasch, A.W. *et al.* Protein design and variant prediction using autoregressive generative models. *Nat Commun* 12, 2403 (2021).

Embedding

[0391] In one example, the machine learning module may use embedding to provide a lower dimensional representation, such as a vector, of features to organize them based off respective similarities. In some situations, these vectors can become massive. In the case of massive vectors, particular values may become very sparse among a large number of values (e.g., a single instance of a value among 50,000 values). Because such vectors are difficult to work with, reducing the size of the vectors, in some instances, is necessary. A machine learning module can learn the embeddings along with the model parameters. In an embodiment, features such as a plurality of AAV capsid polypeptide sequences comprising a modification for binding to a transferrin receptor can be mapped to vectors implemented in embedding methods. In an embodiment, embedded semantic meanings are utilized. Embedded semantic meanings are values of respective similarity. For example, the distance between two vectors, in vector space, may imply two values located elsewhere with the same distance are categorically similar. Embedded semantic meanings can be used with similarity analysis to rapidly return similar values. In an embodiment, a transferrin receptor sequence and/or a plurality of AAV capsid polypeptide sequences comprising a modification for binding to a transferrin receptor is embedded. In an embodiment, the methods herein are developed to identify meaningful portions of the vector and extract semantic meanings between that space.

Natural Language Processing

[0392] In an embodiment, the plurality of AAV capsid polypeptide sequences comprising a modification for binding to a transferrin receptor is processed with a natural language processing (NLP) model. NLP is a computational approach to evaluating text. General applications of NLP may comprise information retrieval, information extraction, question-answering, summarization, machine translation, dialogue systems. Information extraction comprises of recognition, tagging, and extraction into a structured representation. For example, key elements of information, such as persons, companies, locations, and/or organizations, are extracted from collections of text. Question-answering (QA) comprises of generating a list of documents corresponding to a user's query. For example, QA may comprise generating either just the text of the answer or answer-providing passages. Summarization comprises of reducing a large amount of text into a smaller amount of text. For example, summarization may comprise of an abbreviated narrative representation of the original document. Machine translation (MT) comprises of either rule-based

or probabilistic methods (e.g., machine learning) for translating text to speech or vice versa, or from one language to another. For example, MT attempts to capture contextual, idiomatic and pragmatic nuances of language. Dialogue systems comprise of conversational communications through communication modes such as text, speech, and images.

[0393] NPL may comprise of different levels of language. For example, ascending levels of language may comprise of phonology, morphology, lexical, syntactic, semantic, discourse, and pragmatic. The phonology level comprises of the interpretation of speech sounds. For example, phonological analysis may comprise of phonetic rules, phonemic rules, and prosodic rules. Morphology comprises of the nature of words, which comprise of morphemes (i.e., smallest units of meaning). For example, the suffix -ed to a verb indicates the action of the verb took place in the past. The lexical level comprises of interpreting the meaning of individual words. For example, words may be assigned part-of-speech tags based on context. The syntactic level comprises of analyzing the words in a sentence to extract the grammatical structure of the sentence. For example, syntactic processes attempt to compute the meaning of a sentence from the order and dependency of the words in a sentence. The semantic level comprises of determining the meaning of a sentence by analyzing the interactions between word-level meanings in a sentence. For example, a semantic process may comprise disambiguation wherein words with multiple meanings are reduced into a singular meaning based on the context of the sentence. The discourse level comprises of determining the meaning of more than one sentence. For example, discourse processing may comprise of anaphora resolution and discourse/text structure recognition. The pragmatic level comprises of determining the use of language based on the context of the text, not necessarily the content of the text. For example, pragmatic processes deduce extra meanings read into text that are not necessarily encoded into the text.

[0394] NPL may also comprise different approaches to language processing. For example, the approaches may comprise of symbolic, statistical, connectionist, and hybrid. Symbolic approaches comprise of using explicit representations of facts through well-understood knowledge representation schemes to analyze linguistic phenomena. For example, symbolic approaches may comprise of logic or rule-based systems. Statical approaches comprise of using text corpora to determine generalized models of linguistic phenomena. For example, a statical approach may use a HMM to determine speech recognition, lexical acquisition, parsing, part-of-speech tagging, collocations, statistical machine translation, statistical grammar learning, etc. Connectionist

approach comprises of combining statistical learning with various theories of representation. For example, a connectionist approach may comprise of a network of interconnected local processing units with knowledge stored as weights in the connections between units wherein local interactions may result in an observed global behavior. See e.g., Liddy, E.D. 2001. Natural Language Processing. In Encyclopedia of Library and Information Science, 2nd Ed. NY. Marcel Decker, Inc.

[0395] In a sequence context, NPL may connect patterns among sequences and produce variations that are correlated to the connected pattern.

Training Methods

[0396] In an embodiment, the machine learning module can be trained using techniques such as unsupervised, supervised, semi-supervised, reinforcement learning, transfer learning, incremental learning, curriculum learning techniques, and/or learning to learn. Training typically occurs after selection and development of a machine learning module and before the machine learning module is operably in use. In one aspect, the training data used to teach the machine learning module can comprise input data such as a plurality of AAV capsid polypeptide sequences comprising a modification for binding to a transferrin receptor and the respective target output data such as novel transferrin receptor binding modifications.

[0397] In an embodiment, after the machine learning module has been trained, the trained machine learning module may be deployed. In an embodiment, the trained machine learning module is deployed by one or more computing device. Deploy, deploying, to deploy are terms of art and generally refer to taking a trained machine learning module and using it in a production environment. For example, allowing the machine learning module to receive new data (e.g., data that is not solely for the purpose of training) and generate output data. In another example, deploying a machine learning module includes making the module usable for its intended purpose.

[0398] In an embodiment, the TfR1 targeting machine learning model is trained to identify one or more sequences from the plurality of sequences that decreases transduction of off-target tissues. It is to be understood that one or more sequences from the plurality of sequences has a decreases transduction of off-target tissues if the off-target tissue binding is different than the off-target tissue binding of a sequence not generated by the TfR1 targeting machine learning model. Off-target tissue binding can be determined by means known in the art. By means of example, and without limitation, off-target tissue binding can be determined by calculating binding strength or affinity (such as based on equilibrium constants, K_a , K_d , etc.). In certain embodiments, off-target

bindings increased. In certain embodiments, off-target binding is increased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or at least 100%. In certain embodiments, off-target binding is decreased. In certain embodiments, off-target binding is decreased by at least 5%, preferably at least 10%, more preferably at least 20%, such as at least 30%, at least 40%, at least 50%, at least 60%, at least 70%, at least 80%, at least 90%, or (substantially) 100%.

Unsupervised and Supervised Learning

[0399] In an example embodiment, unsupervised learning is implemented. Unsupervised learning can involve providing all or a portion of unlabeled training data to a machine learning module. The machine learning module can then determine one or more outputs implicitly based on the provided unlabeled training data. In an example embodiment, supervised learning is implemented. Supervised learning can involve providing all or a portion of labeled training data to a machine learning module, with the machine learning module determining one or more outputs based on the provided labeled training data, and the outputs are either accepted or corrected depending on the agreement to the actual outcome of the training data. In some examples, supervised learning of machine learning system(s) can be governed by a set of rules and/or a set of labels for the training input, and the set of rules and/or set of labels may be used to correct inferences of a machine learning module.

Semi-Supervised and Reinforcement Learning

[0400] In an embodiment, semi-supervised learning is implemented. Semi-supervised learning can involve providing all or a portion of training data that is partially labeled to a machine learning module. During semi-supervised learning, supervised learning is used for a portion of labeled training data, and unsupervised learning is used for a portion of unlabeled training data. In an embodiment, reinforcement learning is implemented. Reinforcement learning can involve first providing all or a portion of the training data to a machine learning module and as the machine learning module produces an output, the machine learning module receives a “reward” signal in response to a correct output. Typically, the reward signal is a numerical value and the machine learning module is developed to maximize the numerical value of the reward signal. In addition, reinforcement learning can adopt a value function that provides a numerical value representing an expected total of the numerical values provided by the reward signal over time.

Transfer Learning

[0401] In an embodiment, transfer learning is implemented. Transfer learning techniques can involve providing all or a portion of a first training data to a machine learning module, then, after training on the first training data, providing all or a portion of a second training data. In an embodiment, a first machine learning module can be pre-trained on data from one or more computing devices. The first trained machine learning module is then provided to a computing device, where the computing device is intended to execute the first trained machine learning model to produce an output. Then, during the second training phase, the first trained machine learning model can be additionally trained using additional training data, where the training data can be derived from kernel and non-kernel data of one or more computing devices. This second training of the machine learning module and/or the first trained machine learning model using the training data can be performed using either supervised, unsupervised, or semi-supervised learning. In addition, it is understood transfer learning techniques can involve one, two, three, or more training attempts. Once the machine learning module has been trained on at least the training data, the training phase can be completed. The resulting trained machine learning model can be utilized as at least one of trained machine learning module.

Incremental and Curriculum Learning

[0402] In an embodiment, incremental learning is implemented. Incremental learning techniques can involve providing a trained machine learning module with input data that is used to continuously extend the knowledge of the trained machine learning module. Another machine learning training technique is curriculum learning, which can involve training the machine learning module with training data arranged in a particular order, such as providing relatively easy training examples first, then proceeding with progressively more difficult training examples. As the name suggests, difficulty of training data is analogous to a curriculum or course of study at a school.

Learning to Learn

[0403] In an embodiment, learning to learn is implemented. Learning to learn, or meta-learning, comprises, in general, two levels of learning: quick learning of a single task and slower learning across many tasks. For example, a machine learning module is first trained and comprises of a first set of parameters or weights. During or after operation of the first trained machine learning module, the parameters or weights are adjusted by the machine learning module. This process occurs iteratively on the success of the machine learning module. In another example, an optimizer,

or another machine learning module, is used wherein the output of a first trained machine learning module is fed to an optimizer that constantly learns and returns the final results. Other techniques for training the machine learning module and/or trained machine learning module are possible as well.

Contrastive Learning

[0404] In example embodiment, contrastive learning is implemented. Contrastive learning is a self-supervised model of learning in which training data is unlabeled is considered as a form of learning in-between supervised and unsupervised learning. This method learns by contrastive loss, which separates unrelated (i.e., negative) data pairs and connects related (i.e., positive) data pairs. For example, to create positive and negative data pairs, more than one view of a datapoint, such as rotating an image or using a different time-point of a video, is used as input. Positive and negative pairs are learned by solving dictionary look-up problem. The two views are separated into query and key of a dictionary. A query has a positive match to a key and negative match to all other keys. The machine learning module then learns by connecting queries to their keys and separating queries from their non-keys. A loss function, such as those described herein, is used to minimize the distance between positive data pairs (e.g., a query to its key) while maximizing the distance between negative data points. See e.g., Tian, Yonglong, et al. “What makes for good views for contrastive learning?” *Advances in Neural Information Processing Systems* 33 (2020): 6827-6839.

Pre-Trained Learning

[0405] In an embodiment, the machine learning module is pre-trained. A pre-trained machine learning model is a model that has been previously trained to solve a similar problem. The pre-trained machine learning model is generally pre-trained with similar input data to that of the new problem. A pre-trained machine learning model further trained to solve a new problem is generally referred to as transfer learning, which is described herein. In some instances, a pre-trained machine learning model is trained on a large dataset of related information. The pre-trained model is then further trained and tuned for the new problem. Using a pre-trained machine learning module provides the advantage of building a new machine learning module with input neurons/nodes that are already familiar with the input data and are more readily refined to a particular problem. For example, a machine learning module previously trained using a plurality of AAV capsid polypeptide sequences may be further trained to estimate novel transferrin receptor binding

modifications. See e.g., Diamant N, *et al.* Patient contrastive learning: A performant, expressive, and practical approach to electrocardiogram modeling. PLoS Comput Biol. 2022 Feb 14;18(2):e1009862.

[0406] In some examples, after the training phase has been completed but before producing predictions expressed as outputs, a trained machine learning module can be provided to a computing device where a trained machine learning module is not already resident, in other words, after training phase has been completed, the trained machine learning module can be downloaded to a computing device. For example, a first computing device storing a trained machine learning module can provide the trained machine learning module to a second computing device. Providing a trained machine learning module to the second computing device may comprise one or more of communicating a copy of trained machine learning module to the second computing device, making a copy of trained machine learning module for the second computing device, providing access to trained machine learning module to the second computing device, and/or otherwise providing the trained machine learning system to the second computing device. In an embodiment, a trained machine learning module can be used by the second computing device immediately after being provided by the first computing device. In some examples, after a trained machine learning module is provided to the second computing device, the trained machine learning module can be installed and/or otherwise prepared for use before the trained machine learning module can be used by the second computing device.

[0407] After a machine learning model has been trained it can be used to output, estimate, infer, predict, generate, produce, or determine, for simplicity these terms will collectively be referred to as results. A trained machine learning module can receive input data and operably generate results. As such, the input data can be used as an input to the trained machine learning module for providing corresponding results to kernel components and non-kernel components. For example, a trained machine learning module can generate results in response to requests. In an embodiment, a trained machine learning module can be executed by a portion of other software. For example, a trained machine learning module can be executed by a result daemon to be readily available to provide results upon request.

[0408] In an embodiment, a machine learning module and/or trained machine learning module can be executed and/or accelerated using one or more computer processors and/or on-device co-processors. Such on-device co-processors can speed up training of a machine learning module

and/or generation of results. In some examples, trained machine learning module can be trained, reside, and execute to provide results on a particular computing device, and/or otherwise can make results for the particular computing device.

[0409] Input data can include data from a computing device executing a trained machine learning module and/or input data from one or more computing devices. In an embodiment, a trained machine learning module can use results as input feedback. A trained machine learning module can also rely on past results as inputs for generating new results. In an embodiment, input data can comprise a plurality of AAV capsid polypeptide sequences comprising a modification for binding to a transferrin receptor and, when provided to a trained machine learning module, results in output data such as novel transferrin receptor binding modifications.

Algorithms

[0410] Different machine-learning algorithms have been contemplated to carry out the embodiments discussed herein. For example, Boltzmann machines, Bayesian networks, autoregressive models, variational auto encoders (VAEs), diffusion models, energy-based models, flow-based models, generative adversarial networks (GANs), mixture models, hidden Markov models, or large language models (LLMs) for a classical or deep generative model that generates AAV capsid polypeptide sequences comprising a modification for binding to a TfR1. Additional examples include convolutional neural network (CNN), recurrent neural networks (RNNs), long short-term memory models (LSTMs), gated recurrent units (GRUs), capsule networks, attention mechanisms, or transformer networks. For example, linear regression (LiR), logistic regression (LoR), Bayesian networks (for example, naive-bayes), random forest (RF) (including decision trees), neural networks (NN) (also known as artificial neural networks), matrix factorization, a hidden Markov model (HMM), support vector machines (SVM), K-means clustering (KMC), K-nearest neighbor (KNN), a suitable statistical machine learning algorithm, and/or a heuristic machine learning system for classifying or evaluating one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

Generative Adversarial Network

[0411] In an example embodiment, the machine learning network comprises a generative adversarial network (GAN). A generative model, in general, can summarize the distribution of input variables and generates new input variables in the input distribution. For example, a (GAN) may have learned a distribution, such as a Gaussian distribution, for a variable. The GAN can then

summarize the data distribution, and then generate new input variables that fit into the distribution. GANs may include two components: a generator and a discriminator. A generator creates new input variables from the learned space. In particular, the generator takes a, for example, fixed-length random vector, which is drawn from a Gaussian distribution, as input and generates a sample in the domain. The multidimensional vector space forms a compressed representation of the data distribution. A discriminator determines whether an input variable is real (i.e., a user input) or fake (i.e., generated) using a binary class label. In some instance, a discriminator can be used in transfer learning. The discriminator “trains” the generator to produce input variables that are indiscernible from the user input data. See e.g., Goodfellow, I. J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative Adversarial Networks. arXiv 2014. In an embodiment, the input data comprises of a transferrin receptor sequence and the GAN generates one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

[0412] In an example embodiment, the generative adversarial network comprises a Wasserstein Generative Adversarial Network (WGAN). A WGAN modifies a GAN by using a critic that scores the realness or fakeness of an input instead of a discriminator that may classify the realness or fakeness of an input. The advantage to a critic over a discriminator may be, for example, minimizing the distance between the distribution of the data observed in the user input dataset and the distribution observed in generated input. In some instances, a continuous distribution is desired when a finer sensitivity to the input is needed. In some instance, the critic model comprises of a linear activation function. In an embodiment, the input data comprises of a transferrin receptor sequence and the WGAN generates one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

Convolutional Autoencoder

[0413] In an embodiment, convolutional autoencoder (CAE) is implemented. A CAE is a type of neural network and comprises, in general, two main components. First, the convolutional operator that filters an input signal to extract features of the signal. Second, an autoencoder that learns a set of signals from an input and reconstructs the signal into an output. By combining these two components, the CAE learns the optimal filters that minimize reconstruction error resulting an improved output. CAEs are trained to only learn filters capable of feature extraction that can be used to reconstruct the input. Generally, convolutional autoencoders implement unsupervised

learning. In an embodiment, the convolutional autoencoder is a variational convolutional autoencoder. In an embodiment, features from a transferrin receptor sequence are used as an input signal into a CAE which reconstructs that signal into an output such as a one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

Bayesian Network

[0414] In one example embodiment, a Bayesian Network is implemented. BNs are used in machine learning to make predictions through Bayesian inference from probabilistic graphical models. In BNs, input features are mapped onto a directed acyclic graph forming the nodes of the graph. The edges connecting the nodes contain the conditional dependencies between nodes to form a predicative model. For each connected node the probability of the input features resulting in the connected node is learned and forms the predictive mechanism. The nodes may comprise the same, similar or different probability functions to determine movement from one node to another. The nodes of a Bayesian network are conditionally independent of its non-descendants given its parents thus satisfying a local Markov property. This property affords reduced computations in larger networks by simplifying the joint distribution.

[0415] There are multiple methods to evaluate the inference, or predictability, in a BN but only two are mentioned for demonstrative purposes. The first method involves computing the joint probability of a particular assignment of values for each variable. The joint probability can be considered the product of each conditional probability and, in some instances, comprises the logarithm of that product. The second method is Markov chain Monte Carlo (MCMC), which can be implemented when the sample size is large. MCMC is a well-known class of sample distribution algorithms and will not be discussed in detail herein.

[0416] The assumption of conditional independence of variables forms the basis for Naïve Bayes classifiers. This assumption implies there is no correlation between different input features. As a result, the number of computed probabilities is significantly reduced as well as the computation of the probability normalization. While independence between features is rarely true, this assumption exchanges reduced computations for less accurate predictions, however the predictions are reasonably accurate. In an embodiment, a transferrin receptor sequence is mapped to the BN graph to train the BN machine learning module, which, after training, is used to estimate one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

Hidden Markov Model

[0417] In an embodiment, a hidden Markov model is implemented. A HMM takes advantage of the statistical Markov model to predict an outcome. A Markov model assumes a Markov process, wherein the probability of an outcome is solely dependent on the previous event. In the case of HMM, it is assumed an unknown or “hidden” state is dependent on some observable event. A HMM comprises a network of connected nodes. Traversing the network is dependent on three model parameters: start probability; state transition probabilities; and observation probability. The start probability is a variable that governs, from the input node, the most plausible consecutive state. From there each node i has a state transition probability to node j . Typically the state transition probabilities are stored in a matrix M_{ij} wherein the sum of the rows, representing the probability of state i transitioning to state j , equals 1. The observation probability is a variable containing the probability of output o occurring. These too are typically stored in a matrix N_{oj} wherein the probability of output o is dependent on state j . To build the model parameters and train the HMM, the state and output probabilities are computed. This can be accomplished with, for example, an inductive algorithm. Next, the state sequences are ranked on probability, which can be accomplished, for example, with the Viterbi algorithm. Finally, the model parameters are modulated to maximize the probability of a certain sequence of observations. This is typically accomplished with an iterative process wherein the neighborhood of states is explored, the probabilities of the state sequences are measured, and model parameters updated to increase the probabilities of the state sequences. In an embodiment, a transferrin receptor sequence is used to train the nodes/states of the HMM machine learning module, which, after training, is used to estimate one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

Recurrent Neural Network (RNN)

[0418] In an example embodiment, a recurrent neural network is implemented. RNNs are class of NNs further attempting to replicate the biological neural networks of the brain. RNNs comprise of delay differential equations on sequential data or time series data to replicate the processes and interactions of the human brain. RNNs have “memory” wherein the RNN can take information from prior inputs to influence the current output. RNNs can process variable length sequences of inputs by using their “memory” or internal state information. Where NNs may assume inputs are independent from the outputs, the outputs of RNNs may be dependent on prior elements with the

input sequence. For example, input such as a transferrin receptor sequence is received by a RNN, which determines one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1. See Sherstinsky, Alex. "Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network." *Physica D: Nonlinear Phenomena* 404 (2020): 132306.

Long Short-term Memory (LSTM)

[0419] In an example embodiment, a Long Short-term Memory is implemented. LSTM are a class of RNNs designed to overcome vanishing and exploding gradients. In RNNs, long term dependencies become more difficult to capture because the parameters or weights either do not change with training or fluctuate rapidly. This occurs when the RNN gradient exponentially decreases to zero, resulting in no change to the weights or parameters, or exponentially increases to infinity, resulting in large changes in the weights or parameters. This exponential effect is dependent on the number of layers and multiplicative gradient. LSTM overcomes the vanishing/exploding gradients by implementing “cells” within the hidden layers of the NN. The “cells” comprise three gates: an input gate, an output gate, and a forget gate. The input gate reduces error by controlling relevant inputs to update the current cell state. The output gate reduces error by controlling relevant memory content in the present hidden state. The forget gate reduces error by controlling whether prior cell states are put in “memory” or forgotten. The gates use activation functions to determine whether the data can pass through the gates. While one skilled in the art would recognize the use of any relevant activation function, example activation functions are sigmoid, tanh, and RELU. See Zhu, Xiaodan, *et al.* "Long short-term memory over recursive structures." International Conference on Machine Learning. PMLR, 2015.

Convolutional Neural Network (CNN)

[0420] In an example embodiment, a convolutional neural network is implemented. CNNs is a class of NNs further attempting to replicate the biological neural networks, but of the animal visual cortex. CNNs process data with a grid pattern to learn spatial hierarchies of features. Wherein NNs are highly connected, sometimes fully connected, CNNs are connected such that neurons corresponding to neighboring data (e.g., pixels) are connected. This significantly reduces the number of weights and calculations each neuron must perform.

[0421] In general, input data, such as a transferrin receptor sequence, comprises of a multidimensional vector. A CNN, typically, comprises of three layers: convolution, pooling, and

fully connected. The convolution and pooling layers extract features and the fully connected layer combines the extracted features into an output, such as one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

[0422] In particular, the convolutional layer comprises of multiple mathematical operations such as of linear operations, a specialized type being a convolution. The convolutional layer calculates the scalar product between the weights and the region connected to the input volume of the neurons. These computations are performed on kernels, which are reduced dimensions of the input vector. The kernels span the entirety of the input. The rectified linear unit (i.e., ReLu) applies an elementwise activation function (e.g., sigmoid function) on the kernels.

[0423] CNNs can optimized with hyperparameters. In general, there three hyperparameters are used: depth, stride, and zero-padding. Depth controls the number of neurons within a layer. Reducing the depth may increase the speed of the CNN but may also reduce the accuracy of the CNN. Stride determines the overlap of the neurons. Zero-padding controls the border padding in the input.

[0424] The pooling layer down-samples along the spatial dimensionality of the given input (i.e., convolutional layer output), reducing the number of parameters within that activation. As an example, kernels are reduced to dimensionalities of 2x2 with a stride of 2, which scales the activation map down to 25%. The fully connected layer uses inter-layer-connected neurons (i.e., neurons are only connected to neurons in other layers) to score the activations for classification and/or regression. Extracted features may become hierarchically more complex as one layer feeds its output into the next layer. See O'Shea, K.; Nash, R. An Introduction to Convolutional Neural Networks. arXiv 2015 and Yamashita, R., *et al* Convolutional neural networks: an overview and application in radiology. *Insights Imaging* **9**, 611–629 (2018).

Linear Regression (LiR)

[0425] In one example embodiment, linear regression machine learning is implemented. LiR is typically used in machine learning to predict a result through the mathematical relationship between an independent and dependent variable, such as AAV capsid polypeptide sequences and AAV capsid polypeptide sequences comprising a modification for binding to a TfR1, respectively. A simple linear regression model would have one independent variable (x) and one dependent variable (y). A representation of an example mathematical relationship of a simple linear regression model would be $y = mx + b$. In this example, the machine learning algorithm tries

variations of the tuning variables m and b to optimize a line that includes all the given training data.

[0426] The tuning variables can be optimized, for example, with a cost function. A cost function takes advantage of the minimization problem to identify the optimal tuning variables. The minimization problem preposes the optimal tuning variable will minimize the error between the predicted outcome and the actual outcome. An example cost function may comprise summing all the square differences between the predicted and actual output values and dividing them by the total number of input values and results in the average square error.

[0427] To select new tuning variables to reduce the cost function, the machine learning module may use, for example, gradient descent methods. An example gradient descent method comprises evaluating the partial derivative of the cost function with respect to the tuning variables. The sign and magnitude of the partial derivatives indicate whether the choice of a new tuning variable value will reduce the cost function, thereby optimizing the linear regression algorithm. A new tuning variable value is selected depending on a set threshold. Depending on the machine learning module, a steep or gradual negative slope is selected. Both the cost function and gradient descent can be used with other algorithms and modules mentioned throughout. For the sake of brevity, both the cost function and gradient descent are well known in the art and are applicable to other machine learning algorithms and may not be mentioned with the same detail.

[0428] LiR models may have many levels of complexity comprising one or more independent variables. Furthermore, in an LiR function with more than one independent variable, each independent variable may have the same one or more tuning variables or each, separately, may have their own one or more tuning variables. The number of independent variables and tuning variables will be understood to one skilled in the art for the problem being solved. In an embodiment, a transferrin receptor sequence is used as the independent variables to train a LiR machine learning module, which, after training, is used to estimate, for example, one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

Logistic Regression (LoR)

[0429] In one example embodiment, logistic regression machine learning is implemented. Logistic Regression, often considered a LiR type model, is typically used in machine learning to classify information, such as a transferrin receptor sequence into categories such as one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1. LoR takes

advantage of probability to predict an outcome from input data. However, what makes LoR different from a LiR is that LoR uses a more complex logistic function, for example a sigmoid function. In addition, the cost function can be a sigmoid function limited to a result between 0 and 1. For example, the sigmoid function can be of the form $f(x)=1/(1+e^{-x})$, where x represents some linear representation of input features and tuning variables. Similar to LiR, the tuning variable(s) of the cost function are optimized (typically by taking the log of some variation of the cost function) such that the result of the cost function, given variable representations of the input features, is a number between 0 and 1, preferably falling on either side of 0.5. As described in LiR, gradient descent may also be used in LoR cost function optimization and is an example of the process. In an embodiment, a transferrin receptor sequence is used as the independent variables to train a LoR machine learning module, which, after training, is used to estimate, for example, one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

Random Forest

[0430] In one example embodiment, random forest (RF) is implemented. RF consists of an ensemble of decision trees producing individual class predictions. The prevailing prediction from the ensemble of decision trees becomes the RF prediction. Decision trees are branching flowchart-like graphs comprising of the root, nodes, edges/branches, and leaves. The root is the first decision node from which feature information is assessed and from it extends the first set of edges/branches. The edges/branches contain the information of the outcome of a node and pass the information to the next node. The leaf nodes are the terminal nodes that output the prediction. Decision trees can be used for both classification as well as regression and is typically trained using supervised learning methods. Training of a decision tree is sensitive to the training data set. An individual decision tree may become over or under-fit to the training data and result in a poor predictive model. Random forest compensates by using multiple decision trees trained on different data sets. In an embodiment, a transferrin receptor sequence is used to train the nodes of the decision trees of a RF machine learning module, which, after training, is used to estimate one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

Gradient Boosting

In an example embodiment, gradient boosting is implemented. Gradient boosting is a method of strengthening the evaluation capability of a decision tree node. In general, a tree is fit on a modified version of an original data set. For example, a decision tree is first trained with equal weights

across its nodes. The decision tree is allowed to evaluate data to identify nodes that are less accurate. Another tree is added to the model and the weights of the corresponding underperforming nodes are then modified in the new tree to improve their accuracy. This process is performed iteratively until the accuracy of the model has reached a defined threshold or a defined limit of trees has been reached. Less accurate nodes are identified by the gradient of a loss function. Loss functions must be differentiable such as a linear or logarithmic functions. The modified node weights in the new tree are selected to minimize the gradient of the loss function. In an example embodiment, a decision tree is implemented to determine one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1 and gradient boosting is applied to the tree to improve its ability to accurately determine the one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

Neural Networks

[0431] In one example embodiment, Neural Networks are implemented. NNs are a family of statistical learning models influenced by biological neural networks of the brain. NNs can be trained on a relatively-large dataset (e.g., 50,000 or more) and used to estimate, approximate, or predict an output that depends on a large number of inputs/features. NNs can be envisioned as so-called “neuromorphic” systems of interconnected processor elements, or “neurons”, and exchange electronic signals, or “messages”. Similar to the so-called “plasticity” of synaptic neurotransmitter connections that carry messages between biological neurons, the connections in NNs that carry electronic “messages” between “neurons” are provided with numeric weights that correspond to the strength or weakness of a given connection. The weights can be tuned based on experience, making NNs adaptive to inputs and capable of learning. For example, an NN for one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1 is defined by a set of input neurons that can be given input data such as a transferrin receptor sequence. The input neuron weighs and transforms the input data and passes the result to other neurons, often referred to as “hidden” neurons. This is repeated until an output neuron is activated. The activated output neuron produces a result. In an embodiment, a transferrin receptor sequence is used to train the neurons in a NN machine learning module, which, after training, is used to estimate one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

Deep Learning

[0432] In an embodiment, deep learning is implemented. Deep learning expands the neural network by including more layers of neurons. A deep learning module is characterized as having three “macro” layers: (1) an input layer which takes in the input features, and fetches embeddings for the input, (2) one or more intermediate (or hidden) layers which introduces nonlinear neural net transformations to the inputs, and (3) a response layer which transforms the final results of the intermediate layers to the prediction. In an embodiment, a transferrin receptor sequence is used to train the neurons of a deep learning module, which, after training, is used to estimate one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

Matrix Factorization

[0433] In an embodiment, Matrix Factorization is implemented. Matrix factorization machine learning exploits inherent relationships between two entities drawn out when multiplied together. Generally, the input features are mapped to a matrix F which is multiplied with a matrix R containing the relationship between the features and a predicted outcome. The resulting dot product provides the prediction. The matrix R is constructed by assigning random values throughout the matrix. In this example, two training matrices are assembled. The first matrix X contains training input features and the second matrix Z contains the known output of the training input features. First the dot product of R and X are computed and the square mean error, as one example method, of the result is estimated. The values in R are modulated and the process is repeated in a gradient descent style approach until the error is appropriately minimized. The trained matrix R is then used in the machine learning model. In an embodiment, a transferrin receptor sequence are used to train the relationship matrix R in a matrix factorization machine learning module. After training, the relationship matrix R and input matrix F , which comprises vector representations of a transferrin receptor sequence, results in the prediction matrix P comprising one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

Support Vector Machine

[0434] In an embodiment, support vector machines are implemented. SVMs separate data into classes defined by n-dimensional hyperplanes (n-hyperplane) and are used in both regression and classification problems. Hyperplanes are decision boundaries developed during the training process of a SVM. The dimensionality of a hyperplane depends on the number of input features. For example, a SVM with two input features will have a linear (1-dimensional) hyperplane while

a SVM with three input features will have a planer (2-dimensional) hyperplane. A hyperplane is optimized to have the largest margin or spatial distance from the nearest data point for each data type. In the case of simple linear regression and classification a linear equation is used to develop the hyperplane. However, when the features are more complex a kernel is used to describe the hyperplane. A kernel is a function that transforms the input features into higher dimensional space. Kernel functions can be linear, polynomial, a radial distribution function (or gaussian radial distribution function), or sigmoidal. In an embodiment, a transferrin receptor sequence is used to train the linear equation or kernel function of the SVM machine learning module, which, after training, is used to estimate one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

K-means clustering

[0435] In one example embodiment, K-means clustering is implemented. KMC assumes data points have implicit shared characteristics and “clusters” data within a centroid or “mean” of the clustered data points. During training, KMC adds a number of k centroids and optimizes its position around clusters. This process is iterative, where each centroid, initially positioned at random, is re-positioned towards the average point of a cluster. This process concludes when the centroids have reached an optimal position within a cluster. Training of a KMC module is typically unsupervised. In an embodiment, a transferrin receptor sequence is used to train the centroids of a KMC machine learning module, which, after training, is used to estimate one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

K-nearest neighbor

[0436] In one example embodiment, K-nearest neighbor is implemented. On a general level, KNN shares similar characteristics to KMC. For example, KNN assumes data points near each other share similar characteristics and computes the distance between data points to identify those similar characteristics but instead of k centroids, KNN uses k number of neighbors. The k in KNN represents how many neighbors will assign a data point to a class, for classification, or object property value, for regression. Selection of an appropriate number of k is integral to the accuracy of KNN. For example, a large k may reduce random error associated with variance in the data but increase error by ignoring small but significant differences in the data. Therefore, a careful choice of k is selected to balance overfitting and underfitting. Concluding whether some data point belongs to some class or property value k , the distance between neighbors is computed. Common

methods to compute this distance are Euclidean, Manhattan or Hamming to name a few. In some embodiments, neighbors are given weights depending on the neighbor distance to scale the similarity between neighbors to reduce the error of edge neighbors of one class “out-voting” near neighbors of another class. In one example embodiment, k is 1 and a Markov model approach is utilized. In an embodiment, a transferrin receptor sequence is used to train a KNN machine learning module, which, after training, is used to estimate one or more AAV capsid polypeptide sequences comprising a modification for binding to a TfR1.

[0437] To perform one or more of its functionalities, the machine learning module may communicate with one or more other systems. For example, an integration system may integrate the machine learning module with one or more email servers, web servers, one or more databases, or other servers, systems, or repositories. In addition, one or more functionalities may require communication between a user and the machine learning module.

[0438] Any one or more of the modules described herein may be implemented using hardware (e.g., one or more processors of a computer/machine) or a combination of hardware and software. For example, any module described herein may configure a hardware processor (e.g., among one or more hardware processors of a machine) to perform the operations described herein for that module. In some example embodiments, any one or more of the modules described herein may comprise one or more hardware processors and may be configured to perform the operations described herein. In embodiments, one or more hardware processors are configured to include any one or more of the modules described herein.

[0439] Moreover, any two or more of these modules may be combined into a single module, and the functions described herein for a single module may be subdivided among multiple modules. Furthermore, according to various example embodiments, modules described herein as being implemented within a single machine, database, or device may be distributed across multiple machines, databases, or devices. The multiple machines, databases, or devices are communicatively coupled to enable communications between the multiple machines, databases, or devices. The modules themselves are communicatively coupled (e.g., via appropriate interfaces) to each other and to various data sources, to allow information to be passed between the applications so as to allow the applications to share and access common data.

Floating-point Operations Per Second (FLOPS)

[0440] FLOPS are a measure of computation speed. The base level for computing is giga-FLOPS, which is 1 billion (10^9) floating-point operations per second. This would be equivalent to a person with pen and paper or a calculator constantly performing one (1) calculation every one (1) second for approximately thirty-two years (~32) years. As computing floating-point numbers is necessary in fields such as financial applications, scientific applications, visual rendering, and real-time processing, this unit of measurement may be used to describe the methods, systems, and products described herein. As such, and further described herein, the methods, systems, and products are carried out on time scales larger than a couple seconds. Accordingly, these methods, systems, and products are required to be carried out computationally because no person in a life-time could carry out these methods.

[0441] Further embodiments are illustrated in the following Examples which are given for illustrative purposes only and are not intended to limit the scope of the invention.

EXAMPLES**Example 1 – An AAV capsid reprogrammed to bind human Transferrin Receptor mediates brain-wide gene delivery****Main Text:**

[0442] A critical challenge in the development of more efficient delivery vectors for gene therapy is engineering vectors with known mechanisms of action (MOAs) present in human patients. Conventional approaches to engineering vectors such as adeno-associated viruses (AAVs) with enhanced tropisms have largely relied on capsid library selections in animals. These selections have been used by numerous groups over the past two decades because they can be successful without requiring prior knowledge of an MOA (1–7). However, vectors selected in animals mostly have not translated across preclinical models. Despite extensive searches, AAV capsids with clear translational potential and known MOAs for entering the CNS have not yet emerged. Here, Applicants approached this challenge from a different perspective and selected AAV capsids first for a specific MOA, that is binding to the human Transferrin Receptor (TfR1), and showed that these capsids crossed the blood-brain barrier (BBB) and mediated efficient gene delivery throughout the CNS.

[0443] We chose TfR1 as a target because of its high expression on the human BBB; its ability to mediate constitutive, ligand-independent receptor-mediated transcytosis (RMT) across the CNS

vasculature (8–12); and its track record as a target to increase the delivery of biologics into the CNS of mice (13–15), nonhuman primates (NHPs) (16–18), and humans as investigational therapies (19–22) and an approved antibody-based therapeutic for Mucopolysaccharidosis type II (23). We first screened 7-mer-modified AAV9 capsid libraries for their ability to bind to human TfR1 *in vitro*. The top-performing capsid, AAV-BI-hTfR1, exhibited more efficient gene delivery to human brain endothelial cells and improved active transport across a human vascular monolayer. When systemically administered to adult *TFRC* knock-in (KI) mice carrying a chimeric *Tfrc* gene with a humanized extracellular domain, BI-hTfR1 transduced the majority of neurons and astrocytes across multiple brain regions. The enhanced tropism was entirely dependent on an interaction with the humanized TfR1, as Applicants observed no enhancement in wild type (WT) mice. The tropism was also selectively enhanced in the CNS, consistent with the high level of expression of *TFRC* on the CNS vasculature relative to the vasculature in other organs (12).

[0444] To explore its ability to deliver a therapeutically relevant cargo, Applicants intravenously injected *TFRC* KI mice with BI-hTfR1 or AAV9 packaging the human Glucosylceramidase Beta 1 (*GBA1*) gene. Inactivating mutations of *GBA1* are the primary cause of Gaucher's disease, a lysosomal storage disorder often affecting the CNS, and *GBA1* mutations are a genetic risk factor for Parkinson's disease and dementia with Lewy Bodies (24). Systemic delivery of BI-hTfR1:*GBA1* in adult *TFRC* KI mice resulted in robust expression throughout the brain and increased glucocerebrosidase (GCase) activity in the brain and cerebrospinal fluid (CSF), which were all not observed with AAV9. Through its direct engagement of TfR1, a transporter of natural protein ligands as well as therapeutic biologics in humans, BI-hTfR1 represents a promising vector for the development of CNS-targeting human gene therapies.

Results

AAV capsids engineered to bind human TfR1

[0445] To target AAV capsids to TfR1, Applicants screened AAV9-based NNK capsid libraries of variants with random 7-mer insertions between VP1 residues 588–589 for selective binding to human TfR1 using recently described receptor-targeting approach (25) (Fig. 1A). Based on the results of the *in vitro* screens, Applicants chose four capsids that share a common sequence motif for validation: BI-hTfR1, BI-hTfR1-2, BI-hTfR1-3, and BI-hTfR1-4. Applicants used each capsid to individually package a single-stranded genome that expresses nuclear mScarlet and

Luciferase under the control of the ubiquitous CAG promoter (ssAAV-CAG-NLS-mScarlet-P2A-Luciferase-WPRE-SV40pA). The four capsids exhibited elevated association with and transduction of CHO cells that stably express human TfR1 but not the control CHO cells or those that express TfR1 from rhesus macaque, marmoset, or mouse (Fig. 1B and C, and fig. S1). We next tested whether the capsids were capable of increased association with and transduction of human brain endothelial cells that endogenously express *TFRC*. Although the extent of AAV associated with primary human brain microvascular endothelial cells (hBMVECs) and a human brain endothelial cell line (hCMEC/D3) varied (Fig. 1D), each of the four TfR1-binding capsids exhibited significantly enhanced transduction of both brain endothelial cell types compared to AAV9 (Fig. 1E). To further assess the relative binding affinities of the capsids with purified human TfR1, Applicants performed bio-layer interferometry (BLI). We derived kinetic constants using a multi-phasic binding model because a monovalent interaction algorithm led to a poor fit of the data. Factors that may contribute to the multi-phasic binding curves are outlined in the Materials and Methods section. Based on its performance across cell association and transduction assays and its ability to bind human TfR1 at lower concentrations, Applicants chose BI-hTFR1 for further investigation in the rest of this study.

BI-hTFR1 binds the TfR1 apical domain

[0446] To evaluate whether binding to TfR1 is necessary for the increased transduction of brain endothelial cells by BI-hTFR1, Applicants assessed hCMEC/D3 transduction in the presence of two anti-TfR1 antibodies: OKT9, a monoclonal antibody that binds the TfR1 apical domain (26) or AF2474, a polyclonal antibody that competes with Transferrin (Tf) for TfR1 binding. We found that transduction was significantly inhibited by increasing concentrations of OKT9 but not by AF2474 (Fig. 2A), suggesting that an interaction with the human TfR1 apical domain was required for the enhanced transduction mediated by BI-hTFR1 in these cells. Consistent with this result, OKT9 directly competed with BI-hTFR1 for binding to purified human TfR1 (fig. S2B). Notably, binding to the TfR1 apical domain is a common feature of several antibody-based BBB shuttles that have been tested in humans and likely positions the binding site away from the Tf binding site. This is important because surface TfR1 is thought to be mostly occupied by Tf based on the typical 10–20 μ M blood concentration of iron-bound holo-Tf (27) and its sub nanomolar binding affinity for TfR1 (fig. S2C). Given the double digit nanomolar affinity of BI-hTFR1 for human TfR1, Tf would be predicted to outcompete BI-hTFR1 for binding to TfR1 if their binding

epitopes overlap. Using BLI, Applicants found that BI-hTFR1 binds human TfR1 with similar kinetics whether in the presence or absence of a receptor-saturating concentration of holo-Tf (Fig. 2B and fig. S2D) and that BI-hTFR1 is capable of binding a preformed complex of human TfR1 and Tf (Fig. 2C). Consistent with the BLI results, OKT9, but not Tf, inhibited the association of BI-hTFR1 with hCMEC/D3 cells and its colocalization with human TfR1 (Fig. 2D and fig. S3).

Active trafficking of BI-hTFR1 across a human BBB model

[0447] To test AAV transport across a human endothelial cell barrier, Applicants established a BBB transwell model using the hCMEC/D3 cell line and compared the amount of BI-hTFR1 that was transported across the endothelial barrier relative to AAV9 and AAV2. To minimize the effect of well-to-well variation, Applicants mixed BI-hTFR1, AAV9, and AAV2 carrying barcoded genomes that were individually identifiable by qPCR (table S1), added this mixture to the media of the upper chambers which were maintained at 37°C or 4°C, and measured transport to the lower chamber three hours later (Fig. 3A). At 4°C, ATP-dependent transcytosis is suppressed but passive transport, at least in part due to paracellular crossing, can occur, providing a measurement of leakiness that is a known issue with the use of hCMEC/D3 transwell assays (28). As previously reported, AAV9 crossed the transwell barrier more efficiently than AAV2 at 37°C, and the number of vector genomes detected in the bottom chamber was significantly reduced at 4°C for both AAVs as is consistent with ATP-dependent transcytosis (29–31). Significantly more BI-hTFR1 was actively transported to the bottom chamber than AAV9 or AAV2 (Fig. 3B). Passive transport of BI-hTFR1 and AAV2 were reduced relative to AAV9, likely due to increased association with the cells.

[0448] RMT involves endocytosis, vesicular trafficking, and exocytosis (32). We explored whether BI-hTFR1 transcytoses hCMEC/D3 cells via the RMT pathway of TfR1 by assessing its colocalization with various organelle markers. We found that BI-hTFR1 partially colocalized with markers of the early and late endosomal pathway (Rab5 and Rab7) as well as the trans-Golgi network (TGN46), but did not extensively colocalize with markers of the cis-Golgi (Rcas1) or endoplasmic reticulum (KDEL (SEQ ID NO: 1787)) (Fig. 3C and fig. S4). In contrast, AAV2 was not highly colocalized with any of the endosomal or organelle markers with the exception of TGN46. Partial colocalization of both AAVs with TGN46 is consistent with the requirement for trafficking of AAVs to the trans-Golgi for transduction (33). Cell surface TfR1 clustering promotes clathrin-coated pit formation (34) and uptake into a subpopulation of endosomes enriched in Rab5

(35). This is consistent with the observation that BI-hTFR1 colocalized with Rab5 to a greater degree than AAV2. A smaller proportion of Tfr1 colocalizes with Rab7 (36), which decorates endosomes associated with transcytosis (37) and the lysosomal degradation pathway. This is again consistent with the observation of a partial overlap between BI-hTFR1 and Rab7 localization in the cells. Our colocalization assays suggest that BI-hTFR1 engages Tfr1 and is actively transported across the human BBB model via RMT. In contrast, AAV2 is predominantly trafficked to the trans-Golgi (33).

CNS tropism of BI-hTFR1

[0449] Having established that BI-hTFR1 is more effective at transducing human brain endothelial cells via its interaction with Tfr1 and is more efficiently transported across a human BBB model, Applicants next sought to investigate whether BI-hTFR1 could cross the BBB *in vivo* by engaging human Tfr1. To test this, Applicants used *TFRC* knock-in (KI) C57BL/6J mice in which exons 4–19 of the mouse *Tfrc* encoding the extracellular domain were replaced by the corresponding region of human *TFRC* (Fig. 4A) (38). Levels of the mouse-human hybrid gene mRNA and protein products were found to be similar to those in WT mice (fig. S5). Therefore, Applicants deemed these mice suitable for assessing the tropism of AAVs targeted to human Tfr1. We intravenously injected BI-hTFR1 or AAV9:CAG-NLS-mScarlet-P2A-Luciferase-WPRE-SV40pA at 5×10^{11} vg/mouse into adult female C57BL/6J or *TFRC* KI mice (Fig. 4B). Three weeks post-injection, Applicants observed enhanced biodistribution to and transduction of the brain and spinal cord by BI-hTFR1 in the *TFRC* KI but not in control C57BL/6J mice (Fig. 4C and D, and fig. S6). Increased biodistribution and transduction were not observed in any of the other organs that were assessed, indicating that the interaction with human Tfr1 selectively enhanced the CNS tropism. Comparing the *TFRC* KI mice that received BI-hTFR1 to the C57BL/6J mice that received AAV9, Applicants detected 6-fold and 12-fold more viral genomes; 54-fold and 43-fold higher viral mRNA transcript levels; 132-fold and 58-fold greater Luciferase enzymatic activity in the brain and spinal cord, respectively. As expected, AAV9 transduction was not significantly affected by genotype in any of the assessed organs.

[0450] To evaluate the efficiency of transduction of specific cell types within the CNS, Applicants assessed the fraction of cells positive for markers of neurons or astrocytes that expressed mScarlet. In *TFRC* KI mice, BI-hTFR1 transduced a mean of 54, 32, and 68% of NeuN⁺ neurons in the cortex, striatum, and thalamus, respectively (Fig. 5C and D). In contrast, in

C57BL/6J mice, BI-hTFR1 transduced less than 1% of NeuN⁺ cells in the cortex and striatum and thalamus. AAV9 transduced 1–6% of NeuN⁺ cells in these same regions regardless of the mouse strain. BI-hTFR1 transduced more than 80% of the SOX9⁺ astrocytes in the cortex, striatum, and thalamus of *TFRC* KI mice, in comparison with 10% or fewer SOX9⁺ cells in C57BL/6J mice, or 20% or fewer SOX9⁺ cells transduced by AAV9 in C57BL/6J mice (Fig. 5E and F). Consistent with the biodistribution and transduction data, mScarlet levels in the livers of C57BL/6J and *TFRC* KI mice were comparable between AAV9 and BI-hTFR1 (fig. S6A and B). We observed a significant reduction of biodistribution to the DRGs in *TFRC* KI animals treated with BI-hTFR1 relative to AAV9 controls (Fig. 4C); DRG cell transduction by AAV9 or BI-hTFR1 in C57BL/6J or *TFRC* KI mice is shown in fig. S6C.

BI-hTFR1 mediates GBA expression throughout the brain

[0451] We next evaluated BI-hTFR1's ability to deliver the human gene Glucosylceramidase Beta 1 (*GBA1*) to *TFRC* KI mice. We built a single-stranded AAV genome containing a hybrid cytomegalovirus enhancer-chicken beta-Actin (CMV-CBA) promoter driving the expression of human *GBA1* with a C-terminal influenza virus hemagglutinin (HA) tag (Fig. 6A) (39). The genome was packaged into AAV9 or BI-hTFR1 and administered intravenously to *TFRC* KI mice at a dose of 1×10^{14} vg/kg, which is comparable to the weight-adjusted dose of the FDA-approved AAV9-based Zolgensma gene therapy for spinal muscular atrophy. Additionally, Applicants administered BI-hTFR1:*GBA1* to a separate group of *TFRC* KI mice at a 20-fold lower dose (5×10^{12} vg/kg). For the untreated control mice, Applicants used C57BL/6J mice, which possess the same genetic background of the *TFRC* KI mice. Three weeks post-administration, Applicants detected ~30 times more AAV genomes in the brains of mice treated with BI-hTFR1 as compared with those of mice that received AAV9 (Fig. 6B). We observed robust brain-wide expression of GBA-HA as assessed by HA immunostaining (Fig. 6C) and significantly elevated glucocerebrosidase (GCase) activity in brain tissue lysates and the CSF, compared to the mice injected with the high dose of AAV9 or control C57BL/6J mice (Fig. 6D, and fig. S7A and B). In the mice injected with the high dose of AAV9, Applicants observed minimal GBA expression in the brain but similar amounts of viral genomes, GBA expression, and GCase activity in the liver compared to the mice injected with the high dose of BI-hTFR1 (Fig. 6B and 6D, and fig. S7C); as expected, in the mice injected with the low dose of BI-hTFR1, the amount of viral genomes in the liver was reduced relative to the high dose AAV9 group. Reductions in GCase activity in the brain

and CSF have been reported to be associated with the progression of Parkinson's disease (40, 41), and previous studies suggest that *GBA1* delivery to the brain can treat neuronal α -synucleinopathy (42–44). Notably, in the mice that received the low dose of BI-hTFR1, Applicants observed neuronal transduction in regions linked to motor symptoms of Parkinson's disease such as the substantia nigra pars compacta, deep cerebellar nuclei, and other parts of the brainstem (Fig. 6C and Data fig. S7A) (45). Under the same imaging conditions, GBA-HA was not detected in the brains of the mice injected with the high dose of AAV9. This suggests that a 20-fold lower dose of BI-hTFR1 led to greater GBA levels in disease-relevant brain regions compared to AAV9.

Discussion

[0452] In this study, Applicants leveraged a well established mechanism—engagement of the apical domain of human TfR1—to shuttle an engineered AAV capsid, BI-hTFR1, across the BBB. Protein shuttles designed to bind human TfR1 to facilitate RMT are now in the clinic and are showing evidence of CNS entry (NCT04023994) and efficacy (46). The TfR1 apical domain is also a receptor target of natural viruses (47–51), indicating that this iron transporter can be exploited for viral entry. Importantly, Applicants demonstrated that BI-hTFR1 binds human TfR1 and pre-formed TfR1-Tf complexes with comparable kinetics and does not compete with holo-Tf for TfR1 binding. The ability to bind TfR1 in the presence of its natural ligand is critical because AAVs cannot be administered at doses capable of competing with Tf for TfR1 binding. The use of this validated transport mechanism bolsters Applicants confidence that BI-hTFR1 can perform well as a CNS-targeting human gene therapy vehicle.

[0453] We engineered BI-hTFR1 to bind TfR1 by inserting a novel 7-mer TfR1 recognition sequence directly within surface loop VIII of the capsid. Because the approach did not require the introduction or conjugation of bulky protein domains as a means to retarget the AAV (52–58), Applicants do not anticipate additional complexities in the manufacturing of BI-hTFR1 compared to AAV9, one of the top producing natural AAVs. Indeed, Applicants consistently produced and purified BI-hTFR1 with final yields within twofold of AAV9 (table S2).

[0454] As compared to natural AAV capsids such as AAV9 and AAV2 that are used in approved gene therapies, BI-hTFR1 demonstrated the remarkable ability to be actively transported across a human brain endothelial cell layer and to cross the BBB and transduce cells throughout the brains and spinal cords of *TFR1* KI mice after intravenous delivery. BI-hTFR1 exhibited a broad CNS tropism enhancement similar to those observed with engineered mouse BBB-crossing

capsids, e.g., AAV-PHP.B, BI28, AAVF, and 9P31 that bind to mouse GPI-anchored proteins LY6A, LY6C1, or Car4 (25, 59, 60). By engaging highly abundant CNS endothelial cell surface proteins as new receptors, these capsids are able to efficiently cross the BBB and gain access to parenchymal cells.

[0455] A critical advantage of BI-hTFR1 and other capsids developed through target receptor binding screens is that their MOA is well established, thus their species tropism is predictable and can inform large animal experiments. This stands in contrast to capsids identified from conventional selections with unknown MOAs and therefore unknown potential to translate to other species including humans. We do not expect the human TfR1-targeting AAVs reported in this study to exhibit enhanced CNS tropisms in NHPs such as macaques and marmosets. Therefore, tests in these species would only be valuable in terms of profiling biodistribution, immune responses, and toxicity in the absence of the target, human TfR1. Knowing the precise target of the AAV capsid also aids the prediction of its function in human patients. For instance, there are no known missense variants with an allele frequency greater than 1 in a 1000 within the *TFRC* apical domain (61). Therefore, Applicants predict that BI-hTFR1 binding to human TfR1 will not be impacted by common variations within the coding sequence of the human apical domain.

[0456] A second advantage of human receptor-targeting capsids is that gene therapies leveraging these vectors can be tested in KI mouse models possessing both the human target receptor and disease-relevant mutation(s). For instance, disease models are typically created using mice and, in the absence of a capsid with an enhanced CNS tropism that translates across species, studies will need to use a surrogate mouse BBB-crossing capsid to deliver and test their potentially therapeutic payloads in these disease models. With BI-hTFR1, the gene therapy product that would be delivered to human patients can be tested, without switching out the capsid for a surrogate, in a mouse disease model that has been crossed with *TFRC* KI mice. Using a therapeutically relevant payload, Applicants found that BI-hTFR1 could deliver 30 times the number of AAV:*GBAI* genomes to the brains of *TFRC* KI mice and increase brain and CSF GCCase activity as compared to AAV9. Even at a 20-fold lower dose, *GBAI* expression was detectable throughout the brain in BI-hTFR1:*GBAI* treated mice. However, the *TFRC* KI mice were WT with respect to *Gba1* and only the high dose of BI-hTFR1:*GBAI* led to increased brain and CSF GCCase activity that were detectable above the endogenous mouse GCCase activity.

[0457] Collectively, the results demonstrate that BI-hTfR1 can leverage the RMT activity of human TfR1 to cross the BBB and deliver genes widely and efficiently to neurons and glia throughout the CNS. These qualities provide strong support for the further exploration of BI-hTfR1 as a broadly applicable vehicle for human CNS gene therapies.

References and Notes

1. O. J. Müller, F. Kaul, M. D. Weitzman, R. Pasqualini, W. Arap, J. A. Kleinschmidt, M. Trepel, Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. *Nat. Biotechnol.* **21**, 1040–1046 (2003).
2. J. T. Koerber, N. Maheshri, B. K. Kaspar, D. V. Schaffer, Construction of diverse adeno-associated viral libraries for directed evolution of enhanced gene delivery vehicles. *Nat. Protoc.* **1**, 701–706 (2006).
3. W. Li, A. Asokan, Z. Wu, T. Van Dyke, N. DiPrimio, J. S. Johnson, L. Govindaswamy, M. Agbandje-McKenna, S. Leichtle, D. Eugene Redmond Jr, T. J. McCown, K. B. Petermann, N. E. Sharpless, R. J. Samulski, Engineering and Selection of Shuffled AAV Genomes: A New Strategy for Producing Targeted Biological Nanoparticles. *Mol. Ther.* **16**, 1252–1260 (2008).
4. Y. H. Chen, M. Chang, B. L. Davidson, Molecular signatures of disease brain endothelia provide new sites for CNS-directed enzyme therapy. *Nat. Med.* **15**, 1215–1218 (2009).
5. B. E. Deverman, P. L. Pravdo, B. P. Simpson, S. R. Kumar, K. Y. Chan, A. Banerjee, W.-L. Wu, B. Yang, N. Huber, S. P. Pasca, V. Gradinaru, Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nat. Biotechnol.* **34**, 204–209 (2016).
6. M. Nonnenmacher, W. Wang, M. A. Child, X.-Q. Ren, C. Huang, A. Z. Ren, J. Tocci, Q. Chen, K. Bittner, K. Tyson, N. Pande, C. H.-Y. Chung, S. M. Paul, J. Hou, Rapid evolution of blood-brain-barrier-penetrating AAV capsids by RNA-driven biopanning. *Mol Ther Methods Clin Dev.* **20**, 366–378 (2021).
7. K. S. Hanlon, J. C. Meltzer, T. Buzhdygan, M. J. Cheng, M. Sena-Esteves, R. E. Bennett, T. P. Sullivan, R. Razmpour, Y. Gong, C. Ng, J. Nammour, D. Maiz, S. Dujardin, S. H. Ramirez, E. Hudry, C. A. Maguire, Selection of an Efficient AAV Vector for Robust CNS Transgene Expression. *Mol Ther Methods Clin Dev.* **15**, 320–332 (2019).
8. R. S. Ajioka, J. Kaplan, Intracellular pools of transferrin receptors result from constitutive internalization of unoccupied receptors. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 6445–6449 (1986).

9. F. R. Maxfield, T. E. McGraw, Endocytic recycling. *Nat. Rev. Mol. Cell Biol.* **5**, 121–132 (2004).
10. T. Moos, E. H. Morgan, Transferrin and transferrin receptor function in brain barrier systems. *Cell. Mol. Neurobiol.* **20**, 77–95 (2000).
11. W. M. Pardridge, J. Eisenberg, J. Yang, Human blood-brain barrier transferrin receptor. *Metabolism.* **36**, 892–895 (1987).
12. W. A. Jefferies, M. R. Brandon, S. V. Hunt, A. F. Williams, K. C. Gatter, D. Y. Mason, Transferrin receptor on endothelium of brain capillaries. *Nature.* **312**, 162–163 (1984).
13. S. Kida, Y. Koshimura, E. Yoden, A. Yoshioka, H. Morimoto, A. Imakiire, N. Tanaka, S. Tanaka, A. Mori, J. Ito, A. Inoue, R. Yamamoto, K. Minami, T. Hirato, K. Takahashi, H. Sonoda, Enzyme replacement with transferrin receptor-targeted α -L-iduronidase rescues brain pathology in mucopolysaccharidosis I mice. *Mol Ther Methods Clin Dev.* **29**, 439–449 (2023).
14. J. C. Ullman, A. Arguello, J. A. Getz, A. Bhalla, C. S. Mahon, J. Wang, T. Giese, C. Bedard, D. J. Kim, J. R. Blumenfeld, N. Liang, R. Ravi, A. A. Nugent, S. S. Davis, C. Ha, J. Duque, H. L. Tran, R. C. Wells, S. Lianoglou, V. M. Daryani, W. Kwan, H. Solanoy, H. Nguyen, T. Earr, J. C. Dugas, M. D. Tuck, J. L. Harvey, M. L. Reyzer, R. M. Caprioli, S. Hall, S. Poda, P. E. Sanchez, M. S. Dennis, K. Gunasekaran, A. Srivastava, T. Sandmann, K. R. Henne, R. G. Thorne, G. Di Paolo, G. Astarita, D. Diaz, A. P. Silverman, R. J. Watts, Z. K. Sweeney, M. S. Kariolis, A. G. Henry, Brain delivery and activity of a lysosomal enzyme using a blood-brain barrier transport vehicle in mice. *Sci. Transl. Med.* **12** (2020), doi:10.1126/scitranslmed.aay1163.
15. S. M. Hammond, F. Abendroth, L. Goli, J. Stoodley, M. Burrell, G. Thom, I. Gurrell, N. Ahlskog, M. J. Gait, M. J. Wood, C. I. Webster, Antibody-oligonucleotide conjugate achieves CNS delivery in animal models for spinal muscular atrophy. *JCI Insight.* **7** (2022), doi:10.1172/jci.insight.154142.
16. M. S. Kariolis, R. C. Wells, J. A. Getz, W. Kwan, C. S. Mahon, R. Tong, D. J. Kim, A. Srivastava, C. Bedard, K. R. Henne, T. Giese, V. A. Assimon, X. Chen, Y. Zhang, H. Solanoy, K. Jenkins, P. E. Sanchez, L. Kane, T. Miyamoto, K. S. Chew, M. E. Pizzo, N. Liang, M. E. K. Calvert, S. L. DeVos, S. Baskaran, S. Hall, Z. K. Sweeney, R. G. Thorne, R. J. Watts, M. S. Dennis, A. P. Silverman, Y. J. Y. Zuchero, Brain delivery of therapeutic proteins using an Fc fragment blood-brain barrier transport vehicle in mice and monkeys. *Sci. Transl. Med.* **12** (2020), doi:10.1126/scitranslmed.aay1359.

17. L. Rué, T. Jaspers, I. M. S. Degors, S. Noppen, D. Schols, B. De Strooper, M. Dewilde, Novel Human/Non-Human Primate Cross-Reactive Anti-Transferrin Receptor Nanobodies for Brain Delivery of Biologics. *Pharmaceutics*. **15** (2023), doi:10.3390/pharmaceutics15061748.
18. Y. J. Yu, J. K. Atwal, Y. Zhang, R. K. Tong, K. R. Wildsmith, C. Tan, N. Bien-Ly, M. Hersom, J. A. Maloney, W. J. Meilandt, D. Bumbaca, K. Gadkar, K. Hoyte, W. Luk, Y. Lu, J. A. Ernst, K. Searce-Levie, J. A. Couch, M. S. Dennis, R. J. Watts, Therapeutic bispecific antibodies cross the blood-brain barrier in nonhuman primates. *Sci. Transl. Med.* **6**, 261ra154 (2014).
19. B. van Lengerich, L. Zhan, D. Xia, D. Chan, D. Joy, J. I. Park, D. Tatarakis, M. Calvert, S. Hummel, S. Lianoglou, M. E. Pizzo, R. Prorok, E. Thomsen, L. M. Bartos, P. Beumers, A. Capell, S. S. Davis, L. de Weerd, J. C. Dugas, J. Duque, T. Earr, K. Gadkar, T. Giese, A. Gill, J. Gnörich, C. Ha, M. Kannuswamy, D. J. Kim, S. T. Kunte, L. H. Kunze, D. Lac, K. Lechtenberg, A. W.-S. Leung, C.-C. Liang, I. Lopez, P. McQuade, A. Modi, V. O. Torres, H. N. Nguyen, I. Pesämaa, N. Propson, M. Reich, Y. Robles-Colmenares, K. Schlepckow, L. Slemann, H. Solanoy, J. H. Suh, R. G. Thorne, C. Vieira, K. Wind-Mark, K. Xiong, Y. J. Y. Zuchero, D. Diaz, M. S. Dennis, F. Huang, K. Searce-Levie, R. J. Watts, C. Haass, J. W. Lewcock, G. Di Paolo, M. Brendel, P. E. Sanchez, K. M. Monroe, A TREM2-activating antibody with a blood-brain barrier transport vehicle enhances microglial metabolism in Alzheimer's disease models. *Nat. Neurosci.* **26**, 416–429 (2023).
20. P. V. Candelaria, L. S. Leoh, M. L. Penichet, T. R. Daniels-Wells, Antibodies Targeting the Transferrin Receptor 1 (TfR1) as Direct Anti-cancer Agents. *Front. Immunol.* **12**, 607692 (2021).
21. S. J. Barker, M. B. Thayer, C. Kim, D. Tatarakis, M. Simon, R. L. Dial, L. Nilewski, R. C. Wells, Y. Zhou, M. Afetian, A. Chappell, K. S. Chew, J. Chow, A. Clemens, C. B. Discenza, J. Dugas, C. Dwyer, T. Earr, C. Ha, D. Huynh, S. Jayaraman, W. Kwan, C. Mahon, M. Pizzo, E. Roche, L. Sanders, A. Stergioulis, R. Tong, H. Tran, J. Zuchero, A. A. Estrada, K. Gadkar, C. M. M. Koth, P. E. Sanchez, R. G. Thorne, R. J. Watts, T. Sandmann, L. Kane, F. Rigo, M. S. Dennis, J. W. Lewcock, S. L. DeVos, Targeting Transferrin Receptor to Transport Antisense Oligonucleotides Across the Blood-Brain Barrier. *bioRxiv* (2023), p. 2023.04.25.538145.
22. R. Giugliani, A. M. Martins, S. So, T. Yamamoto, M. Yamaoka, T. Ikeda, K. Tanizawa, H. Sonoda, M. Schmidt, Y. Sato, Iduronate-2-sulfatase fused with anti-hTfR antibody, pabinafusp alfa, for MPS-II: A phase 2 trial in Brazil. *Mol. Ther.* **29**, 2378–2386 (2021).

23. T. Okuyama, Y. Eto, N. Sakai, K. Nakamura, T. Yamamoto, M. Yamaoka, T. Ikeda, S. So, K. Tanizawa, H. Sonoda, Y. Sato, A Phase 2/3 Trial of Pabinafusp Alfa, IDS Fused with Anti-Human Transferrin Receptor Antibody, Targeting Neurodegeneration in MPS-II. *Mol. Ther.* **29**, 671–679 (2021).
24. L. Smith, A. H. V. Schapira, GBA Variants and Parkinson Disease: Mechanisms and Treatments. *Cells*. **11** (2022), doi:10.3390/cells11081261.
25. Q. Huang, A. T. Chen, K. Y. Chan, H. Sorensen, A. J. Barry, B. Azari, Q. Zheng, T. Beddow, B. Zhao, I. G. Tobey, C. Moncada-Reid, F.-E. Eid, C. J. Walkey, M. C. Ljungberg, W. R. Lagor, J. D. Heaney, Y. A. Chan, B. E. Deverman, Targeting AAV vectors to the central nervous system by engineering capsid-receptor interactions that enable crossing of the blood-brain barrier. *PLoS Biol.* **21**, e3002112 (2023).
26. S. Ferrero, M. D. Flores, C. Short, C. A. Vazquez, L. E. Clark, J. Ziegenbein, S. Zink, D. Fuentes, C. Payes, M. V. Batto, M. Collazo, C. C. García, J. Abraham, S. M. Cordo, J. A. Rodriguez, G. Helguera, Antibody-Based Inhibition of Pathogenic New World Hemorrhagic Fever Mammarenaviruses by Steric Occlusion of the Human Transferrin Receptor 1 Apical Domain. *J. Virol.* **95**, e0186820 (2021).
27. J. Williams, K. Moreton, The distribution of iron between the metal-binding sites of transferrin human serum. *Biochem. J.* **185**, 483–488 (1980).
28. V. Balzer, P. Poc, E. Puris, S. Martin, M. Aliasgari, S. Auriola, G. Fricker, Re-evaluation of the hCMEC/D3 based in vitro BBB model for ABC transporter studies. *Eur. J. Pharm. Biopharm.* **173**, 12–21 (2022).
29. G. Di Pasquale, J. A. Chiorini, AAV transcytosis through barrier epithelia and endothelium. *Mol. Ther.* **13**, 506–516 (2006).
30. S. F. Merkel, A. M. Andrews, E. M. Lutton, D. Mu, E. Hudry, B. T. Hyman, C. A. Maguire, S. H. Ramirez, Trafficking of adeno-associated virus vectors across a model of the blood-brain barrier; a comparative study of transcytosis and transduction using primary human brain endothelial cells. *J. Neurochem.* **140**, 216–230 (2017).
31. R. Song, K. Pekrun, T. A. Khan, F. Zhang, S. P. Paşca, M. A. Kay, Selection of rAAV vectors that cross the human blood-brain barrier and target the central nervous system using a transwell model. *Mol Ther Methods Clin Dev.* **27**, 73–88 (2022).

32. R. Villaseñor, J. Lampe, M. Schwaninger, L. Collin, Intracellular transport and regulation of transcytosis across the blood-brain barrier. *Cell. Mol. Life Sci.* **76**, 1081–1092 (2019).
33. M. E. Nonnenmacher, J.-C. Cintrat, D. Gillet, T. Weber, Syntaxin 5-dependent retrograde transport to the trans-Golgi network is required for adeno-associated virus transduction. *J. Virol.* **89**, 1673–1687 (2015).
34. A. P. Liu, F. Aguet, G. Danuser, S. L. Schmid, Local clustering of transferrin receptors promotes clathrin-coated pit initiation. *J. Cell Biol.* **191**, 1381–1393 (2010).
35. D. M. Navaroli, K. D. Bellvé, C. Standley, L. M. Lifshitz, J. Cardia, D. Lambricht, D. Leonard, K. E. Fogarty, S. Corvera, Rabenosyn-5 defines the fate of the transferrin receptor following clathrin-mediated endocytosis. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E471–80 (2012).
36. A. S. Haqqani, C. E. Delaney, E. Brunette, E. Baumann, G. K. Farrington, W. Sisk, J. Eldredge, W. Ding, T.-L. Tremblay, D. B. Stanimirovic, Endosomal trafficking regulates receptor-mediated transcytosis of antibodies across the blood brain barrier. *J. Cereb. Blood Flow Metab.* **38**, 727–740 (2018).
37. P. Alam, M. R. Holst, L. Lauritsen, J. Nielsen, S. S. E. Nielsen, P. H. Jensen, J. R. Brewer, D. E. Otzen, M. S. Nielsen, Polarized α -synuclein trafficking and transcytosis across brain endothelial cells via Rab7-decorated carriers. *Fluids Barriers CNS.* **19**, 37 (2022).
38. C. Liu, C. Shang, L. Ma, X. Zhou, C. Guo, Abstract 377: Evaluating delivery of TFR1-targeting therapies to the CNS in a novel humanized TFR1 mouse model. *Cancer Res.* **82**, 377–377 (2022).
39. J. R. Mendell, S. Al-Zaidy, R. Shell, W. D. Arnold, L. R. Rodino-Klapac, T. W. Prior, L. Lowes, L. Alfano, K. Berry, K. Church, J. T. Kissel, S. Nagendran, J. L'Italien, D. M. Sproule, C. Wells, J. A. Cardenas, M. D. Heitzer, A. Kaspar, S. Corcoran, L. Braun, S. Likhite, C. Miranda, K. Meyer, K. D. Foust, A. H. M. Burghes, B. K. Kaspar, Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *N. Engl. J. Med.* **377**, 1713–1722 (2017).
40. C. E. G. Leyns, A. Prigent, B. Beezhold, L. Yao, N. G. Hatcher, P. Tao, J. Kang, E. Suh, V. M. Van Deerlin, J. Q. Trojanowski, V. M. Y. Lee, M. E. Kennedy, M. J. Fell, M. X. Henderson, Glucocerebrosidase activity and lipid levels are related to protein pathologies in Parkinson's disease. *NPJ Parkinsons Dis.* **9**, 74 (2023).

41. L. Oftedal, J. Maple-Grødem, I. Dalen, O.-B. Tysnes, K. F. Pedersen, G. Alves, J. Lange, Association of CSF Glucocerebrosidase Activity With the Risk of Incident Dementia in Patients With Parkinson Disease. *Neurology*. **100**, e388–e395 (2023).
42. G. Morabito, S. G. Giannelli, G. Ordazzo, S. Bido, V. Castoldi, M. Indrigo, T. Cabassi, S. Cattaneo, M. Luoni, C. Cancellieri, A. Sessa, M. Bacigaluppi, S. Taverna, L. Leocani, J. L. Lanciego, V. Broccoli, AAV-PHP.B-Mediated Global-Scale Expression in the Mouse Nervous System Enables GBA1 Gene Therapy for Wide Protection from Synucleinopathy. *Mol. Ther.* **25**, 2727–2742 (2017).
43. K. L. Jackson, C. Viel, J. Clarke, J. Bu, M. Chan, B. Wang, L. S. Shihabuddin, S. P. Sardi, Viral delivery of a microRNA to Gba to the mouse central nervous system models neuronopathic Gaucher disease. *Neurobiol. Dis.* **130**, 104513 (2019).
44. E. M. Rocha, G. A. Smith, E. Park, H. Cao, E. Brown, M. A. Hayes, J. Beagan, J. R. McLean, S. C. Izen, E. Perez-Torres, P. J. Hallett, O. Isacson, Glucocerebrosidase gene therapy prevents α -synucleinopathy of midbrain dopamine neurons. *Neurobiol. Dis.* **82**, 495–503 (2015).
45. K. Seidel, J. Mahlke, S. Siswanto, R. Krüger, H. Heinsen, G. Auburger, M. Bouzrou, L. T. Grinberg, H. Wicht, H.-W. Korf, W. den Dunnen, U. Rüb, The brainstem pathologies of Parkinson's disease and dementia with Lewy bodies. *Brain Pathol.* **25**, 121–135 (2015).
46. Agents shuttle across blood-brain barrier to treat Hunter syndrome. *Nat. Biotechnol.* **39**, 253 (2021).
47. S. R. Radoshitzky, J. Abraham, C. F. Spiropoulou, J. H. Kuhn, D. Nguyen, W. Li, J. Nagel, P. J. Schmidt, J. H. Nunberg, N. C. Andrews, M. Farzan, H. Choe, Transferrin receptor 1 is a cellular receptor for New World haemorrhagic fever arenaviruses. *Nature*. **446**, 92–96 (2007).
48. E. Wang, N. Obeng-Adjei, Q. Ying, L. Meertens, T. Dragic, R. A. Davey, S. R. Ross, Mouse mammary tumor virus uses mouse but not human transferrin receptor 1 to reach a low pH compartment and infect cells. *Virology*. **381**, 230–240 (2008).
49. M. L. Flanagan, J. Oldenburg, T. Reignier, N. Holt, G. A. Hamilton, V. K. Martin, P. M. Cannon, New world clade B arenaviruses can use transferrin receptor 1 (TfR1)-dependent and -independent entry pathways, and glycoproteins from human pathogenic strains are associated with the use of TfR1. *J. Virol.* **82**, 938–948 (2008).

50. L. B. Goodman, S. M. Lyi, N. C. Johnson, J. O. Cifuentes, S. L. Hafenstein, C. R. Parrish, Binding site on the transferrin receptor for the parvovirus capsid and effects of altered affinity on cell uptake and infection. *J. Virol.* **84**, 4969–4978 (2010).
51. J. S. Parker, W. J. Murphy, D. Wang, S. J. O'Brien, C. R. Parrish, Canine and feline parvoviruses can use human or feline transferrin receptors to bind, enter, and infect cells. *J. Virol.* **75**, 3896–3902 (2001).
52. S. Ponnazhagan, G. Mahendra, S. Kumar, J. A. Thompson, M. Castillas Jr, Conjugate-based targeting of recombinant adeno-associated virus type 2 vectors by using avidin-linked ligands. *J. Virol.* **76**, 12900–12907 (2002).
53. J. Reul, A. Muik, C. J. Buchholz, Ligand Coupling to the AAV Capsid for Cell-Specific Gene Transfer. *Methods Mol. Biol.* **1950**, 35–50 (2019).
54. R. C. Münch, H. Janicki, I. Völker, A. Rasbach, M. Hallek, H. Büning, C. J. Buchholz, Displaying high-affinity ligands on adeno-associated viral vectors enables tumor cell-specific and safe gene transfer. *Mol. Ther.* **21**, 109–118 (2013).
55. R. C. Münch, A. Muth, A. Muik, T. Friedel, J. Schmatz, B. Dreier, A. Trkola, A. Plückthun, H. Büning, C. J. Buchholz, Off-target-free gene delivery by affinity-purified receptor-targeted viral vectors. *Nat. Commun.* **6**, 6246 (2015).
56. A. Muik, J. Reul, T. Friedel, A. Muth, K. P. Hartmann, I. C. Schneider, R. C. Münch, C. J. Buchholz, Covalent coupling of high-affinity ligands to the surface of viral vector particles by protein trans-splicing mediates cell type-specific gene transfer. *Biomaterials*. **144**, 84–94 (2017).
57. A. M. Eichhoff, K. Börner, B. Albrecht, W. Schäfer, N. Baum, F. Haag, J. Körbelin, M. Trepel, I. Braren, D. Grimm, S. Adriouch, F. Koch-Nolte, Nanobody-Enhanced Targeting of AAV Gene Therapy Vectors. *Mol Ther Methods Clin Dev.* **15**, 211–220 (2019).
58. M. V. Hamann, N. Beschoner, X.-K. Vu, I. Hauber, U. C. Lange, B. Traenkle, P. D. Kaiser, D. Foth, C. Schneider, H. Büning, U. Rothbauer, J. Hauber, Improved targeting of human CD4⁺ T cells by nanobody-modified AAV2 gene therapy vectors. *PLoS One*. **16**, e0261269 (2021).
59. Q. Huang, K. Y. Chan, I. G. Tobey, Y. A. Chan, T. Poterba, C. L. Boutros, A. B. Balazs, R. Daneman, J. M. Bloom, C. Seed, B. E. Deverman, Delivering genes across the blood-brain barrier: LY6A, a novel cellular receptor for AAV-PHP.B capsids. *PLoS One*. **14**, e0225206 (2019).

60. T. F. Shay, E. E. Sullivan, X. Ding, X. Chen, S. Ravindra Kumar, D. Goertsen, D. Brown, A. Crosby, J. Vielmetter, M. Borsos, D. A. Wolfe, A. W. Lam, V. Gradinaru, Primate-conserved carbonic anhydrase IV and murine-restricted LY6C1 enable blood-brain barrier crossing by engineered viral vectors. *Sci Adv.* **9**, eadg6618 (2023).
61. S. Iqbal, E. Pérez-Palma, J. B. Jespersen, P. May, D. Hoksza, H. O. Heyne, S. S. Ahmed, Z. T. Rifat, M. S. Rahman, K. Lage, A. Palotie, J. R. Cottrell, F. F. Wagner, M. J. Daly, A. J. Campbell, D. Lal, Comprehensive characterization of amino acid positions in protein structures reveals molecular effect of missense variants. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 28201–28211 (2020).
62. W. J. Errington, B. Bruncsics, C. A. Sarkar, Mechanisms of noncanonical binding dynamics in multivalent protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 25659–25667 (2019).
63. A. J. Schoeffler, E. Helgason, N. Popovych, E. C. Dueber, Diagnosing and mitigating method-based avidity artifacts that confound polyubiquitin-binding assays. *Biophys Rep (N Y)*. **1**, 100033 (2021).
64. F.-E. Eid, A. T. Chen, K. Y. Chan, Q. Huang, Q. Zheng, I. G. Tobey, S. Pacouret, P. P. Brauer, C. Keyes, M. Powell, J. Johnston, B. Zhao, K. Lage, A. F. Tarantal, Y. A. Chan, B. E. Deverman, Systematic multi-trait AAV capsid engineering for efficient gene delivery. *bioRxiv* (2022), p. 2022.12.22.521680.
65. R. C. Challis, S. Ravindra Kumar, K. Y. Chan, C. Challis, K. Beadle, M. J. Jang, H. M. Kim, P. S. Rajendran, J. D. Tompkins, K. Shivkumar, B. E. Deverman, V. Gradinaru, Systemic AAV vectors for widespread and targeted gene delivery in rodents. *Nat. Protoc.* **14**, 379–414 (2019).
66. S. Berg, D. Kutra, T. Kroeger, C. N. Straehle, B. X. Kausler, C. Haubold, M. Schiegg, J. Ales, T. Beier, M. Rudy, K. Eren, J. I. Cervantes, B. Xu, F. Beuttenmueller, A. Wolny, C. Zhang, U. Koethe, F. A. Hamprecht, A. Kreshuk, ilastik: interactive machine learning for (bio)image analysis. *Nat. Methods.* **16**, 1226–1232 (2019).
67. D. R. Stirling, M. J. Swain-Bowden, A. M. Lucas, A. E. Carpenter, B. A. Cimini, A. Goodman, CellProfiler 4: improvements in speed, utility and usability. *BMC Bioinformatics.* **22**, 433 (2021).

68. N. K.-H. Lim, V. Moestrup, X. Zhang, W.-A. Wang, A. Møller, F.-D. Huang, An Improved Method for Collection of Cerebrospinal Fluid from Anesthetized Mice. *J. Vis. Exp.* (2018), doi:10.3791/56774.

Supplementary Materials for Example 1

Materials and Methods

Animals

[0458] All procedures were performed as approved by the Broad Institute IACUC. C57BL/6J (000664) were obtained from The Jackson Laboratory (JAX). *TFRC* KI mice (C57BL/6-Tfr1tm1TFR1/Bcgen, B-hTFR1, 110861) were obtained from Biocytogen. Recombinant AAV vectors were administered intravenously via the retro-orbital sinus in young adult male or female mice. Mice were randomly assigned to groups based on predetermined sample sizes. Transgenic animals were genotyped using DNeasy Blood & Tissue Kit (Qiagen, 69504) using primers listed in table S1. Two mice were excluded from analysis in Fig. 4, Fig. 5, fig. S5, and fig. S6: The first was a mouse assigned to the AAV9 in the *TFRC* KI group but was determined to be a WT by genotyping and western blotting. One mouse in the BI-hTFR1-treated C57BL/6J group displayed no evidence of transduction in the biodistribution, transduction, or Luciferase assays, and showed no fluorescent reporter signal in the liver or the brain. Experimenters were not blinded to the sample groups.

[0459] Mice were euthanized with Euthasol and transcardially perfused with PBS at room temperature followed by 4% paraformaldehyde (PFA) in ice cold PBS. Tissues were post-fixed overnight in 4% PFA in PBS and sectioned by vibratome. Immunohistochemistry (IHC) was performed on floating sections with antibodies diluted in PBS containing 5% donkey serum, 0.1% Triton X-100, and 0.05% sodium azide. Primary antibodies NeuN 1:500 (Invitrogen, MA5-33103), SOX9 1:250 (abcam, ab185966) were incubated at room temperature overnight. The sections were then washed and stained with secondary Alexa-conjugated antibodies Alexa Fluor 647 (Invitrogen, A-31573) and Alexa Fluor 555 (Invitrogen, A-21427) at 1:1000 for four hours or overnight.

Plasmids

[0460] AAV9 and the BI-hTFR1 Rep-Cap plasmids were generated by gene synthesis (GenScript). The CAG-WPRE-hGH pA backbone was obtained from Viviana Gradinaru's lab through Addgene (#99122). GFP, GFP-2A-luciferase, and mScarlet cDNAs were synthesized as gBlocks (IDT) or synthesized and cloned (GenScript).

[0461] Full length *TFRC* cDNA expression and lentiviral plasmids were cloned by inserting the open reading frames of *TFRC* (human: NM_003234.4; mouse: NM_011638.4; marmoset: NM_001301847.1, macaque: NM_001257303.1) into the lentiviral backbone pLenti-EF-FH-TAZ-ires-blast (Addgene #52083) with EcoRI/SalI sites.

[0462] For the C-terminal Fc fusion protein, the coding sequence of the extracellular region of human TfR1 (residues 89–760) were amplified by PCR and inserted into pCMV6-XL4 FLAG-NGRN-Fc (Addgene #115773) with EcoRV and XbaI sites. The signal peptide H7 (ATGGAGTTTGGGCTGAGCTGGGTTTTCTCGTTGCTCTTTTATAGAGGTGTCCAGTGT (SEQ ID NO: 1735)) was introduced to N terminal of the *TFRC* sequence to direct protein secretion.

[0463] For the N-terminal Fc fusion protein, the coding sequence of the extracellular region of human TfR1 (residues 89–760) was amplified by PCR. An N-terminal Fc region with an N-terminal H7 signal peptide and a C-terminal GS linker was ordered from IDT. The desired region of plasmid pCMV6-XL4 FLAG-NGRN-Fc (Addgene #115773) was amplified by PCR. These three DNA fragments were assembled with NEBuilder® HiFi DNA Assembly (NEB E2621).

Cell lines and primary cultures

[0464] HEK293T/17 (CRL-11268), Pro5 (CRL-1781), and CHO-K1 (CCL-61) were obtained from ATCC. hBMVECs cells (H-6023) were obtained from Cell Biologics and cultured as directed by the manufacturer. hCMEC/D3 cells were obtained from Millipore-Sigma (SCC066).

Reconstitution of full-length TfR1

[0465] Full length human TfR1 was reconstituted into Peptidiscs (PEPTIDISC LAB, dissolved at 2 mg/mL in 20 mM TrisHCL, pH 8). Briefly, HEK293 cells expressing C-terminally tagged full length human TfR1 were harvested by centrifugation at $300 \times g$. Cells were lysed by sonication and fractionated by centrifugation at $15,000 \times g$ to remove nuclei and cell debris and at $100,000 \times g$ to pellet the cell membrane. Membrane was resuspended in Buffer A (TrisHCl, pH 7.5, 50 mM NaCl, 10% glycerol) to 10 mg/mL, and then supplemented with 0.8% n-Dodecyl- β -D-Maltoside (DDM, Goldbio, no. DDM5) at 4°C with gentle agitation to solubilize the membranes. Anti-FLAG M2 magnetic beads (Sigma-Aldrich) were incubated with the solubilized membrane and washed with buffer A with 0.02% DDM. Beads were incubated for five minutes at room temperature in a solution of 1 mg/mL Peptidisc peptide and then thoroughly washed in buffer A without detergent.

Peptidisc-TfR1 was eluted in TrisHCl, pH 7.5, 150 mM NaCl, 0.1 mg/mL 3XFLAG peptide and analyzed by SDS-PAGE.

Bio-Layer Interferometry

[0466] A GatorPrime instrument was used for the characterization of the binding dynamics for all BLI studies and all analysis was conducted with Gator Bio software v 2.10.4.0713. Experiments were run in black 384 well plates (Greiner Bio-One, 781906). All reagents and buffer wells were prepared with Q buffer and probes from Gator Bio.

[0467] For AAV capsid binding to full length human TfR1, AAVx probes (Gator Bio) were equilibrated in Q buffer and then loaded with either BI-hTfR1, BI-hTfR1-2, BI-hTfR1-3, BI-hTfR1-4, or AAV9 by uniform thresholding to an 8 nm shift. Seven serial dilutions of full length human TfR1-FLAG were associated to each capsid for 120 seconds and then dissociated in Q buffer for 120 seconds. Only the five highest concentrations are shown on sensorgrams. Three trials of the assay were performed. A probe incubated with Q buffer instead of human TfR1 was included in each trial for background subtraction.

[0468] For BI-hTfR1 and AAV9 binding to pre-incubated Tf with full length human TfR1, assays and analysis were conducted as above but with 300 nM holo-Tf (Millipore Sigma T413) pre-incubated with each dilution of TfR1 for 15 minutes.

[0469] For kinetic analysis of capsid binding to TfR1, the sensorgram data from each of three trials using seven different analyte concentrations per capsid displayed clear multi-phasic binding and dissociation characteristics that could not be explained by a monovalent interaction. Therefore, Applicants used the global 2:1 exponential association and decay model. However, Applicants are not implying that the binding between the TfR1-targeting capsids and TfR1 is bivalent. The kinetic constants shown in fig. S2 should be interpreted with caution due to the following factors: First and foremost, the high density of binding epitopes on the AAV surface may lead to local increases in the effective concentration of the receptor analyte, which results in a higher apparent affinity(62). Second, TfR1 itself is bivalent, which could contribute to the aforementioned avidity artifact and also result in the bridging of closely loaded adjacent AAVs. While probes were loaded to less than 60% of their total binding capacity, the uniformity of loading cannot be controlled and a fraction of the measurements may result from 2:1 binding of the receptor dimers to adjacent capsids(63). Third, transitional binding states due to conformational changes or sequential binding events may contribute to the observation of multi-phasic binding kinetics.

[0470] The software-computed rate constants from these replicate fittings were subject to QC thresholding that included $R^2 > 0.99$, Gator-computed error ratios of $<25\%$ of their respective kinetic constants, and fitting residuals $<10\%$. In the cases of both BI-hTFR1-3 and BI-hTFR1-4, one replicate was eliminated due to a k_{on} error ratio of greater than 25% . Trials with QC threshold-passing data were averaged and their standard error was taken.

[0471] For the tandem competition binding assays, a human TfR1 apical domain-targeting antibody, OKT9 (Thermo Fisher Scientific) was chosen as a competitor. The extracellular domain of human TfR1 (residues 89–760) was recombinantly expressed as an N-terminal Fc-fusion protein (Fc-hTfR1) in Expi293 suspension cells and purified by gravity flow with PierceTM Protein A Agarose (Thermo Fisher Scientific). HFCII probes were loaded with Fc-hTfR1 by thresholding to a 1 nm shift. Fc-TfR1-loaded probes were associated in tandem, first to BI-hTFR1 until the signal reached approximately 2 nm, and subsequently to the OKT9 antibody at 250, 100, 50, or 25 nM or in Q buffer alone. The latter was included to demonstrate BI-hTFR1 dissociation in the absence of competitor OKT9. In parallel, a Fc-hTfR1 loaded probe was associated first in Q buffer alone and next in 250 nM of the OKT9 antibody to demonstrate maximum antibody binding in the absence of capsid. The two association phases were 240 and 360 seconds, respectively.

[0472] For the triple tandem binding assays involving holo-Tf, holo-Tf was biotinylated for two hours on ice with an EZ-LinkTM Sulfo-NHS-LC-LC-Biotin kit (Thermo Fisher Scientific) and then buffer exchanged into PBS with a 0.5 mL Pierce Protein Concentrator PES column with a 10K MWCO to remove unreacted biotin. Streptavidin (SA) probes were equilibrated in Q buffer and then loaded to a 2 nm shift with biotinylated holo-Tf or were incubated in Q buffer to serve as a no-ligand control. Three of the probes were submitted to sequential binding steps for 240 seconds each. The probes were incubated in Q buffer for 240 or 120 seconds between each binding step to demonstrate $<10\%$ baseline drift (i.e., loss of binding) before binding the next molecule. In channel 1, the holo-Tf-loaded probe was associated first with 150 nM full length human TfR1, next with Q buffer only, and finally with BI-hTFR1. In channel 2, the holo-Tf loaded probe was associated with TfR1, then 200 nM OKT9 antibody, and finally with BI-hTFR1 capsid. In channel 3, the holo-Tf-loaded probe was associated with Q buffer alone, then the OKT9 antibody, and finally with BI-hTFR1. In channel 4, the unloaded probe was associated with TfR1, then OKT9 antibody, and finally with BI-hTFR1 to reveal any nonspecific binding to the probe.

[0473] For full length hTfR binding to human holo-Tf, MFC probes were equilibrated in Q buffer and then loaded with anti-Tf monoclonal antibody 12A6 (Thermo Fisher Scientific) to a 1 nm shift. Next, holo-Tf was loaded as ligand to saturation. Serial dilutions of full length peptidisc-reconstituted hTfR1-FLAG were associated with holo-Tf for 120 seconds and then dissociated in Q buffer for 120 seconds. Three trials of the experiment were performed. A probe incubated with Q buffer alone was included in each trial for background subtraction in sensorgram curve fittings and kinetics analysis. For subsequent kinetics analysis, the sensorgram data from each of three trials were fitted with global 1:1 exponential association and decay models. The software-computed rate constants from these replicate fittings were subject to QC thresholding that included $R^2 > 0.99$, computed error ratios of $<10\%$ of their respective kinetic constants, and fitting residuals $<10\%$. Kinetics values were averaged and their standard error was taken.

Lentivirus production

[0474] Lentivirus was produced with a third-generation lentivirus system by cotransfection of three packaging plasmids (pMDLg-RRE, pRSV-Rev, and pVSV-G) and vector plasmid (pLenti-EF-FH-TAZ-IRES-blast, a gift from Yutaka Hata, Addgene plasmid # 52083) encoding the entire coding sequences of *TFRC* or *Tfrc* (human: NM_001128148; macaque: NM_001257303; marmoset: NM_001301847; or mouse: NM_011638.4) at a ratio of 4:1:1:6 in HEK293T/17 cells. Lentiviruses were harvested from the media three days after transfection and filtered with 0.45 μm PES to remove cell debris.

Exogenous TfR1 expression in CHO cell lines

[0475] CHO cell lines expressing human *TFRC* or macaque, marmoset, or mouse *Tfrc* were established via random lentiviral integration. Briefly, CHO cells were seeded at 150,000 cells/well in 12-well plates and transduced with lentivirus-containing media for 48 hours, then transferred to 10 cm dishes for selection with 10 $\mu\text{g/mL}$ blasticidin for one week followed by maintenance in media with 1 $\mu\text{g/mL}$ blasticidin. To validate expression of the desired construct, RNA was extracted from cells using the Qiagen RNeasy Kit. Maxima H Minus Reverse Transcriptase (ThermoFisher, EP0753) and oligo dT were used for cDNA synthesis. Species-specific primers were used for qPCR (table S1). CHO cells expressing TfR1 from each species were also assessed by immunostaining with two anti-TfR1 antibodies: ab84036 (Abcam) and OKT9 (eBioscienceTM) (FIG. 7).

AAV production and titering

[0476] The AAV9 variant capsids carrying CAG-NLS-mScarlet-P2A-Luciferase-WPRE-SV40pA or CMV/CBA-*GBAI*-HA-pA were either produced in suspension followed by iodixanol purification and titered as previously described (64) or produced using adherent cells as previously described (65). Production, purification methods, and post-purification yields are provided in table S2. For BLI experiments, AAV vectors were produced in adherent HEK293T cells and purified by static binding (3 to 10 plates per preparation). At 72 hours post-transfection, the cells and supernatant (60–200 mL per prep) were collected in 125–500 mL shake flasks, incubated with 0.1% Triton X-100, 2 mM MgCl₂ and 25 U/mL benzonase (Sigma, E1014-25KU) for 90 minutes at 37°C, and centrifuged for 10 minutes at 3724 ×g for clarification. The clarified lysates were incubated with 150 µL POROS AAV9 resin (Thermofisher, A27353), at 37°C for 90 minutes, under agitation (1.14 ×g). The mixes were loaded into detergent removal spin columns (Thermofisher, 87778) and supernatants were discarded using a vacuum manifold. POROS AAV9 beads were rinsed three times with 5 mL PBS, using a vacuum manifold. AAV vectors were eluted in 15 mL Falcon tubes, with 2 mL elution buffer (0.1 M Glycine, pH2.5) and neutralized with 500 µL 1 M Tris, 0.5 M NaCl, pH 8. AAVs were dialysed against 3 × 2 L of PBS (with calcium and magnesium) supplemented with 0.001% pluronic F68, using float-a-lyzer G2 dialysis cassettes with a MWCO of 8–10 Kd (Repligen, G235067), or buffer exchanged and concentrated in the same final formulation, using centrifugal filters with a MWCO of 100 Kd (Thermofisher, 88532).

Binding assays

[0477] CHO cells stably expressing human *TFRC* or macaque, marmoset, or mouse *Tfrc* were seeded at 30,000 cells per well; hCMEC/D3 cells were seeded at 30,000 cells per well; and hBMVEC cells were seeded at 15,000 cells per well. At 24 hours post-seeding, each well was subjected to a media change with the fresh cold media that contained each AAV variant carrying CAG-NLS-mScarlet-P2A-Luciferase-WPRE-SV40pA at 3,000, 6,000, or 12,000 vg/cell for CHO, hCMEC/D3, or hBMVEC cells, respectively. The plate was then maintained at 4°C with gentle rocking for one hour, followed by five phosphate buffered saline (PBS) washes. Cells were treated with Proteinase K at 56°C for one hour, followed by heat inactivation at 95°C for 10 minutes. Total DNA extraction was performed using the DNeasy kit (Qiagen). The DNA was diluted 1:20 and subjected to qPCR with the mScarlet-qPCR-F and mScarlet-qPCR-R primers (table S1).

Transduction assays

[0478] CHO cells stably expressing human *TFRC* or macaque, marmoset, or mouse *Tfrc* were seeded at 100,000 cells per well; hCMEC/D3 cells were seeded at 10,000 cells per well; and hBMVEC cells were seeded at 5,000 cells per well. At 24 hours post-seeding, each well was subjected to a media change with the fresh media that contained the AAV variant carrying CAG-NLS-mScarlet-P2A-Luciferase-WPRE-SV40pA at 3,000, 6,000, or 12,000 vg/cell for CHO, hCMEC/D3, or hBMVEC cells, respectively, and the plate was maintained at 37°C with 5% CO₂ for 24 hours. Transduction was measured using the Britelite Plus reporter gene assay system (PerkinElmer, 6066766). Luciferase activity was reported in relative light units (RLU) as raw data or normalized to AAV9 controls.

Antibody inhibition assays

[0479] The hCMEC/D3 cells were seeded at 7,500 cells per well. Two days later, 100 µL of media with 3×10^8 vg/mL AAV and the specified concentration of OKT9 or R&D AF2474 antibody was transferred to each well. The plate was maintained at 37°C with 5% CO₂ for 24 hours. Transduction was measured using the Britelite Plus reporter gene assay system.

Transwell assays

[0480] hCMEC/D3 cells (under nine passages) were maintained in EGM2-MV media with VEGF (Lonza, CC4147). Falcon® Permeable Supports for 24-well Plates with 1.0 µm Transparent PET Membranes (Corning, 353104) were prepared by coating with collagen type I (Millipore, 08-115) diluted 1:50 in PBS and incubating at 37°C for two hours. On Day 0, cells were plated at 30,000/cm² on the membrane and grown for two days in 200 µL of EGM2-MV media with VEGF on the top of the membrane and 600 µL of media in the lower chamber. On Day 2 and Day 4, the media at both the top and bottom of the transwell was changed to EGM2-MV media without VEGF, and the media at the bottom of the transwell was changed to Astrocyte Conditioned Media (Sciencell, 1811). Starting from Day 4, TEER values were measured using the EVOM2 meter (World Precision Instruments). TEER values were calculated as the ohmmeter readout in wells with cells minus the ohmmeter readout in wells without cells, multiplied by the surface area in cm² units.

[0481] Transwell permeabilities using this process were previously validated by measuring the permeability of Dextran FITC 4k (Sigma, 46944), 40k (Biotium, 76221-470), and 70k (Biotium, 76221-460). TEER values peaked on Day 6, reaching values $>10 \Omega/\text{cm}^2$, which is when AAVs

were added to the transwells. BI-hTFR1, AAV2, and AAV9 with capsid-specific barcoded transgenes were pooled and added to the top well of the transwell at 25,000 vg/cell (each AAV). After three hours of incubation at 37°C or 4°C, 22 µL of media was extracted from the bottom chamber. To detect the amount of AAV in the media, extracts were treated with DNaseI (NEB, M0303S) for 15 minutes at room temperature, then diluted in TE buffer (5 mM EDTA), and denatured at 70°C for 10 minutes to deactivate DNaseI. DNA extracts were then diluted in 5% Tween 20 with UltraPure™ Salmon Sperm DNA Solution (ThermoFisher, 15632011). The amount of AAV that passed through the transwell was measured by qPCR using a standard curve of known AAV quantities derived from AAV-containing media applied to the top well.

Immunofluorescence staining and colocalization in hCMEC/D3 cells

[0482] For the imaging of BI-hTFR1 or AAV2 binding to hCMEC/D3 cells with or without OKT9 or Tf-647, BI-hTFR1 and AAV2 were added to hCMEC/D3 cells in optical plates (PhenoPlate 96-well, PerkinElmer, 6055302) at 50,000 vg/cell, while Tf-647 (ThermoFisher, T23366) and purified TfR1 antibody (OKT-9, AAT Bioquest, 10713000) at a 1:200 dilution were added to wells to reach a final concentration of 1 µg/mL each. Cells were incubated for one hour at 4°C, rinsed three times with PBS, fixed with 4% paraformaldehyde, and blocked (Avidin/Biotin Blocking Kit, SP-2001, Vector Laboratories) prior to applying Anti-AAV9 (1:200 dilution, ThermoFisher, 7103332100) to the wells treated with BI-hTFR1 or the wells with no virus, or Anti-AAVX (1:200 dilution, ThermoFisher, 7103522100) to the wells treated with AAV2. After three PBS washes to clear away unbound primary antibodies, Streptavidin 555 (1:500 dilution, Thermofisher, S21381) and Alexa Fluor® 488 (1:500 dilution, JacksonImmunoResearch, 115-545-003) were applied for the visualization of AAVs and TfR1, respectively.

[0483] To image AAV colocalization with Tf-647, BI-hTFR1 or AAV2 were added to hCMEC/D3 cells in optical plates at 25,000 vg/cell with Tf-647 (1 µg/mL). Cells were incubated for one hour at 4°C, followed by immunostaining as described above.

[0484] To image AAV and TfR1 colocalization, BI-hTFR1 or AAV2 were added to hCMEC/D3 cells in optical plates at 25,000 vg/cell. Cells were incubated for one hour at 37°C, prepared for immunostaining as described above but with the addition of a permeabilization step prior to the blocking of endogenous biotin to visualize TfR1. Streptavidin 555 (1:500 dilution, Thermofisher, S21381) and goat anti-mouse 647 nm (1:500 dilution, 115-605-062, Jackson ImmunoResearch) were applied.

[0485] For organelle colocalization, the following conjugated reagents were applied: Rab5 ((D-11) FITC, 1:100 dilution, Santa Cruz Biotechnology, sc-46692), Rab7 (1:400 dilution, abcam, ab198337), KDEL (SEQ ID NO: 1787) (1:500 dilution, abcam, ab184819), RCAS1 (1:200 dilution, Cell Signaling Technology, 12290), and TGN46 (1:50 dilution, abcam, ab50595). For unconjugated antibodies (RCAS1 and TGN46), goat anti-rabbit 488 (Alexa Fluor® 488 AffiniPure Goat Anti-Rabbit IgG (H+L), 1:500 dilution, JacksonImmuno, 111-545-144) was applied to visualize the staining. For wells requiring dual mouse antibodies (those co-stained with Rab5 and OKT9), cells were blocked for one hour with mouse-on-mouse blocking reagent (ReadyProbes™ Mouse-on-Mouse IgG Blocking Reagent, 1:30 dilution, ThermoFisher, R37621) between antibody staining procedures.

[0486] Images for all experiments were obtained with a 60X oil objective on a Nikon Ti-e inverted spinning disc confocal microscope, with the exception of TGN46 staining, which was performed using a 63X oil objective on a Zeiss LSM 900 confocal microscope. Within each experiment, images were taken with predetermined, optimized, and fixed exposure times to allow image comparisons.

Imaging and tissue cell counting

[0487] Images of native mScarlet in the whole brain, cortex, striatum, thalamus, spinal cord, liver, or the dorsal root ganglion were taken on a Keyence BZ-X810. For images of whole brain sections, single plane tiled images were taken with autofocus mode on and stitched using the Keyence Analysis Software. For comparisons of native mScarlet fluorescence across groups, imaging settings were optimized for BI-hTFR1 in *TFRC* KI mice and then applied across all samples and all groups within each organ. Adjustments to the levels and gamma in the whole brain, cortex, striatum, thalamus, spinal cord, liver, and dorsal root ganglion images were performed using identical settings across all samples within each brain region or tissue type.

[0488] Stained sections were imaged on a Keyence BZ-X810 using sectioning mode with a Z-stack depth of 10 µm during acquisition. For cell counting, image settings were exposed to detect low mScarlet expressing cells in each sample group within each brain region. The Z-stack was reconstructed into a full-single-stack image using Keyence Analysis Software and exported as separate TIFs for each channel imaged. Images were then imported as input data into ilastik version 1.4.0 (66) where the Pixel Classification within the segmentation workflow was used. All color/intensity, edge, and texture features were selected and used during training. Training was

performed to distinguish foreground and background pixels across images that represented the cortex, striatum, or thalamus in C57BL/6J or *TFRC* KI mice that had received BI-hTFR1 or AAV9 in NeuN or SOX9 stained brain sections. Predictions were then exported as probabilities in TIF format and then imported into CellProfiler 4.2.4 (67). A custom workflow was created to determine percent overlap between positively transduced cells and antibody markers. First, channels were aligned using the Align module. Second, the IdentifyPrimaryObjects module with Otsu two classes thresholding method was used to detect and segment foreground objects (positively transduced cells, NeuN-positive, or SOX9-positive cells) from background in the images. Third, foreground objects were then expanded by 2–5 pixels using the ExpandOrShrinkObjects module. Fourth, the MaskObjects module was used to determine the number of colocalized objects that overlap with a threshold of 40% overlap or more. A separate CellProfiler workflow was created where the IdentifyPrimaryObjects module was used to detect foreground and background objects (two class) and then the OverlayOutlines module was used to overlay an outline of detected foreground objects onto the raw input image of the mScarlet, NeuN, or SOX9 channel. The images were exported with the SaveImages module. The exported image was then used to verify if further training with the Pixel Classification in ilastik was needed in order to detect missed or improperly classified cells. An independent evaluator without prior knowledge of expected outcomes conducted a manual count of randomly selected images using the Photoshop Count Tool to determine the percent overlap of mScarlet-positive cells with NeuN or SOX9-positive cells in stained tissue. The results were evaluated to be within $\pm 5\%$ deviation.

Western blot

[0489] Purified N-terminal flag tagged full length human and mouse TfR1 proteins or whole mouse brain lysates were separated on Bolt 4–12% Bis-Tris Plus gels and transferred onto nitrocellulose membranes. After incubation with anti-TfR1 antibodies (0.5 $\mu\text{g/mL}$ ThermoFisher, Catalog # 13-6890; 0.25 $\mu\text{g/mL}$ R&D, AF2474) or anti- β -actin (1:5000; Abcam, AB20272) followed by incubation with a horseradish peroxidase (HRP)-conjugated secondary antibody. The detection of the HRP signal was by SuperSignal West Femto Maximum Sensitivity Substrate using a Bio-Rad ChemiDoc TM MP system #1708280.

In vivo biodistribution, transduction, and Luciferase assays

[0490] BI-hTFR1 or AAV9 encoding CAG-NLS-mScarlet-P2A-Luciferase-WPRE-SV40pA were administered intravenously to 18 to 21-week-old female C57BL/6J or *TFRC* KI mice at a

dose of 5×10^{11} vg/mouse. After three weeks, the mice were perfused with PBS. Tissue samples from the brain, liver, spinal cord, and DRGs were dropped fixed into 4% PFA. Remaining tissues, including additional samples of the brain, liver, spinal cord, and DRGs were collected snap frozen on dry ice and stored at -80°C . Biodistribution analysis: Samples were processed using a DNeasy 96 Blood & Tissue Kit (Qiagen, 69581). qPCR was performed for the mScarlet transgene and the mouse glucagon gene as previously described (see primer sequences in table S1) (5). For mRNA analysis, RNA was isolated using TRIzol (Invitrogen, 15596026) and purified using an RNeasy 96 Kit (Qiagen, 74171). cDNA synthesis was performed using Maxima H Minus Reverse Transcriptase (Thermo Scientific, EP0753). qPCR for mScarlet and mouse GAPDH was performed (primer sequences in table S1). Luciferase assays: Protein was extracted from tissue samples using the lysis buffer T-Per with 1X Halt protease inhibitor (Invitrogen, 78430) while kept ice cold at all times. 10 μg of total protein was used to assess the reporter gene using the Britelite Plus reporter gene assay system.

GBA1 in vivo delivery experiment

[0491] BI-hTFR1 or AAV9 encoding CMV/CBA-GBA1-HA-pA was intravenously injected into 28 to 30-week-old female TFRC KI mice at a dose of 1×10^{14} or 5×10^{12} vg/kg. After three weeks, cerebrospinal fluid (CSF) was collected from the cisterna magna of anesthetized mice as previously described (68). Briefly, the cisterna magna was surgically exposed and punctured with a sharpened glass capillary, and CSF was allowed to flow into the capillary. Approximately 6–10 μL of CSF was collected from each animal. Then, the thoracic cavity was exposed and 200–400 μL of blood was collected into a microtainer tube with serum separator additive (BD, 365967). The mice were then transcardially perfused with ice cold PBS. Brain, liver, and spinal cord tissue sections were drop fixed into 4% PFA. Additional tissue samples from these organs were collected and snap frozen at -80°C . After a 1.5-hour incubation at room temperature, the blood was centrifuged at $10,000 \times g$ for 90 seconds, and serum was collected and stored at -80°C . IHC was performed on fixed tissues as described in the Animals section of the Methods. GBA-HA was detected with an anti-HA primary antibody (Sigma, 11867423001) at a 1:50 dilution, followed by an AlexaFluor568-conjugated secondary antibody at a 1:500 dilution (Invitrogen, A78946). Vector genomes were quantified using qPCR primers targeting CMV and mouse GAPDH (table S1).

[0492] To measure GCcase activity, whole tissue lysates were prepared. Tissue was thawed on ice, weighed, and homogenized in a $2\times$ volume of ultra-pure distilled water with cComplete

protease inhibitor (Sigma, 4693132001) using the GenoGrinder (SPEX Sample prep). Homogenate was then centrifuged at $10,000 \times g$ for 2 minutes, transferred to a new, pre-chilled 1.5 mL Eppendorf tube, and incubated on ice for 15–30 minutes. Next, the samples were centrifuged at $12,000 \times g$ for 20 minutes. Supernatant was collected and stored at -80°C or used in subsequent analysis. A BCA assay (Thermo Fisher Scientific, 23225) was used to determine total protein concentration in the resulting lysates, as described by the product manufacturer. The SensoLyte Red Glucocerebrosidase (GBA) Activity Assay Kit (AnaSpec, AS-72259) was used to determine enzyme activity in serum and whole tissue lysate samples. Briefly, 20 μL of serum or the volume of tissue lysate containing 30 μg of protein was added into each well. Assay buffer was added to a total volume of 50 μL . Then, 50 μL of 1X enzyme substrate was added to each well. Samples were gently mixed and incubated for one hour at room temperature. The fluorescence signal at an excitation/emission of 570/610 nm was measured in end-point mode using the EnVision 2104 plate reader (PerkinElmer). A resorufin reference fluorescence standard curve was used to determine the amount of reaction product in each well from the measured relative fluorescence units (RFU). The resulting values were used to calculate enzyme activity, expressed in μU per mL of blood or μU per mg of protein. One unit of GCase activity is the amount of enzyme that generates 1 μmol of product per minute under the specified reaction conditions.

Table S1. Primers and barcodes used for qPCR in this study. Each sequence is listed 5' to 3'.

Primer name	Primer sequence
hTfR1-qPCR-F	GCTATGGGACTATTGCTGTGATCGTC (SEQ ID NO: 1736)
hTfR1-qPCR-R	CCAATATAAGCGACGTGCTGCAG (SEQ ID NO: 1737)
msTfR1-qPCR-F	TGTGGAGATGAAACTGGCTGC (SEQ ID NO: 1738)
msTfR1-qPCR-R	AGTGCAATAGCTGCAAAGCAGAG (SEQ ID NO: 1739)
marTfR1-qPCR-F	AACAAGAGCGTGAAGCTCTCTG (SEQ ID NO: 1740)
marTfR1-qPCR-R	CTCCAGGACCCCAGGAGTC (SEQ ID NO: 1741)
macTfR1-qPCR-F	GGAAAACAAGAGTGTGAAGCTCACG (SEQ ID NO: 1742)
macTfR1-qPCR-R	GGAGAGCTGTCCCCACACTG (SEQ ID NO: 1743)
MSCARLET-QPCR-F	CCGTGACCCAGGACACCTC (SEQ ID NO: 1744)
mScarlet-qPCR-R-biod	CTTCTTCTGCATTACGGGGC (SEQ ID NO: 1745)
mScarlet-qPCR-R-binding	GCCATCTTAATGTGCGCCCTTCAG (SEQ ID NO: 1746)
CMV-qPCR-F	TGTTCCCATAGTAACGCCAATAG (SEQ ID NO: 1747)
CMV-qPCR-R	GTACTTGGCATATGATACACTTGATG (SEQ ID NO: 1748)

GAPDH-qPCR-F	ACCACAGTCCATGCCATCAC (SEQ ID NO: 1749)
GAPDH-qPCR-R	TCCACCACCCTGTTGCTGTA (SEQ ID NO: 1750)
mGAPDH-qPCR-F	CGCCCTGATCTGAGGTAAAT (SEQ ID NO: 1751)
mGAPDH-qPCR-R	CGGAGCAACAGATGTGTGTA (SEQ ID NO: 1752)
mGlucagon-qPCR-F	AAGGGACCTTTACCAGTGATGTG (SEQ ID NO: 1753)
mGlucagon-qPCR-R	ACTTACTCTCGCCTTCCTCGG (SEQ ID NO: 1754)
msTfR1-exon1and2-F	CTAACTTGTGTTGGTGGGGAAC (SEQ ID NO: 1755)
msTfR1-exon1and2-R	CTTCATCTGCAGCCAGTTTCAT (SEQ ID NO: 1756)
genotyping-h&mTFRC-F	CTCGTGAGGCTGGATCTCAAAAAGATG (SEQ ID NO: 1757)
genotyping-hTFRC-R	CACCAGGTAAACAAGTCTACCGTTCTTATC (SEQ ID NO: 1758)
genotyping-mTfrc-R	ACTGGGTCTAAGTTACCATTGACTGC (SEQ ID NO: 1759)
BI-hTfR1_barcode_qPCR_F	TCGAGACCATAGATGACTACCC (SEQ ID NO: 1760)
BI-hTfR1_barcode_qPCR_R	GATCTGTGTTGTTGCGGTATAATG (SEQ ID NO: 1761)
AAV2_barcode_qPCR_F	TCGAGATGGAGGAAGGAACT (SEQ ID NO: 1762)
AAV2_barcode_qPCR_R	AGATCTGCTGTAACCCACAAA (SEQ ID NO: 1763)
AAV9_barcode_qPCR_F	TCGAGCATGCGGGTAGAT (SEQ ID NO: 1764)
AAV9_barcode_qPCR_R	CTGTCTTGATGTGATTGCTGTA (SEQ ID NO: 1765)
BI-hTfR1 genome barcode	CTCGAGACCATAGATGACTACCCATCCCATCGAGCGTTGT ACCCTATTTCAGAGGCAGACATTATACCGCAACAACACAG ATCT (SEQ ID NO: 1766)
AAV2 genome barcode	CTCGAGATGGAGGAAGGAACTTTTGC GTGTGAGCGTTGTA CCCTATTTCAGAGTGACACGTTTGTGGGTTACAGCAGATCT (SEQ ID NO: 1767)
AAV9 genome barcode	CTCGAGCATGCGGGTAGATCATGTGGGTAGGAGCGTTGT ACCCTATTTCAGAGGGTTACAGCAATCACATCCAAGACA GATCT (SEQ ID NO: 1768)

Table S2. Production and purification methods used in this study with final post-purification yields.

Study	Genome	Production	Purification	ddPCR target	AAV9	BI-hTFR1	Unit
1	CAG-NLS-GFP-P2A-Luc	adherent HEK293T	static affinity	CAG	4.10E+11	2.06E+11	vg/150mm
2	CAG-NLS-GFP-P2A-Luc	suspension HEK293T	iodixanol	WPRE	2.78E+13	9.93E+12	vg/L
3	CMV-CBA-hGBA-HA	suspension HEK293T	iodixanol	CAG	1.74E+13	1.25E+13	vg/L

Example 2 - Improved polypeptides that bind human Transferrin Receptor and facilitate AAV entry into the CNS

Background:

[0493] A critical challenge in the development of more efficient delivery vectors for gene therapy is engineering vectors with known mechanisms of action (MOAs) present in human patients. Conventional approaches to engineering vectors such as adeno-associated viruses (AAVs) with enhanced tropisms have largely relied on capsid library selections in animals. These selections have been used by numerous groups over the past two decades because they can be successful without requiring prior knowledge of an MOA (1–7). However, vectors selected in animals mostly have not translated across preclinical models. Despite extensive searches, AAV capsids with clear translational potential and known MOAs for entering the CNS have not yet emerged. Here, Applicants approached this challenge from a different perspective and selected AAV capsids first for a specific MOA, that is binding to the human Transferrin Receptor (TfR1), and showed that these capsids crossed the blood-brain barrier (BBB) and mediated efficient gene delivery throughout the CNS.

[0494] Applicants chose TfR1 as a target because of its high expression on the human BBB; its ability to mediate constitutive, ligand-independent receptor-mediated transcytosis (RMT) across the CNS vasculature (8–12); and its track record as a target to increase the delivery of biologics into the CNS of mice (13–15), nonhuman primates (NHPs) (16–18), and humans as investigational therapies (19–22) and an approved antibody-based therapeutic for Mucopolysaccharidosis type II (23). Applicants first screened 7-mer-modified AAV9 capsid libraries for their ability to bind to human TfR1 *in vitro*. The top-performing capsid, AAV-BI-

hTfR1, exhibited more efficient gene delivery to human brain endothelial cells and improved active transport across a human vascular monolayer. When systemically administered to adult *TFRC* knock-in (KI) mice carrying a chimeric *Tfrc* gene with a humanized extracellular domain, BI-hTfR1 transduced the majority of neurons and astrocytes across multiple brain regions. The enhanced tropism was entirely dependent on an interaction with the humanized TfR1, as Applicants observed no enhancement in wild type (WT) mice. The tropism was also selectively enhanced in the CNS, consistent with the high level of expression of *TFRC* on the CNS vasculature relative to the vasculature in other organs (12).

2nd and 3rd Generation TfR1 Binding Polypeptides and Their use for Gene Delivery to the Brain

[0495] To identify TfR1 binding polypeptides that enable improved gene delivery to the CNS when incorporated into the AAV capsid, Applicants generated limited site saturation mutagenesis libraries from several of the TfR1 binding 7-mers (SEQ IDs 32-38), including the BI-hTfR1 polypeptide that Applicants previously described PCT/US2023/070285, hereby incorporated by reference. Applicants performed single or dual site saturation mutagenesis on every combination of two amino acids within the immediate region surrounding the 7-mer insertion (amino acids 586 to 592, AAV9 VP1 numbering). In some cases, saturation mutagenesis was performed within the previously identified 7-mer. The mutations were generated using a pooled DNA oligonucleotide library and were assembled into AAV genomes carrying a Cap gene fragment as previously described PCT/US2023/070285. Each sequence was encoded by two nucleotide sequences, providing internal replicates for each amino acid sequence. The DNA libraries were used to make AAV libraries, and the libraries were then screened for binding to TfR1-Fc fusion proteins and for their ability to cross the BBB and transduce cells within the brain and spinal cord of human *TFRC* KI mice.

[0496] Numerous variants identified through these library screening assays outperformed BI-hTfR1 based on *in vivo* assessments of their ability to transduce cells within the CNS. Table 1 lists human polypeptides that were all validated for binding to human TfR1-Fc fusion proteins and which showed strong evidence of crossing the BBB expressing the packaged Cap gene in the brain and spinal cord. The table provides the indicated SEQ ID for the polypeptides that were inserted in replacement of AA566 through AA592 in SEQ ID 2 and the transduction enrichment scores for each polypeptide sequence in the brain, spinal cord, and liver. Transduction was assessed by Cap gene PCR from cDNA generated from the tissue samples followed by Illumina sequencing as

previously described (24). Enrichment is the log₂ assay reads per million (RPM)/RPM of each variant in the virus library. Enrichment scores are grouped within the following ranges: “--” is < -3; “-” is > -3 and < -1; “+” is > -1 and < 2; “++” is > 2 and < 4; “+++” is > 4 and < 5; “++++” is > 5 and < 6; “+++++” is > 6.

Table 1. Polypeptide sequences that bind human TfR1 and facilitate gene delivery to the brain of human TFRC knock-in mice

[0497] In a second library, top candidates from the above screen were subjected to additional single- or dual-site saturation mutagenesis. A pooled DNA oligonucleotide library containing the designed sequences was assembled into two distinct Cap expressing genomes. In one genome, the Cap gene was placed under the regulatory control of a CBA-CMV promoter (CBA-Cap). In a second genome, the Cap gene was placed under the regulatory control of a human *SYN1* gene fragment (hSyn-Cap). The CBA-Cap library was screened for binding to TfR1-Fc fusion proteins and the CBA-Cap and hSyn-Cap libraries were screened for their ability to cross the BBB and transduce cells within the CNS of control mice or mice with human *TFRC* expression in the brain endothelium. For this experiment the hTFR1 mice were made to express human *TFRC* by administering AAV-BI30:CAG-hTFRC to immune compromised mice prior to injection of the libraries. Use of immune compromised mice such as NSG mice prevents a humeral immune response to the initial AAV-BI30 exposure from neutralizing the activity of the subsequently injected AAV library.

[0498] Table 2 lists TfR1 binding polypeptides that were screened for CNS transduction efficiency as part of this second screening experiment. Each library was screened for CNS transduction in the brains of mice that had been previously made to express human *TFRC* using AAV-BI30 packaging AAV-CAG-hTFRC or no receptor as controls. The columns provide the SEQ ID and transduction assessments for each polypeptide sequence in each assay (hTfR1 or control) with each library (CBA or hSyn). Transduction was assessed by Cap gene PCR from cDNA generated from the tissue samples followed by Illumina sequencing as previously described (24). Enrichment is the log₂ assay reads per million (RPM)/RPM of each variant in the virus library. Enrichment scores are listed within the following ranges: “--” is < -3; “-” is > -3 and < 1; “+” is > 1 and < 2; “++” is > 2 and < 4; “+++” is > 4 and < 5; “++++” is > 5 and < 6; “+++++” is > 6.

[0499] Numerous variants identified through each of these library screens outperformed BI-hTFR1 based on *in vivo* assessments of their ability to transduce cells within the CNS. Tables 1

and 2 provide example sequences from these libraries. All sequences reported in the tables were found to bind a purified human TfR1-Fc fusion protein.

Table 2. Polypeptide sequences that bind human TfR1 and facilitate gene delivery to the brain of mice made to express human TfR1 on the CNS vasculature

[0500] Applicants tested several of the novel TfR1 binding polypeptides modified capsids and compared them to BI-hTFR1 (SEQ ID 21). BI-hTFR1, BI294 (SEQ ID 30) and BI295 (SEQ ID 31) were used to package a ssAAV-CAG-NLS-mScarlet-2A-luc-pA genome, and the AAVs were delivered to adult human *TFRC* KI mice at 5×10^{10} vg/mouse (approximately 2×10^{12} vg/kg). Even at this low dose, each AAV transduced a substantial fraction of cells throughout the brain. BI294 outperformed BI-hTFR1 as assessed by the brightness of mScarlet expression (FIG. 14A) and the fraction of neurons (FIG. 14B) and astrocytes (FIG. 14C) transduced. BI295, which is a second-generation variant of a distinct TfR1 binding sequence, also transduced a greater fraction of neurons and glia in the brain of the *TFRC* KI mice. Notably, as compared to BI-hTFR1, both BI294 and BI295 were detargeted from the liver (FIG. 15). These results in the humanized *TFRC* KI mice suggest that BI294 and BI295 are well suited for mediating efficient CNS gene delivery for gene therapy applications in humans.

[0501] Based on the sequence differences between Family 1 polypeptides exemplified by BI-hTFR1, BI294, and additional novel sequences disclosed within (SEQ ID#s 30, 32-34, and 58-1208) and Family 3 polypeptides exemplified by BI295 and novel sequences disclosed within (SEQ ID#s 31, 36-38, and 1209-1386), Applicants hypothesized that the sequence families used distinct binding mechanisms. To test this, Applicants used the knowledge that the capsids bind to human TfR1 apical domain but not the macaque TfR1 apical domain to map the human-specific amino acids that are critical for binding. Applicants made amino acid mutations within human and macaque TFR1 where Applicants swapped individual or grouped neighboring subsets of amino acids that differ between the human and macaque apical domain sequences (FIG. 16A, B). Applicants made plasmids expressing WT human or macaque *TFRC*, or plasmids where codons for one or more amino acids from the macaque TfR1 sequence were substituted for the corresponding human amino acid coding sequence. The opposite coding mutation swaps were also made where the macaque amino acid(s) was replaced with the human AA(s). These plasmids were used to individually transfect CHO cells, and the transfected cells were subsequently exposed to BI294 or BI295 and transduction was assessed by measuring luciferase activity (FIG. 16C,D). The results clearly demonstrated that human TfR1 binding by BI294 requires the human-specific amino

acids Y247 and/or T248 (FIG. 16C) and that TfR1 binding by BI295 requires a distinct human specific amino acid R208 (FIG. 16D). Therefore, Applicants conclude that the BI-hTfR1/BI294 motif (Family 1) and the BI295 motif (Family 3) bind to two distinct regions of the human TfR1 apical domain using two different sequence motifs. Importantly, mutations at these sites are uncommon in the human population indicating that genetic variation in humans is unlikely to directly impact binding of the AAVs described here to the TfR.

[0502] To evaluate whether the BI-hTfR1 polypeptide has functions outside of the context in which it was originally screened and assessed, Applicants transferred the BI-hTfR1 peptide into AAV7 to create SEQ ID 28. Because the polypeptide may also use additional adjacent AAV9 residues as contact points with TfR1 or contain additional amino acids that stabilize the binding conformation of the BI-hTfR1 7-mer, Applicants also tested a capsid where the BI-hTfR1 7-mer and flanking amino acids (SEQ ID 43) were transferred to AAV7 generating SEQ ID 29. These engineered AAV7 capsids with incorporated TfR1 binding polypeptides were able to transduce CHO cells expressing human TfR1 more effectively than unmodified CHO cells. Notably, the AAV made from SEQ ID 29 performed similarly to BI-hTfR1, while the SEQ ID 28, which only transferred the BI-hTfR1 7-mer without the surrounding residues from AAV9 showed a reduced enhancement in the ability to transduce CHO cells expressing human TfR1. These results indicate that SEQ ID 43 can be used to impart human TfR1 binding and cellular uptake on other macromolecules.

References to Example 2

1. O. J. Müller, F. Kaul, M. D. Weitzman, R. Pasqualini, W. Arap, J. A. Kleinschmidt, M. Trepel, Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. *Nat. Biotechnol.* **21**, 1040–1046 (2003).
2. J. T. Koerber, N. Maheshri, B. K. Kaspar, D. V. Schaffer, Construction of diverse adeno-associated viral libraries for directed evolution of enhanced gene delivery vehicles. *Nat. Protoc.* **1**, 701–706 (2006).
3. W. Li, A. Asokan, Z. Wu, T. Van Dyke, N. DiPrimio, J. S. Johnson, L. Govindaswamy, M. Agbandje-McKenna, S. Leichtle, D. Eugene Redmond Jr, T. J. McCown, K. B. Petermann, N. E. Sharpless, R. J. Samulski, Engineering and Selection of Shuffled AAV Genomes: A New Strategy for Producing Targeted Biological Nanoparticles. *Mol. Ther.* **16**, 1252–1260 (2008).

4. Y. H. Chen, M. Chang, B. L. Davidson, Molecular signatures of disease brain endothelia provide new sites for CNS-directed enzyme therapy. *Nat. Med.* **15**, 1215–1218 (2009).
5. B. E. Deverman, P. L. Pravdo, B. P. Simpson, S. R. Kumar, K. Y. Chan, A. Banerjee, W.-L. Wu, B. Yang, N. Huber, S. P. Pasca, V. Gradinaru, Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nat. Biotechnol.* **34**, 204–209 (2016).
6. M. Nonnenmacher, W. Wang, M. A. Child, X.-Q. Ren, C. Huang, A. Z. Ren, J. Tocci, Q. Chen, K. Bittner, K. Tyson, N. Pande, C. H.-Y. Chung, S. M. Paul, J. Hou, Rapid evolution of blood-brain-barrier-penetrating AAV capsids by RNA-driven biopanning. *Mol Ther Methods Clin Dev.* **20**, 366–378 (2021).
7. K. S. Hanlon, J. C. Meltzer, T. Buzhdygan, M. J. Cheng, M. Sena-Esteves, R. E. Bennett, T. P. Sullivan, R. Razmpour, Y. Gong, C. Ng, J. Nammour, D. Maiz, S. Dujardin, S. H. Ramirez, E. Hudry, C. A. Maguire, Selection of an Efficient AAV Vector for Robust CNS Transgene Expression. *Mol Ther Methods Clin Dev.* **15**, 320–332 (2019).
8. R. S. Ajioka, J. Kaplan, Intracellular pools of transferrin receptors result from constitutive internalization of unoccupied receptors. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 6445–6449 (1986).
9. F. R. Maxfield, T. E. McGraw, Endocytic recycling. *Nat. Rev. Mol. Cell Biol.* **5**, 121–132 (2004).
10. T. Moos, E. H. Morgan, Transferrin and transferrin receptor function in brain barrier systems. *Cell. Mol. Neurobiol.* **20**, 77–95 (2000).
11. W. M. Pardridge, J. Eisenberg, J. Yang, Human blood-brain barrier transferrin receptor. *Metabolism.* **36**, 892–895 (1987).
12. W. A. Jefferies, M. R. Brandon, S. V. Hunt, A. F. Williams, K. C. Gatter, D. Y. Mason, Transferrin receptor on endothelium of brain capillaries. *Nature.* **312**, 162–163 (1984).
13. S. Kida, Y. Koshimura, E. Yoden, A. Yoshioka, H. Morimoto, A. Imakiire, N. Tanaka, S. Tanaka, A. Mori, J. Ito, A. Inoue, R. Yamamoto, K. Minami, T. Hirato, K. Takahashi, H. Sonoda, Enzyme replacement with transferrin receptor-targeted α -L-iduronidase rescues brain pathology in mucopolysaccharidosis I mice. *Mol Ther Methods Clin Dev.* **29**, 439–449 (2023).
14. J. C. Ullman, A. Arguello, J. A. Getz, A. Bhalla, C. S. Mahon, J. Wang, T. Giese, C. Bedard, D. J. Kim, J. R. Blumenfeld, N. Liang, R. Ravi, A. A. Nugent, S. S. Davis, C. Ha, J. Duque, H. L. Tran, R. C. Wells, S. Lianoglou, V. M. Daryani, W. Kwan, H. Solanoy, H. Nguyen,

- T. Earr, J. C. Dugas, M. D. Tuck, J. L. Harvey, M. L. Reyzer, R. M. Caprioli, S. Hall, S. Poda, P. E. Sanchez, M. S. Dennis, K. Gunasekaran, A. Srivastava, T. Sandmann, K. R. Henne, R. G. Thorne, G. Di Paolo, G. Astarita, D. Diaz, A. P. Silverman, R. J. Watts, Z. K. Sweeney, M. S. Kariolis, A. G. Henry, Brain delivery and activity of a lysosomal enzyme using a blood-brain barrier transport vehicle in mice. *Sci. Transl. Med.* **12** (2020), doi:10.1126/scitranslmed.aay1163.
15. S. M. Hammond, F. Abendroth, L. Goli, J. Stoodley, M. Burrell, G. Thom, I. Gurrell, N. Ahlskog, M. J. Gait, M. J. Wood, C. I. Webster, Antibody-oligonucleotide conjugate achieves CNS delivery in animal models for spinal muscular atrophy. *JCI Insight.* **7** (2022), doi:10.1172/jci.insight.154142.
16. M. S. Kariolis, R. C. Wells, J. A. Getz, W. Kwan, C. S. Mahon, R. Tong, D. J. Kim, A. Srivastava, C. Bedard, K. R. Henne, T. Giese, V. A. Assimon, X. Chen, Y. Zhang, H. Solanoy, K. Jenkins, P. E. Sanchez, L. Kane, T. Miyamoto, K. S. Chew, M. E. Pizzo, N. Liang, M. E. K. Calvert, S. L. DeVos, S. Baskaran, S. Hall, Z. K. Sweeney, R. G. Thorne, R. J. Watts, M. S. Dennis, A. P. Silverman, Y. J. Y. Zuchero, Brain delivery of therapeutic proteins using an Fc fragment blood-brain barrier transport vehicle in mice and monkeys. *Sci. Transl. Med.* **12** (2020), doi:10.1126/scitranslmed.aay1359.
17. L. Rué, T. Jaspers, I. M. S. Degors, S. Noppen, D. Schols, B. De Strooper, M. Dewilde, Novel Human/Non-Human Primate Cross-Reactive Anti-Transferrin Receptor Nanobodies for Brain Delivery of Biologics. *Pharmaceutics.* **15** (2023), doi:10.3390/pharmaceutics15061748.
18. Y. J. Yu, J. K. Atwal, Y. Zhang, R. K. Tong, K. R. Wildsmith, C. Tan, N. Bien-Ly, M. Hersom, J. A. Maloney, W. J. Meilandt, D. Bumbaca, K. Gadkar, K. Hoyte, W. Luk, Y. Lu, J. A. Ernst, K. Scarce-Levie, J. A. Couch, M. S. Dennis, R. J. Watts, Therapeutic bispecific antibodies cross the blood-brain barrier in nonhuman primates. *Sci. Transl. Med.* **6**, 261ra154 (2014).
19. B. van Lengerich, L. Zhan, D. Xia, D. Chan, D. Joy, J. I. Park, D. Tatarakis, M. Calvert, S. Hummel, S. Lianoglou, M. E. Pizzo, R. Prorok, E. Thomsen, L. M. Bartos, P. Beumers, A. Capell, S. S. Davis, L. de Weerd, J. C. Dugas, J. Duque, T. Earr, K. Gadkar, T. Giese, A. Gill, J. Gnörich, C. Ha, M. Kannuswamy, D. J. Kim, S. T. Kunte, L. H. Kunze, D. Lac, K. Lechtenberg, A. W.-S. Leung, C.-C. Liang, I. Lopez, P. McQuade, A. Modi, V. O. Torres, H. N. Nguyen, I. Pesämaa, N. Propson, M. Reich, Y. Robles-Colmenares, K. Schlepckow, L. Slemann, H. Solanoy, J. H. Suh, R. G. Thorne, C. Vieira, K. Wind-Mark, K. Xiong, Y. J. Y. Zuchero, D. Diaz, M. S. Dennis, F. Huang, K. Scarce-Levie, R. J. Watts, C. Haass, J. W. Lewcock, G. Di Paolo, M.

- Brendel, P. E. Sanchez, K. M. Monroe, A TREM2-activating antibody with a blood-brain barrier transport vehicle enhances microglial metabolism in Alzheimer's disease models. *Nat. Neurosci.* **26**, 416–429 (2023).
20. P. V. Candelaria, L. S. Leoh, M. L. Penichet, T. R. Daniels-Wells, Antibodies Targeting the Transferrin Receptor 1 (TfR1) as Direct Anti-cancer Agents. *Front. Immunol.* **12**, 607692 (2021).
21. S. J. Barker, M. B. Thayer, C. Kim, D. Tatarakis, M. Simon, R. L. Dial, L. Nilewski, R. C. Wells, Y. Zhou, M. Afetian, A. Chappell, K. S. Chew, J. Chow, A. Clemens, C. B. Discenza, J. Dugas, C. Dwyer, T. Earr, C. Ha, D. Huynh, S. Jayaraman, W. Kwan, C. Mahon, M. Pizzo, E. Roche, L. Sanders, A. Stergioulis, R. Tong, H. Tran, J. Zuchero, A. A. Estrada, K. Gadkar, C. M. M. Koth, P. E. Sanchez, R. G. Thorne, R. J. Watts, T. Sandmann, L. Kane, F. Rigo, M. S. Dennis, J. W. Lewcock, S. L. DeVos, Targeting Transferrin Receptor to Transport Antisense Oligonucleotides Across the Blood-Brain Barrier. *bioRxiv* (2023), p. 2023.04.25.538145.
22. R. Giugliani, A. M. Martins, S. So, T. Yamamoto, M. Yamaoka, T. Ikeda, K. Tanizawa, H. Sonoda, M. Schmidt, Y. Sato, Iduronate-2-sulfatase fused with anti-hTfR antibody, pabinafusp alfa, for MPS-II: A phase 2 trial in Brazil. *Mol. Ther.* **29**, 2378–2386 (2021).
23. T. Okuyama, Y. Eto, N. Sakai, K. Nakamura, T. Yamamoto, M. Yamaoka, T. Ikeda, S. So, K. Tanizawa, H. Sonoda, Y. Sato, A Phase 2/3 Trial of Pabinafusp Alfa, IDS Fused with Anti-Human Transferrin Receptor Antibody, Targeting Neurodegeneration in MPS-II. *Mol. Ther.* **29**, 671–679 (2021).
24. Q. Huang, A. T. Chen, K. Y. Chan, H. Sorensen, A. J. Barry, B. Azari, Q. Zheng, T. Beddow, B. Zhao, I. G. Tobey, C. Moncada-Reid, F.-E. Eid, C. J. Walkey, M. C. Ljungberg, W. R. Lagor, J. D. Heaney, Y. A. Chan, B. E. Deverman, Targeting AAV vectors to the central nervous system by engineering capsid-receptor interactions that enable crossing of the blood-brain barrier. *PLoS Biol.* **21**, e3002112 (2023).

Motif analysis

[0503] X₁X₂S[TS]NGX₃ (SEQ ID NO: 1769), where X₁ is any residue other than D, K, R; X₂ is any residue other than P; and X₃ is a V, I, or R.

[0504] $X_1X_2X_3X_4X_5X_6X_7$, where X_1 and X_2 are aromatic residues (F, W, or Y); X_3 is an S, T, A, H, X_4 is a G, N, H; X_4 is any residue; X_5 is any residue; X_6 is any residue; and X_7 is D, E, A, S, N, H, or P.

[0505] $X_1CX_2X_3CX_4X_5$, where X_1 is L,P,R,A,H,E,S,N, or W; X_2 is G,P,S,T,K,Y,R; X_3 is P,K,D,E,L,G, or S; X_4 is any residue, and X_5 is any residue.

[0506] $FAKX_1X_2X_3X_4$, where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0507] $FARX_1X_2X_3X_4$ where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0508] $FSKX_1X_2X_3X_4$ where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0509] $FSRX_1X_2X_3X_4$ where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0510] $LAKX_1X_2X_3X_4$ where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0511] $LARX_1X_2X_3X_4$ where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0512] $LHKX_1X_2X_3X_4$ where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0513] $LHRX_1X_2X_3X_4$ where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0514] $LSKX_1X_2X_3X_4$ where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0515] $LSRX_1X_2X_3X_4$ where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0516] $YAKX_1X_2X_3X_4$ where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0517] $YARX_1X_2X_3X_4$ where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0518] $YHKX_1X_2X_3X_4$ where X_1 is any residue, X_2 is a G or an N, X_3 is any residue, and X_4 is preferably a N or a D.

[0519] YHRX₁X₂X₃X₄ where X₁ is any residue, X₂ is a G or an N, X₃ is any residue, and X₄ is preferably a N or a D.

[0520] YMKX₁X₂X₃X₄ where X₁ is any residue, X₂ is a G or an N, X₃ is any residue, and X₄ is preferably a N or a D.

[0521] YMRX₁X₂X₃X₄ where X₁ is any residue, X₂ is a G or an N, X₃ is any residue, and X₄ is preferably a N or a D.

[0522] YSKX₁X₂X₃X₄ where X₁ is any residue, X₂ is a G or an N, X₃ is any residue, and X₄ is preferably a N or a D.

[0523] YVKSX₁X₂X₃ (SEQ ID NO: 1770), where X₁ is a D, E, Q, N, or A; X₂ is N, V, T, P, G, S, M, or A; and X₃ is a C, G, I, A, N, R, K, S, or T.

[0524] YVRSX₁X₂X₃ (SEQ ID NO: 1771), where X₁ is a D, E, Q, N, or A; X₂ is N, V, T, P, G, S, M, or A; and X₃ is a C, G, I, A, N, R, K, S, or T.

[0525] YVKTX₁X₂X₃ (SEQ ID NO: 1772), where X₁ is a D, E, Q, N, or A; X₂ is N, V, T, P, G, S, M, or A; and X₃ is a C, G, I, A, N, R, K, S, or T.

[0526] YVRTX₁X₂X₃ (SEQ ID NO: 1773), where X₁ is a D, E, Q, N, or A; X₂ is N, V, T, P, G, S, M, or A; and X₃ is a C, G, I, A, N, R, K, S, or T.

[0527] X₁X₂X₃X₄X₅WP, where X₁ is a R, K, F, I, L, T, Y; and X₂, X₃ and X₄ are any residue; and X₅ is D or E.

[0528] RX₁X₂X₃X₄X₅P, where X₁ and X₂ are any residue; X₃ is D; X₄ is Y, F, V, A, T, S, I, L; and X₅ is Y, F, V, or M.

[0529] RX₁X₂DX₃YP (SEQ ID NO: 1774), where X₁ and X₂ are any residues; and X₃ is A, V, F, T, or S.

[0530] X₁YAKSX₂X₃ (SEQ ID NO: 1775), where X₁ is an S, Y, D, N, or A; X₂ is an L, Y, V, A, Q, or E; and X₃ is an L, A, P, D, N, or I.

[0531] X₁KSSX₂X₃W (SEQ ID NO: 1776), where X₁ is V, P, D, S, L; X₂ is N, G, V, S, H, A, or R; X₃ is V, N, S, H, D, or R.

[0532] **Table S3. Sequence motif families.** The table provides sequence motifs that define specific capsid families and genera, the total number of possible sequence species within the motif, the number of unique 7-mer sequences within each family or genus, and number and fraction of possible sequences from each motif that were tested in the library. Amino acids presented in brackets indicate amino acids acceptable at the specific position. For example, Genus 1.11

FSRLG[AHNLVSPT]N (SEQ ID NO: 1777) has the sequence F-S-R-L-G-X1-N (SEQ ID NO: 1777), where X1 can be any amino acid with the brackets (AHNLVSPT).

Family	Genus	Defining sequence motif	Sequences with the motif family/genus	Assayed	Fraction assayed
Family 1		[YMFL][SHTA][KR][AGILMNQSTV][NG][ATHNFIPLY][IDN]	11520	1608	13.96%
Family 1	Genus 1.1	[YML][SHTA][KR][AGILMNQSTV][N][ATHNFIPLY][IDN]	4320	399	9.24%
Family 1	Genus 1.2	[YML][SHTA][KR][AGILMNQSTV][G][AGFHILNPSTVY][IDN]	5760	825	14.32%
Family 1	Genus 1.3	[LY][AHS][KR][AGHILMNQSTV][GPD]	132	23	17.42%
Family 1	Genus 1.4	[LY][AHS][KR][AGHILMNQSTV][GPN]	132	71	53.79%
Family 1	Genus 1.5	YSK[AILMNQSTV][G]YPTQVFLHSAEDIM[N] (SEQ ID NO: 1778)	126	122	96.83%
Family 1	Genus 1.6	[LY][SRK][VITLA][SA][PRYFHIKWID]	320	42	13.13%
Family 1	Genus 1.7	[FLMY][H][KR][ALM][AGLMNQST][ADEFHILMNQSTVY][IDN]	4224	172	4.07%
Family 1	Genus 1.8	[ILMY][H][KR][ALM][AGLMNQST][ADEFHILMNQSTVY][N]	1584	61	3.85%
Family 1	Genus 1.9	[ILMY][H][KR][ALM][AGLMNQST][ADEFHILMNQSTVY][D]	1584	87	5.49%
Family 1	Genus 1.10	[ILMY][H][KR][L][SQGTNL][PTVIQLA][D]	252	38	15.08%
Family 1	Genus 1.11	FSRLG[AHNLVSP]T[N] (SEQ ID NO: 1777)	8	8	100.00%
Family 1	Genus 1.12	FAR[TSN][GYFH]P[A][N] (SEQ ID NO: 1779)	15	15	100.00%
Family 1	Genus 1.13	FH[KR][LGI][PS][ND] (SEQ ID NO: 1780)	12	5	41.67%
Family 2	Genus 2.1	[YL][HA][KRNA][GQLS][PLINDT][N]	192	17	8.85%
Family 3		[AFHILNPRSTVY][ADEFGHILKLMNQRTSWY][S][ST][NG][IRV]	1224	657	53.68%
Family 3	Genus 3.1	[FLY][DEHNQST][S][ST][NG][IV]	84	63	75.00%
Family 3	Genus 3.2	[VP][STHA][ADEFGHILKLMNQRTSWY][S][ST][NG][IV]	448	315	70.31%
Family 3	Genus 3.3	[AFILPSTVY][DENQST][TS][NGR] (SEQ ID NO: 1781)	90	35	38.89%
Family 3	Genus 3.4	R[EDQT][S][ST][NG][IV] (SEQ ID NO: 1782)	16	9	56.25%
Family 4		R[TLMSGDNE][GEDIFHSAMPVY][D][AVST][YF][AP]	1536	184	11.98%
Family 4	Genus 4.1	R[TMLS][YSAMFFP][D][AVST][YF][P]	224	69	30.80%
Family 4	Genus 4.2	R[TMLS][YSAMFFP][D][AVST][YF][A]	224	13	5.80%
Family 4	Genus 4.3	R[GTDSNE][EDPSG][DTEHN][VAT][YF][P]	900	132	14.67%
Family 4	Genus 4.4	R[GLTDS][DPSG][DEHN][VT][YF][P]	320	53	16.56%
Family 4	Genus 4.5	TR[KYW][EWYQST][GD][PA][AP]	96	11	11.46%
Family 5		LC[KR][PC][LSDANQHPV][ETGADNS] (SEQ ID NO: 1783)	126	32	25.40%
Family 6	Genus 6.1	[YF][WFX][TS][G][ITVQMH][I][PHLMAQTV][ASDE]	2592	445	17.17%

Family6	Genus6.2	[YF][WFX][TS][G][TVQMH][KR][ASDE]	576	89	15.45%
Family6	Genus6.3	[YF][WFX][TS][G][KR][PHLMAQT][ASDE]	864	201	23.26%
Family6	Genus6.4	YFTG[KRQMH][TRHKVLJE (SEQ ID NO: 1784)	36	27	75.00%
Family6	Genus6.5	YF[TSHA][GKRT][PHLMAQT][DN]	192	58	30.21%
Family6	Genus6.6	YWTG[KMVT][PVIHQTML][ED] (SEQ ID NO: 1785)	64	37	57.81%
Family6	Genus6.7	[YF][SHAQ][GKQR][VILKHRQM][E]	192	60	31.25%
Family7	Genus7.1	[KR][DSN][EFTKSRV][TSGV][FWY]	360	77	21.39%
Family7	Genus7.2	[KR][DD][FY][TSVH][TSGVA][FWY]	240	68	28.33%
Family7	Genus7.3	[SRGNTMQ][DG][TVSNY][SKR][W (SEQ ID NO: 1786)	70	32	45.71%
Family7	Genus7.4	[RVD][QK][MLG][SEARQ][DR][TASGKN][W	540	44	8.15%
Family7	Genus7.5	[DIEQV][SK][LMGP][EASDQT][SA][R][DSETGA][W	2016	194	9.62%
Family7	Genus7.6	G[ESPGA][DEPN][GHTSN][V][RKS][WY]	600	60	10.00%
Family7	Genus7.7	RE[DEPN][GHTSN][V][RKS][WY]	120	17	14.17%
Family7	Genus7.8	G[GS][GESAPD][TGS][S][THKRAN][W	252	74	29.37%
Family8		[VIRND][AGS][L][TSHG][KRE][GIWRA][IDG]	1800	50	2.78%
Family9		[LMW][FRWKTY][IDS][GT][PGSNAR][APSY]	864	56	6.48%
Family10		[PNK][YF][ARK][SPVARILSE][EDML]	384	23	5.99%

Example 3 – Improved polypeptides that bind human Transferrin Receptor and facilitate AAV entry into the CNS

Second generation BI-hTFR capsids out perform BI-hTFR1 and AAV9 in TFRC KI mice

[0533] Applicants next assessed the capabilities of TfR1-binding capsids BI-hTFR1v2 (loop VIII modification, SEQ ID 170; EAQYSRIGPNNQAAQ), BI-hTFR1v3 (loop VIII modification, SEQ ID 243; GAQYSRIGPNPQPE), and BI-hTFR2v2 (loop VIII modification, SEQ ID 1215; SADFRSTNGVAQAE) compared to AAV9 (unmodified loop VIII, SEQ ID 41; AQAQAQ), BI-hTFR1 (loop VIII modification, SEQ ID 45; SAQYSRIGPNAQAQ), and AAV-PHP.eB (loop VIII modification, SEQ ID 43; SDGTLAVPFKAQAQ) when dosed intravenously (retro-orbital route) into adult humanized *TFRC* knock-in mice (B-hTFR1 mice, Biocytogen). Each capsid was used to package a single-stranded AAV-CAG-NLS-mScarlet-P2A-ffLuciferase-WPRE-pA genome and was intravenously (IV) injected into 10 to 16-week-old humanized *TFRC* knock-in mice at a dose of 2e12 vg/kg. Animals were perfused with PBS 21 days later and tissues were collected for biodistribution, luciferase, and immunohistochemistry.

Biodistribution

[0534] BI-hTFR1, BI-hTFR1v2, BI-hTFR1v3 and BI-hTFR2v2 showed an increased brain biodistribution (vg/diploid genome) of 86-fold, 170-fold, 206-fold, and 131-fold viral genomes relative to AAV9, respectively (**FIG. 22**). In the liver, BI-hTFR1 delivered 1.4-fold more viral genomes per cell than AAV9, while the second generation capsids, BI-hTFR1v2, BI-hTFR1v3 and BI-hTFR2v2, delivered 3.0-fold, 3.9-fold, and 21-fold fewer viral genomes per cell respectively relative to AAV9. BI-hTFR1v2 achieved an average of 2.80 vg/cell, 5.20 vg/cell, or 4.74 vg/cell in the brain, spinal cord, or liver, respectively, indicating that BI-hTFR1v2 could be well suited for the delivery of single or dual vector cargo for CNS gene therapy or CNS gene editing applications.

Transduction (luciferase activity)

[0535] To assess gene expression from the TfR1-binding capsids, Applicants next measured the amount of luciferase activity present within tissue lysates made from the same cohort. Relative to AAV9, BI-hTFR1, BI-hTFR1v2, BI-hTFR1v3 and BI-hTFR2v2 provided a 152-fold, 1200-fold, 1080-fold, or 465-fold increase, respectively, in luciferase expression in whole brain lysate (**FIG. 23**). Luciferase expression was determined by measuring the relative luminescence units (RLU) per microgram of protein lysate (Revvity - 6066761 & Thermo Scientific - 23227). In the

liver, BI-hTFR1, BI-hTFR1v2, BI-hTFR1v3 and BI-hTFR2v2, showed a 3.1-fold, 1.2-fold, 3.3-fold and 4.3-fold decrease in RLU per microgram of protein lysate relative to AAV9, respectively.

BI-hTFR1v2 efficiently transduces CNS neurons.

[0536] To assess the ability of the TFR1 binding capsids to transduce individual cell types throughout the CNS, Applicants assessed mScarlet expression in whole sagittal sections of brains from TFRC KI mice for each of the capsids. Notably, BI-hTFR1v2 and BI-hTFR1v3 provided greater mScarlet fluorescence across the CNS as compared with AAV9 or BI-hTFR1 (**FIG. 24**). In the cortex, BI-hTFR1v2 and BI-hTFR1v3 were more efficient at transducing NeuN⁺ cells as compared to AAV9 or BI-hTFR1 (**FIG. 25**). BI-hTFR1v2 transduced 80% and 55% of neurons in the cortex and striatum, respectively (**FIG. 26**). AAV9 transduced only the occasional NeuN⁺ cell. This work establishes these human Tfr1-targeting capsids as promising vectors for human CNS gene therapy.

[0537] Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the disclosure will be apparent to those skilled in the art without departing from the scope and spirit of the disclosure. Although the disclosure has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the disclosure as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the disclosure that are obvious to those skilled in the art are intended to be within the scope of the disclosure. This application is intended to cover any variations, uses, or adaptations of the disclosure following, in general, the principles of the disclosure and including such departures from the present disclosure come within known customary practice within the art to which the disclosure pertains and may be applied to the essential features herein before set forth.

[0538] Citation or identification of any document in this application is not an admission that such a document is available as prior art to the present disclosure.

[0539] References and citations to other documents, such as patents, patent applications, patent publications, journals, books, papers, web contents, have been made in this disclosure. All such documents are hereby incorporated herein by reference in their entirety for all purposes. Any material, or portion thereof, that is said to be incorporated by reference herein, but which conflicts with existing definitions, statements, or other disclosure material explicitly set forth herein is only

incorporated to the extent that no conflict arises between that incorporated material and the present disclosure material. In the event of a conflict, the conflict is to be resolved in favor of the present disclosure as the preferred disclosure.

Table 1 Referenced in Example 2

AA Full	SEQ ID	Brain	Spinal Cord	Liver	AA Full	SEQ ID	Brain	Spinal Cord	Liver
SAQTLAVPFKAQAQ	42	+++	++	----	SAQYSRIGPNIQCQ	628	++	++	----
SAQTLAVPFKAQAQ	42	+++	+++	----	SAQYSRIGPNIQCQ	628	++	+	similar
SDGTLAVPFKAQAQ	43	++	++	----	SAQYSRIGPNIQCQ	629	+	+	----
SDGTLAVPFKAQAQ	43	++	++	----	SAQYSRIGPNIQFQ	630	+	+	--
SAQKSVGSVYAQAQ	44	++	+	----	SAQYSRIGPNIQFQ	630	+	+	+
SAQKSVGSVYAQAQ	44	++	++	----	SAQYSRIGPNIQSQ	631	++	+	----
SAQYSRIGPNAQAQ	45	+	+	similar	SAQYSRIGPNIQSQ	631	++	++	----
SAQYSRIGPNAQAQ	45	+	+	+	SAQYSRIGPNIQTQ	632	++	++	----
AAQYSRIGPNAQAQ	60	++	++	----	SAQYSRIGPNIQTQ	632	++	++	----
AAQYSRIGPNAQAQ	60	+	+	----	SAQYSRIGPNIQVQ	633	++	+	----
AAQYSRIGPNAEAQ	61	+	+	----	SAQYSRIGPNIQVQ	633	+	+	----
AAQYSRIGPNAEAQ	61	+	+	----	SAQYSRIGPNIQAQ	634	++	++	----
AAQYSRIGPNAQAQ	62	+++	++	--	SAQYSRIGPNIQAQ	634	+	+	----
AAQYSRIGPNAQAQ	62	++	++	----	SAQYSRIGPNIQAQ	636	+++	++	----
AAQYSRIGPNAQAQ	63	+	+	----	SAQYSRIGPNIQAQ	636	+++	+++	----
AAQYSRIGPNAQAQ	64	+++	+++	----	SAQYSRIGPNIQAQ	638	+	-	----
AAQYSRIGPNAQAQ	64	+++	+++	----	SAQYSRIGPNIQAQ	639	+	+	--
AAQYSRIGPNAQAQ	65	++	++	+	SAQYSRIGPNIQAQ	639	+	+	----
AAQYSRIGPNAQAQ	65	++	++	+	SAQYSRIGPNIQAQ	640	+	-	----
AAQYSRIGPNAQAQ	66	+	+	similar	SAQYSRIGPNIQAQ	641	+	+	similar
AAQYSRIGPNAQAQ	66	+	+	similar	SAQYSRIGPNIQAQ	642	++	++	----
AAQYSRIGPNAQAQ	67	++	++	--	SAQYSRIGPNIQAQ	642	++	++	--
AAQYSRIGPNAQAQ	67	++	++	----	SAQYSRIGPNIQAQ	643	++	+	----
AAQYSRIGPNAQAQ	68	++	+	----	SAQYSRIGPNIQAQ	643	+	+	----
AAQYSRIGPNAQAQ	68	++	++	----	SAQYSRIGPNIQAQ	644	+	+	----
AAQYSRIGPNAQAQ	69	++	++	----	SAQYSRIGPNIQAQ	644	+	+	----
AAQYSRIGPNAQAQ	69	++	++	similar	SAQYSRIGPNIQAQ	645	+	+	----
AAQYSRIGPNAQAQ	70	+	-	----	SAQYSRIGPNIQAQ	645	+	++	----
AAQYSRIGPNAQAQ	70	+	+	----	SAQYSRIGPNIQAQ	646	++	+	----

AAQYSRIGPNAQAQ	71	++	+	similar	SAQYSRIGPNNQIQ	646	+	-	----
AAQYSRIGPNAQAQ	71	+	+	similar	SAQYSRIGPNNMQ	648	+	+	----
AAQYSRIGPNAQAS	72	++	++	similar	SAQYSRIGPNNMQ	648	+	+	----
AAQYSRIGPNAQAS	72	++	++	+	SAQYSRIGPNNMPAQ	649	+	+	-
AAQYSRIGPNAQAT	73	+++	+++	similar	SAQYSRIGPNNMQAA	650	+	+	----
AAQYSRIGPNAQAT	73	+++	+++	similar	SAQYSRIGPNNMQAA	650	+	+	similar
AAQYSRIGPNAQAV	74	+	++	---	SAQYSRIGPNNMQAG	651	+	+	----
AAQYSRIGPNAQAV	74	+	++	-	SAQYSRIGPNNMQAG	651	+	+	----
AAQYSRIGPNAQCCQ	75	+++	++	----	SAQYSRIGPNNMQAL	652	+	+	----
AAQYSRIGPNAQCCQ	75	+++	++	----	SAQYSRIGPNNMQAL	652	+	-	----
AAQYSRIGPNAQGGQ	76	+	+	----	SAQYSRIGPNNMQAQ	653	+	+	----
AAQYSRIGPNAQGGQ	76	+	+	----	SAQYSRIGPNNMQAQ	653	+	+	similar
AAQYSRIGPNAQPQ	77	++	++	-	SAQYSRIGPNNMQAS	654	++	+	-
AAQYSRIGPNAQPQ	77	++	++	----	SAQYSRIGPNNMQAS	654	++	+	similar
AAQYSRIGPNAQSQ	78	++	++	----	SAQYSRIGPNNMQAT	655	+++	++	----
AAQYSRIGPNAQSQ	78	++	++	-	SAQYSRIGPNNMQAT	655	++	++	similar
AAQYSRIGPNAQITQ	79	+++	+++	----	SAQYSRIGPNNMQCQ	656	++	+	----
AAQYSRIGPNAQITQ	79	+++	++	----	SAQYSRIGPNNMQCQ	656	+	+	----
AAQYSRIGPNAQIVQ	80	++	++	similar	SAQYSRIGPNNMQPQ	657	+	+	----
AAQYSRIGPNAQIVQ	80	++	+	----	SAQYSRIGPNNMQSQ	658	+	+	----
AAQYSRIGPNAQASQ	81	+	+	----	SAQYSRIGPNNMQSQ	658	+	+	----
AAQYSRIGPNAQASQ	81	+	+	----	SAQYSRIGPNNMQITQ	659	++	++	----
AAQYSRIGPNAQATQ	83	++	++	----	SAQYSRIGPNNMQITQ	659	+	++	----
AAQYSRIGPNAQATQ	83	++	+++	----	SAQYSRIGPNNAAQ	663	+++	+++	----
AAQYSRIGPNDQDAQ	88	++	++	----	SAQYSRIGPNNAAQ	663	+++	+++	-
AAQYSRIGPNDQDAQ	88	++	++	similar	SAQYSRIGPNNDAQ	665	+++	+++	----
AAQYSRIGPNEQAAQ	89	++	++	similar	SAQYSRIGPNNDAQ	665	+++	+++	----
AAQYSRIGPNEQAAQ	89	++	++	similar	SAQYSRIGPNEAQ	668	+++	+++	----
AAQYSRIGPNEHQAAQ	90	++	-	----	SAQYSRIGPNEAQ	668	+++	+++	----
AAQYSRIGPNIQAAQ	92	++	+	-	SAQYSRIGPNNFAQ	674	++	-	----
AAQYSRIGPNIQAAQ	92	++	+	+	SAQYSRIGPNNGAQ	675	+++	+++	----

AAQYSRIGPNNAQ	95	+++	+++	+++	+	SAQYSRIGPNNAQ	675	+++	+++	+++	----
AAQYSRIGPNNAQ	95	+++	+++	+++	----	SAQYSRIGPNNAQ	676	++	++	-	----
AAQYSRIGPNPQAQ	101	++++	++++	++++	similar	SAQYSRIGPNNAQ	677	++	-	-	----
AAQYSRIGPNPQAQ	101	+++	+++	+++	----	SAQYSRIGPNNAQ	678	++++	++++	++++	----
AAQYSRIGPNQQAQ	106	+	+	+	----	SAQYSRIGPNNAQ	678	+++	+++	+++	----
AAQYSRIGPNQQAQ	106	+	+	+	----	SAQYSRIGPNNAQ	679	+++	+++	+++	----
AAQYSRIGPNQQAQ	107	+	+	+	----	SAQYSRIGPNNAQ	679	+++	+++	+++	similar
AAQYSRIGPNQQAQ	107	+	+	+	----	SAQYSRIGPNNAQ	680	++++	++++	++++	----
AAQYSRIGPNYQAQ	109	++	-	-	----	SAQYSRIGPNNAQ	680	++++	++++	++++	----
AAQYSRIGPNYQAQ	109	+	-	-	----	SAQYSRIGPNNAQ	681	++++	++++	++++	----
AFQYSRIGPNNAQ	110	+++	+++	+++	----	SAQYSRIGPNNAQ	681	++++	++++	++++	similar
AFQYSRIGPNNAQ	110	++	++	++	----	SAQYSRIGPNNAQ	682	+	+	+	----
AMQYSRIGPNNAQ	111	+++	+++	+++	----	SAQYSRIGPNNAQ	682	+	+	+	----
AMQYSRIGPNNAQ	111	++	++	++	----	SAQYSRIGPNNAQ	683	++	++	++	similar
APQYSRIGPNNAQ	112	+++	+++	+++	----	SAQYSRIGPNNAQ	683	++	++	++	similar
APQYSRIGPNNAQ	112	+++	+++	+++	----	SAQYSRIGPNNAQ	684	+++	+++	+++	similar
AVQYSRIGPNNAQ	116	+	-	-	----	SAQYSRIGPNNAQ	684	+++	+++	+++	----
AVQYSRIGPNNAQ	116	+	-	-	----	SAQYSRIGPNNAQ	685	+++	+++	+++	----
AYQYSRIGPNNAQ	117	+	+	+	----	SAQYSRIGPNNAQ	685	++	+	+	----
CAQYSRIGPNATAQ	118	+	+	+	----	SAQYSRIGPNNAQ	686	+++	+++	+++	----
CAQYSRIGPNATAQ	118	+	+	+	----	SAQYSRIGPNNAQ	686	+++	+++	+++	similar
CAQYSRIGPNQQAQ	119	+	+	+	----	SAQYSRIGPNNAQ	687	++	+	+	----
CAQYSRIGPNPQAQ	120	++	+	+	----	SAQYSRIGPNNAQ	687	++	++	+++	----
CAQYSRIGPNPQAQ	120	+	+	+	----	SAQYSRIGPNNAQ	688	+++	+++	++++	----
EAQYSRIGPNATAQ	122	+	-	-	----	SAQYSRIGPNNAQ	688	+++	+++	+++	----
EAQYSRIGPNATAQ	123	+	+	+	----	SAQYSRIGPNNAQ	689	+++	+++	+++	----
EAQYSRIGPNATAQ	124	+	+	+	----	SAQYSRIGPNNAQ	689	+++	+++	+++	----
EAQYSRIGPNATAQ	124	+	+	+	----	SAQYSRIGPNNAQ	690	+++	+++	+++	----
EAQYSRIGPNATAQ	125	++	++	++	----	SAQYSRIGPNNAQ	690	+++	+++	+++	----
EAQYSRIGPNATAQ	125	++	++	++	----	SAQYSRIGPNNAQ	691	+++	+++	+++	----
EAQYSRIGPNATAQ	126	++	++	++	----	SAQYSRIGPNNAQ	691	+++	+++	+++	----

EAQYSR(GPNAQAG	126	++	++	++				SAQYSR(GPNNQAS	692	++++	+++	++++
EAQYSR(GPNAQAI	127	++	++	++				SAQYSR(GPNNQAS	692	++++	++++	similar
EAQYSR(GPNAQAI	127	+		++				SAQYSR(GPNNQAT	693		+++	----
EAQYSR(GPNAQAL	128	++	++	++				SAQYSR(GPNNQAT	693		++++	similar
EAQYSR(GPNAQAL	128	++	++	++				SAQYSR(GPNNQAV	694	++++	+++	++++
EAQYSR(GPNAQAM	129	+		+				SAQYSR(GPNNQAV	694		+++	++++
EAQYSR(GPNAQAM	129	+		+				SAQYSR(GPNNQAY	695		++	++++
EAQYSR(GPNAQAN	130	++	++	++				SAQYSR(GPNNQAY	695	++	-	-----
EAQYSR(GPNAQAN	130	++	++	++				SAQYSR(GPNNQCQ	696	++++	++++	----
EAQYSR(GPNAQAQ	131	++	++	++				SAQYSR(GPNNQCQ	696	++++	++++	similar
EAQYSR(GPNAQAQ	131	+		+				SAQYSR(GPNNQFQ	697	+++	+++	++++
EAQYSR(GPNAQAR	132	+		-				SAQYSR(GPNNQFQ	697	+++	+++	----
EAQYSR(GPNAQAS	133	++	++	++				SAQYSR(GPNNQGQ	698	++++	++++	++++
EAQYSR(GPNAQAS	133	++	++	++				SAQYSR(GPNNQGQ	698	++++	++++	similar
EAQYSR(GPNAQAT	134	++	++	+++				SAQYSR(GPNNQHQ	699	++++	+++	++++
EAQYSR(GPNAQAT	134	++	++	+++				SAQYSR(GPNNQHQ	699	++++	++++	++++
EAQYSR(GPNAQAV	135	++	++	++				SAQYSR(GPNNQIQ	700	+++	+++	++++
EAQYSR(GPNAQAV	135	++	++	++				SAQYSR(GPNNQIQ	700	+++	+++	----
EAQYSR(GPNAOCQ	136	++	++	++				SAQYSR(GPNNQPQ	701	++++	++++	++++
EAQYSR(GPNAOCQ	136	++	++	++				SAQYSR(GPNNQPQ	701	+++	+++	++++
EAQYSR(GPNAQPP	137	+		+				SAQYSR(GPNNQQQ	702	+++	+++	++++
EAQYSR(GPNAQPP	137	+		++				SAQYSR(GPNNQQQ	702	+++	+++	++++
EAQYSR(GPNAQSQ	138	++	++	++				SAQYSR(GPNNQSQ	703	++++	++++	similar
EAQYSR(GPNAQSQ	138	++	++	++				SAQYSR(GPNNQSQ	703	++++	++++	++++
EAQYSR(GPNAQTQ	139	++	++	++				SAQYSR(GPNNQTQ	704	+++++	+++++	++++
EAQYSR(GPNAQTQ	139	+		+				SAQYSR(GPNNQTQ	704	++++	++++	++++
EAQYSR(GPNAQVQ	140	+		+				SAQYSR(GPNNQVQ	705	++++	+++++	----
EAQYSR(GPNAQVQ	140	+		+				SAQYSR(GPNNQVQ	705		++++	++++
EAQYSR(GPNAFAQ	141	+		+				SAQYSR(GPNNQWQ	706	+++	+++	++++
EAQYSR(GPNEAQ	142	+		+				SAQYSR(GPNNQWQ	706	++	+++	++++
EAQYSR(GPNEAQ	142	+		+				SAQYSR(GPNNQYQ	707	+++	++	++++

EAQYSRIGPNHQAQ	143	+++	+++	+++	-----	SAQYSRIGPNNQYQ	707	+++	++++	-----
EAQYSRIGPNHQAQ	143	+++	+++	++++	-----	SAQYSRIGPNNSAQ	708	++++	++++	-----
EAQYSRIGPNIOAQ	145	+++	+++	+++	-----	SAQYSRIGPNNSAQ	708	++++	++++	similar
EAQYSRIGPNIOAQ	145	+++	+++	+++	-----	SAQYSRIGPNNTAQ	710	++++	+++	-----
EAQYSRIGPNKQAQ	146	++	++	+	-----	SAQYSRIGPNNTAQ	710	+++	+++	-----
EAQYSRIGPNKQAQ	146	++	++	+	-----	SAQYSRIGPNNAQ	720	+++	-	-----
EAQYSRIGPNNAQ	170	++++	++++	++++	-----	SAQYSRIGPNPAQ	722	++++	++++	-----
EAQYSRIGPNNAQ	170	++++	++++	++++	-	SAQYSRIGPNPAQ	722	+++	++++	-----
EAQYSRIGPNPQAQ	198	+	+	++	-----	SAQYSRIGPNPEAQ	730	+++	+++	-----
EAQYSRIGPNPQAQ	198	+	+	++	-----	SAQYSRIGPNPEAQ	730	+++	+++	-----
EAQYSRIGPNQQAQ	199	+	+	+	-----	SAQYSRIGPNPIAQ	743	+++	+	-----
EAQYSRIGPNQQAQ	199	+	+	++	-----	SAQYSRIGPNPIAQ	743	+	-	-----
EAQYSRIGPNRQAQ	200	+	+	-	-----	SAQYSRIGPNPLAQ	746	+++	+++	-----
EAQYSRIGPNSQAQ	202	+	+	+	-----	SAQYSRIGPNPMAQ	749	+++	++++	-----
EAQYSRIGPNSQAQ	202	+	+	+	-----	SAQYSRIGPNPMAQ	749	+++	++++	-----
EAQYSRIGPNYQAQ	203	+++	+++	+++	-----	SAQYSRIGPNPNAQ	752	+++	+++	-----
EAQYSRIGPNYQAQ	203	++	++	+++	-----	SAQYSRIGPNPNAQ	752	+++	+++	-----
ECQYSRIGPNAQAQ	204	+	+	+	-----	SAQYSRIGPNPQAA	756	++++	++++	similar
ECQYSRIGPNAQAQ	204	+	+	++	-----	SAQYSRIGPNPQAA	756	++++	++++	-----
EPQYSRIGPNAQAQ	205	+	+	++	-----	SAQYSRIGPNPQAC	757	++	+	-----
EPQYSRIGPNAQAQ	205	+	+	++	-----	SAQYSRIGPNPQAC	757	+	+	-----
EQYSRIGPNAQAQ	207	++	++	++	-----	SAQYSRIGPNPQAD	758	++	+++	-
EQYSRIGPNAQAQ	207	+	+	+	-----	SAQYSRIGPNPQAD	758	++	++	-----
EMQYSRIGPNAQAQ	210	++	++	+	-----	SAQYSRIGPNPQAE	759	+++	++++	-----
EMQYSRIGPNAQAQ	210	+	+	+	-----	SAQYSRIGPNPQAE	759	+++	++++	-----
EPQYSRIGPNAQAQ	212	++	++	++	-----	SAQYSRIGPNPQAG	760	++++	++++	-----
EPQYSRIGPNAQAQ	212	+	+	++	-----	SAQYSRIGPNPQAG	760	++++	++++	-----
ETQYSRIGPNAQAQ	216	+	+	+	-----	SAQYSRIGPNPQAI	761	++++	+++	-----
ETQYSRIGPNAQAQ	216	+	+	+	-----	SAQYSRIGPNPQAI	761	+++	++++	-----
EVQYSRIGPNAQAQ	224	++	++	+	-----	SAQYSRIGPNPQAL	762	++++	++++	-----
EVQYSRIGPNAQAQ	224	++	++	++	-----	SAQYSRIGPNPQAL	762	++++	++++	-----

FAQYSRIGPNNQAAQ	226	+	+	+	+	+	SAQYSRIGPNNQAAQ	763	++++	++++	++++	++++	++++
GAQYSRIGPNNQAAQ	227	+	+	++	++	++	SAQYSRIGPNNQAAQ	763	++	++	++	++	++
GAQYSRIGPNNQAAQ	227	+	+	+	+	+	SAQYSRIGPNNQAAQ	764	++++	++++	++++	++++	++++
GAQYSRIGPNNQAAQ	229	+	+	+	+	+	SAQYSRIGPNNQAAQ	764	++++	++++	++++	++++	++++
GAQYSRIGPNNQAAQ	229	+	+	+	+	+	SAQYSRIGPNNQAAQ	765	++++	++++	++++	++++	++++
GAQYSRIGPNNQAAQ	239	++++	++++	++++	++++	++++	SAQYSRIGPNNQAAQ	765	++++	++++	++++	++++	+
GAQYSRIGPNNQAAQ	239	+++	+++	+++	+++	+++	SAQYSRIGPNNQAAQ	766	++++	++++	++++	++++	similar
HAQYSRIGPNNQAAQ	254	++	++	+	+	+	SAQYSRIGPNNQAAQ	766	++++	++++	++++	++++	similar
HAQYSRIGPNNQAAQ	254	++	++	+	+	+	SAQYSRIGPNNQAAQ	767	++++	++++	++++	++++	++++
IAQYSRIGPNNQAAQ	256	+	+	+	+	+	SAQYSRIGPNNQAAQ	767	++++	++++	++++	++++	++++
IAQYSRIGPNNQAAQ	256	+	+	+	+	+	SAQYSRIGPNNQAAQ	768	++++	++++	++++	++++	++++
IAQYSRIGPNNQAAQ	257	++	++	++	++	++	SAQYSRIGPNNQAAQ	768	++++	++++	++++	++++	++++
IAQYSRIGPNNQAAQ	257	++	++	-	-	-	SAQYSRIGPNNQAAQ	769	++	++	++	++	+
IAQYSRIGPNNQAAQ	258	++	++	++	++	++	SAQYSRIGPNNQAAQ	770	++++	++++	++++	++++	++++
IAQYSRIGPNNQAAQ	258	++	++	++	++	++	SAQYSRIGPNNQAAQ	770	++++	++++	++++	++++	++++
IAQYSRIGPNNQAAQ	259	++	++	++	++	++	SAQYSRIGPNNQAAQ	772	++++	++++	++++	++++	++++
IAQYSRIGPNNQAAQ	259	+	+	+	+	+	SAQYSRIGPNNQAAQ	772	++	++	++	++	++++
IAQYSRIGPNNQAAQ	260	++	++	++	++	++	SAQYSRIGPNNQAAQ	774	++++	++++	++++	++++	++++
IAQYSRIGPNNQAAQ	260	++	++	++	++	++	SAQYSRIGPNNQAAQ	774	++++	++++	++++	++++	++++
KAQYSRIGPNNQAAQ	261	++	++	++	++	++	SAQYSRIGPNNQAAQ	776	++++	++++	++++	++++	++++
KAQYSRIGPNNQAAQ	261	++	++	++	++	++	SAQYSRIGPNNQAAQ	776	++	++	++	++	++++
KAQYSRIGPNNQAAQ	263	+++	+++	+++	+++	+++	SAQYSRIGPNNQAAQ	778	++++	++++	++++	++++	++++
KAQYSRIGPNNQAAQ	263	++	++	++	++	++	SAQYSRIGPNNQAAQ	778	++++	++++	++++	++++	++++
LAQYSRIGPNNQAAQ	266	+	+	-	-	-	SAQYSRIGPNNQAAQ	780	++++	++++	++++	++++	++++
LAQYSRIGPNNQAAQ	267	+	+	+	+	+	SAQYSRIGPNNQAAQ	780	++++	++++	++++	++++	++++
LAQYSRIGPNNQAAQ	267	+	+	+	+	+	SAQYSRIGPNNQAAQ	782	++++	++++	++++	++++	++++
LAQYSRIGPNNQAAQ	268	+	+	+	+	+	SAQYSRIGPNNQAAQ	782	++	++	++	++	+
LAQYSRIGPNNQAAQ	269	++	++	++	++	++	SAQYSRIGPNNQAAQ	784	++++	++++	++++	++++	++++
LAQYSRIGPNNQAAQ	269	++	++	++	++	++	SAQYSRIGPNNQAAQ	784	++++	++++	++++	++++	++++
LAQYSRIGPNNQAAQ	270	++	++	-	-	-	SAQYSRIGPNNQAAQ	786	++++	++++	++++	++++	++++
LAQYSRIGPNNQAAQ	270	+	+	+	+	+	SAQYSRIGPNNQAAQ	786	++++	++++	++++	++++	++++

LAQYSRIGPNNAQ	271	+++	++	----	SAQYSRIGPNQVQ	788	+++++	++++	----
LAQYSRIGPNNAQ	271	++	+	----	SAQYSRIGPNQVQ	788	++++	++++	----
LAQYSRIGPNPQAQ	272	++	++	----	SAQYSRIGPNPQWQ	789	++	+	----
LAQYSRIGPNPQAQ	272	+	++	----	SAQYSRIGPNPQWQ	789	+	+	----
LAQYSRIGPNQQAQ	273	++	++	----	SAQYSRIGPNPQYQ	790	+++	+++	----
LAQYSRIGPNVQAQ	274	++	+	----	SAQYSRIGPNPQYQ	790	+++	+++	----
LAQYSRIGPNVQAQ	274	+	+	----	SAQYSRIGPNPSAQ	792	+++	+++	----
LPQYSRIGPNQAQ	275	++	-	----	SAQYSRIGPNPSAQ	792	+++	++++	----
LPQYSRIGPNQAQ	275	++	+	----	SAQYSRIGPNPTAQ	800	+++	++++	----
MAQYSRIGPNMAQ	276	+	+	----	SAQYSRIGPNPTAQ	800	+++	+++	----
MAQYSRIGPNAPAQ	277	+	+	----	SAQYSRIGPNPVAQ	808	++++	-	----
MAQYSRIGPNAPAQ	277	+	+	----	SAQYSRIGPNPVAQ	808	++	++	----
MAQYSRIGPNQAQA	278	+	+	similar	SAQYSRIGPNQMAQ	815	+	+	----
MAQYSRIGPNQAAS	279	+	+	similar	SAQYSRIGPNQMAQ	815	+	+	similar
MAQYSRIGPNQAAS	279	+	+	----	SAQYSRIGPNQQAQ	816	+	+	----
MAQYSRIGPNQAAT	280	++	+	similar	SAQYSRIGPNQQAQ	816	+	+	similar
MAQYSRIGPNQAAT	280	+	+	similar	SAQYSRIGPNQQAAS	817	+	+	----
MAQYSRIGPNQCCQ	281	+	+	----	SAQYSRIGPNQQAAS	817	+	+	+
MAQYSRIGPNQAQITQ	282	+	+	----	SAQYSRIGPNQQAAT	818	++	+	similar
MAQYSRIGPNQAQITQ	282	+	+	----	SAQYSRIGPNQQAAT	818	++	+	similar
MAQYSRIGPNATAQ	283	+	-	----	SAQYSRIGPNQQAQV	819	+	+	similar
MAQYSRIGPNATAQ	283	+	++	----	SAQYSRIGPNQQAQV	819	+	+	similar
MAQYSRIGPNEQAQ	285	++	++	similar	SAQYSRIGPNQQAQV	820	++	+	similar
MAQYSRIGPNEQAQ	285	+	+	similar	SAQYSRIGPNQQAQV	820	+	+	similar
MAQYSRIGPNNQAQ	286	++	++	----	SAQYSRIGPNQQAQV	821	+	+	----
MAQYSRIGPNNQAQ	286	++	++	----	SAQYSRIGPNQQAQV	821	+	+	----
MAQYSRIGPNPQAQ	288	++	+++	----	SAQYSRIGPNQQAQV	822	+	+	----
MAQYSRIGPNPQAQ	288	++	++	----	SAQYSRIGPNQQAQV	822	+	+	----
MAQYSRIGPNQAQ	289	+	+	----	SAQYSRIGPNQQAQV	823	+	+	+
MAQYSRIGPNQAQ	289	+	+	----	SAQYSRIGPNQQAQV	823	+	+	+
MPQYSRIGPNQAQ	290	++	+	----	SAQYSRIGPNQQAQV	824	++	++	similar

MPQYSRIGPNAAQ	290	+	+	+	----	SAQYSRIGPNQQTQ	824	++	++	similar
NAQYSRIGPNAMAQ	291	+	-	-	----	SAQYSRIGPNQQVQ	825	+	+	----
NAQYSRIGPNAPAQ	292	++	+	+	----	SAQYSRIGPNQTAQ	826	++	++	----
NAQYSRIGPNAPAQ	292	+	+	+	----	SAQYSRIGPNQTAQ	826	++	++	----
NAQYSRIGPNAAQT	293	+	+	+	similar	SAQYSRIGPNMAQ	829	++	+	----
NAQYSRIGPNAAQT	293	+	+	+	similar	SAQYSRIGPNMAQ	829	+	+	--
NAQYSRIGPNATAQ	294	++	++	++	----	SAQYSRIGPNPAQ	830	+++	+++	----
NAQYSRIGPNATAQ	294	+	+	++	----	SAQYSRIGPNPAQ	830	+++	+++	similar
NAQYSRIGPNDAQ	296	+	+	++	----	SAQYSRIGPNQAQ	831	+	+	similar
NAQYSRIGPNDAQ	296	+	+	+	----	SAQYSRIGPNQAQ	831	+	+	+
NAQYSRIGPNIAQ	298	++	+	+	----	SAQYSRIGPNQAE	832	+	+	similar
NAQYSRIGPNIAQ	298	+	+	+	similar	SAQYSRIGPNQAE	832	+	+	similar
NAQYSRIGPNLQAQ	300	+	+	+	----	SAQYSRIGPNQAQ	833	+	+	similar
NAQYSRIGPNQAQ	302	++	++	++	+	SAQYSRIGPNQAQ	833	+	+	----
NAQYSRIGPNQAQ	302	++	++	++	similar	SAQYSRIGPNQAL	834	+	+	----
NAQYSRIGPNPQAQ	304	+++	+++	+++	----	SAQYSRIGPNQAL	834	+	+	-
NAQYSRIGPNPQAQ	304	+++	+++	+++	----	SAQYSRIGPNQAM	835	+	+	similar
NAQYSRIGPNYQAQ	306	++	+	-	----	SAQYSRIGPNQAQ	836	+	+	----
NFQYSRIGPNAAQ	307	+	-	-	----	SAQYSRIGPNQAS	837	++	++	similar
NFQYSRIGPNAAQ	308	+	+	+	----	SAQYSRIGPNQAS	837	+	+	--
NMQYSRIGPNAAQ	309	+	+	+	----	SAQYSRIGPNQAT	838	++	++	similar
NPQYSRIGPNAAQ	310	+	+	+	----	SAQYSRIGPNQAT	838	++	++	similar
NPQYSRIGPNAAQ	310	+	+	+	----	SAQYSRIGPNQCCQ	839	++	+	similar
NTQYSRIGPNAAQ	311	++	+	+	----	SAQYSRIGPNQCCQ	839	+	+	----
NTQYSRIGPNAAQ	311	+	+	+	----	SAQYSRIGPNQCCQ	840	+	+	----
QAQYSRIGPNAEAQ	312	+	+	+	----	SAQYSRIGPNQSQ	841	++	++	--
QAQYSRIGPNAMAQ	313	+	+	+	----	SAQYSRIGPNQSQ	841	+	+	-
QAQYSRIGPNAMAQ	313	+	++	++	----	SAQYSRIGPNQTTQ	842	++	+	----
QAQYSRIGPNAPAQ	314	++	++	++	similar	SAQYSRIGPNQTTQ	842	++	++	----
QAQYSRIGPNAPAQ	314	+	+	+	----	SAQYSRIGPNQVQ	843	+	+	++
QAQYSRIGPNAAQA	315	+	+	+	similar	SAQYSRIGPNSTAQ	845	++	+	----

QAQYSRIGPNAQA	315	+	+	+	---	SAQYSRIGPNTAQ	845	++	++	----	++
QAQYSRIGPNAQAS	316	+	+	+	---	SAQYSRIGPNTPAQ	850	++	++	----	++
QAQYSRIGPNAQAS	316	+		+	similar	SAQYSRIGPNTPAQ	850	+	+	----	+
QAQYSRIGPNAQAT	317	++	++	+	+	SAQYSRIGPNTQAT	851	++	++	----	similar
QAQYSRIGPNAQAT	317	++	++	++	+	SAQYSRIGPNTQAT	851	++	++	----	similar
QAQYSRIGPNAQCQ	318	+	+	+	+	SAQYSRIGPNTQCQ	852	+	+	----	+
QAQYSRIGPNAQCQ	318	+	+	+	similar	SAQYSRIGPNTQCQ	852	+	+	----	+
QAQYSRIGPNAQBQ	319	+	+	+	----	SAQYSRIGPNTQEQ	853	+	+	----	+
QAQYSRIGPNAQSQ	320	+	+	+	similar	SAQYSRIGPNTQEQ	853	+	+	----	+
QAQYSRIGPNAQTQ	321	++	++	+	---	SAQYSRIGPNTQSQ	854	+	+	----	+
QAQYSRIGPNAQTQ	321	++	++	++	similar	SAQYSRIGPNTQSQ	854	+	+	----	+
QAQYSRIGPNAQAQ	323	+++	+++	+++	----	SAQYSRIGPNTQTQ	855	+	+	----	+
QAQYSRIGPNAQAQ	323	++	++	+	----	SAQYSRIGPNTQTQ	855	+	+	----	+
QAQYSRIGPNDQAQ	327	+	+	+	similar	SAQYSRIGPNNVQAS	858	+	+	----	+
QAQYSRIGPNDQAQ	327	+	+	+	----	SAQYSRIGPNNVQAT	859	++	++	----	+
QAQYSRIGPNEQAQ	328	++	++	+	similar	SAQYSRIGPNNVQAT	859	++	++	----	similar
QAQYSRIGPNEQAQ	328	+	+	+	----	SAQYSRIGPNNVQCQ	860	+	+	----	similar
QAQYSRIGPNIQAQ	330	+	+	+	----	SAQYSRIGPNNVQTQ	861	+	+	----	+
QAQYSRIGPNIQAQ	330	+	+	+	-	SAQYSRIGPNNVQTQ	861	+	+	----	+
QAQYSRIGPNNQAQ	333	+++	+++	+++	----	SAQYSRIGPNNVTAQ	863	++	++	----	+
QAQYSRIGPNNQAQ	333	+++	+++	+++	similar	SAQYSRIGPNNVTAQ	863	++	++	----	similar
QAQYSRIGPNNPQAQ	337	+++	+++	+++	----	SAQYSRIGPNNYAAQ	866	++	++	----	+
QAQYSRIGPNNPQAQ	337	+++	+++	+++	----	SAQYSRIGPNNYAAQ	866	++	++	----	+
QAQYSRIGPNNQAQ	343	+	+	+	+	SAQYSRIGPNNYDAQ	867	++	++	----	++
QAQYSRIGPNNYQAQ	345	++	+	+	----	SAQYSRIGPNNYDAQ	867	+	+	----	++
QAQYSRIGPNNYQAQ	345	++	+	+	----	SAQYSRIGPNNYEAQ	868	+++	+++	----	+++
QAQYSRIGPNAQAQ	346	+	+	+	----	SAQYSRIGPNNYEAQ	868	++	++	----	+++
QAQYSRIGPNAQAQ	346	+	+	+	----	SAQYSRIGPNNYPAQ	869	+++	+++	----	+++
QPQYSRIGPNAQAQ	347	++	++	+	----	SAQYSRIGPNNYPAQ	869	+++	+++	----	+++
QPQYSRIGPNAQAQ	347	++	++	++	----	SAQYSRIGPNNYQAA	870	+++	+++	----	+++
QAQYSRIGPNDQAQ	349	++	++	++	----	SAQYSRIGPNNYQAA	870	+++	+++	----	+

SAQYSRIGPNDAQ	349	+	++	similar	SAQYSRIGPNYQAD	871	++	++	++	similar
SAQYSRIGPNPQAO	350	+	+	----	SAQYSRIGPNYQAD	871	++	++	+	similar
SAQYSRIGPNPQAO	350	+	++	----	SAQYSRIGPNYQAE	872	+++	+++	+++	similar
SAQYSRIGPNAAAA	351	+	+	----	SAQYSRIGPNYQAE	872	+++	+++	+++	similar
SAQYSRIGPNAAAA	351	+	+	----	SAQYSRIGPNYQAG	873	++	++	+	----
SAQYSRIGPNAAAE	352	+	+	similar	SAQYSRIGPNYQAG	873	++	++	+	----
SAQYSRIGPNAAAE	352	+	++	similar	SAQYSRIGPNYQAL	874	+	+	+	----
SAQYSRIGPNAAAN	353	+	+	similar	SAQYSRIGPNYQAM	875	+	+	+	----
SAQYSRIGPNAAAN	353	+	+	similar	SAQYSRIGPNYQAN	876	+	+	+	----
SAQYSRIGPNAAAQ	354	+	+	similar	SAQYSRIGPNYQAO	877	+++	+++	++	----
SAQYSRIGPNAAAQ	354	+	+	----	SAQYSRIGPNYQAO	877	++	++	++	----
SAQYSRIGPNAAAAS	355	++	++	----	SAQYSRIGPNYQAS	878	+++	+++	+++	----
SAQYSRIGPNAAAAS	355	+	+	----	SAQYSRIGPNYQAS	878	+++	+++	+	----
SAQYSRIGPNAAAT	356	++	++	----	SAQYSRIGPNYQAT	879	+++	+++	+	----
SAQYSRIGPNAAAT	356	++	++	----	SAQYSRIGPNYQAT	879	+++	+++	+++	----
SAQYSRIGPNAAAV	357	+	+	----	SAQYSRIGPNYQCO	880	++	++	-	----
SAQYSRIGPNAAAV	357	+	+	----	SAQYSRIGPNYQCO	880	+	+	-	----
SAQYSRIGPNAAOQ	358	+	+	----	SAQYSRIGPNYQPO	881	++	++	++	----
SAQYSRIGPNAAOQ	358	+	++	----	SAQYSRIGPNYQPO	881	++	++	++	----
SAQYSRIGPNAAOQ	359	+	+	----	SAQYSRIGPNYQSO	882	+++	+++	+++	----
SAQYSRIGPNAAATQ	360	+	+	----	SAQYSRIGPNYQITQ	883	+++	+++	+++	----
SAQYSRIGPNAAVQ	361	+	+	----	SAQYSRIGPNYQITQ	883	++	++	+	----
SAQYSRIGPNAAEA	362	+	+	----	SAQYSRIGPNYSAQ	884	+	+	+	----
SAQYSRIGPNAAEA	362	+	+	----	SAQYSRIGPNYTAQ	885	++	++	+	----
SAQYSRIGPNAAEAG	363	+	+	----	SCQYSRIGPNAPAQ	886	+	+	+	----
SAQYSRIGPNAAEAQ	364	+	+	----	SCQYSRIGPNAPAQ	886	+	+	+	----
SAQYSRIGPNAAEAS	365	+	+	----	SCQYSRIGPNQAQAE	887	+	+	+	similar
SAQYSRIGPNAAEAS	365	+	+	----	SCQYSRIGPNQAAT	888	+	+	-	----
SAQYSRIGPNAAEAT	366	++	++	----	SCQYSRIGPNDQAO	889	++	++	+	similar
SAQYSRIGPNAAEAT	366	+	+	----	SCQYSRIGPNQAQ	890	+	+	-	----
SAQYSRIGPNAAEAV	367	+	+	----	SCQYSRIGPNQAQ	890	+	+	+	----

SAQYSRIGPNAEAV	367	+		++			SCQYSRIGPNNAQ	891	++	++	++	
SAQYSRIGPNAECQ	368	+		+			SCQYSRIGPNNAQ	891	++	++	+++	
SAQYSRIGPNAECQ	368	+		+			SCQYSRIGPNNAQ	892	+++	+++	+++	
SAQYSRIGPNAESQ	369	+		+			SCQYSRIGPNNAQ	892	+++	+++	+++	
SAQYSRIGPNAESQ	369	+		++			SFQYSRIGPNAQAQ	893	+	+	+	
SAQYSRIGPNAETQ	370	+		+			SFQYSRIGPNAQAQ	893	+	+	+	
SAQYSRIGPNAETQ	370	+		+			SFQYSRIGPNAQAQ	894	+	+	+	
SAQYSRIGPNAIAA	371	+		-			SFQYSRIGPNAQAQ	895	+	+	+	
SAQYSRIGPNAIAE	372	+++		+++			SFQYSRIGPNAQAQ	896	+++	+++	+++	
SAQYSRIGPNAIAE	372	+++		++			SFQYSRIGPNAQAQ	896	+++	++	+	
SAQYSRIGPNAIAQ	373	+		-			SFQYSRIGPNAQAQ	897	+	+	+	
SAQYSRIGPNAIAT	374	+		+			SFQYSRIGPNAQAQ	897	+	+	+	
SAQYSRIGPNAIAT	374	+		-			SFQYSRIGPNAQAQ	898	+	+	+	
SAQYSRIGPNAICQ	375	++		+			SFQYSRIGPNAQAQ	898	+	+	+	
SAQYSRIGPNAICQ	375	++		+			SFQYSRIGPNAQAQ	899	+	+	+	
SAQYSRIGPNAIFQ	376	+		++			SFQYSRIGPNAQAQ	900	+	+	+	
SAQYSRIGPNAIFQ	376	+		+			SFQYSRIGPNAQAQ	901	+	+	+	
SAQYSRIGPNAIIQ	377	++		+			SFQYSRIGPNAQAQ	901	+	+	-	
SAQYSRIGPNAIIQ	377	++		++			SFQYSRIGPNAQAQ	902	+++	+	+	
SAQYSRIGPNAIPQ	378	+++		+++			SFQYSRIGPNAQAQ	902	+	+	+	
SAQYSRIGPNAIPQ	378	+++		+++			SFQYSRIGPNAQAQ	903	+	+	++	
SAQYSRIGPNAISQ	379	++		+			SFQYSRIGPNAQAQ	904	+++	+++	+++	
SAQYSRIGPNAISQ	379	++		+			SFQYSRIGPNAQAQ	904	++	++	+	
SAQYSRIGPNAITQ	380	++		+			SFQYSRIGPNAQAQ	905	+	+	-	
SAQYSRIGPNAITQ	380	++		+			SFQYSRIGPNAQAQ	906	+	+	+	
SAQYSRIGPNAIVQ	381	+++		+++			SFQYSRIGPNAQAQ	907	++	++	+	
SAQYSRIGPNAKAD	382	+		+			SFQYSRIGPNDQAQ	908	++	++	++	
SAQYSRIGPNAKAE	383	++		++			SFQYSRIGPNDQAQ	908	+	+	+	
SAQYSRIGPNAKAE	383	++		+++			SFQYSRIGPNEQAQ	909	+	+	+	
SAQYSRIGPNAIAA	384	+		++			SFQYSRIGPNEQAQ	909	+	+	+	
SAQYSRIGPNAIAE	385	+++		+++			SFQYSRIGPNEQAQ	911	++	++	+	

SAQYSRIGPNALAE	385	++	+++	----	SEQYSRIGPNQAQ	912	+	-	-----
SAQYSRIGPNALAT	386	++	++	-----	SEQYSRIGPNQAQ	913	++++	++++	-----
SAQYSRIGPNALAT	386	+	+	-----	SEQYSRIGPNQAQ	913	+++	+	-----
SAQYSRIGPNALPQ	387	++	++	-----	SEQYSRIGPNQAQ	914	++	++	-----
SAQYSRIGPNALPQ	387	+	+	-----	SEQYSRIGPNQAQ	914	++	++	-----
SAQYSRIGPNALSQ	388	+	=	-----	SEQYSRIGPNQAQ	915	+	=	-----
SAQYSRIGPNAMAA	389	++	++	-----	SEQYSRIGPNYQAQ	916	+++	+	-----
SAQYSRIGPNAMAA	389	++	++	-----	SEQYSRIGPNAEQAQ	917	+	++	-----
SAQYSRIGPNAMAE	390	+	+	-----	SEQYSRIGPNAMQAQ	918	++	=	-----
SAQYSRIGPNAMAE	390	+	+	=	SEQYSRIGPNAMQAQ	918	++	=	-----
SAQYSRIGPNAMAG	391	+++	++	-----	SEQYSRIGPNAPQAQ	919	+++	+++	-----
SAQYSRIGPNAMAG	391	++	+	-----	SEQYSRIGPNAPQAQ	919	+++	++++	-----
SAQYSRIGPNAMAI	392	+	+	-----	SEQYSRIGPNQAQA	920	+	+	-----
SAQYSRIGPNAMAI	392	+	+	-----	SEQYSRIGPNQAQAD	921	+++	+	-----
SAQYSRIGPNAMAL	393	++	+	-----	SEQYSRIGPNQAQAD	921	++	++	-----
SAQYSRIGPNAMAL	393	++	++	-----	SEQYSRIGPNQAQAE	922	++	++	-----
SAQYSRIGPNAMAM	394	++	++	-----	SEQYSRIGPNQAQAE	922	++	++	-----
SAQYSRIGPNAMAM	394	+	=	-----	SEQYSRIGPNQAQ	923	+	+	-----
SAQYSRIGPNAMAN	395	+	++	-----	SEQYSRIGPNQAQAS	924	++	+	-----
SAQYSRIGPNAMAQ	396	++	++	-----	SEQYSRIGPNQAQAS	924	+	=	-----
SAQYSRIGPNAMAQ	396	++	++	-----	SEQYSRIGPNQAQAT	925	++	+	-----
SAQYSRIGPNAMAS	397	+++	++	----	SEQYSRIGPNQAQAT	925	++	+	-----
SAQYSRIGPNAMAS	397	+++	+++	-----	SEQYSRIGPNQAQSQ	926	+	=	-----
SAQYSRIGPNAMAT	398	+++	+++	----	SEQYSRIGPNQAQSQ	926	+	=	-----
SAQYSRIGPNAMAT	398	+++	+++	-----	SEQYSRIGPNQQAQ	927	+++	+++	-----
SAQYSRIGPNAMAV	399	++	+	-----	SEQYSRIGPNQQAQ	927	++	++	-----
SAQYSRIGPNAMAV	399	++	+	-----	SEQYSRIGPNEQAQ	929	+++	++	-----
SAQYSRIGPNAMICQ	400	+++	++	-----	SEQYSRIGPNEQAQ	929	++	++	similar
SAQYSRIGPNAMICQ	400	++	++	-----	SEQYSRIGPNQQAQ	931	+	+	-----
SAQYSRIGPNAMIGQ	401	+	+	-----	SEQYSRIGPNQQAQ	931	+	+	-----
SAQYSRIGPNAMICQ	402	++	++	-----	SEQYSRIGPNQQAQ	932	+++	++	-----

SAQYSRIGPNAMIQ	402	+	+	+	+	SIQYSRIGPNQAAQ	932	+++	+	-----
SAQYSRIGPNAMPO	403	++	++	++	++	SIQYSRIGPNPQAQ	934	++++	+++	-----
SAQYSRIGPNAMPQ	403	+	+	+	+	SIQYSRIGPNPQAQ	934	++++	++++	-----
SAQYSRIGPNAMSQ	404	++	++	++	++	SIQYSRIGPNQQAQ	935	+	-	-----
SAQYSRIGPNAMISQ	404	++	++	++	++	SKQYSRIGPNQQAQ	936	+	-	-----
SAQYSRIGPNAMITQ	405	+++	+++	+++	+++	SKQYSRIGPNEQAQ	937	++	-	-----
SAQYSRIGPNAMITQ	405	++	++	++	++	SIQYSRIGPNAPAQ	938	+++	++	-----
SAQYSRIGPNAMIVQ	406	++	++	++	++	SIQYSRIGPNAPAQ	938	+++	+++	-----
SAQYSRIGPNAMIVQ	406	++	++	++	++	SIQYSRIGPNAQAA	939	+	-	-----
SAQYSRIGPNANAQ	407	+	+	+	+	SIQYSRIGPNAQAL	940	+	-	-----
SAQYSRIGPNANAT	408	+	+	+	+	SIQYSRIGPNAQAAQ	941	+	-	-----
SAQYSRIGPNANSQ	409	+	+	+	+	SIQYSRIGPNAQAS	942	++	+	-----
SAQYSRIGPNAPAA	410	++	++	++	++	SIQYSRIGPNAQAT	943	+	+	-----
SAQYSRIGPNAPAA	410	+	+	+	+	SIQYSRIGPNQQAQ	944	+	+	-----
SAQYSRIGPNAPAE	411	++	++	++	++	SIQYSRIGPNEQAQ	945	+	+	-----
SAQYSRIGPNAPAE	411	++	++	++	++	SIQYSRIGPNQQAQ	946	+	-	-----
SAQYSRIGPNAPAG	412	++	++	++	++	SIQYSRIGPNQQAQ	947	+++	++++	-----
SAQYSRIGPNAPAG	412	++	++	++	++	SIQYSRIGPNQQAQ	947	+++	+++	-----
SAQYSRIGPNAPAI	413	+	+	+	+	SIQYSRIGPNQQAQ	948	++++	++++	-----
SAQYSRIGPNAPAL	414	+	+	+	+	SIQYSRIGPNQQAQ	948	+++	+++	-----
SAQYSRIGPNAPAM	415	+	+	+	+	SIQYSRIGPNQQAQ	949	++	-	-----
SAQYSRIGPNAPAN	416	++	++	++	++	SIQYSRIGPNSQAQ	950	+	+	-----
SAQYSRIGPNAPAN	416	++	++	++	++	SMQYSRIGPNAAAQ	951	++	+	-----
SAQYSRIGPNAPAQ	417	++	++	++	++	SMQYSRIGPNAAAQ	951	++	++	-----
SAQYSRIGPNAPAQ	417	++	++	++	++	SMQYSRIGPNADAQ	952	+	+	-----
SAQYSRIGPNAPAS	418	++	++	++	++	SMQYSRIGPNAMAQ	953	+	-	-----
SAQYSRIGPNAPAS	418	++	++	++	++	SMQYSRIGPNANAQ	954	++	++	-----
SAQYSRIGPNAPAT	419	+++	+++	+++	+++	SMQYSRIGPNANAQ	954	+	-	-----
SAQYSRIGPNAPAT	419	+++	+++	+++	+++	SMQYSRIGPNAPAQ	955	+++	+++	-----
SAQYSRIGPNAPAV	420	+	+	+	+	SMQYSRIGPNAPAQ	955	+++	+++	-----
SAQYSRIGPNAPAV	420	+	+	+	+	SMQYSRIGPNAQAA	956	++	++	-----

SAQYSRIGPNAPCQ	421	+	+	+	+	similar	SMQYSRIGPNAA	956	++	+	+	similar
SAQYSRIGPNAPCQ	421	+	+	+	+	similar	SMQYSRIGPNAAQAD	957	+	+	+	similar
SAQYSRIGPNAPPQ	422	+	+	+	+	similar	SMQYSRIGPNAAQAD	957	+	+	+	similar
SAQYSRIGPNAPPQ	422	+	+	+	+	similar	SMQYSRIGPNAAQAE	958	+	+	+	similar
SAQYSRIGPNAPSQ	423	++	++	++	++	similar	SMQYSRIGPNAAQAE	958	+	+	+	similar
SAQYSRIGPNAPSQ	423	++	++	++	++	---	SMQYSRIGPNAAQAG	959	++	++	++	---
SAQYSRIGPNAPTQ	424	+	+	+	+	---	SMQYSRIGPNAAQAG	959	+	+	+	---
SAQYSRIGPNAPTQ	424	+	+	+	+	---	SMQYSRIGPNAAQAL	960	++	+	+	---
SAQYSRIGPNAPVQ	425	+	+	+	+	---	SMQYSRIGPNAAQAL	960	++	+	+	---
SAQYSRIGPNAPVQ	425	+	+	+	+	---	SMQYSRIGPNAAQAM	961	++	+	+	---
SAQYSRIGPNAA	426	+	+	+	+	+	SMQYSRIGPNAAQAN	962	+	+	+	---
SAQYSRIGPNAA	426	+	+	+	+	+	SMQYSRIGPNAAQAQ	963	++	++	++	---
SAQYSRIGPNAAE	427	+	+	+	+	+	SMQYSRIGPNAAQAQ	963	+	+	+	---
SAQYSRIGPNAAE	427	+	+	+	+	similar	SMQYSRIGPNAAQAS	964	++	++	++	---
SAQYSRIGPNAAQAG	428	+	+	+	+	similar	SMQYSRIGPNAAQAS	964	++	+	+	---
SAQYSRIGPNAAQAG	428	+	+	+	+	similar	SMQYSRIGPNAAQAT	965	+++	++	++	---
SAQYSRIGPNAAI	429	+	+	+	+	---	SMQYSRIGPNAAQAT	965	+++	+++	+++	---
SAQYSRIGPNAAAL	430	+	+	+	+	similar	SMQYSRIGPNAAQAV	966	+	+	+	---
SAQYSRIGPNAAAL	430	+	+	+	+	-	SMQYSRIGPNAAQAV	966	+	+	+	---
SAQYSRIGPNAAAM	431	+	+	+	+	---	SMQYSRIGPNAAQCQ	967	+	+	+	---
SAQYSRIGPNAAAM	431	+	+	+	+	---	SMQYSRIGPNAAQCQ	967	+	+	+	---
SAQYSRIGPNAAAN	432	+	+	+	+	similar	SMQYSRIGPNAAQPQ	968	+	+	+	---
SAQYSRIGPNAAAS	433	++	++	+	+	+	SMQYSRIGPNAAQSQ	969	++	++	++	---
SAQYSRIGPNAAAS	433	+	+	+	+	similar	SMQYSRIGPNAAQSQ	969	+	+	+	---
SAQYSRIGPNAAAT	434	++	++	++	++	similar	SMQYSRIGPNAAQVQ	970	+	+	+	---
SAQYSRIGPNAAAT	434	++	++	++	++	similar	SMQYSRIGPNASAQ	971	++	+	+	---
SAQYSRIGPNAAQAV	435	++	++	+	+	similar	SMQYSRIGPNASAQ	971	+	++	++	---
SAQYSRIGPNAAQAV	435	+	+	+	+	-	SMQYSRIGPNATAQ	972	+++	+	+	---
SAQYSRIGPNAAQCA	436	++	++	+	+	---	SMQYSRIGPNATAQ	972	+++	++	++	---
SAQYSRIGPNAAQCA	436	+	+	+	+	---	SMQYSRIGPNDDAQ	973	+	+	+	---
SAQYSRIGPNAAQCE	437	++	++	++	++	similar	SMQYSRIGPNDDAQ	973	+	+	+	-

SAQYSRIGPNQAQE	437	++	++	++	similar	SMQYSRIGPNEQAQ	974	+	+	+	----
SAQYSRIGPNQAQQ	438	++	++	+	----	SMQYSRIGPNEQAQ	974	+	+	+	----
SAQYSRIGPNQAQQQ	438	++	++	+	similar	SMQYSRIGPNFQAQ	975	++	-	-	----
SAQYSRIGPNQAQCS	439	+	+	+	----	SMQYSRIGPNFQAQ	975	+	+	+	----
SAQYSRIGPNQAQCS	439	+	+	+	--	SMQYSRIGPNHQAQ	976	+++	-	-	----
SAQYSRIGPNQAQCT	440	+++	++	++	----	SMQYSRIGPNHQAQ	976	++	-	-	----
SAQYSRIGPNQAQCT	440	++	++	+++	----	SMQYSRIGPNQAQ	977	+	+	+	----
SAQYSRIGPNQAQGE	441	+	+	+	similar	SMQYSRIGPNQAQ	977	+	-	-	----
SAQYSRIGPNQAQGE	441	+	+	+	similar	SMQYSRIGPNNQAQ	978	++++	+++++	+	----
SAQYSRIGPNQAQQQ	442	+	+	+	----	SMQYSRIGPNNQAQ	978	++++	++++	+	----
SAQYSRIGPNQAQGT	443	+	+	+	----	SMQYSRIGPNPQAQ	982	++++	++++	+	----
SAQYSRIGPNQAQHE	444	+	+	+	----	SMQYSRIGPNPQAQ	982	++++	++++	+	----
SAQYSRIGPNQAQPA	445	+	+	+	----	SMQYSRIGPNQAQ	985	++	+	+	----
SAQYSRIGPNQAQPA	445	+	+	+	----	SMQYSRIGPNQAQ	985	++	+	+	----
SAQYSRIGPNQAQPE	446	++	++	++	similar	SMQYSRIGPNSQAQ	986	++	++	++	----
SAQYSRIGPNQAQPE	446	++	++	+	similar	SMQYSRIGPNSQAQ	986	++	++	++	----
SAQYSRIGPNQAQPI	447	+	+	+	----	SMQYSRIGPNYQAQ	987	+++	++	++	----
SAQYSRIGPNQAQPI	447	+	+	+	----	SPQYSRIGPNAAAQ	988	++	++	++	----
SAQYSRIGPNQAQPL	448	++	++	++	----	SPQYSRIGPNAAAQ	988	++	+	+	----
SAQYSRIGPNQAQPL	448	++	++	++	----	SPQYSRIGPNAEAQ	989	++	++	++	----
SAQYSRIGPNQAQPN	449	+	+	+	----	SPQYSRIGPNAEAQ	989	+	++	++	----
SAQYSRIGPNQAQPN	449	+	+	+	----	SPQYSRIGPNAIAQ	990	+	+	+	----
SAQYSRIGPNQAQPQ	450	+	+	+	----	SPQYSRIGPNAMAQ	991	+++	+++	+++	----
SAQYSRIGPNQAQPQ	450	+	+	+	----	SPQYSRIGPNAMAQ	991	+++	++	++	----
SAQYSRIGPNQAQPS	451	++	++	+	+	SPQYSRIGPNANAQ	992	+	+	+	----
SAQYSRIGPNQAQPS	451	+	+	+	----	SPQYSRIGPNANAQ	992	+	+	+	----
SAQYSRIGPNQAQPT	452	+++	+++	+++	similar	SPQYSRIGPNAPAQ	993	++++	++++	++++	----
SAQYSRIGPNQAQPT	452	++	++	+++	----	SPQYSRIGPNAPAQ	993	++++	++++	++++	----
SAQYSRIGPNQAQPV	453	++	++	++	----	SPQYSRIGPNAQAA	994	+++	+++	+++	----
SAQYSRIGPNQAQPV	453	++	++	+	----	SPQYSRIGPNAQAA	994	+++	+++	+++	+
SAQYSRIGPNQAQSA	454	+	+	+	similar	SPQYSRIGPNAQAD	995	+++	+++	+++	similar

SAQYSRIGPNAQSA	454	+	+	+	similar	SPQYSRIGPNAQAD	995	+++	+++	similar
SAQYSRIGPNAQSE	455	++	++	++	similar	SPQYSRIGPNAQAE	996	+++	+++	-----
SAQYSRIGPNAQSE	455	+		+	similar	SPQYSRIGPNAQAE	996	+++	+++	-
SAQYSRIGPNAQSG	456	+		+	----	SPQYSRIGPNAQAG	997	+++	+++	-----
SAQYSRIGPNAQSL	457	+		+	----	SPQYSRIGPNAQAG	997	++	++	-----
SAQYSRIGPNAQSL	457	+		+	---	SPQYSRIGPNAQAI	998	+++	+++	-----
SAQYSRIGPNAQSN	458	+		+	similar	SPQYSRIGPNAQAI	998	++	+++	-----
SAQYSRIGPNAQSN	458	+		+	----	SPQYSRIGPNAQAL	999	++	++	-----
SAQYSRIGPNAQSQ	459	++		+	similar	SPQYSRIGPNAQAL	999	++	++	-----
SAQYSRIGPNAQSQ	459	++		++	---	SPQYSRIGPNAQAM	1000	+++	++	-----
SAQYSRIGPNAQSS	460	++		+	similar	SPQYSRIGPNAQAM	1000	++	++	-----
SAQYSRIGPNAQSS	460	++		+	--	SPQYSRIGPNAQAN	1001	++	+	-----
SAQYSRIGPNAQST	461	++		++	similar	SPQYSRIGPNAQAN	1001	++	+	-----
SAQYSRIGPNAQST	461	++		+++	similar	SPQYSRIGPNAQAQ	1002	+++	++	-----
SAQYSRIGPNAQSV	462	+		-	----	SPQYSRIGPNAQAQ	1002	+++	+++	-----
SAQYSRIGPNAQSV	462	+		++	----	SPQYSRIGPNAQAS	1003	+++	+++	similar
SAQYSRIGPNAQTA	463	++		++	--	SPQYSRIGPNAQAS	1003	+++	+++	-----
SAQYSRIGPNAQTA	463	+		+	----	SPQYSRIGPNAQAT	1004	++++	+++	-----
SAQYSRIGPNAQTE	464	++		++	similar	SPQYSRIGPNAQAT	1004	++++	++++	-----
SAQYSRIGPNAQTE	464	++		++	similar	SPQYSRIGPNAQAV	1005	+++	+	-----
SAQYSRIGPNAQTQ	465	++		+++	----	SPQYSRIGPNAQAV	1005	++	+	-----
SAQYSRIGPNAQTQ	465	++		++	----	SPQYSRIGPNAQAY	1006	++	-	-----
SAQYSRIGPNAQTS	466	+		+	----	SPQYSRIGPNAQCQ	1007	+	-	-----
SAQYSRIGPNAQTS	466	+		+	----	SPQYSRIGPNAQPQ	1008	++	+	-----
SAQYSRIGPNAQTT	467	++		++	----	SPQYSRIGPNAQPQ	1008	++	+	-----
SAQYSRIGPNAQTT	467	++		++	----	SPQYSRIGPNAQSQ	1009	+++	++	-----
SAQYSRIGPNAQVA	468	+		+	--	SPQYSRIGPNAQSQ	1009	++	++	-----
SAQYSRIGPNAQVA	468	+		+	----	SPQYSRIGPNAQVQ	1010	++	+	-----
SAQYSRIGPNAQVE	469	+		+	similar	SPQYSRIGPNAQVQ	1010	++	-	-----
SAQYSRIGPNAQVE	469	+		+	similar	SPQYSRIGPNASAQ	1011	++	+	-----
SAQYSRIGPNAQVN	470	+		+	--	SPQYSRIGPNASAQ	1011	++	+++	-----

SAQYSRIGPNQVQ	471	+	+	+	similar	SPQYSRIGPNATAQ	1013	+++	+	----
SAQYSRIGPNQVQ	471	+	+	+	----	SPQYSRIGPNATAQ	1013	+++	+++	----
SAQYSRIGPNQVVS	472	+	+	+	----	SPQYSRIGPNQQAQ	1015	+	+	----
SAQYSRIGPNQVVS	472	+	+	+	---	SPQYSRIGPNQQAQ	1015	+	+	----
SAQYSRIGPNQVVT	473	++	++	++	---	SPQYSRIGPNQQAQ	1016	++++	++++	----
SAQYSRIGPNQVVT	473	++	++	+	----	SPQYSRIGPNQQAQ	1016	++++	++++	----
SAQYSRIGPNASAA	474	+	+	+	---	SPQYSRIGPNQQAQ	1018	++++	++++	----
SAQYSRIGPNASAA	474	+	+	++	----	SPQYSRIGPNQQAQ	1018	++++	++++	-
SAQYSRIGPNASAE	475	+	+	+	+	SPQYSRIGPNQQAQ	1020	+++	+++	----
SAQYSRIGPNASAE	475	+	+	+	similar	SPQYSRIGPNQQAQ	1020	++	++	----
SAQYSRIGPNASAG	476	+	+	+	----	SPQYSRIGPNQQAQ	1021	+++	+	----
SAQYSRIGPNASAN	477	+	+	+	----	SPQYSRIGPNQQAQ	1021	+++	++++	----
SAQYSRIGPNASAN	477	+	+	+	similar	SPQYSRIGPNQQAQ	1022	++	+	----
SAQYSRIGPNASAQ	478	+	+	+	----	SPQYSRIGPNQQAQ	1022	++	-	----
SAQYSRIGPNASAS	479	++	++	+	similar	SPQYSRIGPNQQAQ	1023	+++	++	----
SAQYSRIGPNASAS	479	+	+	+	----	SPQYSRIGPNQQAQ	1023	++	++	----
SAQYSRIGPNASAT	480	++	++	++	----	SPQYSRIGPNQQAQ	1025	++++	++++	----
SAQYSRIGPNASAT	480	++	++	++	similar	SPQYSRIGPNQQAQ	1025	++++	++++	----
SAQYSRIGPNASAV	481	+	+	+	similar	SPQYSRIGPNQQAQ	1029	++++	++++	----
SAQYSRIGPNASAV	481	+	+	++	-	SPQYSRIGPNQQAQ	1029	++++	+++	----
SAQYSRIGPNASCQ	482	+	+	+	----	SPQYSRIGPNQQAQ	1033	+++	+++	----
SAQYSRIGPNASCQ	482	+	+	+	----	SPQYSRIGPNQQAQ	1033	+++	+++	----
SAQYSRIGPNASSQ	483	+	+	+	+	SPQYSRIGPNQQAQ	1035	++	++	----
SAQYSRIGPNASSQ	483	+	+	+	----	SPQYSRIGPNQQAQ	1035	++	++	----
SAQYSRIGPNASTQ	484	+	+	+	----	SPQYSRIGPNQQAQ	1037	++	++	----
SAQYSRIGPNASTQ	484	+	+	+	----	SPQYSRIGPNQQAQ	1037	+	+	----
SAQYSRIGPNASVQ	485	+	+	+	----	SPQYSRIGPNQQAQ	1039	++	++	----
SAQYSRIGPNASVQ	485	+	+	-	----	SPQYSRIGPNQQAQ	1039	++	++	----
SAQYSRIGPNATAA	486	++++	++++	++++	----	SPQYSRIGPNQQAQ	1041	+++	+++	----
SAQYSRIGPNATAA	486	+++	+++	+++	----	SPQYSRIGPNQQAQ	1041	++	++	----
SAQYSRIGPNATAD	487	+++	+++	++	-	SPQYSRIGPNATAQ	1042	++	++	----

SAQYSRIGPNATAD	487	+++	+++	+++	-	SOQYSRIGPNATAQ	1042	+	+	+	----
SAQYSRIGPNATAE	488	+++	+++	+++	-	SOQYSRIGPNATAQ	1043	+	-	+	----
SAQYSRIGPNATAE	488	+++	+++	+++	-	SOQYSRIGPNNAQAQ	1044	+++	+++	+++	similar
SAQYSRIGPNATAG	489	+++	+++	+++	-----	SOQYSRIGPNNAQAQ	1044	+++	+++	+++	----
SAQYSRIGPNATAG	489	+++	+++	++	-----	SOQYSRIGPNPQAQ	1050	++++	+++	+++	----
SAQYSRIGPNATAI	490	++	+	-	-----	SOQYSRIGPNPQAQ	1050	++++	++++	++++	----
SAQYSRIGPNATAI	490	+	+	-	-----	SOQYSRIGPNYQAQ	1057	+	+	+	-----
SAQYSRIGPNATAL	491	++	++	++	-----	SOQYSRIGPNYQAQ	1057	+	+	+	-----
SAQYSRIGPNATAL	491	++	++	+	-----	STQYSRIGPNAQAAT	1058	+	+	+	-----
SAQYSRIGPNATAM	492	++	++	++	-----	STQYSRIGPNATAQ	1059	++	++	++	-----
SAQYSRIGPNATAM	492	++	++	++	-----	STQYSRIGPNATAQ	1059	++	++	++	-----
SAQYSRIGPNATAQ	493	+++	+++	+++	-----	STQYSRIGPNHQAQ	1060	++	+	-	-----
SAQYSRIGPNATAQ	493	+++	+++	+++	-----	STQYSRIGPNHQAQ	1060	+	+	+	-----
SAQYSRIGPNATAS	494	+++	+++	+++	-----	STQYSRIGPNNAQAQ	1063	+++	+++	+++	-----
SAQYSRIGPNATAS	494	+++	+++	+++	-----	STQYSRIGPNNAQAQ	1063	+++	+++	+++	-----
SAQYSRIGPNATAT	495	+++	+++	+++	-----	STQYSRIGPNPQAQ	1069	++++	+++	+++	-----
SAQYSRIGPNATAT	495	++++	++++	++++	-----	STQYSRIGPNPQAQ	1069	+++	+++	+++	-----
SAQYSRIGPNATAV	496	++	++	+	-----	STQYSRIGPNYQAQ	1075	+	+	+	-----
SAQYSRIGPNATAV	496	+	+	-	-----	STQYSRIGPNYQAQ	1075	+	+	+	-----
SAQYSRIGPNATCQ	497	++	++	++	-----	SVQYSRIGPNAAQAQ	1076	+	-	-	-----
SAQYSRIGPNATCQ	497	++	++	++	-----	SVQYSRIGPNADAQ	1077	+	+	+	-----
SAQYSRIGPNATGQ	498	++	++	+	-----	SVQYSRIGPNAEAQ	1078	++	++	++	-----
SAQYSRIGPNATGQ	498	+	+	+	-----	SVQYSRIGPNAEAQ	1078	++	+	+	-----
SAQYSRIGPNATIQ	499	+++	+++	+++	-----	SVQYSRIGPNAMAQ	1079	++	+++	+++	-----
SAQYSRIGPNATIQ	499	++	++	+++	-----	SVQYSRIGPNAMAQ	1079	+	+	+	-----
SAQYSRIGPNATPQ	501	+++	+++	+++	-----	SVQYSRIGPNAPAQ	1080	++++	+++	+++	-----
SAQYSRIGPNATPQ	501	+++	+++	+++	-----	SVQYSRIGPNAPAQ	1080	++++	++++	++++	-----
SAQYSRIGPNATSQ	503	+++	+++	+	-----	SVQYSRIGPNAQAA	1081	++	-	-	-----
SAQYSRIGPNATSQ	503	+++	+++	+++	-----	SVQYSRIGPNAQAA	1081	++	+++	+++	-----
SAQYSRIGPNATTQ	505	+++	+++	+++	-----	SVQYSRIGPNAQAD	1082	++	++	++	-----
SAQYSRIGPNATTQ	505	+++	+++	+++	-----	SVQYSRIGPNAQAD	1082	++	++	++	-----

SAQYSRIGPNATVQ	507	+	+	+	+	SVQYSRIGPNAQAE	1083	+++	+++	+++	+++	+++
SAQYSRIGPNATVQ	507	+	+	-	+	SVQYSRIGPNAQAE	1083	+++	+++	+++	+++	+++
SAQYSRIGPNAAV	508	++	++	++	++	SVQYSRIGPNAQAG	1084	++	++	++	++	++
SAQYSRIGPNAD	509	++	++	++	++	SVQYSRIGPNAQAG	1084	++	++	++	++	++
SAQYSRIGPNAD	509	++	++	+++	+++	SVQYSRIGPNAQAG	1085	+	+	+	+	+
SAQYSRIGPNAE	510	+++	+++	+++	+++	SVQYSRIGPNAQAG	1085	+	+	+	+	+
SAQYSRIGPNAE	510	+++	+++	++++	++++	similar	1086	++	++	++	++	++
SAQYSRIGPNAG	511	+	+	-	-	SVQYSRIGPNAQAS	1086	+	+	+	+	+
SAQYSRIGPNAG	511	+	+	+	+	SVQYSRIGPNAQAT	1087	++	++	++	++	++
SAQYSRIGPNAT	512	++	++	+	+	SVQYSRIGPNAQAT	1087	++	++	++	++	++
SAQYSRIGPNAT	513	++	++	+	+	SVQYSRIGPNAQAT	1088	++	++	++	++	++
SAQYSRIGPNAT	513	++	++	++	++	SVQYSRIGPNAQAT	1088	++	++	++	++	++
SAQYSRIGPNAT	514	+	+	-	-	SVQYSRIGPNAQAT	1089	++	++	++	++	++
SAQYSRIGPNAT	514	+	+	+	+	SVQYSRIGPNAQAT	1089	++	++	++	++	++
SAQYSRIGPNAT	515	++	++	+++	+++	SVQYSRIGPNAQAT	1090	++	++	++	++	++
SAQYSRIGPNAT	515	++	++	+++	+++	SVQYSRIGPNDQAQ	1093	+++	+++	+++	+++	+++
SAQYSRIGPNAT	516	+++	+++	+++	+++	SVQYSRIGPNDQAQ	1093	+++	+++	+++	+++	+++
SAQYSRIGPNAT	516	+++	+++	+++	+++	SVQYSRIGPNEQAQ	1095	+++	+++	+++	+++	+++
SAQYSRIGPNAT	517	++	++	+++	+++	SVQYSRIGPNEQAQ	1095	+++	+++	+++	+++	+++
SAQYSRIGPNAT	518	+++	+++	++	++	SVQYSRIGPNIQAQ	1096	++	++	++	++	++
SAQYSRIGPNAT	518	+++	+++	+	+	SVQYSRIGPNNQAQ	1097	++++	++++	++++	++++	++++
SAQYSRIGPNAT	519	++	++	-	-	SVQYSRIGPNNQAQ	1097	+++	+++	+++	+++	+++
SAQYSRIGPNAT	520	+	+	+	+	SVQYSRIGPNNQAQ	1099	+++	+++	+++	+++	+++
SAQYSRIGPNCQAS	521	+	+	+	+	SVQYSRIGPNNQAQ	1099	+++	+++	+++	+++	+++
SAQYSRIGPNCQAT	522	++	++	+	+	SVQYSRIGPNNQAQ	1101	++	++	++	++	++
SAQYSRIGPNCQAT	522	+	+	+	+	SVQYSRIGPNNQAQ	1102	++	++	++	++	++
SAQYSRIGPNDAAQ	523	+	+	+	+	SVQYSRIGPNNQAQ	1102	++	++	++	++	++
SAQYSRIGPNDAAQ	523	+	+	+	+	SVQYSRIGPNNQAQ	1103	+	+	+	+	+
SAQYSRIGPNDIAQ	524	+	+	+	+	SVQYSRIGPNNQAQ	1104	+	+	+	+	+
SAQYSRIGPNDIAQ	525	++	++	++	++	SVQYSRIGPNDQAQ	1105	+	+	+	+	+
SAQYSRIGPNDIAQ	525	+	+	+	+	SVQYSRIGPNDQAQ	1105	+	+	+	+	+

SAQYSRIGPNDMAQ	526	++	++	++	---	SYQYSRIGPNNQAQ	1106	++++	+++	----
SAQYSRIGPNDMAQ	526	++	++	++	----	SYQYSRIGPNNQAQ	1106	+++	+++	----
SAQYSRIGPNDQAA	527	++	++	++	similar	SYQYSRIGPNNQAQ	1107	+++	+++	----
SAQYSRIGPNDQAA	527	++	++	++	similar	SYQYSRIGPNNQAQ	1107	++	+	----
SAQYSRIGPNDQAC	528	+	+	+	----	SYQYSRIGPNNQAQ	1108	+	-	----
SAQYSRIGPNDQAC	528	+	+	+	----	TAQYSRIGPNNQAQ	1109	++	++	----
SAQYSRIGPNDQAE	529	+	+	+	-	TAQYSRIGPNNQAQ	1109	++	++	----
SAQYSRIGPNDQAE	529	+	+	+	similar	TAQYSRIGPNNQAQ	1110	+	+	----
SAQYSRIGPNDQAG	530	++	++	++	similar	TAQYSRIGPNNQAQ	1112	++	++	----
SAQYSRIGPNDQAG	530	++	++	++	similar	TAQYSRIGPNNQAQ	1112	++	+	----
SAQYSRIGPNDQAI	531	++	++	++	----	TAQYSRIGPNNQAQ	1114	+	+	----
SAQYSRIGPNDQAI	531	++	++	++	-	TAQYSRIGPNNQAQ	1114	+	+	----
SAQYSRIGPNDQAK	532	+++	+++	+++	----	TAQYSRIGPNNQAQ	1115	++	++	similar
SAQYSRIGPNDQAK	532	+++	+++	+++	similar	TAQYSRIGPNNQAQ	1115	++	++	similar
SAQYSRIGPNDQAL	533	++	++	++	---	TAQYSRIGPNNQAQ	1118	++	++	----
SAQYSRIGPNDQAL	533	++	++	++	----	TAQYSRIGPNNQAQ	1118	++	++	----
SAQYSRIGPNDQAM	534	+	+	+	----	TAQYSRIGPNNQAQ	1120	+	+	----
SAQYSRIGPNDQAM	534	+	+	+	----	TAQYSRIGPNNQAQ	1120	+	-	----
SAQYSRIGPNDQAN	535	++	++	++	similar	TPQYSRIGPNNQAQ	1121	+	+	----
SAQYSRIGPNDQAN	535	++	++	++	similar	TPQYSRIGPNNQAQ	1121	+	+	----
SAQYSRIGPNDQAQ	536	++	++	+	similar	VAQYSRIGPNNQAQ	1124	+	+	----
SAQYSRIGPNDQAQ	536	++	++	++	-	VAQYSRIGPNNQAQ	1124	+	+	----
SAQYSRIGPNDQAR	537	++	++	++	----	VAQYSRIGPNNQAQ	1125	+	+	-
SAQYSRIGPNDQAR	537	+	+	+	----	VAQYSRIGPNNQAQ	1125	+	+	----
SAQYSRIGPNDQAS	538	++	++	++	similar	VAQYSRIGPNNQAQ	1126	+	+	----
SAQYSRIGPNDQAS	538	++	++	++	---	VAQYSRIGPNNQAQ	1127	+++	+++	----
SAQYSRIGPNDQAT	539	+++	+++	+++	similar	VAQYSRIGPNNQAQ	1127	+++	+++	----
SAQYSRIGPNDQAT	539	+++	+++	+++	----	VAQYSRIGPNNQAQ	1128	++	++	----
SAQYSRIGPNDQAV	540	+++	+++	+++	similar	VAQYSRIGPNNQAQ	1128	+	+	----
SAQYSRIGPNDQAV	540	+++	+++	+++	similar	VAQYSRIGPNNQAQ	1129	+	+	----
SAQYSRIGPNDQCC	541	++	++	++	----	VAQYSRIGPNNQAQ	1130	++	++	----

SAQYSRIGPNDQOQ	541	++	++	++	----	VAQYSRIGPNQAQ	1130	++	++	++	----
SAQYSRIGPNDQPO	542	++	++	++	----	VAQYSRIGPNQAO	1131	+	+	+	----
SAQYSRIGPNDQPO	542	++	++	++	----	VAQYSRIGPNQAO	1131	+	+	+	----
SAQYSRIGPNDQSO	543	++	++	++	-	VAQYSRIGPNQQAQ	1132	++	+	+	----
SAQYSRIGPNDQSO	543	++	++	++	similar	VAQYSRIGPNQQAQ	1132	++	++	++	----
SAQYSRIGPNDQTO	544	+	+	+	----	VAQYSRIGPNVQAQ	1133	++	++	++	----
SAQYSRIGPNDQTO	544	+	+	+	----	VAQYSRIGPNVQAQ	1133	++	++	++	----
SAQYSRIGPNDQVO	545	+	+	+	----	VAQYSRIGPNVQAQ	1134	+	+	+	----
SAQYSRIGPNDQVO	545	+	+	+	-	WAQYSRIGPNDQAQ	1135	+	-	-	----
SAQYSRIGPNDQAQ	546	+	+	+	----	YAQYSRIGPNAPAQ	1136	+	+	+	----
SAQYSRIGPNDQAQ	546	+	+	+	----	YAQYSRIGPNAPAQ	1136	+	+	+	----
SAQYSRIGPNDTAQ	547	++	++	++	----	SFQYSRNSDNATAQ	1152	++	+	+	----
SAQYSRIGPNDTAQ	547	+	+	+	----	SVQYSRNSDNHQAQ	1153	+	+	+	----
SAQYSRIGPNDVAQ	549	+	+	-	----	SVQYSRNSDNHQAQ	1153	+	+	+	----
SAQYSRIGPNDVAQ	549	+	+	-	----	SVQYSRNSDNHQAQ	1154	+	+	+	----
SAQYSRIGPNEAAQ	550	+	+	+	similar	SVQYSRNSDNHQAQ	1155	+	-	-	----
SAQYSRIGPNEIAQ	551	++	++	++	----	SVQYSRNSDNHQAQ	1155	+	+	+	----
SAQYSRIGPNEIAQ	551	++	++	++	----	WQYSRNSDNAQAQ	1156	+++	++	++	----
SAQYSRIGPNEIAQ	552	++	++	+	----	WQYSRNSDNAQAQ	1156	++	++	+++	----
SAQYSRIGPNEIAQ	552	++	++	++	----	WVQYSRNSDNAQAQ	1157	+++	+++	+++	----
SAQYSRIGPNEIAQ	553	++	++	++	----	WVQYSRNSDNAQAQ	1157	+++	+++	+++	----
SAQYSRIGPNEIAQ	553	++	++	++	----	SADFRSTNGVAQAA	1213	+	+	+	----
SAQYSRIGPNEIAQ	554	+	+	+	similar	SADFRSTNGVAQAA	1213	+	+	+	----
SAQYSRIGPNEIAQ	554	+	+	+	similar	SADFRSTNGVAQAD	1214	+	+	+	----
SAQYSRIGPNEIAQ	555	+	+	+	similar	SADFRSTNGVAQAD	1214	+	+	+	----
SAQYSRIGPNEIAQ	555	+	+	+	similar	SADFRSTNGVAQAE	1215	++	++	++	----
SAQYSRIGPNEIAQ	556	+	+	+	similar	SADFRSTNGVAQAE	1215	++	++	++	----
SAQYSRIGPNEIAQ	556	+	+	+	-	SADFRSTNGVAQAF	1216	+	+	+	----
SAQYSRIGPNEIAQ	557	+++	+++	+++	----	SADFRSTNGVAQAF	1216	+	+	+	----
SAQYSRIGPNEIAQ	557	+++	+++	+++	similar	SADFRSTNGVAQAG	1217	+	+	+	----
SAQYSRIGPNEIAQ	558	+	+	+	-	SADFRSTNGVAQAS	1218	++	++	++	----

SAQYSRIGPNEQAL	558	+	+	+	----	SADFRSTNGVAQAS	1218	+	++	----
SAQYSRIGPNEQAM	559	+	+	+	----	SADFRSTNGVAQAT	1219	++	++	----
SAQYSRIGPNEQAM	559	+	+	+	similar	SADFRSTNGVAQAT	1219	++	++	----
SAQYSRIGPNEQAN	560	+	+	+	similar	SADFRSTNGVAQFQ	1220	++	++	----
SAQYSRIGPNEQAN	560	+	+	+	similar	SADFRSTNGVAQFQ	1220	+	+	----
SAQYSRIGPNEQAQ	561	+	+	+	similar	SADFRSTNGVAQQQ	1222	+	-	----
SAQYSRIGPNEQAQ	561	+	+	+	similar	SADFRSTNGVAQQQ	1222	+	+	----
SAQYSRIGPNEQAR	562	++	++	+++	----	SADFRSTNGVMQAO	1223	+	+	----
SAQYSRIGPNEQAR	562	++	++	++	similar	SADFRSTNGVMQAO	1223	+	+	----
SAQYSRIGPNEQAS	563	++	++	++	similar	SAEFRSTNGVAPAQ	1227	+	+	----
SAQYSRIGPNEQAS	563	+	+	+	similar	SAEFRSTNGVAPAQ	1227	+	+	----
SAQYSRIGPNEQAT	564	++	++	++	similar	SAEFRSTNGVAQAA	1228	+	+	-
SAQYSRIGPNEQAT	564	++	++	++	similar	SAEFRSTNGVAQAE	1229	++	++	----
SAQYSRIGPNEQAV	565	++	++	++	similar	SAEFRSTNGVAQAE	1229	++	++	----
SAQYSRIGPNEQAV	565	++	++	++	similar	SAEFRSTNGVAQAF	1230	+	+	----
SAQYSRIGPNEQCQ	566	++	++	++	similar	SAEFRSTNGVAQAS	1231	+	+	-
SAQYSRIGPNEQCQ	566	++	++	++	similar	SAEFRSTNGVAQAS	1231	+	+	----
SAQYSRIGPNEQOQ	567	+	+	+	-	SAEFRSTNGVAQAT	1232	++	+	----
SAQYSRIGPNEQOQ	567	+	+	+	----	SAEFRSTNGVAQAT	1232	++	++	----
SAQYSRIGPNEQPQ	568	++	++	++	--	SALFRSTNGVAQAE	1233	+	+	----
SAQYSRIGPNEQPQ	568	++	++	++	similar	SANFRSTNGVAQAE	1234	+	++	----
SAQYSRIGPNEQSQ	569	++	++	++	similar	SANFRSTNGVAQAF	1234	+	+	similar
SAQYSRIGPNEQSQ	569	+	+	++	similar	SAQFRSTNGVAAAE	1235	++	++	----
SAQYSRIGPNEQTQ	570	++	++	++	--	SAQFRSTNGVAAAE	1235	+	++	-
SAQYSRIGPNEQTQ	570	++	++	++	----	SAQFRSTNGVADAE	1236	+	+	----
SAQYSRIGPNEQVQ	571	++	++	+	similar	SAQFRSTNGVADAE	1236	+	+	----
SAQYSRIGPNEQVQ	571	+	+	+	similar	SAQFRSTNGVAGAE	1237	+	+	----
SAQYSRIGPNEQAQ	572	+	+	+	similar	SAQFRSTNGVAGAE	1237	+	+	----
SAQYSRIGPNEQAQ	572	+	+	+	----	SAQFRSTNGVAPAA	1238	+	+	----
SAQYSRIGPNETAQ	573	++	++	++	----	SAQFRSTNGVAPAD	1239	+	+	----
SAQYSRIGPNETAQ	573	++	++	++	----	SAQFRSTNGVAPAD	1239	+	+	----

SAQYSRIGPNEVAQ	576	++	++	++	SAQFRSTNGVAPAE	1240	++	++	++
SAQYSRIGPNEVAQ	576	++	++	+	SAQFRSTNGVAPAE	1240	++	++	++
SAQYSRIGPNEFAAQ	577	+		-	SAQFRSTNGVAPAF	1241	+	-	++
SAQYSRIGPNEFAQ	578	+		+	SAQFRSTNGVAPAS	1242	+	+	++
SAQYSRIGPNEFAQ	578	+		+	SAQFRSTNGVAPAS	1242	+	+	++
SAQYSRIGPNEFAQ	579	++	++	+++	SAQFRSTNGVAPAT	1243	++	++	++
SAQYSRIGPNEFAQ	579	++	++	++	SAQFRSTNGVAPAT	1243	++	++	++
SAQYSRIGPNEFQAA	580	++	+	+	SAQFRSTNGVAVAE	1244	+	+	+
SAQYSRIGPNEFQAA	580	+	+	++	SAQFRSTNGVAVAE	1244	+	+	++
SAQYSRIGPNEFQAE	581	+	+	+	SAQFRSTNGVMPAQ	1245	++	++	++
SAQYSRIGPNEFQAE	581	+	+	+	SAQFRSTNGVMPAQ	1245	++	++	++
SAQYSRIGPNEFQAG	582	++	++	+	SAQFRSTNGVMOAE	1246	+	+	similar
SAQYSRIGPNEFQAAQ	583	+	+	+	SAQFRSTNGVMOAE	1246	+	+	similar
SAQYSRIGPNEFQAS	584	++	++	+	SAQFRSTNGVSPAQ	1247	+	+	similar
SAQYSRIGPNEFQAS	584	+	+	+	SAQFRSTNGVQAQAE	1248	+	+	similar
SAQYSRIGPNEFQAT	585	+++	+++	+++	SAQFRSTNGVQAQAE	1248	+	+	++
SAQYSRIGPNEFQAT	585	++	++	+	SSQFRSTNGVADAQ	1261	+	+	++
SAQYSRIGPNEFQSQ	586	++	++	-	SSQFRSTNGVADAQ	1261	+	+	++
SAQYSRIGPNEHAAQ	587	+	+	+	SSQFRSTNGVQAQAE	1262	+	+	++
SAQYSRIGPNEHDAQ	588	++	++	++	SSQFRSTNGVQAQAE	1262	+	+	similar
SAQYSRIGPNEHDAQ	588	++	++	++	SADFYSTNGVAPAQ	1267	+	++	++
SAQYSRIGPNEHEAQ	589	+++	+++	+++	SADFYSTNGVAPAQ	1267	+	+	++
SAQYSRIGPNEHEAQ	589	++	++	+++	SADFYSTNGVQAQAE	1268	++	++	++
SAQYSRIGPNEHNAQ	590	+	+	-	SADFYSTNGVQAQAE	1268	++	++	++
SAQYSRIGPNEHPAQ	591	++	++	++	SAEFYSTNGVQAQAE	1269	+++	+++	++
SAQYSRIGPNEHPAQ	591	+	+	+	SAEFYSTNGVQAQAE	1269	+++	+++	++
SAQYSRIGPNEHQAA	592	++	++	++	SAHFYSTNGVQAQAE	1270	+	+	++
SAQYSRIGPNEHQAA	592	+	+	+	SALFVYSTNGVQAQAE	1271	++	++	++
SAQYSRIGPNEHQAD	593	++	++	++	SALFVYSTNGVQAQAE	1271	++	++	++
SAQYSRIGPNEHQAD	593	++	++	++	SAMFVYSTNGVQAQAE	1272	++	++	++
SAQYSRIGPNEHQAE	594	+++	+++	+++	SAMFVYSTNGVQAQAE	1272	++	++	++

SAQYSRIGPNHQAE	594	+++	+++	+	similar	SAQFVSTNGVAAAE	1273	++	++	++	++	++	++
SAQYSRIGPNHQAG	595	++	++	+	++	SAQFVSTNGVAAAE	1273	++	++	++	++	++	++
SAQYSRIGPNHQAG	595	+	+	+	++	SAQFVSTNGVAGAE	1274	+++	+++	+++	+++	+++	+++
SAQYSRIGPNHQAL	596	+	+	-	++	SAQFVSTNGVAGAE	1274	++	++	++	++	++	++
SAQYSRIGPNHQAN	597	+	+	+	++	SAQFVSTNGVAPAD	1275	++	++	++	++	++	++
SAQYSRIGPNHQAQ	598	++	++	++	++	SAQFVSTNGVAPAD	1275	++	++	++	++	++	++
SAQYSRIGPNHQAQ	598	++	++	+	++	SAQFVSTNGVAPAE	1276	++	++	++	++	++	++
SAQYSRIGPNHQAS	599	++	++	++	++	SAQFVSTNGVAPAE	1276	++	++	++	++	++	++
SAQYSRIGPNHQAS	599	++	++	++	++	SAQFVSTNGVAPAS	1277	+	+	+	+	+	+
SAQYSRIGPNHQAT	600	+++	+++	+	++	SAQFVSTNGVAPAS	1277	+	+	+	+	+	+
SAQYSRIGPNHQAT	600	++	++	+	++	SAQFVSTNGVAPAT	1278	++	++	++	++	++	++
SAQYSRIGPNHQAV	601	++	++	++	++	SAQFVSTNGVAPAT	1278	++	++	++	++	++	++
SAQYSRIGPNHQCQ	602	++	++	+	++	SAQFVSTNGVAQAE	1279	++	++	++	++	++	++
SAQYSRIGPNHQFQ	603	+	+	+	++	SAQFVSTNGVAQAE	1279	++	++	++	++	++	++
SAQYSRIGPNHQGQ	604	++	++	-	++	SAQFVSTNGVAQSE	1280	++	++	++	++	++	++
SAQYSRIGPNHQIQ	605	+	+	-	++	SAQFVSTNGVAQSE	1280	++	++	++	++	++	++
SAQYSRIGPNHQIQ	605	+	+	+	++	SAQFVSTNGVASAE	1281	+	+	+	+	+	+
SAQYSRIGPNHQPO	606	++	++	+	++	SAQFVSTNGVASAE	1281	+	+	+	+	+	+
SAQYSRIGPNHQPO	606	+	+	++	++	SAQFVSTNGVAVAE	1282	++	++	++	++	++	++
SAQYSRIGPNHQSQ	607	++	++	++	++	SAQFVSTNGVAVAE	1282	+	+	+	+	+	+
SAQYSRIGPNHQSQ	607	++	++	+	++	SAQFVSTNGVMPAQ	1283	++	++	++	++	++	++
SAQYSRIGPNHQTO	608	+++	+++	+	++	SAQFVSTNGVMPAQ	1283	++	++	++	++	++	++
SAQYSRIGPNHQTO	608	+++	+++	-	++	SAQFVSTNGVMQAE	1284	+++	+++	+++	+++	+++	+++
SAQYSRIGPNIAAQ	612	+	+	+	similar	SAQFVSTNGVMQAE	1284	+++	+++	+++	+++	+++	+++
SAQYSRIGPNIAAQ	612	+	+	+	similar	SAQFVSTNGVSPAQ	1285	+	+	+	+	+	+
SAQYSRIGPNIEAQ	613	++	++	++	++	SAQFVSTNGVSPAQ	1285	+	+	+	+	+	+
SAQYSRIGPNIEAQ	613	+	+	+	++	SASFVSTNGVAQAE	1286	++	++	++	++	++	++
SAQYSRIGPNIGAQ	614	+	+	++	++	SASFVSTNGVAQAE	1286	++	++	++	++	++	++
SAQYSRIGPNIGAQ	614	+	+	+	++	SATFVSTNGVAQAE	1287	++	++	++	++	++	++
SAQYSRIGPNIIAQ	615	+	+	-	++	SATFVSTNGVAQAE	1287	++	++	++	++	++	++
SAQYSRIGPNIMAQ	616	++	++	++	++	SAVFVSTNGVAQAE	1288	+	+	+	+	+	+

SAQYSRIGPNIMAQ	616	++	++	++	----	SAQFYSTNGVAQAE	1288	+	+	+	-----
SAQYSRIGPNIMAQ	617	+	+	+	-----	SQOFYSTNGVAQAE	1289	+	+	+	-----
SAQYSRIGPNIMAQ	617	+	+	+	-----	SQOFYSTNGVAQAE	1289	+	+	+	-----
SAQYSRIGPNIQAA	618	++	++	++	similar	SDQFYSTNGVAQAE	1290	+	+	+	-----
SAQYSRIGPNIQAA	618	+	++	++	----	SMQFYSTNGVAQAE	1291	++	++	++	-----
SAQYSRIGPNIQAE	619	+	+	+	similar	SMQFYSTNGVAQAE	1291	+	+	++	-----
SAQYSRIGPNIQAE	619	+	+	+	similar	SNQFYSTNGVAQAE	1292	+	+	+	-----
SAQYSRIGPNIQAG	620	+	+	++	----	SQQFYSTNGVAQAE	1293	++	++	++	-----
SAQYSRIGPNIQAG	620	+	+	+	similar	SQQFYSTNGVAQAE	1293	++	++	++	-----
SAQYSRIGPNIQAL	621	+	+	-	----	SSQFYSTNGVAQAAA	1294	+	+	+	-----
SAQYSRIGPNIQAL	621	+	+	+	-----	SSQFYSTNGVAQAAA	1294	+	+	+	-----
SAQYSRIGPNIQAM	622	++	++	+	-----	SSQFYSTNGVAQAD	1295	+	+	+	-----
SAQYSRIGPNIQAM	622	+	+	+	----	SSQFYSTNGVAQAD	1295	+	+	+	-----
SAQYSRIGPNIQAN	623	+	+	+	+	SSQFYSTNGVAQAE	1296	+++	+++	+++	-----
SAQYSRIGPNIQAQ	624	++	++	++	similar	SSQFYSTNGVAQAE	1296	+++	+++	+++	-----
SAQYSRIGPNIQAQ	624	++	++	++	----	SSQFYSTNGVAQAS	1297	+	+	+	-----
SAQYSRIGPNIQAS	625	++	++	+	+	SSQFYSTNGVAQAS	1297	+	+	+	-----
SAQYSRIGPNIQAS	625	++	++	++	+	SSQFYSTNGVAQAT	1298	+	+	++	-----
SAQYSRIGPNIQAT	626	+++	+++	+++	+	SSQFYSTNGVAQAT	1298	+	+	++	-----
SAQYSRIGPNIQAT	626	+++	+++	+++	----	SAQCLCKPCLTNQAA	1389	+	+	+	-----
SAQYSRIGPNIQAV	627	+	+	+	----	SAQCLCKPCLTNQAO	1389	+	+	+	-----

Table 2 Referenced in Example 2

AA sequence	SEQ ID	hTFR1 with CBA-Cap	hTFR1 with hSyn-Cap	Control CBA-Cap	Control hSyn-Cap	AA sequence	SEQ ID	hTFR1 with CBA-Cap	hTFR1 with hSyn-Cap	Control CBA-Cap	Control hSyn-Cap
SAQYSRIGPNQAQ	45	+	-	--	--	SAQYSRIGPNQTSE	828	++	+	--	--
SAQYSRIGPNQAQ	45	+	-	--	--	SAQYSRIGPNQTSE	828	++	++	-	--
AAQYSRIGPNATAE	82	+++	+++	--	--	SAQYSRIGPNSTAE	844	++	++	+	--
AAQYSRIGPNATAE	82	+++	++	--	--	SAQYSRIGPNSTAE	844	++	++	--	--
AAQYSRIGPNATPE	84	+++	+++	-	--	SAQYSRIGPNSTPE	846	++	++	--	--
AAQYSRIGPNATPE	84	++	++	--	--	SAQYSRIGPNSTPE	846	++	++	--	--
AAQYSRIGPNATSE	85	+++	+++	--	--	SAQYSRIGPNSTSE	847	+++	+++	--	--
AAQYSRIGPNATSE	85	++	+++	--	--	SAQYSRIGPNSTSE	847	++	++	--	--
AAQYSRIGPNATTE	86	+++	++	--	--	SAQYSRIGPNSTTE	848	++	++	--	--
AAQYSRIGPNATTE	86	++	++	--	--	SAQYSRIGPNSTTE	848	++	++	--	--
AAQYSRIGPNATVE	87	++	++	--	--	SAQYSRIGPNSTVE	849	+	++	--	--
AAQYSRIGPNATVE	87	++	++	--	--	SAQYSRIGPNSTVE	849	+	++	--	--
AAQYSRIGPNHTAE	91	+++	+++	--	--	SAQYSRIGPNTTSE	856	++	++	--	--
AAQYSRIGPNHTAE	91	+++	+++	--	--	SAQYSRIGPNTTSE	856	++	-	--	--
AAQYSRIGPNITAE	93	++	++	--	--	SAQYSRIGPNTTTE	857	++	++	--	--
AAQYSRIGPNITAE	93	++	++	--	--	SAQYSRIGPNVTAE	862	++	++	--	--
AAQYSRIGPNNEAA	94	++	++	--	--	SAQYSRIGPNVTAE	862	++	++	--	--
AAQYSRIGPNNTAE	96	++++	+++	--	--	SAQYSRIGPNVTSE	864	++	++	--	--
AAQYSRIGPNNTAE	96	+++	+++	--	--	SAQYSRIGPNVTSE	864	++	++	--	--
AAQYSRIGPNPAAE	97	+++	+++	-	--	SAQYSRIGPNVTTE	865	++	++	--	--
AAQYSRIGPNPAAE	97	+++	+++	--	--	SAQYSRIGPNVTTE	865	++	-	--	--
AAQYSRIGPNPHAE	98	++	++	--	--	SFQYSRIGPNETAE	910	++	++	--	--
AAQYSRIGPNPHAE	98	11	11	--	--	SFQYSRIGPNETAE	910	11	11	--	--
AAQYSRIGPNPMAE	99	+++	+++	--	--	SIQYSRIGPNDIAE	928	++	++	--	--
AAQYSRIGPNPMAE	99	+++	++	--	--	SIQYSRIGPNDIAE	928	+	++	--	--
AAQYSRIGPNPQAE	100	++++	+++	--	--	SIQYSRIGPNETAE	930	++	++	--	--
AAQYSRIGPNPQAE	100	+++	++	--	--	SIQYSRIGPNPEAE	933	++	++	--	--
AAQYSRIGPNPQPE	102	++++	+++	--	--	SNOYSRIGPNNTAE	979	++	++	--	--
AAQYSRIGPNPQPE	102	++++	++++	--	--	SNOYSRIGPNNTAE	979	++	+++	--	--

AAQYSRIGPNPQVE	103	++++	+++	--	--	--	SMQYSRIGPNPEAE	980	++	--+	--	--
AAQYSRIGPNPQVE	103	++++	++	--	--	--	SMQYSRIGPNPEAE	980	+	+	--	--
AAQYSRIGPNPSAE	104	+++	+++	--	--	--	SMQYSRIGPNPQAE	981	++	+++	--	--
AAQYSRIGPNPSAE	104	+++	+++	--	--	--	SMQYSRIGPNPQAE	981	++	+++	--	--
AAQYSRIGPNPTAE	105	+++	+++	--	--	--	SMQYSRIGPNPQPE	983	+++	+++	--	--
AAQYSRIGPNPTAE	105	+++	+++	--	--	--	SMQYSRIGPNPQPE	983	+++	+++	--	--
AAQYSRIGPNSTAE	108	+++	+++	--	--	--	SMQYSRIGPNPVAE	984	+++	+++	--	--
AAQYSRIGPNSTAE	108	++	++	--	--	--	SMQYSRIGPNPVAE	984	+++	+++	--	--
APQYSRIGPNATAE	113	+++	+++	--	--	--	SPQYSRIGPNATAE	1012	++++	+++	--	--
APQYSRIGPNATAE	113	+++	+++	--	--	--	SPQYSRIGPNATAE	1012	+++	+++	--	--
APQYSRIGPNPQAE	114	++++	+++	--	--	--	SPQYSRIGPNATPE	1014	+++	+	--	--
APQYSRIGPNPQAE	114	++++	+++	--	--	--	SPQYSRIGPNATPE	1014	++	+	--	--
ATQYSRIGPNPQAE	115	+++	++	--	--	--	SPQYSRIGPNDTAE	1017	+++	+++	--	--
ATQYSRIGPNPQAE	115	+++	++	--	--	--	SPQYSRIGPNDTAE	1017	+++	+++	--	--
DRQYSRIGPNPQAE	121	++	++	--	--	--	SPQYSRIGPNETAE	1019	++++	+++	--	--
DRQYSRIGPNPQAE	121	++	++	--	--	--	SPQYSRIGPNETAE	1019	++++	+++	--	--
EAQYSRIGPNHTAE	144	++++	+++	--	--	--	SPQYSRIGPNNEAA	1024	++	+	--	--
EAQYSRIGPNHTAE	144	+++	+++	--	--	--	SPQYSRIGPNNEAA	1024	++	-	--	--
EAQYSRIGPNKTAE	147	++++	+++	--	--	--	SPQYSRIGPNNTAE	1026	++++	+++	--	--
EAQYSRIGPNKTAE	147	++++	+++	--	--	--	SPQYSRIGPNNTAE	1026	++++	+++	--	--
EAQYSRIGPNNAAM	148	++	++	--	--	--	SPQYSRIGPNPEAE	1027	++	+++	--	--
EAQYSRIGPNNAAM	148	++	+	--	--	--	SPQYSRIGPNPEAE	1027	++	+	--	--
EAQYSRIGPNNAAQ	149	++	++	--	--	--	SPQYSRIGPNPQAE	1028	+++++	+++	--	--
EAQYSRIGPNNAAQ	149	++	++	--	--	--	SPQYSRIGPNPQAE	1028	+++++	+++	--	--
EAQYSRIGPNNAAT	150	+++	+++	--	--	--	SPQYSRIGPNPQPE	1030	+++++	+++	--	--
EAQYSRIGPNNAAT	150	++	++	--	--	--	SPQYSRIGPNPQPE	1030	++++	+++	--	--
EAQYSRIGPNNHAE	151	++	+++	--	--	--	SPQYSRIGPNPQVE	1031	+++	+++	--	--
EAQYSRIGPNNHAE	151	++	++	--	--	--	SPQYSRIGPNPQVE	1031	+++	+++	--	--
EAQYSRIGPNNIAE	152	++	+++	--	--	--	SPQYSRIGPNPTAE	1032	++++	+++	--	--
EAQYSRIGPNNIAE	152	++	++	--	--	--	SPQYSRIGPNPTAE	1032	+++	+++	--	--
EAQYSRIGPNNKAE	153	++	++	--	--	--	SPQYSRIGPNQTAE	1034	++++	+++	--	--

[illegible]

EAQYSRIGPNNQAN	169	+++	+++	+++	---	---	STQYSRIGPNHTAE	1061	++	---	---	---
EAQYSRIGPNNQAO	170	+++	+++	---	---	---	STQYSRIGPNNEAA	1062	++	---	---	---
EAQYSRIGPNNQAO	170	+++	+++	---	---	---	STQYSRIGPNNEAA	1062	++	---	---	---
EAQYSRIGPNNQAS	171	+++	+++	---	---	---	STQYSRIGPNNNTAE	1064	+++	---	---	---
EAQYSRIGPNNQAS	171	+++	+++	---	---	---	STQYSRIGPNNNTAE	1064	+++	---	---	---
EAQYSRIGPNNQAT	172	+++	+++	---	---	---	STQYSRIGPNFAAE	1065	+++	---	---	---
EAQYSRIGPNNQAI	172	+++	+++	---	---	---	STQYSRIGPNFAAE	1065	+++	---	---	---
EAQYSRIGPNNQGE	173	++	++	---	---	---	STQYSRIGPNPEAE	1066	++	---	---	---
EAQYSRIGPNNQGE	173	++	++	---	---	---	STQYSRIGPNPEAE	1066	++	---	---	---
EAQYSRIGPNNQPD	174	++	+++	---	---	---	STQYSRIGPNMAE	1067	+++	---	---	---
EAQYSRIGPNNQPD	174	+++	++	---	---	---	STQYSRIGPNMAE	1067	+++	---	---	---
EAQYSRIGPNNQPE	175	+++	+++	---	---	---	STQYSRIGPNQPAE	1068	+++	---	---	---
EAQYSRIGPNNQPE	175	+++	+++	---	---	---	STQYSRIGPNQPAE	1068	+++	---	---	---
EAQYSRIGPNNQPN	176	++	+++	---	---	---	STQYSRIGPNQSE	1070	+++	---	---	---
EAQYSRIGPNNQPN	176	+++	+++	---	---	---	STQYSRIGPNQSE	1070	++	---	---	---
EAQYSRIGPNNQPQ	177	+++	+++	---	---	---	STQYSRIGPNQTE	1071	++	---	---	---
EAQYSRIGPNNQPQ	177	+++	+++	---	---	---	STQYSRIGPNQTE	1071	++	---	---	---
EAQYSRIGPNNQPS	178	+++	+++	---	---	---	STQYSRIGPNQVE	1072	++	---	---	---
EAQYSRIGPNNQPS	178	+++	+++	---	---	---	STQYSRIGPNQVE	1072	++	---	---	---
EAQYSRIGPNNQPT	179	+++	+++	---	---	---	STQYSRIGPNPSAE	1073	+++	---	---	---
EAQYSRIGPNNQPT	179	+++	+++	---	---	---	STQYSRIGPNPSAE	1073	+++	---	---	---
EAQYSRIGPNNQSE	180	+++	+++	---	---	---	STQYSRIGPNPTAE	1074	+++	---	---	---
EAQYSRIGPNNQSE	180	+++	+++	---	---	---	STQYSRIGPNPTAE	1074	+++	---	---	---
EAQYSRIGPNNQTE	181	++	+++	---	---	---	SVQYSRIGPNATAE	1091	++	---	---	---
EAQYSRIGPNNQTE	181	+++	+++	---	---	---	SVQYSRIGPNATAE	1091	+	---	---	---
EAQYSRIGPNNQVE	182	++	+++	---	---	---	SVQYSRIGPNATPE	1092	+++	---	---	---
EAQYSRIGPNNQVE	182	+++	+++	---	---	---	SVQYSRIGPNATPE	1092	++	---	---	---
EAQYSRIGPNNNSAM	183	++	++	---	---	---	SVQYSRIGPNDDTAE	1094	++	---	---	---
EAQYSRIGPNNNSAM	183	++	+	---	---	---	SVQYSRIGPNDDTAE	1094	++	---	---	---
EAQYSRIGPNNNSAQ	184	++	++	---	---	---	SVQYSRIGPNPEAE	1098	++	---	---	---
EAQYSRIGPNNNSAQ	184	++	++	---	---	---	SVQYSRIGPNQPOE	1100	+++	---	---	---

EAQYSRIGPNNSAS	185	++	++	++	++	---	SVQYSRIGPNPQPE	1100	+++	---	---	---
EAQYSRIGPNNSAS	185	++	++	++	++	---	TAQYSRIGPNHTAE	1111	+++	++	++	---
EAQYSRIGPNNSAT	186	++	+++	++	++	---	TAQYSRIGPNHTAE	1111	++	++	++	---
EAQYSRIGPNNSAT	186	+++	++	++	++	---	TAQYSRIGPNITAE	1113	++	++	++	---
EAQYSRIGPNNSPQ	187	++	++	++	++	---	TAQYSRIGPNITAE	1113	++	++	++	---
EAQYSRIGPNNSPQ	187	++	++	++	++	---	TAQYSRIGPNNTAE	1116	+++	---	---	---
EAQYSRIGPNNTAA	188	+++	+++	++	++	---	TAQYSRIGPNNTAE	1116	+++	++	++	---
EAQYSRIGPNNTAA	188	++	+++	++	++	---	TAQYSRIGPNPIAE	1117	++	---	---	---
EAQYSRIGPNNTAD	189	+++	+++	++	++	---	TAQYSRIGPNPIAE	1117	+++	++	++	---
EAQYSRIGPNNTAD	189	++	+++	++	++	---	TAQYSRIGPNPVAE	1119	++	++	++	---
EAQYSRIGPNNTAE	190	+++	+++	++	+	---	TAQYSRIGPNPVAE	1119	+++	++	++	---
EAQYSRIGPNNTAE	190	+++	+++	++	++	---	TPQYSRIGPNATAE	1122	+++	---	---	---
EAQYSRIGPNNTAM	191	++	++	++	++	---	TPQYSRIGPNATAE	1122	+++	---	---	---
EAQYSRIGPNNTAM	191	++	++	++	++	---	TQQYSRIGPNPQAE	1123	+++	---	---	---
EAQYSRIGPNNTAQ	192	++	++	++	++	---	TQQYSRIGPNPQAE	1123	+++	---	---	---
EAQYSRIGPNNTAQ	192	++	++	++	++	---	EAQYSRIGANNQAO	1137	++	+	++	---
EAQYSRIGPNNTAS	193	+++	++++	++	++	---	EAQYSRIGANNQAO	1137	++	---	---	---
EAQYSRIGPNNTAS	193	++	+++	++	++	---	EAQYSRIGVNNQAO	1138	++	+	+	---
EAQYSRIGPNNTAT	194	+++	+++	++	++	---	EAQYSRIGVNNQAO	1138	++	++	++	---
EAQYSRIGPNNTAT	194	+++	+++	++	++	---	EAQYSRIGVNNQAO	1139	+++	---	---	---
EAQYSRIGPNNTPQ	195	+++	++	++	++	---	EAQYSRIGVNNQAO	1139	+++	---	---	---
EAQYSRIGPNNTPQ	195	++	+++	++	++	---	SAQHSRIGPNNOAT	1140	++	---	---	---
EAQYSRIGPNNVAE	196	++	++++	++	++	---	SAQHSRIGPNNOAT	1140	++	---	---	---
EAQYSRIGPNNVAE	196	+++	++	++	++	---	SAQHSRIGPNPQAE	1141	+++	++	++	---
EAQYSRIGPNNVAQ	197	++	++	++	++	---	SAQHSRIGPNPQAE	1141	+++	++	++	---
EAQYSRIGPNNVAQ	197	++	+	++	++	---	SAQYSGIGPNNOAT	1142	+++	---	---	---
EAQYSRIGPNRTAE	201	++++	++++	++	++	---	SAQYSGIGPNNOAT	1142	+++	+	+	---
EAQYSRIGPNRTAE	201	++++	++++	++	++	---	SAQYSRIGANPQAE	1143	++	---	---	---
EFQYSRIGPNATAE	206	++	+++	++	++	---	SAQYSRIGANPQAE	1143	++	---	---	---
EFQYSRIGPNATAE	206	++	+++	++	++	---	SAQYSRIGINPQAE	1144	++	+	+	---
IQYSRIGPNATAE	208	++	++	++	++	---	SAQYSRIGINPQAE	1144	+	+	+	---

[illegible]

GAQYSRIGPNPAE	231	+++	+++	+++	-	---	SAQLHRLGPNQPE	1167	++++	++++	---	---
GAQYSRIGPNPAE	231	++++	+++	+++	-	---	SAQLHRLGPNQPE	1167	+++++	+++	---	---
GAQYSRIGPNPEAE	232	++	+++	+++	---	---	SAQLHRLGPNQPL	1168	+++	+++	---	---
GAQYSRIGPNPEAE	232	+++	+++	+++	---	---	SAQLHRLGPNQPL	1168	++	---	---	---
GAQYSRIGPNPHAE	233	++	++	++	---	---	SAQLHRLGPNQPM	1169	+++	+++	---	---
GAQYSRIGPNPHAE	233	++	++	++	---	---	SAQLHRLGPNQPM	1169	+++	+++	---	---
GAQYSRIGPNPIAE	234	++	+++	+++	---	---	SAQLHRLGPNQPO	1170	+++	+++	---	---
GAQYSRIGPNPIAE	234	++	+++	+++	---	---	SAQLHRLGPNQPO	1170	++	---	---	---
GAQYSRIGPNPLAE	235	+++	++++	++++	---	---	SAQLHRLGPNQPT	1171	++++	++++	---	---
GAQYSRIGPNPLAE	235	+++	+++	+++	---	---	SAQLHRLGPNQPT	1171	+++++	+++	---	---
GAQYSRIGPNMAE	236	++++	+++	+++	---	---	SAQLHRLGPNQSD	1172	+++	+++	---	---
GAQYSRIGPNMAE	236	++++	+++	+++	-	---	SAQLHRLGPNQSD	1172	+++	+++	---	---
GAQYSRIGPNNAE	237	++	++	++	-	---	SAQLHRLGPNQSE	1173	+++	+++	---	---
GAQYSRIGPNNAE	237	++	++	++	---	---	SAQLHRLGPNQSE	1173	+++++	+++	-	---
GAQYSRIGPNPOAE	238	+++	+++	+++	---	---	SAQLHRLGPNQTD	1174	+++	+++	---	---
GAQYSRIGPNPOAE	238	+++	+++	++	---	---	SAQLHRLGPNQTD	1174	+++	+++	---	---
GAQYSRIGPNPQFE	240	++	++	++	---	---	SAQLHRLGPNQTE	1175	+++++	+++	---	---
GAQYSRIGPNPQFE	240	++	++	++	---	---	SAQLHRLGPNQTE	1175	+++++	+++	---	---
GAQYSRIGPNPQGE	241	+++	+++	+++	---	---	SAQLHRLGPNQVE	1176	+++	+++	---	---
GAQYSRIGPNPQGE	241	+++	+++	+++	---	---	SAQLHRLGPNQVE	1176	+++++	+++	---	---
GAQYSRIGPNPQHE	242	+++	+++	+++	---	---	SAQLHRLGPNQPPQ	1177	+++	+++	---	---
GAQYSRIGPNPQHE	242	+++	+++	+++	---	---	SAQLHRLGPNQPPQ	1177	+++	+++	---	---
GAQYSRIGPNPQPE	243	++++	++++	++++	-	---	SAQLHRLGPNQEQ	1178	+++++	+++	---	---
GAQYSRIGPNPQPE	243	++++	+++	+++	-	---	SAQLHRLGPNQEQ	1178	+++++	+++	---	---
GAQYSRIGPNPQPE	244	+++	+++	+++	---	---	SAQLHRLGPNQFAD	1179	+++	+++	---	---
GAQYSRIGPNPQPE	244	+++	+++	++	---	---	SAQLHRLGPNQFAD	1179	+++	+++	---	---
GAQYSRIGPNPQSE	245	++++	+++	+++	-	---	SAQLHRLGPNHQAE	1180	+++	+++	---	---
GAQYSRIGPNPQSE	245	++++	+++	+++	---	---	SAQLHRLGPNHQAE	1180	+++	+++	---	---
GAQYSRIGPNPQTE	246	++++	+++	+++	---	---	SAQLHRLGPNIQAD	1181	++	++	---	---
GAQYSRIGPNPQTE	246	++++	+++	+++	---	---	SAQLHRLGPNIQAD	1181	++	++	---	---
GAQYSRIGPNPQVE	247	++++	+++	+++	---	---	SAQLHRLGPNIQAE	1182	+++	+++	---	---

GAQYSRIGPNPQVE	247	++++	++	--	--	--	SAQLHRLGPNLQAE	1182	+++	--+	--	--
GAQYSRIGPNPQYE	248	++	++	--	--	--	SAQLHRLGPNLQAD	1183	+++	+++	--	--
GAQYSRIGPNPQYE	248	++	++	--	--	--	SAQLHRLGPNLQAD	1183	+++	+++	--	--
GAQYSRIGPNPSAE	249	+++	+++	--	--	--	SAQLHRLGPNLQAE	1184	+++	+++	--	--
GAQYSRIGPNPSAE	249	+++	+++	--	--	--	SAQLHRLGPNLQAE	1184	+++	+++	--	--
GAQYSRIGPNPTAE	250	++++	++++	-	-	-	SAQLHRLGPNNQAD	1185	++	-	--	--
GAQYSRIGPNPTAE	250	++++	++++	--	--	--	SAQLHRLGPNNQAD	1185	++++	+++	--	--
GAQYSRIGPNPVAE	251	+++	++++	--	--	--	SAQLHRLGPNNQAE	1186	++++	+++	--	--
GAQYSRIGPNPVAE	251	++++	+++	--	--	--	SAQLHRLGPNNQAE	1186	++++	+++	-	--
GPOYSRIGPNPQAE	252	++++	++++	--	--	--	SAQLHRLGPNNQAD	1187	+++	+	--	--
GPOYSRIGPNPQAE	252	++++	++++	--	--	--	SAQLHRLGPNNQAD	1187	+++	+	--	--
GVQYSRIGPNPQAE	253	+++	+++	-	-	-	SAQLHRLGPNNQAE	1188	++++	+++	--	--
GVQYSRIGPNPQAE	253	+++	+++	-	-	-	SAQLHRLGPNNQAE	1188	++++	+++	--	--
HAQYSRIGPNPQPE	255	+++	++	--	--	--	SAQLHRLGPNNQPO	1189	+++	+++	--	--
HAQYSRIGPNPQPE	255	++	+	--	--	--	SAQLHRLGPNNQPO	1189	++++	-	--	--
KAQYSRIGPNDTAE	262	+++	+++	--	--	--	SAQLHRLGPNNQVQ	1190	++++	+++	--	--
KAQYSRIGPNDTAE	262	+++	+++	--	--	--	SAQLHRLGPNPQAD	1191	++++	+++	--	--
KAQYSRIGPNETAE	264	++++	+++	--	--	--	SAQLHRLGPNPQAD	1191	+++	+++	--	--
KAQYSRIGPNETAE	264	+++	++	--	--	--	SAQLHRLGPNPQAE	1192	+++++	+++	--	--
KAQYSRIGPNPEAE	265	++	++	--	--	--	SAQLHRLGPNPQAE	1192	+++++	+++	--	--
MAQYSRIGPNATPE	284	+++	++	--	--	--	SAQLHRLGPNPQPO	1193	+++	+++	--	--
MAQYSRIGPNATPE	284	++	++	--	--	--	SAQLHRLGPNPQPO	1193	+++	+++	--	--
MAQYSRIGPNNTAE	287	++++	+++	--	--	--	SAQLHRLGPNQQAE	1194	++++	+++	--	--
MAQYSRIGPNNTAE	287	+++	+++	--	--	--	SAQLHRLGPNQQAE	1194	++++	+++	--	--
NAQYSRIGPNATPE	295	++	++	--	--	--	SAQLHRLGPNQQAT	1195	++++	+++	--	--
NAQYSRIGPNATPE	295	++	+	--	--	--	SAQLHRLGPNQQAT	1195	++++	+++	--	--
NAQYSRIGPNHTAE	297	++	+++	--	--	--	SAQLHRLGPNQQIQ	1196	+++	+	--	--
NAQYSRIGPNHTAE	297	++	++	--	--	--	SAQLHRLGPNQQPO	1197	+++	+++	--	--
NAQYSRIGPNITAE	299	+++	++	--	--	--	SAQLHRLGPNQQPO	1197	+++	+++	--	--
NAQYSRIGPNITAE	299	++	++	--	--	--	SAQLHRLGPNQQVQ	1198	++++	+++	--	--
NAQYSRIGPNLTAE	301	++	++	--	--	--	SAQLHRLGPNQQVQ	1198	++++	++	--	--

NAQYSRIGPNLTAE	301	+	++	--	--	--	SAQLHRLGPNVQAA	1199	+++	--+	--	--
NAQYSRIGPNNTAE	303	+++	++	--	--	--	SAQLHRLGPNVQAA	1199	+++	+	--	--
NAQYSRIGPNNTAE	303	++	++	+	--	--	SAQLHRLGPNVQAE	1200	++++	+	--	--
NAQYSRIGPNPQPE	305	++	+++	--	--	--	SAQLHRLGPNVQAE	1200	+++	+	--	--
NAQYSRIGPNPQPE	305	++	++	--	--	--	SAQLHRLGPNVQAL	1201	+++	+	--	--
QAQYSRIGPNATAE	322	++	++	-	--	--	SAQLHRLGPNVQAL	1201	+++	+	--	--
QAQYSRIGPNATAE	322	++	-	--	--	--	SAQLHRLGPNVQAM	1202	+++	+	--	--
QAQYSRIGPNATPE	324	++	++	-	--	--	SAQLHRLGPNVQAM	1202	+++	+	--	--
QAQYSRIGPNATPE	324	++	++	-	--	--	SAQLHRLGPNVQAO	1203	+++	+	--	--
QAQYSRIGPNATSE	325	++	++	--	--	--	SAQLHRLGPNVQAO	1203	+++	+	--	--
QAQYSRIGPNATSE	325	++	++	--	--	--	SAQLHRLGPNVQAT	1204	+++	+++	--	--
QAQYSRIGPNATTE	326	++	++	--	--	--	SAQLHRLGPNVQAT	1204	+++	+	--	--
QAQYSRIGPNATTE	326	++	++	--	--	--	SAQLHRLGPNVQAD	1205	++++	+++	--	--
QAQYSRIGPNHTAE	329	+++	+++	--	--	--	SAQLHRLGPNVQAD	1205	++++	+	--	--
QAQYSRIGPNHTAE	329	+++	++	-	--	--	SAQLHRLGPNVQAE	1206	+++++	+++	--	--
QAQYSRIGPNITAE	331	++++	++	--	--	--	SAQLHRLGPNVQAE	1206	+++	++	--	--
QAQYSRIGPNITAE	331	+++	++	++	--	--	SPQLHRLGPNVQAD	1207	+++	+++	--	--
QAQYSRIGPNNEAA	332	++	+++	--	--	--	SPQLHRLGPNVQAD	1207	+++	+++	--	--
QAQYSRIGPNNEAA	332	++	++	--	--	--	SPQLHRLGPNVQAE	1208	+++	+++	--	--
QAQYSRIGPNNTAE	334	++++	+++	--	--	--	SPQLHRLGPNVQAE	1208	+++	+++	--	--
QAQYSRIGPNNTAE	334	+++	+++	--	--	--	SPQLHRLGPNVQAO	1209	+++	+++	--	--
QAQYSRIGPNPMAE	335	+++	+++	--	--	--	SPQLHRLGPNVQAO	1209	+++	+++	--	--
QAQYSRIGPNPMAE	335	++++	++	--	--	--	YAQLHRLGPNVQAE	1210	+++	++	--	--
QAQYSRIGPNPQAE	336	+++	+++	--	--	--	YAQLHRLGPNVQAE	1210	+++	+	--	--
QAQYSRIGPNPQAE	336	+++	++	-	--	--	SADFRSTNGVAGHE	1211	+++	+	--	--
QAQYSRIGPNPQPE	338	+++	+++	--	--	--	SADFRSTNGVAGHE	1211	++	+	--	--
QAQYSRIGPNPQPE	338	+++	+++	-	--	--	SADFRSTNGVAGQE	1212	+++	+++	--	--
QAQYSRIGPNPQSE	339	+++	+++	-	--	--	SADFRSTNGVAGQE	1212	+++	+++	--	--
QAQYSRIGPNPQSE	339	+++	++	--	--	--	SADFRSTNGVQAT	1219	+++	+	--	--
QAQYSRIGPNPQVE	340	+++	++	-	--	--	SADFRSTNGVQAT	1219	++	+	--	--
QAQYSRIGPNPQVE	340	++	++	--	--	--	SADFRSTNGVQQE	1221	+++	+++	--	--

SAQYSRIGPNPTAE	341	+++	+++	+++	---	---	SADFRSTNGVAQOE	1221	++	++	---	---
SAQYSRIGPNPTAE	341	+++	+++	---	---	---	SADFRSTNGVSAOE	1224	+++	+	---	---
SAQYSRIGPNPAE	342	+++	---	---	---	---	SADFRSTNGVSAOE	1224	+++	+	---	---
SAQYSRIGPNPAE	342	++	---	---	---	---	SADFRSTNGVSOQE	1225	++	---	---	---
SAQYSRIGPNVTAE	344	++	---	---	---	---	SADFRSTNGVSOQE	1226	+++	---	---	---
SAQYSRIGPNVTAE	344	++	---	---	---	---	SADFRSTNGVSOQE	1226	+++	+	---	---
QPQYSRIGPNATAE	348	++++	---	---	---	---	SASFRSTNGVAQAE	1248	++	+	---	---
QPQYSRIGPNATAE	348	+++	---	---	---	---	SASFESTNGVAQAE	1248	++	+	---	---
SAQYSRIGPNALAE	372	++	---	---	---	---	SKDFRSTNGVAPAE	1249	+++	---	---	---
SAQYSRIGPNALAE	372	++	---	---	---	---	SKDFRSTNGVAPAE	1249	+++	+	---	---
SAQYSRIGPNALAE	385	++	---	---	---	---	SMDFRSTNGVAQFE	1250	++	+	---	---
SAQYSRIGPNALAE	385	++	---	---	---	---	SMDFRSTNGVAQFE	1250	++	---	---	---
SAQYSRIGPNAMAE	390	++	---	---	---	---	SNDFRSTNGVAQOE	1251	++	+	---	---
SAQYSRIGPNAMAE	390	++	---	---	---	---	SNDFRSTNGVAQOE	1251	++	+	---	---
SAQYSRIGPNATAD	487	++	---	---	---	---	SQDFRSTNGVMQAE	1252	+	---	---	---
SAQYSRIGPNATAD	487	++	---	---	---	---	SQDFRSTNGVMQAE	1252	+	+	---	---
SAQYSRIGPNATAE	488	+++	---	---	---	---	SRDFRSTNGVAPAE	1253	+++	---	---	---
SAQYSRIGPNATAE	488	++	---	---	---	---	SRDFRSTNGVAPAE	1253	+++	---	---	---
SAQYSRIGPNATPE	500	+++	---	---	---	---	SSDFRSTNGVAHAIE	1254	+++	+	---	---
SAQYSRIGPNATPE	500	++	---	---	---	---	SSDFRSTNGVAHAIE	1254	+++	---	---	---
SAQYSRIGPNATSE	502	+++	---	---	---	---	SSDFRSTNGVAQAE	1255	+++	---	---	---
SAQYSRIGPNATSE	502	++	---	---	---	---	SSDFRSTNGVAQAE	1255	+++	+	---	---
SAQYSRIGPNATTE	504	++	---	---	---	---	SSDFRSTNGVAQFE	1256	++++	+	---	---
SAQYSRIGPNATTE	504	++	---	---	---	---	SSDFRSTNGVAQFE	1256	++++	---	---	---
SAQYSRIGPNATIVE	506	+++	---	---	---	---	SSDFRSTNGVAQOE	1257	+++	---	---	---
SAQYSRIGPNATIVE	506	++	---	---	---	---	SSDFRSTNGVAQOE	1257	++++	---	---	---
SAQYSRIGPNAAVE	510	+++	---	---	---	---	SSDFRSTNGVAQSE	1258	+++	+	---	---
SAQYSRIGPNAAVE	510	+++	---	---	---	---	SSDFRSTNGVAQSE	1258	+++	+	---	---
SAQYSRIGPNDOAT	539	++	---	---	---	---	SSDFRSTNGVMQAE	1259	++	+	---	---
SAQYSRIGPNDOAT	539	++	---	---	---	---	SSDFRSTNGVMQAE	1259	++	+	---	---
SAQYSRIGPNDTPE	548	++	---	---	---	---	SSDFRSTNGVSAOE	1260	+++	+	---	---

SAQ'YSR(GPNDITPE	548	++	++	++	++	SSDFSTNGVQAQE	1260	+++	++	++	++	++
SAQ'YSR(GPNETTPE	574	++	+++	++	++	STDFRSTNGVAQAE	1263	+++	++	++	++	++
SAQ'YSR(GPNETTPE	574	++	++	++	++	STDFRSTNGVAQAE	1263	+++	++	++	++	++
SAQ'YSR(GPNETTTE	575	++	++	++	++	STDPRSTNGVAQFE	1264	++	++	++	++	++
SAQ'YSR(GPNETTTE	575	++	++	++	++	STDPRSTNGVAQFE	1264	++	++	++	++	++
SAQ'YSR(GPNNHQAQE	594	+++	++	++	++	STDPRSTNGVAQSE	1265	+++	++	++	++	++
SAQ'YSR(GPNNHQAQE	594	+++	++	++	++	STDPRSTNGVAQSE	1265	+++	++	++	++	++
SAQ'YSR(GPNNHTAE	609	++++	+++	++	++	SVDFRSTNGVAPAE	1266	+++	++	++	++	++
SAQ'YSR(GPNNHTAE	609	+++	+++	+++	++	SVDFRSTNGVAPAE	1266	+++	++	++	++	++
SAQ'YSR(GPNNHTPE	610	+++	+++	++	++	SADFRSTNGKAQAE	1299	++	++	++	++	++
SAQ'YSR(GPNNHTPE	610	++	++	++	++	SADFRSTNGKAQAE	1299	++	++	++	++	++
SAQ'YSR(GPNNHTVE	611	++	++++	++	++	SADFRSTNGLAQAE	1300	+++	++	++	++	++
SAQ'YSR(GPNTIAE	635	+++	++	++	++	SADFRSTNGLAQAE	1300	+++	++	++	++	++
SAQ'YSR(GPNTIAE	635	+++	++	++	++	SADFRSTNGMAQAE	1301	++	++	++	++	++
SAQ'YSR(GPNTITPE	637	++	++	++	++	SADFRSTNGMAQAE	1301	++	++	++	++	++
SAQ'YSR(GPNTITPE	637	++	++	++	++	SADFRSTNGRAQAE	1302	++	++	++	++	++
SAQ'YSR(GPNTITPE	647	+++	++	++	++	SADFRSTNGRAQAE	1302	++	++	++	++	++
SAQ'YSR(GPNTITPE	647	+++	++	++	++	SSQFASSTNGRAQAE	1303	++	++	++	++	++
SAQ'YSR(GPNNMTAE	660	+++	++	++	++	SSQFASSTNGRAQAE	1303	++	++	++	++	++
SAQ'YSR(GPNNMTAE	660	+++	++	++	++	SSQFDSSTNGRAQAE	1304	++	++	++	++	++
SAQ'YSR(GPNNMTPE	661	+++	++	++	++	SSQFDSSTNGRAQAE	1304	+++	++	++	++	++
SAQ'YSR(GPNNMTPE	661	++	++	++	++	SSQFHSTNGKAQAE	1305	++	++	++	++	++
SAQ'YSR(GPNNMTSE	662	+++	++	++	++	SSQFHSTNGKAQAE	1305	++	++	++	++	++
SAQ'YSR(GPNNMTSE	662	+++	++	++	++	SSQFHSTNGVAQQE	1306	++	++	++	++	++
SAQ'YSR(GPNNDDAA	664	++	++	++	++	SSQFHSTNGVAQQE	1306	++	++	++	++	++
SAQ'YSR(GPNNDDAA	664	+	+	+	++	SSQFISTNGRAQAE	1307	++	++	++	++	++
SAQ'YSR(GPNNDDAT	666	++	++	++	++	SSQFISTNGKAQAE	1308	++	++	++	++	++
SAQ'YSR(GPNNDDAT	666	++	++	++	++	SSQFISTNGKAQAE	1308	++	++	++	++	++
SAQ'YSR(GPNNNEAA	667	++	+++	++	++	SSQFISTNGRAQAE	1309	++	++	++	++	++
SAQ'YSR(GPNNNEAA	667	++	++	++	++	SSQFISTNGRAQAE	1309	++	++	++	++	++
SAQ'YSR(GPNNNEAQ	668	+++	+++	+++	++	SSQFOSTNGKAQAE	1310	++	++	++	++	++

SAQYSRIGPNNEAQ	668	+++	++	--	--	--	SSQFQSTNGKAQAE	1310	++	--+	--	--
SAQYSRIGPNNEAT	669	++	+++	--	--	--	SSQFQSTNGRAQAE	1311	+++	+	--	--
SAQYSRIGPNNEAT	669	++	++	--	--	--	SSQFQSTNGRAQAE	1311	++	+	--	--
SAQYSRIGPNNEPA	670	++	++	--	--	--	SSQFQSTNGVAQOE	1312	++	+	--	--
SAQYSRIGPNNEPA	670	++	++	--	--	--	SSQFQSTNGVAQOE	1312	++	+	--	--
SAQYSRIGPNNEPA	671	++	++	--	--	--	SSQFSSSTNGKAQAE	1313	++	+	-	--
SAQYSRIGPNNEPA	671	++	++	--	--	--	SSQFSSSTNGKAQAE	1313	+++	+	+	--
SAQYSRIGPNNEPA	672	++	+++	--	--	--	SSQFSSSTNGRAQAE	1314	++	+	--	--
SAQYSRIGPNNETA	672	+	++	--	--	--	SSQFSSSTNGRAQAE	1314	++	+	-	--
SAQYSRIGPNNEVA	673	++++	+++++	--	--	--	SSQFSSSTNGVAQOE	1315	+++	+	--	--
SAQYSRIGPNNEVA	673	+++	+++	--	--	--	SSQFSSSTNGVAQOE	1315	++	+	--	--
SAQYSRIGPNNPAQ	680	+++	+++	--	--	--	SSQFTSTNGKAQAE	1316	++	+	-	--
SAQYSRIGPNNPAQ	680	+++	+++	--	--	--	SSQFTSTNGKAQAE	1316	+++	+	-	--
SAQYSRIGPNNQAE	684	+++	++	-	--	--	SSQFTSTNGRAQAE	1317	++	+	--	--
SAQYSRIGPNNQAE	684	+++	++	--	--	--	SSQFTSTNGRAQAE	1317	++	+	-	--
SAQYSRIGPNNQAT	693	+++	+++	--	--	--	SSQFTSTNGVAQOE	1318	+++	++	--	--
SAQYSRIGPNNQAT	693	+++	+++	--	--	--	SSQFTSTNGVAQOE	1318	++	+	--	--
SAQYSRIGPNNTAE	709	++++	+++	--	--	--	SSQFTSTNGVMQAE	1319	++	+	-	--
SAQYSRIGPNNTAE	709	++++	+++	--	--	--	SSQFTSTNGVMQAE	1319	++	+	-	--
SAQYSRIGPNNTIE	711	++	++	--	--	--	SAFVESTNGRAPAQ	1320	++	+	--	--
SAQYSRIGPNNTIE	711	++	++	--	--	--	SAFVESTNGRAPAQ	1320	+	+	--	--
SAQYSRIGPNNTIE	712	+++	+++	-	--	--	SAHVESTNGRAPAQ	1321	++	+	--	--
SAQYSRIGPNNTIE	712	+++	+++	--	--	--	SAHVESTNGRAPAQ	1321	++	+	--	--
SAQYSRIGPNNTIE	713	++	+++	--	--	--	SAHVESTNGRAQAA	1322	++	+	-	--
SAQYSRIGPNNTIE	713	++	+++	--	--	--	SAHVESTNGRAQAA	1322	++	+	--	--
SAQYSRIGPNNTIE	714	+++	++	--	--	--	SAHVESTNGRAQAE	1323	++	+	-	--
SAQYSRIGPNNTIE	714	+++	++	--	--	--	SAHVESTNGRAQAE	1323	++	+	+	-
SAQYSRIGPNNTIE	715	++++	++++	--	--	--	SAHVESTNGRAQAS	1324	++	-	--	--
SAQYSRIGPNNTIE	715	+++	++++	--	--	--	SAHVESTNGRAQAS	1324	++	-	-	--
SAQYSRIGPNNTSE	716	++++	+++	--	--	--	SAHVESTNGRAQAT	1325	++	+	-	--
SAQYSRIGPNNTSE	716	++++	+++	--	--	--	SAHVESTNGRAQAT	1325	++	-	--	--

SAQYSRIGPNNTTE	717	++++	++++	++	++	++	SAVESTNGRAQAT	1326	++	++	++	+	+
SAQYSRIGPNNTTE	717	++++	++++	++	++	++	SAVESTNGRAQAT	1326	++	++	++	+	+
SAQYSRIGPNNTTE	718	++++	++++	++	++	++	SAVESTNGRMQAQ	1327	++	++	++	++	++
SAQYSRIGPNNTTE	718	++++	++++	++	++	++	SAVESTNGRMQAQ	1327	++	++	++	++	++
SAQYSRIGPNNTTE	719	++	++	++	++	++	SAVESTNGRAPAQ	1328	++	++	++	++	++
SAQYSRIGPNPAE	721	++++	++++	++	++	++	SAVESTNGRAPAQ	1328	++	++	++	++	++
SAQYSRIGPNPAE	721	++++	++++	++	++	++	SAVESTNGRAQAA	1329	++	++	++	++	++
SAQYSRIGPNPAGE	723	+++	+++	++	++	++	SAVESTNGRAQAA	1329	+++	+++	+++	++	++
SAQYSRIGPNPAGE	723	++	++	++	++	++	SAVESTNGRAQAE	1330	+++	+++	+++	++	++
SAQYSRIGPNPAHE	724	++	++	++	++	++	SAVESTNGRAQAE	1330	+++	+++	+++	++	++
SAQYSRIGPNPAHE	724	++	++	++	++	++	SAVESTNGRAQAF	1331	++	++	+++	++	++
SAQYSRIGPNPAPE	725	++++	++++	++	++	++	SAVESTNGRAQAF	1331	++	++	++	++	++
SAQYSRIGPNPAPE	725	++++	++++	++	++	++	SAVESTNGRAQAS	1332	+++	+++	++	++	++
SAQYSRIGPNPASE	726	++++	++++	++	++	++	SAVESTNGRAQAS	1332	+++	+++	++	++	++
SAQYSRIGPNPASE	726	+++	+++	++	++	++	SAVESTNGRAQAT	1333	+++	+++	+++	++	++
SAQYSRIGPNPATE	727	+++	+++	++	++	++	SAVESTNGRAQAT	1333	+++	+++	++	++	++
SAQYSRIGPNPATE	727	++++	++++	++	++	++	SAVESTNGRAQFQ	1334	++	++	++	++	++
SAQYSRIGPNPAVE	728	+++	+++	++	++	++	SAVESTNGRAQFQ	1334	++	++	++	++	++
SAQYSRIGPNPAVE	728	+++	+++	++	++	++	SAVESTNGRAWAQ	1335	++	++	++	++	++
SAQYSRIGPNPEAE	729	++	++	++	++	++	SAVESTNGRMQAQ	1336	++	++	++	++	++
SAQYSRIGPNPEAE	729	++	++	++	++	++	SAVESTNGRMQAQ	1336	++	++	++	++	++
SAQYSRIGPNPEFE	731	++	++	++	++	++	SAMVESTNGRAQAA	1337	++	++	++	++	++
SAQYSRIGPNPEGE	732	+++	+++	+++	++	++	SAMVESTNGRAQAA	1337	++	++	++	++	++
SAQYSRIGPNPEGE	732	++	++	++	++	++	SAMVESTNGRAQAE	1338	++	++	++	++	++
SAQYSRIGPNPEPE	733	+++	+++	+++	++	++	SAMVESTNGRAQAE	1338	++	++	++	++	++
SAQYSRIGPNPEPE	733	++	++	+++	++	++	SAMVESTNGRAQAS	1339	++	++	++	++	++
SAQYSRIGPNPETE	734	++	++	+++	++	++	SAMVESTNGRAQAS	1339	++	++	++	++	++
SAQYSRIGPNPETE	734	++	++	+	++	++	SAMVESTNGRAQAT	1340	++	++	+++	++	++
SAQYSRIGPNPEVE	735	++	++	+++	++	++	SAMVESTNGRAQAT	1340	++	++	++	++	++
SAQYSRIGPNPEVE	735	+++	+++	++	++	++	SANVESTNGRAQAT	1341	++	++	++	++	++
SAQYSRIGPNPGAE	736	++	++	++	++	++	SANVESTNGRAQAT	1341	++	++	++	++	++

SAQYSR(GPNPGAE	736	++	++	++	++	++	SAQVESTNGRAPAA	1342	++	++	++	++	++
SAQYSR(GPNPGPE	737	++	++	++	++	++	SAQVESTNGRAPAA	1342	++	++	++	++	++
SAQYSR(GPNPGPE	737	+++	++	++	++	++	SAQVESTNGRAPAD	1343	+++	++	++	++	++
SAQYSR(GPNPGTE	738	++	++	++	++	++	SAQVESTNGRAPAD	1343	++	++	++	++	++
SAQYSR(GPNPGTE	738	++	++	++	++	++	SAQVESTNGRAPAE	1344	++	++	++	++	++
SAQYSR(GPNPGVE	739	++	++	+	++	++	SAQVESTNGRAPAE	1344	++	++	++	++	++
SAQYSR(GPNPGVE	739	++	++	-	++	++	SAQVESTNGRAPAM	1345	++	++	++	++	++
SAQYSR(GPNPHAE	740	++	++	++	++	++	SAQVESTNGRAPAM	1345	++	++	++	++	++
SAQYSR(GPNPHAE	740	++	++	++	++	++	SAQVESTNGRAPAS	1346	++	++	++	++	++
SAQYSR(GPNPHPE	741	++	++	++	++	++	SAQVESTNGRAPAS	1346	++	++	++	++	++
SAQYSR(GPNPHPE	741	++	++	++	++	++	SAQVESTNGRAPAT	1347	++	++	++	++	++
SAQYSR(GPNPIAE	742	++	++	+++	++	++	SAQVESTNGRAPAT	1347	++	++	++	++	++
SAQYSR(GPNPIAE	742	++	++	++	++	++	SAQVESTNGRAPAY	1348	++	++	++	++	++
SAQYSR(GPNPIPE	744	+++	+++	+++	++	++	SAQVESTNGRAPAY	1348	++	++	++	++	++
SAQYSR(GPNPIPE	744	+++	+++	+++	++	++	SAQVESTNGRAQAE	1349	+++	++	++	++	++
SAQYSR(GPNPIAE	745	+++	+++	+++	++	++	SAQVESTNGRAQAE	1349	++	++	++	++	++
SAQYSR(GPNPIAE	745	+++	+++	+++	++	++	SAQVESTNGRAQAT	1350	++	++	++	++	++
SAQYSR(GPNPLPE	747	++++	++++	++++	++	++	SAQVESTNGRAQAT	1350	++	++	++	++	++
SAQYSR(GPNPLPE	747	++++	++++	++++	++	++	SAQVESTNGRAOSE	1351	+++	++	++	++	++
SAQYSR(GPNPMAE	748	+++	+++	+++	++	++	SAQVESTNGRAOSE	1351	++	++	++	++	++
SAQYSR(GPNPMAE	748	+++	+++	+++	++	++	SAQVESTNGRAOST	1352	++	++	++	++	++
SAQYSR(GPNPMPE	750	+++	+++	+++	-	++	SAQVESTNGRAOST	1352	++	++	++	++	++
SAQYSR(GPNPMPE	750	+++	+++	+++	++	++	SAQVESTNGRAWAE	1353	++	++	++	++	++
SAQYSR(GPNPNAE	751	++	+++	+++	++	++	SAQVESTNGRAWAE	1353	++	++	++	++	++
SAQYSR(GPNPNAE	751	+++	+++	+++	++	++	SAQVESTNGRLPAQ	1354	++	++	++	++	++
SAQYSR(GPNPNS	753	++	++	++	++	++	SAQVESTNGRLPAQ	1354	++	++	++	++	++
SAQYSR(GPNPNS	753	++	++	++	++	++	SAQVESTNGRMOAE	1355	++	++	++	++	++
SAQYSR(GPNPNTE	754	++	++	++	++	++	SAQVESTNGRMOAE	1355	++	++	++	++	++
SAQYSR(GPNPNVE	755	++	++	++	++	++	SAQVESTNGRMOAS	1356	+	++	++	++	++
SAQYSR(GPNPNVE	755	++	++	+	++	++	SAQVESTNGRMOAS	1356	+	++	++	++	++
SAQYSR(GPNPQAD	758	+++	+++	+++	++	++	SAQVESTNGRMOAT	1357	++	++	++	++	++

SAQYSRIGPNQAD	758	+++	++	--	--	--	SAQVESTNGRMQAT	1357	++	--+	--	--
SAQYSRIGPNQAE	759	++++	+++	-	-	--	SAQVESTNGRSPAQ	1358	++	+	--	--
SAQYSRIGPNQAE	759	++++	+++	-	-	--	SAQVESTNGRSPAQ	1358	++	+	--	--
SAQYSRIGPNQFE	771	++	++	--	--	--	SARVESTNGRAPAQ	1359	++	+	-	--
SAQYSRIGPNQFE	771	++	++	--	--	--	SARVESTNGRAPAQ	1359	++	++	--	--
SAQYSRIGPNQGE	773	++++	+++	--	--	--	SARVESTNGRAQAE	1360	++	+	-	--
SAQYSRIGPNQGE	773	++++	+++	--	--	--	SARVESTNGRAQAE	1360	++	-	--	--
SAQYSRIGPNQHE	775	+++	+++	--	--	--	SAVVESTNGRAQAT	1361	++	++	-	--
SAQYSRIGPNQHE	775	+++	++	--	--	--	SAVVESTNGRAQAT	1361	++	+	-	--
SAQYSRIGPNQJE	777	++	+++	--	--	--	SAYVESTNGRAPAQ	1362	++	+	--	--
SAQYSRIGPNQJE	777	+++	+++	--	--	--	SAYVESTNGRAPAQ	1362	++	-	--	--
SAQYSRIGPNQPE	779	++++	++++	--	--	--	SAYVESTNGRAQAT	1363	++	++	-	--
SAQYSRIGPNQPE	779	++++	++++	--	--	--	SAYVESTNGRAQAT	1363	++	+	-	--
SAQYSRIGPNQQE	781	++	+++	--	--	--	SDQVESTNGRAWAQ	1364	++	++	-	--
SAQYSRIGPNQQE	781	++	++	--	--	--	SDQVESTNGRAWAQ	1364	++	++	-	--
SAQYSRIGPNQSE	783	+++++	++++	--	--	--	SDQVESTNGRMQAO	1365	++	+	-	--
SAQYSRIGPNQSE	783	++++	++++	--	--	--	SDQVESTNGRMQAO	1365	++	+	-	--
SAQYSRIGPNQTE	785	++++	++++	--	--	--	SNQVESTNGRAQAE	1366	+++	+	-	--
SAQYSRIGPNQTE	785	++++	++++	--	--	--	SNQVESTNGRAQAE	1366	+++	++	-	--
SAQYSRIGPNQVE	787	+++	+++	-	-	--	SNQVESTNGRAQAT	1367	++	+	-	--
SAQYSRIGPNQVE	787	++++	+++	--	--	--	SNQVESTNGRAQAT	1367	++	++	--	-
SAQYSRIGPNPSAE	791	++++	++++	--	--	--	SNQVESTNGRMQAO	1368	++	++	-	--
SAQYSRIGPNPSAE	791	++++	+++	-	-	--	SNQVESTNGRMQAO	1368	++	+	-	--
SAQYSRIGPNPSGE	793	++	++	--	--	--	SQQVESTNGRAQAT	1369	++	+	--	--
SAQYSRIGPNPSGE	793	++	++	--	--	--	SQQVESTNGRAQAT	1369	++	+	-	--
SAQYSRIGPNPSHE	794	++	++	--	--	--	SQQVESTNGRMQAO	1370	++	++	-	--
SAQYSRIGPNPSHE	794	++	+	--	--	--	SQQVESTNGRMQAO	1370	+	+	-	--
SAQYSRIGPNPSTE	795	++++	++++	--	--	--	SSQVESTNGRAPAQ	1371	++	++	--	--
SAQYSRIGPNPSTE	795	++++	++++	--	--	--	SSQVESTNGRAPAQ	1371	++	+	--	--
SAQYSRIGPNPSSSE	796	++++	++++	--	--	--	SSQVESTNGRAQAA	1372	+++	++	-	--
SAQYSRIGPNPSSSE	796	++++	+++	--	--	--	SSQVESTNGRAQAA	1372	+++	++	+	++

SAQYSR(GPNPSTE	797	+++	+++	+++	---	SSQVESTNGRAQAD	1373	+++	---	+	---
SAQYSR(GPNPSTE	797	+++	+++	---	---	SSQVESTNGRAQAD	1373	++	---	+	---
SAQYSR(GPNPSTE	798	+++	+++	---	---	SSQVESTNGRAQAE	1374	+++	---	+	---
SAQYSR(GPNPSTE	798	+++	+++	---	---	SSQVESTNGRAQAE	1374	+++	---	+	---
SAQYSR(GPNPTAE	799	+++	+++	-	---	SSQVESTNGRAQAF	1375	+++	---	+	---
SAQYSR(GPNPTAE	799	+++	+++	---	---	SSQVESTNGRAQAF	1375	+++	---	+	---
SAQYSR(GPNPTGE	801	+++	+++	---	---	SSQVESTNGRAQAG	1376	+++	---	+	---
SAQYSR(GPNPTGE	801	+++	+++	---	---	SSQVESTNGRAQAG	1376	++	---	+	---
SAQYSR(GPNPTHE	802	+++	+++	---	---	SSQVESTNGRAQAM	1377	+++	---	+	---
SAQYSR(GPNPTHE	802	++	+++	---	---	SSQVESTNGRAQAM	1377	++	---	+	---
SAQYSR(GPNPTPE	803	+++	+++	---	---	SSQVESTNGRAQAN	1378	+++	---	+	---
SAQYSR(GPNPTPE	803	+++	++++	---	---	SSQVESTNGRAQAN	1378	++	---	+	---
SAQYSR(GPNPTSE	804	+++	+++	---	---	SSQVESTNGRAQAQ	1379	++	---	+	---
SAQYSR(GPNPTSE	804	+++	+++	---	---	SSQVESTNGRAQAQ	1379	++	---	+	---
SAQYSR(GPNPTTE	805	+++	+++	---	---	SSQVESTNGRAQAS	1380	+++	---	+	---
SAQYSR(GPNPTTE	805	+++	+++	---	---	SSQVESTNGRAQAS	1380	+++	---	+	---
SAQYSR(GPNPTVE	806	+++	++++	---	---	SSQVESTNGRAQAT	1381	+++	---	+	---
SAQYSR(GPNPTVE	806	+++	+++	---	---	SSQVESTNGRAQAT	1381	+++	---	+	---
SAQYSR(GPNPVAE	807	+++	+++	---	---	SSQVESTNGRAQAV	1382	++	---	+	---
SAQYSR(GPNPVAE	807	++	+++	---	---	SSQVESTNGRAQAV	1382	++	---	+	---
SAQYSR(GPNPVGE	809	+++	+++	---	---	SSQVESTNGRAQAY	1383	++	---	+	---
SAQYSR(GPNPVGE	809	+++	+++	---	---	SSQVESTNGRAQAY	1383	++	---	+	---
SAQYSR(GPNPVHE	810	+++	+++	---	---	SSQVESTNGRAQFQ	1384	++	---	+	---
SAQYSR(GPNPVHE	810	+++	+++	---	---	SSQVESTNGRAQFQ	1384	++	---	+	---
SAQYSR(GPNPVPE	811	+++	+++	---	---	SSQVESTNGRAQQQ	1385	++	---	+	---
SAQYSR(GPNPVPE	811	+++	+++	---	---	SSQVESTNGRAQQQ	1385	+	---	+	---
SAQYSR(GPNPVQE	812	++	+++	---	---	SSQVESTNGRAQSQ	1386	++	---	+	---
SAQYSR(GPNPVQE	812	+++	+++	---	---	SSQVESTNGRAQSQ	1386	++	---	+	---
SAQYSR(GPNPVSE	813	+++	+++	---	---	SSQVESTNGRAWAQ	1387	++	---	+	---
SAQYSR(GPNPVSE	813	+++	+++	---	---	SSQVESTNGRAWAQ	1387	+	---	+	---
SAQYSR(GPNPVVE	814	+++	+++	---	---	SSQVESTNGRMOAQ	1388	++	---	+	---

SAQYSRIGPNPYVE	814	++	+++	--	--	SSQVESTNGRMQAQ	1388	++	--+	--	--	--
SAQYSRIGPNQIPE	827	++	++	--	--	SAQLCKPCLTNQAQ	1389	+++	+	--	--	--
SAQYSRIGPNQIPE	827	++	+	--	--	SAQLCKPCLTNQAQ	1389	++	+	--	--	--

CLAIMS

What is claimed is:

1. An engineered adeno associated virus (AAV) capsid polypeptide comprising a transferrin receptor (TfR1) binding modification defined by the formula $X_1-X_2-X_3$ -[7-mer]- $X_4-X_5-X_6-X_7$, wherein the 7-mer is one of YSRIGPN (SEQ ID NO: 1390), YSRNSDN (SEQ ID NO: 1391), LHRLGPN (SEQ ID NO: 1392), FRSTNGV (SEQ ID NO: 1393), FVSTNGV (SEQ ID NO: 1394), FZ₁STNGZ₂ (SEQ ID NO: 1395), FRSTNGZ₃ (SEQ ID NO: 1396), or VESTNGR (SEQ ID NO: 1397), and wherein the 7-mer is inserted between amino acids 588 and 589 of an AAV9 capsid polypeptide, or in an analogous position of a capsid polypeptide of another AAV serotype, and wherein X_1 , X_2 , X_3 , X_4 , X_5 , X_6 , X_7 indicate an amino acid modification at one or more amino acid positions in the capsid polypeptide flanking the inserted 7-mer.

2. The engineered AAV capsid polypeptide of claim 1, wherein the 7-mer comprises or consists of YSRIGPN (SEQ ID NO: 1390), YSRNSDN (SEQ ID NO: 1391), or LHRLGPN (SEQ ID NO: 1392), and wherein

X_1 is A, G, E, L, N, Q, S, W, or M;

X_2 is A, F, I, L, M, N, Q, P, T, V, or Y;

X_3 is Q;

X_4 is any amino acid;

X_5 is D, F, G, I, L, M, Q, P, S, T, or V;

X_6 is A, C, F, G, H, I, P, S, T, V, W, or Y;

X_7 is D, E, Q, or T;

or any combination thereof.

3. The engineered AAV capsid polypeptide of claim 2, wherein the TfR1 binding modification comprises or consists of the amino acid sequence of one of SEQ ID Nos: 60-1210.

4. The engineered AAV capsid polypeptide of claim 2, wherein the n-mer comprises or consists of YSRIGPN (SEQ ID NO: 1390), and wherein

X_1 is A, G, E, L, N, Q, S, W, or M;

X_2 is of A, F, I, L, M, N, Q, P, T, V, or Y;

X₃ is of Q;
X₄ is of A, E, F, H, I, L, M, N, P, Q, V, Y, D, or G;
X₅ is of D, F, G, I, L, M, Q, P, S, T, or V;
X₆ is of A, C, F, G, H, I, P, S, T, V, W, or Y;
X₇ is of D, E, Q, or T;
or any combination thereof.

5. The engineered AAV capsid polypeptide of claim 4, wherein the TfR1 binding modification comprises or consists of the amino acid sequence of one of SEQ ID Nos: 60-1136.

6. The engineered AAV capsid polypeptide of claim 2, wherein the n-mer comprises or consists of YSRNSDN (SEQ ID NO: 1391) and

X₁ is of S, or W;
X₂ is of V, I, or F;
X₃ is of Q;
X₄ is of any amino acid;
X₅ is of Q or T;
X₆ is of A;
X₇ is of Q;
or any combination thereof.

7. The engineered AAV capsid polypeptide of claim 6, wherein the TfR1 binding modification comprises or consists of the amino acid sequence of one of SEQ ID Nos: 1152-1157.

8. The engineered AAV capsid polypeptide of claim 2, wherein the n-mer comprises or consists of LHRLGPN (SEQ ID NO: 1392) and

X₁ is of S, A, L, or M;
X₂ is of A or P;
X₃ is of Q;
X₄ is of A, E, F, H, I, L, M, N, P, Q, V, or Y;
X₅ is of Q;
X₆ is of A, P, S, or T;

X₇ is of D, E, Q, or T;
or any combination thereof.

9. The engineered AAV capsid polypeptide of claim 8, wherein the TfR1 binding modification comprises or consists of the amino acid sequence of one of SEQ ID Nos: 1158-1210.

10. The engineered AAV capsid polypeptide of claim 1, wherein the 7-mer comprises or consists of FRSTNGV (SEQ ID NO: 1393), FVSTNGV (SEQ ID NO: 1394), FZ₁STNGZ₂ (SEQ ID NO: 1395), FRSTNGZ₃ (SEQ ID NO: 1396), or VESTNGR (SEQ ID NO: 1397) and wherein

X₁ is of S;
X₂ is of A, S, M, or D;
X₃ is of F, H, I, L, M, N, Q, R, Y, D, or E;
X₄ is of A, S, or M;
X₅ is of Q or P;
X₆ is of A, F, H, Q, or S;
X₇ is of A, D, E, F, Q, S, or T;
or any combination thereof.

11. The engineered AAV capsid polypeptide of claim 9, wherein the TfR1 binding modification comprises or consists of the amino acid sequence of one of SEQ ID Nos: 1211-1389.

12. The engineered AAV capsid polypeptide of claim 9, wherein the n-mer comprises or consists of FRSTNGV (SEQ ID NO: 1393), and

X₁ is of S;
X₂ is of A or S;
X₃ is of D;
X₄ is of A, or S;
X₅ is of Q or P;
X₆ is of A, F, H, Q, or S;
X₇ is of D, E, Q, or T;
or any combination thereof.

13. The engineered AAV capsid polypeptide of claim 12, wherein the TfR1 binding modification comprises or consists of the amino acid sequence of one of SEQ ID Nos: 1211-1266.
14. The engineered AAV capsid polypeptide of claim 9, wherein the n-mer comprises or consists of FVSTNGV (SEQ ID NO: 1394), and
- X₁ is of S;
 - X₂ is of A, S, or M;
 - X₃ is of Q, E, or D;
 - X₄ is of A, or M;
 - X₅ is of Q or P;
 - X₆ is of A;
 - X₇ is of E;
 - or any combination thereof.
15. The engineered AAV capsid polypeptide of claim 14, wherein the TfR1 binding modification comprises or consists of the amino acid sequence of one of SEQ ID Nos: 1267-1298.
16. The engineered AAV capsid polypeptide of claim 9, wherein the n-mer comprises or consists of FZ₁STNGZ₂ (SEQ ID NO: 1395) or FRSTNGZ₃ (SEQ ID NO: 1396), and
- X₁ is of S;
 - X₂ is of A or S;
 - X₃ is of Q or D;
 - X₄ is of A;
 - X₅ is of Q;
 - X₆ is of A;
 - X₇ is of E;
 - or any combination thereof.
17. The engineered AAV capsid polypeptide of claim 16, wherein the TfR1 binding modification comprises or consists of the amino acid sequence of one of SEQ ID Nos: 1299-1319.

18. The engineered AAV capsid polypeptide of claim 9, wherein the 7-mer comprises or consists of VESTNGR (SEQ ID NO: 1397), and
- X₁ is of S;
 - X₂ is of S or D;
 - X₃ is of F, H, I, L, M, N, Q, R, or Y;
 - X₄ is of A;
 - X₅ is of Q or P;
 - X₆ is of A;
 - X₇ is of A, D, E, F, Q, S, or T;
 - or any combination thereof.
19. The engineered AAV capsid polypeptide of claim 17, wherein the TfR1 binding modification comprises or consists of the amino acid sequence of one of SEQ ID Nos: 1320-1388.
20. The engineered AAV capsid polypeptide of any one or the preceding claims, wherein Z₁ is selected from the group consisting of A, D, H, N, Q, and S; Z₂ is selected from the group consisting of K and R; and Z₃ is selected from the group consisting of L, M, and R.
21. The engineered AAV capsid polypeptide of claim 1, wherein the TfR1 binding modification comprises or consists of an amino acid sequence of EAQYSRIGPNNQAQ; SADFRSTNGVAQAE; or GAQYSRIGPNPQPE.
22. The engineered AAV capsid polypeptide of any of the preceding claims, wherein the viral capsid protein is VP1, VP2, VP3, or a combination thereof.
23. The engineered AAV capsid polypeptide of any of the preceding claims, wherein the other AAV serotype is AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAVrh8, AAV rh.74, or AAV rh. 10.

24. An engineered AAV particle comprising the engineered AAV capsid polypeptide of any one of the preceding claims, and further comprising a recombinant AAV genome encoding a transgene.
25. The engineered AAV particle of claim 24, wherein the transgene encodes a therapeutic polypeptide, an antibody or fragment thereof, a siRNA, a CRISPR-Cas system, Transcription Activator-like Effector (TALE)- or Zinc Finger Protein (ZFP)-based transcriptional activator; repressor; or epigenomic silencer, a RNA encoding a partial gene fragment designed for transplacing into an endogenous RNA, one or more transfer RNAs, or a component thereof, or an OMEGA system or any component thereof.
26. The engineered AAV particle of claims 24 or 25, wherein the transgene is operably linked to a regulatory sequence that promotes expression in the CNS.
27. A pharmaceutical composition comprising the recombinant engineered AAV particle of any one of claims 24 to 26 and an acceptable carrier.
28. A method of delivering a polypeptide or polynucleotide to the CNS of a subject comprising administering the pharmaceutical composition of claim 27 to the subject.
29. The method of claim 28, wherein the pharmaceutical composition is delivered at a dosage between 0.1×10^{12} vg/kg to 1×10^{14} vg/kg.
30. The method of claim 29, wherein the dosage is between 0.1×10^{12} vg/kg to 100×10^{12} vg/kg.
31. The method of claim 30, wherein the dosage is between 1×10^{12} vg/kg to 10×10^{12} vg/kg.
32. The method of claim 31, wherein the dosage is 5×10^{12} vg/kg.
33. The method of claim 31, wherein the dosage is 2×10^{12} vg/kg.

34. The method of any one of claims 28 to 33, wherein the pharmaceutical composition is administered systemically or directly to the CNS.
35. A method of manufacturing a recombinant engineered AAV, said method comprising:
culturing mammalian cells comprising; (1) a polynucleotide encoding the engineered AAV capsid polypeptide of any one of claims 1 to 26, (2) a polynucleotide encoding a recombinant AAV genome comprising a transgene operably linked to a regulatory sequence and flanked by AAV ITR sequences, and optionally (3) polynucleotide encoding adenoviral helper genes, under conditions sufficient for production of recombinant engineered AAV particles; and
recovering the recombinant engineered AAV particles from said culture.
36. A host cell for production of recombinant engineered AAV particles which comprises a polynucleotide encoding the engineered AAV capsid polypeptide of any one of claims 1 to 26, a polynucleotide encoding a recombinant AAV genome comprising a transgene operably linked to a regulatory sequence and flanked by AAV ITR sequences, and optionally a polynucleotide encoding adenoviral helper genes.
37. An AAV library comprising a population of variant engineered recombinant AAV particles wherein each member of the population of variant engineered recombinant AAV particles comprises a variant recombinant AAV capsid polypeptide having a modification defined by the formula $X_1-X_2-X_3$ -[7-mer]- $X_4-X_5-X_6-X_7$, and wherein the 7-mer is inserted between amino acids 588 and 589 of an AAV9 capsid polypeptide, or in an analogous position of a capsid polypeptide of another AAV serotype, and wherein X_1 , X_2 , X_3 , X_4 , X_5 , X_6 , X_7 indicate modifications at one or more amino acid positions in the capsid polypeptide flanking the inserted 7-mer, wherein the modification has been selected for binding of TfR1 and/or increased tropism for the CNS relevant to a reference AAV particle without the modification.
38. The AAV library of claim 37, wherein the AAV serotype is AAV1, AAV2, AAV3, AAV4, AAV5, AAV6, AAV7, AAV8, AAVrh8, AAV rh.74, or AAV rh. 10.
39. A method of screening an AAV library for a recombinant AAV particle which binds TfR1 and/or has increased tropism for the CNS, said method comprising assaying the AAV library

of claim 37 for increased binding of Tfr1 and/or tropism for the CNS relative to an AAV vector with a reference capsid and selecting those recombinant AAV vectors which have increased binding of Tfr1 and/or tropism for the CNS.

40. A method for training a machine learning algorithm comprising:
- receiving, by at least one computing device, a plurality of AAV capsid polypeptide sequences comprising a modification for binding to a Tfr1;
 - training, by at least one computing device, with the plurality of AAV capsid polypeptide sequences comprising a modification for binding to a Tfr1, a transferrin receptor (TfR1) targeting machine learning model; and
 - deploying, by at least one computing device, the TfR1 targeting machine learning algorithm.
41. The method of claim 40, wherein the TfR1 targeting machine learning model is trained to identify one or more sequences from the plurality of sequences within increased binding of Tfr1.
42. The method of any one of claims 40-41, wherein the TfR1 targeting machine learning model is trained to identify one or more sequences from the plurality of sequences that decreases transduction of off-target tissues.
43. The method of any one of claims 40-42, wherein the TfR1 targeting machine learning model is trained to identify one or more sequences from the plurality of sequences that binds to a extracellular domain of TFRC.
44. The method of any one of claims 40-43, wherein the TfR1 targeting machine learning model is trained to identify one or more sequences from the plurality of sequences that binds to a apical domain.
45. The method of any one of claims 40-44, wherein the modification is defined by the formula $X_1-X_2-X_3-[7\text{-mer}]-X_4-X_5-X_6-X_7$.

46. The method of any of claims 40-45, wherein the training comprises unsupervised learning, supervised learning, semi-supervised learning, reinforcement learning, transfer learning, incremental learning, curriculum learning, learning to learn, or contrastive learning.

47. The method of claim 40-46, wherein the transferrin receptor (TfR1) targeting machine learning model comprises of linear classifiers, logistic classifiers, random forest, artificial neural networks, matrix factorization, support vector machines, K-means clustering, or K-nearest neighbor.

48. The method of claim 40-46 wherein the transferrin receptor (TfR1) targeting machine learning model comprises of Boltzmann machines, Bayesian networks, autoregressive models, variational auto encoders (VAEs), diffusion models, energy-based models, flow-based models, generative adversarial networks (GANs), mixture models, hidden Markov models, or large language models (LLMs).

49. The method of claim 40-46 wherein the transferrin receptor (TfR1) targeting machine learning model comprises of convolutional neural networks (CNNs), recurrent neural networks (RNNs), long short-term memory models (LSTMs), gated recurrent units (GRUs), capsule networks, attention mechanisms, or transformer networks.

50. The method of claim 40-47, the transferrin receptor (TfR1) targeting machine learning model is a pre-trained, and further trained to predict transferrin receptor (TfR1) targeting by a plurality of sequences.