nature methods

ARTICLES

https://doi.org/10.1038/s41592-020-0799-7

Check for updates

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

Sripriya Ravindra Kumar[1], Timothy F. Miles[1,5], Xinhong Chen[1,5], David Brown[1], Tatyana Dobreva[1], Qin Huang[1,2], Xiaozhe Ding[1], Yicheng Luo[1], Pétur H. Einarsson[1], Alon Greenbaum[1,3,4], Min J. Jang[1], Benjamin E. Deverman[1,2] and Viviana Gradinaru[1] ✉

Recombinant adeno-associated viruses (rAAVs) are efficient 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 Cre-recombination-based AAV targeted evolution (CREATE) platform, we developed Multiplexed-CREATE (M-CREATE) to identify variants of interest in a given selection landscape through multiple positive and negative selection criteria. M-CREATE incorporates next-generation sequencing, synthetic library generation and a dedicated analysis pipeline. We have identified capsid 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, the M-CREATE methodology accelerates the discovery of capsids for use in neuroscience and gene-therapy applications.

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 to identify a complete toolbox of efficient and specific vectors. Previously, we evolved the AAV-PHP.B and AAV-PHP.eB variants from AAV9 using a selection method called CREATE[26,27]. This method applies positive selective pressure for capsids capable of infecting a target cell population by pairing a viral genome containing lox sites with in vivo selection in transgenic mice expressing Cre in the cell type of interest. This combination allows a Cre–Lox recombination-dependent PCR amplification of only those capsids which successfully deliver their genomes to the nuclei of the target cell type.

To more efficiently expand the AAV toolbox, we developed Multiplexed-CREATE (M-CREATE) (Fig. 1a and Supplementary Fig. 1a,b), which compares 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 an 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 one (R1) synthetic pool library with equal variant representation and (3) the reduction of false positives by including codon replicates of each selected variant in the pool. These improvements allow interpretation of variants' relative infection efficiencies across a broad range of enrichments in multiple positive selections and enable post-hoc negative screening by comparing capsid libraries recovered from multiple target cell types or organs. Collectively, these features allow prioritization of capsid variants for validation and characterization.

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

Results

Analysis of capsid libraries during round-1 selection. M-CREATE was developed to enable the analysis of capsid variants' behavior within and across in vivo selections. By doing so, we aimed to identify capsids with diverse tropisms, as well as reveal the capsid sequence diversity within a given tropism. M-CREATE achieves these aims by incorporating NGS and a synthetic capsid library for round-2 (R2) in vivo selection along with a dedicated analysis pipeline to assign capsid enrichment values.

During DNA- and virus-library generation there is potential for biased accumulation and over-representation of certain capsid variants, obscuring their true enrichment during in vivo selection. These deviations may result from PCR amplification bias in the DNA library or sequence bias in the efficiency of virus production across various steps such as capsid assembly, genome packaging

[1]Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA. [2]Present address: Stanley Center for Psychiatric Research, Broad Institute, Cambridge, MA, USA. [3]Present address: Joint Department of Biomedical Engineering, North Carolina State University, Raleigh, NC, USA. [4]Present address: University of North Carolina at Chapel Hill, Chapel Hill, NC, USA. [5]These authors contributed equally: Timothy F. Miles, Xinhong Chen. ✉e-mail: viviana@caltech.edu

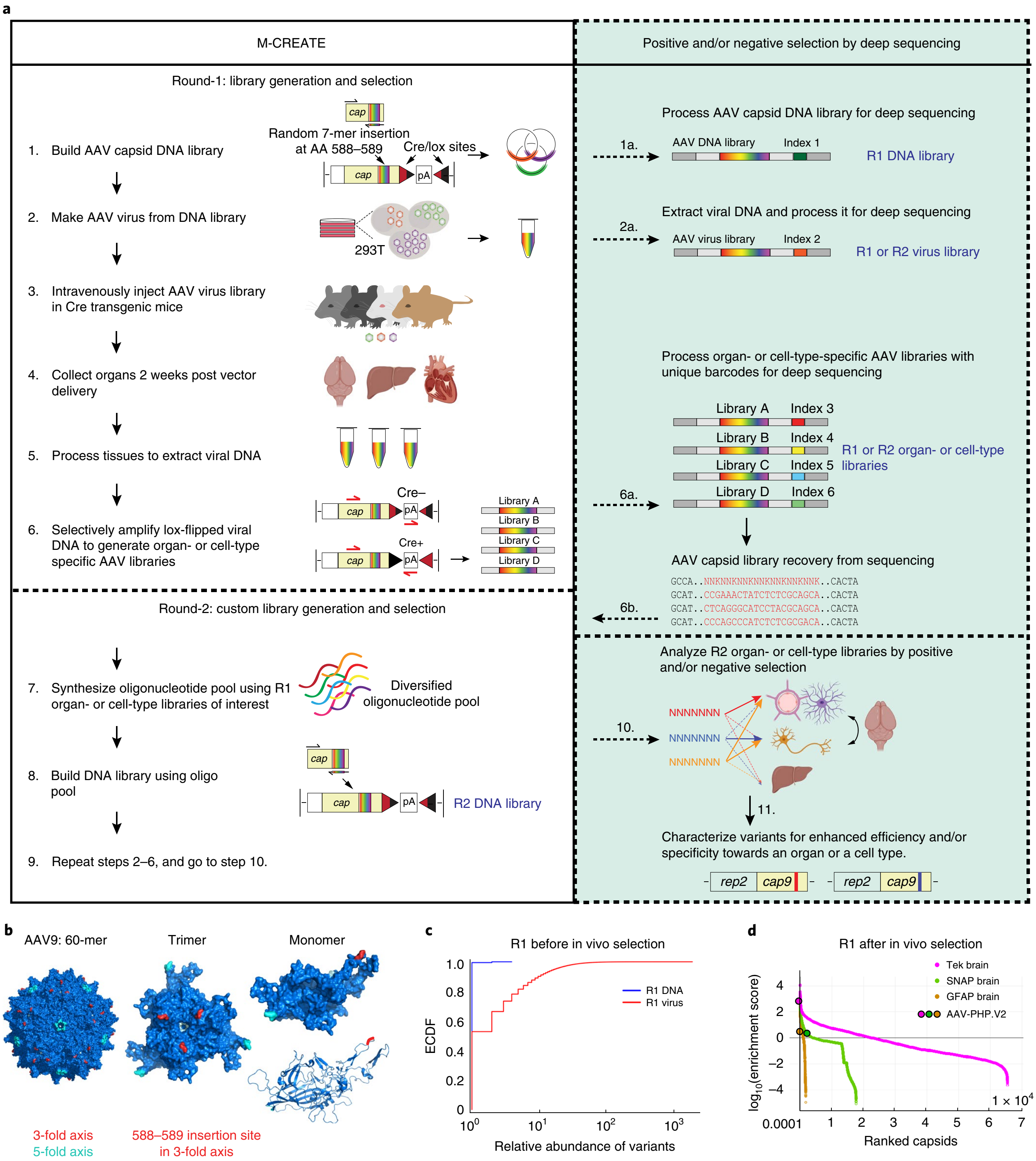

a
M-CREATE
Positive and/or negative selection by deep sequencing
Round-1: library generation and selection
1. Build AAV capsid DNA library
cap
Random 7-mer insertion at AA 588–589
Cre/lox sites
cap
pA
2. Make AAV virus from DNA library
293T
3. Intravenously inject AAV virus library in Cre transgenic mice
4. Collect organs 2 weeks post vector delivery
5. Process tissues to extract viral DNA
6. Selectively amplify lox-flipped viral DNA to generate organ- or cell-type specific AAV libraries
Cre–
cap
pA
Cre+
cap
pA
Library A
Library B
Library C
Library D
Round-2: custom library generation and selection
7. Synthesize oligonucleotide pool using R1 organ- or cell-type libraries of interest
Diversified oligonucleotide pool
8. Build DNA library using oligo pool
cap
cap
pA
R2 DNA library
9. Repeat steps 2–6, and go to step 10.
Process AAV capsid DNA library for deep sequencing
1a.
AAV DNA library
Index 1
R1 DNA library
Extract viral DNA and process it for deep sequencing
2a.
AAV virus library
Index 2
R1 or R2 virus library
Process organ- or cell-type-specific AAV libraries with unique barcodes for deep sequencing
Library A
Index 3
Library B
Index 4
R1 or R2 organ- or cell-type libraries
Library C
Index 5
6a.
Library D
Index 6
AAV capsid library recovery from sequencing
GCCA..NNKNNKNNKNNKNNKNNKNNK..CACTA
GCAT..CCGAAACTATCTCTCGCAGCA..CACTA
GCAT..CTCAGGGCATCCTACGCAGCA..CACTA
GCAT..CCCAGCCCATCTCTCGCGACA..CACTA
6b.
Analyze R2 organ- or cell-type libraries by positive and/or negative selection
10.
NNNNNNN
NNNNNNN
NNNNNNN
11.
Characterize variants for enhanced efficiency and/or specificity towards an organ or a cell type.
rep2
cap9
rep2
cap9
b
AAV9: 60-mer
Trimer
Monomer
3-fold axis
5-fold axis
588–589 insertion site in 3-fold axis
c
R1 before in vivo selection
ECDF
1.0
0.8
0.6
0.4
0.2
0
R1 DNA
R1 virus
10^0
10^1
10^2
10^3
Relative abundance of variants
d
R1 after in vivo selection
$\log_{10}$(enrichment score)
4
2
0
−2
−4
Tek brain
SNAP brain
GFAP brain
AAV-PHP.V2
1 × 10^4
0.0001
1
2
3
4
5
6
7
Ranked capsids

Fig. 1 | Workflow of M-CREATE and analysis of 7-mer-i selection in R1. a, A multiplexed selection approach to identify capsids with specific and broad tropisms. Steps 1–6 describe the workflow in R1 selection, steps 7–9 describe R2 selection using the 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. The genes *rep2* and *cap9* in step 11 refer to *rep* from AAV2 and *cap* from AAV9, respectively, and the colored bar within *cap9* represents the targeted mutation. **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 libraries from Tek-Cre, SNAP25-Cre and GFAP-Cre brain tissue ($n = 2$ per Cre line) are shown with capsid libraries, sorted by decreasing order of the enrichment score. The enrichment scores of the AAV-PHP.V2 variant are mapped as well.

and purification. We investigated this with a 7-mer-i (i for insertion) library, in which a randomized 7-amino-acid (AA) library is inserted between AA 588 and 589 of AAV9 (Fig. 1a,b) in the rAAV-ΔCap9-in-cis-Lox2 plasmid (Methods and Supplementary Fig. 1a; theoretical library size, 3.4×10^{10} unique nucleotide sequences, and an estimated $\sim1\times10^{8}$ nucleotide sequences upon transfection). We sequenced the libraries after DNA assembly and after virus purification to a depth of 10–20 million (M) reads, which was adequate to capture the bias among variants during virus production (Fig. 1c and Supplementary Fig. 1b–d). 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. Thus, even permissive sites like 588–589 will impose biological constraints on sampled sequence space.

For in vivo selection, we intravenously administered the 7-mer-i viral library in transgenic mice expressing Cre in astrocytes (GFAP-Cre), neurons (SNAP25-Cre) or endothelial cells (Tek-Cre) at a dose of 2×10^{11} vector genomes (vg) per adult mouse ($n=2$ mice per Cre transgenic line). Two weeks after intravenous (i.v.) injection, we collected 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 $\sim8\times10^{4}$ unique nucleotide variants in brain tissue samples (~48% of which were identified in the sequenced portion of the virus library) and <50 variants in spinal-cord samples across the transgenic lines, and each variant was represented with an enrichment score that reflects the change in relative abundance between the brain and the starting virus library (Methods and Fig. 1d).

Two features of this dataset stand out. First, the variants recovered from brain tissue were disproportionately represented in the sequenced fraction of the viral library, demonstrating how production biases can skew selection results. Second, the distribution of capsid RCs reveals that more than half of the unique variants recovered after selection appear at low RCs. These variants may either have arisen spontaneously from errors during experimental manipulation or retain AAV9's basal levels of central nervous system (CNS) transduction (Supplementary Fig. 1e).

Unbiased 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 R1 output (synthetic pool library) via oligonucleotide pools. We compared this library with a library PCR amplified directly from the recovered R1 DNA (PCR pool library) (Fig. 2a and Supplementary Table 1).

The synthetic pool library design comprised: (1) equimolar amounts of ~8,950 capsid variants present at high read counts in at least one of the R1 selections from brain and spinal cord (Supplementary Fig. 1e); (2) alternative codon replicates of those ~8,950 variants (optimized for mammalian codons) to reduce false positives; and (3) a spike-in library of controls (Supplementary Note 1 and Supplementary Dataset 1), resulting in a total library size of 18,000 nucleotide variants.

Both R2 virus libraries produced a high titer ($\sim6\times10^{11}$ vg per 10 ng of R2 DNA library per 150-mm dish; Supplementary Fig. 2a), and ~99% of variants of the R2 DNA were found after viral production (Fig. 2b). However, the distribution of the DNA and virus libraries from both designs differed notably. The PCR pool library carries forward the R1 selection biases (Fig. 2c and 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 Cre lines ($n=2$ mice per Cre transgenic line, GFAP, SNAP25, Tek), as well as the Syn-Cre line (for neurons). Two weeks after i.v. injection, we extracted, selectively amplified and deep-sequenced rAAV genomes from brain samples (as in R1). The synthetic pool library produced a greater number of enriched capsid variants than the PCR pool brain library (for example, ~1,700 versus ~700 variants per tissue library at the AA level in GFAP-Cre mice) (Fig. 2d and Supplementary Fig. 2d). In the synthetic pool, ~90% of the variants from the spike-in library were enriched (Supplementary Fig. 2d, middle panel, and Supplementary Dataset 1).

The degree of correlation between variant enrichment scores for PCR and synthetic pool libraries varies in each Cre transgenic line, indicating the presence of noise within experiments (Supplementary Fig. 2e and 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 and Supplementary Fig. 3a,b). This is a substantial advantage over the PCR pool design, allowing us to confidently interpret enrichment scores in a given selection.

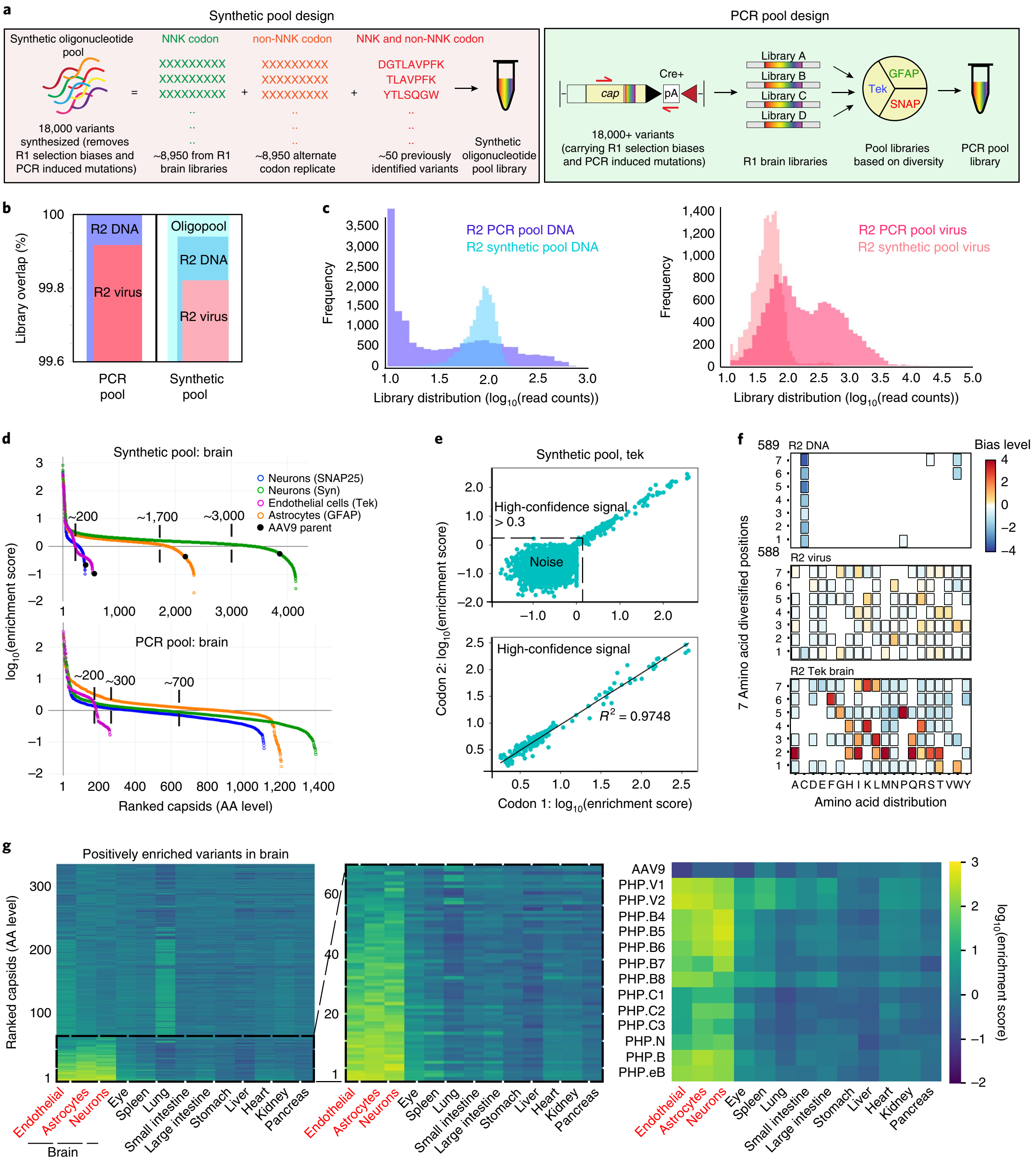
Analysis of capsid libraries after round-2 selections. Whereas the amino acid distribution of the DNA library closely matched the Oligopool design, virus production selected for a motif within the 7-AA diversified insertion (between AA 588 and AA 589), with Asn at position 2, β-branched amino acids (I, T, V) at position 4 and positively charged amino acids (K, R) at position 5 (Fig. 2f and

Fig. 2 | R2 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 theoretical composition. **c**, Histograms of DNA and virus libraries from the two methods, where the variants in a library are binned by their RCs (in $\log_{10}$ 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, in which the libraries are sorted in decreasing order of enrichment score ($\log_{10}$ 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 ($\log_{10}$ scale) of two alternate codon replicates for 8,462 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 two codons and the regression line (best fit). The coefficient of determination r^2 is shown. **f**, Heat maps representing the magnitude ($\log_2$(fold change)) of a given amino acid's relative enrichment or depletion at each position given statistical significance is reached (boxed if $P\leq0.0001$, two-sided, two-proportion z-test, P values corrected for multiple comparisons using Bonferroni correction). R2 DNA normalized to oligopool (top, ~9,000 sequences), R2 virus normalized to R2 DNA (middle, n = ~9,000 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**, Heat map 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 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.

Supplementary Fig. 3c). Fitness for BBB crossing resulted in a different pattern. For instance, variants highly enriched after recovery from brain tissue (across all Cre lines) shared preferences for Pro in position 5, and Phe in position 6.

By assessing enrichment score reproducibility within the synthetic pool design, we next determined the brain-enriched variants' distribution across peripheral organs (Fig. 2g, left). About 60 variants that are highly enriched in brain are comparatively depleted across other organs (Fig. 2g, middle). Encouraged by the expected behavior of spike-in control variants (AAV9, PHP.B, PHP.eB), we chose eleven additional variants 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).

We chose these variants due to their enrichments and where they fall in sequence space. We noticed that the enriched variants cluster into distinct families based on sequence similarity. The most enriched variants form a distinct family across selections with a com-



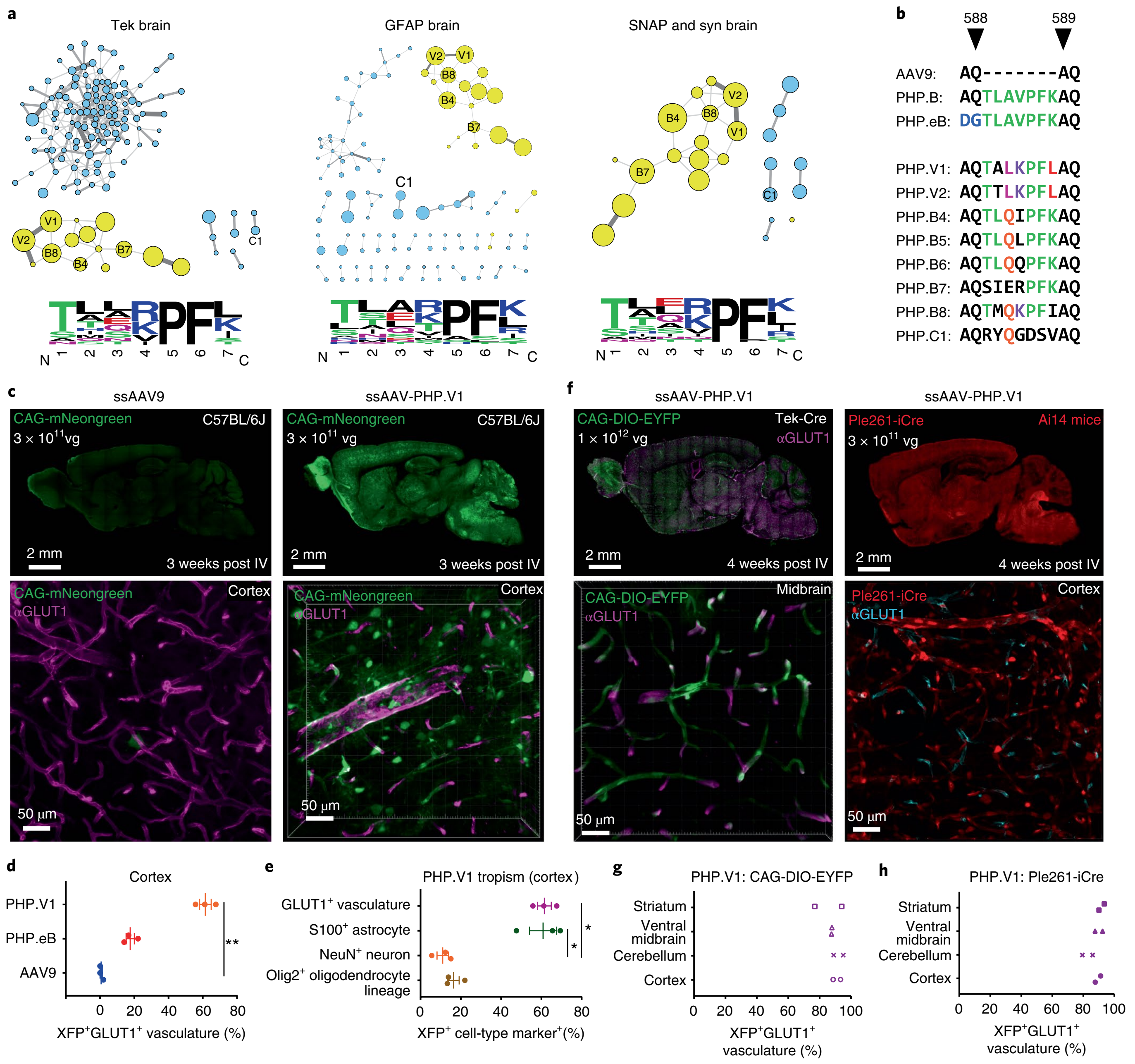


Fig. 3 | Selected AAV capsids form sequence families and include variants for brain-wide transduction of vasculature. a, Clustering analysis of variants from synthetic pool brain libraries after enrichment in Tek-Cre (left), GFAP-Cre (middle) and combined SNAP-Cre and Syn-Cre (right) selections. The size of the nodes represents relative enrichment in the brain. Thickness of the edges (connecting lines) represents the 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-AA 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 i.v. dose per mouse, imaged under the same settings) in sagittal sections of brain (top) with higher-magnification image from cortex (bottom). Magenta, αGLUT1 antibody staining for vasculature. **d**, Percentage of vasculature stained with αGLUT1 that overlaps with mNeongreen (XFP) expression in cortex. One-way analysis of variance (ANOVA) non-parametric Kruskal–Wallis test ($P=0.0036$), and follow-up multiple comparisons using uncorrected Dunn's test ($P=0.0070$ for AAV9 versus PHP.V1) are reported. $^{**}P\leq0.01$ is shown, $P>0.05$ is not shown; data are mean ± s.e.m., $n=3$ mice per AAV variant, cells quantified from 2–4 images per mouse per cell type. **e**, Percentage of cells stained with each cell-type specific marker (αGLUT1, αS100 for astrocytes, αNeuN for neurons and αOlig2 for oligodendrocyte lineage cells) that overlaps with mNeongreen (XFP) expression in cortex. Kruskal–Wallis test ($P=0.0078$), and uncorrected Dunn's test ($P=0.0235$ for neuron versus vascular cells, and 0.0174 for neuron versus astrocyte) are reported. $^{*}P\leq0.05$ is shown, and $P>0.05$ is not shown; data are mean ± s.e.m., $n=3$ mice, cells quantified from 2–4 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 i.v. dose per mouse), and by ssAAV-PHP.V1:Ple261-iCre in Ai14 reporter mice (right) ($n=2$, 3 weeks of expression, 3×10^{11} vg i.v. dose per 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).

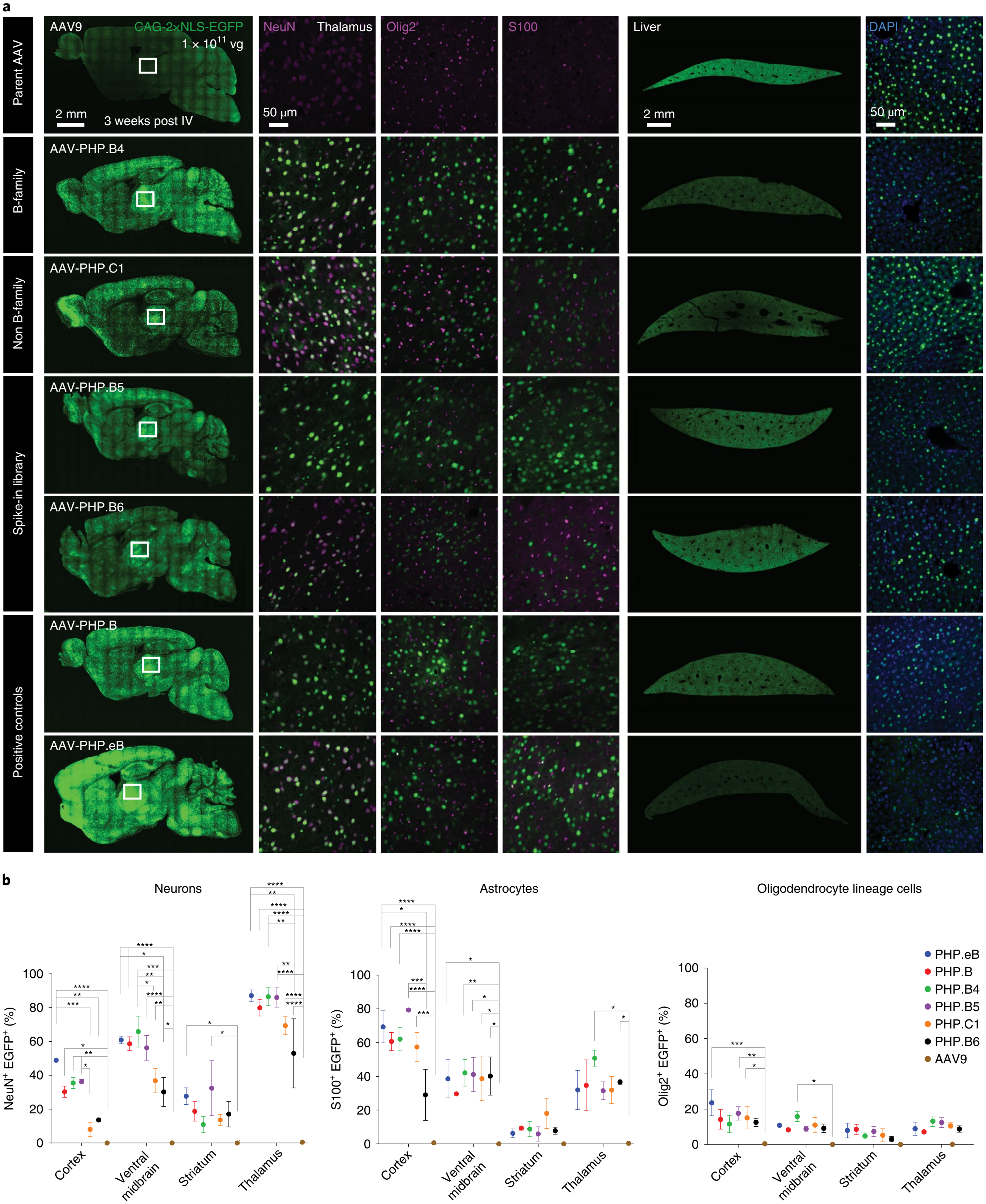


Fig. 4 | Characterization of R2 brain libraries and identification of capsids with broad CNS tropism. a, Transduction by AAV-PHP.B4-B6 and C1 variants, as well as B, eB and AAV9 controls in sagittal brain and liver sections (each column was imaged under the same settings). White box, thalamus (this is 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 i.v. dose per 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 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 (****$P \leq 0.0001$, ***$P \leq 0.001$, **$P \leq 0.01$, *$P \leq 0.05$, is shown, and $P > 0.05$ is not shown on the plot; 95% confidence interval (CI), data are mean ± s.e.m. The dataset comprises a mean of two images per region per cell-type marker per mouse).

mon 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 and Supplementary Fig. 3d). This amino acid pattern closely resembles the TLAVPFK motif in the previously identified variant AAV-PHP.B (ref. [26]) (Supplementary Note 3). Given the sequence similarity among members of this family, we next tested whether selected variants can cross the BBB and target the CNS with similar efficiency and tropism.

AAV9 variants with enhanced BBB entry and CNS transduction. Given the dominance of the PHP.B family in the R2 selection, we characterized its most enriched member, harboring a TALKPFL motif and henceforth referred to as AAV-PHP.V1 (Fig. 3a,b). Despite its sequence similarity to AAV.PHP.B, the tropism of AAV-PHP.V1 is biased toward transducing brain vascular cells (Fig. 3c and 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 with ~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 i.v. 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),g and Supplementary Video 1), (2) in fluorescent reporter mice where Cre is delivered with an endothelial-cell-type specific MiniPromoter (Ple261)[30] (Fig. 3f (right),h and 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 difference in tropism between AAV-PHP.V1 and AAV-PHP.B or AAV-PHP.eB, we characterized several additional variants within the PHP.B-like family. One variant, AAV-PHP.V2, harboring the TTLKPFL 7-mer sequence and differing by only one amino acid from AAV-PHP.V1, has a similar tropism (Supplementary Fig. 5 and Supplementary Note 4). Three other variants with sequences of roughly equal deviation from both AAV.PHP.V1 and AAV.PHP.B, AAV-PHP.B4 with TLQIPFK, AAV-PHP.B7 with SIERPFK and AAV-PHP.B8 with TMQKPFI (Figs. 3a,b and 4a,b) have PHP.B-like tropism with biased transduction toward neurons and astrocytes (Fig. 4b and Supplementary Fig. 6a–c). Similar variants among the spike-in library, AAV-PHP.B5 with TLQLPFK and AAV-PHP.B6 with TLQQPFK, also shared this tropism (Figs. 3b and 4a,b, Supplementary Fig. 6a and Supplementary Note 5).

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

Collectively, our characterization of these AAV variants suggests several key points. First, within a diverse sequence family, there is room for both functional redundancy and the emergence of alternative 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, those variants found exclusively within the PCR pool library are more likely to be false positives.

The ability to confidently predict in vivo transduction from a pool of 18,000 nucleotide variants in R2 across multiple mice and Cre-lines is a substantial advance in the selection process and demonstrates the power of M-CREATE for the evolution of individual vectors.

An AAV9 variant that specifically transduces neurons. Using NGS, we re-investigated a 3-mer-s (s for substitution) PHP.B library generated by the prior CREATE methodology and that yielded AAV-PHP.eB[27] (Fig. 5a and Supplementary Note 6). We deep sequenced the libraries recovered from brain (using Cre-dependent PCR) and an R2 library from the livers of wild-type mice (processed via PCR for all capsid sequences regardless of Cre-mediated inversion) and identified 150–200 capsids enriched in brain tissue (Fig. 5b and Supplementary Fig. 7a,b).

Variants that were enriched in brain and underrepresented in liver show a significant bias towards certain amino acids such as G, D and E at position 1; G and S at position 2 (which includes the AAV-PHP.eB motif, DG); and S, N and P at position 9, 10 and 11 (Fig. 5c and Supplementary Fig. 7c; $P \leq 0.0001$, two-sided, two-proportion z-test, P values were corrected for multiple comparisons using Bonferroni correction). We clustered variants that were enriched in the brain according to their sequence similarities and ranked them by their underrepresentation in liver (represented by node size in clusters). A distinct family referred to as N emerged with the common motif SNP at positions 9–11 in the PHP.B backbone (Fig. 5d and Supplementary Fig. 7d).

The core variant of the N-family cluster, with the AQTLAVPFSNP motif, was highly abundant in R1 and R2 selections, had higher enrichment score in Vglut2 and Vgat brain tissues compared to GFAP, and was underrepresented in liver tissue (Fig. 5b and 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 and Supplementary Fig. 7e,f).

Thus, by re-examining the 3-mer-s library we identified several useful variants, including one with notable cell-type-specific tropism (Supplementary Note 7).

Investigation of capsid families beyond the C57BL/6J mouse strain. The enhanced CNS tropism of AAV-PHP.eB and AAV-PHP.B relative to AAV9 is absent in a subset of mouse strains. Their CNS transduction is highly efficient in C57BL/6J, FVB/NCrl, DBA/2 and SJL/J, with intermediate enhancement in 129S1/SvimJ, and no apparent enhancement over AAV9 in BALB/cJ and several additional strains[32–37]. This pattern holds for the two variants from the PHP.B family that we characterized further, AAV-PHP.V1 and AAV-PHP.N (Fig. 6a and Supplementary Table 3). These variants 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 with that following AAV9 and AAV-PHP.eB transduction (Supplementary Fig. 8a) however, suggesting mechanistic complexity.

Notably, M-CREATE revealed many non-PHP.B-like sequence families that enriched through selection for transduction of cells

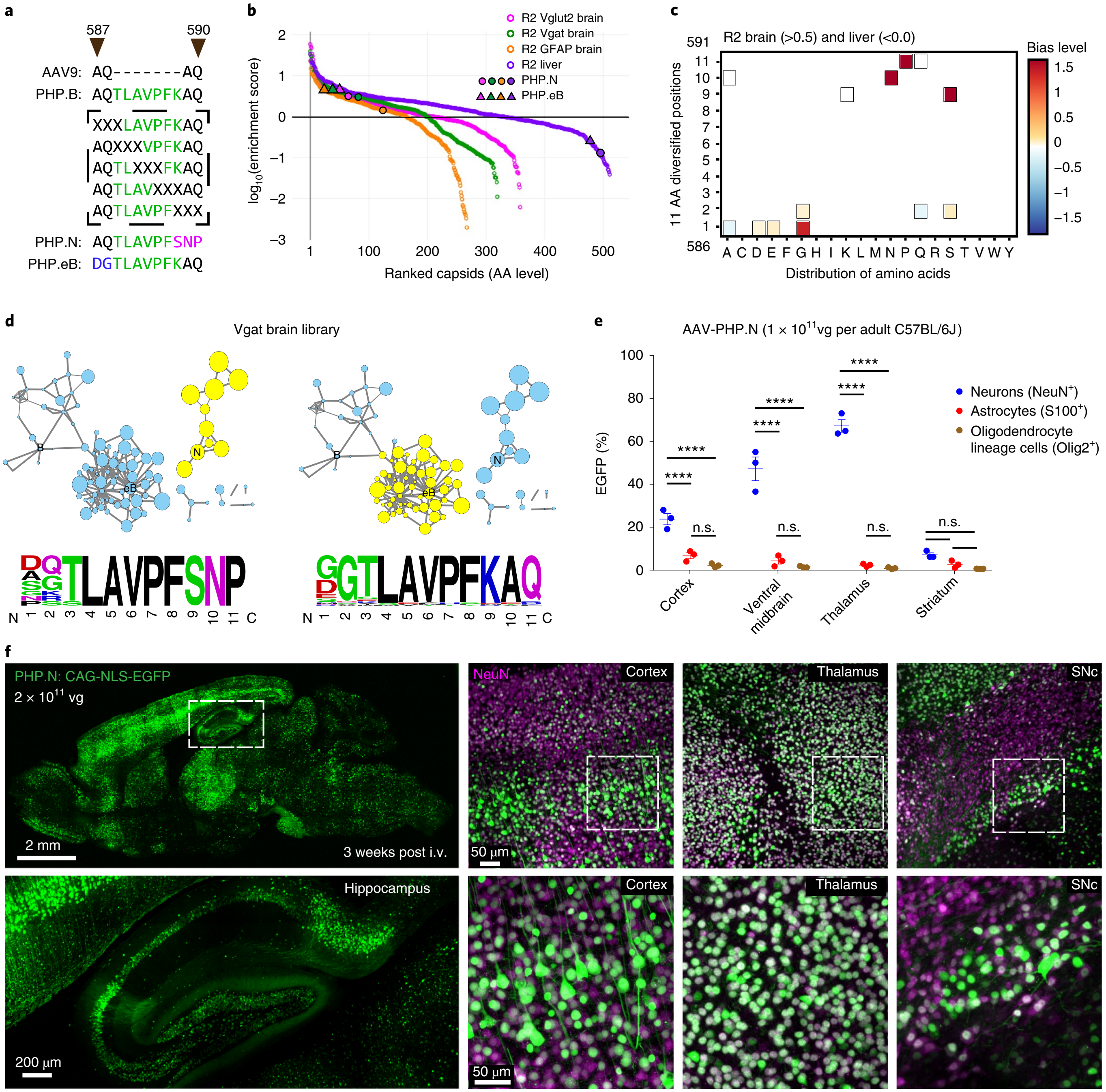


Fig. 5 | Recovery of AAV-PHP.B variants including one with high specificity for neurons. **a**, Design of the 3-mer-s PHP.B library with combinations of three AA diversification between AA 587–597 of AAV-PHP.B (corresponding to AA 587–590 of AAV9). Shared amion acid identity with the parent AAV-PHP.B (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 the amino acid 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 is mapped on the plot. **c**, Heat map represents the magnitude ($\log_2$(fold change)) of a given amino acid's relative enrichment or depletion at each position across the diversified region, only if statistical significance is reached on fold change (boxed if $P \leq 0.0001$, two-sided, two-proportion z-test, P 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 underrepresented in liver (<0.0) (32 amino acid sequences). **d**, Clustering analysis of enriched variants from Vgat brain library is shown. Node size represents the degree of depletion in liver. Thickness of edges (connecting lines) represents degree of relatedness between nodes. Two distinct families are highlighted in yellow and their corresponding amino acid frequency logos are shown below (amino acid size represents prevalence, and color encodes amino acid 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 i.v. 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 ×20 objective covering the entire regions). A two-way ANOVA with correction for multiple comparisons using Tukey's test gave adjusted P values reported as ****$P \leq 0.0001$, n.s. for $P > 0.05$, 95% CI. **f**, Transduction by ssAAV-PHP.N:CAG-NLS-EGFP ($n=2$, 2×10^{11} vg i.v. 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).

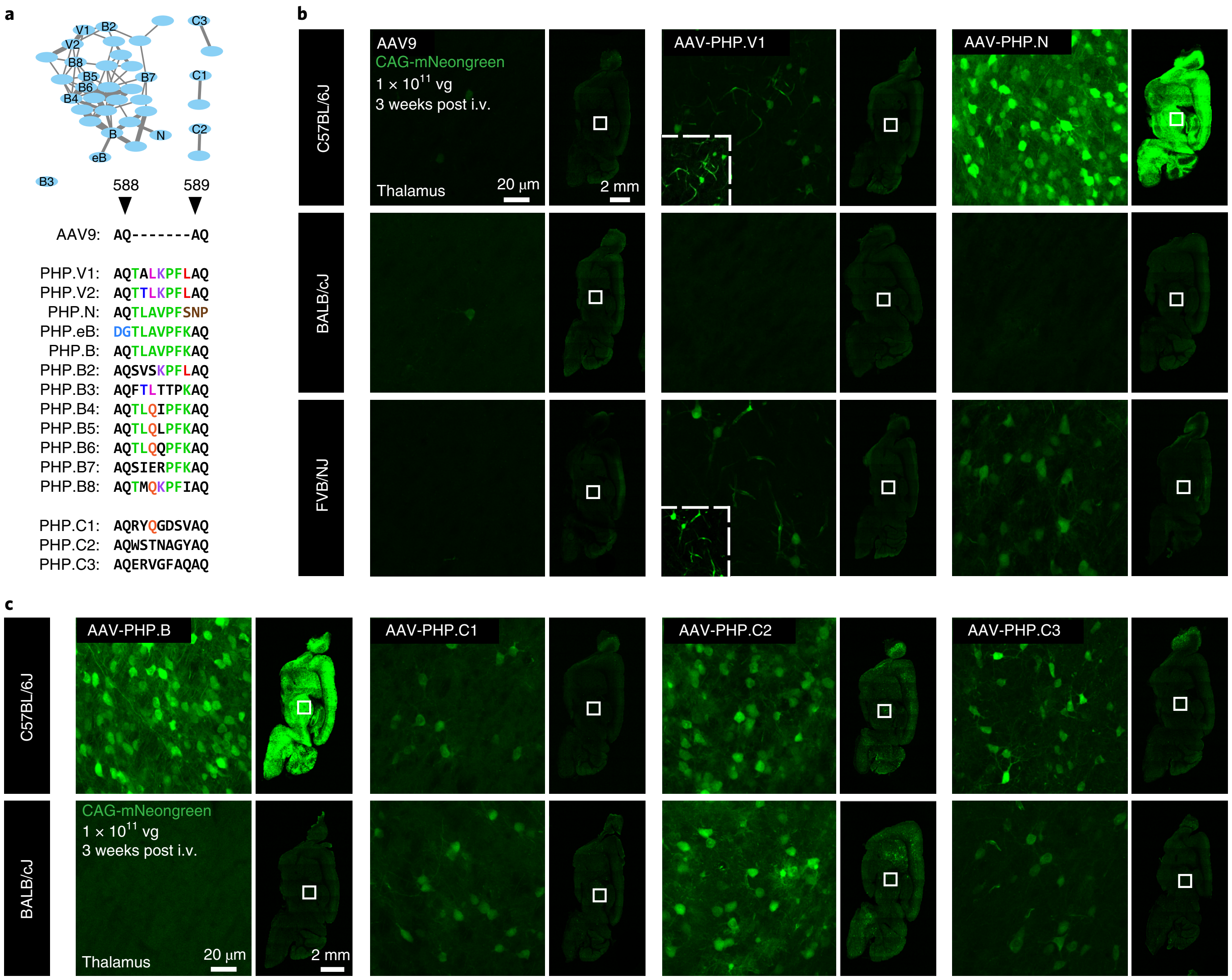


Fig. 6 | Tropism of variants from distinct families across mouse strains. a, Clustering analysis showing the brain-enriched sequence families of variants identified in prior studies (PHP.B-B3, PHP.eB) or in the current study (PHP.B4-B8, PHP.V1-2, PHP.C1-3). Thickness of edges (connecting lines) represents 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 the 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**, White box, thalamus (this is not the precise area that is zoomed-in on the figure to the left). All sagittal sections and thalamus regions were acquired under same image settings. 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 i.v. dose per 6- to 8-week-old adult mouse, 3 weeks of expression. The data reported in **b** and **c** are from one experiment 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 experiment for BALB/cJ, $n = 2$ per group).

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 and Supplementary Fig. 8b). Collectively, these studies suggest that M-CREATE is capable of finding capsid variants with diverse mechanisms of BBB entry that do not exhibit strain specificity.

Discussion

This work outlines the development and validation of the M-CREATE platform 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 useful 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 when delivered intravenously. 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 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[27].

We believe that the variants tested in vivo and their families will find broad application in neuroscience, including studies involving the BBB[38], neural circuits[39], neuropathologies[40] and therapeutics[41]. AAV-PHP.V1 or AAV-PHP.N are well-suited for studies requiring gene delivery for optogenetic or chemogenetic manipulations[42], or in rare monogenic disorders (targeting brain endothelial cells, for example GLUT1-deficiency syndrome, NLS1-microcephaly[40]; or targeting neurons, for example mucopolysaccharidosis type IIIC[22]).

The outcomes from our experiments using M-CREATE opens several promising lines of inquiry, such as the assessment of identified capsid families across species, the investigation of the mechanistic properties that underlie the ability to cross specific barriers (such as the BBB) or target specific cell populations and further evolution of the identified variants for improved efficiency and specificity. In addition, the datasets generated by M-CREATE could be used 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[43] offer hope in this regard. In summary, M-CREATE will serve as a next-generation capsid-selection platform that can open directions in vector engineering and potentially broaden the AAV toolbox for various applications in science and in therapeutics.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41592-020-0799-7.

Received: 8 April 2019; Accepted: 10 March 2020;
Published online: 20 April 2020

References

1. Wu, Z., Asokan, A. & Samulski, R. J. Adeno-associated virus serotypes: vector toolkit for human gene therapy. *Mol. Ther. J. Am. Soc. Gene Ther.* **14**, 316–327 (2006).
2. Naso, M. F., Tomkowicz, B., Perry, W. L. & Strohl, W. R. Adeno-associated virus (AAV) as a vector for gene therapy. *Biodrugs* **31**, 317–334 (2017).
3. Daya, S. & Berns, K. I. Gene therapy using adeno-associated virus vectors. *Clin. Microbiol. Rev.* **21**, 583–593 (2008).
4. Gaj, T., Epstein, B. E. & Schaffer, D. V. Genome engineering using adeno-associated virus: basic and clinical research applications. *Mol. Ther.* **24**, 458–464 (2016).
5. Deverman, B. E., Ravina, B. M., Bankiewicz, K. S., Paul, S. M. & Sah, D. W. Y. Gene therapy for neurological disorders: progress and prospects. *Nat. Rev. Drug Discov.* **17**, 767 (2018).
6. Sen, D. Improving clinical efficacy of adeno associated vectors by rational capsid bioengineering. *J. Biomed. Sci.* **21**, 103 (2014).
7. Lee, E. J., Guenther, C. M. & Suh, J. Adeno-associated virus (AAV) vectors: rational design strategies for capsid engineering. *Curr. Opin. Biomed. Eng.* **7**, 58–63 (2018).
8. Bartlett, J. S., Kleinschmidt, J., Boucher, R. C. & Samulski, R. J. Targeted adeno-associated virus vector transduction of nonpermissive cells mediated by a bispecific F(ab'gamma)2 antibody. *Nat. Biotechnol.* **17**, 181–186 (1999).
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. USA* **116**, 27053–27062 (2019).
10. Bedbrook, C. N., Deverman, B. E. & Gradinaru, V. Viral strategies for targeting the central and peripheral nervous systems. *Annu. Rev. Neurosci.* **41**, 323–348 (2018).
11. Kotterman, M. A. & Schaffer, D. V. Engineering adeno-associated viruses for clinical gene therapy. *Nat. Rev. Genet.* **15**, 445–451 (2014).
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).
13. Maheshri, N., Koerber, J. T., Kaspar, B. K. & Schaffer, D. V. Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. *Nat. Biotechnol.* **24**, 198–204 (2006).
14. Excoffon, K. J. D. A. et al. Directed evolution of adeno-associated virus to an infectious respiratory virus. *Proc. Natl Acad. Sci. USA* **106**, 3865–3870 (2009).
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).
17. Müller, O. J. et al. Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. *Nat. Biotechnol.* **21**, 1040–1046 (2003).
18. Ogden, P. J., Kelsic, E. D., Sinai, S. & Church, G. M. Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. *Science* **366**, 1139–1143 (2019).
19. Pekrun, K. et al. Using a barcoded AAV capsid library to select for clinically relevant gene therapy vectors. *JCI Insight* **4**, pii: 131610 (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, A. S. 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 (2015). discussion 225.
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).
23. Ojala, D. S. 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, D. G. R. et al. A designer AAV variant permits efficient retrograde access to projection neurons. *Neuron* **92**, 372–382 (2016).
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).
26. Deverman, B. E. et al. Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nat. Biotechnol.* **34**, 204–209 (2016).
27. Chan, K. Y. et al. Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. *Nat. Neurosci.* **20**, 1172–1179 (2017).
28. Lee, Y., Messing, A., Su, M. & Brenner, M. GFAP promoter elements required for region-specific and astrocyte-specific expression. *Glia* **56**, 481–493 (2008).
29. Kisanuki, Y. Y. et al. Tie2-Cre transgenic mice: a new model for endothelial cell-lineage analysis in vivo. *Dev. Biol.* **230**, 230–242 (2001).
30. de Leeuw, C. N. et al. rAAV-compatible MiniPromoters for restricted expression in the brain and eye. *Mol. Brain* **9**, 52 (2016).
31. Rincon, M. Y. 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).
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).
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).
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).
35. Challis, R. C. et al. Systemic AAV vectors for widespread and targeted gene delivery in rodents. *Nat. Protoc.* **14**, 379–414 (2019).
36. Batista, A. R. 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.* **31**, 90–102 (2019).
37. Hordeaux, J. et al. The neurotropic properties of AAV-PHP.B are limited to C57BL/6J mice. *Mol. Ther.* **26**, 664–668 (2018).
38. Sweeney, M. D., Zhao, Z., Montagne, A., Nelson, A. R. & Zlokovic, B. V. Blood–brain barrier: from physiology to disease and back. *Physiol. Rev.* **99**, 21–78 (2019).
39. Betley, J. N. & Sternson, S. M. Adeno-associated viral vectors for mapping, monitoring and manipulating neural circuits. *Hum. Gene Ther.* **22**, 669–677 (2011).
40. Sweeney, M. D., Sagare, A. P. & Zlokovic, B. V. Blood–brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nat. Rev. Neurol.* **14**, 133–150 (2018).
41. Lykken, E. A., Shyng, C., Edwards, R. J., Rozenberg, A. & Gray, S. J. Recent progress and considerations for AAV gene therapies targeting the central nervous system. *J. Neurodev. Disord.* **10**, 16 (2018).
42. Vlasov, K., Van Dort, C. J. & Solt, K. in *Methods in Enzymology* vol. 603 (eds. Eckenhoff, R. G. & Dmochowski, I. J.) 181–196 (Academic Press, 2018).
43. Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 96 (2018).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Methods

Plasmids. *Library generation.* The rAAV-ΔCap-in-cis-Lox2 plasmid (Supplementary Fig. 1a, plasmid available upon request at Caltech CLOVER Center) is a modification of the rAAV-ΔCap-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-ΔCap plasmid (Supplementary Fig. 1a, plasmid available upon request at Caltech CLOVER Center) was modified from the AAV2/9 REP-AAP plasmid[26] (Supplementary Note 8).

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

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] (Addgene ID: 99134), 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) (Supplementary Note 9).

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 G. Gao (University of Massachusetts Medical School) and scAAV:CAG-EGFP[44] from Addgene (Addgene ID:83279) (Supplementary Note 9).

AAV capsid library generation. *Round-1 AAV capsid DNA library.* Mutagenesis strategy. The randomized (21-base) heptamer insertion was designed using the NNK saturation mutagenesis strategy, involving degenerate primers containing mixed bases (Integrated DNA Technologies). N can be an A, C, G or T base, and K can be G or T. Using this strategy, we obtained combinations of all 20 amino acids at each position of the heptamer peptide using 33 codons, resulting in a theoretical library size of 1.28 billion at the level of amino acid combinations. The mutagenesis strategy for the 3-mer-s PHP.B library is described in our prior work[27].

Library cloning. The 480-bp AAV capsid fragment (450–592 amino acids) with the randomized heptamer insertion between amino acids 588 and 589 was generated by conventional PCR methods using the pCRII-9Cap-XE[26] template by Q5 Hot Start High-Fidelity 2X Master Mix (NEB; M0494S) with forward primer, XF and reverse primer, 7xMNN-588i (Supplementary Table 4 and Supplementary Note 10).

The rAAV-ΔCap-in-cis-Lox2 plasmid (6,960 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.

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).

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 of assembly reaction.

Quality control. See Supplementary Note 11.

Round-2 AAV capsid DNA library. PCR pool design. To maintain proportionate pooling, we mathematically determined the fraction of each sample or library that needs to be pooled based on an individual library's diversity (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 and 588-R2lib-R (Supplementary Table 4). Similar to R1 library generation, the PCR product was assembled into the rAAV-ΔCap-in-cis-Lox2 plasmid and the virus was produced (Supplementary Note 13).

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 (Supplementary Fig. 1e and Supplementary Note 12). This came to about 9,000 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 and 11-mer-588i (Supplementary Table 4) were used. In 11-mer-588i primer, 'XXXXXXMNNMNNMNNMNNMNNMNNMNNXXXXXX' was replaced with unique nucleotide sequence of a heptamer tissue recovered variant (7xMNN) along with modification of two adjacent codons flanking on either end of the heptamer insertion site (6xX), which are residues 587–588 'AQ' and residues 589–590 'AQ' on AAV9 capsid. Since the spike-in library has 11-mer or 33-base oligonucleotide mutated variants, we used the same primer design where 'XXXXXXMNNMNNMNNMNNMNNMNNMNNXXXXXX' was replaced with a specific nucleotide sequence of a 33-base oligonucleotide 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 (Supplementary Note 13).

AAV virus library production, purification and genome extraction. To prevent capsid mosaic formation of the 7-mer-i library in 293 T producer cells, we transfected only 10 ng of assembled library per 150-mm dish along with other required reagents for AAV vector production (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. The C57BL/6J (000664), Tek-Cre[29] (8863), SNAP25-Cre[45] (23525), GFAP-Cre[46] (012886), Syn1-Cre[47] (3966), and Ai14[48] (007908) mouse lines used in this study were purchased from the Jackson Laboratory (JAX). The i.v. injection of rAAVs was into the retro-orbital sinus of 6- to 8-week-old male or female mice. For testing the transduction phenotypes of novel rAAVs, 6- to 8-week-old, male C57BL/6J or Tek-Cre or Ai14 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 mice, 6- to 8-weeks-old, were i.v. administered with a viral vector dose of 2×10^{11} vg per mouse for the R1 selection, and a dose of 1×10^{12} vg per mouse for the R2 selection. The dose was determined on the basis of 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.

rAAV genome extraction from tissue. Optimization. See Supplementary Note 16.

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]. The extracted DNA by Trizol method was then treated with 3–6 µL of $10\,\mu g\,\mu L^{-1}$ RNase Cocktail Enzyme Mix (ThermoFisher Scientific; AM2286) to remove RNA. The mixture was also digested with SmaI restriction enzyme to improve rAAV genome recovery by PCR (Supplementary Fig. 1h). The treated mixture was then finally purified with a Zymo DNA Clean and Concentrator kit (D4033). From deep-sequencing data analysis, we observed that the amount of tissue processed was sufficient for rAAV genome recovery.

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 (Supplementary Fig. 1b). We used the primers 71F and CDF/R 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 (Supplementary Table 4).

Total rAAV genome recovery by PCR (Cre-independent). To recover all rAAV genomes from a tissue, we used the primers XF and 588-R2lib-R 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 (Supplementary Table 4).

Sample preparation for NGS. We processed the DNA library, the virus library and the tissue libraries following in vivo selection to add flow cell adaptors around the diversified 21-nucleotide insertion region (Supplementary Fig. 1b).

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 and

588i-lib-PCR1-R that are positioned around 50 bases from the randomized 21-nucleotide insertion on the capsid, and that contain the Read1 and Read2 flow cell sequences on the 5′ end (Supplementary Table 4 and Supplementary Note 17). 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 (Supplementary Note 18).

Preparation of rAAV tissue DNA library. The PCR-amplified rAAV DNA library from tissue (see sections: rAAV genome recovery by Cre-dependent PCR and total rAAV genome recovery by PCR (Cre-independent)) was further amplified with a 1:100 dilution of this DNA as a template to the primers 1527 and 1532 that are positioned around 50 bases from the randomized 21-nucleotide insertion on the capsid, and that contain the Read1 and Read2 sequences on the 5' end (see Supplementary Table 4). 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 ten-cycle reaction with the same temperature cycle as described above (for DNA and virus library preparation), and followed additional processing and validation before sequencing (Supplementary Note 18).

In vivo characterization of AAV vectors. *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-base oligonucleotide substitution (in case of 7-mer-i variants, the flanking amino acids from AAV9 capsid AA 587–588 'AQ' and AA 589–590 '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. The designed primers cover the entire fragment that is inserted into the linearized pUCmini-iCAP-PHP.B backbone. Hence these primers are simply self-annealed using PCR to synthesize double-stranded DNA fragment without the use of a template DNA. They are amplified by Q5 Hot Start High-Fidelity 2X Master Mix for 20 cycles of 98 °C for 10 s, 60 °C for 30 s and 72 °C for 10 s. This fragment was then assembled into the MscI/AgeI digested pUCmini-iCAP-PHP.B backbone by the Gibson assembly method. There is a second MscI site on the backbone; however, this was blocked by methylation. The assembled plasmids were then transformed into NEB Stable competent *E. coli* (New England Biolabs; C3040H), and colonies were selected on carbenicillin/ampicillin-LB agar plates. A list of primers used to clone AAV-PHP variants is provided (Supplementary Table 5).

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.

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^{11}–10×10^{11} vg with 3–4 weeks of in vivo expression times unless mentioned otherwise in the figures or legends (Supplementary Note 19).

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.

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 (Supplementary Note 20).

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[49] protocol (Supplementary Note 21).

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 (Supplementary Note 22). The acquired images were processed in the respective microscope software 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).

Tissue clearing. Brain hemispheres were cleared using iDISCO[50] method and tissues over 500 µm thickness were optically cleared using ScaleS4(0)[51] (Supplementary Note 23).

Tissue processing and imaging for quantification of rAAV transduction in vivo. For quantification of rAAV transduction, 6- to 8-week-old male mice were i.v. 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 were compared directly across groups were acquired and processed with the same microscope and settings (Supplementary Note 24).

In vitro characterization of AAV vectors. Human brain microvascular endothelial cells (HBMEC) (ScienCell Research Laboratories, cat. no. 1000) were cultured as per the instructions provided by the vendor. HBMEC were cultured from a frozen stock vial in fibronectin-coated T-75 flask (7,000–9,000 cells per cm^2 seeding density) using the endothelial cell medium (cat. no. 1001). The cells were subcultured in fibronectin-coated 48-well plates (0.95 cm^2 growth area) at the recommended seeding density and incubated at 37 °C for ~24–48 h until the cells were completely adherent with ~70–80% confluence. The viral vectors packaging pAAV:CAG-mNeongreen were added to the cell culture at a dose of either 1×10^8 or 1×10^{10} vg per well (3 wells per dose per vector). The medium was changed 24 h later, and the culture was assessed for fluorescence expression at 3 d post infection. Per vendor recommendation, the culture medium was changed every other day to maintain the cell culture.

Data analysis. *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 (Supplementary Note 25). The experimenter was not blinded during data analysis.

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 26).

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:

Enrichment score = $\log_{10}$((variant 1 RC in tissue library1 / sum of variants N RC in library1) / (variant 1 RC in virus library / sum of variants N RC in virus library))

To consistently represent library recovery between R1 and R2 selected variants, we estimated the enrichment score of the variants in R1 selection (Supplementary Note 27).

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

Standard score = (read count_i – mean) / s.d.

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. The s.d. is the s.d. 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)[52] was modeled in PyMOL.

Heat map generation. The relative amino acid distributions of the diversified regions are plotted as heat maps. The plots were generated using the Python Plotly plotting library. The heat map values were generated from custom scripts written in Python, using functions in the custom "pepars" Python package (see Supplementary Note 28).

Clustering analysis. Using custom scripts written in MATLAB (version R2017b; MathWorks), we determined the reverse Hamming distances representing the number of shared amino acids between two peptides. Cytoscape (version 3.7.1[53]) software was then used to cluster the variants. The amino acid frequency plot representing the highlighted cluster was created using Weblogo (Version 2.8.2)[54,55] (Supplementary Note 29).

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 versus known template frequency (NNK), and two-proportion z-test for two-library comparisons were performed. P values are corrected for multiple comparisons using a 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 with 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% 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 s.e.m. All experiments were validated in more than one independent trial unless otherwise noted.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The NGS datasets using the synthetic pool and PCR pool selection methods that are reported in this article are available under the SRA accession code PRJNA610987. 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, and viruses may be available for commonly packaged genomes. Other plasmids or viruses not available at Addgene may be requested from Caltech, CLOVER Center (http://clover.caltech.edu/). GenBank: AAV-PHP.V1: MT162422, AAV-PHP.V2: MT162423, AAV-PHP.N: MT162424, AAV- PHP.C1: MT162425, AAV-PHP.C2: MT162426, AAV-PHP.C3: MT162427, AAV-PHP.B4: MT162428, AAV-PHP.B5: MT162429, AAV-PHP.B6: MT162430, AAV-PHP.B7: MT162431 and AAV-PHP.B8: MT162432.

Code availability

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.

References

44. Paulk, N. K. et al. Bioengineered viral platform for intramuscular passive vaccine delivery to human skeletal muscle. *Mol. Ther. Methods Clin. Dev.* **10**, 144–155 (2018).
45. Harris, J. A. et al. Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. *Front. Neural Circuits* **8**, 76 (2014).
46. Garcia, A. D. R., Doan, N. B., Imura, T., Bush, T. G. & Sofroniew, M. V. GFAP-expressing progenitors are the principal source of constitutive neurogenesis in adult mouse forebrain. *Nat. Neurosci.* **7**, 1233–1241 (2004).
47. 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).
48. 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).
49. Choi, H. M. T. et al. Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).
50. Renier, N. et al. iDISCO: a simple, rapid method to immunolabel large tissue samples for volume imaging. *Cell* **159**, 896–910 (2014).
51. Hama, H. et al. ScaleS: an optical clearing palette for biological imaging. *Nat. Neurosci.* **18**, 1518–1529 (2015).
52. DiMattia, M. A. et al. Structural insight into the unique properties of adeno-associated virus serotype 9. *J. Virol.* **86**, 6947–6958 (2012).
53. Shannon, P. et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
54. Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).
55. Crooks, G. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

Acknowledgements

We thank the following alumni and current members of the Gradinaru group for their assistance in this study: K. Y. Chan and R. Challis for performing mouse injections, R. Hurt for performing preliminary characterization of vectors, Y. Lei for assistance with cloning, K. Beadle for vector production, E. Sullivan for tissue sectioning, E. Mackey for tissue sectioning and mouse colony management, N. Flytzanis and N. Goeden for their contributions towards histology, imaging, data analysis and manuscript preparation, P. Anguiano for administrative assistance, and the entire Gradinaru group for discussions. We thank L. V. Sibener at Stanford University for sharing the Matlab scripts used in amino acid clustering analysis. We thank the Biological Imaging Facility at Caltech (supported by Caltech Beckman Institute and the Arnold and Mabel Beckman Foundation). We also thank the Millard and Muriel Jacobs Genetics and Genomics Laboratory at Caltech; and Integrative Genomics Core at City of Hope for providing sequencing service. This work was primarily supported by grants from the National Institutes of Health (NIH) to V.G.: NIH Director's New Innovator DP2NS087949 and PECASE, NIH BRAIN R01MH117069, NIH Pioneer DP1OD025535 and SPARC 1OT2OD024899. Additional funding includes the Vallee Foundation (V.G.), the Moore Foundation (V.G.), the CZI Neurodegeneration Challenge Network (V.G.), and the NSF NeuroNex Technology Hub grant 1707316 (V.G.), the Heritage Medical Research Institute (V.G.) and the Beckman Institute for CLARITY, Optogenetics and Vector Engineering Research (CLOVER) for technology development and dissemination (V.G.).

Author contributions

S.R.K., B.E.D., T.F.M. and V.G. designed the experiments. S.R.K., B.E.D., X.C., T.F.M., Y.L., A.G., Q.H. and M.J.J. performed experiments. X.C. assisted with virus production and characterization of AAV-PHP variants in mice. Q.H. assisted with method optimization, cloning, virus production and tissue harvest. Y.L. assisted with method optimization and processed tissues for deep sequencing for 3-mer-s library. T.F.M. performed the clustering analysis, contributed to experiments related to NGS data validation, variant assessment across mice strains and amino acid bias heat map analysis. A.G. processed and imaged cleared brain hemisphere, and compiled the Supplementary Video 1 with input from S.R.K., and V.G. D.B., T.D. and P.H.E. built the software to process the NGS raw data for analysis with input from B.E.D., T.F.M., V.G. and S.R.K. M.J.J. performed the HCR experiments. X.D. produced structural models for AAV9 and contributed to the data analysis pipeline. S.R.K. prepared the figures with input from all authors. S.R.K., T.F.M., B.E.D. and V.G. wrote the manuscript with input from all authors. V.G. supervised all aspects of the work.

Competing interests

The California Institute of Technology has filed and licensed a patent application for the work described in this manuscript with S.R.K., B.E.D., and V.G. listed as inventors (Caltech disclosure reference no. CIT 8198).

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41592-020-0799-7.

Correspondence and requests for materials should be addressed to V.G.

Peer review information Nina Vogt was the primary editor on this article and managed its editorial process and peer review in collaboration with the rest of the editorial team nature research

Corresponding author(s): Gradinaru

Last updated by author(s): Jan 21, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
□	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
□	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
□	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
□	☒	A description of all covariates tested
□	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
□	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
□	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	□	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	□	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	□	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Python custom codes were used to generate the NGS data. The microscope images were generated in the respective microscope softwares Zen Black 2.3 SP1 (Zeiss), BZ-X Analyzer (Keyence), Lavision BioTec.
Data analysis	Python and Matlab based custom codes used for data analysis are available here: https://github.com/GradinaruLab/mCREATE. Adobe Photoshop CC 2018 Count Tool, Plotly, GraphPad PRISM 7.05, Matplotlib, Seaborn, MatlabR2017b, Weblogo (Version 2.8.2), Cytoscape v3.7.11, PyMOL2.2.0, Microsoft Excel 2016, Keyence Hybrid Cell Count software (BZ-H3C), ImageJ, Imaris (Bitplane), TeraStitcher, Zen Black 2.3 SP1 (Zeiss), BZ-X Analyzer (Keyence), Lavision BioTec. BioRender for illustrations.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

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/).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

A sample size of 3 mice or 3 biological replicates per experimental group was chosen for in vivo or in vitro validation of AAV variants given smaller range of variation that existed within each group. A sample size of 2 mice per Cre line for round-2 in vivo selection was chosen given the higher confidence in the selection design that produced minimal variation across each selection.

Data exclusions

No data was excluded from analysis.

Replication

The major claims in this study surrounding the tropism of newly identified capsids were validated in multiple trials, and sometimes with different experimental setups to validate the findings across various systems. The experiments differed in starting DNA and virus preparation, packaged genomes, mouse dosage, mouse strains, and performed by different co-authors. The number of such trials for PHP.N is 5, PHP.V1 is 8, PHP.V2 is 4, PHP.C1,C2,C3 is 2, where all attempts to validate the tropism of these variants were successful.

Randomization

Allocation of organisms to separate groups was random.

Blinding

The investigator was not blinded during analysis as the differences between control and variant groups were too obvious and blinding would not have helped in these experiments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☐	☒	Antibodies
☐	☒	Eukaryotic cell lines
☒	☐	Palaeontology
☐	☒	Animals and other organisms
☒	☐	Human research participants
☒	☐	Clinical data

Methods

n/a	Involved in the study	
☒	☐	ChIP-seq
☒	☐	Flow cytometry
☒	☐	MRI-based neuroimaging

Antibodies

Antibodies used

The primary antibodies used in this study were rabbit polyclonal anti-S100 (1:400, Abcam, ab868), rabbit monoclonal anti-Olig2 (1:400, Abcam, ab109186, EPR2673), rabbit monoclonal anti-NeuN (1:400, Abcam, ab177487, EPR12763), rabbit polyclonal anti-GLUT1 (1:400, Millipore Sigma, 07-1401), and chicken polyclonal anti-GFP (1:200, Aves Labs, GFP-1020). The secondary antibody used in this study were polyclonal AlexaFluor 647 AffiniPure donkey anti-rabbit IgG (H+L) (1:500, Jackson ImmunoResearch Lab, 711-605-152), and the goat anti-Chicken IgY (H+L) cross-adsorbed, Alexa Fluor 633 (1:200, ThermoFisher Scientific, A-21103).

Validation

All the antibodies used in this study were validated in prior publications as listed in the manufacturers' websites and on CiteAb.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

293T cells from ATCC, Human Brain Microvascular Endothelial Cells (HBMEC) from ScienCell Research Laboratories.

Authentication

No authentication was performed as the source was reliable. 293T cells are routinely used in the lab, from multiple stocks from ATCC with similar performance.

Mycoplasma contamination

Cell lines tested negative for mycoplasma contamination.

Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	C57BL/6J (000664), Tek-Cre25 (8863), SNAP25-Cre34 (23525), GFAP-Cre35 (012886), Syn1-Cre36 (3966), Ai1437 (007908), FVB/NJ (001800), and BALB/cJ (000651) mouse lines used in this study were purchased from the Jackson Laboratory (JAX). For in vivo selection, 6- to 8-week-old adult male and female mice were intravenously injected with viral libraries. Both genders were used for capsid selection to recover capsid variants with minimal gender bias. For testing the transduction phenotypes of novel rAAVs, 6- to 8-week-old C57BL/6J, Tek-Cre, Ai-14, BALB/cJ, or FVB/NJ adult male mice were randomly assigned. The gender was kept consistent during this validation to avoid potential discrepancies that exist in viral expression due to gender differences.
Wild animals	No wild animals were involved in this study.
Field-collected samples	The study did not involve any field-collected samples.
Ethics oversight	All animal procedures performed in this study were approved by the California Institute of Technology Institutional Animal Care and Use Commitee (IACUC).

Note that full information on the approval of the study protocol must also be provided in the manuscript.