Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types

Sripriya Ravindra Kumar[1], **Timothy F. Miles**[1], **Xinhong Chen**[1], **David Brown**[1], **Tatyana Dobreva**[1,4], **Qin Huang**[1,4], **Xiaozhe Ding**[1], **Yicheng Luo**[1], **Pétur H. Einarsson**[1], **Alon Greenbaum**[1,2,3], **Min J. Jang**[1], **Benjamin E. Deverman**[1,4], **Viviana Gradinaru**[1]

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, California, USA.

[2]Present addresses: Joint Department of Biomedical Engineering, North Carolina State University, Raleigh, North Carolina, USA.

[3]Present addresses: University of North Carolina at Chapel Hill, Chapel Hill, North Carolina, USA.

[4]Present addresses: Stanley Center for Psychiatric Research, Broad Institute, Cambridge, Massachusetts, USA.

Abstract

Recombinant adeno-associated viruses (rAAVs) are efficient, non-invasive gene delivery vectors via intravenous delivery, however, natural serotypes display a finite set of tropisms. To expand their utility, we evolved AAV capsids to efficiently transduce specific cell types in adult mouse brains. Building upon our previous Cre recombination-based AAV targeted evolution (CREATE) platform, we developed Multiplexed-CREATE (M-CREATE) to quickly and accurately identify variants of interest in a given selection landscape through multiple positive and negative selection criteria by incorporating next-generation sequencing, synthetic library generation, and a novel analysis pipeline. *In vivo* selections for brain endothelial cell-, astrocyte-, and neuron-transducing capsids have identified variants that can transduce the central nervous system broadly, exhibit bias

---

toward vascular cells and astrocytes, target neurons with greater specificity, or cross the blood-brain barrier across diverse murine strains. Collectively, M-CREATE methodology accelerates the discovery of novel capsids for use in neuroscience and gene therapy applications.

INTRODUCTION:

Recombinant adeno-associated viruses (rAAVs) are widely used as gene delivery vectors in scientific research and therapeutic applications due to their ability to transduce both dividing and non-dividing cells, their long-term persistence as episomal DNA in infected cells, and their low immunogenicity[1–5]. However, gene delivery by natural AAV serotypes is limited by dose-limiting safety constraints and largely overlapping tropisms. AAV capsids engineered by rational design[6–9] or directed evolution[10–20] have yielded vectors with improved efficiencies for select cell populations[21–27], yet much work remains. Previously, we evolved AAV-PHP.B/eB variants from AAV9 using a selection method called CREATE: Cre recombination-based AAV targeted evolution[26]. This method relies on applying positive selective pressure for functional capsids by pairing Cre expression in defined cell populations with Cre-Lox recombination-dependent PCR amplification of capsid variants.

To more efficiently expand the AAV toolbox, we developed Multiplexed-CREATE (M-CREATE) (Fig. 1a, Supplementary Fig. 1a, b), named for its ability to accurately compare the enrichment profiles of thousands of capsid variants across multiple cell types and organs within a single experiment. This method improves upon its predecessor by capturing the breadth of capsid variants at every stage of the selection process. M-CREATE supports: (1) the calculation of a true enrichment score for each variant by using next generation sequencing (NGS) to correct for biases in viral production prior to selection, (2) reduced propagation of bias in successive rounds of selection through the creation of a post-round 1 synthetic pool library with equal variant representation, (3) the reduction of false positives by including codon replicates of each selected variant in the pool. Combined, these improvements allow confident interpretation across a broad range of enrichments in multiple positive selections and enable post-hoc negative screening by comparing deep sequencing of recovered capsid libraries among multiple targets (cells types or organs). Collectively, these features transform our ability to identify variants worthy of validation and characterization *in vivo*.

To demonstrate the ability of M-CREATE to reveal interesting variants missed by its predecessor (CREATE), we used the capsid library design that yielded AAV-PHP.B, identifying several AAV9 variants with distinct tropisms including ones that have biased transduction of brain vascular cells or that can cross the blood-brain barrier (BBB) without strain specificity.

RESULTS:

Multiplexed-CREATE allows detailed characterization of the capsid libraries during Round-1 selection.

During DNA and virus library generation there is potential for accumulation of biases that over-represent certain capsid variants, obscuring their true enrichment during *in vivo* selection. These biases may result from PCR amplification bias in the DNA library or sequence bias in the efficiency of virus production across various steps: capsid assembly, genome packaging and stability during purification. We investigated this with a *7-mer-i (i-insertion)* library, a randomized 7-mer library inserted between positions 588–589 of AAV9 (Fig. 1a,b) in rAAV- Cap9-in-cis-Lox2 plasmid (Supplementary Fig. 1a; theoretical library size: 3.4×10^{10} unique nucleotides, and an estimated $\sim 1 \times 10^{8}$ upon transfection; see Methods). Sequencing libraries after DNA assembly and virus purification to a depth of 10 – 20 million (M) reads was adequate to capture the bias among variants during virus production (Fig. 1c; despite ~1% variant overlap among these libraries; Supplementary Fig. 1c,d), demonstrating that even permissive sites like 588–589 will impose biological constraints on sampled sequence space. The DNA library had a uniform distribution of 9.6 M unique variants within ~10 M total reads (read count (RC) mean = 1.0, S.D. = 0.074), indicating minimal bias. In contrast, the virus library had 3.6 M unique variants within ~20 M depth (RC mean = 4.59, S.D. = 11.15) indicating enrichment of a subset of variants during viral production.

For *in vivo* selection, we intravenously administered the *7-mer-i* viral library at a dose of 2×10^{11} vg per adult transgenic mouse expressing Cre in different brain cell types: GFAP-Cre mice for astrocytes, SNAP25-Cre mice for neurons, and Tek-Cre mice for endothelial cells (*n* = 2 mice per Cre transgenic line, see Methods). Two weeks after intravenous (IV) injection, we harvested brain, spinal cord, and liver tissues. We extracted the rAAV genomes from tissues and selectively amplified the capsids that transduced Cre-expressing cells (Supplementary Fig. 1e–i). Upon deep sequencing, we observed $\sim 8 \times 10^{4}$ unique nucleotide variants recovered from brain tissues and < 50 variants in spinal cords (~48% of which were identified in virus library) across the transgenic lines, and each variant was represented with an enrichment score reflecting the change in relative abundance between the brain and the starting virus library (Fig. 1d, see Methods).

Two features of this dataset stand out. First, the recovered variants in brain tissue were disproportionately represented among the fraction of the transformed capsid library observed by sequencing after viral production demonstrating how production biases skew selection results. Second, the distribution of capsid read counts (RCs) reveals that more than half of the unique recovered variants after selection appear at remarkably low read counts. These variants may either be unintended mutants from experimental manipulation or AAV9-like variants with low basal level of CNS transduction (Supplementary Fig. 1e, see Methods).

A novel Round-2 library design improves the selection outcome

Concerned that the sequence bias during viral production and recovery would propagate across selection rounds despite our post-hoc enrichment scoring, we designed an unbiased

library based on the round-1 (R1) output (*synthetic pool* library) via oligo pools (Twist Bioscience). This library was compared to a library PCR amplified directly from the recovered R1 DNA (*PCR pool* library) (Fig. 2a, Supplementary Table 1, see Methods).

The *synthetic pool* library design comprised: (1) equimolar amounts of ~8950 capsid variants present at high read counts in at least one of the R1 selections from brain and spinal cord (Supplementary Fig. 1e, see Methods); (2) alternative codon replicates of those ~8950 variants (optimized for mammalian codons) to reduce false positives; and (3) a "*spike-in*" library of controls (Supplementary Note 1, Supplementary Dataset 1), resulting in a total library size of 18,000 nucleotide variants.

As anticipated, both round-2 (R2) virus libraries produced a high titer (~6×10^{11} vg per 10 ng of R2 DNA library per 150 mm dish; Supplementary Fig. 2a), and ~99% of variants from the R2 DNA were found after viral production (Fig. 2b). However, the distribution of the DNA and virus libraries from both designs differed significantly. The *PCR pool* library carries forward the R1 selection biases (Fig. 2c, Supplementary Fig. 2b,c) where the abundance reflects prior enrichment across tissues in R1 as well as bias from viral production and sample mixing. Comparatively, the *synthetic pool* DNA library is more evenly distributed, minimizing bias amplification across selection rounds.

For *in vivo* selection, we intravenously administered a dose of 1×10^{12} vg per adult transgenic mouse into three of the previously used lines ($n = 2$ mice per Cre transgenic line – GFAP, SNAP25, Tek), as well as the Syn-Cre line (for neurons). Two weeks after IV injection, rAAV genomes from brain samples were extracted, selectively amplified, and deep sequenced (as in R1). The *synthetic pool* library produced a greater number of positively enriched capsid variants than the *PCR pool* brain library (e.g. ~1700 versus ~700 variants/ tissue library at amino acid (AA) level in GFAP-Cre) (Fig. 2d, Supplementary Fig. 2d). In the *synthetic pool*, ~90% of the variants from the *spike-in* library were positively enriched as expected (Supplementary Fig. 2d, middle panel; Supplementary Dataset 1).

The degree of correlation for enrichment scores of variants recovered from both *PCR* and *synthetic pool* libraries varies in each Cre transgenic line, demonstrating the presence of noise within experiments (Supplementary Fig. 2e, Supplementary Note 2). The *synthetic pool*'s codon replicate feature addresses this predicament by pinpointing the level of enrichment needed within each selection to rise above noise (Fig. 2e, Supplementary Fig. 3a,b). This is a significant advantage over the PCR pool design, allowing researchers to confidently interpret enrichment scores in a given selection.

Analysis of AAV capsid libraries after Round-2 selections

Whereas the AA distribution of the DNA library closely matched the Oligopool design, virus production selected for a motif with Asn (N) at position 2, β-branched AAs (I, T, V) at position 4, and positively charged AAs (K, R) at position 5 (Fig. 2f, Supplementary Fig. 3c). Fitness for BBB crossing resulted in a very different pattern. In comparison to the R2 virus library, highly enriched variants share preferences, for example, proline (P) in position 5, and phenylalanine (F) in position 6.

Confident in our ability to assess enrichment score reproducibility within the *synthetic pool* design*, we* determined the distribution of the positively enriched variants from brain across all peripheral organs (Fig. 2g, left). About 60 variants that are highly enriched in brain are comparatively depleted across all other organs (Fig. 2g, middle). Encouraged by the expected behavior of *spike-in* control variants (AAV9, PHP.B, PHP.eB), eleven novel variants were chosen for further validation (Fig. 2g, right), including several that would have been overlooked if the choice had been based on *PCR pool* or CREATE (Supplementary Table 2).

These variants were chosen due to their enrichments and where they fall in sequence space. We noticed that the positively enriched variants cluster into distinct families based on sequence similarity (see Methods). In agreement with the heatmaps discussed above, the most enriched variants form a distinct family across selections that share a common motif: T in position 1, L in position 2, P in positive 5, F in position 6, and K or L in position 7 (Fig. 3a, Supplementary Fig. 3d). This AA pattern closely resembles the previously identified variant, AAV-PHP.B – TLAVPFK (Supplementary Note 3). Given the sequence similarity among members, we predicted that they may similarly cross the BBB and target the central nervous system.

Capsid recovery from Round-2 selection yields a pool of AAV9 variants with enhanced BBB entry and CNS transduction

Given the dominance of the PHP.B-family in this particular selection, we tested its most enriched member, TALKPFL (Fig. 3a,b) henceforth referred to as AAV-PHP.V1. Somewhat surprisingly given its sequence similarity to AAV.PHP.B, the tropism of AAV-PHP.V1 is biased toward transducing brain vascular cells (Fig. 3c, Supplementary Fig. 4a). When delivered intravenously, AAV-PHP.V1 carrying a fluorescent reporter under the control of the ubiquitous CAG promoter transduces ~60% of GLUT1$^+$ cortical brain vasculature compared to ~20% with AAV-PHP.eB and almost no transduction with AAV9 (Fig. 3c,d). In addition to the vasculature, AAV-PHP.V1 also transduced ~60% of cortical S100$^+$ astrocytes (Fig. 3e). However, AAV-PHP.V1 is not as efficient for astrocyte transduction as the previously reported AAV-PHP.eB (when packaged with an astrocyte specific GfABC1D promoter[28], Supplementary Fig. 4b).

For applications requiring endothelial cell-restricted transduction via intravenous delivery, AAV-PHP.V1 vectors can be used in three different systems: (1) in endothelial cell-type specific Tek-Cre[29] mice with a Cre-dependent expression vector (Fig. 3f (left), 3g, Supplementary Video 1), (2) in fluorescent reporter mice where Cre is delivered with an endothelial cell-type specific MiniPromoter (Ple261)[30] (Fig. 3f (right), 3h, Supplementary Fig. 4c–e), and (3) in wild-type mice by packaging a self-complementary genome (scAAV) containing a ubiquitous promoter (Supplementary Fig. 4f). The mechanism of endothelial cell-specific transduction by AAV-PHP.V1 using scAAV genomes is unclear, but shifts in vector tropism when packaging scAAV genomes have been reported for another capsid[31].

Given the dramatic difference in tropism between AAV-PHP.V1 and AAV-PHP.B/eB, we tested several additional variants within the PHP.B-like family. One variant, AAV-PHP.V2 – TTLKPFL, differed by only one AA from AAV-PHP.V1, has a similar tropism

(Supplementary Fig. 5, Supplementary Note 4). Three other variants with sequences of roughly equal deviation from both AAV.PHP.V1 and AAV.PHP.B, AAV-PHP.B4 – TLQIPFK, AAV-PHP.B7 – SIERPFK, and AAV-PHP.B8 – TMQKPFI (Fig. 3a,b, 4a,b), have PHP.B-like tropism with biased transduction toward neurons and astrocytes (Fig. 4b, Supplementary Fig. 6a–c). Similar variants among the spike-in library, AAV-PHP.B5 – TLQLPFK and AAV-PHP.B6 – TLQQPFK, also shared this tropism (Fig. 3b, 4a,b; Supplementary Fig. 6a; Supplementary Note 5).

We next investigated a series of variants selected to verify M-CREATE's predictive power outside this family: (1) A highly enriched variant with a completely unrelated sequence, AAV-PHP.C1 – RYQGDSV (Fig. 3a,b, 4a,b), transduced astrocytes at a similar efficiency and neurons at lower efficiency compared to other tested variants from B-family (Fig. 4b). (2) Two variants found in high abundance in the R2 *synthetic pool* virus library and negatively enriched in brain (with both codon replicates in agreement), AAV-PHP.X1 – ARQMDLS and AAV-PHP.X2 – TNKVGNI (Supplementary Fig. 2b, right), poorly transduced the CNS (Supplementary Fig. 6b). (3) Two variants that were found in higher abundance in brain libraries from the *PCR pool* R2, AAV-PHP.X3 – QNVTKGV and AAV-PHP.X4 - LNAIKNI also failed to outperform AAV9 in the brain (Supplementary Fig. 6d).

Collectively, our characterization of these AAV variants demonstrates several key points. First, within a diverse sequence family, there is room for both functional redundancy and the emergence of novel tropisms. Second, highly enriched sequences outside the dominant family are also likely to possess enhanced function. Third, buoyed by codon replicate agreement in the synthetic pool, a variant's enrichment across tissues may be predictive. Fourth, while the *synthetic pool* R2 library contains a subset of the sequences that are in the *PCR pool* R2 and may thereby lack some enhanced variants, the excluded *PCR pool* population is enriched in false positives.

The ability to confidently predict *in vivo* transduction from a pool of 18,000 variants across mice is a significant advance in the selection process and demonstrates the power of M-CREATE for the evolution of individual vectors.

Re-investigation of capsid selection that yielded AAV.PHP.eB reveals variant that specifically transduces neurons

Using NGS, we re-investigated a *3-mer-s (s-substitution)* PHP.B library generated by the prior CREATE methodology that yielded AAV-PHP.eB[27] (Fig. 5a, Supplementary Note 6). We deep sequenced the brain libraries using Cre-dependent PCR and a R2 liver library from wild-type mice (processed via PCR for all capsid sequences regardless of Cre-mediated inversion) and identified 150 – 200 positively enriched capsids in brain tissue (Fig. 5b, Supplementary Fig.7a,b).

Variants that were positively enriched in brain and negatively enriched in liver show a significant bias towards certain AAs: G, D, E at position 1; G, S at position 2 (which includes the AAV-PHP.eB motif, DG); and S, N, P at position 9, 10, 11 (Fig. 5c, Supplementary Fig. 7c, see Methods). Variants that were positively enriched in the brain were clustered according to their sequence similarities and ranked by their negative

enrichment in liver (represented by node size in clusters; see Methods). A distinct family referred to as N emerged with a common motif "SNP" at positions 9–11 on PHP.B backbone (Fig. 5d, Supplementary Fig. 7d).

The core variant of the N-family cluster: AQTLAVPFSNP was found in high abundance in R1 and R2 selections, had higher enrichment score in Vglut2 and Vgat brain tissues compared to GFAP, and had negative enrichment in liver tissue (Fig. 5b, Supplementary Fig. 7a–d). Unlike AAV-PHP.eB, this variant (AAV-PHP.N) specifically transduced NeuN[+] neurons even when packaged with a ubiquitous CAG promoter, although the transduction efficiency varied across brain regions (from ~10–70% in NeuN[+] neurons, including both VGLUT1[+] excitatory and GAD1[+] inhibitory neurons, Fig. 5e,f; Supplementary Fig. 7e,f).

Thus, by re-examining the *3-mer-s* library we were able to identify several novel variants, including one with notable cell-type-specific tropism (Supplementary Note 7).

Investigation of capsid families beyond C57BL/6J mouse strain

The enhanced CNS tropism of AAV-PHP.eB is absent in a subset of mouse strains. It is highly efficient in C57BL/6J, FVB/NCrl, DBA/2, and SJL/J, with intermediate enhancement in 129S1/SvimJ, and no enhancement in BALB/cJ and several additional strains[32–36]. This pattern holds for the two newly identified variants from the PHP.B family, AAV-PHP.V1 and AAV-PHP.N (Fig. 6a, Supplementary Table 3), which did not transduce the CNS in BALB/cJ, yet transduced the FVB/NJ strain (Fig. 6b). AAV-PHP.V1 transduced Human Brain Microvascular Endothelial Cell (HBMEC) culture, resulting in increased mean fluorescent intensity compared to AAV9 and AAV-PHP.eB (Supplementary Fig. 8a) however, suggesting the potential for mechanistic complexity.

Importantly, M-CREATE revealed many non-PHP.B-like sequence families that enriched through selection for transduction of cells in the CNS. We tested the previously mentioned AAV-PHP.C1: RYQGDSV, as well as AAV-PHP.C2: WSTNAGY, and AAV-PHP.C3: ERVGFAQ (Fig. 6a). These showed enhanced BBB crossing irrespective of mouse strain, with roughly equal CNS transduction in BALB/cJ and C57BL/6J (Fig. 6c, Supplementary Fig. 8b). Collectively, these preliminary studies suggest that M-CREATE is capable of finding capsid variants with diverse mechanisms of BBB entry that lack strain-specificity.

DISCUSSION:

This work outlines the development and validation of an improved platform, M-CREATE, for multiplexed viral capsid selection. M-CREATE incorporates multiple internal controls to monitor sequence progression, minimize bias, and accelerate the discovery of capsid variants with novel tropisms. Utilizing M-CREATE, we have identified both individual capsids and distinct families of capsids that are biased toward different cell-types of the adult brain. The outcome from *7-mer-i* selection demonstrates the possibility of finding AAV capsids with improved efficiency and specificity towards one or more cell types. Patterns of CNS infectivity across mouse strains suggest that M-CREATE may also identify multiple capsids with distinct mechanisms of BBB crossing. With additional rounds of evolution as shown in

the *3-mer-s* selection, the specificity or efficiency of *7-mer-i* library variants may be improved, as was observed with AAV-PHP.N. or AAV-PHP.eB (from prior study).

We believe that the variants tested *in vivo* and their families will find broad application in neuroscience, including studies involving the BBB[37], neural circuits[38], neuropathologies[39], and therapeutics[40]. AAV-PHP.V1 or AAV-PHP.N are well-suited for studies requiring gene delivery for optogenetic or chemogenetic manipulations[41], or rare monogenic disorders (targeting brain endothelial cells: e.g., GLUT1-deficiency syndrome, NLS1-microcephaly[39]; or targeting neurons: e.g., mucopolysaccharidosis type IIIC (MPSIIIC)[22]).

The outcome from M-CREATE will open several promising lines of inquiry: (1) assessment of identified capsid families across species, (2) investigation of the mechanistic properties that underlie the ability to cross specific barriers (BBB) or target specific cell populations, (3) further evolution of the identified variants for improved efficiency and specificity, and (4) using the datasets generated by M-CREATE as training sets for *in silico* selection by machine learning models. M-CREATE is presently limited by the low throughput of vector characterization *in vivo*, however RNA sequencing technologies[42] offer hope in this regard. In summary, M-CREATE will serve as a next-generation capsid selection platform that can open new directions in vector engineering and potentially broaden the AAV toolbox for various applications in science and in therapeutics.

ONLINE METHODS

Plasmids

A. Library generation—The rAAV- Cap-in-cis-Lox2 plasmid (Supplementary Fig. 1a, plasmid available upon request at Caltech CLOVER Center) is a modification of the rAAV- Cap-in-cis-Lox plasmid[26]. For *7-mer-i* library fragment generation, we used the pCRII-9Cap-XE plasmid[26] as a template. The AAV2/9 REP-AAP- Cap plasmid (Supplementary Fig. 1a, plasmid available upon request at Caltech CLOVER Center) was modified from the AAV2/9 REP-AAP plasmid[26] (See Supplementary Note 8).

B. Capsid characterization

(i) AAV capsids: The AAV capsid variants with 7-mer insertions or 11-mer substitutions were made between positions 587–597 of AAV-PHP.B capsid using the pUCmini-iCAP-PHP.B backbone[26] (Addgene ID: 103002).

(ii) ssAAV genomes: To characterize the AAV capsid variants, we used the single stranded (ss) rAAV genomes. We used genomes such as pAAV:CAG-mNeonGreen[27] (equivalent plasmid, pAAV: CAG-eYFP[35]; Addgene ID: 104055), pAAV:CAG-NLS-EGFP[26] (equivalent version with one NLS is on Addgene ID 104061), pAAV:CAG-DIO-EYFP[35] (Addgene ID: 104052), pAAV: GfABC1D-2xNLS-mTurquoise2[35] (Addgene ID: 104053), and pAAV-Ple261-iCre[30] (Addgene ID 49113) (See Supplementary Note 9).

(iii) scAAV genomes: To characterize the AAV capsid variant, AAV-PHP.V1, using self-complementary (sc) rAAV genomes, we used scAAV genomes from different sources.

scAAV:CB6-EGFP was a gift from Dr. Guangping Gao and scAAV:CAG-EGFP[43] from Addgene (Addgene ID:83279) (See Supplementary Note 9).

AAV capsid library generation

A. Round-1 AAV capsid DNA library

(i) Mutagenesis strategy: The 7-mer randomized insertion was designed using the NNK saturation mutagenesis strategy, involving degenerate primers containing mixed bases (Integrated DNA Technologies, Inc.). N can be A, C, G, or T bases and K can be G, or T. Using this strategy, we obtained combinations of all 20 AAs at each position of the 7-mer peptide using 33 codons, resulting in a theoretical library size of 1.28 billion at the level of AA combinations. The mutagenesis strategy for the *3-mer-s* PHP.B library is described in our prior work[27].

(ii) Library cloning: The 480 bp AAV capsid fragment (450–592 AAs) with the 7-mer randomized insertion between AAs 588 and 589 was generated by conventional PCR methods using the pCRII-9Cap-XE template by Q5 Hot Start High-Fidelity 2X Master Mix (NEB; M0494S) with forward primer, XF: 5'-ACTCATCGACCAATACTTGTACTATCTCTCTAGAAC-3' and reverse primer, 7xMNN-588i: 5'-GTATTCCTTGGTTTTGAACCCAACCGGTCTGCGCCTGTGCMNNMNNMNNMNNMNNMNNMNNTTGGGCACTCTGGTGGTTTGTG-3' (See Supplementary Note 10).

The rAAV- Cap-in-cis-Lox2 plasmid (6960 bp) was linearized with the restriction enzymes AgeI and XbaI, and the amplified library fragment was assembled into the linearized vector at 1:2 molar ratio using the NEBuilder HiFi DNA Assembly Master Mix (NEB; E2621S) by following the NEB recommended protocol.

(iii) Library purification: The assembled library was then subjected to Plasmid Safe (PS) DNase I (Epicentre; E3105K) treatment, or alternatively, Exonuclease V (RecBCD) (NEB; M0345S) following the recommended protocols, to purify the assembled product by degrading the un-assembled DNA fragments from the mixture. The resulting mixture was purified with a PCR purification kit (DNA Clean and Concentrator kit, Zymo Research; D4013).

(iv) Library yield: With an assembly efficiency of 15% – 20% post-PS treatment, we obtained a yield of about 15 – 20 ng per 100 ng of input DNA per 20 μL reaction.

(v) Quality control: See Supplementary Note 11.

B. Round-2 AAV capsid DNA library

(i) *PCR pool* design: To maintain proportionate pooling, we mathematically determined the fraction of each sample/library that needs to be pooled based on an individual library's diversity (see Supplementary Note 12).

The pooled sample was used as a template for further amplification with 12 cycles of 98°C for 10 s, 60°C for 20 s, and 72°C for 30 s by Q5 polymerase, using the primers 588-R2lib-F: 5'-CACTCATCGACCAATACTTGTACTATCTCTCT-3' and 588-R2lib-R: 5'-GTATTCCTTGGTTTTGAACCCAACCG-3'. Similar to R1 library generation, the PCR product was assembled into the rAAV- Cap-in-cis-Lox2 plasmid and the virus was produced (see Supplementary Note 13).

(ii) *Synthetic pool* **design:** As described in the PCR pool strategy, we chose high-confidence variants whose RCs were above the error-dominant noise slope from the plot of library distribution (see Supplementary Fig. 1e and Supplementary Note 12). This came to about 9000 sequences from all brain and spinal cord samples of all Cre lines. We used similar primer design as mentioned in the description of the R1 library generation. Primers XF: 5'-ACTCATCGACCAATACTTGTACTATCTCTCTAGAAC-3' and 11-mer-588i: 5'-GTATTCCTTGGTTTTGAACCCAACCGGTCTGCGCxrefMNNMNNMNNMNNMNNMNNMNNMNNxrefACTCTGGTGGTTTGTG-3', where "xrefMNNMNNMNNMNNMNNMNNMNNxref" was replaced with unique nucleotide sequence of a 7-mer tissue recovered variant (7xMNN) along with modification of two adjacent codons flanking on either end of the 7-mer insertion site (6xX), which are residues 587–588 "AQ" and residues 589–590 "AQ" on AAV9 capsid. Since *spike-in* library has 11-mer mutated variants, we used the same primer design where "xrefMNNMNNMNNMNNMNNMNNMNNxref" was replaced with a specific nucleotide sequence of a 11-mer variant. A duplicate of each sequence in this library was designed with different codons optimized for mammals. The primers were designed using a custom-built Python based script. The custom-designed oligopool was synthesized in an equimolar ratio by Twist Biosciences. The oligopool was used to minimally amplify the pCRII-XE Cap9 template over 13 cycles of 98°C for 10 s, 60°C for 20 s, and 72°C for 30 s. To obtain a higher yield for large-scale library preparation, the product of the first PCR was used as a template for the second PCR using the primers XF and 588-R2lib-R (described above) and minimally amplified for 13 cycles. Following PCR, we assembled the R2 *synthetic pool* DNA library and produced the virus as described in R1 (see Supplementary Note 13).

C.   AAV virus library production, purification and genome extraction—To prevent capsid mosaic formation of the *7-mer-i* library in 293T producer cells, we transfected only 10 ng of assembled library per 150 mm dish along with other required reagents for AAV vector production (see Supplementary Note 14). For the rAAV DNA extraction from purified rAAV viral library, ~10% of the purified viral library was used to extract the viral genome by proteinase K treatment (see Supplementary Note 15).

Animals

All animal procedures performed in this study were approved by the California Institute of Technology Institutional Animal Care and Use Committee (IACUC), and we have complied with all relevant ethical regulations. C57BL/6J (000664), Tek-Cre[29] (8863), SNAP25-Cre[44] (23525), GFAP-Cre[45] (012886), Syn1-Cre[46] (3966), and Ai14[47] (007908) mice lines used in this study were purchased from the Jackson Laboratory (JAX). The IV injection of rAAVs was into the retro-orbital sinus of adult mice. For testing the transduction phenotypes of

novel rAAVs, 6- to 8-week-old C57BL/6J or Tek-Cre or Ai14 adult male mice were randomly assigned. The experimenter was not blinded for any of the experiments performed in this study.

In vivo selection

The *7-mer-i* viral library selections were carried out in different lines of Cre transgenic adult mice: Tek-Cre, SNAP25-Cre, and GFAP-Cre for the R1 selections, and those three plus Syn1-Cre for the R2 selections. Male and female adult mice were intravenously administered with a viral vector dose of 2×10^{11} vg/mouse for the R1 selection, and a dose of 1×10^{12} vg/mouse for the R2 selection. The dose was determined based on the virus yield which was different across selection rounds (Supplementary Fig. 2a). Both genders were used to recover capsid variants with minimal gender bias. Two weeks post-injection, mice were euthanized and all organs including brain were collected, snap frozen on dry ice, and stored at −80°C.

A. rAAV genome extraction from tissue

(i) Optimization—See Supplementary Note 16.

(ii) rAAV genome extraction with the Trizol method—Half of a frozen brain hemisphere (0.3 g approx.) was homogenized with a 2 ml glass homogenizer (Sigma Aldrich; D8938) or a motorized plastic pestle (Fisher Scientific;12-141-361, 12-141-363) (for smaller tissues) or beads using BeadBug homogenizers (1.5–3.0 mm zirconium or steel beads per manufacturer recommendations) (Homogenizers, Benchmark Scientific, D1032–15, D1032–30, D1033–28) and processed using Trizol as described in our prior work[26] (also see Supplementary Note 17). From deep sequencing data analysis, we observed that the amount of tissue processed was sufficient for rAAV genome recovery.

(iii) rAAV genome recovery by Cre-dependent PCR—rAAV genomes with Lox sites flipped by Cre recombination were selectively recovered and amplified using PCR with primers that yield a PCR product only if the Lox sites are flipped (see Supplementary Fig. 1b). We used the primers 71F: 5'-CTTCCAGTTCAGCTACGAGTTTGAGAAC-3' and CDF/R: 5'- CAAGTAAAACCTCTACAAATGTGGTAAAATCG-3' and amplified the Cre-recombined genomes over 25 cycles of 98°C for 10 s, 58°C for 30 s, and 72°C for 1 min, using Q5 DNA polymerase.

(iv) Total rAAV genome recovery by PCR (Cre-independent)—To recover all rAAV genomes from a tissue, we used the primers XF (5'-ACTCATCGACCAATACTTGTACTATCTCTCTAGAAC-3') and 588-R2lib-R (5'-GTATTCCTTGGTTTTGAACCCAACCG-3') to amplify the genomes over 25 cycles of 98°C for 10 s, 60°C for 30 s, and 72°C for 30 min, using Q5 DNA polymerase.

Sample preparation for NGS

We processed the DNA library, the virus library, and the tissue libraries post-*in vivo* selection to add flow cell adaptors around the diversified 7-mer insertion region (see Supplementary Fig. 1b).

A. Preparation of rAAV DNA and Viral DNA library—The Gibson-assembled rAAV DNA library and the DNA extracted from the viral library were amplified by Q5 DNA polymerase using the primers 588i-lib-PCR1–6bpUID-F: 5'-CACGACGCTCTTCCGATCTAANNNNNNNAGTCCTATGGACAAGTGGCCACA-3' and 588i-lib-PCR1-R: 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTGGTTTTGAACCCAACCG-3' that are positioned around 50 bases from the randomized 7-mer insertion on the capsid, and that contain the Read1 and Read2 flow cell sequences on the 5' end (See Supplementary Note 18). Using 5–10 ng of template DNA in a 50 μl reaction, the DNA was minimally amplified for 4 cycles of 98°C for 10 s, 60°C for 30 s, and 72°C for 10 s. The mixture was then purified with a PCR purification kit. The eluted DNA was then used as a template in a second PCR to add the unique indices (single or dual) via the recommended primers (NEB; E7335S, E7500S, E7600S) in a 12-cycle reaction using the same temperature cycle as described above. The samples were then sent for deep sequencing following additional processing and validation (see Supplementary Note 19).

B. Preparation of rAAV tissue DNA library—The PCR-amplified rAAV DNA library from tissue (see section A: iii and iv) was further amplified with a 1:100 dilution of this DNA as a template to the primers 1527: 5'-ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACAAGTGGCCACAAACCACCAG-3' and 1532: 5'-GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTTGGTTTTGAACCCAACCG-3' that are positioned around 50 bases from the randomized 7-mer insertion on the capsid, and that contain the Read1 and Read2 sequences on the 5' end. The DNA was amplified by Q5 DNA polymerase for 10 cycles of 98°C for 10 s, 59°C for 30 s, and 72°C for 10 s. The mixture was purified with a PCR purification kit. The eluted DNA was then used as a template in a second PCR to add the unique indices (single or dual) using the recommended primers (NEB; E7335S, E7500S, E7600S) in a 10-cycle reaction with the same temperature cycle as described above (for DNA and virus library preparation), and followed additional processing and validation before sequencing (see Supplementary Note 19).

In vivo characterization of AAV vectors

A. Cloning AAV capsid variants—The AAV capsid variants were cloned into a pUCmini-iCAP-PHP.B backbone (Addgene ID: 103002) using overlapping forward and reverse primers with 11-mer substitution (in case of *7-mer-i* variants, the flanking AA from AAV9 capsid AA587–588 "AQ" and AA589–590 "AQ" were subjected to codon modification) that spans from the MscI site (at position 581 AA) to the AgeI site (at position 600 AA) on the pUCmini plasmid. The primers were designed for all capsid variants using a custom Python script and cloned using standard molecular techniques (see Supplementary Note 20). List of primers used to clone AAV-PHP variants is provided (Supplementary Table 4).

B. AAV vector production—Using an optimized protocol[35], we produced AAV vectors from 5–10 150 mm plates, which yielded sufficient amounts for administration to adult mice.

C. AAV vector administration, dosage and expression time.—AAV vectors were administered intravenously to adult male mice (6 – 8 weeks of age) via retro-orbital injection at doses of $1 – 10 \times 10^{11}$ vg with 3–4 weeks of *in vivo* expression times unless mentioned otherwise in the figures/legends (also see Supplementary Note 21).

D. Tissue processing—After 3 weeks of expression (unless noted otherwise), the mice were anesthetized with Euthasol (pentobarbital sodium and phenytoin sodium solution, Virbac AH) and transcardially perfused with 30 – 50 mL of 0.1 M phosphate buffered saline (PBS) (pH 7.4), followed by 30 – 50 ml of 4% paraformaldehyde (PFA) in 0.1 M PBS. After this procedure, all organs were harvested and post-fixed in 4% PFA at 4°C overnight. The tissues were then washed and stored at 4°C in 0.1 M PBS and 0.05% sodium azide. All solutions used for this procedure were freshly prepared. For the brain and liver, 100-μm thick sections were cut on a Leica VT1200 vibratome.

For vascular labeling, the mice were anesthetized and transcardially perfused with 20 mL of ice-cold PBS, followed by 10 mL of ice-cold PBS containing Texas Red-labeled Lycopersicon Esculentum (Tomato) Lectin (1:100, Vector laboratories, TL-1176) or DyLight 594 labeled Tomato Lectin (1:100, Vector laboratories, DL-1177), and then placed in 30 mL of ice-cold 4% PFA for fixation.

E. Immunohistochemistry—Immunohistochemistry was performed on 100-μm thick tissue sections to label different cell-type markers such as NeuN (1:400, Abcam, ab177487) for neurons, S100 (1:400, Abcam, ab868) for astrocytes, Olig2 (1:400; Abcam, ab109186) for oligodendrocyte lineage cells, and GLUT-1 (1:400; Millipore Sigma, 07–1401) for brain endothelial cells using optimized protocols (See Supplementary Note 22).

F. Hybridization chain reaction (HCR) based RNA labeling in tissues— Fluorescence *in situ* hybridization chain reaction (FITC-HCR) was used to label excitatory neurons with VGLUT1 and inhibitory neurons with GAD1 to characterize the AAV capsid variant AAV-PHP.N in brain tissue using an adapted third-generation HCR[48] protocol (See Supplementary Note 23).

G. Imaging and image processing—All images in this study were acquired either with a Zeiss LSM 880 confocal microscope using the objectives Fluar 5× 0.25 M27, Plan-Apochromat 10× 0.45 M27 (working distance 2.0 mm), and Plan-Apochromat 25× 0.8 Imm Corr DIC M27 multi-immersion; or with a Keyence BZ-X700 microscope (see Supplementary Note 24). The acquired images were processed in the respective microscope softwares Zen Black 2.3 SP1 (Zeiss), BZ-X Analyzer (Keyence), Keyence Hybrid Cell Count software (BZ-H3C), ImageJ, Imaris (Bitplane) and with Photoshop CC 2018 (Adobe). The images were compiled in Illustrator CC 2018 (Adobe).

H. Tissue clearing—Brain hemispheres were cleared using iDISCO[49] method and tissues over 500 μm thickness were optically cleared using ScaleS4(0)[50] (See Supplementary Note 25).

I. Tissue processing and imaging for quantification of rAAV transduction *in vivo*—For quantification of rAAV transduction, 6- to 8-week-old male mice were intravenously injected with the virus, which was allowed to express for 3 weeks (unless specified otherwise). The mice were randomly assigned to groups and the experimenter was not blinded. The mice were perfused and the organs were fixed in PFA. The brains and livers were cut into 100-μm thick sections and immunostained with different cell-type-specific antibodies, as described above. The images were acquired either with a 25× objective on a Zeiss LSM 880 confocal microscope or with a Keyence BZ-X700 microscope; images that are compared directly across groups were acquired and processed with the same microscope and settings (See Supplementary Note 26).

In vitro characterization of AAV vectors

Human Brain Microvascular Endothelial Cells (HBMEC) (ScienCell Research Laboratories, Cat. 1000) were cultured as per the instructions provided by the vendor (also see Supplementary Note 27 for AAV transduction protocol).

Data analysis

A. Quantification of rAAV vector transduction—Manual counting was performed with the Adobe Photoshop CC 2018 Count Tool for cell types in which expression and/or antibody staining covered the whole cell morphology. The Keyence Hybrid Cell Count software (BZ-H3C) was used where the software could reliably detect distinct cells in an entire dataset. To maintain consistency in counting across different markers and groups, one person was assigned to quantify across all groups in all brain areas (see Supplementary Note 28). The experimenter was not blinded during any of the analysis.

B. NGS data alignment and processing—The raw fastq files from NGS runs were processed with custom built scripts that align the data to AAV9 template DNA fragment containing the diversified region 7xNNK (for R1) or 11xNNN (for R2 since it was synthesized as 11xNNN) (see Supplementary Note 29).

C. NGS data analysis—The aligned data were then further processed via a custom data-processing pipeline, with scripts written in Python.

The enrichment scores of variants (Total = N) across different libraries were calculated from the read counts (RCs) according to the following formula:

$$\text{Enrichment score} = \log 10 \left[\frac{(\text{Variant 1 RC in tissue library1}/\text{Sum of variants N RC in library1})}{(\text{Variant 1 RC in virus library}/\text{Sum of variants N RC in virus library})} \right]$$

To consistently represent library recovery between R1 and R2 selected variants, we estimated the enrichment score of the variants in R1 selection (see Supplementary Note 30).

The standard score of variants in a specific library was calculated using this formula:

$$\text{Standard score} = (\text{read count\_i} - \text{mean})/\text{standard deviation}$$

Where read count_i is raw copy number of a variant i,

Mean is the mean of read counts of all variants across a specific library,

Standard deviation is the standard deviation of read counts of all variants across a specific library.

The plots generated in this article were using the following software - Plotly, GraphPad PRISM 7.05, Matplotlib, Seaborn, and Microsoft Excel 2016. The AAV9 capsid structure (PDB 3UX1)[51] was modeled in PyMOL.

D. Heatmap generation—The relative AA distributions of the diversified regions are plotted as heatmaps. The plots were generated using the Python Plotly plotting library. The heatmap values were generated from custom scripts written in Python, using functions in the custom "pepars" Python package (see Supplementary Note 31).

E. Clustering analysis—Using custom scripts written in MATLAB (version R2017b; MathWorks) the reverse Hamming distances representing the number of shared AAs between two peptides was determined. Cytoscape (version 3.7.1[52]) software was then used to cluster the variants. The AA frequency plot representing the highlighted cluster was created using Weblogo (Version 2.8.2)[53,54] (see Supplementary Note 32).

Statistics and reproducibility:

Statistical tests were performed using GraphPad PRISM or Python scripts. All correlation analyses reported were carried out using a linear least-squares regression method by an inbuilt Python function from SciPy library "*scipy.stats.linregress*", and the coefficient of determination (R^2) is reported. Tests evaluating the significance of amino acid bias were done using statsmodels Python library. A one-proportion z-test for a library vs known template frequency (NNK), and two-proportion z-test for two library comparisons were performed. P-values are corrected for multiple comparisons using Bonferroni correction. For datasets with two experimental group comparisons, a Mann-Whitney test was used and two-tailed exact P-values are reported. For more than two experimental group comparisons with one variable, a one-way ANOVA non-parametric Kruskal-Wallis test was performed and correction for multiple comparisons using uncorrected Dunn's test was performed. Exact P-values are reported from both tests (unless indicated otherwise). For experimental group comparisons with two variables, a two-way ANOVA with Tukey's test for multiple comparisons reporting corrected P-values were performed with 95% confidence interval (CI).

All quantitative data reported in graphs are from biological replicates (mouse or tissue culture replicates), where each data point from a biological replicate is the mean from technical replicates (raw data such as images of a specific brain region). Statistical analyses were performed on datasets with at least three biological replicates. Error bars in the figures denote standard errors of mean (S.E.M.). All experiments were validated in more than one independent trial unless otherwise noted.

Reporting Summary:

Includes additional information on the methods and reproducibility.

ACCESSION CODES:

GenBank: AAV-PHP.V1:, AAV-PHP.N:, AAV-PHP.V2:, AAV-PHP.B4:, AAV-PHP.B5, AAV-PHP.B6:, AAV-PHP.B7:, AAV-PHP.B8:, AAV-PHP.C1:, AAV-PHP.C2, and AAV-PHP.C3.

DATA AVAILABILITY STATEMENT:

Data beyond what has been provided in the article and supplementary documents are available from the corresponding author upon request. The following vector plasmids are deposited on Addgene for distribution (http://www.addgene.org) AAV-PHP.V1: 127847, AAV-PHP.V2: 127848, AAV-PHP.B4: 127849, and AAV-PHP.N: 127851. Requests for other reagents can be made at Caltech – CLOVER Center (http://clover.caltech.edu/).

CODE AVAILABILITY STATEMENT:

The codes used for M-CREATE data analysis were written in python or MATLAB and are made available on GitHub: https://github.com/GradinaruLab/mCREATE. The custom MATLAB scripts to generate HCR probes is accessible through GitHub on a different repository: https://github.com/GradinaruLab/HCRprobe.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS:

REFERENCES (for main text only):

1. Wu Z, Asokan A & Samulski RJ Adeno-associated virus serotypes: vector toolkit for human gene therapy. Mol. Ther. J. Am. Soc. Gene Ther 14, 316–327 (2006).

2. Naso MF, Tomkowicz B, Perry WL & Strohl WR Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. Biodrugs 31, 317–334 (2017). [PubMed: 28669112]

3. Daya S & Berns KI Gene Therapy Using Adeno-Associated Virus Vectors. Clin. Microbiol. Rev 21, 583–593 (2008). [PubMed: 18854481]
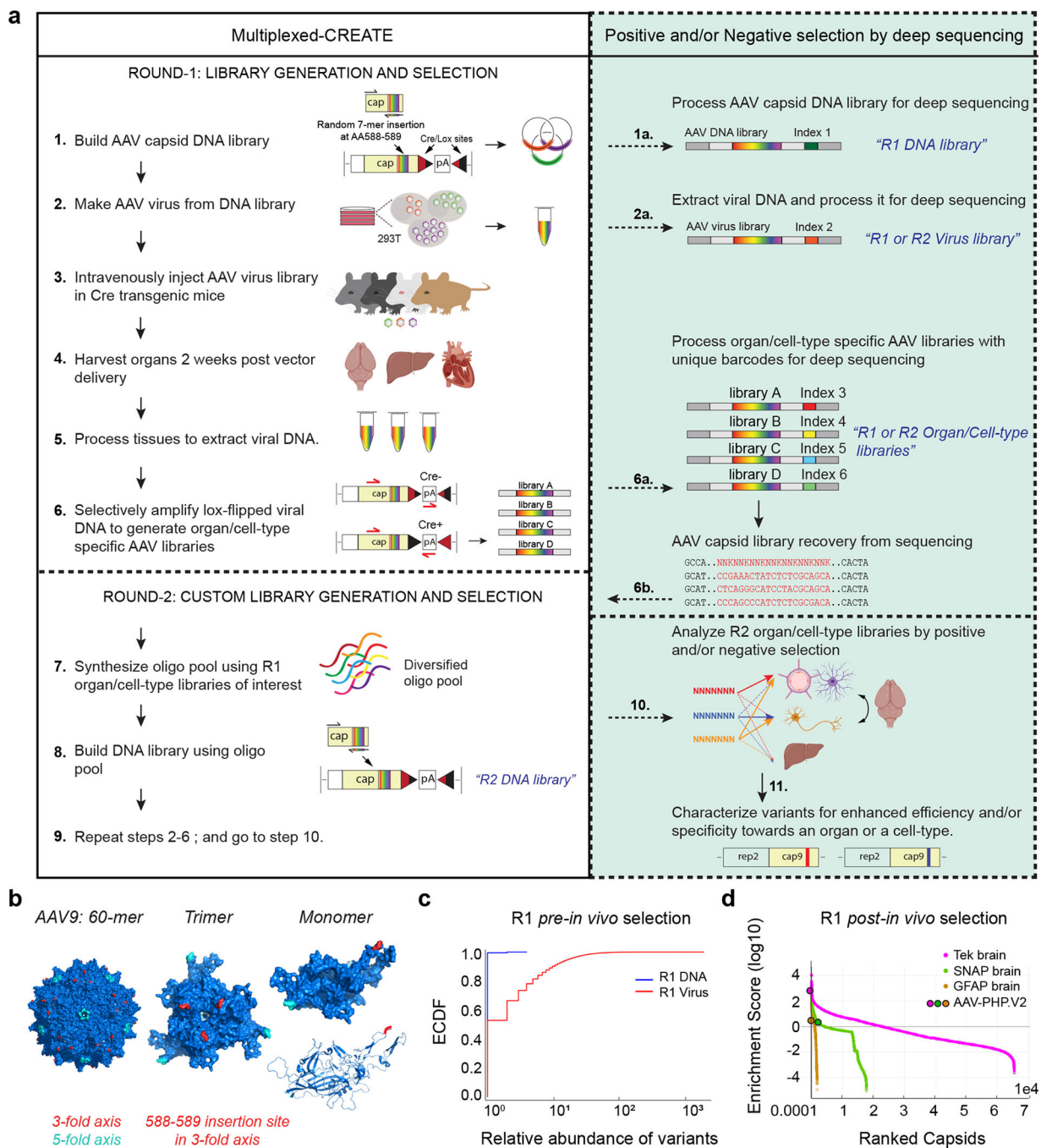
4. Gaj T, Epstein BE & Schaffer DV Genome Engineering Using Adeno-associated Virus: Basic and Clinical Research Applications. Mol. Ther 24, 458–464 (2016). [PubMed: 26373345]

5. Deverman BE, Ravina BM, Bankiewicz KS, Paul SM & Sah DWY Gene therapy for neurological disorders: progress and prospects. Nat. Rev. Drug Discov 17, 767 (2018). [PubMed: 30206384]

6. Sen D Improving clinical efficacy of adeno associated vectors by rational capsid bioengineering. J. Biomed. Sci 21, (2014).

7. Lee EJ, Guenther CM & Suh J Adeno-associated virus (AAV) vectors: Rational design strategies for capsid engineering. Curr. Opin. Biomed. Eng 7, 58–63 (2018). [PubMed: 31106283]

8. Bartlett JS, Kleinschmidt J, Boucher RC & Samulski RJ Targeted adeno-associated virus vector transduction of nonpermissive cells mediated by a bispecific F(ab'gamma)2 antibody. Nat. Biotechnol 17, 181–186 (1999). [PubMed: 10052356]

9. Davidsson M et al. A systematic capsid evolution approach performed in vivo for the design of AAV vectors with tailored properties and tropism. Proc. Natl. Acad. Sci 116, 27053–27062 (2019).

10. Bedbrook CN, Deverman BE & Gradinaru V Viral Strategies for Targeting the Central and Peripheral Nervous Systems. Annu. Rev. Neurosci 41, 323–348 (2018). [PubMed: 29709207]

11. Kotterman MA & Schaffer DV Engineering adeno-associated viruses for clinical gene therapy. Nat. Rev. Genet 15, 445–451 (2014). [PubMed: 24840552]

12. Grimm D et al. In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. J. Virol 82, 5887–5911 (2008). [PubMed: 18400866]

13. Maheshri N, Koerber JT, Kaspar BK & Schaffer DV Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. Nat. Biotechnol 24, 198–204 (2006). [PubMed: 16429148]

14. Excoffon KJDA et al. Directed evolution of adeno-associated virus to an infectious respiratory virus. Proc. Natl. Acad. Sci. U. S. A 106, 3865–3870 (2009). [PubMed: 19237554]

15. Pulicherla N et al. Engineering liver-detargeted AAV9 vectors for cardiac and musculoskeletal gene transfer. Mol. Ther. J. Am. Soc. Gene Ther 19, 1070–1078 (2011).

16. Ying Y et al. Heart-targeted adeno-associated viral vectors selected by in vivo biopanning of a random viral display peptide library. Gene Ther. 17, 980–990 (2010). [PubMed: 20393510]

17. Müller OJ et al. Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. Nat. Biotechnol 21, 1040–1046 (2003). [PubMed: 12897791]

18. Ogden PJ, Kelsic ED, Sinai S & Church GM Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. Science 366, 1139–1143 (2019). [PubMed: 31780559]

19. Pekrun K et al. Using a barcoded AAV capsid library to select for clinically relevant gene therapy vectors. JCI Insight 4, (2019).

20. Dalkara D et al. In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous. Sci. Transl. Med 5, 189ra76 (2013).

21. Davis AS et al. Rational design and engineering of a modified adeno-associated virus (AAV1)-based vector system for enhanced retrograde gene delivery. Neurosurgery 76, 216–225; discussion 225 (2015). [PubMed: 25549186]

22. Tordo J et al. A novel adeno-associated virus capsid with enhanced neurotropism corrects a lysosomal transmembrane enzyme deficiency. Brain 141, 2014–2031 (2018). [PubMed: 29788236]

23. Ojala DS et al. In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the SVZ. Mol. Ther. J. Am. Soc. Gene Ther 26, 304–319 (2018).

24. Tervo DGR et al. A Designer AAV Variant Permits Efficient Retrograde Access to Projection Neurons. Neuron 92, 372–382 (2016). [PubMed: 27720486]

25. Körbelin J et al. A brain microvasculature endothelial cell-specific viral vector with the potential to treat neurovascular and neurological diseases. EMBO Mol. Med 8, 609–625 (2016). [PubMed: 27137490]

26. Deverman BE et al. Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat. Biotechnol 34, 204–209 (2016). [PubMed: 26829320]

27. Chan KY et al. Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. Nat. Neurosci 20, 1172–1179 (2017). [PubMed: 28671695]

28. Lee Y, Messing A, Su M & Brenner M GFAP promoter elements required for region-specific and astrocyte-specific expression. Glia 56, 481–93 (2008). [PubMed: 18240313]

29. Kisanuki YY et al. Tie2-Cre transgenic mice: a new model for endothelial cell-lineage analysis in vivo. Dev. Biol 230, 230–242 (2001). [PubMed: 11161575]

30. de Leeuw CN et al. rAAV-compatible MiniPromoters for restricted expression in the brain and eye. Mol. Brain 9, (2016).

31. Rincon MY et al. Widespread transduction of astrocytes and neurons in the mouse central nervous system after systemic delivery of a self-complementary AAV-PHP.B vector. Gene Ther. 25, 83 (2018). [PubMed: 29523880]

32. Hordeaux J et al. The GPI-Linked Protein LY6A Drives AAV-PHP.B Transport across the Blood-Brain Barrier. Mol. Ther 27, 912–921 (2019). [PubMed: 30819613]

33. Matsuzaki Y et al. Neurotropic Properties of AAV-PHP.B Are Shared among Diverse Inbred Strains of Mice. Mol. Ther 27, 700–704 (2019). [PubMed: 30842039]

34. Huang Q et al. Delivering genes across the blood-brain barrier: LY6A, a novel cellular receptor for AAV-PHP.B capsids. PLOS ONE 14, e0225206 (2019). [PubMed: 31725765]

35. Challis RC et al. Systemic AAV vectors for widespread and targeted gene delivery in rodents. Nat. Protoc 1 (2019) doi:10.1038/s41596-018-0097-3.

36. Batista AR et al. Ly6a Differential Expression in Blood–Brain Barrier Is Responsible for Strain Specific Central Nervous System Transduction Profile of AAV-PHP.B. Hum. Gene Ther (2019) doi:10.1089/hum.2019.186.

37. Sweeney MD, Zhao Z, Montagne A, Nelson AR & Zlokovic BV Blood-Brain Barrier: From Physiology to Disease and Back. Physiol. Rev 99, 21–78 (2019). [PubMed: 30280653]

38. Betley JN & Sternson SM Adeno-Associated Viral Vectors for Mapping, Monitoring, and Manipulating Neural Circuits. Hum. Gene Ther 22, 669–677 (2011). [PubMed: 21319997]

39. Sweeney MD, Sagare AP & Zlokovic BV Blood-brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. Nat. Rev. Neurol 14, 133–150 (2018). [PubMed: 29377008]

40. Lykken EA, Shyng C, Edwards RJ, Rozenberg A & Gray SJ Recent progress and considerations for AAV gene therapies targeting the central nervous system. J. Neurodev. Disord 10, 16 (2018). [PubMed: 29776328]

41. Vlasov K, Van Dort CJ & Solt K Chapter Eleven - Optogenetics and Chemogenetics in Methods in Enzymology (eds. Eckenhoff RG & Dmochowski IJ) vol. 603 181–196 (Academic Press, 2018). [PubMed: 29673525]

42. Hwang B, Lee JH & Bang D Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp. Mol. Med 50, 96 (2018). [PubMed: 30089861]

REFERENCES (for Online Methods only)

43. Paulk NK et al. Bioengineered Viral Platform for Intramuscular Passive Vaccine Delivery to Human Skeletal Muscle. Mol. Ther. - Methods Clin. Dev 10, 144–155 (2018). [PubMed: 30101152]

44. Harris JA et al. Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. Front. Neural Circuits 8, 76 (2014). [PubMed: 25071457]

45. Garcia ADR, Doan NB, Imura T, Bush TG & Sofroniew MV GFAP-expressing progenitors are the principal source of constitutive neurogenesis in adult mouse forebrain. Nat. Neurosci 7, 1233–1241 (2004). [PubMed: 15494728]

46. Zhu Y et al. Ablation of NF1 function in neurons induces abnormal development of cerebral cortex and reactive gliosis in the brain. Genes Dev. 15, 859–876 (2001). [PubMed: 11297510]

47. Madisen L et al. A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat. Neurosci 13, 133–140 (2010). [PubMed: 20023653]

48. Choi HMT et al. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development 145, dev165753 (2018). [PubMed: 29945988]

49. Renier N et al. iDISCO: A Simple, Rapid Method to Immunolabel Large Tissue Samples for Volume Imaging. Cell 159, 896–910 (2014). [PubMed: 25417164]

50. Hama H et al. Sca/eS: an optical clearing palette for biological imaging. Nat. Neurosci 18, 1518–1529 (2015). [PubMed: 26368944]

51. DiMattia MA et al. Structural Insight into the Unique Properties of Adeno-Associated Virus Serotype 9. J. Virol 86, 6947–6958 (2012). [PubMed: 22496238]

52. Shannon P et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Res. 13, 2498–2504 (2003). [PubMed: 14597658]

53. Schneider TD & Stephens RM Sequence logos: a new way to display consensus sequences. Nucleic Acids Res. 18, 6097–6100 (1990). [PubMed: 2172928]

54. Crooks GE WebLogo: A Sequence Logo Generator. Genome Res. 14, 1188–1190 (2004). [PubMed: 15173120]

Figure 1: Workflow Of Multiplexed-CREATE And Analysis Of *7-mer-i* Selection In Round-1. a, A multiplexed selection approach to identify capsids with specific and broad tropisms. Steps 1–6 describe the workflow in Round-1 (R1) selection, steps 7–9 describe Round-2 (R2) selection using *synthetic pool* method, steps 1a, 2a, and 6a-b show the incorporation of deep sequencing to recover capsids after R1 and R2 selection, and steps 10–11 describe positive and/or negative selection criteria followed by variant characterization. **b**, Structural model of the AAV9 capsid (PDB 3UX1) with the insertion site for the *7-mer-i* library highlighted in red in the 60-meric (left), trimeric (middle), and monomeric (right) forms. **c**, Empirical
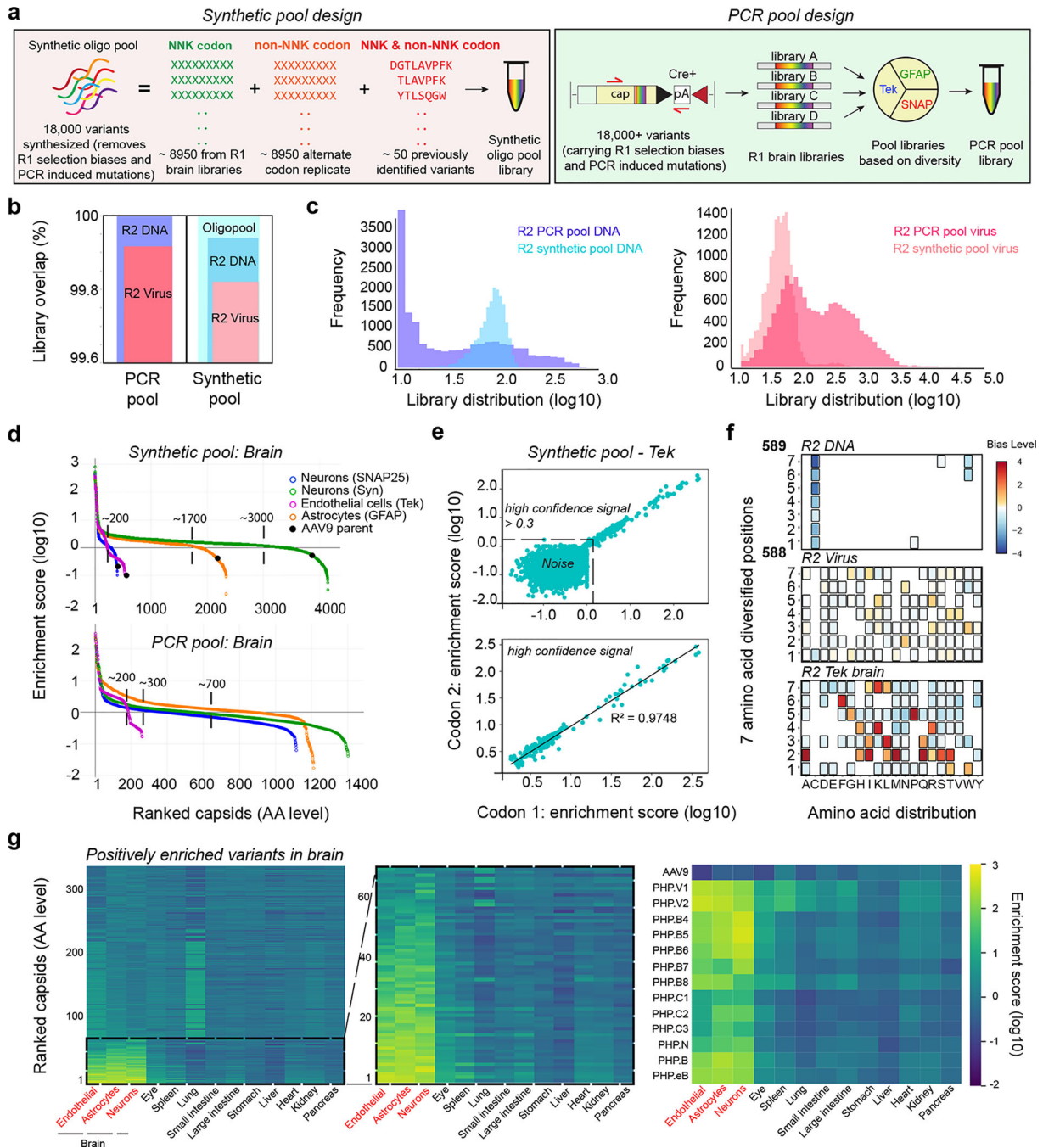
Cumulative Distribution Frequency (ECDF) of R1 DNA and virus libraries that were recovered by deep sequencing post Gibson assembly and virus production, respectively. **d**, Distributions of variants recovered from three R1 brain tissue libraries, Tek, SNAP25, and GFAP (n = 2 per Cre line), are shown with capsid libraries sorted by decreasing order of the enrichment score. The enrichment score of AAV-PHP.V2 variant, described later, is mapped on this plot.
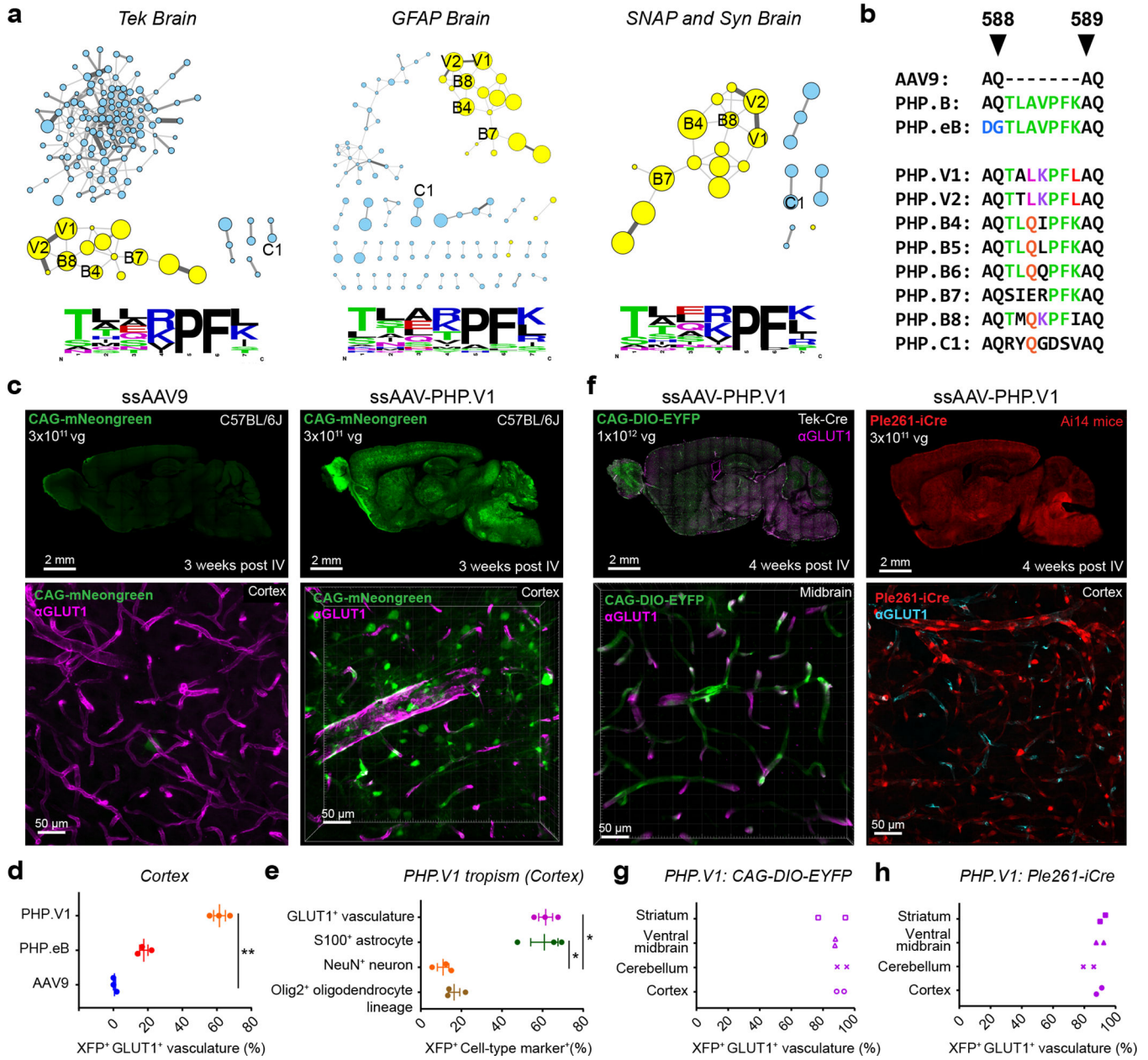
Figure 2: Round-2 Capsid Selections By *Synthetic Pool* And *PCR Pool* Methods. a,
Schematic of R2 *synthetic pool* (left) and *PCR pool* (right) library design. **b**, Overlapping
bar chart showing the percentage of library overlap between the mentioned libraries and
their library design with an assumption of 100% starting input library. **c**, Histograms of
DNA and virus libraries from the two methods, where the variants in a library are binned by
their read counts (in log10 scale) and the height of the histogram is proportional to their
frequency. **d**, Distributions of R2 brain libraries from all Cre transgenic lines (n= 2 mice per
Cre Line, mean is plotted) and both methods, where the libraries are sorted in decreasing
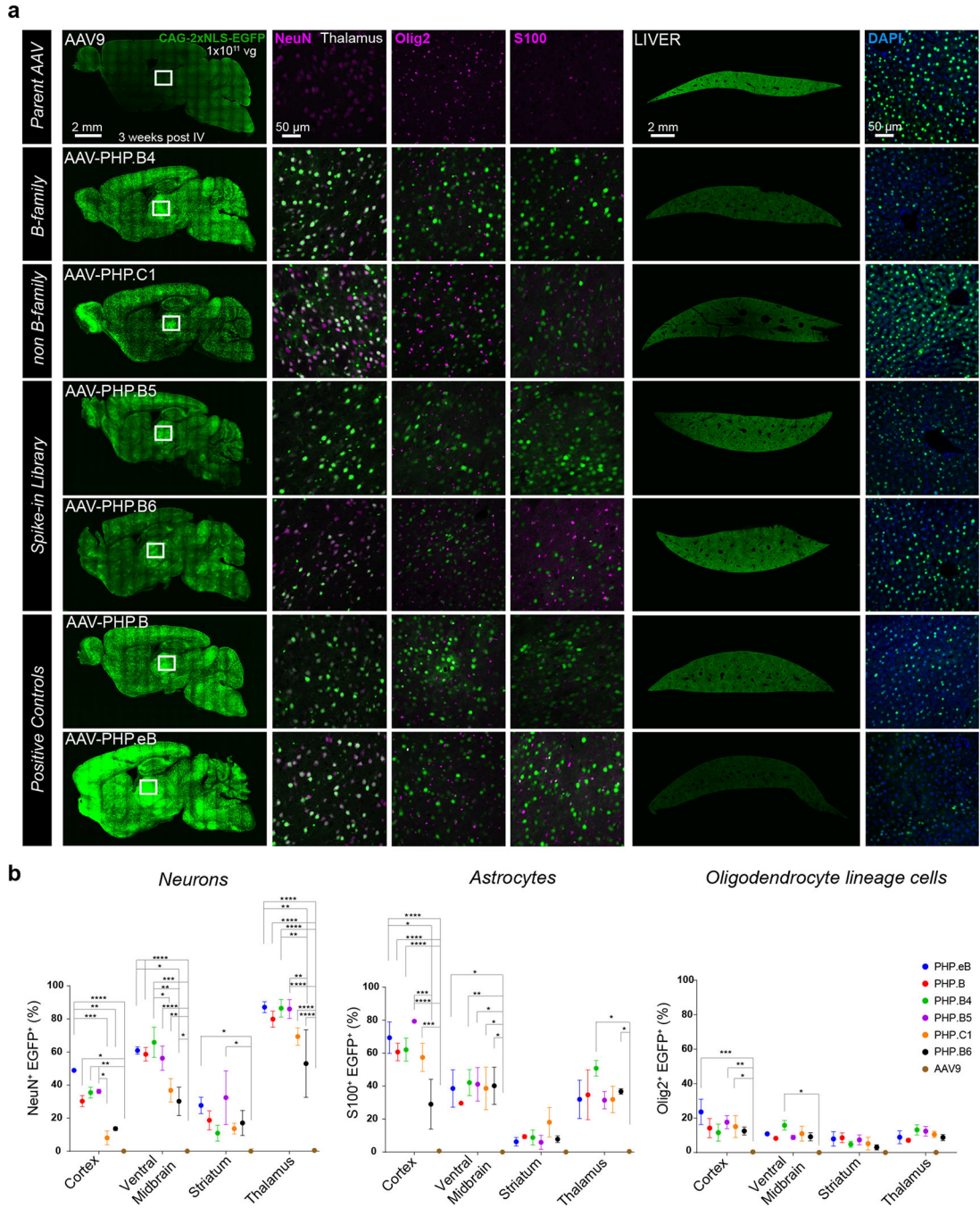
order of enrichment score (log10 scale). The total number of positively enriched variants from these libraries are highlighted by dotted straight lines and AAV9's relative enrichment is mapped on the synthetic pool plot..**e**, Comparison of the enrichment scores (log10 scale) of two alternate codon replicates for 8462 variants from the Tek-Cre brain library (n= 2 mice, mean is plotted). The broken line separates the high-confidence signal (>0.3) from noise. For the high-confidence signal (below), a linear least-squares regression is determined between the 2 codons and the regression line (best fit). The coefficient of determination R^2 is shown. **f**, Heatmaps representing the magnitude (log2 fold change) of a given AA's relative enrichment or depletion at each position given statistical significance is reached (boxed if P-value   0.0001, two-sided, two-proportion z-test, p-values corrected for multiple comparisons using Bonferroni correction). R2 DNA normalized to oligopool (top, ~9000 AA sequences), R2 virus normalized to R2 DNA (middle, n = ~9000 sequences), R2 Tek brain library with enrichment over 0.3 (high-confidence signal) from synthetic pool method normalized to R2 virus (bottom, 154 sequences) are shown (n= 2 for brain library, one per mouse. All other libraries, n = 1). **g**, Heatmap of Cre-independent relative enrichment across organs (n = 2 mice per Cre line, mean across 6 samples from 3 Cre lines is plotted) for variants positively enriched in the brain tissue of at least one Cre-dependent *synthetic pool* selection (red text, n = 2 mice per cell-type, mean is plotted) (left). Zoom-in of the most CNS-enriched variants (middle), and of the variants that are characterized in the current study along with *spike-in* library controls (right) are shown.

Figure 3: Selected AAV Capsids Form Distinct Sequence Families And Include PHP.B-Like Variants For Brain Wide Transduction Of Vasculature. a, Clustering analysis of positively enriched variants from Tek (left), GFAP (middle) and SNAP/Syn (right) *synthetic pool* brain libraries with size of nodes representing their relative enrichment in brain, and the thickness of edges (connecting lines) representing degree of relatedness. Distinct families (yellow) with the corresponding AA frequency logos (AA size represents prevalence and color encodes AA properties) are shown. **b**, The 7-mer insertion peptide sequences of AAV-PHP variants between AA positions 588–589 of AAV9 capsid are shown. AAs are colored by shared identity to AAV-PHP.B and eB (green) or among new variants (unique color per position). **c**, AAV9 (left) and AAV-PHP.V1 (right) mediated expression using ssAAV:CAG-mNeongreen genome (green, n = 3, 3 weeks of expression in
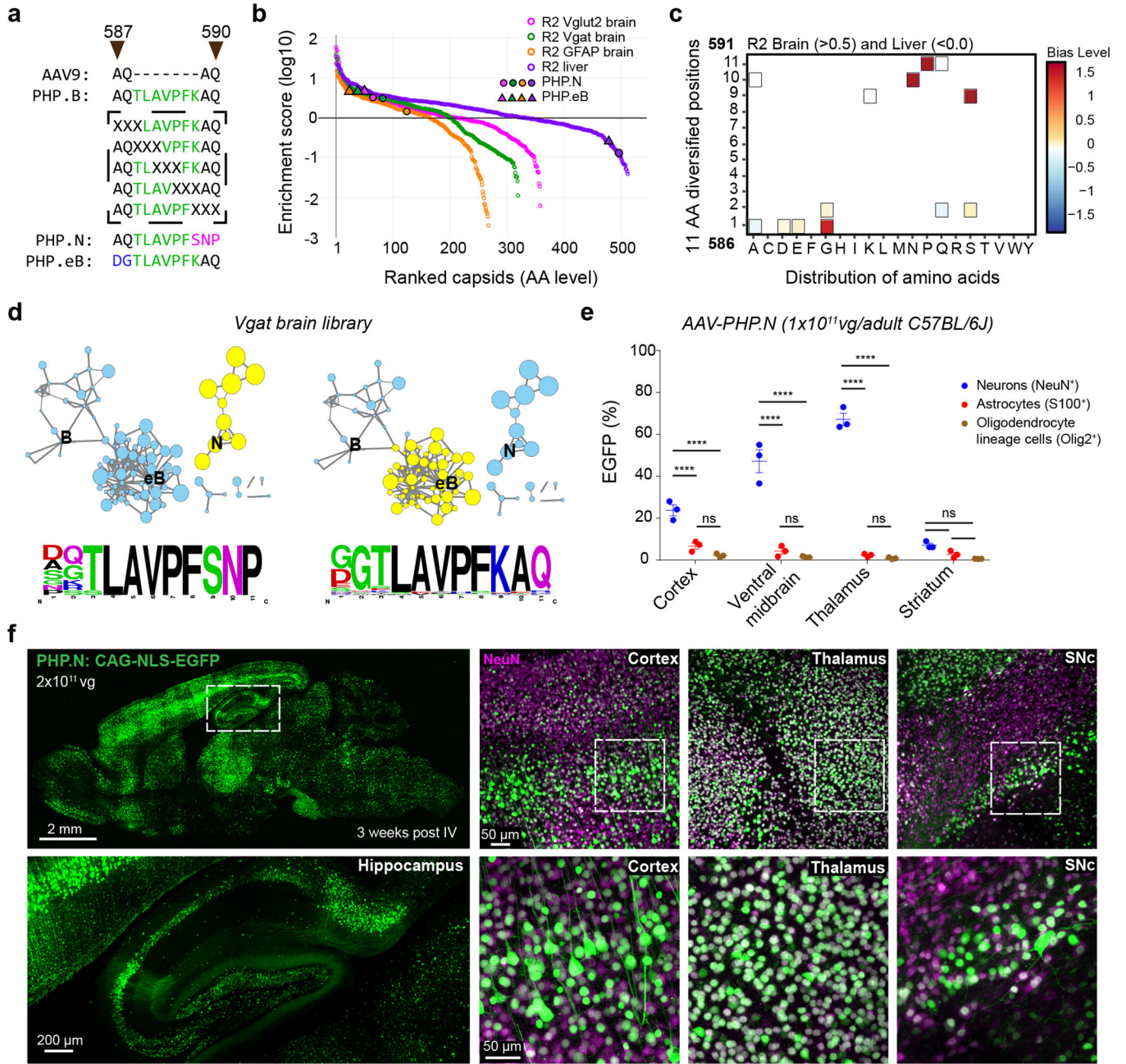
C57BL/6J adult mice with 3×10^{11} vg IV dose/mouse) is matched in fluorescence intensity in sagittal sections of brain (above) with higher magnification image from cortex (below). Magenta is αGLUT1 antibody staining for vasculature. **d**, Percentage of vasculature stained with αGLUT1 that overlaps with mNeongreen (XFP) expression in cortex. One-way ANOVA non-parametric Kruskal-Wallis test (P-value 0.0036), and follow-up multiple comparisons using uncorrected Dunn's test (P-value of 0.0070 for AAV9 vs PHP.V1) are reported. **P 0.01 is shown, P > 0.05 is not shown; data is mean ± S.E.M, n= 3 mice per AAV variant, cells quantified from 4–2 images per mouse per cell-type. **e**, Percentage of cells stained with each cell-type specific marker (αGLUT1, αS100 for astrocytes, αNeuN for neurons, αOlig2 for oligodendrocyte lineage cells) that overlaps with mNeongreen (XFP) expression in cortex. Kruskal-Wallis test (P-value of 0.0078), and uncorrected Dunn's test (P-value of 0.0235 for neuron vs vascular cells, and 0.0174 for neuron vs astrocyte, respectively) are reported. *P 0.05 is shown, and P > 0.05 is not shown; data is mean ± S.E.M, n= 3 mice, cells quantified from 4–2 images per mouse per cell-type. **f**, Vascular transduction by ssAAV-PHP.V1:CAG-DIO-EYFP in Tek-Cre adult mice (left) ($n = 2$, 4 weeks of expression, 1×10^{12} vg IV dose/mouse), and by ssAAV-PHP.V1:Ple261-iCre in Ai14 reporter mice (right) ($n = 2$, 3 weeks of expression, 3×10^{11} vg IV dose/mouse). Tissues are stained with αGLUT1 (magenta (left) and cyan (right)). **g**, Efficiency of vascular transduction (as described in **d**) in Tek-Cre mice (n= 2, mean from 3 images per mouse per brain region). **h**, Efficiency of vascular transduction in Ai14 mice (n= 2, a mean from 4 images per mouse per brain region).

Figure 4: Characterization Of Round-2 Brain Libraries Has Identified Additional Capsids Exhibiting Broad CNS Tropism. a,
Transduction by AAV-PHP.B4–B6 and C1 variants, as well as B, eB, and AAV9 controls in sagittal brain and liver sections. Fluorescence intensity is matched with AAV-PHP.eB across each set of images (column-wise). The white box on the sagittal brain images marks the thalamus and not the precise region of the figures to the right. Vectors are packaged with ssAAV:CAG-2xNLS-EGFP genome ($n = 3$ per group, 1×10^{11} vg IV dose/adult C57BL/6J mouse, 3 weeks of expression). Tissues are stained with cell-type specific markers (magenta): αNeuN for neurons, αS100 for astrocytes and αOlig2 for oligodendrocyte

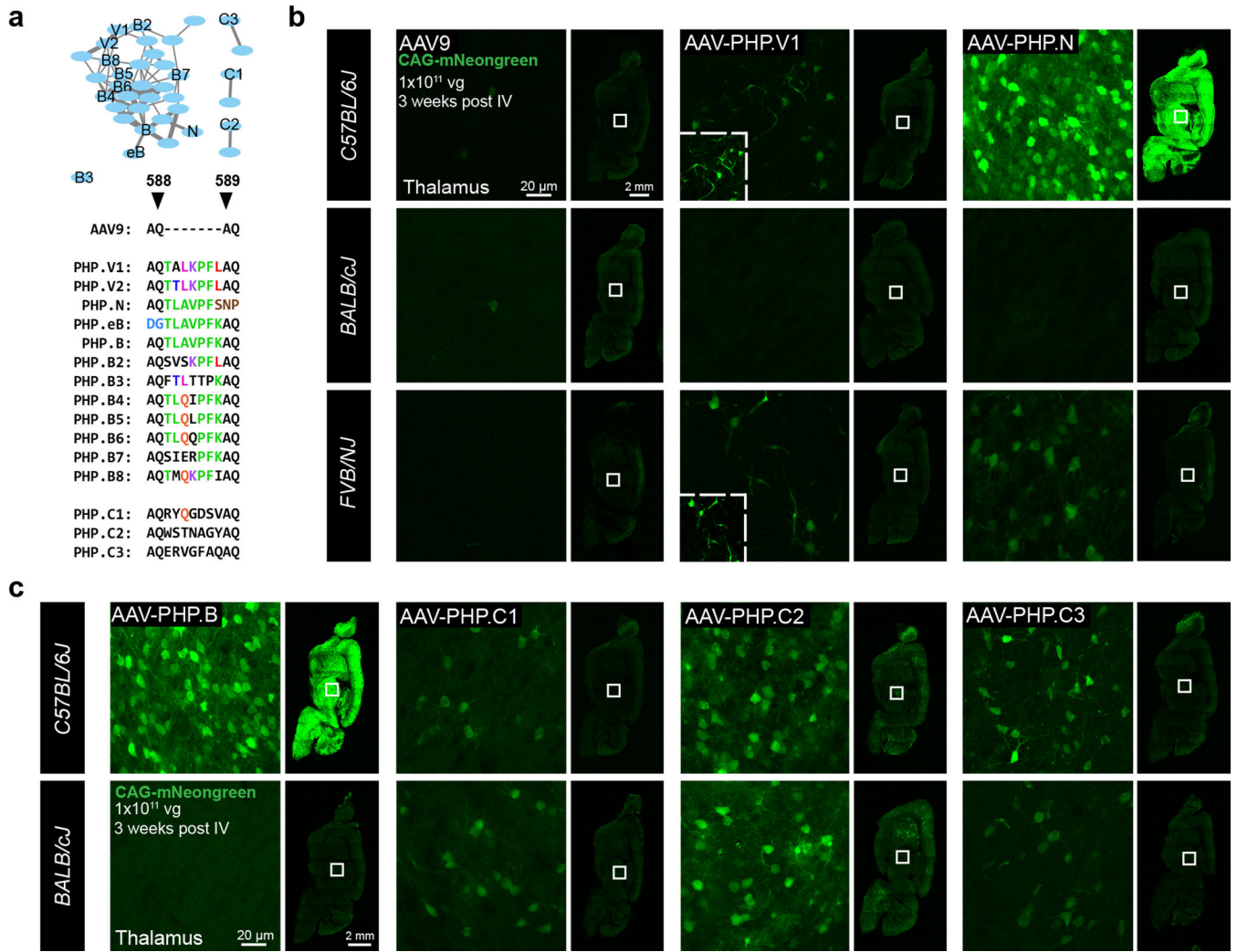lineage cells. Liver tissues are stained with a DNA stain, DAPI (blue). **b,** The percentage of αNeuN$^+$, αS100$^+$ and αOlig2$^+$ cells with detectable nuclear-localized EGFP in the indicated brain regions are shown (n=3 per group, 1×10^{11} vg dose). A two-way ANOVA with correction for multiple comparisons using Tukey's test is reported with adjusted P-values (****P ≤ 0.0001, ***P ≤ 0.001, **P ≤ 0.01, *P ≤ 0.05, is shown, and P > 0.05 is not shown on the plot; 95% CI, data is mean ± S.E.M. The dataset comprises a mean of 2 images per region per cell-type marker per mouse).

Figure 5: Recovery Of Several AAV-PHP.B Variants Including One Exhibiting Higher Specificity For Neurons. a,
The design of the *3-mer-s* PHP.B library with combinations of three AA diversification between AA 587–597 of AAV-PHP.B (or corresponding AA 587–590 of AAV9). Shared AA identity with the parent AAV-PHP.B (green) is shown along with unique motifs for AAV-PHP.N (pink) and AAV-PHP.eB (blue). **b,** Distributions of R2 brain and liver libraries (at AA level) by enrichment score (normalized to R2 virus library, with variants sorted in decreasing order of enrichment score). The enrichment of AAV-PHP.eB and AAV-PHP.N across all libraries are mapped on the plot. **c,** Heatmap represents the magnitude (log2 fold change) of a given AA's relative enrichment or depletion at each position across the diversified region,

only if statistical significance is reached on fold change (boxed if p-value 0.0001, two-sided, two-proportion z-test, p-values corrected for multiple comparisons using Bonferroni correction). Plot includes variants that were highly enriched in brain (>0.5 mean enrichment score, where mean is drawn across Vglut2, Vgat and GFAP, n = 1 library per mouse line (sample pooled from 2 mice per line)) and negatively enriched in liver (<0.0) (32 AA sequences). **d**, Clustering analysis of positively enriched variants from Vgat brain library is shown with node size representing the degree of negative enrichment in liver and the thickness of edges (connecting lines) representing degree of relatedness between nodes. Two distinct families are highlighted in yellow and their corresponding AA frequency logos are shown below (AA size represents prevalence and color encodes AA properties). **e**, The percentage of neurons, astrocytes and oligodendrocyte lineage cells with ssAAV-PHP.N:CAG-2xNLS-EGFP in the indicated brain regions is shown ($n = 3$, 1×10^{11} vg IV dose per adult C57BL/6J mouse, 3 weeks of expression, data is mean±S.E.M, 6–8 images for cortex, thalamus and striatum, and 2 images for ventral midbrain, per mouse per cell-type marker using 20x objective covering the entire regions). A two-way ANOVA with correction for multiple comparisons using Tukey's test gave adjusted P-values reported as ****P 0.0001, ns for P > 0.05, 95% CI. **f**, Transduction by ssAAV-PHP.N:CAG-NLS-EGFP ($n = 2$, 2×10^{11} vg IV dose per adult C57BL/6J mouse, 3 weeks of expression) is shown with NeuN staining (magenta) across three brain areas (cortex, SNc (substantia nigra pars compacta) and thalamus).

Figure 6: Summary Of Engineered AAV Capsids And Investigation Of Variants From Distinct Families Across Mouse Strains. a, Clustering analysis showing the brain-enriched sequence families of all variants described herein, either identified in prior studies (PHP.B-B3, PHP.eB) or in the current study (PHP.B4–B8, PHP.V1–2, PHP.C1–3). The thickness of edges (connecting lines) representing degree of relatedness between nodes. The AA sequences inserted between 588–589 (of AAV9 capsid) for all the variants discussed are shown below. **b**, Transduction of AAV9, AAV-PHP.V1 and AAV-PHP.N across three different mouse strains: C57BL/6J, BALB/cJ and FVB/NJ are shown in sagittal brain sections (right), along with a higher magnification image of the thalamus brain region (left). **c**, Transduction by AAV-PHP.B, AAV-PHP.C1–C3 in C57BL/6J and BALB/cJ mice are shown in sagittal brain sections (right), along with a higher magnification image of the thalamus brain region (left). **b,c,** The white box on the sagittal brain images represents the location of thalamus and not the precise area that is zoomed-in on the figure to the left. The fluorescence intensity is matched across all sagittal sections and across all thalamus regions acquired. The insets in AAV-PHP.V1 are zoom-ins with enhanced brightness. The indicated capsids were used to package ssAAV:CAG-mNeongreen ($n = 2$–3 per group, 1×10^{11} vg IV dose per 6–8 weeks old adult mouse, 3

weeks of expression. The data reported in **b,c** are from one independent trial where all viruses were freshly prepared and titered in the same assay for dosage consistency. AAV-PHP.C2 and AAV-PHP.C3 were further validated in an independent trial for BALB/cJ, n = 2 per group).