SUPPLEMENTAL INFORMATION

SCHEMA computational design of virus capsid chimeras: calibrating how genome packaging, protection, and transduction correlate with calculated structural disruption

Michelle L. Ho[1], Benjamin A. Adler[1,2], Michael L. Torre[1], Jonathan J. Silberg[1,2*], and Junghae Suh[1*]

[1]Department of Bioengineering, Rice University, Houston, TX 77005

[2]Department of Biochemistry and Cell Biology, Rice University, Houston, TX 77005

*Corresponding authors: joff@rice.edu and jsuh@rice.edu

MLH: michelle.liane.ho@gmail.com

BAA: badler91@gmail.com

MLT: mltorre3@gmail.com

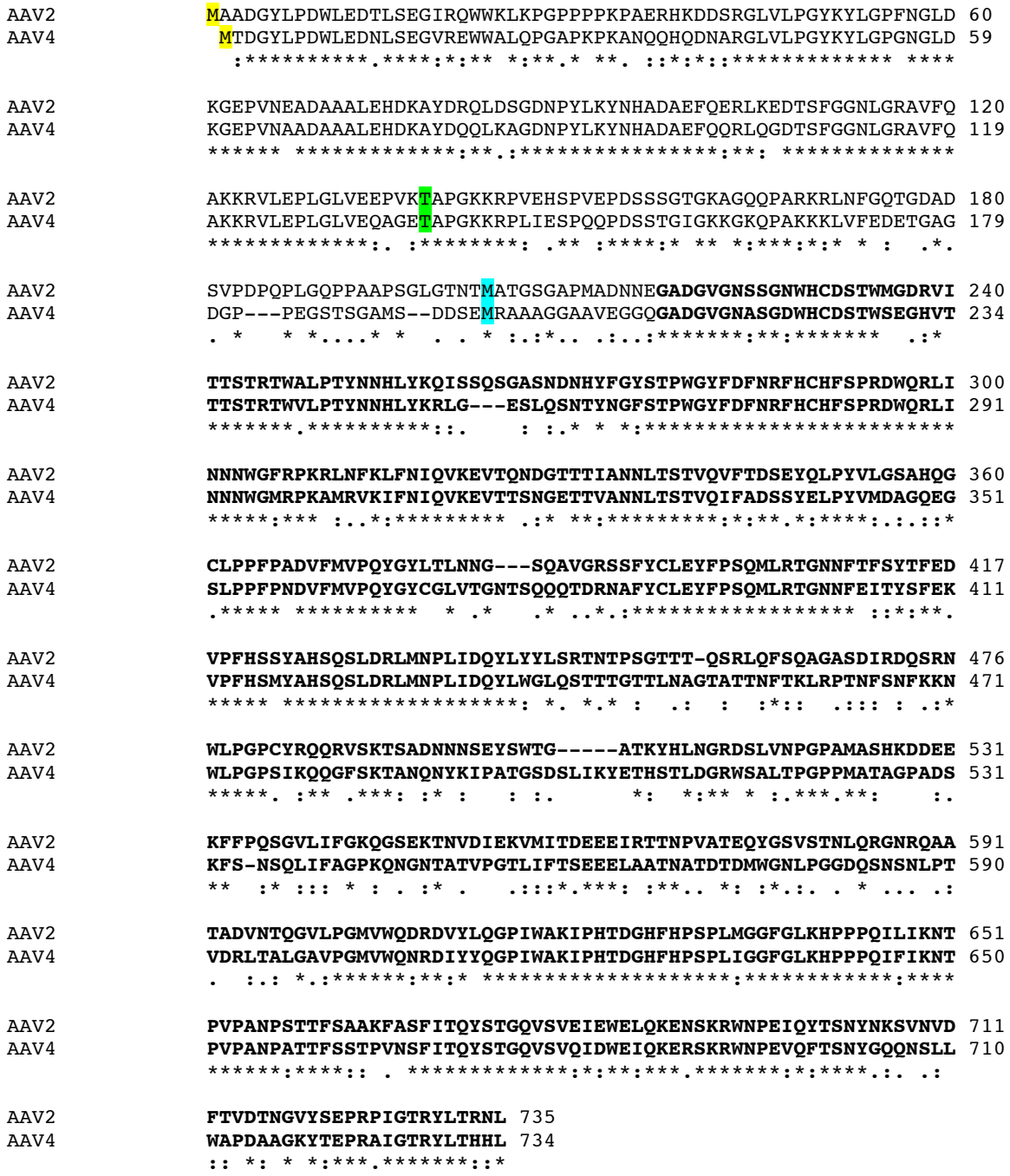

Figure S1. Sequence alignment of the AAV capsid proteins recombined. The alignment of VP subunits from AAV2 and AAV4 was generated using BLAST. Numbering for each residue is based on each VP1. The highlighting of the first residues of VP1 (yellow), VP2 (green), and VP3 (cyan) shows how the majority of the VP primary sequences are identical within these proteins. The bold residues represent those considered by SCHEMA because they represent the portion of the AAV4 VP sequence for which structural coordinates are available.

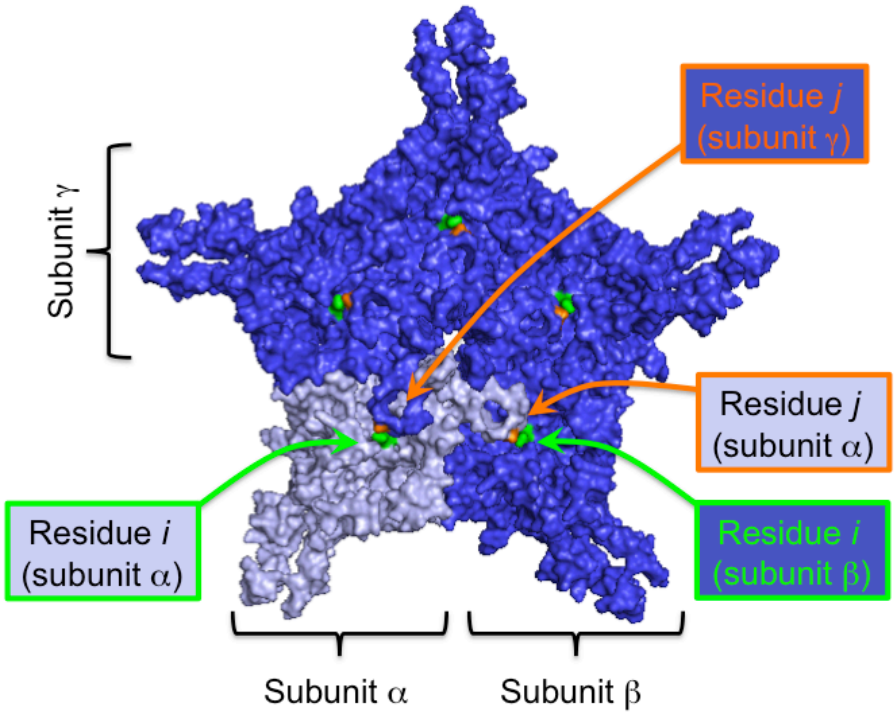

Figure S2. Scaling intermolecular contacts from subunits to capsids. A residue-residue contact made by VP residues *i* and j are shown for five subunits at a 5-fold axis of symmetry within the capsid structure, including subunits α, β, and γ. When counting the number of *ij* residue-residue contacts broken by recombination that are intermolecular on a per subunit basis, two contacts are counted, including the contact made between residue *i* within the α subunit and residue *j* within the β subunit and the contact made between residue *j* within the α subunit and residue *i* with the γ subunit. For this reason, total disruption per subunit ($E_{subunit}$) was calculated as the sum of E_{intra} and 0.5 x E_{inter}.

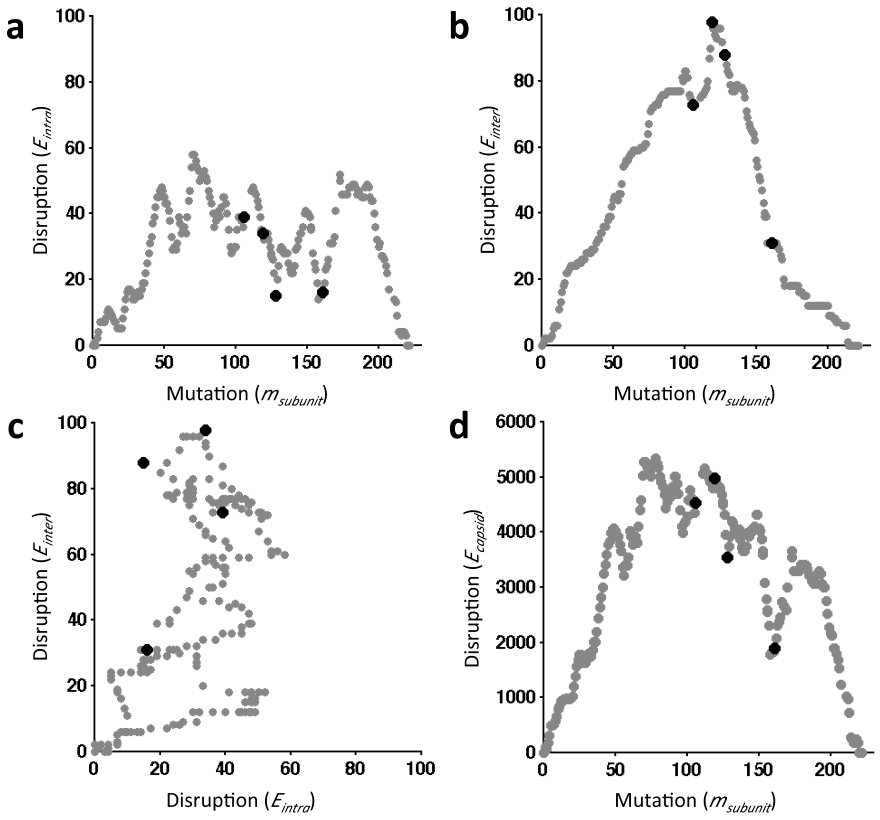

Figure S3. E_{intra} and E_{inter} for single-crossover chimeras. (**a**) Intramolecular and (**b**) intermolecular disruption are plotted versus amino acid substitution level for each possible single-crossover chimera (gray circles) as well as each chimera generated and characterized (black circles). (**c**) The relative contributions that E_{intra} and E_{inter} make to disruption in each chimeric subunit varies. (**d**) The per subunit disruption is shown for each possible single-crossover chimera.

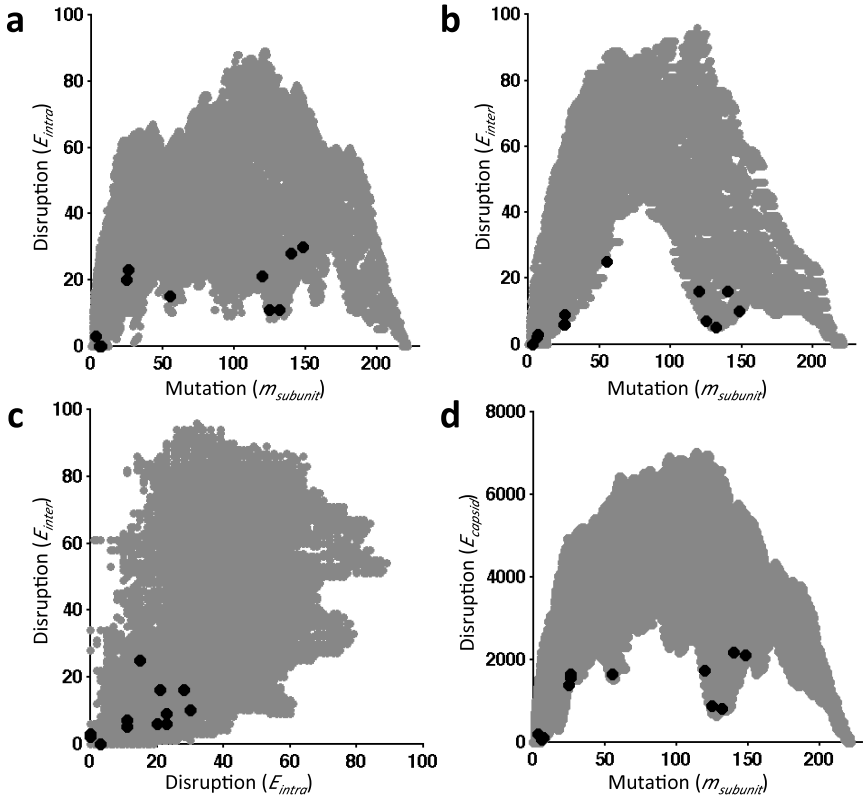

Figure S4. E_{intra} and E_{inter} for double-crossover chimeras. (**a**) Intramolecular and (**b**) intermolecular disruption are plotted versus amino acid substitution level for each possible double-crossover chimera (gray circles) as well as each chimera generated and characterized (black circles). (**c**) The relative contributions that E_{intra} and E_{inter} make to disruption in each chimeric subunit varies. (**d**) The per subunit disruption is shown for each possible double-crossover chimera.

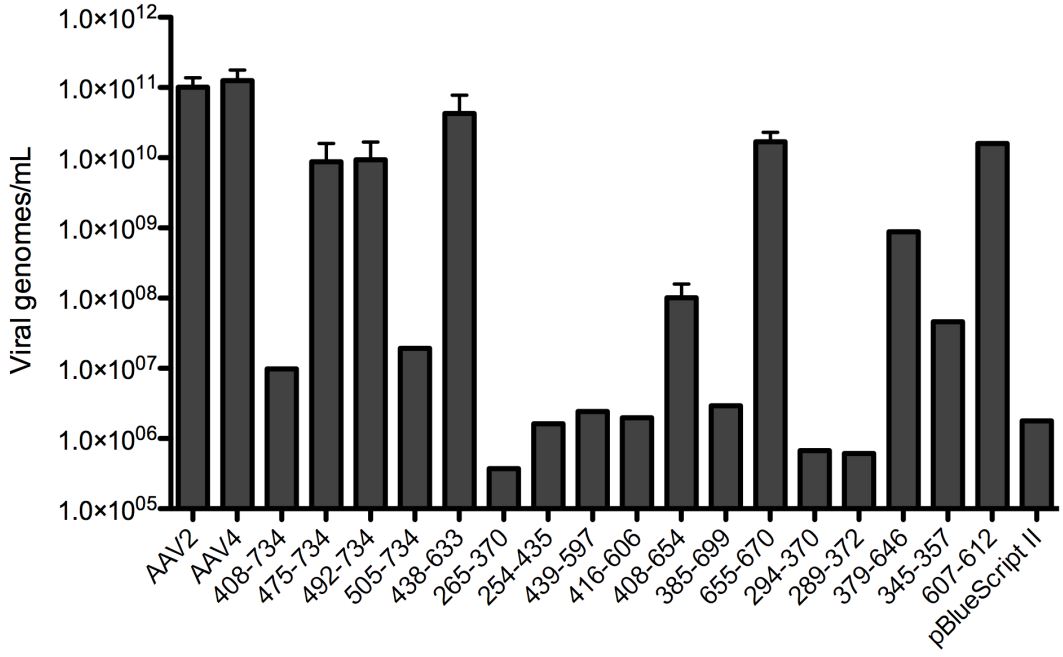

Figure S5. Genomic titers of chimeras. Q-PCR was used to measure viral genomic titers of AAV2, AAV4, and chimeric viruses extracted from the 40% iodixanol layer upon centrifugation. pBlueScriptII served as a negative control to determine the background threshold (black dotted line) for detectable genomic titers. Genomic titers that were within 1-log fold of our negative control titer (pBlueScriptII) were considered not detectable above background, *i.e.*, those below 2 x10^7. Error bars represent the SEM calculated from duplicate sample measurements from one to six independent experiments.

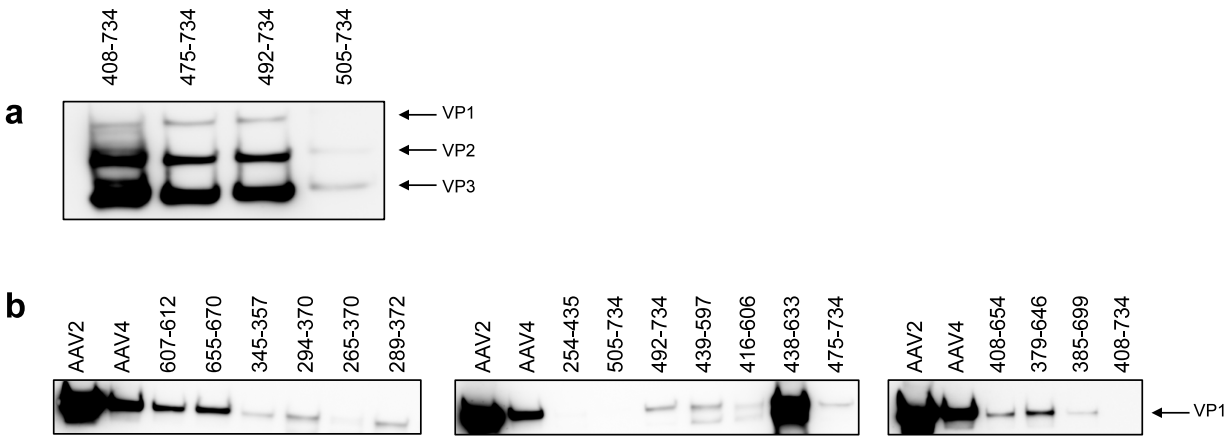

Figure S6. Immunoblot analysis of chimeric VP expression. Western blot analysis of chimeric capsids extracted from the 25-40% iodixanol interface after separation using an iodixanol gradient. Viral proteins were detected with (**a**) monoclonal anti-VP (B1) antibody that only binds to the C-termini of VP1, VP2, and VP3 from AAV2 and (**b**) monoclonal anti-VP1 (A1) antibody that binds to both AAV2 and AAV4 VP1. The B1 antibody staining is stronger primarily due to recognition of VP3, which is expressed at higher stoichiometric ratios than VP1 and VP2. This antibody was limited to analyzing the four chimeras with AAV2 C-termini in their VPs. For this reason, the A1 antibody was used to measure expression of all VP1 chimeras.

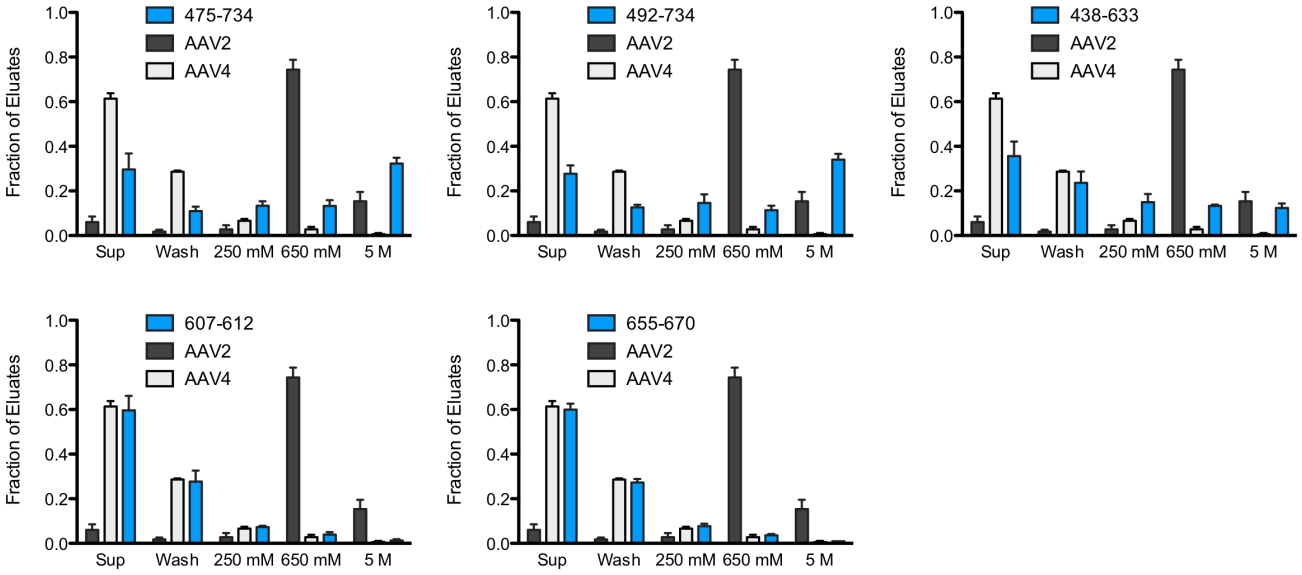

Figure S7. Binding of chimeric capsids to heparin. Heparin affinity chromatography was performed for five chimeras (*blue*), AAV2 (*gray*), and AAV4 (*white*). Q-PCR was used to quantify the number of viruses that did not bind to the column, eluted when the column was washed with buffer, and eluted when the column was washed with buffer containing 250 mM, 650 mM, and 5 M NaCl. Each elution is reported as a fraction of total viruses detected in the five eluates: supernatant (Sup), wash (Wash), low salt wash (250 mM), intermediate salt wash (650 mM), and high salt wash (5 M). Error bars indicate the SEM calculated from three independent experiments.

tcgcgcgtttcggtgatgacggtgaaaacctctgacacatgcagctcccggagacggtcacagcttgtctgtaagcggatgccgggagcagacaagcccgtcag
ggcgcgtcagcgggtgttggcggtgtcggggctggcttaactatgcggcatcagagcagattgtactgagagtgcaccatatgcggtgtgaaataccgcacaga
tgcgtaaggagaaaataccgcatcaggaaattgtaaacgttaatatttgttaaaattcgcgttaaatttttgttaaatcagctcatttttttaaccaataggccg
aaatcggcaaaatcccttataaatcaaaagaatagaccgagatagggttgagtgttgttccagtttggacaagagtccactattaaagaacgtggactccaacg
tcaaagggcgaaaaaccgtctatcagggcgatggcccactacgtgaaccatcaccctaatcaagtttttggggtcgaggtgccgtaaagcataaatcggaacc
ctaaagggagccccccgatttagagcttgacggggaaagccggcgaacgtggcgaggaaggaagggaagaaagcgaaaggagcgggcgctagggcgctggcaagt
gtagcggtcacctgcgcgtaaccaccacacccgccgcgcttaatgcgccgctacagggcgcgtcgcgccattcgccattcaggctgcgcaactgttgggaaggg
cgatcggtgcgggcctcttcgctattacgccagcggcgaaaggggatgtgctgcaaggcgattaagttgggtaacgccagggttttcccagtcacgacgttgt
aaaacgacggccagtgaattgtaatacgactcactatagggcgaattcgagctcggtcccctagagtcctgtattagaggtcacgtgagtgttttgcgacattt
tgcgacaccatgtggtcacgctgggtatttaagcccgagtgagcacgcagggtctccattttgaagcgggaggtttgaaccgcagccgccatgccggggtttta
cgagattgtgattaaggtccccagcgaccttgacgggcatctgcccggcatttctgacagctttgtgaactgggtggccgagaaggaatgggagttgccgccaa
ttctgacatggatctgaatctgattgagcaggcacccctgaccgtggccgagaagctgcagcgcgactttctgacggaatggcgccgtgtgagtaaggccccgg
aggccctttttctttgtgcaattgagaagggagagagctacttccacatgcacgtgctcgtggaaaccaccgggggtgaaatccatggttttgggacgtttcctga
gtcagattcgcgaaaaactgattcagagaatttaccgcgggatcggccgactttgccaaactggttcgcggtcacaaagaccagaaatggcgccggaggcggga
acaaggtggtggatgagtgctacatccccaattacttgctccccaaaacccagcctgagctccagtggcgtggactaatatggaacagtatttaagcgcctgtt
tgaatctcacggagcgtaaacggttggtggcgcagcatctgacgcacgtgtcgcagacgcaggagcagaacaaagagaatcagaatcccaatctgatgcgccgg
tgatcagatcaaaaacttcagccaggtacatggagctggtcgggtggctcgtggacaaggggattacctcggagaagcagtggatccaggaggaccaggcctca
tacatctcctcaatgcggcctccaactcgcggtcccaaatcaaggctgccttggacaatgcgggaaagattatgagcctgactaaaaccgccccgactacctg
gtgggccagcagcccgtggaggacatttccagcatcggatttataaaatttggaactaaacgggtacgatccccaatatgcggcttccgtctttctgggatgg
gccacgaaaaagttcggcaagaggaacaccatctggctgtttgggcctgcaactacgggaagaccaacatcgcggaggccatagcccacactgtgcccttctac
gggtgcgtaaactggaccaatgagaactttcccttcaacgactgtgtcgacaagatggtgatctggtgggaggaggggaagatgaccgccaaggtcgtggagtc
ggccaaagccattctcggaggaagcaaggtgcgcgtggaccagaaatgcaagtcctcggcccagatagacccgactcccgtgatcgtcacctccaacaccaaca
tgtgcgccgtgattgacgggaactcaacgaccttcgaacaccagcagccgttgcaagaccggatgttcaaatttgaactcacccgccgtctggatcatgacttt
gggaaggtcaccaagcaggaagtcaaagacttttccggtgggcaaaggatcacgtggttgaggtggagcatgaattctacgtcaaaaagggtggagccaagaa
aagacccgccccagtgacgcagatataagtgagcccaaacgggtgcgcgagtcagttgcgcagccatcgacgtcagacgcggaagcttcggtggactacgcgg
acaggtaccaaaacaaatgttctcgtcacgtgggtatgaatctgatgctttttccctgccggcaatgcgagagaatgaatcagaatgtggacatttgcttcacg
cacggggtcatggactgtgccgagtgcttccccgtgtcagaatctcaacccgtgtctgtcgtcagaaagcggacgtatcagaaactgtgtccgattcatcacat
catggggagggcgcccgaggtggcctgctcggcctgcgaactggccaatgtggacttggatgactgtgacatggaacaataaatgactcaaaccagatatgact
gacggttaccttccagattggctagaggacaacctctctgaaggcgttcgagagtggtgggcgctgcaacctggagcccctaaacccaaggcaaatcaacaaca
tcaggacaacgctcggggtcttgtgcttccgggttacaaatacctcggacccggcaacggactcgacaagggggaacccgtcaacgcagcggacgcggcagccc
tcgagcacgacaaggcctacgaccagcagctcaaggccggtgacaacccctacctcaagtacaaccacgccgacgcggagttccagcagcggcttcagggcgac
ACTAGTaaaaATGCATaaaaGCTAGCcctgttaatcaataaaccggtttattcgtttcagttgaactttggtgtcgcggccgctcgataagctttgttccctt
tagtgagggttaattccgagctggcgtaatcatggtcatagctgtttcctgtgtgaaattgttatccgctcacaattccacacaacatacgagccggaagcata
aagtgtaaagcctggggtgcctaatgagtgagctaactcacattattgcgttgcgctcactgcccgctttccagtcgggaaacctgtcgtgccagctgcattaa
tgaatcggccaacgcgcggggagaggcggtttgcgtattgggcgctcttccgcttcctcgctcactgctcgctgcgctcggtcgttcggctgcggcgagcggta
tcagctcactcaaaggcggtaatacggttatccacagaatcaggggataacgcaggaaagaacatgtgagcaaaaggccagcaaaaggccggaaccgtaaaaag
gccgcgttgctggcgtttttccataggctccgccccctgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagata
ccaggcgttcccctggaagctccctcgtgcgctctcctgttccgaccctgccgcttaccggatacctgtccgcctttctcccttcgggaagcgtggcgctttc
tcatagctcacgctgtaggtatctcagttcggtaggtcgttcgctccaagctgggctgtgtgcacgaaccccccgttcagcccgaccgctgcgccttatccgg
taactatcgtcttgagtccaacccggtaagacacgacttatcgccactggcagcaccactggtaacaggattagcagagcgaggtatgtaggcggtgctacaga
gttcttgaagtggtggcctaactacggctacactagaaggacagtatttggtatctgcgctctgctgaagccagttacttcggaaaaagagttggtagctcttg
atccggcaaacaaaccaccgctggtagcggtggttttttgtttgcaagcagcagattacgcgcagaaaaaaaggatctcaagaagatcctttgatcttttac
ggggtctgacgctcagtggaacgaaaactcacgttaagggattttggtcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatgaagtt
ttaaatcaatctaaagtataatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttcgttcatccatag
ttgcctgactccccgtcgtgtagataactacgatacgggagggttaccatctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatt
tatcagcaataaaccagccagccggaagggccgagcgcagaagtggtcctgcaactttatccgcctcatccagtctattaattgttgccgggaagctagagtaa
gtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatggcttcattcgctccggttcccaac
gatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctccttcggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactc
atggttatgcagcactgcataattctcttactgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatg
cggcgaccgagttgctcttgcccggcgtcaaacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaaacgttcttcggggcgaaaa
ctctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcaccaactgatcttcagcatcttttactttcaccagcgtttctgggtgagca
aaaacaggaaggcaaaatgccgcaaaaaagggaataagggcgacacggaaatgttgaatactcatactcttccttttcaatattattgaagcatttatcagggt
tattgtctcatgagcggatacatatttgaatgtatttagaaaaataaacaaatagggttccgcgcacatttccccgaaaagtgccacctgacgtctaagaacc
attattatcatgacattaacctataaaaataggcgtatcacgaggccctttcgtc

Figure S8. pXR4-nocap plasmid sequence. The *cap* gene fragment that is constant in all chimeras is highlighted in yellow, and the restriction sites used for subcloning all assembled chimeric genes are red and underlined.

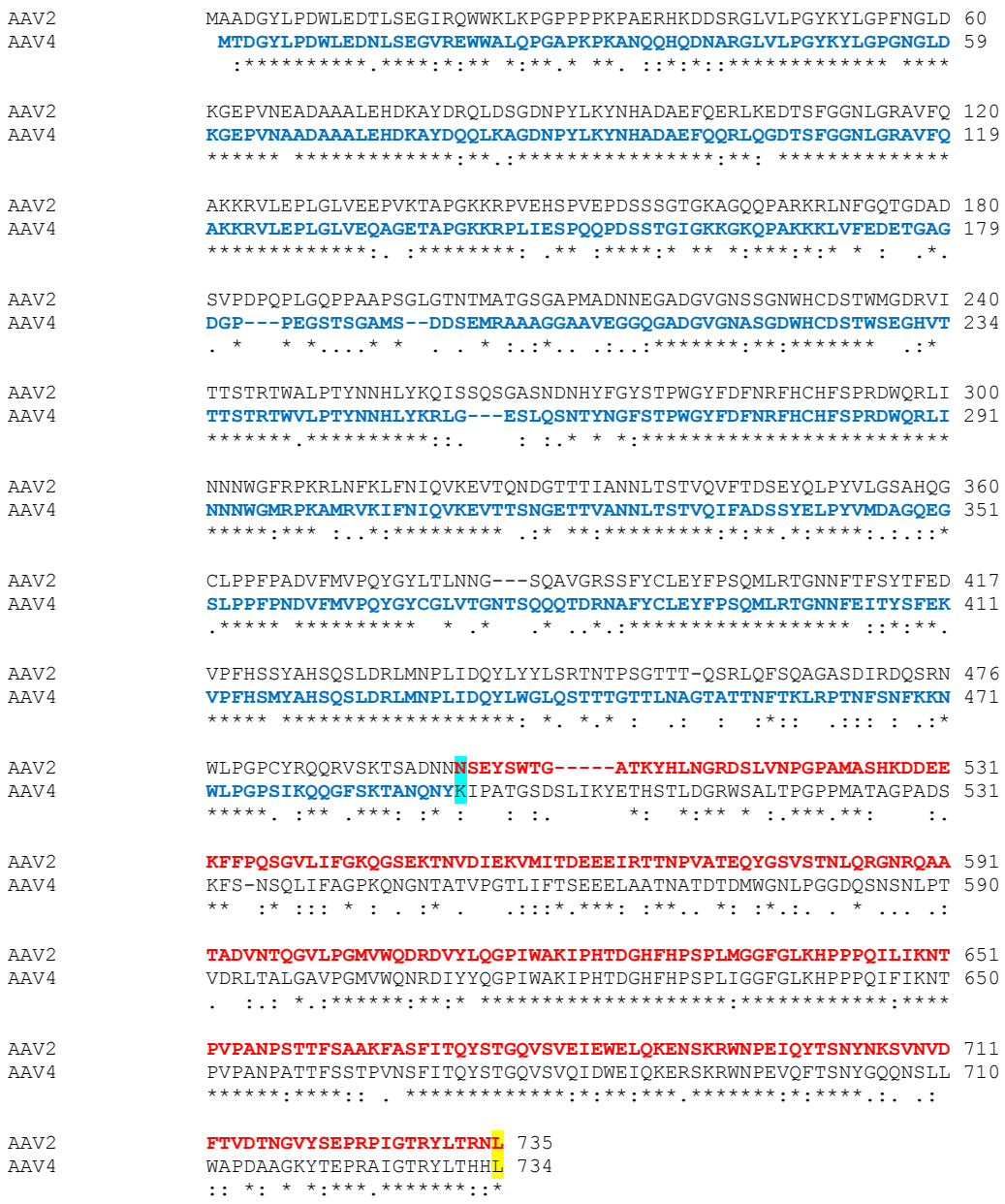

Figure S9. AAV2-AAV4 chimera nomenclature. Chimera names are based on the inheritance of AAV2 (red) and AAV4 (blue) sequence and are numbered based on the residues inherited from AAV2. As illustrated in this figure for chimera 492–734, the numbering represents the location in AAV4 that aligns with the AAV2 residues inherited in each chimera. The first AAV2 residue in this representative chimera, 492 by AAV4 VP1 numbering, is highlighted cyan, while the final AAV2 residue, 734 by AAV4 VP1 numbering, is highlighted yellow.

VP protein	$m_{subunit}$	E_{intra}	E_{inter}	$E_{subunit}$	E_{capsid}
AAV4	0	0	0	0	0
AAV2	223	0	0	0	0
607-612	3	3	0	3	180
655-670	6	0	2	1	60
345-357	7	0	3	1.5	90
294-370	25	20	6	23	1,380
265-370	26	24	6	27.5	1,620
289-372	26	21	9	25.5	1,530
254-435	55	9	25	21.5	1,290
505-734	106	29	69	75.5	3,810
492-734	119	32	96	80	4,800
439-597	120	17	16	25	1,500
416-606	125	12	7	15.5	930
438-633	127	9	5	11.5	690
475-734	128	14	86	57	3,420
408-654	132	12	5	14.5	870
379-646	140	28	16	36	2,160
385-699	148	30	10	35	2,100
408-734	161	15	31	31.5	1,830

Table S1. Subunit SCHEMA disruption values used to calculate E_{capsid}. The E_{intra} and E_{inter} values used to calculate $E_{subunit}$ and E_{capsid} are provided for each chimera analyzed in this study. For each chimera, the per subunit amino acid substitution level ($m_{subunit}$) was calculated relative to AAV4 (m = 0).

Sample	Titer before Amicon (genomes/mL)	Titer after Amicon (genomes/mL)
AAV2	6.12×10^{11}	2.78×10^{12}
AAV4	3.80×10^{11}	1.65×10^{12}
438-633	5.51×10^{11}	2.26×10^{11}

Table S2. Genomic titer of viruses prepared for TEM before and after concentration. AAV2, AAV4, and chimera 438-633 viruses were diluted in 1 x GB containing 0.001% Pluronic F-68 and spun in an Amicon Ultra-4 Centrifugal Filter Unit to simultaneously exchange buffers and concentrate viruses for TEM. Q-PCR was used to measure genomic titers of viruses before and after Amicon concentration to ensure no significant loss of sample occurred and that virus concentrations remained adequate for imaging.

Virus	Genomic Titer	VP Subunit Expression	% Nuclease-Protection	% of Heparin-Bound Eluates	% Transduction Efficiency
AAV4	◆◆◆◆	Y	109.0 ± 16.4	10.4 ± 0.9	52.1 ± 4.8
AAV2	◆◆◆◆	Y	113.8 ± 3.4	92.0 ± 1.1	79.4 ± 1.6
607-612	◆◆◆◆	Y	101.6 ± 5.0	12.3 ± 0.7	38.3 ± 2.8
655-670	◆◆◆◆	Y	98.2 ± 9.4	12.7 ± 0.9	32.8 ± 3.8
345-357	◆	Y	108.5 ± 5.0	n/a	n/a
294-370	n.d.	Y	n/a	n/a	n/a
265-370	n.d.	Y	n/a	n/a	n/a
289-372	n.d.	Y	n/a	n/a	n/a
254-435	n.d.	N	n/a	n/a	n/a
505-734	◆	Y	n/a	n/a	n/a
492-734	◆◆◆	Y	54.3 ± 2.3	60.0 ± 1.5	1.9 ± 0.2
439-597	n.d.	Y	n/a	n/a	n/a
416-606	n.d.	Y	n/a	n/a	n/a
438-633	◆◆◆◆	Y	1.1 ± 0.2	41 ± 2.5	1.6 ± 0.1
475-734	◆◆◆	Y	31.3 ± 4.4	59 ± 2.6	1.9 ± 0.2
408-654	◆	Y	1.7 ± 0.5	n/a	n/a
379-646	◆◆	Y	11.5 ± 9.9	n/a	n/a
385-699	n.d.	Y	n/a	n/a	n/a
408-734	n.d.	Y	n/a	n/a	n/a

Table S3. Summary of virus characteristics. For each virus analyzed, the finding from each method of characterization is listed, including: Q-PCR analysis of genome packaging (genomic titer), immunoblot detection of VP production (VP subunit expression), capsid protection of genomes from nucleases (% nuclease protection), capsid binding to heparin (% of Heparin-Bound Eluates) and cellular transduction (% Transduction Efficiency). In the case of the genomic titer, the black diamonds represent titers relative to the detection limit and the titers obtained with AAV2 and AAV4. Each diamond represents approximately 1-log in titer over background and n.d. means not detected above background, *i.e.*, within 1-log of the signal obtained with pBlueScript II plasmid lacking the AAV *cap* gene. n/a = not analyzed because titer was too low.