Molecular Therapy
Original Article

In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the SVZ

David S. Ojala,[1] Sabrina Sun,[1] Jorge L. Santiago-Ortiz,[1] Mikhail G. Shapiro,[2,6] Philip A. Romero,[3] and David V. Schaffer[1,2,4,5]

[1]Department of Chemical and Biomolecular Engineering, University of California, Berkeley, Berkeley, CA, USA; [2]Department of Bioengineering, University of California, Berkeley, Berkeley, CA, USA; [3]Department of Biochemistry, University of Wisconsin, Madison, WI, USA; [4]The Helen Wills Neuroscience Institute, University of California, Berkeley, Berkeley, CA, USA; [5]Department of Molecular and Cell Biology, University of California, Berkeley, Berkeley, CA, USA

Directed evolution continues to expand the capabilities of complex biomolecules for a range of applications, such as adeno-associated virus vectors for gene therapy; however, advances in library design and selection strategies are key to develop variants that overcome barriers to clinical translation. To address this need, we applied structure-guided SCHEMA recombination of the multimeric adeno-associated virus (AAV) capsid to generate a highly diversified chimeric library with minimal structural disruption. A stringent in vivo Cre-dependent selection strategy was implemented to identify variants that transduce adult neural stem cells (NSCs) in the subventricular zone. A novel variant, SCH9, infected 60% of NSCs and mediated 24-fold higher GFP expression and a 12-fold greater transduction volume than AAV9. SCH9 utilizes both galactose and heparan sulfate as cell surface receptors and exhibits increased resistance to neutralizing antibodies. These results establish the SCHEMA library as a valuable tool for directed evolution and SCH9 as an effective gene delivery vector to investigate subventricular NSCs.

INTRODUCTION

Gene therapy based on engineered viruses has emerged as a highly promising and general approach to treat human disease. In particular, recombinant vehicles or vectors based on adeno-associated viruses (AAVs) have emerged as safe and effective gene therapy vectors.[1,2] Among their favorable properties are a strong clinical safety profile, ability to promote high transgene expression in both dividing and non-dividing cells, minimal risk of integration, and low immunogenicity. As a result, AAV gene therapies have been increasingly successful for the treatment of monogenic disorders in the liver and retina,[3,4] as well as for basic biological investigations in a broad range of organisms.

However, efficient application to a broad range of cell and tissue targets has been limited by delivery challenges that necessitate molecular engineering of this complex, multimeric biomolecule. These problems include pre-existing neutralizing antibodies against the AAV capsid, poor transduction of clinically and biologically important cell types, infection of off-target tissues, and the need for high vector doses. In particular, transduction of the CNS is difficult due to biological barriers that limit vector distribution.[5] It is unsurprising that natural AAV serotypes suffer such shortcomings given that the selective pressures driving natural AAV evolution are often at odds with our needs for enhanced biological tools and/or human therapeutics. However, the AAV capsid can be engineered to overcome these challenges through directed evolution, an iterative process of mutation and selection for improved function that mimics natural evolution but utilizes protein diversification strategies and selective pressures designed to address clinical needs. We have previously applied AAV-directed evolution to engineer novel AAV variants for enhanced in vivo gene delivery to the CNS[6–10] and other tissues.[11–14] However, intricate multimeric structures, such as the AAV capsid, pose protein engineering challenges, including both library generation and phenotypic selection.

Optimization of directed evolution begins with the design of a library. In particular, the goal of AAV library generation is to create a large, diverse pool of capsid variant sequences for subsequent selection, while minimizing the potential for disruption of the core virus functions necessary for packaging, transduction, and gene expression. In general, the majority of mutations introduced via random mutagenesis techniques are deleterious and lead to non-functional proteins.[15] An alternative method, DNA shuffling, recombines naturally occurring, homologous sequences into a library of chimeras in which sequence diversity has been exchanged between functional parent sequences.[16] By shuffling functional sequences, a far greater number of mutations can be introduced relative to random mutagenesis

Received 18 June 2017; accepted 3 September 2017;
https://doi.org/10.1016/j.ymthe.2017.09.006.

[6]Present address: Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA

Correspondence: David V. Schaffer, Department of Chemical and Biomolecular Engineering, University of California, Berkeley, 274 Stanley Hall, Berkeley, CA 94720-1462, USA.
E-mail: schaffer@berkeley.edu

(e.g., error-prone PCR) techniques. However, it can be difficult to rationally predict the crossover locations among a family of proteins that will be least disruptive to folding of chimeric proteins, particularly in multimeric protein and protein-nucleic acid structures as complex as ribosomes, transcriptional machinery, or viral capsids. Moreover, traditional DNA shuffling methods, which typically involve random nuclease digestion of the target sequences followed by homology-based reassembly, are biased to place crossovers in regions of high homology.[17]

SCHEMA is a computational method to predict the optimal crossover points for DNA shuffling of chimeric proteins.[18] This approach uses a contact map representation of protein structures and calculates the number of residue-residue contacts that are disrupted when homologous proteins are recombined. Libraries of chimeric proteins can be designed by balancing the average SCHEMA disruption with the number of mutations introduced to identify a set of crossovers that maximizes the library's overall functional diversity. SCHEMA has been applied to a range of functionally diverse proteins,[19–22] but importantly the extension of SCHEMA to multimeric proteins, which could include cellular transcriptional complexes, translational machinery, self-assembling structural proteins, viruses, and other multiprotein assemblies, affords opportunities to engineer proteins with modular structures and/or complex regulatory domains.[23,24] SCHEMA can be used to design libraries that are high in theoretical diversity, and previous studies have characterized the properties of 10^1–10^3 individual library members. Building upon these capabilities, we harnessed AAV-directed evolution to screen millions of capsid chimeras for improved function. One recent study used SCHEMA to generate and characterize the in vitro properties of 17 individual chimeras of AAV2 and AAV4, each containing a single crossover event.[25] Here, we construct a SCHEMA AAV capsid library comprised of six parent serotypes (AAV2, AAV4, AAV5, AAV6, AAV8, and AAV9) and seven crossover positions, yielding a total library size of more than 1.6 million variants. Moreover, we developed a Cre-dependent selection strategy to screen the entire library and drive convergence to chimeric AAVs that infect adult neural stem cells (NSCs) of the subventricular zone (SVZ), a cell type of clinical importance that is poorly transduced by natural AAV serotypes.

The SVZ of the lateral ventricles is the largest germinal region in the adult mammalian brain.[26] Resident adult NSCs of the SVZ produce new neurons and oligodendrocytes throughout life[27,28] and are therefore an attractive experimental model to study mechanisms of adult neurogenesis and stem cell response to brain injury and disease. Whereas in vitro NSC culture offers opportunities for high-throughput screening of small-molecule libraries that modulate cell proliferation, differentiation, and toxicity,[29,30] cell culture does not fully recapitulate the rich signaling environment of the neurogenic niche. In vivo manipulation of NSCs is thus key to gaining a deeper understanding of the regulatory mechanisms that contribute to NSC fate decisions and to unlocking the potential of endogenous NSCs for therapies. However, gene delivery to the SVZ remains challenging, and no single strategy has enabled efficient delivery while limiting disruption of the SVZ microenvironment.[31,32]

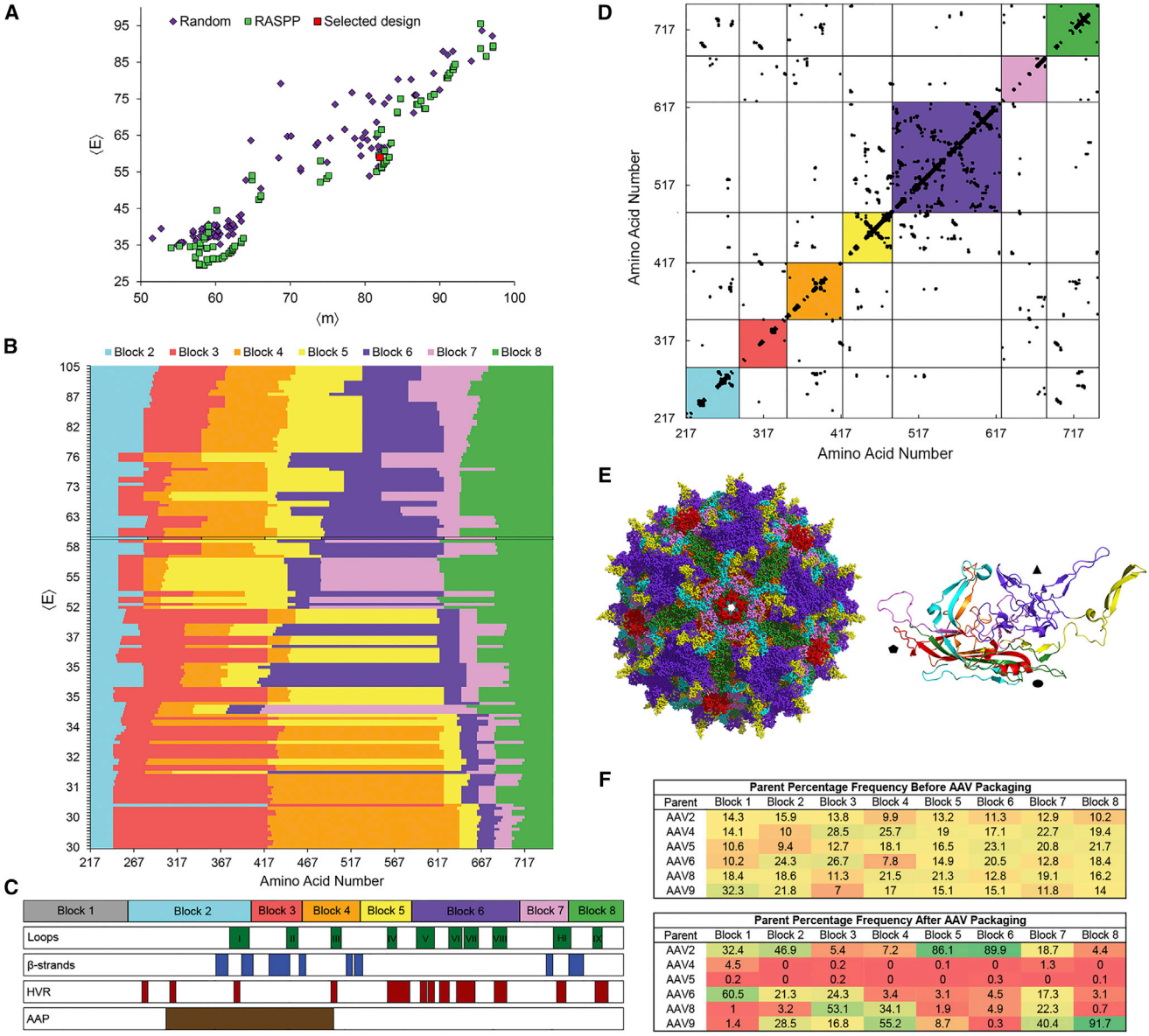
Adult NSCs of the SVZ therefore represent an ideal target to evaluate the efficacy of the SCHEMA AAV library and in vivo Cre-dependent selection strategy. Application of these advanced directed evolution methods resulted in selection of SCH9, a SCHEMA AAV variant that efficiently transduces NSCs throughout the entire SVZ in both hemispheres after a unilateral injection into the lateral ventricle.

RESULTS

SCHEMA-Guided Design of a Chimeric AAV Library

We designed a chimeric AAV library that recombines six natural serotypes (AAV2, 4, 5, 6, 8, and 9) that represent multiple phylogenetic clades,[33] have diverse receptor binding properties,[1] and have enjoyed some success in the clinic.[2] We used the capsid crystal structures to calculate contacting intra- and inter-subunit residue positions, wherein a contact was defined as two residues containing nonhydrogen atoms within a spatial cutoff of 4.5 Å. The final contact map contained residue pairs that were contacting in at least 50% of these six parent structures. To achieve high library diversity, we aimed to design a library containing six crossovers within the crystallized region of the capsid and a seventh in the uncrystallized VP1 region (amino acids 1–216) at position 128 based on a previous example of successful recombination at that location.[34] A library containing eight capsid protein blocks from six parent serotypes yields a theoretical library diversity of more than 1.6 million (6^8) chimeric variants. We additionally modified the SCHEMA scoring function to search for crossover locations that were amenable to combinatorial golden gate assembly for library construction, which requires 4-nt stretches that are conserved across all AAV parent sequences. In order to increase the number of possible crossovers sites, and thereby probe a larger sequence space in silico, we included 4-nt stretches that could be silently mutated during library assembly to be identical in all parent sequences.

After specifying these design parameters, we applied the RASPP method (Recombination as a Shortest Path Problem)[35] to rapidly identify 160 of the least disruptive library designs (sets of seven crossover positions) over a range of mutation levels. For each of these designs, the average library disruption score $\langle E \rangle$ and number of amino acid mutations introduced $\langle m \rangle$ relative to the closest parent serotype were calculated (Figure 1A), and the crossover locations of all RASPP designs are presented in Figure 1B. A final design with an average disruption score $\langle E \rangle$ of 59 and average number of mutations $\langle m \rangle$ of 82 per subunit in the crystallized region of the capsid (Figures 1A–1C) was chosen for several reasons. First, this design was in a cluster of RASPP libraries (Figure 1A) that represented a relative minimum in $\langle E \rangle$ at high mutation levels. Second, the selected design shuffled key capsid structural features, which include surface-exposed loops and hypervariable regions that represent the most divergent regions in the evolution of natural AAV serotypes (Figure 1C). Recombination within these contact-rich regions results in greater disruption, but is also more likely to generate AAV



Parent Percentage Frequency Before AAV Packaging

Parent	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8
AAV2	14.3	15.9	13.8	9.9	13.2	11.3	12.9	10.2
AAV4	14.1	10	28.5	25.7	19	17.1	22.7	19.4
AAV5	10.6	9.4	12.7	18.1	16.5	23.1	20.8	21.7
AAV6	10.2	24.3	26.7	7.8	14.9	20.5	12.8	18.4
AAV8	18.4	18.6	11.3	21.5	21.3	12.8	19.1	16.2
AAV9	32.3	21.8	7	17	15.1	15.1	11.8	14

Parent Percentage Frequency After AAV Packaging

Parent	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8
AAV2	32.4	46.9	5.4	7.2	86.1	89.9	18.7	4.4
AAV4	4.5	0	0.2	0	0.1	0	1.3	0
AAV5	0.2	0	0.2	0	0	0.3	0	0.1
AAV6	60.5	21.3	24.3	3.4	3.1	4.5	17.3	3.1
AAV8	1	3.2	53.1	34.1	1.9	4.9	22.3	0.7
AAV9	1.4	28.5	16.8	55.2	8.7	0.3	40.4	91.7

Figure 1. SCHEMA AAV Library Design

(A) The RASPP optimization algorithm generates library designs that are lower in ⟨*E*⟩ at various mutation levels. The RASPP library design selected for construction is indicated in red. (B) RASPP library designs over a range of ⟨*E*⟩ levels. The library selected for construction is indicated by a black border. Block 1 is omitted because it lies outside of the crystallized region of the capsid for which the SCHEMA analysis can be conducted. Parameters for library design were a minimum block length of 20 amino acids and a maximum length of 250 amino acids. (C) Schematic of *cap* crossover positions in the AAV library design selected for construction. An alignment of capsid loops, β strands that form the anti-parallel β-barrel motif, hypervariable regions (HVR), and the assembly-activating protein (AAP) are provided to indicate known structure-function relationships in the AAV capsid. (D) Protein contact map of the selected library design. All possible residue-residue contacts are displayed as black dots. The colored squares represent the sequence blocks that are shuffled. Contacts retained within the colored squares are preserved during recombination, while contacts outside of these squares may be broken depending on the identity of the parent sequences at each block. (E) Three-dimensional models of the selected capsid design. The shuffled blocks are represented by the corresponding colors used in (B)–(D) and mapped onto the AAV2 crystal structure (PDB: 1LP3) in PyMOL. The full biological assembly and a single asymmetric subunit with shapes indicating the axes of symmetry are shown. (F) The percentage frequency of each AAV parent before and after viral packaging of the assembled SCHEMA library are presented as a heatmap.

chimeras with new and interesting functions. For example, significantly lower disruption scores could be achieved by combining blocks 5 and 6, but doing so would generate capsids with surface-exposed loop regions derived from a single parent sequence. Finally, this set of crossover positions was selected because it provided a relatively even distribution of block sizes. We programmed RASPP to consider

a range of permissible block sizes from 20 to 250 amino acids. The majority of the lowest ⟨*E*⟩ designs contained two long blocks (>175 amino acids for blocks 3 and 4) followed by a series of short blocks (<30 amino acids for blocks 5–7) (Figure 1B). In contrast, our chosen set of crossover positions (Figure 1C) offers a more even distribution of block sizes, ensuring shuffling throughout the capsid as opposed to confining crossovers within a few regions that are of limited diversity in the parent sequences.

The selected library design was assembled by combinatorial golden gate cloning,[36] electroporated into *E. coli* to yield over 5×10^6 transformants, and packaged into AAV virions. The frequency of parent serotypes at each block position was analyzed by deep sequencing before and after viral packaging (Figure 1F). We found that each parent serotype sequence was well represented and distributed at each block location prior to viral packaging, but packaging presumably imposed a significant selective pressure for stable capsids and thereby resulted in substantial changes in library composition. For example, the frequency of AAV4 and AAV5 decreased by an average of 348- and 372-fold, respectively, across the packaged library, likely because of the low average amino acid sequence identity (AAV4: 60%, AAV5: 65%) of these serotypes with the other AAV parents used for library assembly. Changes in library composition upon packaging were also reflected in the decrease in the average disruption score ⟨*E*⟩ per crystallized subunit from 59 to 4 and in the average number of mutations ⟨*m*⟩ from 82 to 28. In agreement with prior applications of SCHEMA,[19,37] lower ⟨*E*⟩ chimeras were thus heavily enriched in the library. Interestingly, we also observed a preference for AAV2 at blocks 5 and 6 and AAV9 at block 8, trends that could be considered in the future to guide rational capsid engineering.

A Cre-Dependent Selection Strategy for AAV-Directed Evolution

Library diversification generates millions of mutant capsids, each a possible solution to a therapeutic gene delivery challenge. It is therefore crucial to design a stringent directed evolution strategy that can rapidly drive convergence from millions of variants to optima in a fitness landscape. NSCs are an especially difficult target given that they are rare cells that reside in a small niche within a complex tissue protected by biological barriers, including the ependymal cell layer lining the ventricle and the blood-brain barrier. NSCs can be isolated and cultured as neurospheres, but in vitro culture poorly recapitulates the spatial organization of the three-dimensional niche, its cell-cell interactions, the extracellular environment, and access to the cerebrospinal fluid (CSF) and vasculature. It would thus be advantageous to conduct evolution in vivo. However, in vivo-directed evolution strategies typically rely on extraction of genomic DNA from whole tissue and PCR to amplify the target gene, methods that are not selective for any particular cell type and may result in false positives that are enriched through transduction of cells that form biological barriers (e.g., ependymal and endothelial cells). To specifically target NSCs, we designed an in vivo Cre-dependent directed evolution and selection strategy to drive positive selection of AAV variants that infect NSCs in the SVZ. A conceptually analogous but distinct Cre-dependent system was reported during the course of this study.[38]

More than 300 transgenic mice that drive Cre expression under the control of a cell-type-specific promoter have been developed.[39] We harnessed the cell-type specificity of Cre expression to mediate selective recovery of the AAV *cap* gene by flanking the *cap* gene with a pair of loxP sites. AAV infection of a Cre-expressing cell followed by second-strand AAV genome synthesis leads to the inversion of the floxed *cap*, and PCR primers that serve as a forward and reverse pair only in the inverted gene template are used to selectively recover the Cre-inverted *cap* genes from the brain tissue (Figures 2A and 2B). Mutant loxP sites lox66 and lox71[40] were utilized to drive the equilibrium of Cre recombination toward unidirectional inversion. We initially attempted to insert the loxP sites in the 3′ UTR of *cap*, where they flanked short stuffer sequences containing the target sequence for the reverse primer used for Cre-dependent recovery. We found that recombination occurred at low levels during bacterial plasmid propagation, even in Sure2 recombinase-deficient *E. coli* (Figure S1). To prevent this undesired recovery of inverted *cap* during in vivo selections, we repositioned the loxP sites to flank *cap* such that artifactual inversion during bacterial propagation of the vector plasmid library would result in an inverted *cap* sequence that does not encode viral proteins and thus would not subsequently package in 293 cells, a provision not included in an alternate design.[38] Note that insertion of loxP sites flanking the *cap* gene alters the reading frame of the *rep* gene. The translation initiation codons of *rep* were thus removed, the viral promoter that drives *cap* expression was maintained (Figure 2A), and *rep* was instead supplied in *trans* for viral packaging by transient transfection of a separate *rep*-encoding helper. These modifications to the viral packaging plasmids resulted in a high AAV viral genomic yield as quantified by qPCR (Figure 2C).

Adult NSCs in the SVZ express glial markers including glial fibrillary acidic protein (GFAP),[28] glutamate aspartate transporter (GLAST),[41] and brain lipid-binding protein (BLBP).[42] To select for adult NSC transduction, we utilized the GFAP-Cre 73.12 mouse line in which Cre recombinase expression is controlled by the mouse GFAP promoter. Cre expression is observed in adult GFAP-expressing NSCs and mature astrocytes.[43] Although Cre is expressed in astrocytes in addition to NSCs, the intracerebroventricular (i.c.v.) route of administration results in preferential transduction of the SVZ where the NSCs reside, and GFAP serves as an important marker of NSC identity.[28] To validate Cre-dependent recovery of *cap*, we delivered AAV libraries containing floxed *cap* genes (pSub2FloxCap) to GFAP-Cre 73.12 or C57BL/6J control mice through an i.c.v. injection. Inverted *cap* could only be amplified from brain tissue of mice expressing Cre, while non-inverted *cap* was present in both groups (Figure 2D). For Cre recombination to occur, the AAV genome must be in double-stranded form, as required for expression of a therapeutic transgene. It is therefore likely that the non-inverted pool of *cap* genes amplified from the GFAP-Cre 73.12 mice represents capsids that failed to infect GFAP-positive cells, were defective in some aspect of the viral life cycle (e.g., capsid uncoating, endosomal escape), or did not complete second-strand synthesis. The Cre-dependent selection strategy thus exclusively recovers capsid variants that complete all steps necessary for robust transgene expression in the target cell type.

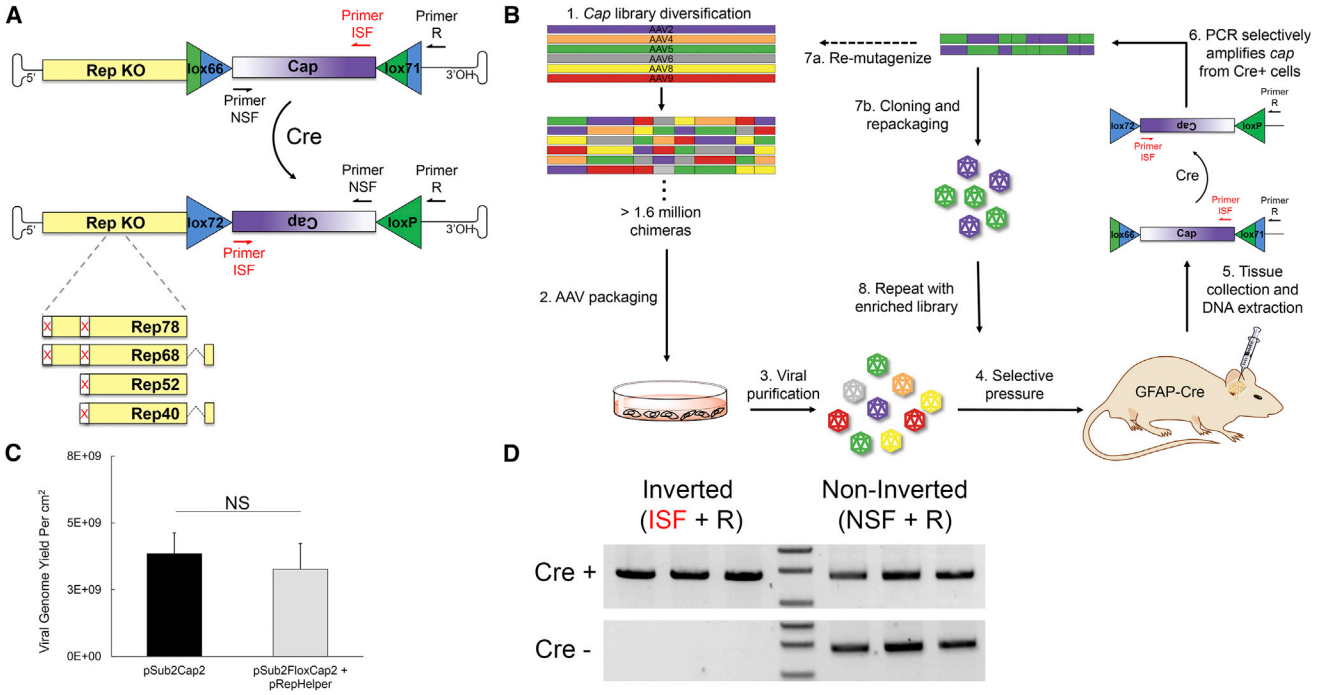


Figure 2. Design of a Cre-Dependent Selection Strategy for AAV-Directed Evolution

(A) Modifications to the AAV viral genome enable Cre-dependent selection. Mutant loxP sites lox66 and lox71 flank the *cap* gene. Upon Cre inversion of *cap*, primer ISF changes from a reverse primer to a forward primer, and primer NSF changes from a forward primer to a reverse primer. Primer R remains a reverse primer. Amplification of inverted *cap* is achieved using the primer pair ISF and R, whereas primers NSF and R selectively amplify non-inverted *cap*. The translation initiation codons for the Rep open reading frames are knocked out. (B) Cre-dependent AAV selection strategy to target adult neural stem cells. AAV libraries were generated through DNA shuffling or other methods, packaged into AAV virions, and administered to GFAP-Cre 73.12 mice through a unilateral i.c.v. injection. Three weeks later, genomic DNA was extracted from brain tissue of the contralateral hemisphere, and inverted *cap* variants were selectively amplified using the primers ISF and R for the next round of selection. (C) Viral genomic titers are not significantly different when *rep* is supplied in *trans* to package wild-type AAV2. Data are presented as mean ± SEM; n = 3. (D) Selective amplification of inverted *cap* from GFAP-Cre mice. NS, not significant.

In Vivo Library Selections Converge on a Dominant SCHEMA AAV Variant

After validating Cre-dependent recovery of *cap*, we initiated in vivo selections using an equimolar mixture of six AAV libraries, each containing 10^6–10^7 unique variants: (1) the new SCHEMA AAV; (2) error-prone AAV9; (3) ancestral AAV;[44] (4) shuffled AAV generated by DNase I digestion and reassembly of AAV1, 2, 4, 5, 6, 8, and 9;[12] (5) error-prone AAV2;[45] and (6) AAV2 7-mer peptide insertion at amino acid 588.[46] Libraries 3–6 have previously yielded highly infectious clones in our directed evolution selections[9–12,44] and provide evolutionary competition for the SCHEMA library. The libraries were combined and injected via i.c.v. administration into the right lateral ventricle of adult GFAP-Cre mice (n = 3) to transduce NSCs throughout the entire SVZ in both hemispheres. In contrast, direct SVZ injection is more disruptive to the local tissue and could require multiple injections to cover the same tissue volume.

Three weeks after injection the contralateral brain hemisphere was harvested, genomic DNA was extracted, and Cre-recombined AAV *cap* variants were recovered from GFAP-expressing cells by PCR. The contralateral hemisphere was harvested to ensure that *cap* variants were not recovered from transduction associated with the injection tract through the cortex superior to the lateral ventricle. After three rounds of in vivo selection, Sanger sequencing analysis of 24 clones revealed convergence on two variants originating from the SCHEMA library. SCH9 (chimera 6, 9, 8, 9, 9, 2, 9, 9; $\langle E \rangle$ 9, $\langle m \rangle$ 49) represented 54% of the clones recovered, whereas SCH2 (chimera 6, 9, 8, 9, 2, 2, 9, 9; $\langle E \rangle$ 4, $\langle m \rangle$ 37) represented 33%. The remaining clones were derived from the AAV2 7-mer insertion (8%) and ancestral libraries (4%). SCH9 differs from the closest parent, AAV9, by 58 total mutations (92% amino acid identity). 49 of these ($\langle m \rangle$) are in the crystallized region of the capsid, and 9 are in the uncrystallized region. An amino acid alignment of sequences of SCH9, SCH2, and multiple parent AAV serotypes are presented in Figures S2 and S3, respectively. The two SCHEMA variants differ only at block 5, resulting in a difference of 18 amino acids. A model of the three-dimensional structure of SCH9 shows AAV9 at loop VR-IV on the capsid surface, AAV2 at loops V–VIII, and AAV8 at the 5-fold pore structure (Figure 3). Based on these intriguing features, and its dominance of the selected pool, we chose to focus primarily on in vivo characterization of SCH9.

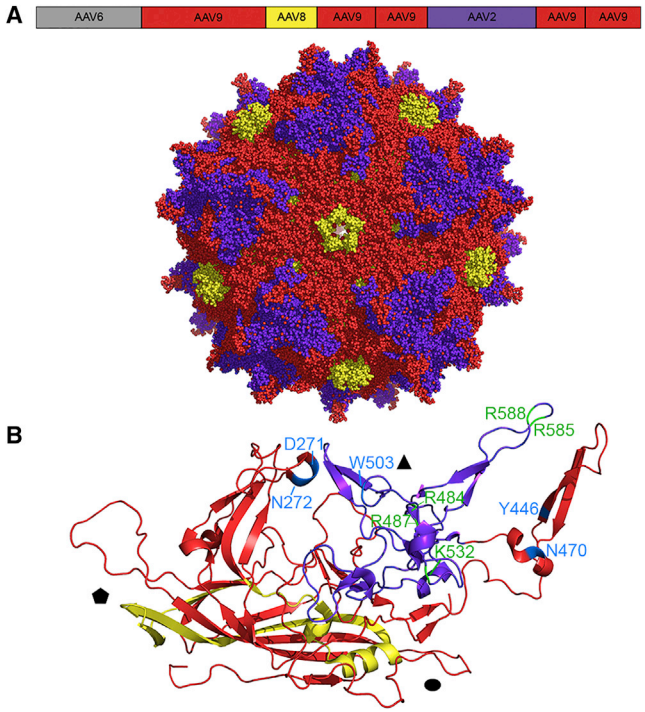


Figure 3. Three-Dimensional Models of the SCH9 Capsid

Block boundaries are represented on a schematic of the *cap* gene. Each parent serotype is represented by a different color (AAV9: red, AAV8: yellow, AAV2: purple) and mapped onto the AAV9 crystal structure (PDB: 3UX1) in PyMOL. (A and B) The full biological assembly (A) and a single asymmetric subunit (B) with shapes indicating the axes of symmetry are shown. Residues involved in heparin or galactose binding are annotated in green and blue, respectively, on the individual subunit.

SCH9 Efficiently Transduces Adult NSCs in the SVZ of Adult Mice

To assess the transduction profile of SCH9 in the SVZ, we successfully packaged (recombinant AAV packaging yields are reported in Figure S4) and delivered recombinant AAV (rAAV) carrying a self-complementary CAG-GFP cassette to the right lateral ventricle of adult C57BL/6J mice. SCH9 was benchmarked against AAV9 because of its broad use in the CNS and capacity to transduce the brain parenchyma from the CSF after intrathecal injection.[47,48] Moreover, of the natural serotypes, AAV9 is the most closely related sequence to SCH9.

Transduction of the contralateral hemisphere was analyzed 4 weeks after injection, and GFP expression was primarily associated with the region surrounding the ventricle, with greatest intensity in the SVZ (Figure 4A). Transduction efficiency was evaluated by both the intensity of GFP expression and the total volume of the SVZ that was positive for GFP. The integrated GFP fluorescence intensity for SCH9 was 24-fold higher, and GFP was expressed in a 12-fold greater SVZ transduction volume, compared to AAV9 (Figures 4B and 4C).

The SVZ is composed of multiple cell types including ependymal cells, adult NSCs (B cells), transit amplifying cells (type C cells), neuroblasts (type A cells), and mature astrocytes.[26] To evaluate the efficiency of NSC transduction in the SVZ, we first assessed molecular markers that are selectively expressed within NSCs. Although most markers are expressed in multiple cell types in the SVZ, reflecting the continuum of gene expression during lineage progression, vascular cell adhesion molecule 1 (VCAM1) specifically localizes to the endfeet of NSCs that contact the ventricle.[49] As an initial characterization, we found GFP/GFAP/VCAM1-positive adult NSCs that were transduced by SCH9 in the SVZ (Figure 4D).

Recombinant AAV genomes are maintained episomally and are progressively lost during the cell divisions characteristic of adult neurogenesis in the SVZ. Specifically, lineage progression from a NSC to an olfactory bulb interneuron involves over seven cell divisions.[50] As a result of the accompanying AAV genome dilution, at late time points after injection the majority of cells that continue to express transgene are slowly dividing NSCs or post-mitotic cells. Moreover, prior studies using integrating retroviral vectors indicate that the time required for neuroblasts to traverse the rostral migratory stream to the olfactory bulb is 9 days, and that all transit amplifying cells and neuroblasts present in the SVZ at the time of injection differentiate and/or migrate to the olfactory bulb and established dendrites by 30 days post-injection.[51,52] These results indicate that neuroblasts present in the rostral migratory stream at late time points after injection are derived from NSCs, a conclusion that was previously used to establish lentiviral or non-viral transduction of NSCs in the SVZ.[53,54] We designed a similar lineage analysis strategy to determine the number of migrating neuroblasts expressing tdTomato 30 days post-injection as an indication of NSC transduction. Recombinant SCH9 or AAV9 encoding Cre recombinase was injected into the right lateral ventricle of adult Ai9 floxed STOP tdTomato mice,[55] within which Cre activity would result in tdTomato expression in transduced cells and their progeny. The majority (injected right hemisphere 83.2% ± 3.6%, left hemisphere 50.3% ± 4.4%) of neuroblasts were positive for tdTomato in the rostral migratory stream 30 days post-injection of SCH9-expressing Cre (Figures 4E and 4G), exceeding AAV9 transduction by more than 4-fold. Furthermore, large numbers of tdTomato-positive neuroblasts were observed migrating radially in the olfactory bulb and adopting the morphology of granule cell neurons (Figure 4F).

To further characterize NSC transduction, we administered the thymidine analog BrdU (5-bromo-2′-deoxyuridine) to label dividing cells in the SVZ prior to injection of single-stranded SCH9 CAG-Cre. After a washout period of 2 weeks, we analyzed colocalization of tdTomato expression with BrdU incorporation into GFAP$^+$ NSCs (Figure 4H). The percentage of adult NSCs (GFAP$^+$, BrdU$^+$, doublecortin$^-$), transit amplifying cells (GFAP$^-$, BrdU$^+$, doublecortin$^-$), and neuroblasts (GFAP$^-$, BrdU$^+$, doublecortin$^+$) expressing tdTomato in the SVZ were quantified (Figure 4I). Approximately 60% of NSCs were transduced in both hemispheres, supporting the efficacy of SCH9 for gene delivery to NSCs using both single-stranded and self-complementary genome formats.

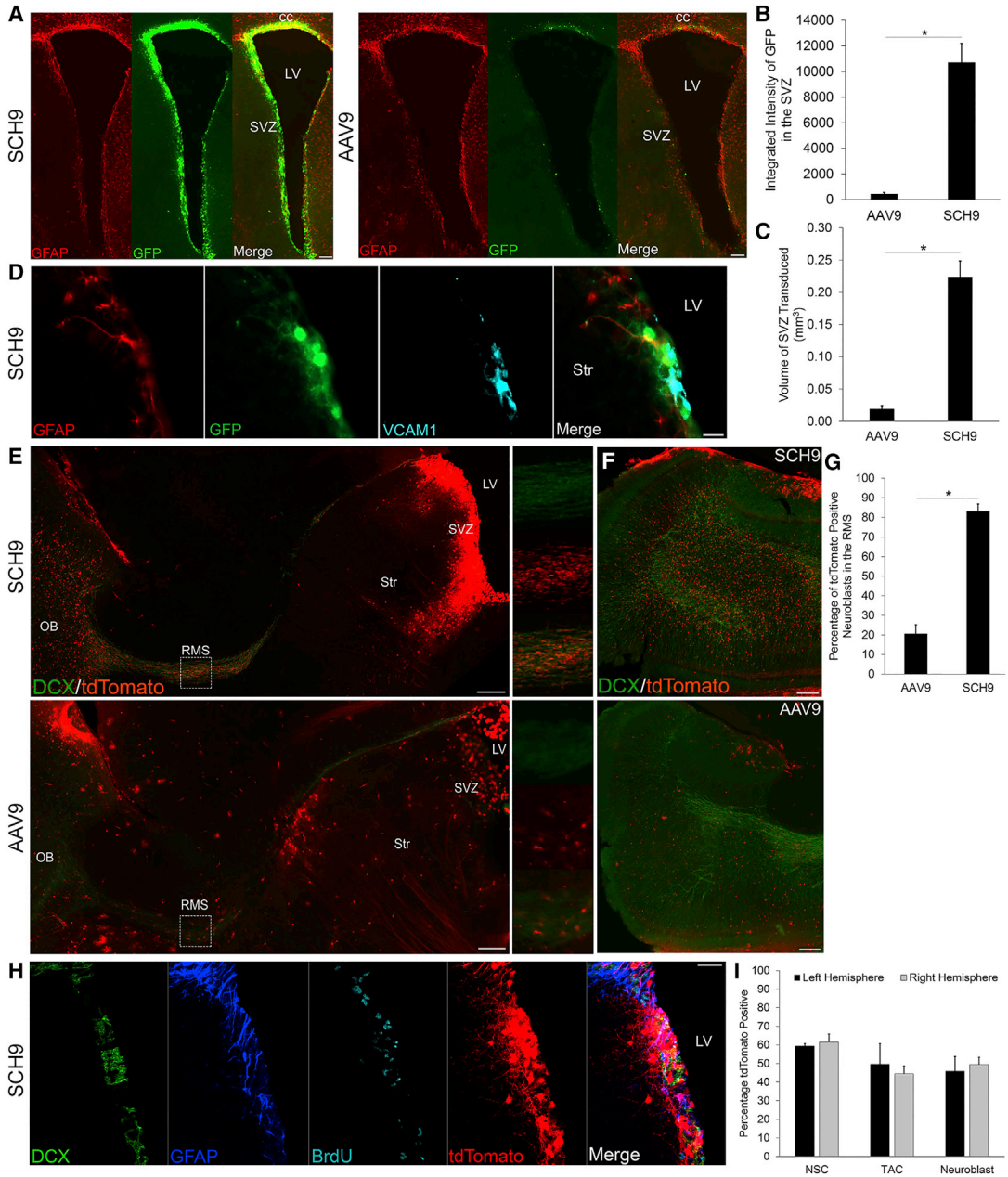


Figure 4. SCH9 Efficiently Transduces Neural Stem Cells in the Subventricular Zone

(A) SCH9 strongly transduces the SVZ contralateral to the injection site. Coronal sections were stained for GFAP (red) and GFP (green). Scale bars indicate 100 μm. (B and C) SCH9 mediates higher GFP fluorescence intensity (B) and transduction volume (C) in the SVZ than AAV9. Data are presented as mean ± SEM; n = 3. Statistical difference of *$p < 0.005$ by two-tailed Student's t test. (D) Neural stem cells infected with SCH9 express GFP (green) and the neural stem cell markers GFAP (red) and VCAM1 (blue) in coronal sections. Scale bar indicates 20 μm. (E) Labeling of tdTomato-positive migrating neuroblasts in the SVZ-RMS-OB pathway 30 days post-injection of SCH9 (top) or AAV9 (bottom) expressing Cre recombinase in Ai9 floxed STOP tdTomato mice. Sagittal sections were stained for the neuronal migration protein doublecortin (DCX, green), and tdTomato fluorescence (red) is native. Dashed rectangles indicate the region of the rostral migratory stream that is magnified in the right panels. Scale bars indicate 200 μm. (F) TdTomato-positive neuroblasts in multiple stages of radial migration and differentiation into granule cells were observed in the olfactory bulb (top: SCH9, bottom: AAV9). Sagittal sections were stained for DCX (green), and tdTomato fluorescence (red) is native. Scale bars indicate 200 μm. (G) SCH9 transduction results in a higher percentage of $DCX^+/tdTomato^+$ neuroblasts in the rostral migratory stream than with AAV9. Data are presented as mean ± SEM; n = 4–5. Statistical difference of *$p < 0.005$ by two-tailed Student's t test. (H) Representative images of marker colocalization with $BrdU^+$ cells. Sagittal sections were stained for DCX (green), GFAP (blue), BrdU (cyan), and tdTomato (red). Scale bar indicates 50 μm. (I) Quantification of $BrdU^+/GFAP^+$ NSCs in the left and right hemispheres. Data are presented as mean ± SD; n = 5. CC, corpus callosum; LV, lateral ventricle; OB, olfactory bulb; RMS, rostral migratory stream; SVZ, subventricular zone; Str, striatum.

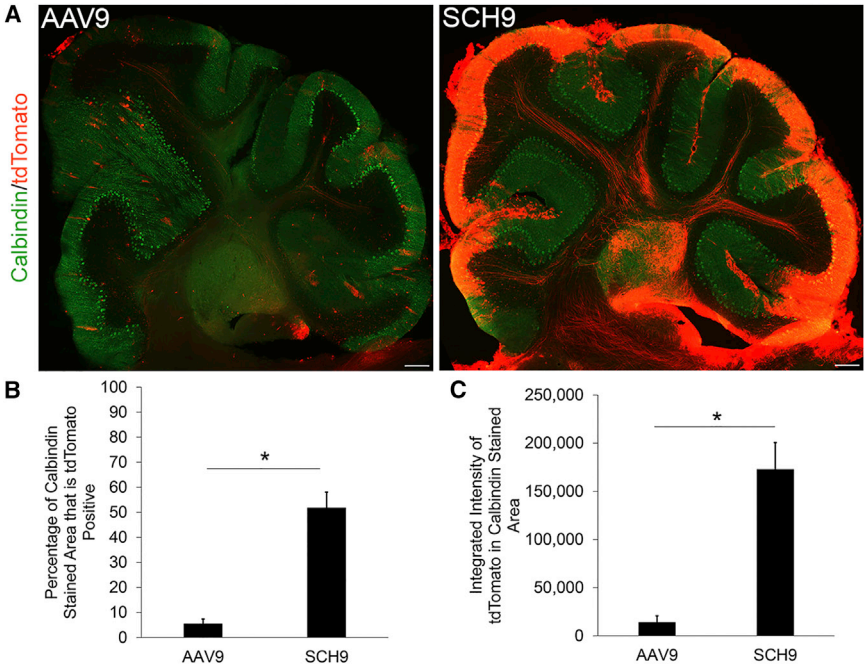


Figure 5. SCH9 Mediates High Transduction of Purkinje Cells in the Cerebellum

(A) Representative images of the cerebellum 30 days after injection of recombinant AAV9 and SCH9 expressing Cre to activate tdTomato (red). Sagittal sections were stained for the Purkinje cell marker calbindin (green), and tdTomato fluorescence is native. Scale bars indicate 200 μm. (B) Integrated intensity of tdTomato fluorescence in calbindin-expressing Purkinje cells. Data are presented as mean ± SEM; n = 4–5. Statistical difference of $^*p < 0.05$ by two-tailed Student's t test. (C) Percentage of calbindin-positive cells expressing tdTomato. Data are presented as mean ± SEM; n = 4–5. Statistical difference of $^*p < 0.05$ by two-tailed Student's t test.

SCH9 Also Displays Tropism for Purkinje Cells in the Cerebellum

Capsid mutations that enhance infection of the target cell type can simultaneously improve transduction in other regions of the brain. Although SCH9 transduction following i.c.v. injection was primarily associated with the SVZ, we also observed increased reporter expression in Purkinje cells of the cerebellum, a region of the brain directly accessible to vector circulating in the CSF (Figure 5A). Purkinje cells are a key target of gene therapies for neurodegenerative diseases including spinocerebellar ataxias.[56] Delivery of SCH9-Cre activated tdTomato reporter expression that was 12.2-fold more intense and covered 9.3-fold greater calbindin-positive area than AAV9-Cre (Figure 5B) as quantified by CellProfiler.

The success of SCH9 in transducing Purkinje cells from the CSF suggests its potential as a gene delivery vector for the cerebellum. Cerebellar gene therapies have employed rAAV delivery to the deep cerebellar nuclei, a major hub in cerebellar circuitry that receives inhibitory inputs from Purkinje cells.[57,58] By harnessing this circuitry, a single injection of rAAV into the deep cerebellar nuclei can transduce Purkinje cells throughout the cerebellar cortex through retrograde transport of the vector. We compared transduction patterns of SCH9 with AAV1, the most commonly used serotype for gene delivery to the cerebellum, after unilateral injection into the deep cerebellar nuclei of the right hemisphere (Figure S5). Both vectors supported strong transduction of Purkinje cells throughout the cerebellum in the ipsilateral hemisphere, indicating that SCH9 can be transported in the retrograde direction. Finally, we evaluated the tropism of SCH9 and AAV1 after direct injection into the striatum (Figure S6). SCH9 (80.7% neurons, 19.3% astrocytes, SD 2.6%) and AAV1 (84.6% neurons, 15.4% astrocytes, SD 1.5%) were similarly neurotropic.

SCH9 Can Utilize Both Heparan Sulfate Proteoglycans and Galactose for Cell Transduction

Given the promising infectious properties of SCH9, we next examined whether its chimeric nature may have conferred a selective advantage to SCH9 by modulating the receptor binding capabilities of its multiple parent serotypes. Block 6 of SCH9 contains the heparin binding pocket of the AAV2 capsid.[59] In addition, blocks 2 and 5 contain the galactose binding residues D271, N272, N470, and Y446 of AAV9, whereas block 6 conserves residue W503.[60] In contrast, SCH2 lacks two of the key galactose-binding residues because of substitution of AAV2 for AAV9 at block 5.

We first employed chromatography to demonstrate that the heparin affinity of both SCHEMA variants was comparable to AAV2, indicating that the chimeric sequence context outside of the heparin pocket did not significantly influence binding affinity (Figure 6A). We next evaluated the potential for dual utilization of heparan sulfate proteoglycans (HSPG) and galactose by infecting CHO-Lec2 cells that express terminal galactose residues and HSPG on the cell surface. As previously described,[61] addition of *Erythrina cristagalli* lectin (ECL) blocks terminal galactose, whereas virus incubation with soluble heparin competitively inhibits AAV serotypes that utilize HSPG for cell entry. As expected, the AAV2 and AAV9 control vectors utilized HSPG and galactose, respectively. Interestingly, SCH2 was solely dependent on HSPG, whereas SCH9 was able to use both HSPG and galactose, and actually required that both be blocked to prevent cell transduction (Figure 6B). After characterizing the different glycan-binding properties of SCH2 and SCH9, we examined whether both variants retained utilization of AAVR, a newly described protein receptor that is critical for AAV infection in natural AAV serotypes.[62] SCH2, SCH9, and the AAV2 control were all clearly dependent on AAVR (Figure S7).

Finally, because DNA shuffling has been shown to disrupt neutralizing antibody epitopes,[14,63] we quantified the resistance of SCH9 to human intravenous immunoglobulin (IVIG), a polyclonal mixture

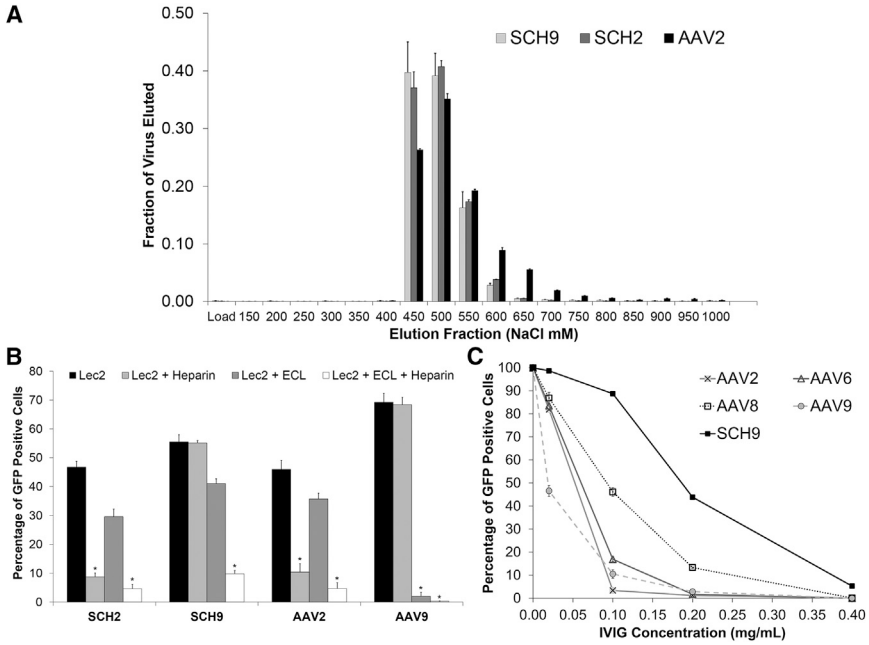


Figure 6. Characterization of SCH9 Glycan Binding and Resistance to Neutralizing Antibodies

(A) SCH9 and SCH2 bind a heparin column with similar affinity as AAV2. Samples were loaded onto a heparin affinity column and eluted with increasing NaCl concentration. The load fraction represents the virus recovered in the column flow-through after sample loading in 150 mM NaCl. Data are represented as mean ± SD; n = 3. (B) SCH9 utilizes both galactose and heparan sulfate proteoglycans for cell entry. Both HSPG and galactose receptors must be blocked to prevent SCH9 infection of Lec2 cells. The controls AAV2 and AAV9 utilize HSPG and galactose, respectively. A modest decrease in infectivity was observed after addition of ECL in all samples, potentially because of steric blocking of receptors on the cell surface. Data are presented as mean ± SEM; n = 3. Statistical difference of *$p < 0.005$ by two-tailed Student's t test. (C) SCH9 is less susceptible to neutralizing antibodies than the parent serotypes AAV2, AAV6, AAV8, and AAV9. Data are presented after being normalized to the fraction of GFP-expressing cells in the absence of IVIG as mean ± SEM; n = 3.

of antibodies against natural AAV serotypes. We found that the antibody titer required to neutralize SCH9 was 2- to 10-fold higher (Table S1) than the parent sequences from which it is derived (Figure 6C). Notably, the greatest fold improvement was relative to AAV9, the most closely related parent sequence.

DISCUSSION

Significant progress has been made in elucidating the molecular mechanisms that govern NSC maintenance and neurogenesis in the SVZ, yet the ability to genetically manipulate endogenous stem cell populations in situ remains challenging because of inefficient vehicles for gene delivery. To overcome the challenges of transducing NSCs in vivo and to broadly address the need for better AAV vectors, we first developed improved technologies for AAV-directed evolution. The SCHEMA scoring algorithm was applied to the multimeric AAV capsid to generate a large library of chimeric variants with unique functional properties. A novel variant, SCH9, was selected from this library using a Cre-dependent strategy and shown to efficiently infect NSCs within the SVZ. This work further establishes the importance of library design and targeted selection strategies for directed evolution, and provides a new tool for manipulation of NSCs in the SVZ.

Our results demonstrate that SCHEMA-guided recombination of the AAV capsid yields chimeric variants that meet or exceed the efficacy of parent serotypes. SCHEMA library design simultaneously minimizes structural disruption while maximizing sequence diversity, two objectives that are infeasible for randomized gene recombination methods. Moreover, the number of mutations introduced far exceeds the diversification achieved with standard mutagenesis techniques including error-prone PCR and peptide insertion. By utilizing the high throughput of AAV-directed evolution, we screened the largest SCHEMA library constructed to date. SCHEMA variants that package efficiently and incorporate sequence elements from multiple AAV parents may represent new starting points in the protein fitness landscape that are promising templates for additional mutagenesis and selections.

In addition to yielding infectious clones, the defined locations of crossovers in the SCHEMA library permit identification of block motifs with desirable properties and assessment of sequence-function relationships.[64] For example, although the library design shuffles blocks 2 through 4 that span the alternative open reading frame of the 204-amino-acid-long nonstructural assembly-activating protein (AAP), we did not observe enrichment of consecutive blocks from the same parent serotype within this AAP region. AAP promotes capsid formation in the nucleus, but recent evidence suggests that it is not essential for capsid assembly of AAV4 and AAV5, and that AAPs enable promiscuous cross-packaging of other serotypes.[65,66] Indeed, we found that overexpression of AAPs from all six AAV parents during packaging did not significantly rescue AAV4 or AAV5 blocks in the library (data not shown). Interestingly, Ho et al.[25] showed that their sample of AAV2/AAV4 SCHEMA chimeras with crossovers in AAP were less likely to assemble into genome-containing capsids. These contrasting outcomes may result from the significant differences in number of parent sequences, crossover locations, and/or library size.

We also found that incorporation of capsid loop structures from both AAV2 and AAV9 conferred dual heparin and galactose affinity to SCH9. Both of these glycans are present on NSCs. Specifically, HSPG function as co-receptors for the binding of key NSC growth

factors including FGF-2 and Wnt.[67] Also, LewisX, a glycan motif containing β1-4-linked galactose, is commonly used to isolate neural stem and progenitor cells from the mouse brain[68] and is enriched in undifferentiated NSCs.[69] The presence of HSPG and galactose on NSCs may contribute to SCH9 transduction, although the low infectivity observed with AAV9 indicates that galactose binding alone is not sufficient. Interestingly, AAV2g9, a rationally designed variant of AAV2 with the galactose binding residues of AAV9 replacing the corresponding residues on the AAV2 capsid, exhibits higher transgene expression in murine cardiac and muscle tissues than AAV2.[61] The improved infectious properties of SCH9 and AAV2g9 suggest that engineering dual glycan binding is a promising approach to increase vector potency.

In the course of completing our selections, a similar Cre-dependent selection strategy was reported to select a variant of AAV9 that crosses the blood-brain barrier more efficiently,[38] highlighting the effectiveness of this technique in overcoming challenging biological barriers. This innovative work positioned loxP sites in the 3′ UTR of *cap*, a location that we had initially tested but found to result in low levels of incidental recombination during bacterial propagation even in recombinase-deficient *E. coli*. We mitigated the likelihood of false positives by introducing loxP sites that flank the *cap* gene such that artifactual recombination prevents translation of functional capsid proteins.

Cre-dependent recovery of *cap* variants offers several advantages over other selection strategies for targeting specific cell populations in the CNS. Fluorescence-activated cell sorting of brain tissue is challenging due to the dense packing of multiple cell types, interconnected cellular processes that are sensitive to shear forces, and large sample sizes that require long sorting times.[70] Magnetic-activated cell sorting is more suitable for processing the entire brain tissue, but is highly dependent on the affinity and specificity of antibodies for a target cell type. Cre-dependent recovery can be performed after simple bulk tissue homogenization, reducing the number and complexity of post-mortem processing steps that can contribute to degradation of AAV episomes. Furthermore, Cre enforces additional stringency by requiring the AAV capsid to traffic to the nucleus, uncoat, and convert the single-stranded genome to a double-stranded form before inversion can occur. Importantly, these are the same criteria that are necessary for expression of a therapeutic transgene. Hundreds of transgenic Cre mouse strains are commercially available,[39] highlighting the broad applicability of this strategy to target a variety of cell types and tissues. Additionally, a GFP-dependent Cre recombinase has recently been developed,[71] further expanding the utility of this system to transgenic GFP reporter mice, of which there are over 1,300 labeling specific cell populations in the CNS.[72] Finally, intersectional strategies using two recombinase systems (Cre/*loxP* and Flp/*FRT*)[73] could be developed to further specify a cell type of interest for positive selection. Collectively, these strategies have strong potential for selections targeting specific cell types within tissues. However, an important limitation is that capsid variants selected in mice may not be effective in larger animal models because of species-specific differences.[9]

Although Cre-dependent selections impose a stringent positive selective pressure, the resulting variants have the potential to transduce off-target cells. Such additional expression may have therapeutic relevance, as in the transduction of the cerebellum described in this work. However, as required for a given application, gene expression may be further restricted by additional mutagenesis and screening for detargeted clones, or by inclusion of a cell-type-specific promoter in the expression cassette. For example, a minimal GFAP promoter, which could further restrict expression to NSCs and astrocytes, has been developed.[74]

Natural AAV serotypes have been considered for gene delivery to the SVZ in vivo, because of the limitations of non-viral-,[54,75,76] adeno-viral-,[77] retroviral-,[78] and lentiviral-based[53] methods. AAV2 exhibits poor NSC transduction following direct injection into the SVZ.[79] AAV4 displays selective tropism for ependymal cells after delivery into the lateral ventricle, while injection directly into the SVZ results in transduction of mature astrocytes.[80] Interestingly, AAV4.18, an engineered mutant of AAV4 with affinity for 2,8-linked polysialic acid (PSA), selectively infects migrating progenitors after i.c.v. injection in neonatal mice.[81] Although this study did not examine the tropism of AAV4.18 in adult mice, it may exhibit preferential tropism for type A neuronal precursors that strongly express PSA-NCAM, a polysialylated glycoprotein that is not expressed on adult NSCs or transit amplifying progenitor cells during adult neurogenesis.[82] Finally, direct injection of AAV1 into the SVZ results in efficient transduction of transit amplifying cells and neuroblasts, but not the NSC population.[83] Moreover, AAV1-mediated transgene expression spanned only one-third of the SVZ, highlighting the need for approaches that achieve widespread transduction of the SVZ from a single vector administration.

SCH9 offers a number of advantages over the gene delivery methods discussed above. A single i.c.v. injection promotes transduction of the entire SVZ in both hemispheres, a significant improvement in transduction volume relative to direct injection into the SVZ. Although i.c.v. injection creates a needle tract through the cortex superior to the ventricle, it does not directly damage the SVZ, and there is negligible impact on the contralateral hemisphere. Widespread transduction of the SVZ enables measurements of total neurogenesis, as well as the effect of genetic modifications on downstream processes including olfaction and response to brain injury. Moreover, SCH9 can provide spatial and temporal control of gene expression without the need to generate bigenic animals. Finally, the transient expression profile of AAV in dividing cells of the germinal niche, distinct from integrating viral vectors, could offer the opportunity for transient expression of regulatory factors during stem cell fate determination and reduce the risk of off-target cutting[84] when using gene editing tools including zinc finger nucleases, TALENs, and CRISPR/Cas9. Future work to determine the minimal vector dose that supports robust transduction of the SVZ can further improve the safety and efficacy of this approach.

NSCs that generate neuronal precursors have also been observed in the SVZ of the adult human brain.[85] Despite their germinal capacity,

chains of migrating neuroblasts are not observed in the human rostral migratory stream beyond 2 years of age,[86] although recent data suggest that some precursors migrate into the striatum and differentiate into interneurons in adult humans.[87] Levels of neurogenesis are low in healthy adults, but pathological conditions including stroke,[88] epilepsy,[89] and Huntington's disease[90] have been shown to induce increased neurogenesis in humans, raising the possibility of therapeutic intervention to further stimulate neurogenesis in response to neurodegeneration. Such interventions are likely to be informed by our understanding of molecular regulation of the murine SVZ. In addition to its role in neurogenesis, the SVZ has also been implicated as a potential oncogenic niche for cancer stem cells that contribute to tumor initiation and resistance in high-grade gliomas. For example, tumor-initiating cells that are resistant to supramaximal chemotherapy doses have been isolated from the SVZ of glioblastoma (GBM) patients[91] and may play a key role in the high recurrence of GBM after chemotherapy. Indeed, irradiation of the SVZ prolongs progression-free survival in GBM patients,[92] suggesting that targeted therapeutic approaches including AAV-mediated delivery of anticancer therapeutics[93] to the SVZ could be effective. Moreover, AAV vectors could be applied to knockout (KO) tumor suppressors or overexpress oncogenes in mice to study the fundamental mechanisms underlying primary tumor formation in the SVZ.[94]

In summary, we describe the development of vector engineering strategies that can be broadly applied to develop AAV variants that overcome challenges as effective biology tools and/or clinical delivery vehicles. The novel variant SCH9 supports gene delivery to NSCs throughout the entire SVZ and will enable future elucidation of molecular mechanisms underlying neurogenesis, as well as the development of therapeutic strategies for diseases that impact the SVZ and striatum.

MATERIALS AND METHODS

SCHEMA Library Design

A library of chimeric AAVs was designed using the SCHEMA scoring function and the RASPP algorithm.[18,35] The amino acid sequences of AAV2, 4, 5, 6, 8, and 9 were aligned using MUSCLE[95] to generate the parent sequence alignment. We modified SCHEMA to consider both intra- and inter-subunit amino acid contacts in the multimeric AAV capsid, wherein a pair of residues is contacting if they contain nonhydrogen atoms within 4.5 Å. We used the crystal structures for AAV2 (1LP3), AAV4 (2G8G), AAV5 (3NTT), AAV6 (3OAH), AAV8 (2QA0), and AAV9 (3UX1) to calculate contacting residue positions. The final contact map contained residue pairs that were contacting in at least 50% of these six parent structures. A chimeric capsid's SCHEMA disruption $\langle E \rangle$ is the number of contacts that contain new amino acid combinations that are not present in any of the parent sequences. A chimeric capsid's $\langle m \rangle$ is the number of mutations from the closest parent sequence.

We used the RASPP algorithm to design libraries that balance the average structural disruption $\langle E \rangle$ and average sequence diversity $\langle m \rangle$. We altered the algorithm to include crossover locations that were amenable to combinatorial golden gate assembly for library construction. Golden gate assembly requires 4-nt stretches that are conserved across all AAV parent sequences. In order to increase the number of possible crossovers sites, and thereby probe a larger sequence space in silico, we included 4-nt stretches that could be silently mutated during library assembly to be identical in all parent sequences. For the library design, we considered a minimum sequence block length of 20 amino acids and maximum length of 250 amino acids. The final library was chosen based on its low $\langle E \rangle$, its uniform block size, and recombination of key capsid structural features.

SCHEMA Library Construction

In order to facilitate combinatorial golden gate cloning with the type IIS restriction enzyme *Bsa*I, we silently mutated all *Bsa*I recognition sites found in pBluescript SK (+), AAV2, 4, 5, 6, 8, and 9, by QuikChange site-directed mutagenesis (Table S2). The 48 DNA sequences corresponding to each shuffled block were PCR amplified from the parent *cap* genes using PCR primers designed in j5, a DNA assembly design automation software[96] (Tables S3 and S4). Primers were designed to incorporate silent mutations at block junctures to facilitate golden gate cloning into the pBluescript vector backbone (Table S5). Equimolar amounts of blocks 1–4 and 5–8 were assembled separately and then combined in a final golden gate reaction using previously described methods.[36] The golden gate assembly was desalted by drop dialysis on a nitrocellulose membrane (Millipore catalog no. VSWP02500) and then transformed into electrocompetent DH10B *E. coli* to achieve a library size greater than the theoretical diversity of 6^8 clones. The library was then subcloned from pBluescript to the AAV packaging plasmid pSub2FloxCap using the restriction enzymes *Hind*III and *Not*I.

The SCHEMA library, before and after packaging, was analyzed using Illumina sequencing. A 2.5-kb fragment containing the AAV *cap* gene was cut out of the pSub2FloxCap vector using the *Hind*III and *Not*I sites and gel extracted. These gel-extracted inserts were used as inputs to the Nextera XT DNA Sample Prep Kit (Illumina). Each sample was barcoded using a different index primer. The resulting libraries were quantified using a high-sensitivity Bioanalyzer chip (Agilent), a Qubit Assay Kit (Invitrogen), and finally qPCR (Kapa Biosystems). The average sequence fragment was ~1,400 bp. The two libraries were pooled in equimolar proportions and sequenced using a MiSeq, version 3, 2 × 300 run with a 5% PhiX control spike-in. Sequencing reads were mapped to all AAV parents using Bowtie2,[97] and the specific sequence blocks present were determined considering the read position and sequence identity to the parents.

Design of AAV Constructs for Cre-Dependent Selections

PCR primers used for construct design and amplification of *cap* are presented in Table S2. pSub2RepKO and pRepHelper were generous gifts of Timothy Day and Dr. John Flannery of University of California (UC), Berkeley. pSub2RepKO, a *rep* KO in the AAV packaging plasmid pSub2,[14] was generated by digestion with *Sgra*I and *Bam*HI, Klenow reaction, and blunt-end ligation. pRepHelper, used to supply

Rep in *trans* during AAV packaging, was created by sequential digestion of pAAV2/rh10 with *Pme*I and *Bsm*I, Klenow reaction, and blunt-end ligation. To insert the lox66 site 5′ of *cap*, we introduced a unique *Bgl*II site into pSub2RepKO by site-directed mutagenesis using the primers *Bgl*IIFwd and *Bgl*IIRev. Oligonucleotides Lox66Fwd and Lox66Rev were annealed and ligated into the *Bgl*II and *Hind*III sites of pSub2RepKO to form pSub2Lox66. To insert the lox71 site 3′ of *cap*, we introduced unique *Xho*I and *Kpn*I sites into pSub2Lox66 by site-directed mutagenesis with the primers XhoIFwd/XhoIRev and KpnIFwd/KpnIRev, respectively. Oligonucleotides SOELox71Fwd and SOELox71Rev were assembled by splice overlap extension and amplified with Lox71Fwd and Lox71Rev. The resulting fragment and pSub2Lox66 were digested with *Xho*I and *Kpn*I and ligated to create pSub2Flox. pSub2Flox and the AAV *cap* libraries used in this selection were digested with *Hind*III and *Not*I and ligated to generate pSub2FloxCap libraries for viral packaging.

AAV Vector Production

HEK293T cells were obtained from the American Type Culture Collection (Manassas, VA, USA) and cultured in DMEM (GIBCO) with 10% fetal bovine serum (Invitrogen) and 1% penicillin/streptomycin (Invitrogen) at 37°C and 5% CO_2. AAV libraries and self-complementary recombinant AAV vectors driving expression of GFP or Cre recombinase under the control of a cytomegalovirus early enhancer/chicken beta actin (CAG) promoter were packaged in HEK293T cells as previously described.[12,14] Briefly, AAV vectors were produced by triple transient transfection with polyethylenimine (23966-1; Polysciences), purified by iodixanol density centrifugation, and buffer exchanged into PBS with 0.001% Tween 20 (Sigma-Aldrich) by Amicon filtration (UFC910024; EMD Millipore). DNase-resistant viral genomic titers were measured by real-time qPCR using a Bio-Rad iCycler (Bio-Rad, Hercules, CA, USA).

In Vivo Selections and Characterization of SCHEMA AAV Variants

Seven-week-old GFAP-Cre 73.12 (Jackson Laboratory Stock 012886), C57BL/6J (Jackson Laboratory Stock 000664), or Ai9 tdTomato mice (Jackson Laboratory Stock 007909) were anesthetized with isoflurane and placed in a stereotaxic apparatus. An incision was made to expose the skull, and a hole was drilled for injection. For library selections, 5 μL of an equimolar mixture of AAV libraries (1×10^{10} viral genomes/μL) was stereotaxically injected into the right lateral ventricle of GFAP-Cre mice (n = 3) at the coordinates 0.05 mm posterior and 1.0 mm lateral to the bregma at a depth of 2.5 mm using a Hamilton syringe as previously described.[98] Injection coordinates were selected using a mouse brain atlas and adjusted after test injections with 0.1% Fast Green dye (Sigma). Injection accuracy throughout the study was confirmed by reporter expression in the choroid plexus and surrounding the contralateral ventricle. Mice were sacrificed 3 weeks after injection, and brain tissue was harvested. The hemisphere contralateral to the injection site was homogenized on dry ice using a mortar and pestle. Homogenized tissue was digested in Hirt lysis buffer with proteinase K (New England Biolabs) and RNase A (Thermo Fisher) at 55°C for 3 hr, and extrachromosomal DNA was isolated using the Hirt method as previously described.[99] The PCR primers Cap_ISF and Cap_R were used to amplify inverted *cap*, whereas primers Cap_NSF and Cap_R specifically amplify non-inverted *cap*. The primers Internal_Cap_ISF and Internal_Cap_R may be used for nested PCR if amplification of inverted *cap* is challenging. After three rounds of selection, capsid sequences were determined by Sanger sequencing (UC Berkeley DNA Sequencing Facility), and dominant variants were digested with *Hind*III and *Not*I and ligated into pXX2Not for recombinant AAV packaging.

To characterize SCH9 and AAV9 in vivo, we stereotaxically injected 5 μL of self-complementary recombinant vector (1×10^{10} viral genomes/μL) expressing GFP or Cre into the right lateral ventricle of C57BL/6 or Ai9 tdTomato mice, respectively, at the coordinates 0.05 mm posterior and 1.0 mm lateral to the bregma at a depth of 2.5 mm using a Hamilton syringe. Ai9 mice received injections of 50 mg/kg BrdU (Sigma-Aldrich) for 3 consecutive days prior to injection of single-stranded SCH9 CAG-Cre. For injections of the deep cerebellar nuclei, 4 μL of recombinant AAV vector (2×10^{9} viral genomes/μL) expressing GFP was stereotaxically injected into the right hemisphere with coordinates 6.0 mm posterior and 2.0 mm lateral to the bregma at a depth of 2.2 mm from the cerebellar surface using a Hamilton syringe. For injections of the striatum, 3 μL of recombinant AAV vector (6.7×10^{8} viral genomes/μL) expressing GFP was stereotaxically injected into the right hemisphere with coordinates 0.5 mm anterior and 2.0 mm lateral to the bregma at a depth of 3.0 mm from the cerebellar surface using a Hamilton syringe. Animal procedures were approved by the UC Berkeley Laboratory Animal Care and Use Committee and conducted in accordance with NIH guidelines for animal care.

Immunohistochemistry

Mice were anesthetized by intraperitoneal injection of 100 mg/kg ketamine and 10 mg/kg xylazine, and were transcardially perfused with 0.9% saline followed by 4% paraformaldehyde. Brains were post-fixed overnight in 4% paraformaldehyde at 4°C, washed in PBS, and stored in 30% sucrose until they sank. Serial coronal or sagittal sections were cut at 40 μm thickness on a Series 8000 sliding microtome (Bright) and stored in cryoprotectant at −20°C until use. Free-floating sections were washed three times in PBS, incubated with blocking solution (10% donkey serum and 1% Triton X-100 in PBS) for 2 hr at room temperature, and stained with primary antibodies in blocking solution for 72 hr at 4°C. The following primary antibodies were used in this study: mouse anti-Calbindin (1:2,000; ab82812; Abcam), rabbit anti-GFP (1:1,000; A-11122; Life Technologies), goat anti-GFAP (1:750; ab53554; Abcam), guinea pig anti-DCX (1:1,000; AB2253; EMD Millipore), rat anti-VCAM1 (1:50; MAB2627; EMD Millipore), chicken anti-GFAP (1:750; ab4674; Abcam), rat anti-BrdU (1:750; ab6326; Abcam), and rabbit anti-tdTomato (1:750; 600-401-379; Rockland). After three washes in PBS, sections were incubated with secondary antibodies for 2 hr at room temperature and stained with DAPI

(Thermo Fisher) for 10 minutes. Stained sections were washed three times in PBS and mounted onto slides using VectaShield HardSet Antifade Mounting Medium (Vector Laboratories).

Imaging and Analysis

Images were acquired using a Zeiss Axio Scan.Z1 or a confocal Zeiss LSM 880 NLO AxioExaminer (UC Berkeley Molecular Imaging Center). All image analyses were conducted on original images acquired with equivalent settings. Data are presented as mean ± SEM, and statistical significance was established by two-tailed Student's t test.

To determine transduction volume in the SVZ, we quantified the surface area of GFP expression in the SVZ from thresholded images using CellProfiler[100] in six coronal sections spanning the SVZ from the anterior horn of the lateral ventricle to the anterior commissure, with three mice per group. The total surface area was multiplied by the section thickness (40 μm) and the distance between sections to obtain the transduction volume. The same thresholded images were used for quantification of integrated intensity of GFP expression using CellProfiler.

To quantify the percentage of tdTomato-positive neuroblasts in the rostral migratory stream, we applied the cell segmentation capabilities of CellProfiler to threshold, segment, and score doublecortin and tdTomato-positive cell bodies in the rostral migratory stream. Measurements were taken from two to five sagittal tissue sections containing the rostral migratory stream in each animal, with four to five mice in each group. To evaluate transduction of adult NSCs, we scored the identities of all BrdU-positive cells in the SVZ by colocalization with tdTomato and GFAP or DCX. Counts were performed on confocal images of every fifth sagittal section spanning the SVZ in five mice with four to five sections per animal.

To calculate the percentage of calbindin-stained area that is tdTomato-positive, we employed a CellProfiler pipeline to generate a thresholded mask of the calbindin stain. This mask was applied to the thresholded tdTomato image, and the tdTomato-positive area was divided by the total calbindin area. The integrated intensity of thresholded tdTomato within the calbindin mask was also recorded. Measurements were taken from four to seven 40 μm sagittal tissue sections spanning the cerebellum, with four to five mice in each group.

In Vitro Characterization of SCHEMA AAV Variants

Unless otherwise noted, all cell lines were cultured in DMEM (GIBCO) supplemented with 10% fetal bovine serum (Invitrogen) and 1% penicillin/streptomycin (Invitrogen) at 37°C and 5% CO_2. The heparin affinities of SCH9, SCH2, and wild-type AAV2 were determined as previously described.[6] A 1 mL HiTrap heparin column (GE Healthcare Sciences) was equilibrated with 150 mM NaCl and 50 mM Tris (at pH 7.5). 1×10^{11} purified viral genomic particles were loaded onto the column and eluted by 50 mM stepwise increases in NaCl up to a final concentration of 950 mM, followed by a 1 M NaCl wash. 10 μL of each elution was used to infect HEK293T cells, and the percentage of GFP-positive cells was quantified 48 hr after infection using a Guava EasyCyte 6HT flow cytometer (EMD/Millipore) (UC Berkeley Stem Cell Center, Berkeley, CA, USA).

AAV utilization of galactose and HSPG for cell transduction was characterized as previously described.[61] CHO-Lec2 cells presenting terminal galactose residues on their surface were obtained from the tissue culture facility at the University of California, Berkeley and cultured in MEM α nucleosides (GIBCO) supplemented with 10% fetal bovine serum (Invitrogen) and 1% penicillin/streptomycin (Invitrogen) at 37°C and 5% CO_2. One day after seeding, cells were incubated at 4°C for 30 min followed by a complete media change into MEM with or without 100 μg/mL *Erythrina cristagalli* lectin (ECL) (Vector Labs). Self-complementary rAAV CAG-GFP virions were treated with soluble heparin (500 μg/mL) in PBS or mock-treated for 1 hr and then used to infect cells at a genomic MOI of 12,000 (n = 3). After a 1 hr incubation with virus, Lec2 cells were washed three times in cold PBS to remove unbound AAV, and the percentage of GFP-expressing cells was quantified 72 hr after infection by flow cytometry.

To analyze antibody evasion properties, we incubated SCH9, AAV2, AAV6, AAV8, and AAV9 at 37°C for 1 hr with serial dilutions of IVIG (Gammagard Liquid 10% NDC 0944-2700-03) and then used them to infect HEK293T cells at a genomic MOI of 8,000 (n = 3) as previously described.[44] IVIG was incubated at 56°C for 30 min to heat-inactivate complement prior to use. The percentage of GFP-expressing cells was quantified 48 hr after infection by flow cytometry. Neutralizing antibody titers were recorded as the first IVIG concentration at which a 50% or greater reduction in GFP expression was observed.

To study dependence on AAVR, we infected wild-type HeLa or $AAVR^{KO}$ cells (clone KIAA0319L, a generous gift from Dr. Jan Carette, Stanford University, CA, USA) at a genomic MOI of 20,000 (n = 6) with SCH9, SCH2, or AAV2 carrying self-complementary CAG-GFP. The percentage of GFP-expressing cells was quantified 72 hr after infection by flow cytometry.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and five tables and can be found with this article online at https://doi.org/10.1016/j.ymthe.2017.09.006.

AUTHOR CONTRIBUTIONS

D.S.O. conceived and performed experiments, analyzed and interpreted the data, and wrote and edited the manuscript. S.S. performed experiments and edited the manuscript, and J.L.S.-O. performed experiments. M.G.S. contributed to conception and design of experiments. P.A.R. designed software, analyzed and interpreted the data, and wrote and edited the manuscript. D.V.S. designed experiments, supervised the project through all stages, and wrote and edited the manuscript.

CONFLICTS OF INTEREST

D.V.S., D.S.O., and J.L.S.-O. are inventors on patents involving AAV-directed evolution. D.V.S. is also a co-founder of a company developing AAV vectors for clinical gene therapy.

ACKNOWLEDGMENTS

D.S.O. was supported by a National Science Foundation Graduate Fellowship and a UC Berkeley Dissertation Fellowship. S.S. was supported by a National Defense Science and Engineering Graduate Fellowship. J.L.S.-O. was supported by a National Science Foundation Graduate Fellowship and a UC Berkeley Graduate Division Fellowship. This work was also supported by NIH grant R01 EY022975. The authors are grateful to Timothy Day and Dr. John Flannery of UC Berkeley for providing pRepHelper and pSub2RepKO, Dr. Jan Carette of Stanford for the AAVR cell line, Dr. Marla Feller of UC Berkeley for the Ai9 tdTomato mice, and Dr. Thomas Gaj for helpful discussions.

REFERENCES

1. Asokan, A., Schaffer, D.V., and Samulski, R.J. (2012). The AAV vector toolkit: poised at the clinical crossroads. Mol. Ther. *20*, 699–708.
2. Kotterman, M.A., Chalberg, T.W., and Schaffer, D.V. (2015). Viral vectors for gene therapy: translational and clinical outlook. Annu. Rev. Biomed. Eng. *17*, 63–89.
3. Spencer, H.T., Riley, B.E., and Doering, C.B. (2016). State of the art: gene therapy of haemophilia. Haemophilia *22* (*Suppl 5*), 66–71.
4. Pierce, E.A., and Bennett, J. (2015). The status of RPE65 gene therapy trials: safety and efficacy. Cold Spring Harb. Perspect. Med. *5*, a017285.
5. Ojala, D.S., Amara, D.P., and Schaffer, D.V. (2015). Adeno-associated virus vectors and neurological gene therapy. Neuroscientist *21*, 84–98.
6. Jang, J.H., Koerber, J.T., Kim, J.S., Asuri, P., Vazin, T., Bartel, M., Keung, A., Kwon, I., Park, K.I., and Schaffer, D.V. (2011). An evolved adeno-associated viral variant enhances gene delivery and gene targeting in neural stem cells. Mol. Ther. *19*, 667–675.
7. Kotterman, M.A., Vazin, T., and Schaffer, D.V. (2015). Enhanced selective gene delivery to neural stem cells in vivo by an adeno-associated viral variant. Development *142*, 1885–1892.
8. Koerber, J.T., Klimczak, R., Jang, J.H., Dalkara, D., Flannery, J.G., and Schaffer, D.V. (2009). Molecular evolution of adeno-associated virus for enhanced glial gene delivery. Mol. Ther. *17*, 2088–2095.
9. Dalkara, D., Byrne, L.C., Klimczak, R.R., Visel, M., Yin, L., Merigan, W.H., Flannery, J.G., and Schaffer, D.V. (2013). In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous. Sci. Transl. Med. *5*, 189ra76.
10. Tervo, D.G., Hwang, B.Y., Viswanathan, S., Gaj, T., Lavzin, M., Ritola, K.D., Lindo, S., Michael, S., Kuleshova, E., Ojala, D., et al. (2016). A designer AAV variant permits efficient retrograde access to projection neurons. Neuron *92*, 372–382.
11. Steines, B., Dickey, D.D., Bergen, J., Excoffon, K.J., Weinstein, J.R., Li, X., Yan, Z., Abou Alaiwa, M.H., Shah, V.S., Bouzek, D.C., et al. (2016). CFTR gene transfer with AAV improves early cystic fibrosis pig phenotypes. JCI Insight *1*, e88728.
12. Koerber, J.T., Jang, J.H., and Schaffer, D.V. (2008). DNA shuffling of adeno-associated virus yields functionally diverse viral progeny. Mol. Ther. *16*, 1703–1709.
13. Klimczak, R.R., Koerber, J.T., Dalkara, D., Flannery, J.G., and Schaffer, D.V. (2009). A novel adeno-associated viral variant for efficient and selective intravitreal transduction of rat Müller cells. PLoS ONE *4*, e7467.
14. Maheshri, N., Koerber, J.T., Kaspar, B.K., and Schaffer, D.V. (2006). Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. Nat. Biotechnol. *24*, 198–204.
15. Moore, J.C., and Arnold, F.H. (1996). Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents. Nat. Biotechnol. *14*, 458–467.
16. Crameri, A., Raillard, S.A., Bermudez, E., and Stemmer, W.P. (1998). DNA shuffling of a family of genes from diverse species accelerates directed evolution. Nature *391*, 288–291.
17. Chaparro-Riggers, J.F., Loo, B.L., Polizzi, K.M., Gibbs, P.R., Tang, X.S., Nelson, M.J., and Bommarius, A.S. (2007). Revealing biases inherent in recombination protocols. BMC Biotechnol. *7*, 77.
18. Voigt, C.A., Martinez, C., Wang, Z.G., Mayo, S.L., and Arnold, F.H. (2002). Protein building blocks preserved by recombination. Nat. Struct. Biol. *9*, 553–558.
19. Meyer, M.M., Hochrein, L., and Arnold, F.H. (2006). Structure-guided SCHEMA recombination of distantly related beta-lactamases. Protein Eng. Des. Sel. *19*, 563–570.
20. Otey, C.R., Landwehr, M., Endelman, J.B., Hiraga, K., Bloom, J.D., and Arnold, F.H. (2006). Structure-guided recombination creates an artificial family of cytochromes P450. PLoS Biol. *4*, e112.
21. Romero, P.A., Stone, E., Lamb, C., Chantranupong, L., Krause, A., Miklos, A.E., Hughes, R.A., Fechtel, B., Ellington, A.D., Arnold, F.H., and Georgiou, G. (2012). SCHEMA-designed variants of human Arginase I and II reveal sequence elements important to stability and catalysis. ACS Synth. Biol. *1*, 221–228.
22. Heinzelman, P., Komor, R., Kanaan, A., Romero, P., Yu, X., Mohler, S., Snow, C., and Arnold, F. (2010). Efficient screening of fungal cellobiohydrolase class I enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination. Protein Eng. Des. Sel. *23*, 871–880.
23. Orelle, C., Carlson, E.D., Szal, T., Florin, T., Jewett, M.C., and Mankin, A.S. (2015). Protein synthesis by ribosomes with tethered subunits. Nature *524*, 119–124.
24. Menzella, H.G., Reid, R., Carney, J.R., Chandran, S.S., Reisinger, S.J., Patel, K.G., Hopwood, D.A., and Santi, D.V. (2005). Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes. Nat. Biotechnol. *23*, 1171–1176.
25. Ho, M.L., Adler, B.A., Torre, M.L., Silberg, J.J., and Suh, J. (2013). SCHEMA computational design of virus capsid chimeras: calibrating how genome packaging, protection, and transduction correlate with calculated structural disruption. ACS Synth. Biol. *2*, 724–733.
26. Lim, D.A., and Alvarez-Buylla, A. (2016). The adult ventricular-subventricular zone (V-SVZ) and olfactory bulb (OB) neurogenesis. Cold Spring Harb. Perspect. Biol. *8*, a018820.
27. Gage, F.H. (2000). Mammalian neural stem cells. Science *287*, 1433–1438.
28. Doetsch, F., Caillé, I., Lim, D.A., García-Verdugo, J.M., and Alvarez-Buylla, A. (1999). Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. Cell *97*, 703–716.
29. Liu, Y., Lacson, R., Cassaday, J., Ross, D.A., Kreamer, A., Hudak, E., Peltier, R., McLaren, D., Muñoz-Sanjuan, I., Santini, F., et al. (2009). Identification of small-molecule modulators of mouse SVZ progenitor cell proliferation and differentiation through high-throughput screening. J. Biomol. Screen. *14*, 319–329.
30. Nierode, G.J., Perea, B.C., McFarland, S.K., Pascoal, J.F., Clark, D.S., Schaffer, D.V., and Dordick, J.S. (2016). High-throughput toxicity and phenotypic screening of 3D human neural progenitor cell cultures on a microarray chip platform. Stem Cell Reports *7*, 970–982.
31. Agrawal, S., and Schaffer, D.V. (2005). In situ stem cell therapy: novel targets, familiar challenges. Trends Biotechnol. *23*, 78–83.
32. Heidenreich, M., and Zhang, F. (2016). Applications of CRISPR-Cas systems in neuroscience. Nat. Rev. Neurosci. *17*, 36–44.
33. Gao, G., Vandenberghe, L.H., Alvira, M.R., Lu, Y., Calcedo, R., Zhou, X., and Wilson, J.M. (2004). Clades of adeno-associated viruses are widely disseminated in human tissues. J. Virol. *78*, 6381–6388.
34. Excoffon, K.J., Koerber, J.T., Dickey, D.D., Murtha, M., Keshavjee, S., Kaspar, B.K., Zabner, J., and Schaffer, D.V. (2009). Directed evolution of adeno-associated virus to an infectious respiratory virus. Proc. Natl. Acad. Sci. USA *106*, 3865–3870.
35. Endelman, J.B., Silberg, J.J., Wang, Z.G., and Arnold, F.H. (2004). Site-directed protein recombination as a shortest-path problem. Protein Eng. Des. Sel. *17*, 589–594.

36. Engler, C., and Marillonnet, S. (2013). Combinatorial DNA assembly using Golden Gate cloning. Methods Mol. Biol. *1073*, 141–156.

37. Otey, C.R., Silberg, J.J., Voigt, C.A., Endelman, J.B., Bandara, G., and Arnold, F.H. (2004). Functional evolution and structural conservation in chimeric cytochromes p450: calibrating a structure-guided approach. Chem. Biol. *11*, 309–318.

38. Deverman, B.E., Pravdo, P.L., Simpson, B.P., Kumar, S.R., Chan, K.Y., Banerjee, A., Wu, W.L., Yang, B., Huber, N., Pasca, S.P., and Gradinaru, V. (2016). Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat. Biotechnol. *34*, 204–209.

39. Heffner, C.S., Herbert Pratt, C., Babiuk, R.P., Sharma, Y., Rockwood, S.F., Donahue, L.R., Eppig, J.T., and Murray, S.A. (2012). Supporting conditional mouse mutagenesis with a comprehensive cre characterization resource. Nat. Commun. *3*, 1218.

40. Albert, H., Dale, E.C., Lee, E., and Ow, D.W. (1995). Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. Plant J. *7*, 649–659.

41. Platel, J.C., Gordon, V., Heintz, T., and Bordey, A. (2009). GFAP-GFP neural progenitors are antigenically homogeneous and anchored in their enclosed mosaic niche. Glia *57*, 66–78.

42. Giachino, C., Basak, O., Lugert, S., Knuckles, P., Obernier, K., Fiorelli, R., Frank, S., Raineteau, O., Alvarez-Buylla, A., and Taylor, V. (2014). Molecular diversity subdivides the adult forebrain neural stem cell population. Stem Cells *32*, 70–84.

43. Garcia, A.D., Doan, N.B., Imura, T., Bush, T.G., and Sofroniew, M.V. (2004). GFAP-expressing progenitors are the principal source of constitutive neurogenesis in adult mouse forebrain. Nat. Neurosci. *7*, 1233–1241.

44. Santiago-Ortiz, J., Ojala, D.S., Westesson, O., Weinstein, J.R., Wong, S.Y., Steinsapir, A., Kumar, S., Holmes, I., and Schaffer, D.V. (2015). AAV ancestral reconstruction library enables selection of broadly infectious viral variants. Gene Ther. *22*, 934–946.

45. Koerber, J.T., Maheshri, N., Kaspar, B.K., and Schaffer, D.V. (2006). Construction of diverse adeno-associated viral libraries for directed evolution of enhanced gene delivery vehicles. Nat. Protoc. *1*, 701–706.

46. Müller, O.J., Kaul, F., Weitzman, M.D., Pasqualini, R., Arap, W., Kleinschmidt, J.A., and Trepel, M. (2003). Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. Nat. Biotechnol. *21*, 1040–1046.

47. Samaranch, L., Salegio, E.A., San Sebastian, W., Kells, A.P., Foust, K.D., Bringas, J.R., Lamarre, C., Forsayeth, J., Kaspar, B.K., and Bankiewicz, K.S. (2012). Adeno-associated virus serotype 9 transduction in the central nervous system of nonhuman primates. Hum. Gene Ther. *23*, 382–389.

48. Schuster, D.J., Dykstra, J.A., Riedl, M.S., Kitto, K.F., Belur, L.R., McIvor, R.S., Elde, R.P., Fairbanks, C.A., and Vulchanova, L. (2014). Biodistribution of adeno-associated virus serotype 9 (AAV9) vector after intrathecal and intravenous delivery in mouse. Front. Neuroanat. *8*, 42.

49. Kokovay, E., Wang, Y., Kusek, G., Wurster, R., Lederman, P., Lowry, N., Shen, Q., and Temple, S. (2012). VCAM1 is essential to maintain the structure of the SVZ niche and acts as an environmental sensor to regulate SVZ lineage progression. Cell Stem Cell *11*, 220–230.

50. Ponti, G., Obernier, K., and Alvarez-Buylla, A. (2013). Lineage progression from stem cells to new neurons in the adult brain ventricular-subventricular zone. Cell Cycle *12*, 1649–1650.

51. Petreanu, L., and Alvarez-Buylla, A. (2002). Maturation and death of adult-born olfactory bulb granule neurons: role of olfaction. J. Neurosci. *22*, 6106–6113.

52. Lois, C., and Alvarez-Buylla, A. (1994). Long-distance neuronal migration in the adult mammalian brain. Science *264*, 1145–1148.

53. Consiglio, A., Gritti, A., Dolcetta, D., Follenzi, A., Bordignon, C., Gage, F.H., Vescovi, A.L., and Naldini, L. (2004). Robust in vivo gene transfer into adult mammalian neural stem cells by lentiviral vectors. Proc. Natl. Acad. Sci. USA *101*, 14835–14840.

54. Barnabé-Heider, F., Meletis, K., Eriksson, M., Bergmann, O., Sabelström, H., Harvey, M.A., Mikkers, H., and Frisén, J. (2008). Genetic manipulation of adult mouse neurogenic niches by in vivo electroporation. Nat. Methods *5*, 189–196.

55. Madisen, L., Zwingman, T.A., Sunkin, S.M., Oh, S.W., Zariwala, H.A., Gu, H., Ng, L.L., Palmiter, R.D., Hawrylycz, M.J., Jones, A.R., et al. (2010). A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. Nat. Neurosci. *13*, 133–140.

56. Orr, H.T. (2012). Cell biology of spinocerebellar ataxia. J. Cell Biol. *197*, 167–177.

57. Keiser, M.S., Kordower, J.H., Gonzalez-Alegre, P., and Davidson, B.L. (2015). Broad distribution of ataxin 1 silencing in rhesus cerebella for spinocerebellar ataxia type 1 therapy. Brain *138*, 3555–3566.

58. Dodge, J.C., Haidet, A.M., Yang, W., Passini, M.A., Hester, M., Clarke, J., Roskelley, E.M., Treleaven, C.M., Rizo, L., Martin, H., et al. (2008). Delivery of AAV-IGF-1 to the CNS extends survival in ALS mice through modification of aberrant glial cell activity. Mol. Ther. *16*, 1056–1064.

59. Kern, A., Schmidt, K., Leder, C., Müller, O.J., Wobus, C.E., Bettinger, K., Von der Lieth, C.W., King, J.A., and Kleinschmidt, J.A. (2003). Identification of a heparin-binding motif on adeno-associated virus type 2 capsids. J. Virol. *77*, 11072–11081.

60. Bell, C.L., Gurda, B.L., Van Vliet, K., Agbandje-McKenna, M., and Wilson, J.M. (2012). Identification of the galactose binding domain of the adeno-associated virus serotype 9 capsid. J. Virol. *86*, 7326–7333.

61. Shen, S., Horowitz, E.D., Troupes, A.N., Brown, S.M., Pulicherla, N., Samulski, R.J., Agbandje-McKenna, M., and Asokan, A. (2013). Engraftment of a galactose receptor footprint onto adeno-associated viral capsids improves transduction efficiency. J. Biol. Chem. *288*, 28814–28823.

62. Pillay, S., Meyer, N.L., Puschnik, A.S., Davulcu, O., Diep, J., Ishikawa, Y., Jae, L.T., Wosen, J.E., Nagamine, C.M., Chapman, M.S., and Carette, J.E. (2016). An essential receptor for adeno-associated virus infection. Nature *530*, 108–112.

63. Grimm, D., Lee, J.S., Wang, L., Desai, T., Akache, B., Storm, T.A., and Kay, M.A. (2008). In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. J. Virol. *82*, 5887–5911.

64. Heinzelman, P., Romero, P.A., and Arnold, F.H. (2013). Efficient sampling of SCHEMA chimera families to identify useful sequence elements. Methods Enzymol. *523*, 351–368.

65. Sonntag, F., Köther, K., Schmidt, K., Weghofer, M., Raupp, C., Nieto, K., Kuck, A., Gerlach, B., Böttcher, B., Müller, O.J., et al. (2011). The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes. J. Virol. *85*, 12686–12697.

66. Earley, L.F., Powers, J.M., Adachi, K., Baumgart, J.T., Meyer, N.L., Xie, Q., Chapman, M.S., and Nakai, H. (2017). Adeno-associated Virus (AAV) assembly-activating protein is not an essential requirement for capsid assembly of AAV serotypes 4, 5, and 11. J. Virol. *91*, e01980-16.

67. Lanctot, P.M., Gage, F.H., and Varki, A.P. (2007). The glycans of stem cells. Curr. Opin. Chem. Biol. *11*, 373–380.

68. Hennen, E., and Faissner, A. (2012). LewisX: a neural stem cell specific glycan? Int. J. Biochem. Cell Biol. *44*, 830–833.

69. Yagi, H., Saito, T., Yanagisawa, M., Yu, R.K., and Kato, K. (2012). Lewis X-carrying N-glycans regulate the proliferation of mouse embryonic neural stem cells via the Notch signaling pathway. J. Biol. Chem. *287*, 24356–24364.

70. Saxena, A., Wagatsuma, A., Noro, Y., Kuji, T., Asaka-Oba, A., Watahiki, A., Gurnot, C., Fagiolini, M., Hensch, T.K., and Carninci, P. (2012). Trehalose-enhanced isolation of neuronal sub-types from adult mouse brain. Biotechniques *52*, 381–385.

71. Tang, J.C., Rudolph, S., Dhande, O.S., Abraira, V.E., Choi, S., Lapan, S.W., Drew, I.R., Drokhlyansky, E., Huberman, A.D., Regehr, W.G., and Cepko, C.L. (2015). Cell type-specific manipulation with GFP-dependent Cre recombinase. Nat. Neurosci. *18*, 1334–1341.

72. Heintz, N. (2004). Gene expression nervous system atlas (GENSAT). Nat. Neurosci. *7*, 483.

73. Dymecki, S.M., Ray, R.S., and Kim, J.C. (2010). Mapping cell fate and function using recombinase-based intersectional strategies. Methods Enzymol. *477*, 183–213.

74. Lee, Y., Messing, A., Su, M., and Brenner, M. (2008). GFAP promoter elements required for region-specific and astrocyte-specific expression. Glia *56*, 481–493.

75. Lemkine, G.F., Mantero, S., Migné, C., Raji, A., Goula, D., Normandie, P., Levi, G., and Demeneix, B.A. (2002). Preferential transfection of adult mouse neural stem cells and their immediate progeny in vivo with polyethylenimine. Mol. Cell. Neurosci. *19*, 165–174.

76. Falk, A., Holmström, N., Carlén, M., Cassidy, R., Lundberg, C., and Frisén, J. (2002). Gene delivery to adult neural stem cells. Exp. Cell Res. *279*, 34–39.

77. Yoon, S.O., Lois, C., Alvirez, M., Alvarez-Buylla, A., Falck-Pedersen, E., and Chao, M.V. (1996). Adenovirus-mediated gene delivery into neuronal precursors of the adult mouse brain. Proc. Natl. Acad. Sci. USA *93*, 11974–11979.

78. Rogelius, N., Ericson, C., and Lundberg, C. (2005). In vivo labeling of neuroblasts in the subventricular zone of rats. J. Neurosci. Methods *142*, 285–293.

79. Passini, M.A., Lee, E.B., Heuer, G.G., and Wolfe, J.H. (2002). Distribution of a lysosomal enzyme in the adult brain by axonal transport and by cells of the rostral migratory stream. J. Neurosci. *22*, 6437–6446.

80. Liu, G., Martins, I.H., Chiorini, J.A., and Davidson, B.L. (2005). Adeno-associated virus type 4 (AAV4) targets ependyma and astrocytes in the subventricular zone and RMS. Gene Ther. *12*, 1503–1508.

81. Murlidharan, G., Corriher, T., Ghashghaei, H.T., and Asokan, A. (2015). Unique glycan signatures regulate adeno-associated virus tropism in the developing brain. J. Virol. *89*, 3976–3987.

82. Gascon, E., Vutskits, L., and Kiss, J.Z. (2010). The role of PSA-NCAM in adult neurogenesis. Adv. Exp. Med. Biol. *663*, 127–136.

83. Bockstael, O., Melas, C., Pythoud, C., Levivier, M., McCarty, D., Samulski, R.J., De Witte, O., and Tenenbaum, L. (2012). Rapid transgene expression in multiple precursor cell types of adult rat subventricular zone mediated by adeno-associated type 1 vectors. Hum. Gene Ther. *23*, 742–753.

84. Cho, S.W., Kim, S., Kim, Y., Kweon, J., Kim, H.S., Bae, S., and Kim, J.S. (2014). Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. *24*, 132–141.

85. Sanai, N., Tramontin, A.D., Quiñones-Hinojosa, A., Barbaro, N.M., Gupta, N., Kunwar, S., Lawton, M.T., McDermott, M.W., Parsa, A.T., Manuel-García Verdugo, J., et al. (2004). Unique astrocyte ribbon in adult human brain contains neural stem cells but lacks chain migration. Nature *427*, 740–744.

86. Sanai, N., Nguyen, T., Ihrie, R.A., Mirzadeh, Z., Tsai, H.H., Wong, M., Gupta, N., Berger, M.S., Huang, E., Garcia-Verdugo, J.M., et al. (2011). Corridors of migrating neurons in the human brain and their decline during infancy. Nature *478*, 382–386.

87. Ernst, A., Alkass, K., Bernard, S., Salehpour, M., Perl, S., Tisdale, J., Possnert, G., Druid, H., and Frisén, J. (2014). Neurogenesis in the striatum of the adult human brain. Cell *156*, 1072–1083.

88. Macas, J., Nern, C., Plate, K.H., and Momma, S. (2006). Increased generation of neuronal progenitors after ischemic injury in the aged adult human forebrain. J. Neurosci. *26*, 13114–13119.

89. Liu, Y.W., Curtis, M.A., Gibbons, H.M., Mee, E.W., Bergin, P.S., Teoh, H.H., Connor, B., Dragunow, M., and Faull, R.L. (2008). Doublecortin expression in the normal and epileptic adult human brain. Eur. J. Neurosci. *28*, 2254–2265.

90. Curtis, M.A., Penney, E.B., Pearson, A.G., van Roon-Mom, W.M., Butterworth, N.J., Dragunow, M., Connor, B., and Faull, R.L. (2003). Increased cell proliferation and neurogenesis in the adult human Huntington's disease brain. Proc. Natl. Acad. Sci. USA *100*, 9023–9027.

91. Piccirillo, S.G., Spiteri, I., Sottoriva, A., Touloumis, A., Ber, S., Price, S.J., Heywood, R., Francis, N.J., Howarth, K.D., Collins, V.P., et al. (2015). Contributions to drug resistance in glioblastoma derived from malignant cells in the sub-ependymal zone. Cancer Res. *75*, 194–202.

92. Evers, P., Lee, P.P., DeMarco, J., Agazaryan, N., Sayre, J.W., Selch, M., and Pajonk, F. (2010). Irradiation of the potential cancer stem cell niches in the adult brain improves progression-free survival of patients with malignant glioma. BMC Cancer *10*, 384.

93. Santiago-Ortiz, J.L., and Schaffer, D.V. (2016). Adeno-associated virus (AAV) vectors in cancer gene therapy. J. Control. Release *240*, 287–301.

94. Abel, T.W., Clark, C., Bierie, B., Chytil, A., Aakre, M., Gorska, A., and Moses, H.L. (2009). GFAP-Cre-mediated activation of oncogenic K-ras results in expansion of the subventricular zone and infiltrating glioma. Mol. Cancer Res. *7*, 645–653.

95. Edgar, R.C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. *32*, 1792–1797.

96. Hillson, N.J., Rosengarten, R.D., and Keasling, J.D. (2012). j5 DNA assembly design automation software. ACS Synth. Biol. *1*, 14–21.

97. Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol. *10*, R25.

98. Lai, K., Kaspar, B.K., Gage, F.H., and Schaffer, D.V. (2003). Sonic hedgehog regulates adult neural progenitor proliferation in vitro and in vivo. Nat. Neurosci. *6*, 21–27.

99. Arad, U. (1998). Modified Hirt procedure for rapid purification of extrachromosomal DNA from mammalian cells. Biotechniques *24*, 760–762.

100. Carpenter, A.E., Jones, T.R., Lamprecht, M.R., Clarke, C., Kang, I.H., Friman, O., Guertin, D.A., Chang, J.H., Lindquist, R.A., Moffat, J., et al. (2006). CellProfiler: image analysis software for identifying and quantifying cell phenotypes. Genome Biol. *7*, R100.

YMTHE, Volume *26*

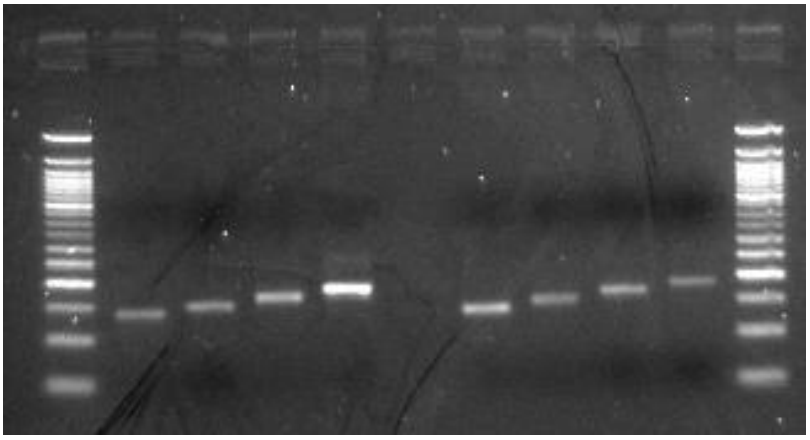
Supplemental Information

In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the SVZ

David S. Ojala, Sabrina Sun, Jorge L. Santiago-Ortiz, Mikhail G. Shapiro, Philip A. Romero, and David V. Schaffer

SUPPLEMENTARY INFORMATION

Figure S1. Inversion of *cap* in recombinase deficient Sure2 *E. coli* was detectable by PCR but not restriction digest. Lox66 and lox71 sites flanking stuffer sequences of 83, 95, 110, or 125 base pairs were cloned between the *Xho*I and *Kpn*I sites of pSub2RepKO. Constructs were transformed into competent *E. coli* expressing Cre recombinase (StrataClone) or recombinase-deficient Sure2 (Agilent), plasmid DNA was purified, and PCR was performed to amplify inverted genomes. Inverted amplicons were observed for both bacterial strains, indicating low levels of recombination that could lead to false positives during selections. Recombination was not detected by restriction digest.



Lane	1	2	3	4	5	6	7	8	9	10	11
Sample	50 bp DNA Ladder (NEB N3236L)	83 bp stuffer, Cre +	95 bp stuffer, Cre +	110 bp stuffer, Cre +	125 bp stuffer, Cre +		83 bp stuffer, recombinase deficient Sure2	95 bp stuffer, recombinase deficient Sure2	110 bp stuffer, recombinase deficient Sure2	125 bp stuffer, recombinase deficient Sure2	50 bp DNA Ladder (NEB N3236L)

Figure S2. SCH9 and SCH2 *cap* amino acid sequences.

SCH9:

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYLGPFNGL
DKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTSFGGNLGRAV
FQAKKRVLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDT
ESVPDPQPIGEPPAAPSGVGSLTMASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRVI
TTSTRTWALPTYNNHLYKQISNSTSGGSSNDNAYFGYSTPWGYFDFNRFHCHFSPRDWQRLI
NNNWGFRPKRLSFKLFNIQVKEVTQNEGTKTIANNLTSTIQVFTDSDYQLPYVLGSAHEGCL
PPFPADVFMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYEFENVPFHSSY
AHSQSLDRLMNPLIDQYLYYLSKTINGSGQNQQTLKFSVAGPSNMAVQGRNWLPGPCYRQ
QRVSKTSADNNNSEYSWTGATKYHLNGRDSLVNPGPAMASHKDDEEKFFPQSGVLIFGKQ
GSEKTNVDIEKVMITDEEEIRTTNPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMVW
QDRDVYLQGPIWAKIPHTDGNFHPSPLMGGFGMKHPPPQILIKNTPVPADPPTAFNKDKLNS
FITQYSTGQVSVEIEWELQKENSKRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRY
LTRNL

SCH2:

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYLGPFNGL
DKGEPVNAADAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTSFGGNLGRAV
FQAKKRVLEPLGLVEEAAKTAPGKKRPVEQSPQEPDSSAGIGKSGAQPAKKRLNFGQTGDT
ESVPDPQPIGEPPAAPSGVGSLTMASGGGAPVADNNEGADGVGSSSGNWHCDSQWLGDRVI
TTSTRTWALPTYNNHLYKQISNSTSGGSSNDNAYFGYSTPWGYFDFNRFHCHFSPRDWQRLI
NNNWGFRPKRLSFKLFNIQVKEVTQNEGTKTIANNLTSTIQVFTDSDYQLPYVLGSAHEGCL
PPFPADVFMIPQYGYLTLNDGSQAVGRSSFYCLEYFPSQMLRTGNNFQFSYTFEDVPFHSSY
AHSQSLDRLMNPLIDQYLYYLSRTNTPSGTTTQSRLQFSQAGASDIRDQSRNWLPGPCYRQQ
RVSKTSADNNNSEYSWTGATKYHLNGRDSLVNPGPAMASHKDDEEKFFPQSGVLIFGKQGS
EKTNVDIEKVMITDEEEIRTTNPVATEQYGSVSTNLQRGNRQAATADVNTQGVLPGMVWQ
DRDVYLQGPIWAKIPHTDGNFHPSPLMGGFGMKHPPPQILIKNTPVPADPPTAFNKDKLNSFI
TQYSTGQVSVEIEWELQKENSKRWNPEIQYTSNYYKSNNVEFAVNTEGVYSEPRPIGTRYLT
RNL

Figure S3. Alignment of the SCH9 and SCH2 AAV *cap* amino acid sequences with the parent AAV serotypes. Capsid sequences were aligned using the Geneious program (Biomatters). Colored amino acids represent differences relative to the reference SCH9 *cap* sequence. Amino acids involved in heparin or galactose binding are annotated in green and blue, respectively, above the SCH9 sequence.

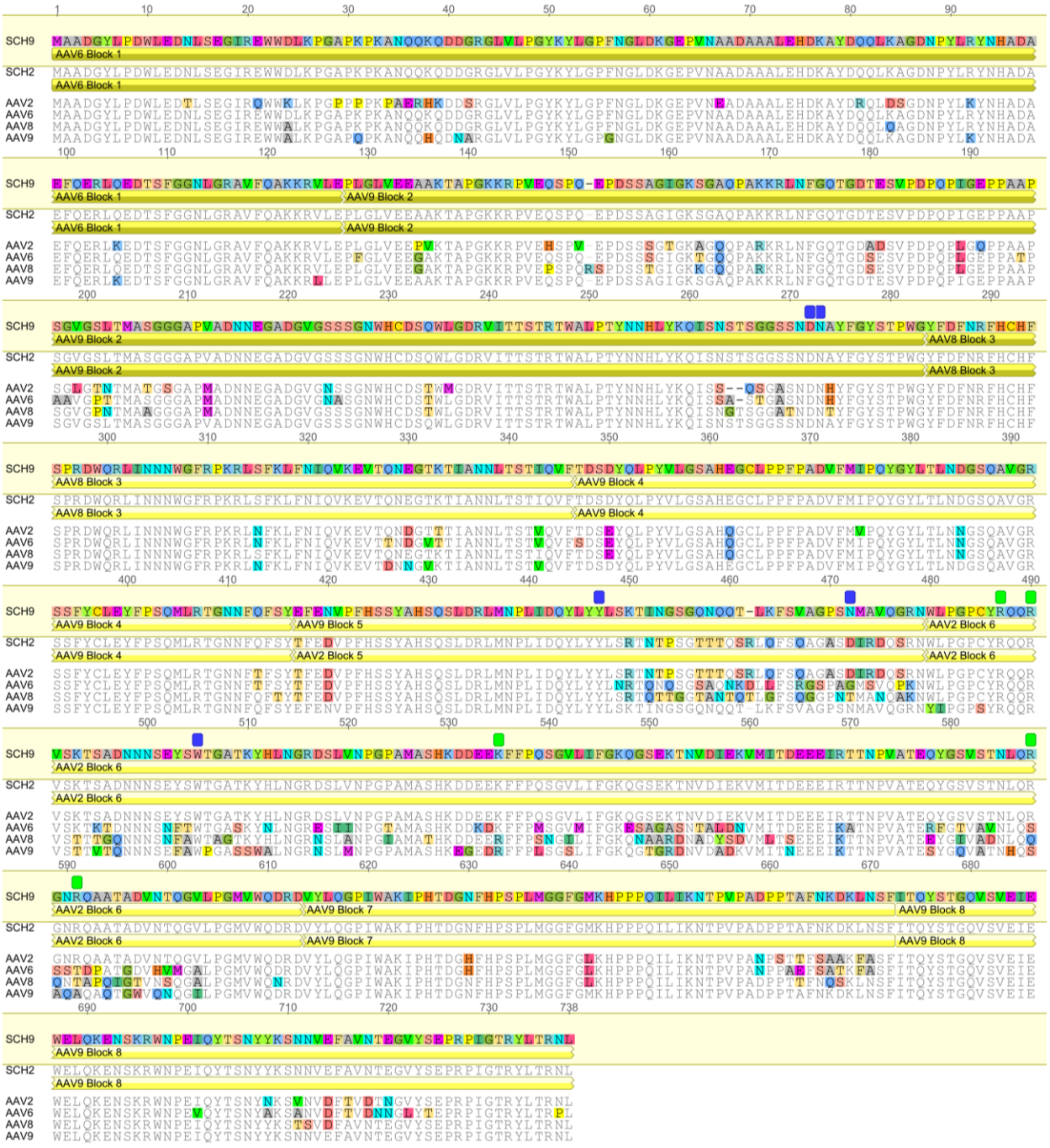


Figure S4. Viral genomic yield of recombinant self-complementary AAV vectors. Data are normalized to the cell culture surface area and are presented as mean ± SD, n=3.

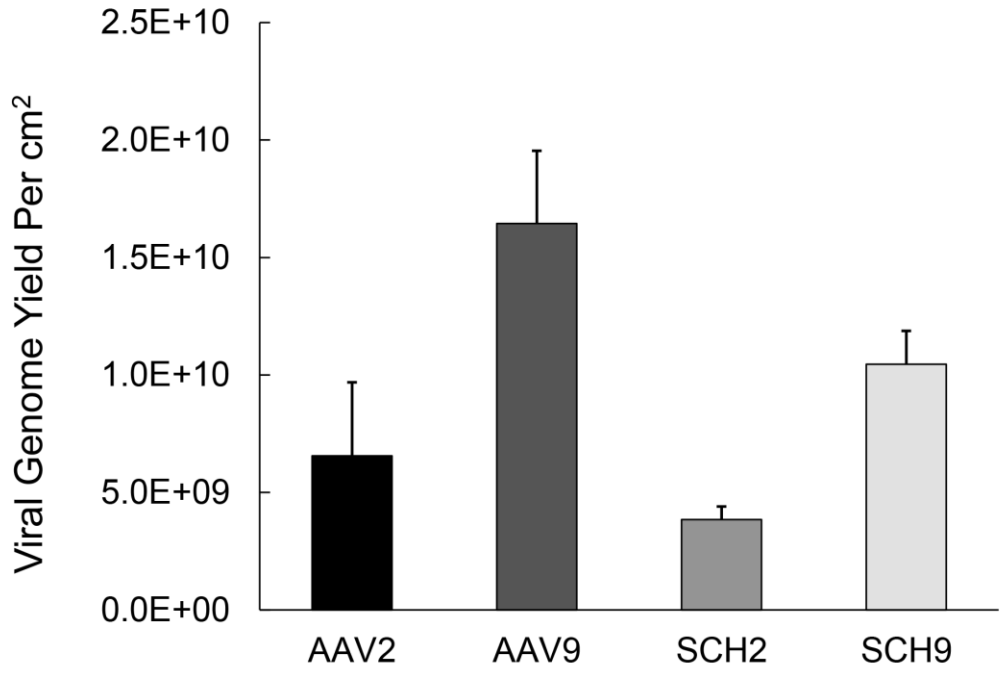


Figure S5. GFP expression in the cerebellum three weeks after unilateral injection of recombinant AAV1 or SCH9 into the deep cerebellar nuclei. Coronal sections were stained for GFP (green) and the Purkinje cell marker calbindin (purple). Scale bars indicate 500 µm.

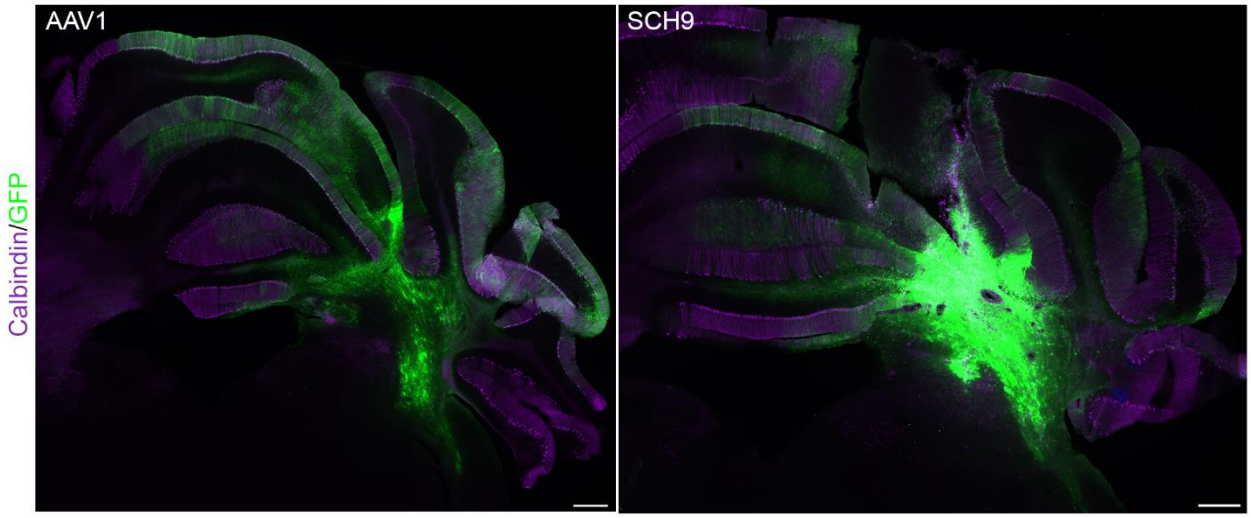


Figure S6. Confocal analysis of GFP expression in the striatum three weeks after injection of recombinant AAV1 or SCH9. Coronal sections were stained for GFP (green), the astrocyte marker GFAP (), and the neuronal marker NeuN (red). Scale bars indicate 25 µm. Image analysis was performed by counting colocalized cells in ImageJ (over 225 cells counted per mouse).

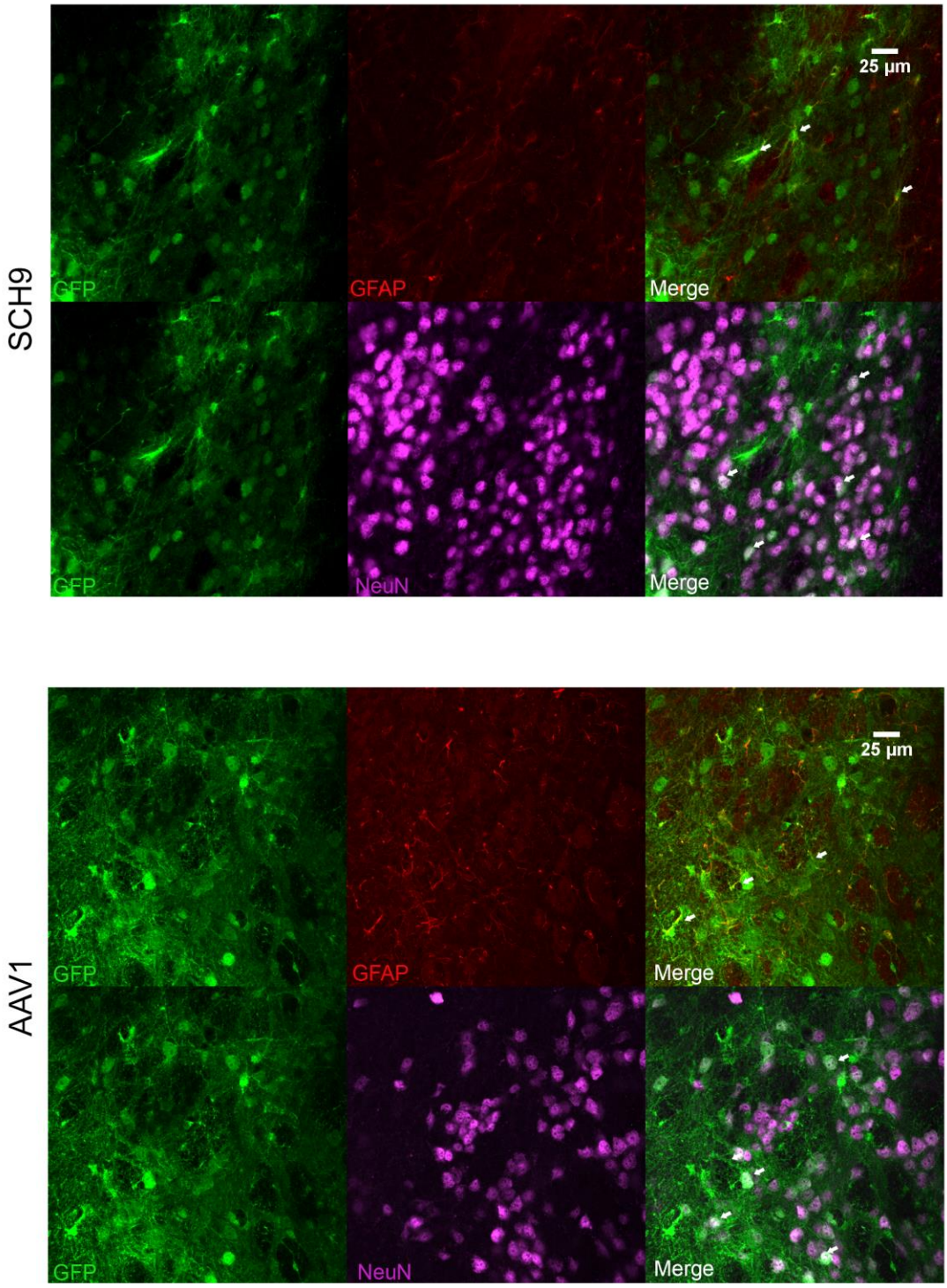


Figure S7. SCH2 and SCH9 are unable to infect a HeLa AAVR knockout cell line. The infectivity of SCH2 and SCH9 was compared with AAV2, a control that is known to utilize AAVR. All three variants efficiently transduce wild type HeLa cells, but not the AAVR knockout line. Data are presented as mean ± SEM, n = 6. *, statistical difference of $P < 0.005$ by two-tailed Student's t-test.

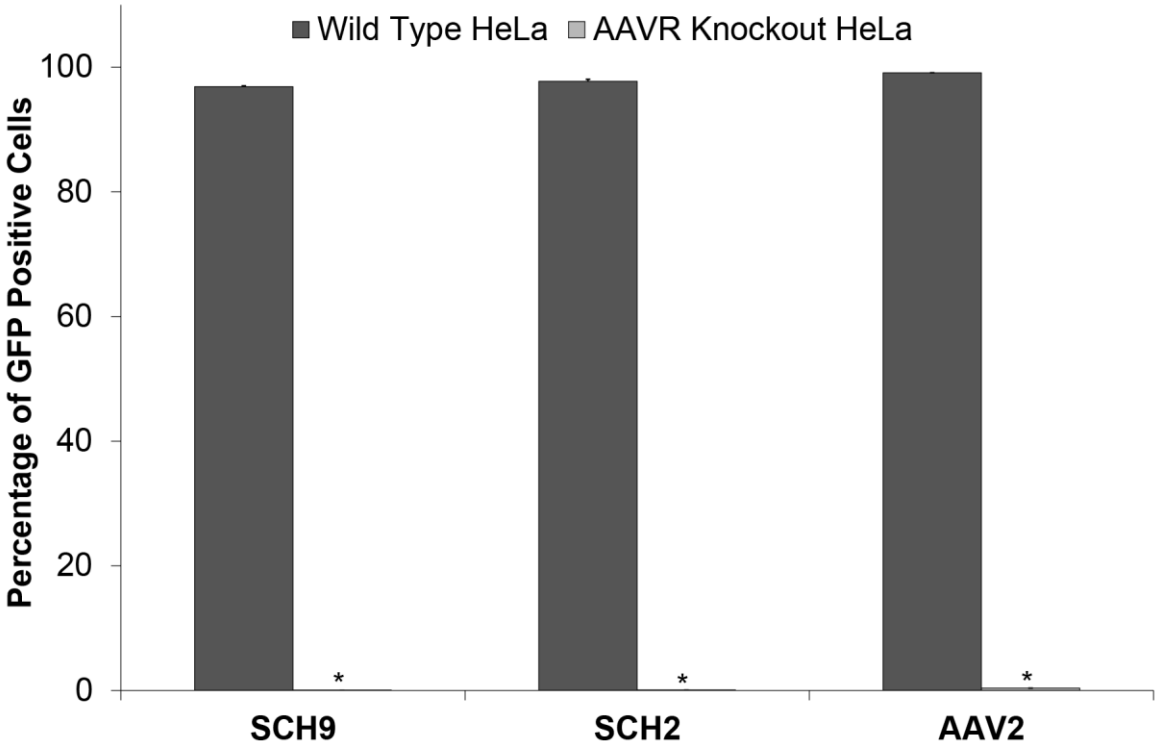


Table S1. Neutralizing IVIG titers of SCH9 and the parent serotypes from which it is derived. The neutralizing titers represent the first IVIG concentration at which 50% or greater reduction in GFP expression was observed.

Variant	Neutralizing IVIG Concentration (mg/mL)	SCH9 Fold Improvement
SCH9	0.20	N/A
AAV2	0.10	2
AAV6	0.10	2
AAV8	0.10	2
AAV9	0.02	10

Table S2. Primer sequences used in this study to design constructs and amplify the AAV *cap* gene.

Primer name	**Primer sequence (5'-3')**
QC_pBluescript_Fwd	GCTGCAATGATACCGCGAAACCCACGCTC
QC_pBluescript_Rev	GAGCGTGGGTTTCGCGGTATCATTGCAGC
QC_AAV4_Fwd	GCTCCTGGAAAGAAGAGGCCGTTGATTGAATCCCC
QC_AAV4_Rev	GGGGATTCAATCAACGGCCTCTTCTTTCCAGGAGC
QC_AAV5_Fwd	GACCCGGAAACGGACTCGATCGAGGAG
QC_AAV5_Rev	CTCCTCGATCGAGTCCGTTTCCGGGTC
QC_AAV6_Fwd	GTTTAGCCGGGGCTCTCCAGCTGGC
QC_AAV6_Rev	GCCAGCTGGAGAGCCCCGGCTAAAC
QC_AAV8_Fwd	CTCCTGGAAAGAAGAGGCCGGTAGAGCCATCAC
QC_AAV8_Rev	GTGATGGCTCTACCGGCCTCTTCTTTCCAGGAG
BglIIFwd	CAAGCGGCCGCGTAAGCTTAGATCTCTGACGTCGATGGCTG CG
BglIIRev	CGCAGCCATCGACGTCAGAGATCTAAGCTTACGCGGCCGCT TG
Lox66Fwd	GATCTATAACTTCGTATAGCATACATTATACGAACGGTACG
Lox66Rev	CGTACCGTTCGTATAATGTATGCTATACGAAGTTATTTCGA
XhoIFwd	CCGCTTGTTAATCAATAAACCGTTTAATTCGTTTCAGTTGAC TCGAGGTCTCTGCGTATTTCTTTCT
XhoIRev	AGAAAGAAATACGCAGAGACCTCGAGTCAACTGAAACGAA TTAAACGGTTTATTGATTAACAAGCGG
KpnIFwd	CGTAGATAAGTAGCATGGCGGGTTAATCAGGTACCACAAG GAACCCCTAG
KpnIRev	CTAGGGGTTCCTTGTGGTACCTGATTAACCCGCCATGCTACT TATCTACG
SOELox71Fwd	GTCAGCCTCGAGATAACTTCGTATAATGTATGCTATACGAA CGGTACTGTGGTCGTCATTGGCAACTACACCTGTTCG
SOELox71Rev	CGTCACGGTACCTGTGGAATTGTGAGCGCTCACAATTCCAC AGCTAGCCTATTTACCGATACCACACGAACAGGTGTAGTTG CCAATGACG
Lox71Fwd	GTCAGCCTCGAGATAACTTCG
Lox71Rev	CGTCACGGTACCTGTGG
Cap_ISF	CATGGAAACTAGATAAGAAAGA
Cap_NSF	GGTACGAAGCTTCGATCAACTACGCAG
Cap_R	AGCTAGCCTATTTACCGATAC
Internal_Cap_ISF	AAGTTCAACTGAAACGAATTA
Internal_Cap_R	CACACGAACAGGTGTAGTT

Table S3. Primer sequences designed in j5 to amplify each sequence block for combinatorial golden gate assembly of the SCHEMA AAV library.

Primer name	Primer sequence (5'-3')
DO_02_(Vector_Backbone)_forward	CACACCAGGTCTCATTGCGCGCTTGGCGTAATCATGG
DO_03_(Vector_Backbone)_reverse	CACACCAGGTCTCATTATAGTGAGTCGTATTACGCGC
DO_04_(AAV2_b1)_forward	CACACCAGGTCTCAATAAGGCGAATTGGGTACCG
DO_05_(AAV2_b1)_reverse	CACACCAGGTCTCAGTTCAAGAACCCTCTTTTTCGC
DO_06_(AAV2_b2)_forward	CACACCAGGTCTCAGAACCTCTGGGCCTGGTTGAG
DO_07_(AAV2_b2)_reverse	CACACCAGGTCTCAACCCCCAAGGGGTGCTGTAG
DO_08_(AAV2_b3)_forward	CACACCAGGTCTCAGGGTATTTTGACTTCAACAGATTCCACTGC
DO_09_(AAV2_b3)_reverse	CACACCAGGTCTCAAAAGACCTGAACCGTGCTGG
DO_10_(AAV2_b4)_forward	CACACCAGGTCTCACTTTACTGACTCGGAGTACCAGC
DO_11_(AAV2_b4)_reverse	CACACCAGGTCTCAGTAGCTGAAGGTAAAGTTGTTTCC
DO_12_(AAV2_b5)_forward	CACACCAGGTCTCACTACACTTTTGAGGACGTTCC
DO_13_(AAV2_b5)_reverse	CACACCAGGTCTCAAGTTCCTAGACTGGTCCCGAATG
DO_14_(AAV2_b6)_forward	CACACCAGGTCTCAAACTGGCTTCCTGGACCCTG
DO_15_(AAV2_b6)_reverse	CACACCAGGTCTCAGTCTCTGTCCTGCCAGACCATG
DO_16_(AAV2_b7)_forward	CACACCAGGTCTCAAGACGTGTACCTTCAGGGGC
DO_17_(AAV2_b7)_reverse	CACACCAGGTCTCAATGAAGGAAGCAAACTTTGCCG
DO_18_(AAV2_b8)_forward	CACACCAGGTCTCATCATCACACAGTACTCCACGG
DO_19_(AAV2_b8)_reverse	CACACCAGGTCTCAGCAAGCGCGCAATTAACCCTC
DO_20_(AAV4_b1)_reverse	CACACCAGGTCTCAGTTCAAGAACCCTCTTTTTGGC
DO_21_(AAV4_b2)_forward	CACACCAGGTCTCAGAACCTCTTGGTCTGGTTGAG
DO_22_(AAV4_b2)_reverse	CACACCAGGTCTCAACCCCCAGGGGGTGGAGAAT
DO_23_(AAV4_b3)_forward	CACACCAGGTCTCAGGGTATTTTGACTTCAACCGCTTCC
DO_24_(AAV4_b3)_reverse	CACACCAGGTCTCAAAAGATCTGAACCGTGCTGG
DO_25_(AAV4_b4)_forward	CACACCAGGTCTCACTTTGCGGACTCGTCGTACG
DO_26_(AAV4_b4)_reverse	CACACCAGGTCTCAGTAGGTAATTTCAAAGTTGTTGCC
DO_27_(AAV4_b5)_forward	CACACCAGGTCTCACTACAGTTTTGAGAAGGTGCCT
DO_28_(AAV4_b5)_reverse	CACACCAGGTCTCAAGTTCTTTTTAAAGTTGGAAAAGTTGGT
DO_29_(AAV4_b6)_forward	CACACCAGGTCTCAAACTGGCTGCCCGGGCCTTC
DO_30_(AAV4_b6)_reverse	CACACCAGGTCTCAGTCTCTGTTTTGCCAGACCATTC
DO_31_(AAV4_b7)_forward	CACACCAGGTCTCAAGACATTTACTACCAGGGTCCC
DO_32_(AAV4_b7)_reverse	CACACCAGGTCTCAATGAAGGAGTTTACCGGAGTAGAG
DO_33_(AAV4_b8)_forward	CACACCAGGTCTCATCATTACTCAGTACAGCACTGGC
DO_34_(AAV4_b8)_reverse	CACACCAGGTCTCAGCAAGCGCGCAATTAACCCTCACTAAAGG
DO_35_(AAV5_b1)_reverse	CACACCAGGTCTCAGTTCGAGAACCCTTTTCTTGGC
DO_36_(AAV5_b2)_forward	CACACCAGGTCTCAGAACCTTTTGGCCTGGTTGAA
DO_37_(AAV5_b2)_reverse	CACACCAGGTCTCAACCCCCAGGGGGTGCTGTAT
DO_38_(AAV5_b3)_forward	CACACCAGGTCTCAGGGTATTTTGACTTTAACCGCTTCC
DO_39_(AAV5_b3)_reverse	CACACCAGGTCTCAAAAGACTTGGACGGTGGAGG
DO_40_(AAV5_b4)_forward	CACACCAGGTCTCACTTTACGGACGACGACTACC
DO_41_(AAV5_b4)_reverse	CACACCAGGTCTCAGTAGGTAAACTCAAAGTTGTTGCC
DO_42_(AAV5_b5)_forward	CACACCAGGTCTCACTACAACTTTGAGGAGGTGCC
DO_43_(AAV5_b5)_reverse	CACACCAGGTCTCAAGTTTTTGTAGGTGTTGGCGTATCTCC
DO_44_(AAV5_b6)_forward	CACACCAGGTCTCAAACTGGTTCCCGGGGCCCAT
DO_45_(AAV5_b6)_reverse	CACACCAGGTCTCAGTCTCTCTCCATCCACACGC
DO_46_(AAV5_b7)_forward	CACACCAGGTCTCAAGACGTGTACCTCCAAGGACC
DO_47_(AAV5_b7)_reverse	CACACCAGGTCTCAATGAAGCTGCTGACGGGCAC
DO_48_(AAV5_b8)_forward	CACACCAGGTCTCATCATCACCCAGTACAGCACC
DO_49_(AAV6_b1)_reverse	CACACCAGGTCTCAGTTCGAGAACCCTCTTCTTGG
DO_50_(AAV6_b2)_forward	CACACCAGGTCTCAGAACCTTTTGGTCTGGTTGAGG
DO_51_(AAV6_b2)_reverse	CACACCAGGTCTCAACCCCCAGGGGGTGCTGTAGC
DO_52_(AAV6_b3)_forward	CACACCAGGTCTCAGGGTATTTTGATTTCAACAGATTCCACTGC
DO_53_(AAV6_b3)_reverse	CACACCAGGTCTCAAAAGACTTGAACCGTGCTGG
DO_54_(AAV6_b4)_forward	CACACCAGGTCTCACTTTTCGGACTCGGAGTACC
DO_55_(AAV6_b4)_reverse	CACACCAGGTCTCAGTAGCTGAAGGTAAAGTTATTGCC
DO_56_(AAV6_b5)_forward	CACACCAGGTCTCACTACACCTTCGAGGACGTGC
DO_57_(AAV6_b5)_reverse	CACACCAGGTCTCAAGTTTTTGGGCTGAACAGACATGC
DO_58_(AAV6_b6)_forward	CACACCAGGTCTCAAACTGGCTACCTGGACCCTG
DO_59_(AAV6_b6)_reverse	CACACCAGGTCTCAGTCTCTGTCTTGCCACACCATTC
DO_60_(AAV6_b7)_forward	CACACCAGGTCTCAAGACGTATACCTGCAGGGTCC
DO_61_(AAV6_b7)_reverse	CACACCAGGTCTCAATGAATGAAGCAAACTTTGTAGCC
DO_62_(AAV6_b8)_forward	CACACCAGGTCTCATCATCACCCAGTATTCCACAGG
DO_63_(AAV8_b1)_reverse	CACACCAGGTCTCAGTTCGAGAACCCGCTTCTTGG
DO_64_(AAV8_b2)_forward	CACACCAGGTCTCAGAACCTCTCGGTCTGGTTGAG
DO_65_(AAV8_b2)_reverse	CACACCAGGTCTCAACCCCCAGGGGGTGCTGTAG
DO_66_(AAV8_b3)_forward	CACACCAGGTCTCAGGGTATTTTGACTTTAACAGATTCCACTGC
DO_67_(AAV8_b3)_reverse	CACACCAGGTCTCAAAAGACCTGGATGGTGCTGG
DO_68_(AAV8_b4)_forward	CACACCAGGTCTCACTTTACGGACTCGGAGTACC
DO_69_(AAV8_b4)_reverse	CACACCAGGTCTCAGTAGGTAAACTGGAAGTTGTTGC
DO_70_(AAV8_b5)_reverse	CACACCAGGTCTCAAGTTCTTTGCCTGATTGGCC
DO_71_(AAV8_b6)_forward	CACACCAGGTCTCAAACTGGCTGCCAGGACCCTG
DO_72_(AAV8_b6)_reverse	CACACCAGGTCTCAGTCTCGGTTCTGCCAGACCATAC
DO_73_(AAV8_b7)_forward	CACACCAGGTCTCAAGACGTGTACCTGCAGGGTCC
DO_74_(AAV8_b7)_reverse	CACACCAGGTCTCAATGAAAGAGTTCAGCTTTGACTGG
DO_75_(AAV8_b8)_forward	CACACCAGGTCTCATCATCACGCAATACAGCACCG
DO_76_(AAV9_b1)_reverse	CACACCAGGTCTCAGTTCAAGAAGCCTCTTTTTGGC
DO_77_(AAV9_b2)_forward	CACACCAGGTCTCAGAACCTCTTGGTCTGGTTGAGG
DO_78_(AAV9_b3)_reverse	CACACCAGGTCTCAAAAGACCTGGACCGTGCTGG
DO_79_(AAV9_b4)_forward	CACACCAGGTCTCACTTTACGGACTCAGACTATCAGC
DO_80_(AAV9_b4)_reverse	CACACCAGGTCTCAGTAGCTGAACTGGAAGTTGTTACC
DO_81_(AAV9_b5)_forward	CACACCAGGTCTCACTACGAGTTTGAGAACGTACC
DO_82_(AAV9_b5)_reverse	CACACCAGGTCTCAAGTTTCTTCCCTGGACAGCC
DO_83_(AAV9_b6)_forward	CACACCAGGTCTCAAACTACATACCTGGACCCAGC
DO_84_(AAV9_b6)_reverse	CACACCAGGTCTCAGTCTCTGTCCTGCCAAACCATACC
DO_85_(AAV9_b7)_forward	CACACCAGGTCTCAAGACGTGTACCTGCAAGGAC
DO_86_(AAV9_b7)_reverse	CACACCAGGTCTCAATGAAAGAGTTCAGCTTGTCCTTG
DO_87_(AAV9_b8)_forward	CACACCAGGTCTCATCATCACCCAGTATTCTACTGGC

Table S4. PCR reactions for combinatorial golden gate cloning of the SCHEMA AAV library.

Block ID Number	Primary Template	Forward Primer	Reverse Primer	Amplicon length (bp)
0	pBluescript SK +	DO_02_(Vector_Backbone)_forward	DO_03_(Vector_Backbone)_reverse	2843
1	AAV2	DO_04_(AAV2_b1)_forward	DO_05_(AAV2_b1)_reverse	784
2	AAV2	DO_06_(AAV2_b2)_forward	DO_07_(AAV2_b2)_reverse	491
3	AAV2	DO_08_(AAV2_b3)_forward	DO_09_(AAV2_b3)_reverse	220
4	AAV2	DO_10_(AAV2_b4)_forward	DO_11_(AAV2_b4)_reverse	242
5	AAV2	DO_12_(AAV2_b5)_forward	DO_13_(AAV2_b5)_reverse	222
6	AAV2	DO_14_(AAV2_b6)_forward	DO_15_(AAV2_b6)_reverse	433
7	AAV2	DO_16_(AAV2_b7)_forward	DO_17_(AAV2_b7)_reverse	211
8	AAV2	DO_18_(AAV2_b8)_forward	DO_19_(AAV2_b8)_reverse	298
9	AAV4	DO_04_(AAV2_b1)_forward	DO_20_(AAV4_b1)_reverse	784
10	AAV4	DO_21_(AAV4_b2)_forward	DO_22_(AAV4_b2)_reverse	467
11	AAV4	DO_23_(AAV4_b3)_forward	DO_24_(AAV4_b3)_reverse	220
12	AAV4	DO_25_(AAV4_b4)_forward	DO_26_(AAV4_b4)_reverse	251
13	AAV4	DO_27_(AAV4_b5)_forward	DO_28_(AAV4_b5)_reverse	225
14	AAV4	DO_29_(AAV4_b6)_forward	DO_30_(AAV4_b6)_reverse	445
15	AAV4	DO_31_(AAV4_b7)_forward	DO_32_(AAV4_b7)_reverse	211
16	AAV4	DO_33_(AAV4_b8)_forward	DO_34_(AAV4_b8)_reverse	298
17	AAV5	DO_04_(AAV2_b1)_forward	DO_35_(AAV5_b1)_reverse	787
18	AAV5	DO_36_(AAV5_b2)_forward	DO_37_(AAV5_b2)_reverse	467
19	AAV5	DO_38_(AAV5_b3)_forward	DO_39_(AAV5_b3)_reverse	220
20	AAV5	DO_40_(AAV5_b4)_forward	DO_41_(AAV5_b4)_reverse	248
21	AAV5	DO_42_(AAV5_b5)_forward	DO_43_(AAV5_b5)_reverse	204
22	AAV5	DO_44_(AAV5_b6)_forward	DO_45_(AAV5_b6)_reverse	442
23	AAV5	DO_46_(AAV5_b7)_forward	DO_47_(AAV5_b7)_reverse	208
24	AAV5	DO_48_(AAV5_b8)_forward	DO_19_(AAV2_b8)_reverse	298
25	AAV6	DO_04_(AAV2_b1)_forward	DO_49_(AAV6_b1)_reverse	784
26	AAV6	DO_50_(AAV6_b2)_forward	DO_51_(AAV6_b2)_reverse	494
27	AAV6	DO_52_(AAV6_b3)_forward	DO_53_(AAV6_b3)_reverse	220
28	AAV6	DO_54_(AAV6_b4)_forward	DO_55_(AAV6_b4)_reverse	242
29	AAV6	DO_56_(AAV6_b5)_forward	DO_57_(AAV6_b5)_reverse	222
30	AAV6	DO_58_(AAV6_b6)_forward	DO_59_(AAV6_b6)_reverse	433
31	AAV6	DO_60_(AAV6_b7)_forward	DO_61_(AAV6_b7)_reverse	211
32	AAV6	DO_62_(AAV6_b8)_forward	DO_19_(AAV2_b8)_reverse	298
33	AAV8	DO_04_(AAV2_b1)_forward	DO_63_(AAV8_b1)_reverse	784
34	AAV8	DO_64_(AAV8_b2)_forward	DO_65_(AAV8_b2)_reverse	500
35	AAV8	DO_66_(AAV8_b3)_forward	DO_67_(AAV8_b3)_reverse	220
36	AAV8	DO_68_(AAV8_b4)_forward	DO_69_(AAV8_b4)_reverse	242
37	AAV8	DO_56_(AAV6_b5)_forward	DO_70_(AAV8_b5)_reverse	222
38	AAV8	DO_71_(AAV8_b6)_forward	DO_72_(AAV8_b6)_reverse	433
39	AAV8	DO_73_(AAV8_b7)_forward	DO_74_(AAV8_b7)_reverse	211
40	AAV8	DO_75_(AAV8_b8)_forward	DO_19_(AAV2_b8)_reverse	298
41	AAV9	DO_04_(AAV2_b1)_forward	DO_76_(AAV9_b1)_reverse	787
42	AAV9	DO_77_(AAV9_b2)_forward	DO_51_(AAV6_b2)_reverse	491
43	AAV9	DO_08_(AAV2_b3)_forward	DO_78_(AAV9_b3)_reverse	220
44	AAV9	DO_79_(AAV9_b4)_forward	DO_80_(AAV9_b4)_reverse	242
45	AAV9	DO_81_(AAV9_b5)_forward	DO_82_(AAV9_b5)_reverse	219
46	AAV9	DO_83_(AAV9_b6)_forward	DO_84_(AAV9_b6)_reverse	433
47	AAV9	DO_85_(AAV9_b7)_forward	DO_86_(AAV9_b7)_reverse	211
48	AAV9	DO_87_(AAV9_b8)_forward	DO_34_(AAV4_b8)_reverse	298

Table S5. Unique block junctures specified during primer design to ensure efficient golden gate assembly.

Block Identity	Overhang with Previous Block	Overhang with Next Block
(Vector_Backbone)	TTGC	ATAA
(AAV2_b1)	ATAA	GAAC
(AAV2_b2)	GAAC	GGGT
(AAV2_b3)	GGGT	CTTT
(AAV2_b4)	CTTT	CTAC
(AAV2_b5)	CTAC	AACT
(AAV2_b6)	AACT	AGAC
(AAV2_b7)	AGAC	TCAT
(AAV2_b8)	TCAT	TTGC
(AAV4_b1)	ATAA	GAAC
(AAV4_b2)	GAAC	GGGT
(AAV4_b3)	GGGT	CTTT
(AAV4_b4)	CTTT	CTAC
(AAV4_b5)	CTAC	AACT
(AAV4_b6)	AACT	AGAC
(AAV4_b7)	AGAC	TCAT
(AAV4_b8)	TCAT	TTGC
(AAV5_b1)	ATAA	GAAC
(AAV5_b2)	GAAC	GGGT
(AAV5_b3)	GGGT	CTTT
(AAV5_b4)	CTTT	CTAC
(AAV5_b5)	CTAC	AACT
(AAV5_b6)	AACT	AGAC
(AAV5_b7)	AGAC	TCAT
(AAV5_b8)	TCAT	TTGC
(AAV6_b1)	ATAA	GAAC
(AAV6_b2)	GAAC	GGGT
(AAV6_b3)	GGGT	CTTT
(AAV6_b4)	CTTT	CTAC
(AAV6_b5)	CTAC	AACT
(AAV6_b6)	AACT	AGAC
(AAV6_b7)	AGAC	TCAT
(AAV6_b8)	TCAT	TTGC
(AAV8_b1)	ATAA	GAAC
(AAV8_b2)	GAAC	GGGT
(AAV8_b3)	GGGT	CTTT
(AAV8_b4)	CTTT	CTAC
(AAV8_b5)	CTAC	AACT
(AAV8_b6)	AACT	AGAC
(AAV8_b7)	AGAC	TCAT
(AAV8_b8)	TCAT	TTGC
(AAV9_b1)	ATAA	GAAC
(AAV9_b2)	GAAC	GGGT
(AAV9_b3)	GGGT	CTTT
(AAV9_b4)	CTTT	CTAC
(AAV9_b5)	CTAC	AACT
(AAV9_b6)	AACT	AGAC
(AAV9_b7)	AGAC	TCAT
(AAV9_b8)	TCAT	TTGC