Title
Improved Strategies for Directed Evolution of Adeno-Associated Virus Gene Therapy Vectors

Author
Ojala, David Stephen

Publication Date
2017

Peer reviewed|Thesis/dissertation

Improved Strategies for Directed Evolution of Adeno-Associated Virus Gene Therapy Vectors

By

David Stephen Ojala

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Chemical Engineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David V. Schaffer, Chair
Professor John G. Flannery
Professor Sanjay Kumar

Spring 2017

Abstract

Improved Strategies for Directed Evolution of Adeno-Associated Virus Gene Therapy Vectors

by

David Stephen Ojala

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor David V. Schaffer, Chair

Gene therapy, the introduction of genetic material into a patient to address the underlying causes of disease, has shown strong clinical potential to treat a variety of genetic disorders. There is no scarcity of disease applications for gene therapy and gene editing; rather, the grand challenge in this field is the development of technology platforms for gene delivery that can safely and efficiently mediate stable gene expression. Viral vectors based on adeno-associated virus (AAV) have emerged as successful gene delivery vehicles due to their natural efficacy at circumventing biological barriers and achieving high transduction efficiency. AAV is thus a promising gene therapy vehicle; however, human therapeutic needs demand delivery properties that at best may have conferred AAVs with no selective advantages during natural evolution and at worst may be at odds with natural selection. As a result, there have been considerable efforts to engineer AAV vectors to meet biomedical needs. One successful approach, directed evolution, emulates how viruses naturally evolve – iterative rounds of genetic diversification and selection for improved function – but with selective pressures that can be designed to result in therapeutically useful viruses. Directed evolution has been applied to generate new viral variants with altered gene delivery specificities, but existing AAV libraries and selection strategies must be improved to overcome remaining clinical challenges. These include biological transport barriers that limit viral access to target tissues, poor infectivity of clinically important cell types, and inability to control which tissues are transduced. Such challenges have motivated my dissertation research to engineer new AAV libraries, design improved selection strategies, and evolve AAV vectors that address unmet therapeutic needs.

Significant improvements to naturally occurring AAV variants are needed to address the clinical challenges highlighted above, yet these naturally occurring variants often lack the evolutionary plasticity required to tolerate new mutations. In contrast, ancestral sequences are by definition highly evolvable, having given rise to modern AAV variants. I computationally designed and experimentally constructed an ancestral AAV library to access novel viral sequences with enhanced infectious properties, and to gain insights into AAV's evolutionary history. I found that ancestral AAV variants were broadly infectious, a property that may have enhanced vector spread during the natural evolution of the virus. Moreover, variants selected for muscle tropism mediated up to 31-fold higher gene expression in muscle compared to AAV1, a serotype clinically utilized for muscle delivery, highlighting their potential for gene therapy.

Another method to engineer AAV variants with new biological properties is to shuffle the DNA sequences of naturally occurring AAVs. I applied the computational algorithm SCHEMA to predict the optimal crossover locations for DNA recombination of the AAV capsid gene. SCHEMA calculates the number of amino acid interactions in the capsid crystal structure that are broken upon creation of a chimeric AAV. I designed and constructed a SCHEMA library of over 1.6 million chimeric AAV variants, each representing a possible solution to a therapeutic challenge. I next developed an *in vivo* Cre-dependent selection strategy to drive convergence from millions of variants to a select few that target adult neural stem cells within the central nervous system. Adult neural stem cells confer plasticity to the central nervous system and have neuroregenerative capabilities that may be harnessed to treat disease. Significant progress has been made in elucidating the molecular mechanisms that govern neural stem cell maintenance and neurogenesis in the subventricular zone (SVZ), the largest germinal niche in the adult mammalian brain. Our understanding of stem cell biology has improved, yet the ability to genetically manipulate endogenous stem cell populations *in situ* remains challenging due to inefficient vehicles for gene delivery. After three rounds of *in vivo* selection I identified a novel chimeric AAV, SCH9, which mediates 24-fold higher GFP expression and 12-fold greater transduction volume than AAV9 in the SVZ, and efficiently infects adult neural stem cells. Interestingly, I found that SCH9 combines properties from multiple AAV parent serotypes in its ability to utilize both galactose and heparan sulfate proteoglycans as receptors for cell transduction. Moreover, SCH9 is less susceptible to neutralizing antibodies than the AAV serotypes from which it is derived. In summary, my dissertation research has resulted in new technologies for AAV directed evolution and novel vectors with therapeutic promise.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I have been fortunate to work with many intelligent, kind, and generous colleagues that have played an instrumental role in my development as a scientist and a person. Foremost among these is my research advisor, Professor David Schaffer, who has been an incredible scientific mentor. He has fostered a positive lab culture, encouraged me to pursue creative research directions, taught me to communicate scientific research effectively, and was unequivocally supportive and understanding during difficult times. I would also like to recognize Professor John Flannery and Professor Sanjay Kumar of my dissertation committee for their suggestions and support throughout the years. In addition, I wish to thank my scientific collaborators Phil Romero and Palmer Yu for their valuable contributions to our work. I am also grateful for funding from the National Science Foundation Graduate Research Fellowship and UC Berkeley Dissertation Fellowship.

It has been a great privilege to work with Schaffer Lab members both past and present. Noem Ramey is the glue that holds the Schaffer Lab together. She puts up with last minute orders, poorly timed disasters, and countless favors — all while looking out for the health and emotional well-being of everyone in lab. Collaborating with Jorge Santiago-Ortiz was a highlight of my graduate experience. If I can capture just a tenth of his optimism and outgoing personality I will be a better man. It has been a true pleasure to learn from Thomas Gaj, I greatly enjoyed our scientific discussions and have no doubt that he will make an outstanding PI. I am lucky to have shared a lab bench with Prajit Limsirichai, whose work ethic, generosity, and ability to fix things with a zip tie and rubber bands is second to none. I am also grateful to Dominic Amara for his contributions as an undergraduate researcher, I know he will make a great doctor. Thanks to Leah Byrne and Tim Day for many stimulating discussions about AAV and for happily accommodating countless requests for an aliquot of this or that. Thank you to Melissa Kotterman, Sisi Chen, and Mikhail Shapiro for their valuable mentorship throughout the years, and to Lukasz Bugaj, Bum-Yeol Hwang, and Dawn Spelke for helping me to get started in lab. I am grateful to many others who have made my time at Berkeley both productive and enjoyable: Eddy Kim, Kevin Metcalf, Stanley Herrmann, Maroof Adil, Goncalo Rodrigues, Anthony Conway, and Ashley Fritz.

I am also thankful for the loving support of my parents Jeff and Bev Ojala and my brother Matt Ojala, who believed in my ability and instilled in me the core values I have drawn upon over the years. Finally, to Sabrina Sun, who came into my life when I needed her most and has been unconditional in her support. I could not have done it without you.

Chapter 1: Adeno-Associated Virus Vectors and Neurological Gene Therapy

This chapter is adapted from a manuscript published as

David S. Ojala, Dominic P. Amara, David V. Schaffer. AAV Vectors and Neurological Gene Therapy. *The Neuroscientist* **21**, 84-98 (2015).

1.1 Gene Therapy

Gene therapy is the delivery of genetic material for the expression of therapeutic molecules to treat disease. We have long understood that mutations in genes can cause disease and that by providing a functional copy of a gene or by correcting a mutation through gene editing we can develop cures. The study of genetic disorders has led to elucidation of the molecular mechanisms underlying many diseases. Moreover, the development of sophisticated gene editing tools has resulted in an unprecedented ability to modify the genome with specificity and efficacy. The major limitation preventing clinical translation of these technologies has been the inability to safely and efficiently deliver genetic cargo into a patient. Gene delivery vehicles can be grouped into viral and non-viral technologies [1]. Although non-viral methods have attractive safety features, they are much less efficient than viruses, primarily due to the challenging biological transport barriers present in cells [2]. Viruses evolved over millions of years to overcome these barriers and efficiently deliver genetic cargo to cells. As a result, approximately two-thirds of the 2,000 gene therapy clinical trials completed worldwide have utilized viral vectors, with an increasing emphasis on adeno-associated virus (AAV) [1].

1.2 AAV Vectors in Gene Therapy

AAV vectors in particular have been increasingly successful due to their gene delivery efficacy, lack of pathogenicity, and strong safety profile [3]. As a result of these properties, AAV vectors have enabled clinical successes in a number of recent clinical trials that have established the promise of gene therapy in general, including for the treatment of diseases such as Leber's congenital amaurosis (LCA) [4, 5], where over four Phase I and I/II clinical trials have demonstrated safety and long-term (over five years) improvement in retinal and visual function; hemophilia B, targeted in several Phase I and Phase I/II clinical trials that have shown long-term efficacy and no toxic effects [6, 7]; and the Sanfilippo B syndrome, where gene expression and consequently improved cognitive development have been sustained for at least a year and are still ongoing (Pasteur Institute Phase I/II trial, unpublished). Moreover, alipogene tiparvovec (Glybera; uniQure), a gene therapy for lipoprotein lipase deficiency (LPLD) that employs an AAV vector, received regulatory approval by the European Medicines Agency in 2012 [8]. AAV vectors may also offer a strong potential for the treatment of cancer, and their excellent gene delivery properties have been harnessed for *in vitro* cancer studies (i.e. cultured cells), *in vivo* pre-clinical cancer models (i.e. animal models of cancer), and cancer clinical trials under development [9].

1.3 AAV Biology

Adeno-associated viruses are non-enveloped, single-stranded DNA viruses that replicate only in the presence of a helper virus, primarily adenovirus [2]. The AAV viral genome is 4.7 kilobases and contains two inverted terminal repeats (ITRs) that flank the viral genes *rep* and *cap* [2]. Multiple proteins are produced from each viral gene through the use of alternative splicing and start codons (Figure 1.1). The *rep* open reading frame (ORF) encodes four proteins necessary for AAV replication, site-specific integration, transcriptional regulation of viral promoters, and viral assembly. The *cap* ORF serves as a template for the production of three structural proteins, VP1-3, that assemble to form a 60-mer viral capsid approximately 25 nm in diameter (Figure 1.2) [2]. Finally, an alternative ORF nested in *cap* encodes assembly-activating protein (AAP), which interacts with the capsid proteins VP1-3 and is necessary for capsid assembly [10].

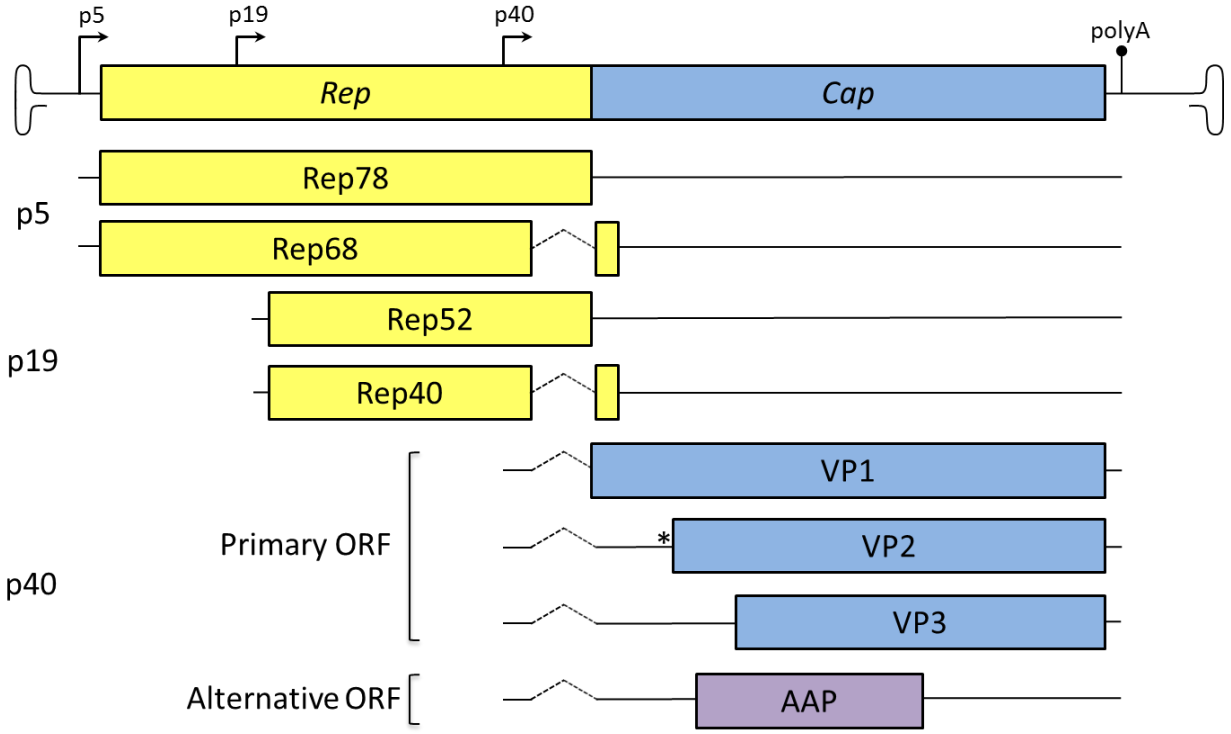


Figure 1.1. Key features of the AAV genome. Through alternative splicing, *rep* encodes four proteins – Rep78, Rep68, Rep52, and Rep40 – that are involved in viral genome replication. The *cap* open reading frame (ORF) encodes three capsid proteins, VP1-3, that assemble to form a 60-mer viral capsid. There are two splice forms of the *cap* mRNA, where the longer form encodes VP1. In the shorter splice form the start codon for VP1 is removed, and translation initiates either from the VP3 start codon or an alternative ACG* start codon for VP2. The protein AAP, which assists in viral capsid assembly, is encoded by an alternative open reading frame with a nonconventional CUG start codon. Gene expression is driven by the p5, p19, or p40 promoter as indicated. VP1; viral capsid protein 1, VP2; viral capsid protein 2, VP3; viral capsid protein 3, AAP; assembly-activating protein.

Vectors based on AAV are particularly promising gene delivery vehicles in large part because they exhibit low immunogenicity, can mediate long-term gene expression in both dividing and non-dividing cells, and have a low risk of insertional mutagenesis [11]. To generate a

recombinant AAV for gene delivery, the viral genome is removed and replaced with the therapeutic transgene and a promoter. The viral ITRs are the only *cis* elements needed to ensure packaging of the genetic payload into particles, and *rep* and *cap* can be supplied separately during virus production [12]. Infection with recombinant AAV vectors results in primarily episomal transgene expression that is persistent in non-dividing cells for at least 10 years [13]. AAV has been used in an increasing number of clinical trials, including recent successes for hemophilia B and the retinal degenerative disorder Leber's congenital amaurosis type 2 (LCA2) [4, 6, 14, 15], as well as efforts that led to the first approval of a gene therapy product (for lipoprotein lipase deficiency) in Western nations late last year [16].

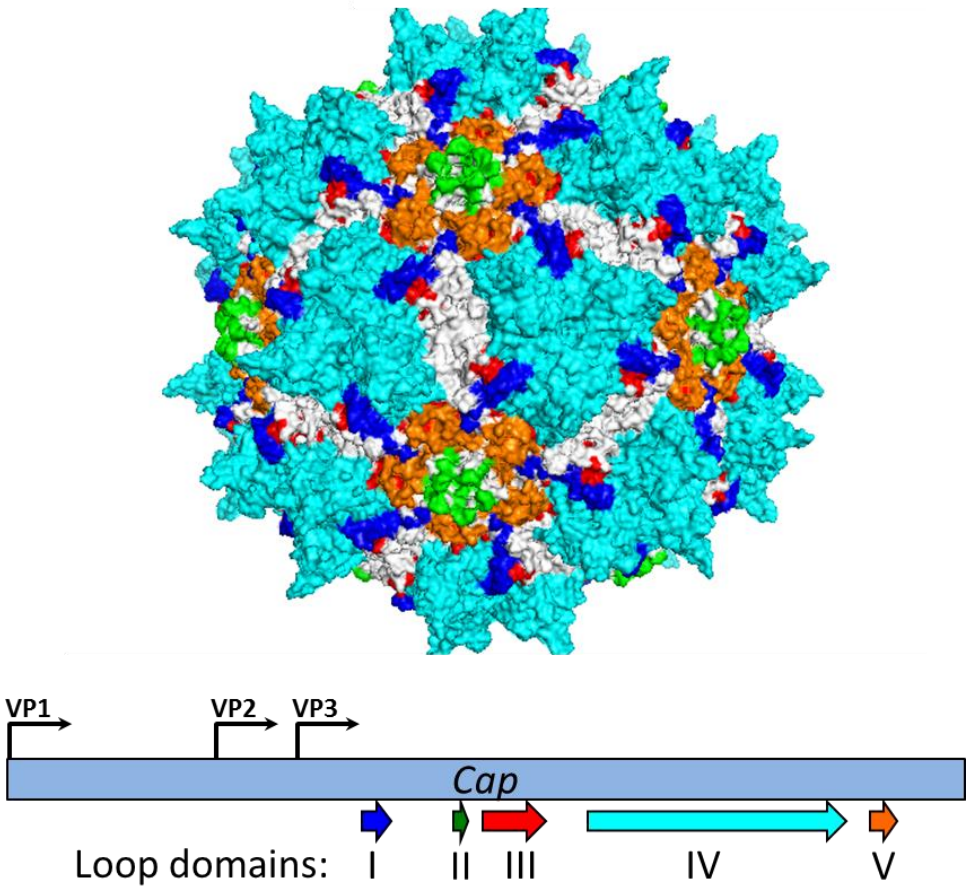


Figure 1.2. AAV capsid structure. The locations of interstrand loop domains are indicated by the colored arrows on the *cap* gene and mapped to the three-dimensional AAV capsid structure, and these loop domains are highly diversified regions of *cap* that play a role in many virus-host interactions. The right-angle arrows indicate the translation start sites for viral proteins 1-3. This image was generated in Pymol [17] based on the crystal structure coordinates of AAV2 (Protein data bank accession number 1LP3).

1.4 Targeting the Central Nervous System

Because recombinant AAV vectors are well-suited for gene therapy in the central nervous system, where most cells are post-mitotic and many chronic neurological diseases necessitate long-term transgene expression, AAV has entered into numerous CNS clinical studies (Table 1.1) [18].

These trials have highlighted clinical challenges to successful gene therapy in the CNS. Biological transport barriers limit viral access to the brain parenchyma. For example, the blood-brain barrier (BBB) prevents global transduction of the CNS following intravascular vector administration, and clinical trials to date have thus employed intracranial injections that result primarily in localized transduction due to limited vector dispersion in the brain parenchyma. In addition, the ability to target specific cell types and regions in the CNS that are implicated in disease is important for safe and effective therapies for several indications; however, it is difficult to control AAV tropism and transgene expression. Finally, AAV vectors have an excellent safety record, though the implementation of new routes of administration and serotypes could increase the risk of an immune response to the vector and/or transgene. These challenges motivate the careful selection of natural AAV serotypes – and engineering of novel AAV variants – to yield transduction profiles that are optimally suited to specific therapeutic needs.

Disease	Clinical Trial	Serotype	Promoter	Transgene Product	ClinicalTrials.gov identifier	Refs.
Alzheimer's	Phase 2	AAV2	CAG	Nerve growth factor (NGF)	NCT00876863	[13]
Canavan	Phase 1	AAV2	NSE	Aspartoacylase (ASPA)	NA	[14]
Late infantile neuronal ceroidlipofuscinosis	Phase 1	AAV2	CAG	Human CLN2	NCT00151216	[15]
Late infantile neuronal ceroidlipofuscinosis	Phase 1/2 recruiting	AAVrh.10	CAG	Human CLN2	NCT01414985	
Parkinson's	Phase 1/2	AAV2	CAG	Neurturin (NTN)	NCT00400634 NCT00985517	[16]
Parkinson's	Phase 1 recruiting	AAV2	NA	Glial cell derived neurotrophic factor (GDNF)	NCT01621581	
Parkinson's	Phase 2	AAV2	CAG	Glutamate decarboxylase 1 (GAD)	NCT00643890	[17]
Sanfilippo syndrome (MPSIIIB)	Phase 1/2 recruiting	AAV5	NA	Human α-N-acetylglucosaminidase	ISRCTN* 19853672	[18]

CAG, chicken beta-actin promoter with CMV enhancer; CLN2, also known as tripeptidyl peptidase 1; CMV, cytomegalovirus; NA, not available; NSE, neuron specific enolase.
* International Standard Randomized Controlled Trial Number

Table 1.1. Summary of AAV clinical trials in the central nervous system.

1.5 Transduction Properties of Natural AAV Serotypes in the CNS

There are 11 naturally occurring AAV serotypes and over 100 variants of AAV [19], which were isolated from human and non-human primate tissue, and which mediate a range of different cellular transduction and vector spread profiles in the central nervous system following direct injection into brain parenchyma. Most serotypes studied to date preferentially transduce neurons after intraparenchymal injection, and AAV2 for example has particularly strong neuronal tropism [20]. However, while this vector has been favored in clinical trials due to its established safety record and historical use in the early AAV clinical studies, research has shown AAV1 and AAV5 to be more efficient in transducing neurons – and to a lesser extent some glia – in a number of rat

and non-human primate (NHP) brain regions [21]. By comparison, AAV4, one of the more phylogenetically distant AAV serotypes, preferentially transduces ependymal cells, even after intrastriatal injection [22, 23]. Finally, among more recently isolated AAVs, intracerebral injection of AAV7, 8, 9, and rh.10 results in primarily neuronal transduction, and vector spread is greatest with AAV9 and rh.10 [24].

In addition to neurons, glia represent potential targets. These cells play important supporting roles for neurons, including functions that could be enhanced via gene delivery. They are also involved in disorders such as amyotrophic lateral sclerosis and are thus potential direct disease targets [25]. Recently discovered serotypes hu.32, hu.11, pi.2, hu.48R3, and rh.8 are able to transduce both astrocytes and oligodendrocytes with varying efficiencies in the adult mouse brain [26].

In addition to cellular transduction, vector spread through tissue is a significant challenge, particularly for disorders that afflict large regions of the CNS such as lysosomal storage diseases. Vector can undergo transport either extracellularly or intracellularly. Intracellularly, axonal transport of AAV can occur in the retrograde [27] or the anterograde direction, and AAV serotypes differ in their potential for such transport [28]. AAV1, 9, and rh.10 can all be disseminated along axonal projections in both the retrograde and anterograde directions after injection into the ventral tegmental area, a region dense in efferent and afferent projections [29]. Axonal transport has been proposed as a mechanism to enhance therapeutic efficacy of AAV by protecting potentially both presynaptic neurons and their projective fields [27]. Such axonal transport can also be utilized to trace and manipulate neural circuits in the CNS [30].

The discovery that AAV9 can cross the blood-brain barrier (BBB) in both neonatal and adult mice [31] has raised the exciting possibility that intravascular (IV) AAV administration could mediate widespread CNS gene expression. IV delivery leads to both neuronal and astrocytic transduction in neonates, though expression was primarily restricted to astrocytes in adults. Since intracranial injection results in mixed astrocytic and neuronal transduction, this astrocytic tropism may be due to astrocytic endfeet interfacing with endothelial cells within the BBB. Follow-up studies have demonstrated that rAAVrh.10, rAAVrh.39, rAAVrh.43 are also capable of parenchymal cell transduction after intravascular infusion in neonatal mice [32]. One challenge associated with IV administration, however, is that the majority of the human population harbors neutralizing antibodies against one or more AAV serotypes due to natural exposure, which can severely limit IV-mediated AAV delivery [33]. Finally, intracerebroventricular (ICV) administration (see Therapeutic Routes of Administration below) may also enable AAV to gain access to large volumes of the CNS, though ICV administration of AAV2, 4, or 5 shifts tropism primarily to ependymal cells with sparse transduction in the parenchyma [22].

1.6 AAV Vector Engineering

AAV directed evolution

AAV is thus a promising gene therapy vehicle; however, human therapeutic needs demand delivery properties that at best probably conferred AAVs with no selective advantages during natural evolution (i.e. CNS infection is not a prominent feature of AAV's life cycle) and at worst may be at odds with natural selection (e.g. a primary viral receptor that is promiscuously expressed at high levels can limit broad viral dispersal within tissue [34]). As a result, there have been

considerable efforts to engineer viral capsid proteins to meet biomedical needs [2]. However, the complexity of structure-function relationships in multimeric, 4 MDa viral particles renders rational design efforts difficult. An alternative approach, directed evolution, emulates how viruses naturally evolve – iterative rounds of genetic diversification and selection for improved function – but with selective pressures that can be designed to result in therapeutically useful viruses. Directed evolution has been applied to generate new viral variants with altered gene delivery specificities and enhanced evasion of neutralizing antibodies [35-42].

Directed evolution of AAV (Figure 1.3) first involves the generation of large libraries of mutated *cap* genes, using techniques such as error-prone polymerase chain reaction (PCR), DNA shuffling, or the staggered extension process [2]. These genetic libraries are then converted into viral particles, where each particle contains a viral genome that encodes its capsid shell, thereby linking the virus' genotype and phenotype. After applying a selective pressure for a specific improved function, *cap* genes can be recovered, amplified by PCR, and used for additional rounds of selection. Iteration drives library convergence towards top performers, which are then benchmarked to the best available natural serotypes. In addition to providing novel and useful vehicles, the resulting mutations can be reverse engineered to elucidate new basic structure-function relationship information. Moreover, beneficial mutations identified in different studies may also be combined on a single vector to address multiple therapeutic needs.

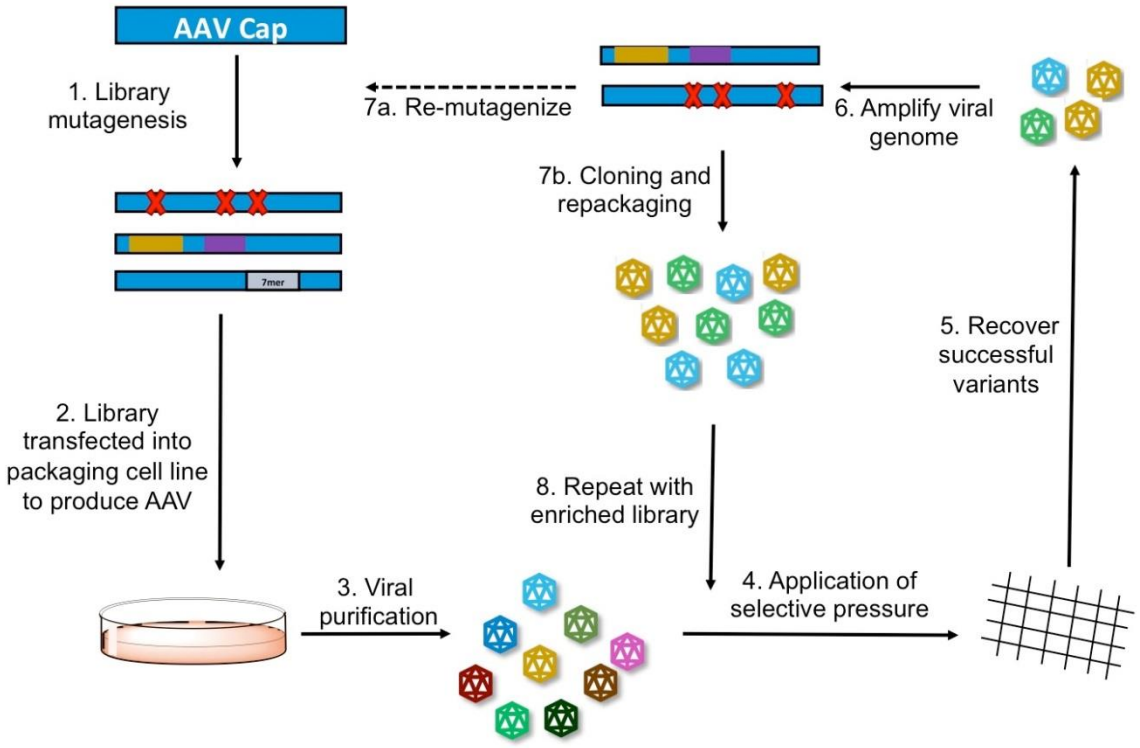


Figure 1.3. AAV directed evolution algorithm. 1. The AAV *cap* gene is mutagenized by techniques such as error prone PCR, DNA shuffling, or the staggered extension process. 2. The mutagenized library is transfected into a packaging cell line (HEK293T) to produce viral particles. 3. Viral libraries are harvested and purified. 4. A selective pressure is applied. 5. Successful variants are recovered. 6. *Cap* genes are amplified by PCR. 7a. Additional mutagenesis can be conducted to increase library diversity. 7b. The enriched library is repackaged into viral particles. 8. The process is iterated to increase viral fitness.

Numerous *in vitro* selections to target cell types of the central nervous system have been conducted. As discussed above, most AAV vectors predominantly transduce neurons in the central nervous system after intraparenchymal injection. Koerber *et al.* evolved AAV variants to infect human and rat astrocytes *in vitro* up to 15-fold more efficiently than their corresponding parent serotypes [41]. In addition, several evolved variants transduced astrocytes up to 5.5-fold more efficiently than parental serotypes in the rat striatum, despite the *in vitro* nature of the original selection. Directed evolution has also been applied to target neural stem cells (NSCs), which natural AAV serotypes do not efficiently infect. After seven rounds of selection on rat hippocampal NSCs, the variant AAV r3.45 was recovered [42]. This AAV, which included a peptide insertion on the capsid surface, was 15 to 50-fold more infectious on rat NSCs relative to AAV2 and AAV5. In addition, though it was selected on rat NSCs, AAV r3.45 also supported efficient gene delivery to both murine and human NSCs. Furthermore, the vector was harnessed to mediate gene correction by homologous recombination, an advantageous feature of AAV gene delivery [43]. Finally, this variant is selective for NSCs within the adult rodent brain (data not shown).

Glioblastoma multiforme is the most common brain cancer and has a very poor prognosis. Maguire *et al.* conducted seven rounds of selection using a shuffled AAV library selected on glioblastoma U87 cells. A resulting, chimeric capsid with elements from AAV1, 2, rh.8, rh.10, and several point mutations transduced 97% of glioblastoma U87 cells at a dose of 10^4 genome copies/cell and also outperformed the AAV2 parental serotype on multiple glioma cell lines [44]. Future *in vivo* analysis may explore the promise of such variants for gene therapy in solid tumors.

While *in vitro* selections have been the primary focus of work to date, numerous cellular and tissue complexities of the CNS cannot be emulated in culture. Consequently, several studies have shifted towards animal models that better represent the transport barriers that gene therapy vectors must overcome. For example, Gray *et al.* [45] employed directed evolution to select variants that crossed a seizure-compromised blood-brain barrier. A shuffled library of AAV serotypes 1-6, 8, and 9 was injected via tail vein into rats 24 hours after kainic acid-induced limbic seizure, and AAV variants were recovered from seizure-sensitive brain sites. After three rounds of selection, two clones primarily composed of AAV1, 8, and 9 were found to selectively transduce regions in the ventral hippocampus and piriform cortex where seizures had compromised the BBB, but they did not cross the intact BBB. The evolved vectors displayed a transduction profile similar to AAV8, infecting mostly neurons and oligodendrocytes with few astrocytes or microglia transduced. Finally, the biodistribution of evolved clones was detargeted from peripheral organs when compared with the parent serotypes AAV 1, 8, and 9.

The retina is also part of the CNS, and the retinal degenerative disorder Leber's congenital amaurosis type 2 has been successfully treated in gene therapy clinical trials [14, 15]. The majority of monogenic retinal diseases involve mutations in genes expressed in photoreceptors and the retinal pigment epithelium (RPE), which lie several hundred microns deep within the retina, and transducing these targets with existing AAV vehicles requires an injection into subretinal space between photoreceptors and RPE. The resulting transient retinal detachment can damage retinas already undergoing degeneration, and subretinal injections only transduce cells that come into contact with the "bleb" of injected liquid. Dalkara *et al.* applied directed evolution to engineer an AAV variant that can reach the outer retina after injection into the readily accessible vitreous humor [38]. Six rounds of selection in adult mice led to a dominant variant containing a seven-amino acid sequence inserted into loop 4 of the capsid. The evolved variant (7m8) mediated

widespread transduction of the outer retina and was able to rescue disease phenotypes in two mouse models of eye diseases, X-linked retinoschisis and Leber's congenital amaurosis. Finally, the vector also showed promising clinical potential in its ability to transduce photoreceptors from the vitreous in non-human primate.

Rational design of AAV vectors

In some cases where specific capsid structure-function relationships are known, rational design can be effective. For example, tyrosine residues on the capsid surface are subject to phosphorylation by tyrosine kinases, leading to capsid ubiquitination and proteasomal degradation [46]. Work by Srivastava and colleagues showing that mutation of tyrosines to phenylalanines can overcome this problem and enable more efficient gene delivery [47], and Dalkara *et al.* [48] built upon this work by introducing Y to F substitutions at two highly conserved, surface-exposed residues on the AAV9 capsid. The resulting tyrosine mutant AAV9-scCAG-GFP vector was administered by tail vein injection in neonatal mice, and transduced CNS cells included both neurons and astrocytes in the hippocampus, hypothalamus, cortex, and cerebellum – a pattern similar to but more efficient than wild type AAV9. Tyrosine mutant AAV9 vectors have also been demonstrated to significantly enhance gene delivery to the CNS after intracardiac injections in adult mice [49]. Tyrosine mutations may not substantially shift the natural tropism of parent serotypes; however, they may enable a reduction in vector dose and thereby lower the risk of an immune response.

1.7 Therapeutics Routes of Administration

Intracranial

To date, intracranial administration of AAV – which involves insertion of flexible fused-silica infusion catheters through burr holes into the brain parenchyma, followed by slow infusion of vector [50] – has been the most commonly employed route for gene delivery to the brain parenchyma [51-56]. This approach circumvents the biological transport barriers that render other administration routes challenging (Figure 1.4). In addition, this route of administration does reduce the risk of vector neutralization by circulating antibodies, though anti-AAV neutralizing antibody titers in the brain parenchyma can reach 1% of levels found in systemic circulation [57].

Intracranial injections do have significant drawbacks. As described above (AAV Vector Design), poor vector spread limits transgene expression to the vicinity of the injection site, a major shortcoming for diseases that affect large CNS regions such as Parkinson's or Alzheimer's diseases, or the entire CNS such as lysosomal storage disorders. It is estimated that complete transduction of the entire human infant brain for treatment of lysosomal storage disorders would require 50-350 injection tracts based on the limited diffusion distance (~1-3 mm) of AAV vectors [58, 59]. Each injection presents a risk of hemorrhaging, edema, and bacterial contamination. The spread of AAV vector throughout the brain parenchyma can be improved by convection-enhanced delivery (CED) [59, 60], which increases vector transport through the interstitial fluid by inducing convective flow in addition to diffusion. Since bulk flow depends only on the pressure gradient, the injection pressure is maintained at a sufficient level to overcome the hydrostatic pressure of the interstitial fluid and thereby distribute the vector throughout the brain, though flow rates should be conservative to avoid uncontrolled vector spread along paths of least resistance such as white

matter tracts [61]. CED has been used safely in a clinical trial for Parkinson's disease [62], and the development of new cannula designs to prevent reflux [63] and sophisticated magnetic resonance imaging (MRI) guidance systems to ensure accurate cannula placement [64, 65] will further improve safety and efficacy.

Intravascular

Intravascular (IV) administration in principle offers the potential for noninvasive transduction of the entire brain with a single vector infusion, given the high density of CNS capillaries [66]. However, AAV transcytosis to the brain parenchyma is obviously limited by the blood-brain barrier, whose tight junctions between endothelial cells preclude paracellular transport of AAV. Serotypes such as AAV9 that do cross the BBB are thought to undergo receptor-mediated transcytosis in endothelial cells [67].

There are, however, several disadvantages of IV delivery: the vector circulates throughout the entire body where it is exposed to circulating antibodies and can transduce peripheral organs in addition to the CNS. The latter both represents a loss of vector, can lead to off-target side effects, and increases the risk of an immune response. In addition, due to the inefficiency in crossing the BBB, only a small fraction of injected virus reaches the brain and spinal cord, necessitating in principle high vector doses on the order of 10^{15} viral genomes to treat an adult human [68]. Moreover, transduction is primarily limited to astrocytes in adult organisms as previously discussed (AAV vector design), though vector engineering may broaden tropism.

Intra-cerebrospinal fluid

Delivery to the cerebrospinal fluid (CSF) places vector near the CNS parenchyma, has potential to reduce peripheral off-target transduction, and limits exposure to serum neutralizing antibodies. That said, analogous to the BBB, tight junctions between ependymal cells limit the efficiency of vector penetration into the brain parenchyma.

Samaranch *et al.* [68] demonstrated that AAV9 infusion into the cisterna magna of non-human primates promotes significantly stronger transgene expression throughout the cortex and cerebellum compared to intravascular delivery. Transgene expression was observed primarily in astrocytes, scattered pyramidal neurons, and almost no microglia or oligodendrocytes. Moderate serum titers of pre-existing anti-AAV antibodies (1:200) did prevent brain transduction, indicating that delivery to the CSF did not offer complete immunological protection.

Gray *et al.* [69] compared the transduction profiles of AAV2.5 [70] and AAV9 after injection into the cisterna magna and lumbar intrathecal space in non-human primates. Both vectors achieved widespread transduction of neurons and astrocytes in the brain and spinal cord, and intrathecal injections were more effective than intracisternal injections in promoting gene delivery to the dorsal root ganglia. Cross-reactivity of neutralizing antibodies between AAV2.5 and AAV9 was not observed in two of four monkeys tested, supporting the possibility of serotype switching for a second administration of vector. Furthermore, circulating neutralizing antibody titers up to 1:128, which prevent gene delivery after IV administration [71], had no inhibitory effect on CNS gene transfer.

Retrograde transport

Non-invasive gene delivery to the CNS can also be accomplished via vector administration to peripheral muscle tissue and retrograde transport along motor neuron projections to cell bodies residing in the CNS [28]. This approach is particularly relevant for diseases such as amyotrophic lateral sclerosis and spinal muscular atrophy that primarily afflict motor neurons. Hollis II *et al.* investigated the efficiency of retrograde transport of AAV serotypes 1-6 after peripheral injection into either extensor carpi muscle or sciatic nerve [72]. AAV1 performed best in retrograde infection of lower motor neurons (1-4.1% of all motor neurons were transduced) after both intramuscular and intranerve injection, and the latter was more efficient.

AAV8 is also capable of retrograde transport in both neonatal and adult mice [73, 74]. In adult mice AAV8 does not cross the blood-nerve barrier, limiting systemic dissemination to peripheral organs [74]. In another study AAV9 was shown to undergo retrograde transport after injection into the gastrocnemius muscle of adult mice, transducing both neurons and astrocytes equally well with up to 43% of total motor neurons transduced [75]. Spread of vector from injected muscle to the CNS and peripheral organs was also observed, likely due to the ability of AAV9 to cross the blood-nerve barrier.

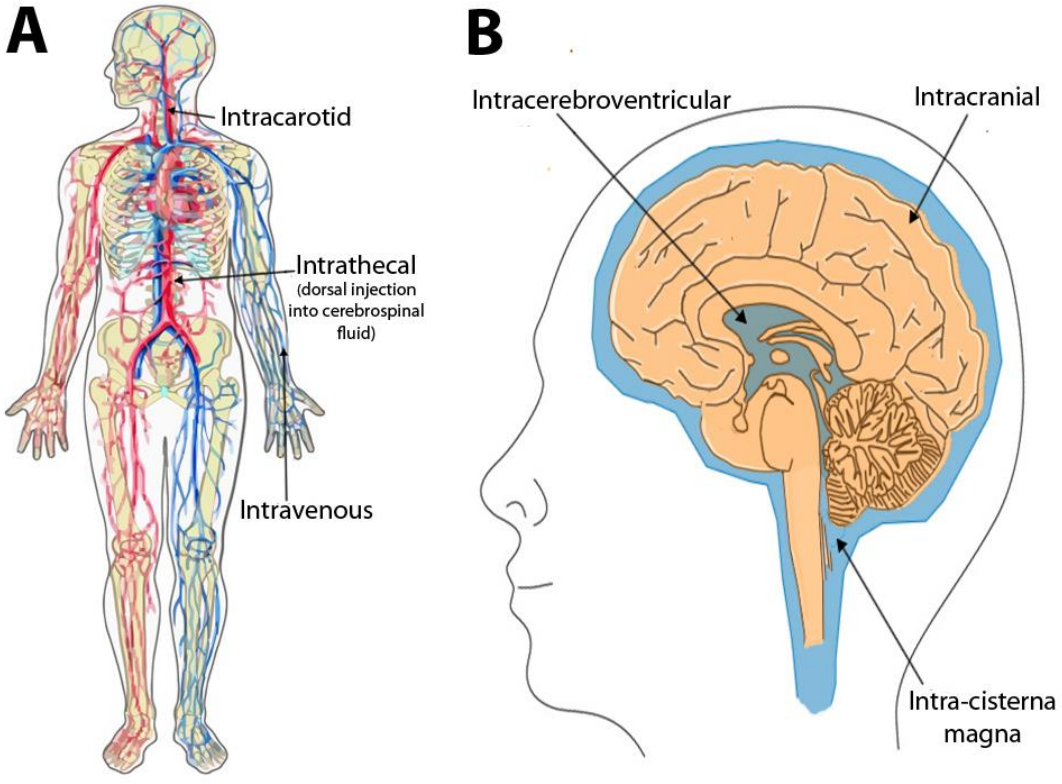


Figure 1.4. Therapeutic routes of administration. Figure 1.4A illustrates intracarotid and intravenous injections into the bloodstream. The intrathecal route is a dorsal injection into the cerebrospinal fluid. Figure 1.4B shows the more invasive intracranial, intra-cisterna magna, and intracerebroventricular injections into the cerebrospinal fluid (blue) or brain tissue. Illustrations adapted from ChemBioDraw [76].

1.8 Immune Responses to AAV in the CNS

Recombinant AAV vectors do not encode viral genes, but immune responses can be mounted against the viral capsid and/or the transgene product. For example, pre-existing neutralizing antibodies generated from a prior exposure to AAV can opsonize and inactivate the vector. The extent of vector neutralization depends on the route of administration. Intravascular routes that expose the vector to circulating neutralizing antibodies are sensitive to low antibody titers [71], whereas relatively high titers of circulating anti-AAV neutralizing antibodies do not appear to significantly impact transduction after intracranial AAV administration in immune-primed mice [57]. Phase I clinical trials employing intracranial delivery of recombinant AAV have confirmed preclinical results, showing a minimal humoral response and limited adaptive immune response [52, 77].

After viral transduction and the onset of transgene expression, immune responses to the transgene product can lead to clearing of transduced cells with subsequent loss of gene expression and inflammation in the CNS. Most CNS gene therapy trials have employed AAV2, a vector with neuronal tropism, but newer vectors with broader tropism such as AAV9 may increase the likelihood of an immune response by infecting antigen-presenting cells (APCs). Ciesielska *et al*. [78] administered AAV9 encoding human aromatic L-amino acid decarboxylase (hAADC) or green fluorescent protein into rat striatum or thalamus. Both vectors provoked a full immune response, with upregulation of MHC II in glia, lymphocytic infiltration, and inflammation leading to significant loss of neurons and generation of antibodies against the transgenes. A significant fraction of the MHC II positive cells were identified as microglia 1 week and 3 weeks after injection, and a smaller population of MHC II positive astrocytes was also observed at the 3-week time point. In contrast to AAV9, AAV2 mediated delivery of hAADC has elicited no safety concerns in human or animal studies [62, 79]. The authors speculated that transduction of antigen-presenting cells and subsequent presentation of transgene antigen led to a full adaptive immune response. Cell-mediated responses have also been observed after intracerebral infusion of an AAV1-GFP vector in NHP [80] and intracranial injection of AAV5-α-l-iduronidase in a dog model of Hurler syndrome [81]. Therefore, immune responses to non-self proteins expressed from AAV vectors may limit therapeutic options for patients that lack immune tolerance due to null mutations in the endogenous gene. In light of these findings, engineered vectors that both overcome CNS transport barriers and do not infect APCs are needed.

Strategies for limiting off-target transduction

Some natural serotypes are selective for specific cell types, and directed evolution may be used to fundamentally re-engineer cell tropism [82]. In addition, a cell type specific promoter can also be employed to restrict expression to specific transduced cells. One challenge is that such a promoter must be sufficiently small for it and the transgene to fit within the limited carrying capacity of the virus, roughly 5 kb ssDNA or 2.5 kb self-complementary dsDNA [83]. Neuron selective promoters that have been explored include synapsin-1, neuron-specific enolase (NSE), and human platelet-derived growth factor (PDGF) [84]. In addition, the 2.5 kb rat tyrosine

hydroxylase promoter has been used to drive transgene expression in midbrain dopamine neurons [85]. Likewise, expression can be restricted to astrocytes and oligodendrocytes using glial fibrillary acidic protein (GFAP) or myelin basic protein (MBP) promoters, respectively [86]. Ideally, a matching pair of cell type specific promoter and viral capsid would be developed for each disease target.

An alternate approach to reduce transgene expression in off-target tissues is the use of microRNA (miRNA) elements. MicroRNAs are non-coding regulatory RNAs that contribute to post-transcriptional gene silencing. miRNA target sequences matching miRNAs expressed only in peripheral organs can be introduced into AAV expression cassettes to reduce transgene mRNA levels in off-target tissues. For example, Qiao *et al.* [87] demonstrated that five copies of liver-specific miR-122 in the 3' untranslated region of AAV9 vectors reduced expression of the reporter genes luciferase and β-galactosidase by 50-fold and 70-fold, respectively, in liver tissue. Expression was not reduced in cardiac and muscle tissues, demonstrating the specificity of silencing. An analogous study [88] incorporated three copies of the same miR-122 element in AAV9 and provided additional evidence of reduced transgene expression in both hepatic cell lines and mouse liver. The discovery and cataloging of new miRNA elements [89] will further expand applications of this technology.

1.9 Gene Therapy for Neurological Disorders

Lysosomal storage disorders

Lysosomal storage disorders are a family of inherited diseases involving deficiencies in enzymes that metabolize lipids, glycoproteins, or mucopolysaccharides. These deficiencies lead to the accumulation of undigested macromolecules in lysosomes, resulting in disruption of cellular function and clinical manifestations. Over 50 lysosomal storage disorders have been identified and collectively affect approximately 1 in 7,700 births (Table 1.2) [90]. Enzyme replacement therapies have been approved for the treatment of LSDs [91] but are ineffective in the CNS since LSD enzymes do not cross the BBB. Given the short half-life of LSD enzymes, repeat intracranial infusions of enzyme would be necessary to achieve a therapeutic effect. In contrast, AAV vectors can provide sustained expression of LSD enzymes with a single vector dose. Moreover, cross-correction of non-transduced cells is possible since many LSD enzymes can be secreted and internalized by neighboring cells.

Disease	Mutated Gene	Size of Coding Region (bp)*	Disease Prevalence**
MPSI-Hurler	Iduronidase, alpha-L (IDUA)	1,962	1:100,000 (severe); 1:500,000 (attenuated)
MPSII-Hunter	Iduronate 2-sulfatase (IDS)	1032	1:100,000 males
MPSIIIB-Sanfilippo type B	N-acetylglucosaminidase, alpha (NAGLU)	2,231	1:200,000
MPSVII-Sly	Glucuronidase, beta (GUSB)	1,956	1:250,000
Neuronal ceroid lipofuscinosis (Batten)*	Ceroid lipofuscinosis neuronal 2 gene (CLN2)	1,669	1:25,000
Tay-Sachs	Hexosaminidase A, alpha polypeptide (HEXA)	1,590	1:3500 (Ashkenazi Jewish population); 1:320,000 (general population)

*PubMed nucleotide
**Estimates from U.S. National Library of Medicine genetics home reference

Table 1.2. Several lysosomal storage disorders affecting the central nervous system.

Late infantile neuronal ceroid lipofuscinosis (LINCL, also known as Batten disease) is caused by mutations in the ceroid lipofuscinosis neuronal 2 gene (CLN2). A phase 1 clinical trial for LINCL, involving AAV2 vector delivery of the CLN2 gene to ten children via intracranial injection, has been completed [53]. Gene delivery yielded a statistically significant slowing of disease progression as measured by a clinical rating scale. A secondary variable, neuroimaging results based on quantitative MRI parameters, was suggestive of improvement but did not yield a statistically significant change relative to the control group. The most common serious adverse events included seizures and the loss of one subject with severe LINCL 49 days after administration following development of status epilepticus. However, none of these adverse events were unequivocally attributed to the vector.

Alternative administration routes and vectors with increased spread (AAVrh.10 [92]) are being explored to achieve the robust CNS transduction needed for correction of LSDs. Haurigot *et al.* studied [93] the impact of intra-CSF delivery of AAV9 vectors encoding sulfamidase in a mouse model of MPS IIIA (Sanfilippo syndrome type A). A high dose (5×10^{10} vg per adult mouse) mediated sulfamidase activity in all brain regions with up to 11% to 36% of normal expression levels in males and 7% to 39% of expression levels in females. Gene delivery resulted in reduced glycosaminoglycan accumulation in most brain regions, correction of behavioral responses, and extended survival. Intracisternal administration of the same vector in dogs resulted in primarily neuronal with some scattered astrocytic transduction, though approximately 3-5% of hepatocytes were also transduced, indicating that the vector exited the CSF. Expression of canine sulfamidase was sustained for weeks, but delivery of human sulfamidase caused expression to

peak after 3 weeks and then decline, implicating an immune response to non-self protein in the dog. As intracisternal injections are not favored in pediatric patients, authors also evaluated intracerebroventricular administration, which resulted in widespread AAV9 vector distribution and transgene expression comparable to the intracisternal route. Serum antibody titers rose rapidly after vector exposure but remained low in the CSF (<1:10) in the absence of severe CNS inflammation. Treatment of dogs with pre-existing immunity against AAV resulted in moderate gene expression in the CNS and severely reduced transduction of peripheral tissues, likely because CSF antibody titers (1:1 to 1:3.1) were much lower than in the periphery (1:1000).

Amyotrophic lateral sclerosis

Amyotrophic lateral sclerosis (ALS) is a progressive neurodegenerative disease characterized by the death of motor neurons. Of familial ALS cases (which collectively account for 10% of ALS), the most common inheritance pattern is autosomal dominant, and several genes have been implicated. Approximately 20% of these familial cases are traced to mutations in superoxide dismutase 1 (SOD1) [94]. The mechanism of SOD1 toxicity is controversial, but it is thought to involve misfolded SOD1 aggregates.

Foust *et al.* [95] intravascularly injected AAV9 encoding SOD1 short hairpin RNA (shRNA) at P1 and P21 in a mouse model of ALS. Transgene expression was robust in astrocytes (P1: 34 ± 2%, P21 54 ± 3%) and motor neurons (P1: 62 ± 1%, P21 8 ± 1%) and persisted throughout the lifespan of the mice. One shRNA mediated 60% and 45% reductions in mutant SOD1 protein in P1 and P21 injected mice, and it extended survival times by 39% when treatment was initiated at birth, one of the longest extensions of survival reported in this mouse model. In addition, vector administration was impressively able to slow disease progression even after disease onset (injection at P85). The same vector administered via intrathecal injection in cynomolgus macaques led to widespread gene expression in both neurons and astrocytes in the grey and white matter of the spinal cord. The percentage of ChAT+ motor neurons expressing the transgene was 50% in the cervical region, 65% in the thoracic region, and 80% in the lumbar region. SOD1 knockdown matched this pattern, with a 60% decrease in SOD1 mRNA in the cervical region, 70% decrease in thoracic region, and 88% decrease in lumbar region. These promising results provide support for human clinical trials. In another approach, Kaspar *et al.* treated a mouse model of ALS with an intramuscular injection of an AAV vector encoding insulin-like growth factor 1. Retrograde gene delivery improved motor neuron survival, delayed deterioration of motor function, and extended overall survival [28].

Neuropathic pain

AAV-mediated delivery of shRNAs has also recently been applied to treat injury-induced neuropathic pain. $Na_v1.3$ channels are upregulated in DRG sensory neurons after injuries to the nervous system [96], and are therefore potential targets for treatment of chronic pain. Samad *et al.* injected an AAV2/5 vector encoding shRNA against $Na_v1.3$ into the lumbar dorsal root ganglion of adult rats with spared nerve injury [97]. Two independent AAV-shRNA vectors were able to transduce ~45% of neurons and achieve ~50% knockdown of $Na_v1.3$ compared to a scrambled shRNA control. Importantly, off-target effects against other sodium channels were not observed.

Moreover, $Na_v1.3$ downregulation resulted in a significant, partial attenuation of mechanical allodynia, with up to a sixfold recovery of pain threshold; a promising proof of principle for gene therapy approaches to treat chronic neuropathic pain.

Parkinson's disease

Parkinson's disease is a neurodegenerative disorder characterized by loss of motor function accompanying the death of dopaminergic neurons in the substantia nigra. In contrast to monogenic disorders, the mechanisms of pathogenesis in Parkinson's disease involve multiple genes and environmental factors [98], and consequently gene therapy is not as straightforward as supplying a functional copy of a defective gene. Instead, clinical trials have utilized neurotrophic factors that prevent neuronal cell death, or alternatively neurotransmitter synthetic enzymes to modulate neuronal activity (Figure 1.5). Translation to the clinic has demonstrated a strong safety profile and is progressing towards therapy efficacy.

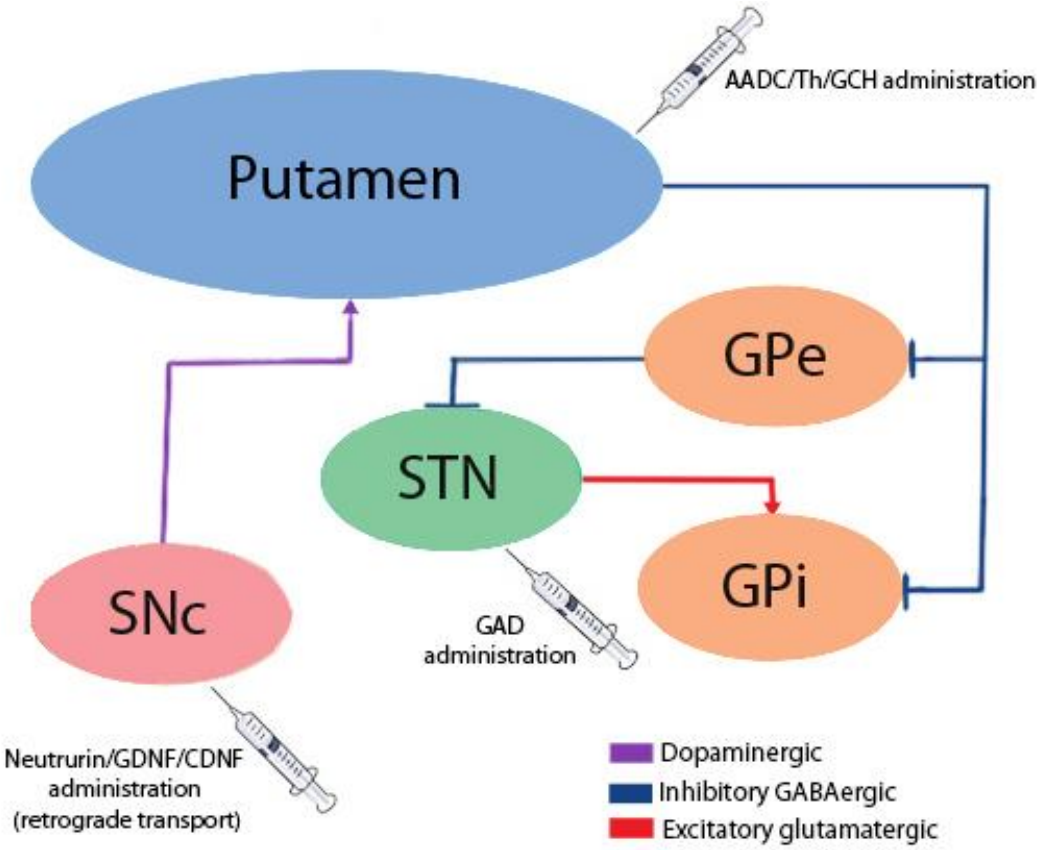


Figure 1.5. Therapeutic strategies for Parkinson's disease. Delivery of neurotrophic factors to the substantia nigra compacta (SNc) for retrograde delivery to the putamen. Gene therapy using glutamic acid decarboxylase (GAD) to quiet neurons in the subthalamic nucleus (STN). Delivery of neurotransmitter synthetic enzymes involved in dopamine production to the putamen. AADC, aromatic L-amino acid decarboxylase; Th, tyrosine hydroxylase; GCH, guanosine triphosphate cyclohydrolase; GDNF, glial derived neurotrophic factor; CDNF, cerebral derived neurotrophic factor, GPe, globus pallidus external; GPi, globus pallidus internal. Illustrations adapted from ChemBioDraw [76].

In a clinical trial conducted by Ceregene, delivery of CERE-120 – an AAV2 vector encoding the neurotrophic factor neurturin under a constitutive CAG promoter – to the putamen resulted in an excellent safety profile in a phase 1 clinical trial [54]. A multicenter, double-blind, sham surgery controlled phase 2 trial strengthened claims of vector safety but failed to yield a statistically significant benefit on the primary endpoint, the motor-off component of the Unified Parkinson's Disease Rating Scale (UPDRS) evaluated after 12 months. Some clinical benefits were evident: 19 of 25 efficacy endpoints were favorable compared with the sham control. Autopsy analysis of two subjects (who died due to unrelated causes) confirmed neurturin expression in the putamen with co-localized tyrosine hydroxylase. Interestingly, neurturin staining was sparse in the substantia nigra, indicating that retrograde transport of the AAV vector and neurturin was inefficient. While unexpected based on preclinical studies in animal models, this result may indicate that Parkinson's disease in humans is characterized by accelerated degeneration of axonal transport capabilities well before the death of the neuron [99].

Based on these results Ceregene initiated a phase 1/2b study incorporating direct injection of the substantia nigra and a 4-fold increased dosage to the putamen. Given the strong placebo effect observed in the first trial, the time point for evaluation of the primary endpoint was increased to 15 months. Phase 1/2b results indicated no significant safety issues with gene delivery to the substantia nigra. Despite these modifications, initial phase 2b efficacy results did not demonstrate a statistically significant benefit on the primary endpoint [100]. Similar to the first trial, a number of secondary endpoints produced statistically significant benefits, and a strong placebo effect was observed in the sham surgery control group. Additional analysis identified a better response on the primary endpoint in patients diagnosed 5 years prior to treatment compared with those diagnosed 10 years prior [101]. Although the efficacy results were disappointing, the field should be encouraged by the exceptional safety profile and progress in vector delivery.

Other neurotrophic factors are being considered to treat Parkinson's disease. While a phase 1/2 double-blind trial with direct daily infusion of glial derived neurotrophic factor (GDNF) protein into the putamen did not produce a statistically significant benefit [102], an AAV-GDNF vector may perform better than repeated GDNF infusions. A phase 1 clinical trial employing an AAV2-GDNF vector is currently recruiting [103]. In addition, multiple groups have shown that administration of cerebral dopamine neurotrophic factor (CDNF) to the striatum can prevent deterioration of midbrain dopamine neurons in a rat model of Parkinson's disease [104, 105], indicating the promise of this molecule.

A different therapeutic strategy is to directly target the enzymatic pathway for dopamine production. L-dopa administration is currently the most effective treatment for relieving symptoms of Parkinson's disease; however, most patients experience a decrease in therapeutic efficacy after taking the medication for several years. One possible reason for such diminishing returns is reduced levels of aromatic L-amino acid decarboxylase (AADC), the enzyme that converts L-dopa to dopamine, and gene delivery of AADC could thus improve the response to L-dopa. Furthermore, the degree of therapy can be controlled by modifying the dosing regimen of L-dopa. Preclinical studies in parkinsonian monkeys demonstrated increased L-dopa conversion lasting for over 7 years after convection-enhanced delivery of AAV2-AADC into the striatum [106]. Furthermore, an open label phase 1 clinical trial utilizing an AAV2 vector for AADC delivery to the putamen improved the mean UPDRS rating scale score by 30% in the on and off states [62, 79]. Given the

strong placebo effect observed in the CERE-120 trials, double-blinded sham surgery controls will likely be included in future trials.

Additional enzymes involved in dopamine production include tyrosine hydroxylase (TH), the rate-limiting enzyme for conversion of L-tyrosine to L-dopa, and guanosine triphosphate cyclohydrolase I (GCH) is the rate-limiting enzyme for production of the TH co-factor tetrahydrobiopterine. These genes have not yet been delivered in a clinical setting with AAV, but preclinical results in rat models are promising [107, 108]. Most studies have delivered each gene in a separate virus, but a new vector designed to co-express both TH and GCH1 from a single virus [109] provided a dose-dependent functional recovery based on enhanced dopamine production in 6-OHDA lesioned rats. Unexpectedly, the same vector administered to parkinsonian monkeys resulted in transgene expression of GCH but not TH for unknown reasons.

Another approach involves delivery of the gene glutamic acid decarboxylase (GAD) – the rate-limiting enzyme for GABA production – into the subthalamic nucleus (STN). Increases in GABA production reduce output from the STN, an overactive region in Parkinson's disease. In a completed double-blind sham surgery controlled clinical trial, administration of AAV2-GAD vector to the STN via intracranial injection was well-tolerated and resulted in improvements in the UPDRS motor score of 23.1% in patients receiving the vector compared to 12.7% in the sham control group after 6 months, the first statistically significant difference in the primary endpoint of a double-blind phase 2 trial employing AAV to treat Parkinson's disease [55]. Despite benefit in the primary endpoint, most quality of life measures did not improve, and as a result the trial sponsor (Neurologix) was unable to proceed to a phase 3 trial.

1.10 Conclusion

Promising clinical trials and the first gene therapy product approval underscore the exciting potential of gene delivery using AAV. Recombinant AAV vectors can provide long-lasting gene expression to treat chronic neurological disorders and have demonstrated a strong safety profile, making them promising for both monogenic and idiopathic CNS disease. That said, clinical translation in the CNS has been hindered by biological transport barriers, immune responses, and infection of off-target cells. However, these challenges have motivated the development of engineered vectors to further and fully realize the potential of gene therapy in the CNS.

1.11 References

1. Ginn, S.L., et al., *Gene therapy clinical trials worldwide to 2012 - an update.* J Gene Med, 2013. **15**(2): p. 65-77.
2. Schaffer, D.V., J.T. Koerber, and K.I. Lim, *Molecular engineering of viral gene delivery vehicles.* Annu Rev Biomed Eng, 2008. **10**: p. 169-94.
3. Kotterman, M.A. and D.V. Schaffer, *Engineering adeno-associated viruses for clinical gene therapy.* Nat Rev Genet, 2014. **15**(7): p. 445-51.
4. Testa, F., et al., *Three-year follow-up after unilateral subretinal delivery of adeno-associated virus in patients with Leber congenital Amaurosis type 2.* Ophthalmology, 2013. **120**(6): p. 1283-91.

5. Dalkara, D. and J.A. Sahel, *Gene therapy for inherited retinal degenerations.* C R Biol, 2014. **337**(3): p. 185-92.
6. Nathwani, A.C., et al., *Adenovirus-associated virus vector-mediated gene transfer in hemophilia B.* N Engl J Med, 2011. **365**(25): p. 2357-65.
7. Ohmori, T., et al., *New approaches to gene and cell therapy for hemophilia.* J Thromb Haemost, 2015. **13 Suppl 1**: p. S133-42.
8. Carpentier, A.C., et al., *Effect of alipogene tiparvovec (AAV1-LPL(S447X)) on postprandial chylomicron metabolism in lipoprotein lipase-deficient patients.* J Clin Endocrinol Metab, 2012. **97**(5): p. 1635-44.
9. Santiago-Ortiz, J.L. and D.V. Schaffer, *Adeno-associated virus (AAV) vectors in cancer gene therapy.* J Control Release, 2016.
10. Sonntag, F., K. Schmidt, and J.A. Kleinschmidt, *A viral assembly factor promotes AAV2 capsid formation in the nucleolus.* Proc Natl Acad Sci U S A, 2010. **107**(22): p. 10220-5.
11. Kaeppel, C., et al., *A largely random AAV integration profile after LPLD gene therapy.* Nat Med, 2013. **19**(7): p. 889-91.
12. Kwon, I. and D.V. Schaffer, *Designer gene delivery vectors: molecular engineering and evolution of adeno-associated viral vectors for enhanced gene transfer.* Pharm Res, 2008. **25**(3): p. 489-99.
13. Mingozzi, F. and K.A. High, *Immune responses to AAV vectors: overcoming barriers to successful gene therapy.* Blood, 2013. **122**(1): p. 23-36.
14. Bainbridge, J.W., et al., *Effect of gene therapy on visual function in Leber's congenital amaurosis.* N Engl J Med, 2008. **358**(21): p. 2231-9.
15. Cideciyan, A.V., et al., *Human RPE65 gene therapy for Leber congenital amaurosis: persistence of early visual improvements and safety at 1 year.* Hum Gene Ther, 2009. **20**(9): p. 999-1004.
16. Miller, N., *Glybera and the future of gene therapy in the European Union.* Nat Rev Drug Discov, 2012. **11**(5): p. 419.
17. Schrodinger, LLC, *The PyMOL Molecular Graphics System, Version 1.3r1*. 2010.
18. Asokan, A., D.V. Schaffer, and R.J. Samulski, *The AAV vector toolkit: poised at the clinical crossroads.* Mol Ther, 2012. **20**(4): p. 699-708.
19. Wu, Z., A. Asokan, and R.J. Samulski, *Adeno-associated virus serotypes: vector toolkit for human gene therapy.* Mol Ther, 2006. **14**(3): p. 316-27.
20. Bartlett, J.S., R.J. Samulski, and T.J. McCown, *Selective and rapid uptake of adeno-associated virus type 2 in brain.* Hum Gene Ther, 1998. **9**(8): p. 1181-6.
21. Mandel, R.J. and C. Burger, *Clinical trials in neurological disorders using AAV vectors: promises and challenges.* Curr Opin Mol Ther, 2004. **6**(5): p. 482-90.
22. Davidson, B.L., et al., *Recombinant adeno-associated virus type 2, 4, and 5 vectors: transduction of variant cell types and regions in the mammalian central nervous system.* Proc Natl Acad Sci U S A, 2000. **97**(7): p. 3428-32.
23. Markakis, E.A., et al., *Comparative transduction efficiency of AAV vector serotypes 1-6 in the substantia nigra and striatum of the primate brain.* Mol Ther, 2010. **18**(3): p. 588-93.
24. Cearley, C.N. and J.H. Wolfe, *Transduction characteristics of adeno-associated virus vectors expressing cap serotypes 7, 8, 9, and Rh10 in the mouse brain.* Mol Ther, 2006. **13**(3): p. 528-37.

25. Ilieva, H., M. Polymenidou, and D.W. Cleveland, *Non-cell autonomous toxicity in neurodegenerative disorders: ALS and beyond.* J Cell Biol, 2009. **187**(6): p. 761-72.
26. Cearley, C.N., et al., *Expanded repertoire of AAV vector serotypes mediate unique patterns of transduction in mouse brain.* Mol Ther, 2008. **16**(10): p. 1710-8.
27. Kaspar, B.K., et al., *Targeted retrograde gene delivery for neuronal protection.* Mol Ther, 2002. **5**(1): p. 50-6.
28. Kaspar, B.K., et al., *Retrograde viral delivery of IGF-1 prolongs survival in a mouse ALS model.* Science, 2003. **301**(5634): p. 839-42.
29. Cearley, C.N. and J.H. Wolfe, *A single injection of an adeno-associated virus vector into nuclei with divergent connections results in widespread vector distribution in the brain and global correction of a neurogenetic disease.* J Neurosci, 2007. **27**(37): p. 9928-40.
30. Betley, J.N. and S.M. Sternson, *Adeno-associated viral vectors for mapping, monitoring, and manipulating neural circuits.* Hum Gene Ther, 2011. **22**(6): p. 669-77.
31. Foust, K.D., et al., *Intravascular AAV9 preferentially targets neonatal neurons and adult astrocytes.* Nat Biotechnol, 2009. **27**(1): p. 59-65.
32. Zhang, H., et al., *Several rAAV vectors efficiently cross the blood-brain barrier and transduce neurons and astrocytes in the neonatal mouse central nervous system.* Mol Ther, 2011. **19**(8): p. 1440-8.
33. Bartel, M., D. Schaffer, and H. Buning, *Enhancing the Clinical Potential of AAV Vectors by Capsid Engineering to Evade Pre-Existing Immunity.* Front Microbiol, 2011. **2**: p. 204.
34. Nguyen, J.B., et al., *Convection-enhanced delivery of AAV-2 combined with heparin increases TK gene transfer in the rat brain.* Neuroreport, 2001. **12**(9): p. 1961-4.
35. Excoffon, K.J., et al., *Directed evolution of adeno-associated virus to an infectious respiratory virus.* Proc Natl Acad Sci U S A, 2009. **106**(10): p. 3865-70.
36. Koerber, J.T., J.H. Jang, and D.V. Schaffer, *DNA shuffling of adeno-associated virus yields functionally diverse viral progeny.* Mol Ther, 2008. **16**(10): p. 1703-9.
37. Maheshri, N., et al., *Directed evolution of adeno-associated virus yields enhanced gene delivery vectors.* Nat Biotechnol, 2006. **24**(2): p. 198-204.
38. Dalkara, D., et al., *In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous.* Sci Transl Med, 2013. **5**(189): p. 189ra76.
39. Asuri, P., et al., *Directed evolution of adeno-associated virus for enhanced gene delivery and gene targeting in human pluripotent stem cells.* Mol Ther, 2012. **20**(2): p. 329-38.
40. Koerber, J.T., et al., *Construction of diverse adeno-associated viral libraries for directed evolution of enhanced gene delivery vehicles.* Nat Protoc, 2006. **1**(2): p. 701-6.
41. Koerber, J.T., et al., *Molecular evolution of adeno-associated virus for enhanced glial gene delivery.* Mol Ther, 2009. **17**(12): p. 2088-95.
42. Jang, J.H., et al., *An evolved adeno-associated viral variant enhances gene delivery and gene targeting in neural stem cells.* Mol Ther, 2011. **19**(4): p. 667-75.
43. Khan, I.F., R.K. Hirata, and D.W. Russell, *AAV-mediated gene targeting methods for human cells.* Nat Protoc, 2011. **6**(4): p. 482-501.
44. Maguire, C.A., et al., *Directed evolution of adeno-associated virus for glioma cell transduction.* J Neurooncol, 2010. **96**(3): p. 337-47.

45. Gray, S.J., et al., *Directed evolution of a novel adeno-associated virus (AAV) vector that crosses the seizure-compromised blood-brain barrier (BBB).* Mol Ther, 2010. **18**(3): p. 570-8.
46. Zhong, L., et al., *Tyrosine-phosphorylation of AAV2 vectors and its consequences on viral intracellular trafficking and transgene expression.* Virology, 2008. **381**(2): p. 194-202.
47. Qing, K., et al., *Role of tyrosine phosphorylation of a cellular protein in adeno-associated virus 2-mediated transgene expression.* Proc Natl Acad Sci U S A, 1997. **94**(20): p. 10879-84.
48. Dalkara, D., et al., *Enhanced gene delivery to the neonatal retina through systemic administration of tyrosine-mutated AAV9.* Gene Ther, 2012. **19**(2): p. 176-81.
49. Iida, A., et al., *Systemic delivery of tyrosine-mutant AAV vectors results in robust transduction of neurons in adult mice.* Biomed Res Int, 2013. **2013**: p. 974819.
50. Lowery, R.L. and A.K. Majewska, *Intracranial injection of adeno-associated viral vectors.* J Vis Exp, 2010(45).
51. Mandel, R.J., *CERE-110, an adeno-associated virus-based gene delivery vector expressing human nerve growth factor for the treatment of Alzheimer's disease.* Curr Opin Mol Ther, 2010. **12**(2): p. 240-7.
52. McPhee, S.W., et al., *Immune responses to AAV in a phase I study for Canavan disease.* J Gene Med, 2006. **8**(5): p. 577-88.
53. Worgall, S., et al., *Treatment of late infantile neuronal ceroid lipofuscinosis by CNS administration of a serotype 2 adeno-associated virus expressing CLN2 cDNA.* Hum Gene Ther, 2008. **19**(5): p. 463-74.
54. Bartus, R.T., et al., *Advancing neurotrophic factors as treatments for age-related neurodegenerative diseases: developing and demonstrating "clinical proof-of-concept" for AAV-neurturin (CERE-120) in Parkinson's disease.* Neurobiol Aging, 2013. **34**(1): p. 35-61.
55. LeWitt, P.A., et al., *AAV2-GAD gene therapy for advanced Parkinson's disease: a double-blind, sham-surgery controlled, randomised trial.* Lancet Neurol, 2011. **10**(4): p. 309-19.
56. Ellinwood, N.M., et al., *Safe, efficient, and reproducible gene therapy of the brain in the dog models of Sanfilippo and Hurler syndromes.* Mol Ther, 2011. **19**(2): p. 251-9.
57. Treleaven, C.M., et al., *Gene transfer to the CNS is efficacious in immune-primed mice harboring physiologically relevant titers of anti-AAV antibodies.* Mol Ther, 2012. **20**(9): p. 1713-23.
58. Vite, C.H., et al., *Adeno-associated virus vector-mediated transduction in the cat brain.* Gene Ther, 2003. **10**(22): p. 1874-81.
59. Cunningham, J., et al., *Biodistribution of adeno-associated virus type-2 in nonhuman primates after convection-enhanced delivery to brain.* Mol Ther, 2008. **16**(7): p. 1267-75.
60. Kells, A.P., et al., *Efficient gene therapy-based method for the delivery of therapeutics to primate cortex.* Proc Natl Acad Sci U S A, 2009. **106**(7): p. 2407-11.
61. Linninger, A.A., et al., *Prediction of convection-enhanced drug delivery to the human brain.* J Theor Biol, 2008. **250**(1): p. 125-38.
62. Eberling, J.L., et al., *Results from a phase I safety trial of hAADC gene therapy for Parkinson disease.* Neurology, 2008. **70**(21): p. 1980-3.

63. Krauze, M.T., et al., *Reflux-free cannula for convection-enhanced high-speed delivery of therapeutic agents.* J Neurosurg, 2005. **103**(5): p. 923-9.
64. San Sebastian, W., et al., *Safety and tolerability of magnetic resonance imaging-guided convection-enhanced delivery of AAV2-hAADC with a novel delivery platform in nonhuman primate striatum.* Hum Gene Ther, 2012. **23**(2): p. 210-7.
65. van der Bom, I.M., et al., *Frameless multimodal image guidance of localized convection-enhanced delivery of therapeutics in the brain.* J Neurointerv Surg, 2013. **5**(1): p. 69-72.
66. Pardridge, W.M., *The blood-brain barrier: bottleneck in brain drug development.* NeuroRx, 2005. **2**(1): p. 3-14.
67. Shen, S., et al., *Terminal N-linked galactose is the primary receptor for adeno-associated virus 9.* J Biol Chem, 2011. **286**(15): p. 13532-40.
68. Samaranch, L., et al., *Adeno-associated virus serotype 9 transduction in the central nervous system of nonhuman primates.* Hum Gene Ther, 2012. **23**(4): p. 382-9.
69. Gray, S.J., et al., *Global CNS gene delivery and evasion of anti-AAV-neutralizing antibodies by intrathecal AAV administration in non-human primates.* Gene Ther, 2013.
70. Bowles, D.E., et al., *Phase 1 gene therapy for Duchenne muscular dystrophy using a translational optimized AAV vector.* Mol Ther, 2012. **20**(2): p. 443-55.
71. Gray, S.J., et al., *Preclinical differences of intravascular AAV9 delivery to neurons and glia: a comparative study of adult mice and nonhuman primates.* Mol Ther, 2011. **19**(6): p. 1058-69.
72. Hollis, E.R., 2nd, et al., *Efficient retrograde neuronal transduction utilizing self-complementary AAV1.* Mol Ther, 2008. **16**(2): p. 296-301.
73. Foust, K.D., et al., *Neonatal intraperitoneal or intravenous injections of recombinant adeno-associated virus type 8 transduce dorsal root ganglia and lower motor neurons.* Hum Gene Ther, 2008. **19**(1): p. 61-70.
74. Zheng, H., et al., *Efficient retrograde transport of adeno-associated virus type 8 to spinal cord and dorsal root ganglion after vector delivery in muscle.* Hum Gene Ther, 2010. **21**(1): p. 87-97.
75. Benkhelifa-Ziyyat, S., et al., *Intramuscular scAAV9-SMN injection mediates widespread gene delivery to the spinal cord and decreases disease severity in SMA mice.* Mol Ther, 2013. **21**(2): p. 282-90.
76. ChemBioDraw. 2013, Cambridge Software.
77. Kaplitt, M.G., et al., *Safety and tolerability of gene therapy with an adeno-associated virus (AAV) borne GAD gene for Parkinson's disease: an open label, phase I trial.* Lancet, 2007. **369**(9579): p. 2097-105.
78. Ciesielska, A., et al., *Cerebral infusion of AAV9 vector-encoding non-self proteins can elicit cell-mediated immune responses.* Mol Ther, 2013. **21**(1): p. 158-66.
79. Christine, C.W., et al., *Safety and tolerability of putaminal AADC gene therapy for Parkinson disease.* Neurology, 2009. **73**(20): p. 1662-9.
80. Hadaczek, P., et al., *Transduction of nonhuman primate brain with adeno-associated virus serotype 1: vector trafficking and immune response.* Hum Gene Ther, 2009. **20**(3): p. 225-37.
81. Ciron, C., et al., *Gene therapy of the brain in the dog model of Hurler's syndrome.* Ann Neurol, 2006. **60**(2): p. 204-13.
82. Pulicherla, N., et al., *Engineering liver-detargeted AAV9 vectors for cardiac and musculoskeletal gene transfer.* Mol Ther, 2011. **19**(6): p. 1070-8.

83. Wang, Z., et al., *Rapid and highly efficient transduction by double-stranded adeno-associated virus vectors in vitro and in vivo.* Gene Ther, 2003. **10**(26): p. 2105-11.
84. Shevtsova, Z., et al., *Promoters and serotypes: targeting of adeno-associated virus vectors for gene transfer in the rat central nervous system in vitro and in vivo.* Exp Physiol, 2005. **90**(1): p. 53-9.
85. Oh, M.S., et al., *Expression of transgenes in midbrain dopamine neurons using the tyrosine hydroxylase promoter.* Gene Ther, 2009. **16**(3): p. 437-40.
86. von Jonquieres, G., et al., *Glial promoter selectivity following AAV-delivery to the immature brain.* PLoS One, 2013. **8**(6): p. e65646.
87. Qiao, C., et al., *Liver-specific microRNA-122 target sequences incorporated in AAV vectors efficiently inhibits transgene expression in the liver.* Gene Ther, 2011. **18**(4): p. 403-10.
88. Geisler, A., et al., *microRNA122-regulated transgene expression increases specificity of cardiac gene transfer upon intravenous delivery of AAV9 vectors.* Gene Ther, 2011. **18**(2): p. 199-209.
89. Kozomara, A. and S. Griffiths-Jones, *miRBase: integrating microRNA annotation and deep-sequencing data.* Nucleic Acids Res, 2011. **39**(Database issue): p. D152-7.
90. Fuller, M., P.J. Meikle, and J.J. Hopwood, *Epidemiology of lysosomal storage diseases: an overview*, in *Fabry Disease: Perspectives from 5 Years of FOS*, A. Mehta, M. Beck, and G. Sunder-Plassmann, Editors. 2006: Oxford.
91. Ohashi, T., *Enzyme replacement therapy for lysosomal storage diseases.* Pediatr Endocrinol Rev, 2012. **10 Suppl 1**: p. 26-34.
92. Sondhi, D., et al., *Long-term expression and safety of administration of AAVrh.10hCLN2 to the brain of rats and nonhuman primates for the treatment of late infantile neuronal ceroid lipofuscinosis.* Hum Gene Ther Methods, 2012. **23**(5): p. 324-35.
93. Haurigot, V., et al., *Whole body correction of mucopolysaccharidosis IIIA by intracerebrospinal fluid gene therapy.* J Clin Invest, 2013.
94. Nizzardo, M., et al., *Research advances in gene therapy approaches for the treatment of amyotrophic lateral sclerosis.* Cell Mol Life Sci, 2012. **69**(10): p. 1641-50.
95. Foust, K.D., et al., *Therapeutic AAV9-mediated Suppression of Mutant SOD1 Slows Disease Progression and Extends Survival in Models of Inherited ALS.* Mol Ther, 2013.
96. Hains, B.C., et al., *Upregulation of sodium channel Nav1.3 and functional involvement in neuronal hyperexcitability associated with central neuropathic pain after spinal cord injury.* J Neurosci, 2003. **23**(26): p. 8881-92.
97. Samad, O.A., et al., *Virus-mediated shRNA knockdown of Na(v)1.3 in rat dorsal root ganglion attenuates nerve injury-induced neuropathic pain.* Mol Ther, 2013. **21**(1): p. 49-56.
98. Shadrina, M.I., P.A. Slominsky, and S.A. Limborska, *Molecular mechanisms of pathogenesis of Parkinson's disease.* Int Rev Cell Mol Biol, 2010. **281**: p. 229-66.
99. Burke, R.E. and K. O'Malley, *Axon degeneration in Parkinson's disease.* Exp Neurol, 2013. **246**: p. 72-83.
100. Ceregene. *Ceregene reports data from Parkinson's disease phase 2b study*. April 19th, 2013; Available from: http://www.ceregene.com/press_041913.asp.
101. Ceregene. *Ceregene reports additional efficacy data from Parkinson's disease phase 2b study*. May 21st, 2013; Available from: http://www.ceregene.com/press_052113.asp.

102. Lang, A.E., et al., *Randomized controlled trial of intraputamenal glial cell line-derived neurotrophic factor infusion in Parkinson disease.* Ann Neurol, 2006. **59**(3): p. 459-66.
103. Stroke, N.I.o.N.D.a. *AAV2-GDNF for Advanced Parkinson's Disease*. 9-26-2013]; Available from: http://clinicaltrials.gov/ct2/show/NCT01621581.
104. Ren, X., et al., *AAV2-mediated striatum delivery of human CDNF prevents the deterioration of midbrain dopamine neurons in a 6-hydroxydopamine induced parkinsonian rat model.* Exp Neurol, 2013. **248**: p. 148-56.
105. Back, S., et al., *Gene therapy with AAV2-CDNF provides functional benefits in a rat model of Parkinson's disease.* Brain Behav, 2013. **3**(2): p. 75-88.
106. Hadaczek, P., et al., *Convection-enhanced delivery of adeno-associated virus type 2 (AAV2) into the striatum and transport of AAV2 within monkey brain.* Hum Gene Ther, 2006. **17**(3): p. 291-302.
107. Bjorklund, T., et al., *Optimized adeno-associated viral vector-mediated striatal DOPA delivery restores sensorimotor function and prevents dyskinesias in a model of advanced Parkinson's disease.* Brain, 2010. **133**(Pt 2): p. 496-511.
108. Carlsson, T., et al., *Reversal of dyskinesias in an animal model of Parkinson's disease by continuous L-DOPA delivery using rAAV vectors.* Brain, 2005. **128**(Pt 3): p. 559-69.
109. Cederfjall, E., et al., *Continuous DOPA synthesis from a single AAV: dosing and efficacy in models of Parkinson's disease.* Sci Rep, 2013. **3**: p. 2157.

Chapter 2: In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the Subventricular Zone

2.1 Introduction

Recombinant 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 gene 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]. However, efficient clinical translation to a broader range of cell and tissue targets has been limited by delivery challenges including pre-existing neutralizing antibodies against the AAV capsid, poor transduction of clinically important cell types, infection of off-target tissues, and the need for high vector doses. In particular, transduction of the central nervous system 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 central nervous system [6-10] and other tissues [11-14].

Optimization of the AAV directed evolution method begins with design of the capsid library. 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 disrupting 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 (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 protein families as structurally complex as a multimeric AAV capsid. 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]. In addition, 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 [23]. Importantly, the extension of SCHEMA to multimeric proteins affords opportunities to engineer proteins with modular structures and/or complex regulatory domains [24, 25]. 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. We constructed a SCHEMA AAV capsid library comprised of six parent serotypes (AAV2, AAV4, AAV5, AAV6, AAV8, AAV9) and seven crossover positions, yielding a total library size of over 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 of the subventricular zone, a cell type of clinical importance that is poorly transduced by natural AAV serotypes.

The subventricular zone (SVZ) of the lateral ventricles is the largest germinal region in the adult mammalian brain [26]. Resident adult neural stem cells (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. *In vitro* NSC culture offers opportunities for high-throughput screening of small molecule libraries that modulate cell proliferation, differentiation, and toxicity [29, 30], but 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 neural stem cell 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 subventricular zone 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.

2.2 Results

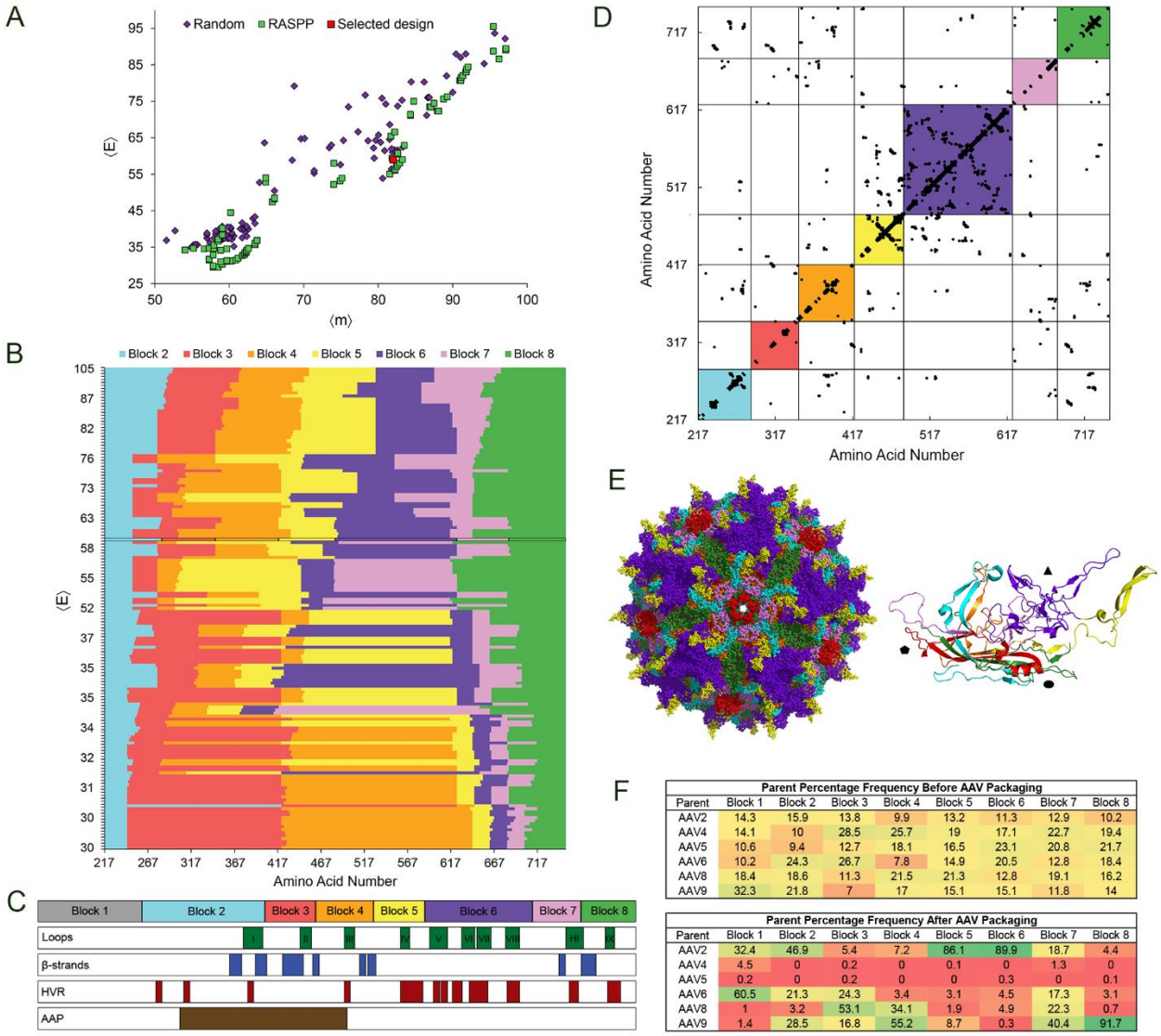
We designed a chimeric AAV library that recombines six natural serotypes – AAV2, 4, 5, 6, 8, and 9 – that represent multiple phylogenetic clades [33], have diverse receptor binding properties [1], and have enjoyed relative success in the clinic [2]. We used the capsid crystal structures to calculate contacting intra- and inter-subunit residue positions, wherein a contact was defined as two residues containing non-hydrogen atoms within a spatial cutoff of 4.5 angstroms. The final contact map contained residue pairs that were contacting in at least 50% of these six parent structures. 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 the six parent serotypes yields a theoretical library diversity of over 1.6 million (6^8) chimeric variants. We additionally modified the SCHEMA algorithm to search for crossover locations that were amenable to combinatorial golden gate assembly for library construction, which requires four nucleotide 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 four nucleotide 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). 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 (Figure 1A-C) 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 chimeras with new and interesting functions. For example, significantly lower disruption scores could be achieved by combining blocks five and six, 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 since it provided a relatively even distribution of block sizes. We programmed RASPP to consider a range of permissible block sizes from 20-250 amino acids. The majority of the lowest $\langle E \rangle$ 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], cloned in electrocompetent *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 dramatic 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 due to 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 $\langle E \rangle$ per crystallized subunit from 59 to 4 and in the average number of mutations $\langle m \rangle$ from 82 to 28. In agreement with prior applications of SCHEMA [19, 37], lower $\langle E \rangle$ chimeras were thus heavily enriched in the

library. Interestingly, we also observed dominance of AAV2 at blocks five and six and AAV9 at block eight. These trends could be used in the future to guide rational capsid engineering.



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 as 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 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) is 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 Figures 1B-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 heat map.

A Cre-dependent selection strategy for AAV directed evolution

Library diversification generates millions of mutant capsids, each a possible solution to a therapeutic 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. Neural stem cells are an especially difficult target given that they 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, extracellular environment, and access to the cerebrospinal fluid (CSF) and vasculature. It would thus be advantageous to conduct evolution *in vivo*; however, *in vivo* AAV directed evolution strategies targeting the central nervous system have relied on extraction of genomic DNA from whole brain tissue and polymerase chain reaction (PCR) to amplify the *cap* gene. These methods 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). In order 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].

Over 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 towards 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* (Supplementary Figure A.1). To prevent false positive recovery of inverted *cap* during *in vivo* selections, the loxP sites were repositioned to flank *cap* such that artefactual 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]. Activated NSCs also express nestin, although the quiescent population does not [43]. 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 neural stem cells and mature astrocytes [44]. Although Cre is expressed in astrocytes in addition to neural stem cells, the intracerebroventricular (ICV) route of administration results in preferential transduction of the SVZ where the neural stem cells reside, and GFAP serves as an important marker of NSC fate [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 intracerebroventricular 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.

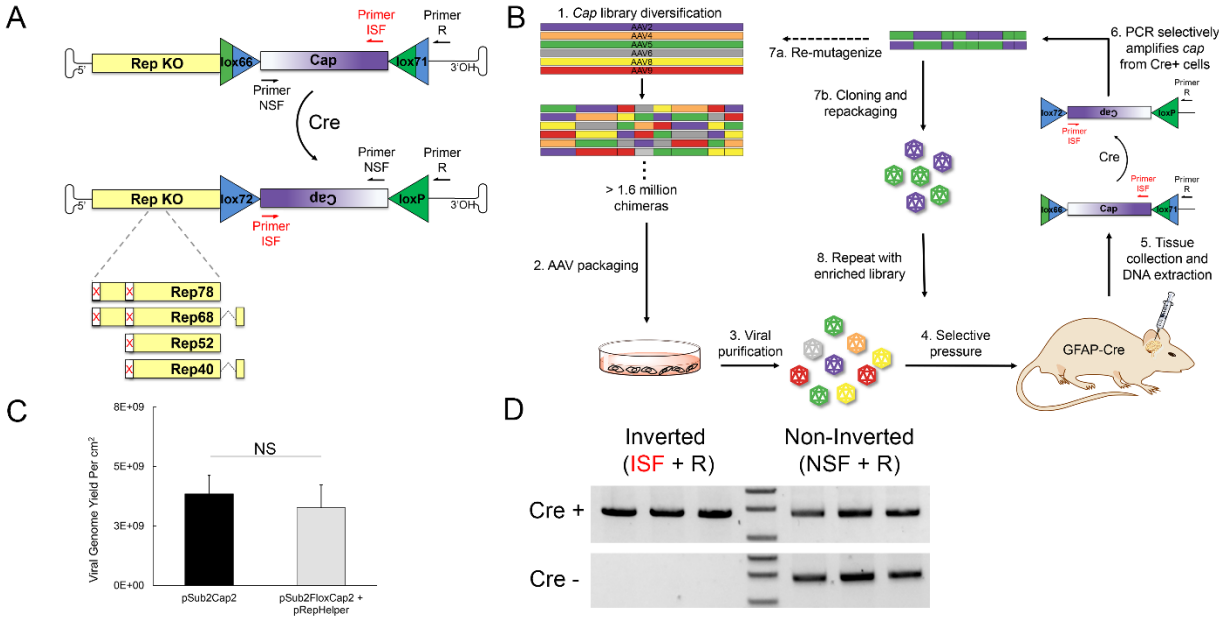


Figure 2. Design of a Cre-dependent selection strategy for AAV directed evolution.
(A) Modifications to the AAV viral genome enable Cre-dependent selection. Mutant loxP sites lox66 and lox71 flank the *cap* gene. Upon Cre inversion of *cap*, Primer ISF changes from a reverse primer to a forward primer 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, while 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 intracerebroventricular 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. NS, not significant.

(D) Selective amplification of inverted *cap* from GFAP-Cre mice.

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 to 10^7 unique variants: (i) the new SCHEMA AAV, (ii) error-prone AAV9, (iii) ancestral AAV [45], (iv) shuffled AAV generated by DNase I digestion and reassembly of AAV1, 2, 4, 5, 6, 8, and 9 [12], (v) error-prone AAV2 [46], and (vi) AAV2 7mer peptide insertion at amino acid 588 [47]. Libraries iii-vi have previously yielded highly infectious clones in our directed evolution selections [9-12, 45] and provide evolutionary competition for the SCHEMA library. The libraries were combined and injected via intracerebroventricular 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 would require multiple injections to cover the same tissue volume due to the limited diffusion distance of AAV in brain tissue.

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, while 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 7mer 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. The amino acid sequences of SCH9, SCH2, and a multiple sequence alignment to the parent AAV serotypes are presented in Supplementary Figures A.2 and A.3, respectively. The two variants differ only at block five, 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 fivefold 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.

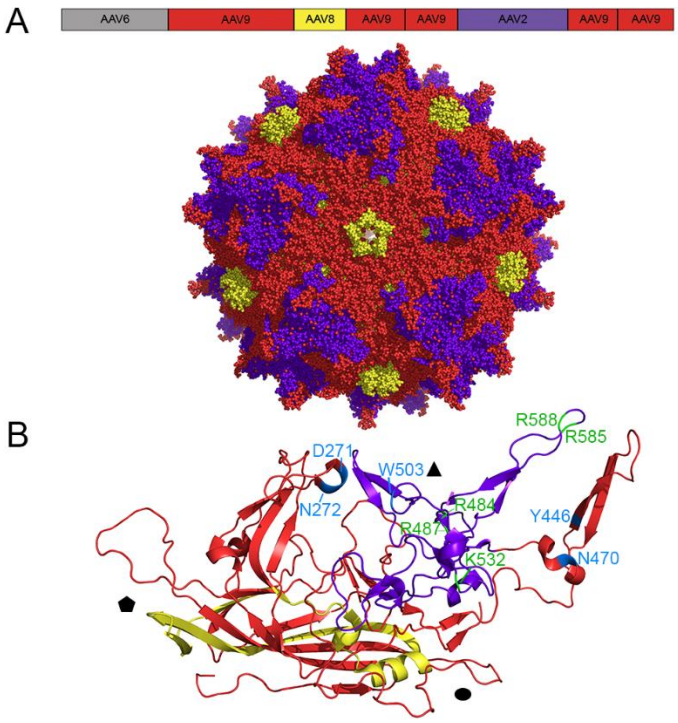


Figure 3. Three-dimensional models of the SCH9 capsid. Block boundaries are represented on a schematic of the *cap* gene. Each parent serotype is represented by a different color (AAV9 red, AAV8 yellow, AAV2 purple) and mapped onto the AAV9 crystal structure (PDB 3UX1) in PyMOL. The full biological assembly (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 neural stem cells in the SVZ of adult mice

To assess the transduction profile of SCH9 in the SVZ, we packaged (recombinant AAV packaging yields are reported in Supplementary Figure A.4) and delivered rAAV carrying a self-complementary CAG-GFP cassette to the right lateral ventricle of adult C57BL/6J mice. SCH9 was benchmarked against AAV9 due to its broad use in the CNS and capacity to transduce the brain parenchyma from the cerebrospinal fluid (CSF) after intrathecal injection [48, 49]. Moreover, of the natural serotypes, AAV9 is the most closely related sequence to SCH9.

Transduction of the contralateral hemisphere was analyzed four weeks after injection, and GFP expression was primarily associated with the region surrounding the ventricle, with greatest intensity in the subventricular zone (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 was 24-fold higher, and expressed in a 12-fold greater SVZ transduction volume, with SCH9 compared to AAV9 (Figures 4B, C).

The SVZ is composed of multiple cell types including ependymal cells, adult NSCs (B1 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 searched for 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 [50]. We therefore identified cells marked as GFP/GFAP/VCAM1 positive as adult neural stem cells transduced by SCH9 in the subventricular zone (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 neural stem cell to an olfactory bulb interneuron involves over seven cell divisions [51]. 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 retrovirus indicate that the time required for neuroblasts to traverse the rostral migratory stream to the olfactory bulb is nine days, and that all transit amplifying cells and neuroblasts present in the SVZ at the time of injection have differentiated and/or migrated to the olfactory bulb and established dendrites by 30 days post-injection [52, 53]. 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 both lentiviral and non-viral transduction of NSCs in the SVZ [54, 55]. 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 [56], within which Cre activity would result in tdTomato expression in transduced cells and their progeny. The majority (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, G), exceeding AAV9 transduction by over 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).

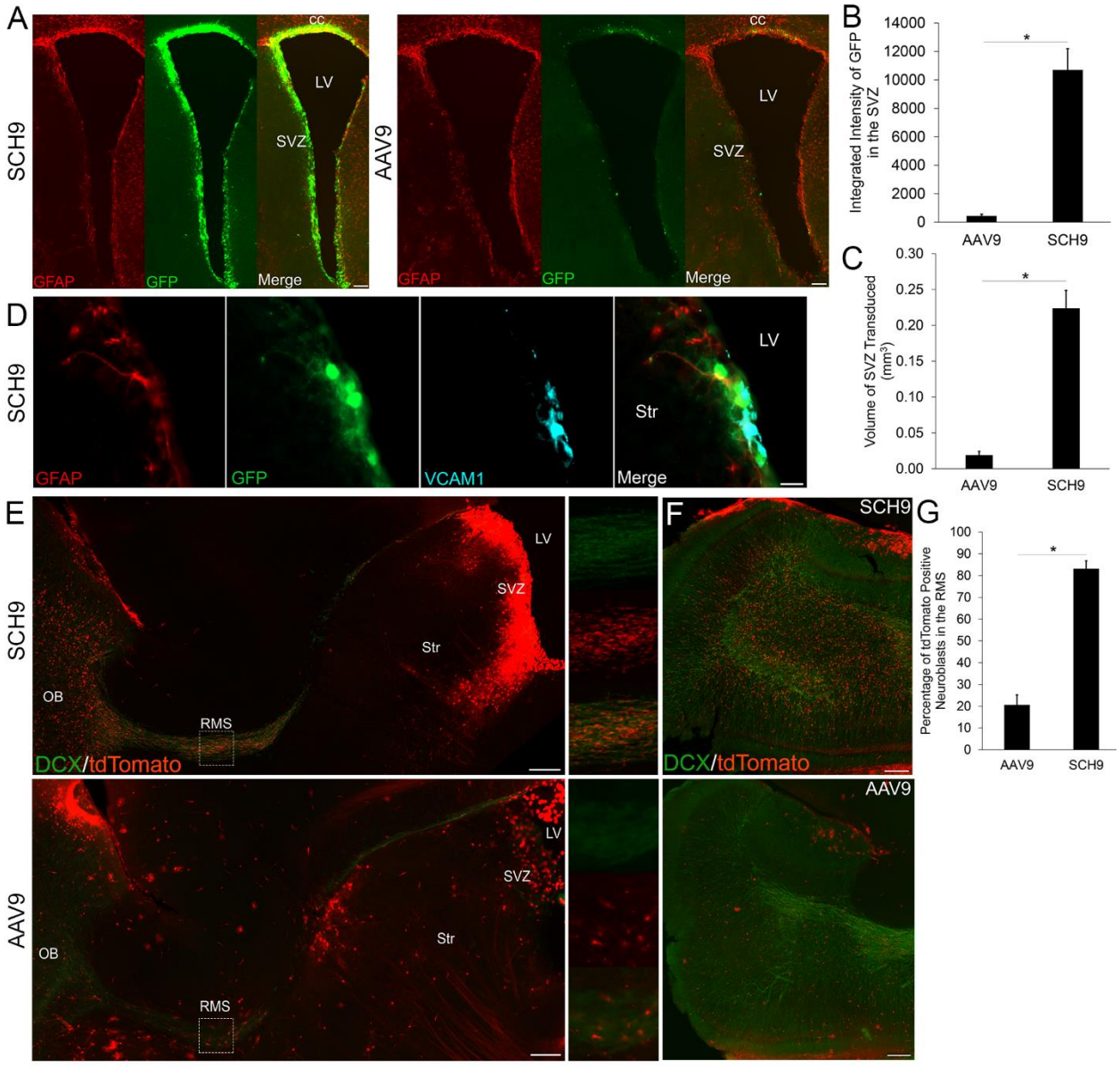


Figure 4. SCH9 efficiently transduces neural stem cells in the subventricular zone.
(A) SCH9 strongly transduces the SVZ contralateral to the injection site. Coronal sections were stained for GFAP (red) and GFP (green). Scale bars indicate 100 µm. LV, lateral ventricle; SVZ, subventricular zone; CC, corpus callosum.

(B, C) SCH9 mediates higher GFP fluorescence intensity and transduction volume 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. LV, lateral ventricle; Str, striatum.

(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. LV, lateral ventricle; SVZ, subventricular zone; Str, striatum; RMS, rostral migratory stream; OB, olfactory bulb.

(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.

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 intracerebroventricular injection was primarily associated with the SVZ, we also observed increased reporter expression in Purkinje cells of the cerebellum, a region of the brain accessible to vector circulating in the cerebrospinal fluid (Figure 5A). Purkinje cells are a key target of gene therapies for neurodegenerative diseases including spinocerebellar ataxias [57]. 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 cerebrospinal fluid 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 [58, 59]. 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 (Supplementary Figure A.5). 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.

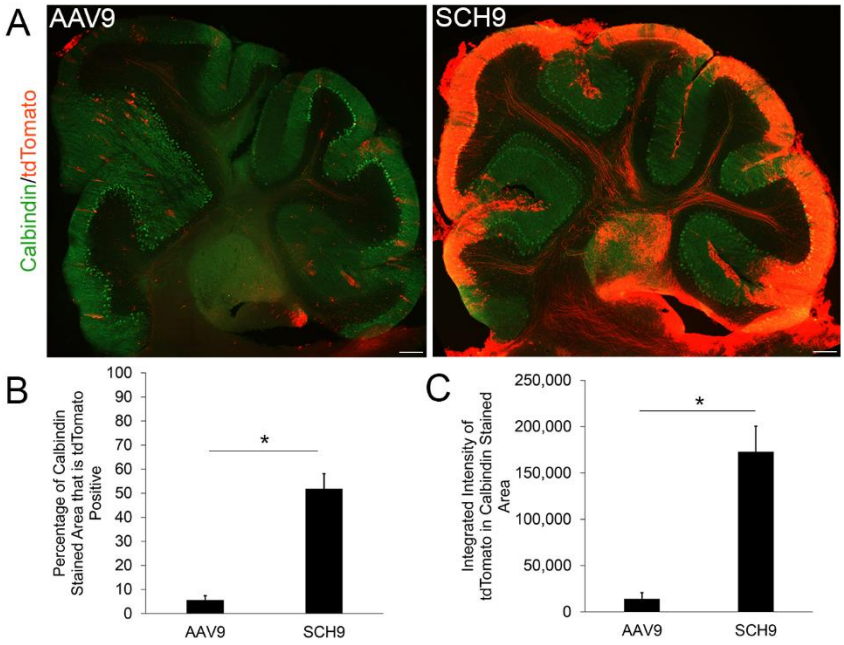


Figure 5. SCH9 mediates high transduction of Purkinje cells in the cerebellum.
(A) Representative images of the cerebellum 30 days after injection of recombinant AAV9 and SCH9 expressing Cre to activate tdTomato (red). Sagittal sections were stained for the Purkinje cell marker calbindin (green), and 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 can utilize both heparan sulfate proteoglycans and galactose for cell transduction

Given the promising infectious properties of SCH9, we next examined whether DNA recombination may have conferred a selective advantage to SCH9 by modulating the receptor binding capabilities of its multiple parent serotypes. Block six of SCH9 contains the heparin binding pocket of the AAV2 capsid [60]. In addition, blocks two and five contain the galactose binding residues D271, N272, N470, and Y446 of AAV9, while block six conserves residue W503 [61]. In contrast, SCH2 lacks two of the key galactose binding residues due to substitution of AAV2 for AAV9 at block five. We first employed affinity 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 [62], 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, while 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 [63]. SCH2, SCH9, and the AAV2 control were all clearly dependent on AAVR (Supplementary Figure A.6).

Finally, since DNA shuffling has been shown to disrupt neutralizing antibody epitopes [14] [64], we quantified the resistance of SCH9 to human intravenous immunoglobulin (IVIG), a polyclonal mixture of antibodies against natural AAV serotypes. We found that the antibody titer required to neutralize SCH9 was two to ten-fold higher (Supplementary Table A.1) than the parent sequences from which it is derived (Figure 6C). Notably, the greatest fold change improvement was relative to AAV9, the most closely related parent sequence.

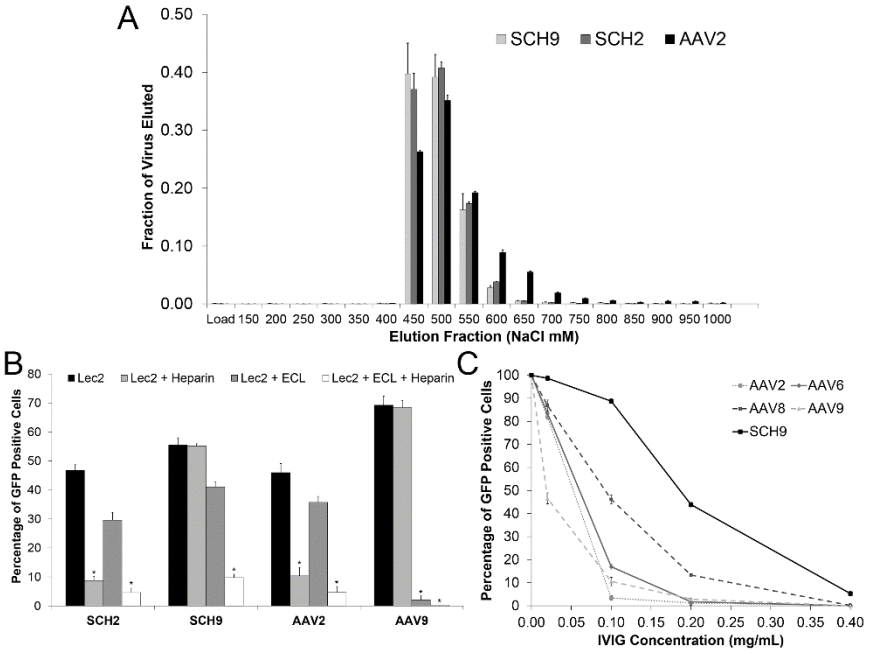


Figure 6. Characterization of SCH9 glycan binding and resistance to neutralizing antibodies.
(A) SCH9 and SCH2 bind a heparin column with similar affinity as AAV2. Samples were loaded onto a heparin affinity column and eluted with increasing NaCl concentration. The load fraction represents the virus 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 due to 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.

2.3 Discussion

Significant progress has been made in elucidating the molecular mechanisms that govern neural stem cell maintenance and neurogenesis in the subventricular zone, yet the ability to genetically manipulate endogenous stem cell populations *in situ* remains challenging due to 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 computational 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 neural stem cells 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 neural stem cells 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, which is not possible using 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 were able screen the largest SCHEMA library constructed to date. SCHEMA variants that package efficiently and incorporate sequence elements from multiple AAV parents 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 permits both the identification of block motifs with desirable properties and the assessment of sequence-function relationships [65]. For example, although the library design shuffles blocks two through four 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 5, and that AAPs enable promiscuous cross-packaging of other serotypes [66, 67]. 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. showed that their AAV2/AAV4 SCHEMA chimeras with crossovers in AAP were less likely to assemble into genome-containing capsids [23]. 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 neural stem cells. Specifically, heparan sulfate proteoglycans function as co-receptors for the binding of key NSC growth factors including FGF-2 and Wnt [68]. Also, LewisX, a glycan motif containing β1-4-linked galactose, is commonly used to isolate neural stem and progenitor cells from the mouse brain [69] and is enriched in undifferentiated NSCs [70]. The presence of HSPG and galactose on neural stem cells may contribute to SCH9 transduction, though 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 [62]. Though AAV2g9 and SCH9 both display dual HSPG and galactose binding, their amino acid sequences share only 88% identity, similar to the level of homology amongst most natural AAV serotypes.

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 Cre-dependent selections in overcoming challenging biological barriers. Deverman *et al*. 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 in recombinase-deficient *E. coli*. Though false positives are unlikely to dominate during successive rounds of selection, we mitigated this concern by introducing loxP sites that flank the *cap* gene such that artefactual 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 central nervous system. 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 [71]. 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 [72], further expanding the utility of this system to transgenic GFP reporter mice, of which there are over 1,300 labelling specific cell populations in the CNS [73]. Finally, intersectional strategies using two recombinase systems (Cre/*loxP* and Flp/*FRT*) [74] 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 as effective in larger animal models due to species-specific differences [9].

Although Cre-dependent selections impose a stringent selective pressure to recover capsids from a target cell type, 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. Alternatively, 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 681 bp GFAP promoter has been developed [75] and may support expression in NSCs.

Several strategies have been employed for gene delivery to adult NSCs *in vivo*. Non-viral gene delivery with linear PEI or by plasmid electroporation supports transfection of NSCs but is

significantly less efficient than viral-mediated gene transfer [55, 76, 77]. By comparison, viral vectors have the potential to be more efficacious but suffer from several drawbacks. For example, adenovirus transduces neuronal precursors in the SVZ but is highly immunogenic [78]. Retrovirus only infects dividing cells, a characteristic of its viral life cycle that results in preferential transduction of rapidly dividing transit-amplifying cells and neuroblasts rather than the relatively quiescent adult neural stem cells [79]. Lentivirus is able to infect both dividing and non-dividing cells and is the most commonly used viral vector for gene delivery to the SVZ [54]. Lentivirus drives gene expression through integration into the host cell genome, a property that has enables lineage tracing studies but carries the risk of insertional oncogenesis. In addition, long-term gene expression may be undesirable when delivering gene editing tools including CRISPR-Cas9 for which off-target activity is a concern. Finally, lentivirus poorly crosses ependymal cell barriers [54], limiting the route of administration to direct injection of the SVZ which can lead to inflammation [80] and reactive astrogliosis [81] that confound analysis of differentiation outcomes.

Natural AAV serotypes have also been considered for gene delivery to the SVZ. AAV2 exhibits poor NSC transduction following direct injection into the SVZ [82]. 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 [83]. Interestingly, AAV4.18, an engineered mutant of AAV4 with affinity for 2,8-linked polysialic acid (PSA), selectively infects migrating progenitors after intracerebroventricular injection in neonatal mice [84]. Although this study did not examine the later 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 [85]. Finally, direct injection of AAV1 into the SVZ results in efficient transduction of transit amplifying cells and neuroblasts but not the neural stem cell population [86]. 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 intracerebroventricular injection promotes transduction of the entire SVZ in both hemispheres, a significant improvement in transduction volume relative to direct injection into the SVZ. Although ICV injection does create 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 when using gene editing tools including zinc finger nucleases, TALENs, and CRISPR/Cas9. Off-target DNA cleavage can cause cellular toxicity and give rise to chromosomal rearrangements that activate oncogenes [87].

Neural stem cells that generate neuronal precursors have also been observed in the subventricular zone of the adult human brain [88]. Despite their germinal capacity, chains of

migrating neuroblasts are not observed in the human rostral migratory stream beyond two years of age [89], though recent data suggests that some precursors migrate into the striatum and differentiate into interneurons in adult humans [90]. Levels of neurogenesis are low in healthy adults, but pathological conditions including stroke [91], epilepsy [92], and Huntington's disease [93] 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 the 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 [94] 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 [95], suggesting that targeted therapeutic approaches including AAV-mediated delivery of anti-cancer therapeutics [96] to the SVZ could be effective. Moreover, AAV vectors could be applied to knockout tumor suppressors or overexpress oncogenes in mice to study the fundamental mechanisms underlying primary tumor formation in the SVZ [97].

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 neural stem cells 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.

2.4 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 [98] to generate the parent sequence alignment. We modified SCHEMA to include 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 angstroms. 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 four nucleotide 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 four nucleotide stretches that could be silently mutated during library assembly to be identical in all parent sequences. For the library design, we considered a minimum allowed 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, all *Bsa*I recognition sites found in pBluescript SK (+), AAV2, 4, 5, 6, 8, and 9 were silently mutated by QuikChange site-directed mutagenesis (Supplementary Table A.2). 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 [99] (Supplementary Tables A.3 and A.4). Primers were designed to incorporate silent mutations at block junctures to facilitate golden gate cloning into the pBluescript vector backbone (Supplementary Table A.5). The golden gate reaction was 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 quantitative PCR (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 [100] 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 Supplementary Table A.2. pSub2RepKO and pRepHelper were generous gifts of Timothy Day and Dr. John Flannery of UC Berkeley. pSub2RepKO, a *rep* knockout 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*, a unique *Bgl*II site was introduced into pSub2RepKO by site-directed mutagenesis using the primers BglIIFwd and BglIIRev. 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*, unique *Xho*I and *Kpn*I sites were introduced 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) and cultured in Dulbecco's Modified Eagle's medium (DMEM, Gibco) with 10% fetal bovine serum (Invitrogen) and 1% penicillin/streptomycin (Invitrogen) at 37 °C and 5% CO_2. AAV libraries or self-complementary recombinant AAV vectors driving expression of green fluorescent protein (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, purified by iodixanol density centrifugation, and buffer exchanged into PBS by Amicon filtration. DNase-resistant viral genomic titers were measured by quantitative real time PCR using a Biorad iCycler (Bio-Rad, Hercules, CA).

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, five microliters 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 [101]. Injection coordinates were selected using a mouse brain atlas (Franklin and Paxinos, 2007) and adjusted after test injections with 0.1% FastGreen dye (Sigma). Injection accuracy throughout the study was confirmed by reporter expression in the choroid plexus and surrounding the contralateral ventricle. Mice were sacrificed three 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 (ThermoFisher) at 55 °C for 3 hours and extrachromosomal DNA was isolated using the Hirt method as previously described [102]. The PCR primers Cap_ISF and Cap_R were used to amplify inverted *cap*, while 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*, five microliters of recombinant vector (1×10^{10} viral genomes/μl) expressing GFP or Cre was stereotaxically injected 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. For injections of the deep cerebellar nuclei, four microliters 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. 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 hours at room temperature, and stained with primary antibodies in blocking solution for 72 hours at 4 °C. The following primary antibodies were used in this study: mouse anti-Calbindin (1:2000; Abcam, ab82812), rabbit anti-GFP (1:1000; Life Technologies, A-11122), goat anti-GFAP (1:1000; Abcam, ab53554), guinea pig anti-DCX (1:1000, EMD Millipore, AB2253), and rat anti-VCAM1 (1:50; EMD Millipore, MAB2627). After three washes in PBS, sections were incubated with secondary antibodies for 2 hours at room temperature and stained with DAPI (Thermo Fisher) for ten 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, the surface area of GFP expression in the SVZ was quantified from thresholded images using CellProfiler [103] 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 the cell segmentation capabilities of CellProfiler were applied 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 calculate the percentage of calbindin stained area that is tdTomato positive, a CellProfiler pipeline was employed to generate a thresholded mask of the calbindin stain. This mask was applied to the thresholded tdTomato image and the tdTomato positive area was dividing 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 affinity of SCH9, SCH2, and wild-type AAV2 were determined as previously described [104]. A 1 ml HiTrap heparin column (GE Healthcare Sciences) was equilibrated with 150 mM NaCl and 50 mM Tris at pH 7.5. 1 x 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 1M NaCl wash. A fraction of each elution was used to infect HEK293T cells, and the percentage of GFP positive cells was quantified 48 hours after infection using a Guava EasyCyte 6HT flow cytometer (EMD/Millipore) (UC Berkeley Stem Cell Center, Berkeley, CA).

AAV utilization of galactose and heparan sulfate proteoglycans for cell transduction was characterized as previously described [62]. 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 minutes 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 hour and then used to infect cells at a genomic MOI of 12,000 (n = 3). After a 1 hour 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 hours after infection by flow cytometry.

To analyze antibody evasion properties, SCH9, AAV2, AAV6, AAV8, and AAV9 were incubated at 37 °C for 1 hour with serial dilutions of heat inactivated IVIG (Gammagard) and then used to infect HEK293T cells at a genomic MOI of 8,000 (n = 3) as previously described [45]. The percentage of GFP-expressing cells was quantified 48 hours 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, wild type HeLa or $AAVR^{KO}$ cells (Clone KIAA0319L, a generous gift from Dr. Jan Carette, Stanford University, CA) were infected 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 hours after infection by flow cytometry.

2.5 Acknowledgements

DSO was supported by a National Science Foundation Graduate Fellowship and a UC Berkeley Dissertation Fellowship. SS was supported by a National Defense Science and Engineering Graduate Fellowship. JSO was supported by a National Science Foundation Graduate Fellowship and a UC Berkeley Graduate Division Fellowship. DVS, DSO, and JSO are inventors on patents involving AAV directed evolution. 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.

2.6 References

1. Asokan, A., D.V. Schaffer, and R.J. Samulski, *The AAV vector toolkit: poised at the clinical crossroads.* Mol Ther, 2012. **20**(4): p. 699-708.
2. Kotterman, M.A., T.W. Chalberg, and D.V. Schaffer, *Viral Vectors for Gene Therapy: Translational and Clinical Outlook.* Annu Rev Biomed Eng, 2015. **17**: p. 63-89.
3. Spencer, H.T., B.E. Riley, and C.B. Doering, *State of the art: gene therapy of haemophilia.* Haemophilia, 2016. **22 Suppl 5**: p. 66-71.
4. Pierce, E.A. and J. Bennett, *The Status of RPE65 Gene Therapy Trials: Safety and Efficacy.* Cold Spring Harb Perspect Med, 2015. **5**(9): p. a017285.
5. Ojala, D.S., D.P. Amara, and D.V. Schaffer, *Adeno-associated virus vectors and neurological gene therapy.* Neuroscientist, 2015. **21**(1): p. 84-98.
6. Jang, J.H., et al., *An evolved adeno-associated viral variant enhances gene delivery and gene targeting in neural stem cells.* Mol Ther, 2011. **19**(4): p. 667-75.
7. Kotterman, M.A., T. Vazin, and D.V. Schaffer, *Enhanced selective gene delivery to neural stem cells in vivo by an adeno-associated viral variant.* Development, 2015. **142**(10): p. 1885-92.
8. Koerber, J.T., et al., *Molecular evolution of adeno-associated virus for enhanced glial gene delivery.* Mol Ther, 2009. **17**(12): p. 2088-95.
9. Dalkara, D., et al., *In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous.* Sci Transl Med, 2013. **5**(189): p. 189ra76.
10. Tervo, D.G., et al., *A Designer AAV Variant Permits Efficient Retrograde Access to Projection Neurons.* Neuron, 2016.
11. Steines, B., et al., *CFTR gene transfer with AAV improves early cystic fibrosis pig phenotypes.* JCI Insight, 2016. **1**(14): p. e88728.
12. Koerber, J.T., J.H. Jang, and D.V. Schaffer, *DNA shuffling of adeno-associated virus yields functionally diverse viral progeny.* Mol Ther, 2008. **16**(10): p. 1703-9.
13. Klimczak, R.R., et al., *A novel adeno-associated viral variant for efficient and selective intravitreal transduction of rat Muller cells.* PLoS One, 2009. **4**(10): p. e7467.
14. Maheshri, N., et al., *Directed evolution of adeno-associated virus yields enhanced gene delivery vectors.* Nat Biotechnol, 2006. **24**(2): p. 198-204.
15. Moore, J.C. and F.H. Arnold, *Directed evolution of a para-nitrobenzyl esterase for aqueous-organic solvents.* Nat Biotechnol, 1996. **14**(4): p. 458-67.
16. Crameri, A., et al., *DNA shuffling of a family of genes from diverse species accelerates directed evolution.* Nature, 1998. **391**(6664): p. 288-91.
17. Chaparro-Riggers, J.F., et al., *Revealing biases inherent in recombination protocols.* BMC Biotechnol, 2007. **7**: p. 77.
18. Voigt, C.A., et al., *Protein building blocks preserved by recombination.* Nat Struct Biol, 2002. **9**(7): p. 553-8.
19. Meyer, M.M., L. Hochrein, and F.H. Arnold, *Structure-guided SCHEMA recombination of distantly related beta-lactamases.* Protein Eng Des Sel, 2006. **19**(12): p. 563-70.

20. Otey, C.R., et al., *Structure-guided recombination creates an artificial family of cytochromes P450.* PLoS Biol, 2006. **4**(5): p. e112.
21. Romero, P.A., et al., *SCHEMA-designed variants of human Arginase I and II reveal sequence elements important to stability and catalysis.* ACS Synth Biol, 2012. **1**(6): p. 221-8.
22. Heinzelman, P., et al., *Efficient screening of fungal cellobiohydrolase class I enzymes for thermostabilizing sequence blocks by SCHEMA structure-guided recombination.* Protein Eng Des Sel, 2010. **23**(11): p. 871-80.
23. Wang, S., et al., *CXCR2 macromolecular complex in pancreatic cancer: a potential therapeutic target in tumor growth.* Transl Oncol, 2013. **6**(2): p. 216-25.
24. Orelle, C., et al., *Protein synthesis by ribosomes with tethered subunits.* Nature, 2015. **524**(7563): p. 119-24.
25. Menzella, H.G., et al., *Combinatorial polyketide biosynthesis by de novo design and rearrangement of modular polyketide synthase genes.* Nat Biotechnol, 2005. **23**(9): p. 1171-6.
26. Lim, D.A. and A. Alvarez-Buylla, *The Adult Ventricular-Subventricular Zone (V-SVZ) and Olfactory Bulb (OB) Neurogenesis.* Cold Spring Harb Perspect Biol, 2016. **8**(5).
27. Gage, F.H., *Mammalian neural stem cells.* Science, 2000. **287**(5457): p. 1433-8.
28. Doetsch, F., et al., *Subventricular zone astrocytes are neural stem cells in the adult mammalian brain.* Cell, 1999. **97**(6): p. 703-16.
29. Liu, Y., et al., *Identification of small-molecule modulators of mouse SVZ progenitor cell proliferation and differentiation through high-throughput screening.* J Biomol Screen, 2009. **14**(4): p. 319-29.
30. Nierode, G.J., et al., *High-Throughput Toxicity and Phenotypic Screening of 3D Human Neural Progenitor Cell Cultures on a Microarray Chip Platform.* Stem Cell Reports, 2016. **7**(5): p. 970-982.
31. Agrawal, S. and D.V. Schaffer, *In situ stem cell therapy: novel targets, familiar challenges.* Trends Biotechnol, 2005. **23**(2): p. 78-83.
32. Heidenreich, M. and F. Zhang, *Applications of CRISPR-Cas systems in neuroscience.* Nat Rev Neurosci, 2016. **17**(1): p. 36-44.
33. Gao, G., et al., *Clades of Adeno-associated viruses are widely disseminated in human tissues.* J Virol, 2004. **78**(12): p. 6381-8.
34. Excoffon, K.J., et al., *Directed evolution of adeno-associated virus to an infectious respiratory virus.* Proc Natl Acad Sci U S A, 2009. **106**(10): p. 3865-70.
35. Endelman, J.B., et al., *Site-directed protein recombination as a shortest-path problem.* Protein Eng Des Sel, 2004. **17**(7): p. 589-94.
36. Engler, C. and S. Marillonnet, *Combinatorial DNA assembly using Golden Gate cloning.* Methods Mol Biol, 2013. **1073**: p. 141-56.
37. Otey, C.R., et al., *Functional evolution and structural conservation in chimeric cytochromes p450: calibrating a structure-guided approach.* Chem Biol, 2004. **11**(3): p. 309-18.
38. Deverman, B.E., et al., *Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain.* Nat Biotechnol, 2016. **34**(2): p. 204-9.
39. Heffner, C.S., et al., *Supporting conditional mouse mutagenesis with a comprehensive cre characterization resource.* Nat Commun, 2012. **3**: p. 1218.

40. Albert, H., et al., *Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome.* Plant J, 1995. **7**(4): p. 649-59.
41. Platel, J.C., et al., *GFAP-GFP neural progenitors are antigenically homogeneous and anchored in their enclosed mosaic niche.* Glia, 2009. **57**(1): p. 66-78.
42. Giachino, C., et al., *Molecular diversity subdivides the adult forebrain neural stem cell population.* Stem Cells, 2014. **32**(1): p. 70-84.
43. Codega, P., et al., *Prospective identification and purification of quiescent adult neural stem cells from their in vivo niche.* Neuron, 2014. **82**(3): p. 545-59.
44. Garcia, A.D., et al., *GFAP-expressing progenitors are the principal source of constitutive neurogenesis in adult mouse forebrain.* Nat Neurosci, 2004. **7**(11): p. 1233-41.
45. Santiago-Ortiz, J., et al., *AAV ancestral reconstruction library enables selection of broadly infectious viral variants.* Gene Ther, 2015. **22**(12): p. 934-46.
46. Koerber, J.T., et al., *Construction of diverse adeno-associated viral libraries for directed evolution of enhanced gene delivery vehicles.* Nat Protoc, 2006. **1**(2): p. 701-6.
47. Muller, O.J., et al., *Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors.* Nat Biotechnol, 2003. **21**(9): p. 1040-6.
48. Samaranch, L., et al., *Adeno-associated virus serotype 9 transduction in the central nervous system of nonhuman primates.* Hum Gene Ther, 2012. **23**(4): p. 382-9.
49. Schuster, D.J., et al., *Biodistribution of adeno-associated virus serotype 9 (AAV9) vector after intrathecal and intravenous delivery in mouse.* Front Neuroanat, 2014. **8**: p. 42.
50. Kokovay, E., et al., *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, 2012. **11**(2): p. 220-30.
51. Ponti, G., K. Obernier, and A. Alvarez-Buylla, *Lineage progression from stem cells to new neurons in the adult brain ventricular-subventricular zone.* Cell Cycle, 2013. **12**(11): p. 1649-50.
52. Petreanu, L. and A. Alvarez-Buylla, *Maturation and death of adult-born olfactory bulb granule neurons: role of olfaction.* J Neurosci, 2002. **22**(14): p. 6106-13.
53. Lois, C. and A. Alvarez-Buylla, *Long-distance neuronal migration in the adult mammalian brain.* Science, 1994. **264**(5162): p. 1145-8.
54. Consiglio, A., et al., *Robust in vivo gene transfer into adult mammalian neural stem cells by lentiviral vectors.* Proc Natl Acad Sci U S A, 2004. **101**(41): p. 14835-40.
55. Barnabe-Heider, F., et al., *Genetic manipulation of adult mouse neurogenic niches by in vivo electroporation.* Nat Methods, 2008. **5**(2): p. 189-96.
56. Madisen, L., et al., *A robust and high-throughput Cre reporting and characterization system for the whole mouse brain.* Nat Neurosci, 2010. **13**(1): p. 133-40.
57. Orr, H.T., *Cell biology of spinocerebellar ataxia.* J Cell Biol, 2012. **197**(2): p. 167-77.
58. Keiser, M.S., et al., *Broad distribution of ataxin 1 silencing in rhesus cerebella for spinocerebellar ataxia type 1 therapy.* Brain, 2015. **138**(Pt 12): p. 3555-66.
59. Dodge, J.C., et al., *Delivery of AAV-IGF-1 to the CNS extends survival in ALS mice through modification of aberrant glial cell activity.* Mol Ther, 2008. **16**(6): p. 1056-64.
60. Kern, A., et al., *Identification of a heparin-binding motif on adeno-associated virus type 2 capsids.* J Virol, 2003. **77**(20): p. 11072-81.
61. Bell, C.L., et al., *Identification of the galactose binding domain of the adeno-associated virus serotype 9 capsid.* J Virol, 2012. **86**(13): p. 7326-33.

62. Shen, S., et al., *Engraftment of a galactose receptor footprint onto adeno-associated viral capsids improves transduction efficiency.* J Biol Chem, 2013. **288**(40): p. 28814-23.
63. Pillay, S., et al., *An essential receptor for adeno-associated virus infection.* Nature, 2016. **530**(7588): p. 108-12.
64. Grimm, D., et al., *In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses.* J Virol, 2008. **82**(12): p. 5887-911.
65. Heinzelman, P., P.A. Romero, and F.H. Arnold, *Efficient sampling of SCHEMA chimera families to identify useful sequence elements.* Methods Enzymol, 2013. **523**: p. 351-68.
66. Sonntag, F., et al., *The assembly-activating protein promotes capsid assembly of different adeno-associated virus serotypes.* J Virol, 2011. **85**(23): p. 12686-97.
67. Earley, L.F., et al., *Adeno-associated Virus (AAV) Assembly-Activating Protein Is Not an Essential Requirement for Capsid Assembly of AAV Serotypes 4, 5, and 11.* J Virol, 2017. **91**(3).
68. Lanctot, P.M., F.H. Gage, and A.P. Varki, *The glycans of stem cells.* Curr Opin Chem Biol, 2007. **11**(4): p. 373-80.
69. Hennen, E. and A. Faissner, *LewisX: a neural stem cell specific glycan?* Int J Biochem Cell Biol, 2012. **44**(6): p. 830-3.
70. Yagi, H., et al., *Lewis X-carrying N-glycans regulate the proliferation of mouse embryonic neural stem cells via the Notch signaling pathway.* J Biol Chem, 2012. **287**(29): p. 24356-64.
71. Saxena, A., et al., *Trehalose-enhanced isolation of neuronal sub-types from adult mouse brain.* Biotechniques, 2012. **52**(6): p. 381-5.
72. Tang, J.C., et al., *Cell type-specific manipulation with GFP-dependent Cre recombinase.* Nat Neurosci, 2015. **18**(9): p. 1334-41.
73. Heintz, N., *Gene expression nervous system atlas (GENSAT).* Nat Neurosci, 2004. **7**(5): p. 483.
74. Dymecki, S.M., R.S. Ray, and J.C. Kim, *Mapping cell fate and function using recombinase-based intersectional strategies.* Methods Enzymol, 2010. **477**: p. 183-213.
75. Lee, Y., et al., *GFAP promoter elements required for region-specific and astrocyte-specific expression.* Glia, 2008. **56**(5): p. 481-93.
76. Lemkine, G.F., et al., *Preferential transfection of adult mouse neural stem cells and their immediate progeny in vivo with polyethylenimine.* Mol Cell Neurosci, 2002. **19**(2): p. 165-74.
77. Falk, A., et al., *Gene delivery to adult neural stem cells.* Exp Cell Res, 2002. **279**(1): p. 34-9.
78. Yoon, S.O., et al., *Adenovirus-mediated gene delivery into neuronal precursors of the adult mouse brain.* Proc Natl Acad Sci U S A, 1996. **93**(21): p. 11974-9.
79. Rogelius, N., C. Ericson, and C. Lundberg, *In vivo labeling of neuroblasts in the subventricular zone of rats.* J Neurosci Methods, 2005. **142**(2): p. 285-93.
80. McCluskey, L., et al., *Inflammatory responses in the rat brain in response to different methods of intra-cerebral administration.* J Neuroimmunol, 2008. **194**(1-2): p. 27-33.
81. Burda, J.E. and M.V. Sofroniew, *Reactive gliosis and the multicellular response to CNS damage and disease.* Neuron, 2014. **81**(2): p. 229-48.

82. Passini, M.A., et al., *Distribution of a lysosomal enzyme in the adult brain by axonal transport and by cells of the rostral migratory stream.* J Neurosci, 2002. **22**(15): p. 6437-46.
83. Liu, G., et al., *Adeno-associated virus type 4 (AAV4) targets ependyma and astrocytes in the subventricular zone and RMS.* Gene Ther, 2005. **12**(20): p. 1503-8.
84. Murlidharan, G., et al., *Unique glycan signatures regulate adeno-associated virus tropism in the developing brain.* J Virol, 2015. **89**(7): p. 3976-87.
85. Gascon, E., L. Vutskits, and J.Z. Kiss, *The role of PSA-NCAM in adult neurogenesis.* Adv Exp Med Biol, 2010. **663**: p. 127-36.
86. Bockstael, O., et al., *Rapid transgene expression in multiple precursor cell types of adult rat subventricular zone mediated by adeno-associated type 1 vectors.* Hum Gene Ther, 2012. **23**(7): p. 742-53.
87. Cho, S.W., et al., *Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases.* Genome Res, 2014. **24**(1): p. 132-41.
88. Sanai, N., et al., *Unique astrocyte ribbon in adult human brain contains neural stem cells but lacks chain migration.* Nature, 2004. **427**(6976): p. 740-4.
89. Sanai, N., et al., *Corridors of migrating neurons in the human brain and their decline during infancy.* Nature, 2011. **478**(7369): p. 382-6.
90. Ernst, A., et al., *Neurogenesis in the striatum of the adult human brain.* Cell, 2014. **156**(5): p. 1072-83.
91. Macas, J., et al., *Increased generation of neuronal progenitors after ischemic injury in the aged adult human forebrain.* J Neurosci, 2006. **26**(50): p. 13114-9.
92. Liu, Y.W., et al., *Doublecortin expression in the normal and epileptic adult human brain.* Eur J Neurosci, 2008. **28**(11): p. 2254-65.
93. Curtis, M.A., et al., *Increased cell proliferation and neurogenesis in the adult human Huntington's disease brain.* Proc Natl Acad Sci U S A, 2003. **100**(15): p. 9023-7.
94. Piccirillo, S.G., et al., *Contributions to drug resistance in glioblastoma derived from malignant cells in the sub-ependymal zone.* Cancer Res, 2015. **75**(1): p. 194-202.
95. Evers, P., et al., *Irradiation of the potential cancer stem cell niches in the adult brain improves progression-free survival of patients with malignant glioma.* BMC Cancer, 2010. **10**: p. 384.
96. Santiago-Ortiz, J.L. and D.V. Schaffer, *Adeno-associated virus (AAV) vectors in cancer gene therapy.* J Control Release, 2016. **240**: p. 287-301.
97. Abel, T.W., et al., *GFAP-Cre-mediated activation of oncogenic K-ras results in expansion of the subventricular zone and infiltrating glioma.* Mol Cancer Res, 2009. **7**(5): p. 645-53.
98. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
99. Hillson, N.J., R.D. Rosengarten, and J.D. Keasling, *j5 DNA assembly design automation software.* ACS Synth Biol, 2012. **1**(1): p. 14-21.
100. Langmead, B., et al., *Ultrafast and memory-efficient alignment of short DNA sequences to the human genome.* Genome Biol, 2009. **10**(3): p. R25.
101. Lai, K., et al., *Sonic hedgehog regulates adult neural progenitor proliferation in vitro and in vivo.* Nat Neurosci, 2003. **6**(1): p. 21-7.
102. Arad, U., *Modified Hirt procedure for rapid purification of extrachromosomal DNA from mammalian cells.* Biotechniques, 1998. **24**(5): p. 760-2.

103. Carpenter, A.E., et al., *CellProfiler: image analysis software for identifying and quantifying cell phenotypes.* Genome Biol, 2006. **7**(10): p. R100.
104. Jang, J.-H., et al., *An evolved adeno-associated viral variant enhances gene delivery and gene targeting in neural stem cells.* Molecular Therapy, 2011. **19**(4): p. 667-675.

Chapter 3: Reconstruction and Characterization of an Ancestral Adeno-Associated Virus (AAV) Library

This chapter is adapted from a manuscript published as

J. Santiago-Ortiz*, D. Ojala*, O. Westesson, J. Weinstein, S. Wong, A. Steinsapir, S. Kumar, I. Holmes, D. Schaffer. AAV Ancestral Reconstruction Library Enables Selection of Broadly Infectious Viral Variants. *Gene Therapy* **22**, 934-946 (2015).

* Indicates co-first authors.

3.1 Introduction

Advances in DNA sequencing, synthesis, and computational phylogenetic analyses are enabling the computational reconstruction and experimental investigation of ancestral protein variants. Following the first ancestral reconstruction study – which resurrected a functional, ancestral digestive ribonuclease from an extinct bovid ruminant using the parsimony principle [1] – reconstructions and functional analyses have been carried out on inferred ancestral proteins belonging to eubacteria, bony vertebrates, mammals, and the least common ancestor of higher primates using several inference methods, including the parsimony, consensus, Bayesian distance, and maximum likelihood methods [2]. Such ancestral reconstructions and subsequent analysis of resurrected variants have yielded insights into the conditions that led to protein evolution as well as the continuous adaption of organisms to changing environmental conditions [3].

Ancestral reconstructions have also been harnessed to incorporate additional sequence diversity into genetic libraries for protein engineering. For instance, small libraries of resurrected ancestral variants were used in evolutionary studies of protein diversification [3-5] and to generate variants that are more tolerant to deleterious mutations. Moreover, inferred ancestral sequences have been combined with extant sequences by swapping residues of interest (e.g. residues in or close to an enzyme's catalytic site) in modern sequences with those of the inferred ancestor. This residue swapping approach was used in basic evolutionary studies [6] as well as to screen for variants with properties such as increased thermostability [7], improved catalytic activity [8], novel substrate binding [9], and higher solubility [10]. Ancestral reconstruction is thus a versatile approach to explore new sequence space for engineering proteins with novel or enhanced properties, and it may likewise offer potential for gene therapy.

This approach has recently been extended to more complex, multimeric proteins including viruses. The evolutionary history of viruses is an especially interesting application given their rapid mutational rates, importance to public health, and promise for gene therapy. For example, ancestral reconstructions of viral proteins have been generated with the goal of developing vaccine candidates against HIV-1 and influenza virus [11, 12], and to study the functionality and properties of the resurrected variants of HIV-1, influenza, and coxsackievirus [13, 14]. These studies demonstrated that viral reconstructions could recapitulate properties of modern variants, including

immunogenicity, packaging, tropism, and cell receptor dependencies. These properties are key to the viral life cycle, and they are also important properties for viruses used as gene therapy vectors.

Adeno-associated virus (AAV) vectors are highly promising for gene therapy. AAVs are non-pathogenic [15] and can transduce numerous dividing and non-dividing cell types, leading to long term expression in the latter [16]. AAV vectors have accordingly been utilized for gene therapy in various tissues, including liver, lung, brain, eye, and muscle [17, 18]. Furthermore, Glybera, the first gene therapy product approved in the European Union in 2012, employs an AAV1 vector [19]. The amino acid composition of the viral capsid, encoded by the *cap* gene, affects AAV tropism, cell receptor usage, and susceptibility to anti-AAV neutralizing antibodies [20]. These key properties in turn impact efficacy in therapeutic gene delivery, which is often limited by poor transduction of numerous cell types, off-target transduction, difficulties with biological transport barriers, and neutralization by pre-existing anti-AAV antibodies [18]. However, extensive engineering of the AAV capsid, via modification of the *cap* gene, promises to improve numerous clinically relevant properties [18].

Given the functional diversity of natural AAV serotypes, availability of numerous genetic sequences, and demonstrated clinical efficacy of recombinant vectors, AAV is an intriguing candidate for ancestral reconstruction, which could further our understanding of its evolutionary history and plasticity. Interesting questions include whether reconstructed variants exhibit higher or lower infectivity on a range of cell types, and whether they are relatively specific for particular cells – an attractive feature for many clinical applications – or are instead promiscuous, as are many extant serotypes. Finally, ancestral sequences and libraries may be useful starting materials for directed evolution studies [8, 21], especially considering that such AAVs likely gave rise to the modern serotypes with their divergent biological properties and tropism.

Motivated by these questions, we conducted ancestral reconstruction of the AAV capsid. Acknowledging and taking advantage of the inherent ambiguity in reconstructing sequences containing highly divergent residues, we synthesized the inferred ancestral capsid not as a single "best guess" sequence, but rather as a large combinatorial library of candidate sequences incorporating degenerate residues at positions of low confidence. We then explored whether phenotypic selection of this ancestral sequence space using five cell lines representative of different tissues would lead to highly infectious variants, and whether these would be promiscuous – i.e. broadly infectious particles - or exhibit specific tropisms. The ancestral library was found to be fit, with packaging and transduction efficiencies that were on par with extant serotypes, and genetically selected variants were found to be broadly infectious on different cell lines. Furthermore, putative ancestral clones exhibited strong *in vivo* gene delivery efficiency, underscoring the potential of such vectors for gene therapy applications.

3.2 Results

Ancestral AAV sequence reconstruction

The goals of ancestral sequence reconstruction are, given a set of extant DNA sequences, to generate a phylogenetic tree and sequence alignment that relates these sequences, and to infer the sequences of ancestral variants at different ancestral nodes. Accurate sequence reconstruction

is challenging due to ambiguity in the evolutionary relationships between extant variants (which affects the phylogenetic tree-building step) as well as sequence divergence at highly variable residues (which affects the sequence alignment and ancestral reconstruction steps).

As a starting point, we reconstructed the phylogeny of human, macaque and rhesus monkey AAV *cap* sequences retrieved from Genbank (n=52) [22]. We used MrBayes [23], which conducts Bayesian Markov chain Monte Carlo (MCMC) simulation of tree space, to estimate the confidence values at each internal node (shown in curly braces in Figs. 3.1a and B.1). This approach generated a phylogenetic tree relating extant sequences, which is essentially a hypothesis concerning the evolutionary history of AAVs. Each branch on this tree depicts the evolutionary direction that diversified the sequences, and each internal node represents a 'splitting' event where two AAV lineages diverged.

With many ancestral nodes to choose from (full tree in Fig. B.1), we selected node 27 (Fig. 3.1a) based on its high confidence value (1.00), which minimizes one potential source of uncertainty (at the level of phylogenetic relationships between entire sequences) and thus improves confidence in the finer-grained downstream reconstruction of individual amino acids' evolutionary histories. This node is also the ancestor of serotypes with demonstrated clinical efficacy (AAV1, Glybera), biomedical interest (AAV6 [24]), or relative resistance to neutralizing antibodies (AAV7 [25]).

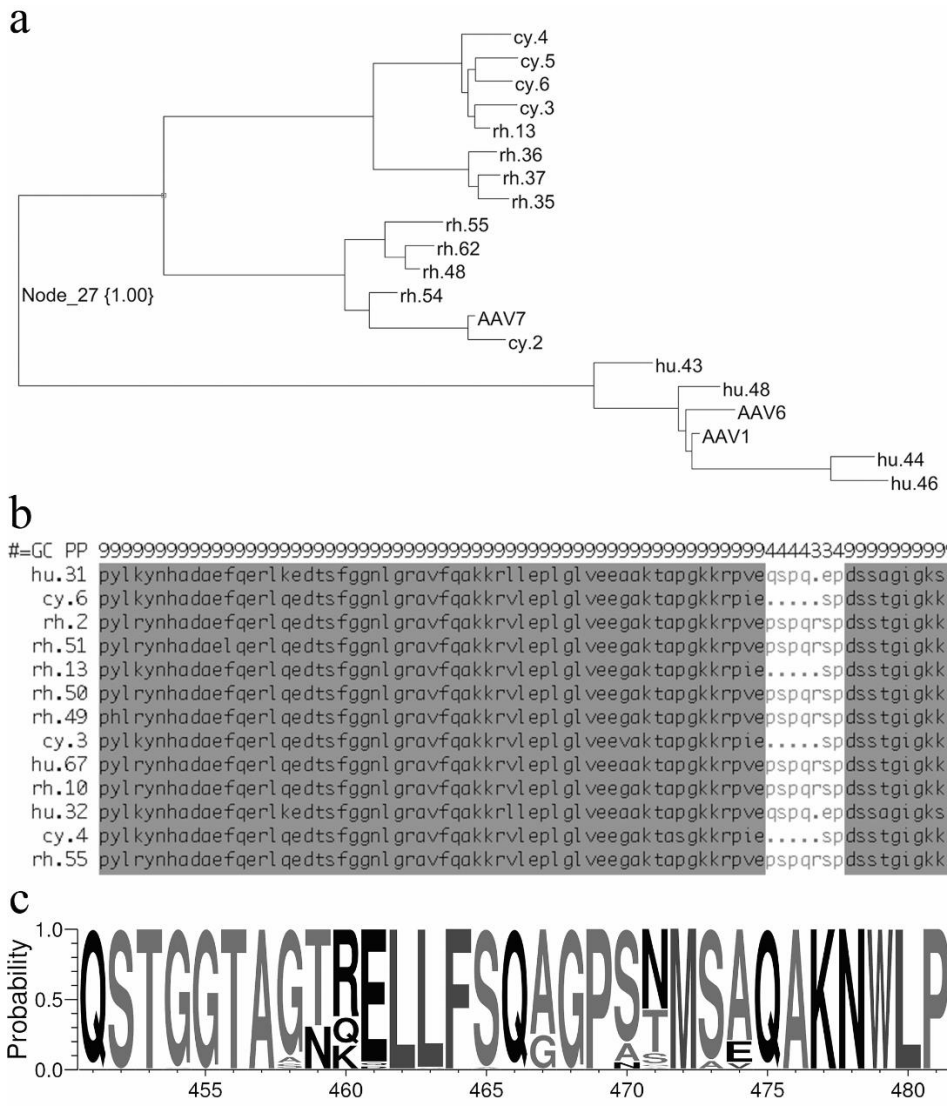


Figure 3.1. Ancestral AAV sequence reconstruction. a) A phylogenetic tree relating a subset of extant AAV variants at node 27. Curly braced numbers indicate clade posterior probabilities[23]. The phylogenetic tree graphic was generated in Dendroscope[26]. b) A multiple sequence alignment of a subset of AAV variants with column-specific confidence annotated along the top with single digits. Confidence ranges from above 0.9 (shaded grey) to 0.3-0.4 (shaded white). c) A distribution of predicted ancestral amino acid sequences for node 27, residues 451-481. The character height of each amino acid is proportional to its posterior probability.

We then used the Markov chain Monte Carlo alignment sampler HandAlign [27] to explore alignment space and predict the ancestral sequence of the most likely alignment at node 27. HandAlign generates a multiple sequence alignment, arranging the sequences of different variants in aligned 'columns' such that residues grouped in a column share a common ancestor (Fig. 3.1b). It also performs the ancestral reconstruction simultaneously with the alignment, and accounts for sequence insertions, deletions, and character substitutions. Figure 3.1c shows the distribution of predicted amino acids as a sequence logo, with character heights proportional to posterior probabilities. The majority of amino acid positions could be predicted with high confidence (≥ 0.90) and thus represented residues highly conserved during evolution. However, as is common in

ancestral reconstruction, other positions were less evolutionarily conserved and could thus be predicted with lower probabilities.

Position	Residue 1	% Freq.	Residue 2	% Freq.	Residue 3	% Freq.
264	T	55	Q	25	A	20
266	A	63	S	37		
268	S	70	A	30		
448	S	71	A	29		
459	T	69	N	31		
460	R	63	Q	20	K	17
467	A	75	G	25		
470	S	85	A	15		
471	N	60	T	32	S	8
474	A	83	E	16		
495	S	75	T	25		
516	D	91	N	9		
533	D	86	E	14		
547	Q	81	E	11	T	8
551	A	50	K	50		
555	T	54	A	46		
557	E	86	D	14		
561	M	62	L	28	I	10
563	S	80	N	19		
577	E	50	Q	50		
583	S	86	D	8	A	6
593	A	45	Q	39	V	16
596	A	81	T	19		
661	A	71	E	19	T	10
662	V	53	T	26	A	22
664	T	66	S	34		
665	P	64	A	26	Q	10
710	T	87	A	13		
717	N	69	D	31		
718	N	60	S	40		
719	E	79	D	21		
723	S	68	T	32		

Table 3.1. Variable positions synthesized in ancestral AAV library.

A DNA library was designed based on these results, and residues above the 0.90 confidence value were fixed, whereas those below this confidence level were varied by introducing the two or three most likely amino acids (above a threshold value of 0.08), such that the fraction of library members containing each amino acid at a given position reflects the probability of that amino acid appearing in the sequence reconstructions. The locations, identities, and synthesis frequencies of the 32 variable residues are presented in Table 2.1, and the most likely full ancestral *cap* amino acid sequence is shown in Fig. B.2 and aligned with extant serotypes in Fig. B.3. The ancestral *cap* library was synthesized (GeneArt, Life Technologies), and analysis of 61 sequenced clones from this library revealed that the amino acid frequencies at variable positions were not significantly different from the theoretical probabilities from the library ($P < 0.001$, see Materials and Methods), highlighting the correctness of the library synthesis.

Phenotypic selection of ancestral AAV library

Given the inherent probabilistic uncertainty of ancestral reconstruction, rather than investigating many possible, candidate ancestral sequences one by one, we selected the library as a whole for functional clones. Specifically, after validating the initial synthesized distribution of amino acids at the 32 variable positions, we probed how those positions would change when subjected to selective pressure for packaging and infectivity, which are key factors for successful

viral replicative fitness during the natural evolution of AAV. The ancestral library was cloned into an AAV packaging plasmid, and viral particles were produced by transfection into human embryonic kidney 293T cells as previously described [28]. The viral genomic titer was comparable to levels obtained when packaging libraries based on extant AAV serotypes (data not shown), indicating that the ancestral library can support robust packaging titers.

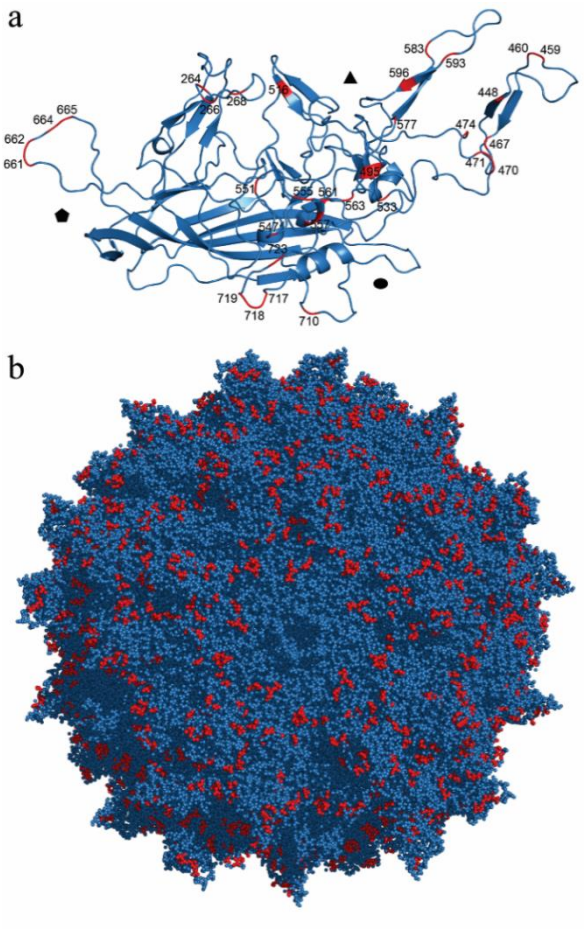


Figure 3.2. Variable residues mapped to the crystal structure of homologous AAV1, the closest AAV relative with an available structure. A three-dimensional molecular model of the AAV1 capsid was generated in PyMOL [29]. An amino acid alignment of the ancestral AAV sequence with AAV1 was used to map the highlighted residues to the a) individual asymmetric unit and b) full biological assembly.

The amino acid distribution at variable positions was only slightly altered by one round of packaging, and we hypothesized that additional selective pressure for infectivity could reveal more about the significance of each variable position. We chose five cell lines representative of different tissues to conduct rounds of selection: C2C12 mouse myoblast cells, IB3-1 human lung epithelial cells, B16-F10 mouse skin melanoma cells, human embryonic kidney 293T cells, and L0 human glioblastoma (GBM) tumor-initiating cells. Briefly, for each round 1 x 10^5 of each cell type were infected with iodixanol-purified, replication-competent AAV libraries at an initial genomic multiplicity of infection (MOI) of 5,000, and successful virions were recovered by superinfecting the cells with adenovirus type 5 two days later. Six rounds of selection were conducted on each

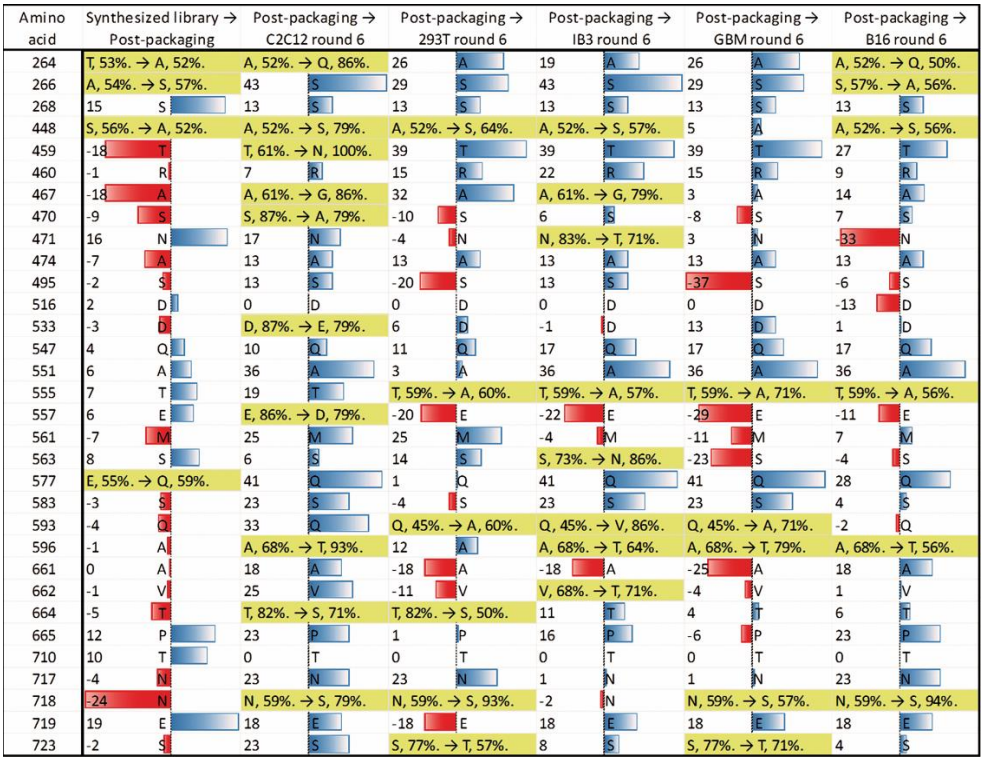
cell line, resulting in five independently selected pools, and the stringency of selection was increased during subsequent rounds by decreasing the genomic MOI (Table B.1).

Amino acid	Theoretical distribution	Synthesized library	Post-packaging	C2C12 round 6	293T round 6	IB3 round 6	GBM round 6	B16 round 6
264	T, 55%	T, 53%	A, 52%	Q, 86%	A, 79%	A, 71%	A, 79%	Q, 50%
266	A, 63%	A, 54%	S, 57%	S, 100%	S, 86%	S, 100%	S, 86%	A, 56%
268	S, 70%	S, 72%	S, 87%	S, 100%	S, 100%	S, 100%	S, 100%	S, 100%
448	S, 71%	S, 56%	A, 52%	S, 79%	S, 64%	S, 57%	A, 57%	S, 56%
459	T, 69%	T, 79%	T, 61%	N, 100%	T, 100%	T, 100%	T, 100%	T, 88%
460	R, 63%	R, 79%	R, 78%	R, 86%	R, 93%	R, 100%	R, 93%	R, 88%
467	A, 75%	A, 79%	A, 61%	G, 86%	A, 93%	G, 79%	A, 64%	A, 75%
470	S, 85%	S, 96%	S, 87%	A, 79%	S, 77%	S, 93%	S, 79%	S, 94%
471	N, 60%	N, 67%	N, 83%	N, 100%	N, 79%	T, 71%	N, 86%	N, 50%
474	A, 83%	A, 94%	A, 87%	A, 100%	A, 100%	A, 100%	A, 100%	A, 100%
495	S, 75%	S, 89%	S, 87%	S, 100%	S, 67%	S, 100%	S, 50%	S, 81%
516	D, 91%	D, 98%	D, 100%	D, 100%	D, 100%	D, 100%	D, 100%	D, 88%
533	D, 86%	D, 90%	D, 87%	E, 79%	D, 93%	D, 86%	D, 100%	D, 88%
547	Q, 81%	Q, 79%	Q, 83%	Q, 93%	Q, 93%	Q, 100%	Q, 100%	Q, 100%
551	A, 50%	A, 58%	A, 64%	A, 100%	A, 67%	A, 100%	A, 100%	A, 100%
555	T, 54%	T, 52%	T, 59%	T, 79%	A, 60%	A, 57%	A, 71%	A, 56%
557	E, 86%	E, 80%	E, 86%	D, 79%	E, 67%	E, 64%	E, 57%	E, 75%
561	M, 62%	M, 75%	M, 68%	M, 93%	M, 93%	M, 64%	M, 57%	M, 75%
563	S, 80%	S, 65%	S, 73%	S, 79%	S, 87%	N, 86%	S, 50%	S, 69%
577	E, 50%	E, 55%	Q, 59%	Q, 100%	Q, 60%	Q, 100%	Q, 100%	Q, 88%
583	S, 86%	S, 80%	S, 77%	S, 100%	S, 73%	S, 100%	S, 100%	S, 81%
593	A, 45%	Q, 49%	Q, 45%	Q, 79%	A, 60%	V, 86%	A, 71%	Q, 44%
596	A, 81%	A, 69%	A, 68%	T, 93%	A, 80%	T, 64%	T, 79%	T, 56%
661	A, 71%	A, 82%	A, 82%	A, 100%	A, 64%	A, 64%	A, 57%	A, 100%
662	V, 53%	V, 69%	V, 68%	V, 93%	V, 57%	T, 71%	V, 64%	V, 69%
664	T, 66%	T, 87%	T, 82%	S, 71%	S, 50%	T, 93%	T, 86%	T, 88%
665	P, 64%	P, 65%	P, 77%	P, 100%	P, 79%	P, 93%	P, 71%	P, 100%
710	T, 87%	T, 90%	T, 100%	T, 100%	T, 100%	T, 100%	T, 100%	T, 100%
717	N, 69%	N, 81%	N, 77%	N, 100%	N, 100%	N, 79%	N, 79%	N, 100%
718	N, 60%	N, 83%	N, 59%	S, 79%	S, 93%	N, 57%	S, 57%	S, 94%
719	E, 79%	E, 63%	E, 82%	E, 100%	E, 64%	E, 100%	E, 100%	E, 100%
723	S, 68%	S, 79%	S, 77%	S, 100%	T, 57%	S, 86%	T, 71%	S, 81%

Figure 3.3. Dominant amino acids at variable positions after six rounds of selection. A heat map was generated based on the frequency of the most common amino acid at each position in the different libraries. The dominant amino acid and frequency at each position were determined based on sequencing results from individual clones n = 61 (synthesized library), n = 23 (post-packaging), and n=14 (for each ancestral library after selection on respective cell lines).

To assess the progression of selection at each variable position, clones were sequenced (*n = 14)* from each library after initial viral packaging (hereafter referred to as post-packaging), after three rounds of selection, and after six rounds of selection. This analysis revealed a range of outcomes for each variable position across the different cell lines. Figure 3.2 shows the positions of the variable amino acids mapped onto the crystal structure of AAV1 (the most homologous serotype with a solved structure), and Figure 3.3 depicts the dominant amino acid at each of these positions for each selected pool as a heat map, with darker shades representing higher convergence. As expected, selection for infection of cell lines led to increased convergence, and Figure 3.4 shows the percentage change in amino acid frequency in rounds 6 relative to post-packaging (increases in blue, decreases in red, and change of amino acid in yellow). Some amino acid positions approached full convergence to the same residue across all cell lines; other positions were divergent, or even acquired specific identities unique to only one cell line. The majority of

residues unique to one cell line are located on the surface of the capsid, and they could for example play a role in altering the affinity of capsid interactions with cell surface receptors.



Amino acid	Synthesized library → Post-packaging	Post-packaging → C2C12 round 6	Post-packaging → 293T round 6	Post-packaging → IB3 round 6	Post-packaging → GBM round 6	Post-packaging → B16 round 6
264	T, 53%. → A, 52%.	A, 52%. → Q, 86%.	26 A	19 A	26 A	A, 52%. → Q, 50%.
266	A, 54%. → S, 57%.	43 S	29 S	43 S	29 S	S, 57%. → A, 56%.
268	15 S	13 S	13 S	13 S	13 S	13 S
448	S, 56%. → A, 52%.	A, 52%. → S, 79%.	A, 52%. → S, 64%.	A, 52%. → S, 57%.	5 A	A, 52%. → S, 56%.
459	-18 T	T, 61%. → N, 100%.	39 T	39 T	39 T	27 T
460	-1 R	7 R	15 R	22 R	15 R	9 R
467	-18 A	A, 61%. → G, 86%.	32 A	A, 61%. → G, 79%.	3 A	14 A
470	-9 S	S, 87%. → A, 79%.	-10 S	6 S	-8 S	7 S
471	16 N	17 N	-4 N	N, 83%. → T, 71%.	3 N	-33 N
474	-7 A	13 A	13 A	13 A	13 A	13 A
495	-2 S	13 S	-20 S	13 S	-37 S	-6 S
516	2 D	0 D	0 D	0 D	0 D	-13 D
533	-3 D	D, 87%. → E, 79%.	6 D	-1 D	13 D	1 D
547	4 Q	10 Q	11 Q	17 Q	17 Q	17 Q
551	6 A	36 A	3 A	36 A	36 A	36 A
555	7 T	19 T	T, 59%. → A, 60%.	T, 59%. → A, 57%.	T, 59%. → A, 71%.	T, 59%. → A, 56%.
557	6 E	E, 86%. → D, 79%.	-20 E	-22 E	-29 E	-11 E
561	-7 M	25 M	25 M	-4 M	-11 M	7 M
563	8 S	6 S	14 S	S, 73%. → N, 86%.	-23 S	-4 S
577	E, 55%. → Q, 59%.	41 Q	1 Q	41 Q	41 Q	28 Q
583	-3 S	23 S	-4 S	23 S	23 S	4 S
593	-4 Q	33 Q	Q, 45%. → A, 60%.	Q, 45%. → V, 86%.	Q, 45%. → A, 71%.	-2 Q
596	-1 A	A, 68%. → T, 93%.	12 A	A, 68%. → T, 64%.	A, 68%. → T, 79%.	A, 68%. → T, 56%.
661	0 A	18 A	-18 A	-18 A	-25 A	18 A
662	-1 V	25 V	-11 V	V, 68%. → T, 71%.	-4 V	1 V
664	-5 T	T, 82%. → S, 71%.	T, 82%. → S, 50%.	11 T	4 T	6 T
665	12 P	23 P	1 P	16 P	-6 P	23 P
710	10 T	0 T	0 T	0 T	0 T	0 T
717	-4 N	23 N	23 N	1 N	1 N	23 N
718	-24 N	N, 59%. → S, 79%.	N, 59%. → S, 93%.	-2 N	N, 59%. → S, 57%.	N, 59%. → S, 94%.
719	19 E	18 E	-18 E	18 E	18 E	18 E
723	-2 S	23 S	S, 77%. → T, 57%.	8 S	S, 77%. → T, 71%.	4 S

Figure 3.4. Change in amino acid frequency at variable positions after six rounds of selection. The percent change in amino acid frequency between the post-packaging library and evolved libraries after six rounds of selection on each cell line was calculated. If the identity of the dominant amino acid did not change, the increase (blue) or decrease (red) in frequency is displayed. If selection resulted in a change in amino acid identity at that position, the new amino acid and frequency is shown (yellow).

To determine whether the changes in amino acid frequencies imparted by phenotypic selection were statistically significantly different from the initial synthesized distribution, we conducted Bayesian Dirichlet-multinomial model comparison tests (as described in Materials and Methods) to calculate the posterior probability that the two sets of variable amino acids come from different distributions. This analysis identified several amino acid positions that are significantly different after selection ($P < 0.05$, shown in green), and many more that are moderately different ($P < 0.5$, shown in yellow) (Fig. 3.5).

			C2C12			293T			IB3			B16			GBM		
Amino acid	Theoretical vs. Synth.	Synth. vs. PP	PP vs. R3	PP vs. R6	R3 vs. R6	PP vs. R3	PP vs. R6	R3 vs. R6	PP vs. R3	PP vs. R6	R3 vs. R6	PP vs. R3	PP vs. R6	R3 vs. R6	PP vs. R3	PP vs. R6	R3 vs. R6
264	0.000	7.6	99.8	100.0	0.2	14.6	0.5	0.4	57.9	54.4	0.1	91.5	95.5	0.1	0.3	37.4	1.3
266	0.000	0.0	81.4	71.8	0.8	0.1	0.5	0.4	0.6	71.8	3.1	0.1	0.1	0.1	0.2	0.5	0.1
268	0.000	0.1	3.2	2.9	0.8	0.1	2.9	9.9	3.1	2.9	0.8	2.9	3.2	0.8	0.1	2.9	1.5
448	0.000	0.0	0.1	0.4	0.2	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.3	0.1	0.1	0.1
459	0.000	0.1	0.1	98.7	64.5	0.5	55.1	2.6	60.9	55.1	0.8	1.4	0.4	0.2	0.2	55.1	6.5
460	0.000	0.0	0.1	0.1	0.1	0.2	0.3	0.1	0.3	8.5	1.5	0.3	0.2	0.2	0.1	0.3	0.2
467	0.000	0.1	0.1	5.3	12.5	0.1	1.4	3.3	0.1	1.3	2.2	0.1	0.1	0.1	0.1	0.1	0.1
470	0.035	0.3	0.2	81.4	3.2	0.3	39.4	0.5	7.1	7.1	0.2	0.1	0.1	0.2	0.1	0.9	0.8
471	0.000	0.1	0.1	5.0	2.8	7.6	0.1	1.8	0.1	17.5	0.8	0.3	1.7	0.3	0.4	0.1	0.6
474	0.000	0.1	3.2	2.9	0.8	3.4	2.9	0.7	0.1	2.9	1.5	0.1	3.2	3.8	3.1	2.9	0.8
495	0.000	0.0	0.1	2.9	1.4	0.2	0.2	0.1	0.1	2.9	3.1	0.1	0.1	0.1	0.1	1.5	1.1
516	0.001	0.5	0.6	0.6	0.8	0.6	0.6	0.7	0.6	0.6	0.8	0.6	3.9	2.8	0.6	0.6	0.8
533	0.000	0.0	1.7	81.4	0.3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	3.1	2.9	0.8
547	0.000	0.0	5.8	0.6	1.7	1.2	0.2	10.2	5.4	5.0	0.8	0.0	5.8	8.8	0.6	5.0	1.5
551	0.000	0.0	0.1	43.3	11.6	0.2	0.1	0.1	0.3	43.3	3.1	1.0	53.7	1.7	0.1	43.3	6.5
555	0.000	0.0	0.1	0.2	0.3	2.4	0.1	0.2	0.4	0.1	0.1	0.1	0.1	0.1	0.1	0.4	0.8
557	0.000	0.0	0.3	75.9	1.3	0.4	0.2	0.1	0.1	0.2	0.1	0.1	0.1	0.1	3.3	0.5	62.7
561	0.000	0.0	1.1	0.8	0.2	0.0	0.9	0.3	31.3	0.1	37.1	0.0	2.7	0.5	0.2	0.0	3.0
563	0.000	0.0	0.1	0.1	0.1	0.1	0.1	0.4	18.6	35.0	100.0	0.1	0.1	0.1	0.5	0.2	0.1
577	0.000	0.1	72.1	61.3	0.8	0.6	0.1	0.5	66.9	61.3	0.8	0.4	0.5	0.1	66.9	61.3	0.8
583	0.000	0.0	11.7	9.6	0.8	0.0	0.0	0.0	10.6	9.6	0.8	0.2	0.1	0.1	10.6	9.6	0.8
593	0.000	0.0	3.2	3.5	0.3	0.1	0.1	0.0	0.1	57.0	98.5	0.0	0.0	0.0	0.0	11.5	0.4
596	0.000	0.0	0.8	66.4	0.5	0.1	0.1	0.1	0.1	0.4	0.8	0.1	0.2	0.3	0.1	3.9	2.2
661	0.000	0.0	0.1	5.5	22.9	0.1	0.0	0.0	0.0	7.0	8.8	0.0	6.4	3.8	0.0	0.1	0.0
662	0.000	0.0	0.3	0.7	0.2	0.1	0.2	0.1	0.0	3.0	1.1	0.0	0.1	0.2	0.1	0.0	0.2
664	0.000	0.0	0.3	13.9	0.3	0.1	2.1	0.5	0.1	0.2	0.2	0.1	0.1	0.1	0.1	0.1	0.1
665	0.000	0.0	0.3	9.6	2.8	0.0	0.6	1.8	0.3	0.4	0.4	0.0	11.7	19.9	0.3	0.4	0.1
710	0.000	2.6	0.6	0.7	0.8	11.1	0.7	6.5	0.6	0.7	0.8	0.7	0.6	0.8	4.2	0.7	3.1
717	0.000	0.0	0.1	9.6	22.9	0.1	9.6	6.5	0.3	0.1	0.3	0.2	11.7	1.7	0.1	0.1	0.1
718	0.000	0.3	0.1	1.0	0.3	0.4	23.4	21.3	0.2	0.1	0.2	0.4	39.3	0.5	0.1	0.1	0.1
719	0.000	0.1	6.4	5.5	0.8	0.5	0.1	0.5	0.2	5.5	1.5	0.2	6.4	1.7	0.2	5.5	1.6
723	0.000	0.0	0.3	9.6	1.4	1.1	89.8	3.5	0.1	0.1	0.4	0.2	0.1	0.2	0.2	5.4	5.7

Figure 3.5. Identification of key variable residues by Bayesian Dirichlet-multinomial model comparison tests. A comparison of the two sets of variable amino acids was conducted to identify positions that changed significantly during selection. The posterior probability that the two sets of amino acids come from two different probability distributions was calculated assuming probability parameters that are Dirichlet-distributed with low pseudocounts to reflect sparse observed sequences. Results colored green indicate a >95% chance that the sets came from different distributions, yellow a >50% chance, red a >5% chance, and no color a <5% chance. Synth, synthesized library; PP, post-packaging; R3, round three of selection; R6, round six of selection.

Transduction efficiency of evolved ancestral libraries

Phenotypic selection could conceivably lead to specific infectivity of a given cell line or may alternatively increase overall infectivity but in a promiscuous manner across all cell types. We investigated these possibilities by evaluating the transduction efficiency of evolved ancestral libraries on the cell line panel. The degree of convergence for each amino acid position after six rounds of selection is shown in Figure 3.3. Selection did not drive full convergence to a single sequence, potentially due to the presence of neutral positions that conferred no selective advantage. Therefore, rather than packaging individual clones, the libraries selected on each cell line were each packaged as a pool of recombinant virus (at a low ratio of AAV helper plasmid per producer cell to minimize mosaic capsids), resulting in five distinct round 6 ancestral libraries; results thus represent overall or average library infectivities. High titer, iodixanol-purified recombinant AAV (rAAV) encoding the green fluorescent protein (GFP) was produced for the ancestral libraries, as well as for natural serotypes AAV1-6, 8, and 9 for comparison of transduction efficiency and tropism. Infection at a genomic MOI of 2,000 (or 32,000 for C2C12s) revealed a range of properties (Fig. 3.6). Functionally selected ancestral libraries mediated high delivery efficiencies most comparable to AAV1 and AAV6 and generally superior to AAV4, AAV5, AAV8, and AAV9. Ancestral libraries were especially successful in infecting C2C12 and GBM cell lines relative to natural serotypes. Importantly, we observed a large increase in infectivity when comparing the synthesized vs. the round 6 ancestral libraries, suggesting phenotypic selection of advantageous amino acids at the variable positions. Interestingly, the libraries in general displayed broad infectivity across all cell lines, indicating that this reconstructed ancestral pool contains

promiscuous AAVs, a property known to be advantageous for natural evolutionary adaptability [30, 31].

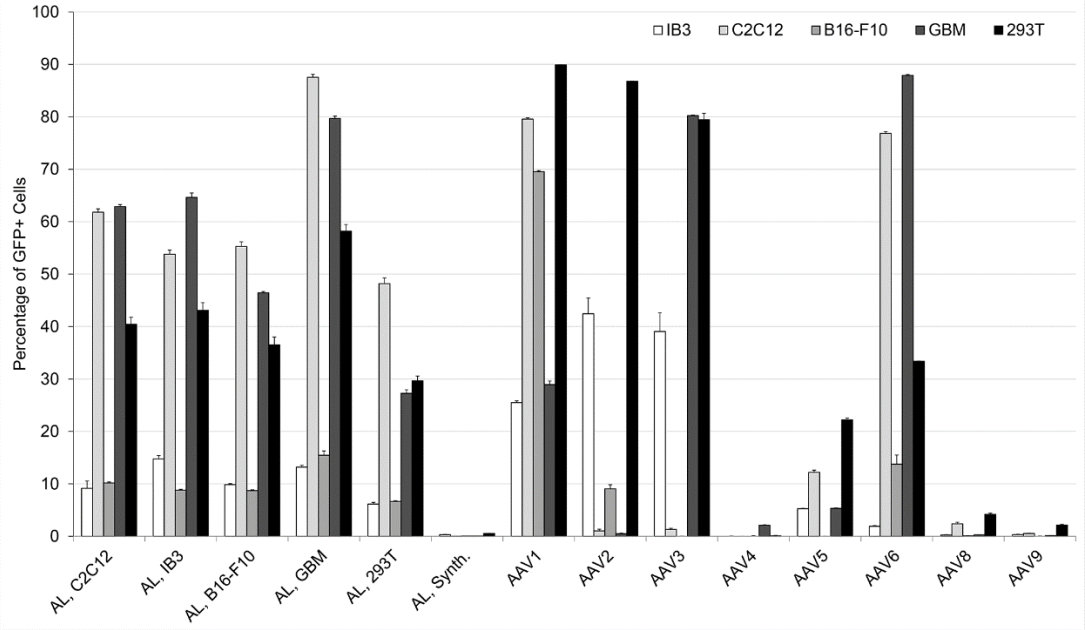


Figure 3.6. Transduction efficiency of ancestral libraries benchmarked against natural AAV serotypes. After six rounds of selection, viral genomic DNA was recovered from ancestral libraries and packaged as rAAV scCMV-GFP along with wild type AAV 1-6, 8, and 9. Cell lines were infected at a genomic multiplicity of infection (MOI) of 2,000 (293T, IB3, B16-F10, GBM) or 32,000 (C2C12). The fraction of GFP expressing cells was quantified by flow cytometry 72 hours later. Data are presented as mean ± SEM, n = 3. AL, ancestral library.

Characterization of the thermostability of candidate ancestral AAV variants

High thermostability and enhanced tolerance to mutations are also properties that could confer an evolutionary advantage to ancestral viral capsids [3, 7, 32]. We benchmarked the thermostability of AAV variants selected from our reconstructed pool against the natural serotypes AAV1, AAV2, AAV5, and AAV6 by assaying their transduction efficiency after heat treatment. Specifically, for initial analysis we chose the ancestral library selected on C2C12 cells and a representative variant from this library, C7. Virions packaged with self-complementary CMV-GFP were treated for 10 minutes at different temperatures using a thermal gradient before being cooled down to 37°C and used to infect 293T cells. We normalized the resulting fraction of GFP expressing cells after treatment at each temperature to the sample incubated at 37° (Fig. 3.7).

Ancestral variants displayed higher thermostability than natural serotypes and showed moderate transduction levels even at the highest treatment temperature, 78°C, which ablated transduction by natural serotypes. The obtained thermostabilities confirm those previously reported for natural serotypes [33], which showed that AAV5 is more stable than AAV1 and that AAV2 is less stable than both. Enhanced thermostability of the ancestral variants in general could enable a higher tolerance to destabilizing mutations, and consequently a higher evolutionary adaptability.

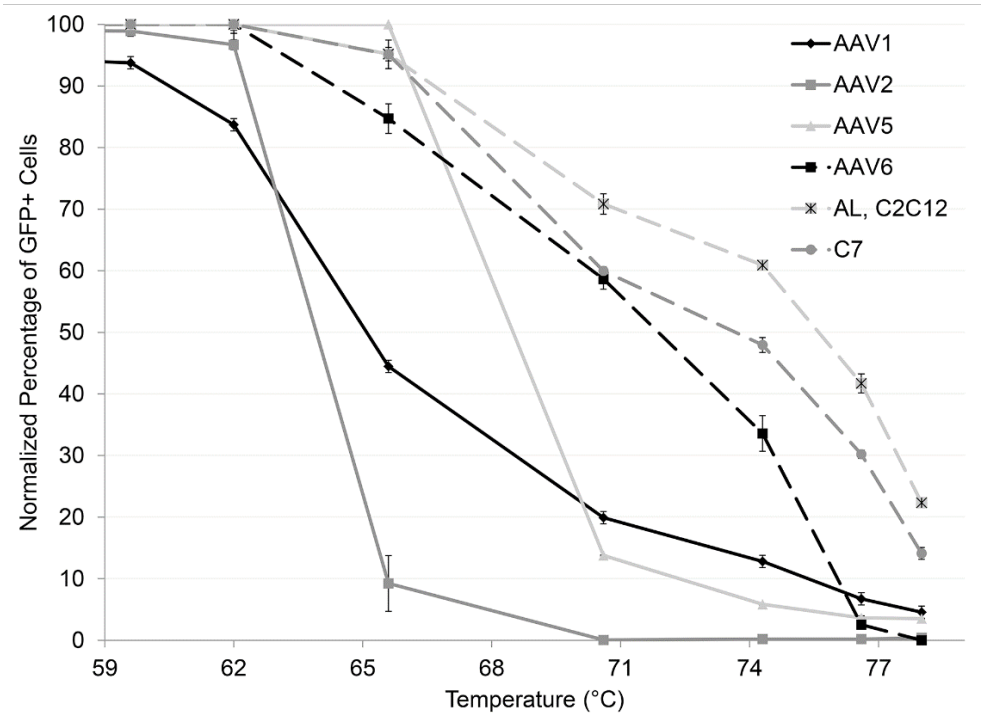


Figure 3.7. Candidate ancestral variants display higher thermostability than natural serotypes. The thermostability of the ancestral library selected on C2C12 cells and of the representative ancestral variant C7 was characterized and compared to that of natural serotypes 1, 2, 5, and 6. Virions packaged with scCMV-GFP were incubated at temperatures ranging from 59.6°C to 78°C for 10 minutes before being cooled down to 37°C and used to infect 293T cells. The fraction of GFP expressing cells was quantified by flow cytometry 72 hours later. Data are presented, after being normalized to the fraction of GFP expressing cells after incubation at 37°, as mean ± SEM, *n*=3.

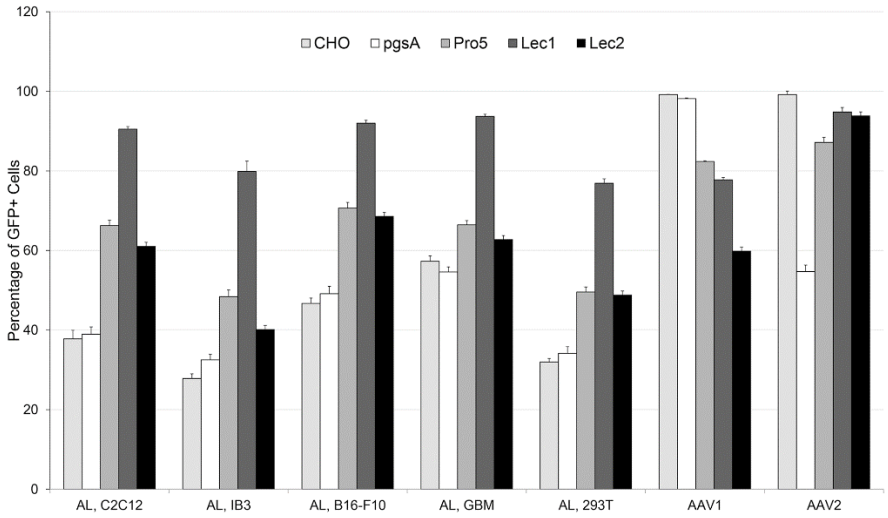
Characterization of ancestral AAV glycan dependencies and susceptibility to neutralizing antibodies

Our *in vitro* transduction experiments demonstrated the broad infectivity of reconstructed variants. Given that ancestral node 27 gave rise to AAV1 and AAV6, we were interested in determining whether the candidate ancestral clones shared the same glycan dependencies, or if those evolved later. AAV1 and AAV6 utilize both alpha 2,3 and alpha 2,6 N-linked sialic acids as their primary receptor, and AAV6 has moderate affinity for heparan sulfate proteoglycans [24]. To probe heparan sulfate proteoglycan (HSPG) usage, we transduced parental CHO-K1 cells and the pgsA CHO variant line deficient in HSPG. To examine sialic acid dependence we transduced parental Pro5 CHO cells presenting glycans with both N- and O-linked sialic acids, a Lec2 CHO variant cell line deficient in all N- and O-linked sialic acids, and a Lec1 line deficient in complex and hybrid type N-glycans including sialic acids [34] (Fig. 3.8b). Interestingly, candidate ancestral AAVs exhibited no dependence on HSPG or N- and O-linked sialic acids (Fig. 3.8a). We also verified that selected individual clones exhibited similar transduction behavior as the evolved libraries (Fig. B.6).

We next examined whether ancestral AAVs were neutralized by antibodies against a broad range of contemporary AAVs, in particular human intravenous immunoglobulin (IVIG) that

contains polyclonal antibodies against extant serotypes due to natural exposure across the human population. *In vitro* incubation with IVIG strongly reduced transduction of ancestral libraries and the AAV1 control (Fig. B.7), indicating that this ancestral pool is not highly serologically distinct from its progeny. Additional capsid engineering may be necessary to address this clinically relevant problem.

a



b

Cell line	Biochemical defect	Glycosylation phenotype	N-glycans	O-GalNAc glycans
Pro5	N/A	Full complement of complex N-glycans capped with sialic acid residues		
Lec1	Loss of N-acetylglucosyl transferase-I activity	N-glycans terminate in mannose, no complex or hybrid N-glycans		
Lec2	Unable to transport CMP-sialic acid into Golgi	N-glycans terminate in galactose, all glycans lack sialic acid		
CHO-K1	N/A	Full complement of heparan sulfate proteoglycans		
pgsA	Loss of xylosyltransferase	Lacks heparan sulfate proteoglycans		

sialic acid ○ galactose ● mannose ■ N-acetylglucosamine □ N-acetylgalactosamine

◥ glucosamine ◇ iduronic acid

Figure 3.8. Glycan dependency of candidate ancestral AAV variants. a) After six rounds of selection, the transduction efficiency of ancestral libraries carrying scCMV-GFP was quantified by flow cytometry 72 hours after

infection at a genomic MOI of 2,000 (Pro5, Lec1, Lec2) and 50,000 (CHO-K1, pgsA). The CHO-K1/pgsA comparison examines heparan sulfate proteoglycan dependence, while Pro5/Lec1 and Pro5/Lec2 probe sialic acid dependence. Data are presented as mean ± SEM, $n = 3$. b) Glycans present on CHO glycosylation mutants. AL, ancestral library.

Characterization of ancestral variants in vivo in mouse gastrocnemius muscle

Upon finding that the ancestral AAV libraries exhibited efficiencies comparable to or in some cases higher than extant serotypes on a panel of cell lines from representative tissues, we next probed *in vivo* infectivity. Based on the high transduction efficiencies of candidate ancestral AAVs on the most nonpermissive cell line (C2C12 mouse myoblasts), we chose to evaluate *in vivo* transduction of mouse gastrocnemius muscle. In particular, individual ancestral variant clones from the selected viral pools (Table B.2) that were closest to the consensus sequences of libraries evolved on C2C12 (clones C4, C7) and glioblastoma cells (clone G4) were chosen, based on the efficiency of these two libraries in transducing C2C12 myoblasts *in vitro*. In addition, these variants were benchmarked against AAV1, given its clinical efficacy in muscle-targeted gene therapy [35].

We generated recombinant AAV vectors expressing firefly luciferase under the control of the hybrid CAG (CMV early enhancer/chicken β-actin/splice acceptor of β-globin gene) promoter. A volume of 30 µl DNase-resistant genomic particles (5×10^{10} vg) was injected into each gastrocnemius muscle of BALB/c mice, and after six weeks, mice were sacrificed and tissue luciferase activities analyzed (Fig. 3.9). Ancestral reconstruction variants yielded 19-31 fold higher transgene expression than AAV1 in gastrocnemius muscle, with variant C7 yielding the highest expression. Interestingly, variant C7 was the most abundant sequence (71%) in the round 6 ancestral library selected on C2C12 cells. These results demonstrate that candidate ancestral AAVs also exhibit high infectivity *in vivo*, and even offer the potential to exceed the performance of the best contemporary natural serotypes in gene therapy applications.

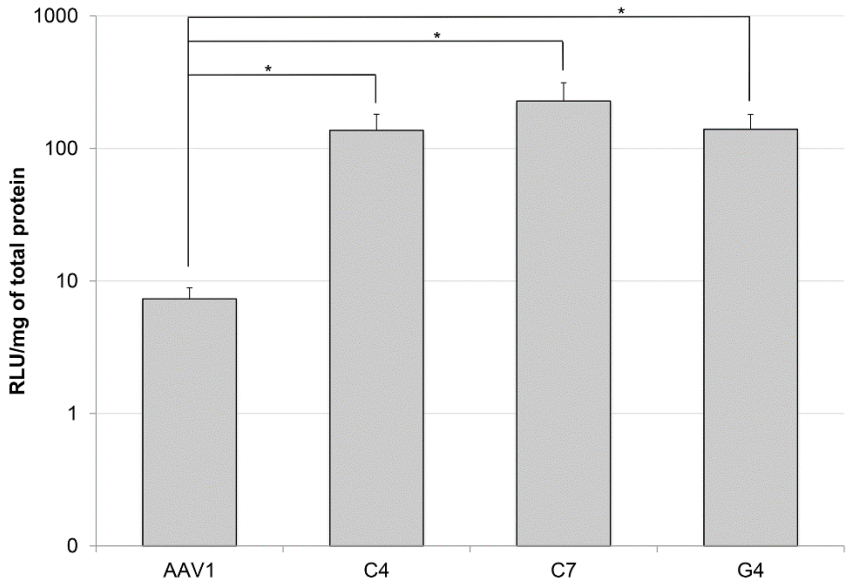


Figure 3.9. Evaluation of gastrocnemius muscle transduction. Luciferase activity measured in relative light units (RLU) per mg protein was determined in gastrocnemius tissue homogenate 48 days after intramuscular administration

of 5 x 10^{10} viral particles of ancestral clones C4, C7, G4, or AAV1 in adult mice. Controls injected with phosphate-buffered saline displayed no activity (data not shown). *, statistical difference of $P < 0.05$ by two-tailed Student's t-test.

3.3 Discussion

Ancestral sequence reconstruction offers unique opportunities to study fundamental biological questions of virus evolution and fitness, including the characterization of ancestral sequence space relative to extant serotypes, the importance of mutational tolerance or evolutionary conservation, and the comparative advantages of promiscuous versus selective tropism. The primary challenge of ancestral reconstruction is to accurately infer an ancestral sequence despite uncertainty arising from sequence divergence within hypervariable regions of extant variants. We have combined sophisticated computational and library synthesis approaches to address this uncertainty and thereby generate a functional ancestral AAV library. We then studied the biological properties of this library to learn more about the evolutionary behavior of AAV and the gene therapy potential of reconstructed ancestral variants.

The posterior probability that an AAV ancestral sequence accurately reflects the actual ancestral virus is the product of the probabilities that each of the amino acids in the capsid protein is correctly predicted. At positions of high evolutionary convergence the posterior probability nears 1.0, yet there are many sites that diverged during evolution and thus cannot be predicted with such high confidence. Our library synthesis approach addressed this concern by introducing the two or three most likely amino acids at the 32 lowest confidence positions in the AAV *cap* protein. Interestingly, the majority of positions varied in our ancestral library have not been previously described in studies of the functional importance of single mutations to the AAV capsid [36, 37]. Unlike previous ancestral reconstructions of enzymes and other proteins, which utilized single best guess ancestral sequences [11, 38], or which sampled only a small fraction of library variants due to the low throughput of enzymatic assays [8, 39], our massively parallel phenotypic selection enabled screening of a large library (and is limited only by the transformation efficiency of electrocompetent bacteria).

The selection strategy applied pressure for efficient packaging and transduction of cell types representing a variety of tissues. By comparing the frequencies of amino acids selected at variable positions to the theoretical ancestral sequence prediction, one can gain insights into both the accuracy of our sequence reconstruction as well as the functional role of each residue in AAV biology. Comparison of sequences from the synthesized library with those recovered after initial library packaging suggested that one round of packaging imposed no statistically significant changes on the amino acid distribution at variable positions (except for a low 0.076 probability change in preference from a threonine to an alanine at residue 264). However, with selection for infectivity on a range of cell types specific positions begin to diverge, and differences between round six and post-packaging sequences were more significant than between round three and post-packaging sequences, likely because six rounds enabled a larger cumulative effect of positive selection. Genetic drift may also play a role, but is unlikely to be the main driving force given that the time to fixation by genetic drift increases with population size [40], and a large number of virions ($>10^8$) was used in each sequential round of selection.

By comparing the ancestral libraries after six rounds of selection with the post-packaging library, we identified several trends in the level of convergence of the amino acid residues, suggesting these positions may have potential roles in modulating properties like capsid stability and infectivity. Some amino acid positions approached full convergence to the same residue across all cell lines (268, 460, 474, 516, 547, 583, 665, 710, 717, 719); these positions are distributed throughout the capsid and may for example be important for core viral functions such as capsid stability, uncoating, or endosomal escape. Others showed more divergent outcomes across different cell lines (264, 467, 593, 664, 723) and may be neutral with respect to overall fitness. Finally, some positions (459, 470, 471, 533, 555, 596, 662, 718) acquired identities specific to a given cell line and may confer an infectious advantage on each respective cell line.

Positions 264 and 459 showed the strongest evidence of change due to selection ($P < 0.05$). Position 459 is prominently exposed on loop IV of the AAV capsid surface. Position 264 is positioned on loop I of the capsid and has been identified as a key determinant of muscle tropism in the rationally engineered variant AAV2.5 [41]. There is also suggestive evidence of selection at positions 266, 470, 533, 551, 557, 577, 596, and 723 in various libraries ($P < 0.5$). Position 533 has been previously described as a key contributor to infectivity and glycan dependence in our previously evolved variant ShH10, a vector differing by only four amino acids from AAV6 but exhibiting unique tropism in the retina[34]. Additionally, Lochrie et al. [42] examined several other of these positions in their thorough mutational analysis of the AAV2 serotype, though AAV2 lies in a different phylogenetic clade than ancestral node 27. The characterization of these variable positions is therefore novel, and lessons learned may inform targeted mutagenesis efforts to improve the fitness of extant variants.

The assembly-activating protein (AAP), which is involved in directing capsid proteins to the nucleolus and in assembly of the viral capsid in this organelle [20], is translated from an alternative open reading frame (ORF) with a non-canonical CTG start codon present within the *cap* gene. This alternate ORF is also present in the ancestral reconstruction of the AAV capsid at node 27 (Fig. B.2b). Three of the variable residues (positions 264, 266, and 268) are present within the AAP ORF, and the putative ancestral AAP sequence is otherwise conserved across the reconstruction. As discussed above, residue 264 is among the positions that showed strong and statistically significant changes after six rounds of packaging and infection on the cell line panel, and it is possible that both capsid and AAP may have undergone functional selective pressure during this process.

The *in vitro* transduction results also demonstrate the importance of utilizing a library approach coupled with selection. A single or small number of best guess sequences could likely include deleterious amino acids that significantly impact fitness. Indeed, our data show that the synthetized ancestral AAV library evaluated prior to rounds of selection reproducibly exhibited dramatically lower infectivity than libraries subjected to selective pressure. This is not surprising, given that numerous directed evolution studies demonstrate that even single point mutations can significantly alter enzyme activity or virus infectivity by several orders of magnitude [36, 43-45].

Interestingly, despite differences in amino acid composition at variable positions, the ancestral libraries selected for infecting a number of individual cell lines subsequently demonstrated broad tropism across all of these cell lines. Such promiscuity may have been

rewarded during the natural evolution of AAVs, since the ability to replicate in different cell and tissue types enhances virus spread. In fact, most natural AAV serotypes exhibit broad tropism[46, 47], indicating that promiscuity continues to be a valued trait for natural evolution.

Such broad tropism, however, has important implications for gene therapy. In cases where disease pathologies affect multiple tissues and cell types (e.g. lysosomal storage disorders), broader infectivity could be an advantageous trait. Expanded tropism may also be useful for infecting cell types refractory to infection by most AAV serotypes, or for *ex vivo* treatments of homogeneous cell populations where off-target infectivity is not a concern. However, in the majority of gene therapy applications it is desirable to limit transgene expression to a target tissue for several important reasons, including risks associated with off-target expression of the transgene, off-target transduction leading to higher immune presentation and reaction, and higher overall dosages needed to overcome vector dilution into multiple tissues. This is true not only when vector is delivered via routes that lead to intentional exposure to multiple tissues (e.g. intravascular delivery) but also for local injection into multiple tissues in which vector leakage into circulation can lead to widespread distribution to multiple organs. For example, biodistribution studies have shown the spread of AAV vectors to sites distant from the target tissue after injecting viral particles through the hepatic artery, intramuscularly, or into the putamen of the brain [48-50].

To address concerns with off-target transgene expression, strategies for controlling gene and protein expression including cell type specific promoters [51] and microRNA elements [52, 53] are being explored to restrict expression to target cells. These approaches are promising, but would not address immune presentation of the capsid protein. Therefore, the optimal scenario is one in which selective AAV tropism is engineered through modification of the capsid protein. Directed evolution can generate vectors capable of targeted gene delivery [34], and evolution for enhanced AAV infectivity of a given target cell in general can enable a reduction in vector dose and thereby reduce the level of off-target transduction. Ancestral variants may be promising starting points for such directed evolution efforts given their high infectivity and representation of a capsid protein sequence space that is different from and complementary to extant serotypes.

High thermostability may also be an advantageous property for AAV engineering. Ancestral sequences have been correlated with increased thermostability in multiple studies [3, 7, 32], and in fact, enriching for seemingly neutral mutations that resemble an ancestral sequence has been shown to increase protein kinetic and thermodynamic stability and to improve the probability of acquiring new function mutations [54]. This work lends additional evidence of the correlation between ancestral sequences and thermostability by demonstrating that candidate ancestral AAV variants are more thermostable than contemporary serotypes.

We also characterized the glycan dependencies of ancestral variants and found that previously studied AAV glycan dependencies including N- and O-linked sialic acids, heparan sulfate proteoglycans, and galactose were not utilized. It is conceivable that these dependences may have arisen more recently in the evolution along these AAV lineages.

In addition, we found that ancestral libraries were as susceptible to neutralizing antibodies as AAV1, suggesting that this ancestral reconstruction pool exhibits immunogenic properties

similar to current serotypes. Multiple antigenic regions have been mapped on natural AAV serotypes, including AAV1, AAV2, AAV5, and AAV8 [55]. Given that AAV1 is a descendant of the node 27 ancestral reconstruction, we aligned known AAV1 epitopes [56] with the ancestral reconstruction sequence. Mapped antigenic regions corresponding to AAV1 residues 496-499, 583, 588-591, and 597 were conserved in the ancestral reconstruction. Additionally, the ancestral sequence is identical to several known AAV2 antigenic regions including residues 272-281, 369-378, and 562-573 [57]. Such conserved regions may contribute to the observed susceptibility of ancestral variants to neutralizing antibodies.

Interestingly, previous studies have also demonstrated cross-seroreactivity between ancestral and extant viral capsids. In particular, antiserum against extant viruses has been shown to neutralize reconstructed ancestral variants [14], and ancestral viruses can elicit neutralizing antibodies that protect against currently circulating strains, a property that has been exploited for the development of vaccine candidates [11, 12]. Neutralizing antibodies may therefore pose a significant clinical challenge for ancestral vectors. Further capsid engineering under a strong selective pressure for evading neutralizing antibodies may enable selection of combinations of mutations that promote antibody evasion [58]. For example, there are variable residues in the ancestral reconstruction that map to antigenic regions corresponding to AAV1 residues 456-459, 494, 582, and 593-595, and to antigenic regions in other serotypes [56]. Mutations in these regions could disrupt the binding of antibodies to capsid epitopes, and could potentially be combined with other mutagenesis strategies to engineer variants with enhanced antibody evasion properties [58].

Ancestral AAVs demonstrated efficient *in vitro* gene transfer to C2C12 mouse myoblast cells comparable to AAV1, a current gold standard for muscle transduction, yet utilized a different receptor for cell entry. This distinction may contribute to their efficient *in vivo* infectivity, which impressively reached 19-31 fold higher levels of expression than AAV1 in mouse gastrocnemius muscle. If the improved expression observed with ancestral reconstruction vectors is reproducible in human muscle tissue, ancestral variants will be auspicious candidates for clinical translation.

In summary, our results indicate that a library of AAV variants representing sequence space around a key ancestral node is rich in broadly infectious variants with potential in gene therapy applications. We have taken initial steps in characterizing this sequence space by varying the amino acids at the lowest confidence positions identified by ancestral sequence reconstruction, followed by phenotypic selection to yield highly functional sets of amino acids at these locations. Sequence analysis of variable residues revealed a variety of outcomes ranging from highly conserved residues to more neutral positions that are pliable to change. Selected variants were promiscuous in their infectivity but showed promise as recombinant vectors *in vitro* and *in vivo*, and the putative mutational tolerance and evolvability of this library could be further harnessed in directed evolution studies to overcome gene therapy challenges such as targeted gene delivery and immune evasion.

3.4 Materials and Methods

Ancestral reconstruction

AAV *cap* sequences (n=52) from Genbank [22], including those from human and non-human primate origin, were incorporated in this analysis, starting from lists of AAV sequences published in previous phylogenetic analyses [59, 60]. The MrBayes package [23] was used to perform Bayesian Markov chain Monte Carlo (MCMC) simulation of tree space and estimate the confidence values at each internal node. We then used the Markov chain Monte Carlo alignment sampler HandAlign [27] to explore alignment space and estimate regional confidence for the most likely alignment at node 27, discarding all but the sequences descended from this node. HandAlign generates a multiple sequence alignment, arranging the sequences of different variants in aligned 'columns' such that residues grouped in a column share a common ancestor. Each alignment column was modeled as a realization of the standard phylogenetic continuous-time Markov process of character evolution, using amino acid and empirical codon substitution rate matrices that were estimated from databases of aligned protein-coding sequence [61]. HandAlign performs the reconstruction simultaneously with the alignment, and accounts for sequence insertions, deletions, and character substitutions. The codon-level model was used to account for the possibility of synonymous substitutions with a phenotype at the DNA level; we also checked for the possibility of dual selection in overlapping reading frames ("overprinted" genes), by reconstructing both ancestral reading frames at the codon level. Neither of these subtle effects appeared significant enough to warrant prioritizing synonymous (silent, DNA-level) variants over the many non-synonymous amino acid variants.

Library construction and vector packaging

The reconstructed ancestral AAV *cap* sequence was synthesized (GeneArt, Life Technologies) with a library size of 5.6×10^{11}, greater than the theoretical diversity of 2.5×10^{11}. The library was digested with *Hin*d III and *Not* I, and ligated into the replication competent AAV packaging plasmid pSub2. The resulting ligation reaction was electroporated into *E. coli* for plasmid production and purification. Replication competent AAV was then packaged and purified by iodixanol density centrifugation as previously described [58, 62]. DNase-resistant genomic titers were obtained via quantitative real time PCR using a Bio-Rad iCycler (Bio-Rad, Hercules, CA) and Taqman probe (Biosearch Technologies, Novato, CA)[62].

Cell culture

C2C12 mouse myoblast, B16-F10 skin melanoma cells, CHO-K1, pgsA, Pro5, Lec1, and Lec2 cells were obtained from the Tissue Culture Facility at the University of California, Berkeley. IB3-1 lung epithelial and human embryonic kidney 293T cells were obtained from American Type Culture Collection (Manassas, VA). Unless otherwise noted all cell lines were cultured in Dulbecco's Modified Eagle's medium (DMEM, Gibco) at 37 °C and 5% CO_2. L0 human glioblastoma tumor initiating cells were kindly provided by Dr. Brent Reynolds (University of Florida, Gainesville), and propagated in neurosphere assay growth conditions [63] with serum-free media (Neurocult NS-A Proliferation kit, Stem Cell Technologies) that contained epidermal

growth factor (EGF, 20 ng/ml, R&D), basic fibroblast growth factor (bFGF, 10 ng/ml, R&D), and heparin (0.2% diluted in phosphate buffered saline, Sigma). IB3-1 cells were cultured in DMEM/F-12 (1:1) (Invitrogen, Carlsbad, CA). CHO-K1 and pgsA cells were cultured in F-12K medium (ATCC), and Pro5, Lec1, and Lec2 cells were cultured in MEM α nucleosides (Gibco). Except for GBM culture, all media were supplemented with 10% fetal bovine serum (Invitrogen) and 1% penicillin/streptomycin (Invitrogen).

Library selection and evolution

All cell lines were seeded in 6-well tissue culture plates at a density of 1 x 10^5 cells per well. One day after seeding, cells were infected with replication competent AAV libraries. After 24 hours of exposure, cells were superinfected with adenovirus serotype 5 (Ad5). Approximately 48 hours later, cytopathic effect was observed, and virions were harvested by three freeze/thaw steps followed by treatment with Benzonase nuclease (1 unit/mL) (Sigma-Aldrich) at 37 °C for 30 minutes. Viral lysates were then incubated at 56°C for 30 minutes to inactivate Ad5. The viral genomic titer was determined as described above. To analyze *cap* sequences, AAV viral genomes were extracted after packaging and rounds 3 and 6 of selection, amplified by PCR, and sequenced at the UC Berkeley DNA Sequencing Facility.

Statistical analysis of variable positions in evolved ancestral libraries

A comparison of the two sets of amino acids at each variable amino acid position was conducted to identify variable positions whose library proportions had changed significantly during selection. The posterior probability that the two sets of variable amino acids come from two different probability distributions was calculated assuming probability parameters that are Dirichlet-distributed with low pseudocounts to reflect sparse observed counts. For comparison of the synthesized and theoretical library, post-synthesis amino acid frequencies distributed via a Dirichlet-multinomial were compared with the theoretical probabilities from the library distributed by a multinomial.

In vitro transduction analysis

After six rounds of selection, ancestral library viral genomes were cloned into the pXX2 recombinant AAV packaging plasmid. To benchmark the infectivity of rAAV ancestral libraries against a panel of natural AAV serotypes, vectors were packaged with a self-complementary CMV-GFP cassette using the transient transfection method previously described [58, 62]. Cell lines (293T, C2C12, IB3-1, B16-F10, CHO-K1, pgsA, Pro5, Lec1, and Lec2) were seeded in 96-well plates at a density of 15,000 cells per well. One day after seeding, cells were infected with rAAV at a genomic MOI of 2,000 (293T, C2C12, IB3-1, B16-F10, GBM), 10,000 (Pro5, Lec1, Lec2), 32,000 (C2C12), or 50,000 (CHO-K1, pgsA) (n = 3). To analyze antibody evasion properties, ancestral rAAV libraries were incubated at 37°C for 1 hour with serial dilutions of heat inactivated IVIG (Gammagard), and then used to infect HEK293T cells at a genomic MOI of 2000 (n = 3).

To characterize thermostability, virions packaged with self-complementary CMV-GFP were diluted with DMEM supplemented with 2% FBS and incubated at temperatures ranging from 59.6°C to 78°C for 10 minutes in a thermocycler (Bio-Rad) before being cooled down to 37°C and used to infect 293T cells at genomic MOIs ranging from 1,500-16,000; MOIs were adjusted to ensure an adequate number of GFP-positive cells for analysis. For all studies, the fraction of GFP-expressing cells 72 hours post-infection was quantified with a Guava EasyCyte 6HT flow cytometer (EMD/Millipore) (UC Berkeley Stem Cell Center, Berkeley, CA).

In vivo animal imaging and quantification of luciferase expression

High-titer rAAV CAG-Luciferase vectors were purified by iodixanol gradient and then concentrated and exchanged into PBS using Amicon Ultra-15 centrifugal filter units (Millipore). To study skeletal muscle transduction 5×10^{10} rAAV-Luc DNase-resistant genomic particles were injected in a volume of 30 µl into each gastrocnemius muscle of 7-week-old female BALB/c mice (Jackson Laboratories, *n = 3*) as previously described [58]. Six weeks after injection, animals were sacrificed, and gastrocnemius muscle was harvested and frozen. Luciferase activity was determined and normalized to total protein as previously described [62]. All animal procedures were approved by the Office of Laboratory Animal Care at the University of California, Berkeley and conducted in accordance with NIH guidelines on laboratory animal care.

3.5 Acknowledgements

The authors are grateful to Professor Brent Reynolds (University of Florida) for kindly providing the L0 human glioblastoma tumor-initiating cells.

3.6 Funding

This work was supported by the National Institutes of Health grant [R01EY022975]. DSO is supported by a National Science Foundation Graduate Fellowship, and JSO is supported by a National Science Foundation Graduate Fellowship and a UC Berkeley Graduate Division Fellowship. IH and OW were supported by the National Human Genome Research Institute grant [HG004483].

Conflict of interest statement. DVS, DSO, and JSO are inventors on patents involving AAV directed evolution.

3.7 References

1. Stackhouse, J., et al., *The ribonuclease from an extinct bovid ruminant.* FEBS Lett, 1990. **262**(1): p. 104-6.
2. Cole, M.F. and E.A. Gaucher, *Utilizing natural diversity to evolve protein function: applications towards thermostability.* Curr Opin Chem Biol, 2011. **15**(3): p. 399-406.
3. Gaucher, E.A., S. Govindarajan, and O.K. Ganesh, *Palaeotemperature trend for Precambrian life inferred from resurrected proteins.* Nature, 2008. **451**(7179): p. 704-7.

4. Ortlund, E.A., et al., *Crystal structure of an ancient protein: evolution by conformational epistasis.* Science, 2007. **317**(5844): p. 1544-8.
5. Ugalde, J.A., B.S. Chang, and M.V. Matz, *Evolution of coral pigments recreated.* Science, 2004. **305**(5689): p. 1433.
6. Afriat-Jurnou, L., C.J. Jackson, and D.S. Tawfik, *Reconstructing a missing link in the evolution of a recently diverged phosphotriesterase by active-site loop remodeling.* Biochemistry, 2012. **51**(31): p. 6047-55.
7. Watanabe, K., et al., *Designing thermostable proteins: ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree.* J Mol Biol, 2006. **355**(4): p. 664-74.
8. Alcolombri, U., M. Elias, and D.S. Tawfik, *Directed evolution of sulfotransferases and paraoxonases by ancestral libraries.* J Mol Biol, 2011. **411**(4): p. 837-53.
9. Chen, F., et al., *Reconstructed evolutionary adaptive paths give polymerases accepting reversible terminators for sequencing and SNP detection.* Proc Natl Acad Sci U S A, 2010. **107**(5): p. 1948-53.
10. Gonzalez, D., et al., *Ancestral mutations as a tool for solubilizing proteins: The case of a hydrophobic phosphate-binding protein.* FEBS Open Bio, 2014. **4**: p. 121-7.
11. Kothe, D.L., et al., *Ancestral and consensus envelope immunogens for HIV-1 subtype C.* Virology, 2006. **352**(2): p. 438-49.
12. Ducatez, M.F., et al., *Feasibility of reconstructed ancestral H5N1 influenza viruses for cross-clade protective vaccine development.* Proc Natl Acad Sci U S A, 2011. **108**(1): p. 349-54.
13. Rolland, M., et al., *Reconstruction and function of ancestral center-of-tree human immunodeficiency virus type 1 proteins.* J Virol, 2007. **81**(16): p. 8507-14.
14. Gullberg, M., et al., *Characterization of a putative ancestor of coxsackievirus B5.* J Virol, 2010. **84**(19): p. 9695-708.
15. Berns, K.I. and R.M. Linden, *The cryptic life style of adeno-associated virus.* Bioessays, 1995. **17**(3): p. 237-45.
16. Ellis, B.L., et al., *A survey of ex vivo/in vitro transduction efficiency of mammalian primary cells and cell lines with Nine natural adeno-associated virus (AAV1-9) and one engineered adeno-associated virus serotype.* Virol J, 2013. **10**: p. 74.
17. Asokan, A., D.V. Schaffer, and R.J. Samulski, *The AAV vector toolkit: poised at the clinical crossroads.* Mol Ther, 2012. **20**(4): p. 699-708.
18. Kotterman, M.A. and D.V. Schaffer, *Engineering adeno-associated viruses for clinical gene therapy.* Nat Rev Genet, 2014. **15**(7): p. 445-51.
19. Gaudet, D., et al., *Efficacy and long-term safety of alipogene tiparvovec (AAV1-LPLS447X) gene therapy for lipoprotein lipase deficiency: an open-label trial.* Gene Ther, 2013. **20**(4): p. 361-9.
20. Sonntag, F., K. Schmidt, and J.A. Kleinschmidt, *A viral assembly factor promotes AAV2 capsid formation in the nucleolus.* Proc Natl Acad Sci U S A, 2010. **107**(22): p. 10220-5.
21. Khersonsky, O., et al., *Directed evolution of serum paraoxonase PON3 by family shuffling and ancestor/consensus mutagenesis, and its biochemical characterization.* Biochemistry, 2009. **48**(28): p. 6644-54.
22. Benson, D.A., et al., *GenBank.* Nucleic Acids Res, 2014. **42**(Database issue): p. D32-7.
23. Huelsenbeck, J.P. and F. Ronquist, *MRBAYES: Bayesian inference of phylogeny.* Bioinformatics, 2001. **17**(17): p. 754-755.

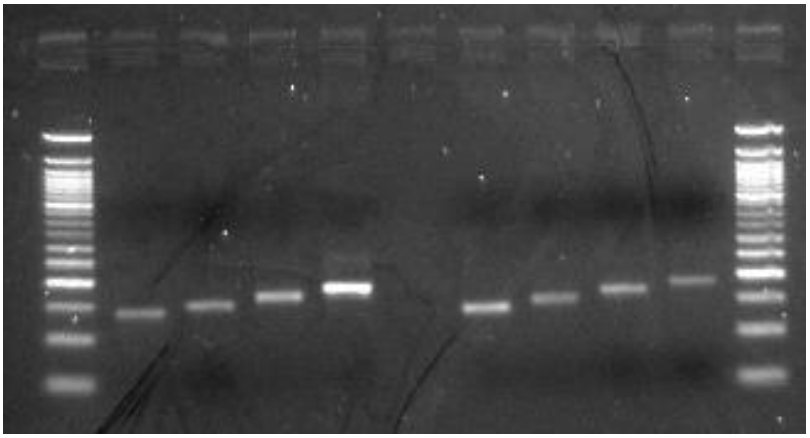
24. Wu, Z., et al., *Alpha2,3 and alpha2,6 N-linked sialic acids facilitate efficient binding and transduction by adeno-associated virus types 1 and 6.* J Virol, 2006. **80**(18): p. 9093-103.
25. Gao, G.P., et al., *Novel adeno-associated viruses from rhesus monkeys as vectors for human gene therapy.* Proc Natl Acad Sci U S A, 2002. **99**(18): p. 11854-9.
26. Huson, D.H. and C. Scornavacca, *Dendroscope 3: an interactive tool for rooted phylogenetic trees and networks.* Syst Biol, 2012. **61**(6): p. 1061-7.
27. Westesson, O., L. Barquist, and I. Holmes, *HandAlign: Bayesian multiple sequence alignment, phylogeny and ancestral reconstruction.* Bioinformatics, 2012. **28**(8): p. 1170-1.
28. Koerber, J.T., et al., *Construction of diverse adeno-associated viral libraries for directed evolution of enhanced gene delivery vehicles.* Nat Protoc, 2006. **1**(2): p. 701-6.
29. Schrodinger, LLC, *The PyMOL Molecular Graphics System, Version 1.3r1*. 2010.
30. Bloom, J.D., et al., *Neutral genetic drift can alter promiscuous protein functions, potentially aiding functional evolution.* Biol Direct, 2007. **2**: p. 17.
31. Aharoni, A., et al., *The 'evolvability' of promiscuous protein functions.* Nat Genet, 2005. **37**(1): p. 73-6.
32. Thornton, J.W., *Resurrecting ancient genes: experimental analysis of extinct molecules.* Nat Rev Genet, 2004. **5**(5): p. 366-75.
33. Rayaprolu, V., et al., *Comparative analysis of adeno-associated virus capsid stability and dynamics.* J Virol, 2013. **87**(24): p. 13150-60.
34. Klimczak, R.R., et al., *A novel adeno-associated viral variant for efficient and selective intravitreal transduction of rat Muller cells.* PLoS One, 2009. **4**(10): p. e7467.
35. Hauck, B. and W. Xiao, *Characterization of tissue tropism determinants of adeno-associated virus type 1.* J Virol, 2003. **77**(4): p. 2768-74.
36. Vandenberghe, L.H., et al., *Naturally occurring singleton residues in AAV capsid impact vector performance and illustrate structural constraints.* Gene Ther, 2009. **16**(12): p. 1416-28.
37. Wu, Z., et al., *Single amino acid changes can influence titer, heparin binding, and tissue tropism in different adeno-associated virus serotypes.* J Virol, 2006. **80**(22): p. 11393-7.
38. Thornton, J.W., E. Need, and D. Crews, *Resurrecting the ancestral steroid receptor: ancient origin of estrogen signaling.* Science, 2003. **301**(5640): p. 1714-7.
39. Voordeckers, K., et al., *Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication.* PLoS Biol, 2012. **10**(12): p. e1001446.
40. Lanfear, R., H. Kokko, and A. Eyre-Walker, *Population size and the rate of evolution.* Trends Ecol Evol, 2014. **29**(1): p. 33-41.
41. Bowles, D.E., et al., *Phase 1 gene therapy for Duchenne muscular dystrophy using a translational optimized AAV vector.* Mol Ther, 2012. **20**(2): p. 443-55.
42. Lochrie, M.A., et al., *Mutations on the external surfaces of adeno-associated virus type 2 capsids that affect transduction and neutralization.* J Virol, 2006. **80**(2): p. 821-34.
43. Schmidt, D.M., et al., *Evolutionary potential of (beta/alpha)8-barrels: functional promiscuity produced by single substitutions in the enolase superfamily.* Biochemistry, 2003. **42**(28): p. 8387-93.
44. Asuri, P., et al., *Directed evolution of adeno-associated virus for enhanced gene delivery and gene targeting in human pluripotent stem cells.* Mol Ther, 2012. **20**(2): p. 329-38.

45. Varadarajan, N., et al., *Engineering of protease variants exhibiting high catalytic activity and exquisite substrate selectivity.* Proc Natl Acad Sci U S A, 2005. **102**(19): p. 6855-60.
46. Zincarelli, C., et al., *Analysis of AAV serotypes 1-9 mediated gene expression and tropism in mice after systemic injection.* Mol Ther, 2008. **16**(6): p. 1073-80.
47. Wu, Z., A. Asokan, and R.J. Samulski, *Adeno-associated virus serotypes: vector toolkit for human gene therapy.* Mol Ther, 2006. **14**(3): p. 316-27.
48. Schenk-Braat, E.A., et al., *An inventory of shedding data from clinical gene therapy trials.* J Gene Med, 2007. **9**(10): p. 910-21.
49. Toromanoff, A., et al., *Safety and efficacy of regional intravenous (r.i.) versus intramuscular (i.m.) delivery of rAAV1 and rAAV8 to nonhuman primate skeletal muscle.* Mol Ther, 2008. **16**(7): p. 1291-9.
50. Ciron, C., et al., *Human alpha-iduronidase gene transfer mediated by adeno-associated virus types 1, 2, and 5 in the brain of nonhuman primates: vector diffusion and biodistribution.* Hum Gene Ther, 2009. **20**(4): p. 350-60.
51. Gray, S.J., et al., *Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors.* Hum Gene Ther, 2011. **22**(9): p. 1143-53.
52. Qiao, C., et al., *Liver-specific microRNA-122 target sequences incorporated in AAV vectors efficiently inhibits transgene expression in the liver.* Gene Ther, 2011. **18**(4): p. 403-10.
53. Geisler, A., et al., *microRNA122-regulated transgene expression increases specificity of cardiac gene transfer upon intravenous delivery of AAV9 vectors.* Gene Ther, 2011. **18**(2): p. 199-209.
54. Bershtein, S., K. Goldin, and D.S. Tawfik, *Intense neutral drifts yield robust and evolvable consensus proteins.* J Mol Biol, 2008. **379**(5): p. 1029-44.
55. Tseng, Y.S. and M. Agbandje-McKenna, *Mapping the AAV Capsid Host Antibody Response toward the Development of Second Generation Gene Delivery Vectors.* Front Immunol, 2014. **5**: p. 9.
56. Gurda, B.L., et al., *Capsid antibodies to different adeno-associated virus serotypes bind common regions.* J Virol, 2013. **87**(16): p. 9111-24.
57. Wobus, C.E., et al., *Monoclonal antibodies against the adeno-associated virus type 2 (AAV-2) capsid: epitope mapping and identification of capsid domains involved in AAV-2-cell interaction and neutralization of AAV-2 infection.* J Virol, 2000. **74**(19): p. 9281-93.
58. Maheshri, N., et al., *Directed evolution of adeno-associated virus yields enhanced gene delivery vectors.* Nat Biotechnol, 2006. **24**(2): p. 198-204.
59. Gao, G., et al., *Clades of Adeno-associated viruses are widely disseminated in human tissues.* J Virol, 2004. **78**(12): p. 6381-8.
60. Takeuchi, Y., R. Myers, and O. Danos, *Recombination and population mosaic of a multifunctional viral gene, adeno-associated virus cap.* PLoS One, 2008. **3**(2): p. e1634.
61. Westesson, O. and I. Holmes, *Developing and applying heterogeneous phylogenetic models with XRate.* PLoS One, 2012. **7**(6): p. e36898.
62. Koerber, J.T., J.H. Jang, and D.V. Schaffer, *DNA shuffling of adeno-associated virus yields functionally diverse viral progeny.* Mol Ther, 2008. **16**(10): p. 1703-9.
63. Deleyrolle, L.P., et al., *Evidence for label-retaining tumour-initiating cells in human glioblastoma.* Brain, 2011. **134**(Pt 5): p. 1331-43.

Appendix A: Supplementary Material for Chapter 2

A.1 Supplementary Figures and Tables

Figure A.1. 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 A.2. 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 A.3. 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.

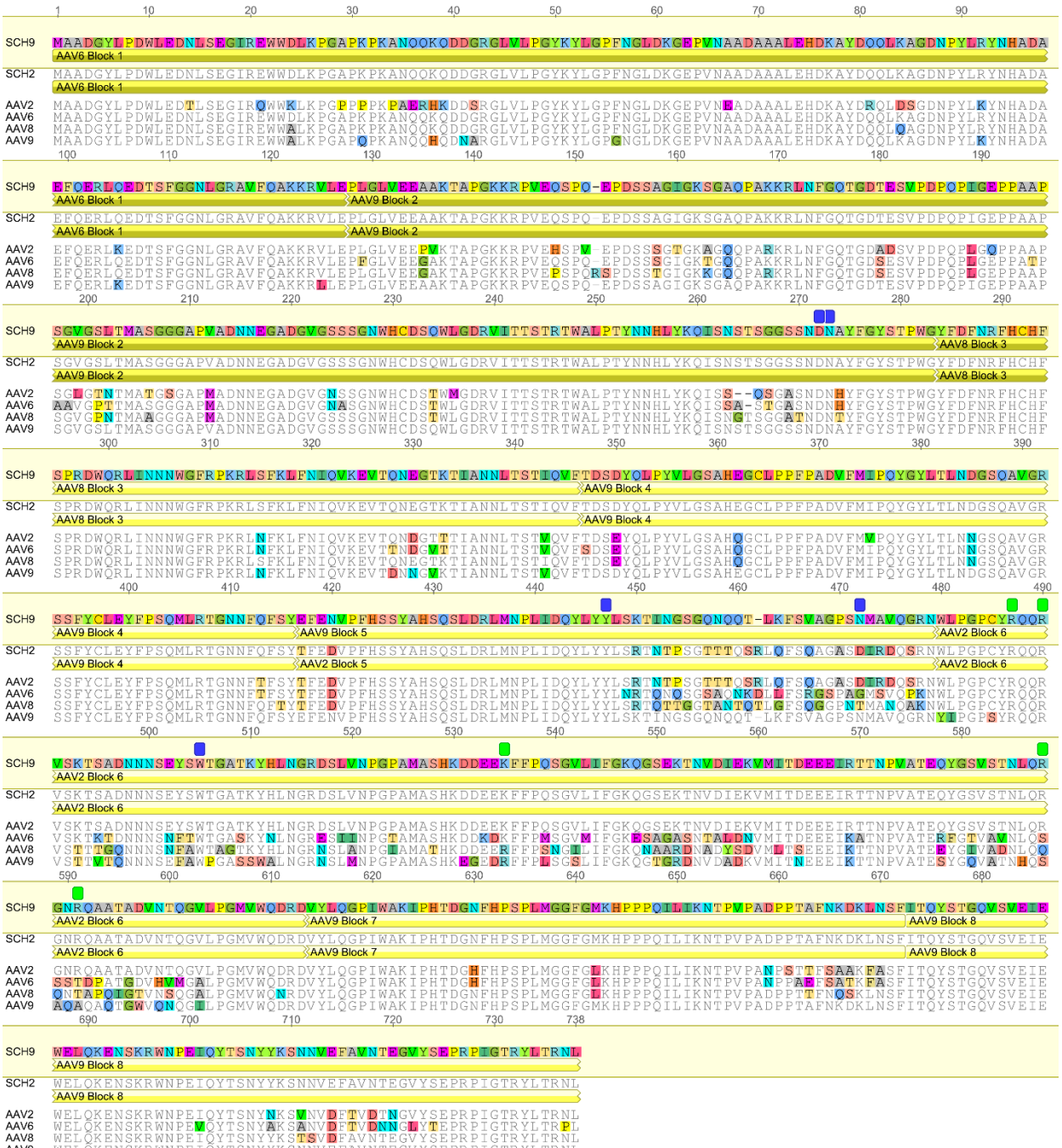


Figure A.4. Viral genomic yield of recombinant self-complementary AAV vectors. Data are normalized to the cell culture surface area and are presented as mean ± SD, n=3.

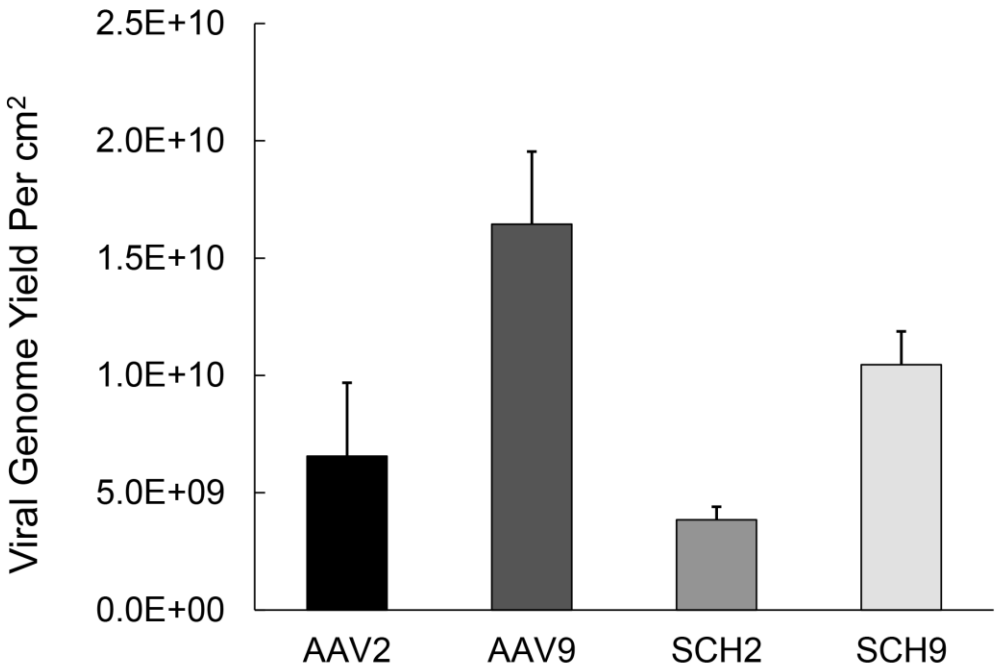


Figure A.5. GFP expression in the cerebellum three weeks after unilateral injection of recombinant AAV1 or SCH9 into the deep cerebellar nuclei. Coronal sections were stained for GFP (green) and the Purkinje cell marker calbindin (purple). Scale bars indicate 500 µm.

Figure A.6. 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.

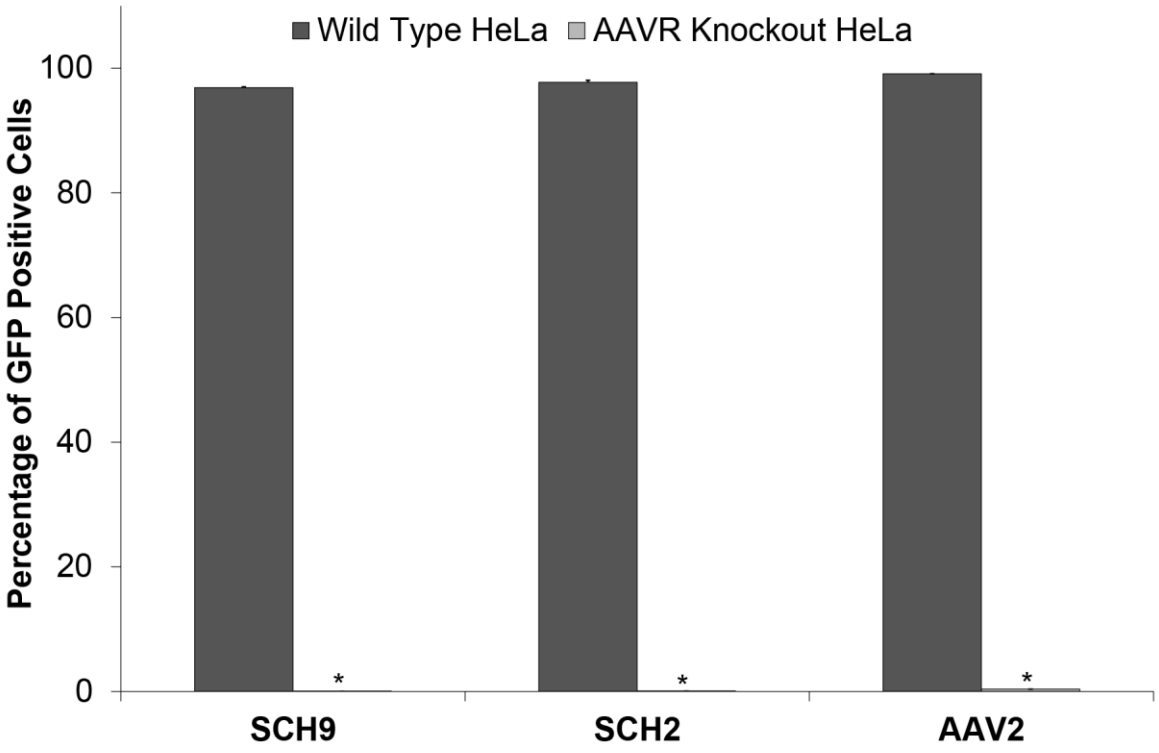


Table A.1. Neutralizing IVIG titers of SCH9 and the parent serotypes from which it is derived. The neutralizing titers represent the first IVIG concentration at which 50% or greater reduction in GFP expression was observed.

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 A.2. 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 A.3. 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 A.4. 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 A.5. 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

Appendix B: Supplementary Material for Chapter 3

This appendix contains material adapted from a manuscript published as

J. Santiago-Ortiz*, D. Ojala*, O. Westesson, J. Weinstein, S. Wong, A. Steinsapir, S. Kumar, I. Holmes, D. Schaffer. AAV Ancestral Reconstruction Library Enables Selection of Broadly Infectious Viral Variants. *Gene Therapy* **22**, 934-946 (2015).

* Indicates co-first authors.

B.1 Supplementary Figures and Tables

Figure B.1. Full phylogenetic tree for AAV ancestral sequence reconstruction. Curly braced numbers indicate clade posterior probabilities. The phylogenetic tree graphic was generated in Dendroscope.

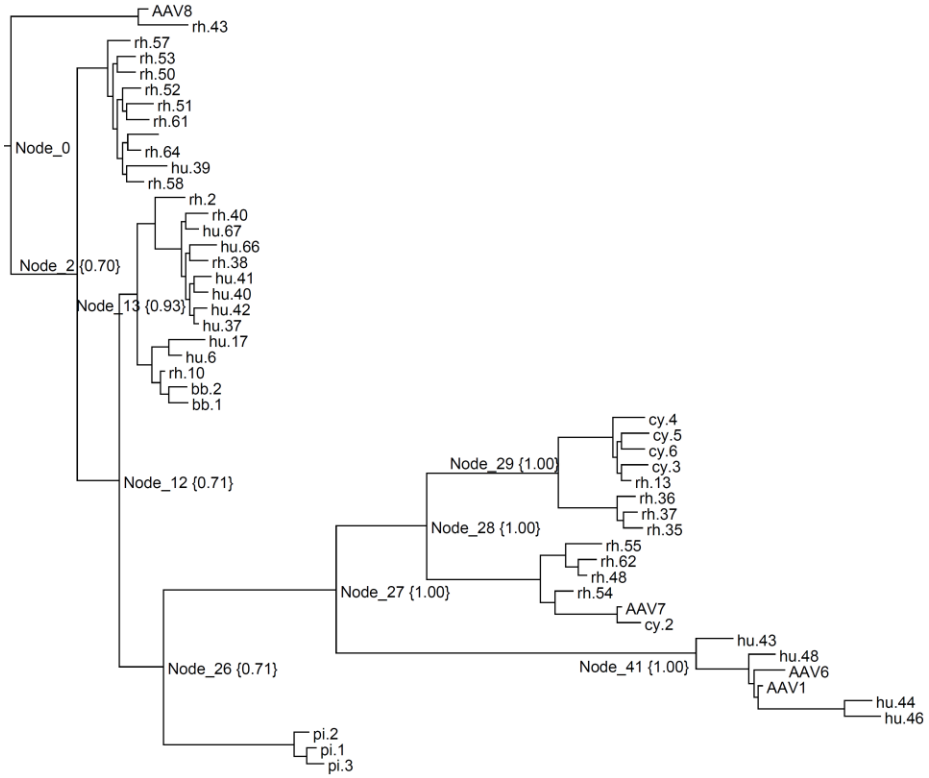


Figure B.2. Amino acid sequences of the ancestral AAV (a) cap and (b) AAP reading frames. Variable residues are labeled with a bold, underlined letter X. In the AAP sequence, the shifted reading frame results in four variable residues corresponding to AAP positions 88, 90, 91, and 92.

a)

MAADGYLPDWLEDNLSEGIREWWDLKPGAPKPKANQQKQDDGRGLVLPGYKYLGPFNGLDKGEPVNAA
DAAALEHDKAYDQQLKAGDNPYLRYNHADAEFQERLQEDTSFGGNLGRAVFQAKKRVLEPLGLVEEGA
KTAPGKKRPVEPSPQRSPDSSTGIGKKGQQPAKKRLNFGQTGDSESVPDPQPLGEPPAGPSGLGSGTMAAG
GGAPMADNNEGADGVGNASGNWHCDSTWLGDRVITTSTRTWALPTYNNHLYKQISS**X**S**X**G**X**TNDNHYF
GYSTPWGYFDFNRFHCHFSPRDWQRLINNNWGFRPKRLNFKLFNIQVKEVTTNDGVTTIANNLTSTVQVFS
DSEYQLPYVLGSAHQGCLPPFPADVFMIPQYGYLTLNNGSQAVGRSSFYCLEYFPSQMLRTGNNFTFSYTF
EDVPFHSSYAHSQSLDRLMNPLIDQYLYYL**X**RTQSTGGTAG**XX**ELLFSQ**X**GP**XX**MS**X**QAKNWLPGPCYRQ
QRVSKTL**X**QNNNSNFAWTGATKYHLNGR**X**SLVNPGVAMATHKDDE**X**RFFPSSGVLIFGK**X**GAG**X**NNT**X**L
XNVM**X**T**X**EEEIKTTNPVATE**X**YGVVA**X**NLQSSNTAP**X**TG**X**VNSQGALPGMVWQNRDVYLQGPIWAKIPH
TDGNFHPSPLMGGFGLKHPPPQILIKNTPVPANPP**XX**F**XX**AKFASFITQYSTGQVSVEIEWELQKENSKRWN
PEIQYTSNYAKS**X**NVDFAV**XXX**GVY**X**EPRPIGTRYLTRNL

b)

LATQSQSPTLNLSENHQQAPLVWDLVQWLQAVAHQWQTITKAPTEWVMPQEIGIAIPHGWATESSPPAPEP
GPCPPTTTTSTSKSPV**X**R**XXX**PTTTTTSATAPPGGILTSTDSTATSHHVTGSDSSTTTGDSGPRDSTSSSSTSRS
RRSRRMTASRPSLITLPARFRSFRTRNTSCRTSSALRTRAACLRSRRTSS

Figure B.3. Alignment of the ancestral AAV Cap protein with natural serotypes. Capsid amino acids were aligned using the Geneious program (Biomatters). Colored amino acids represent differences with the reference ancestral *cap* sequence. The variable positions in the ancestral library are annotated in black and designated with the letter X.

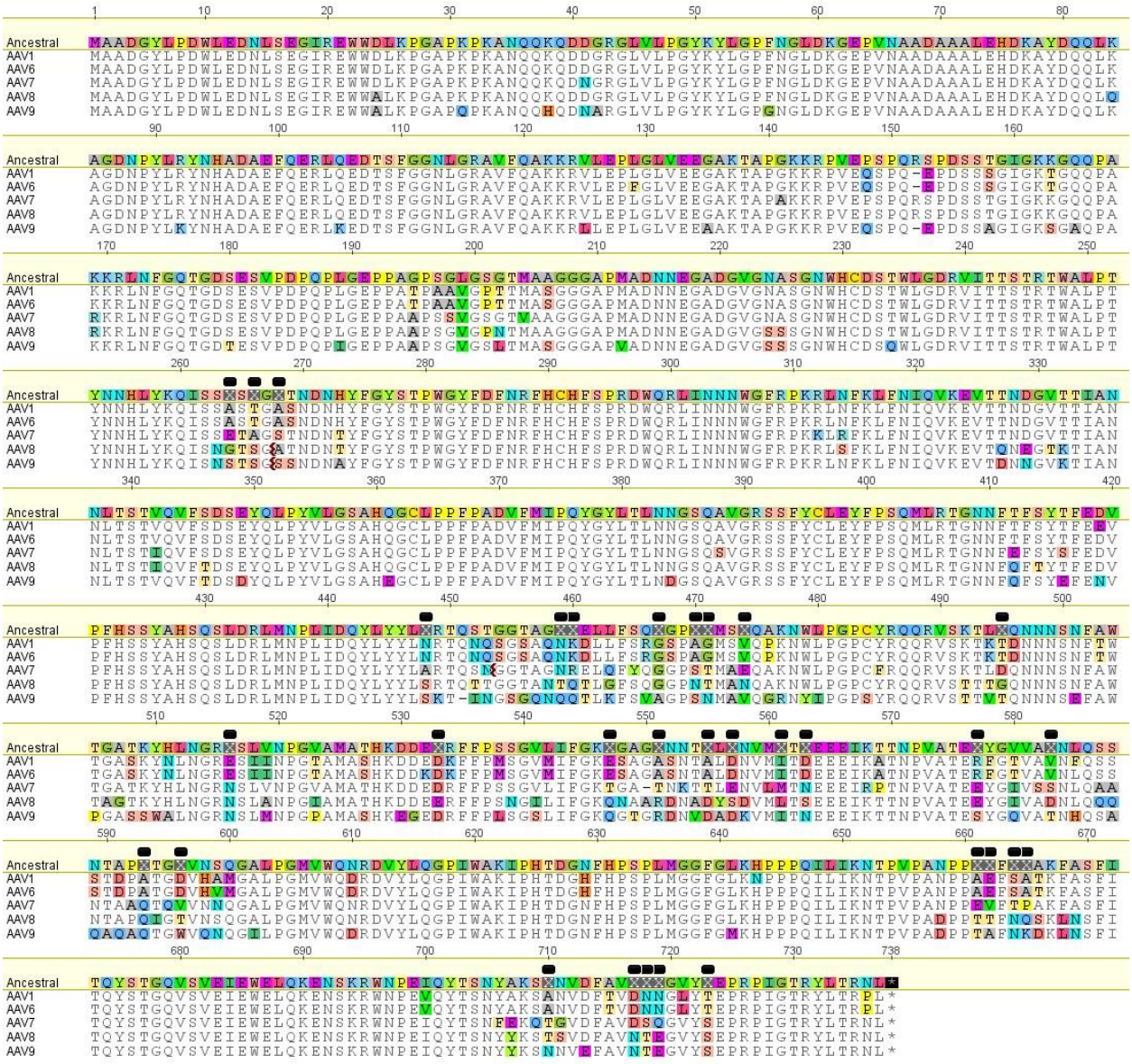
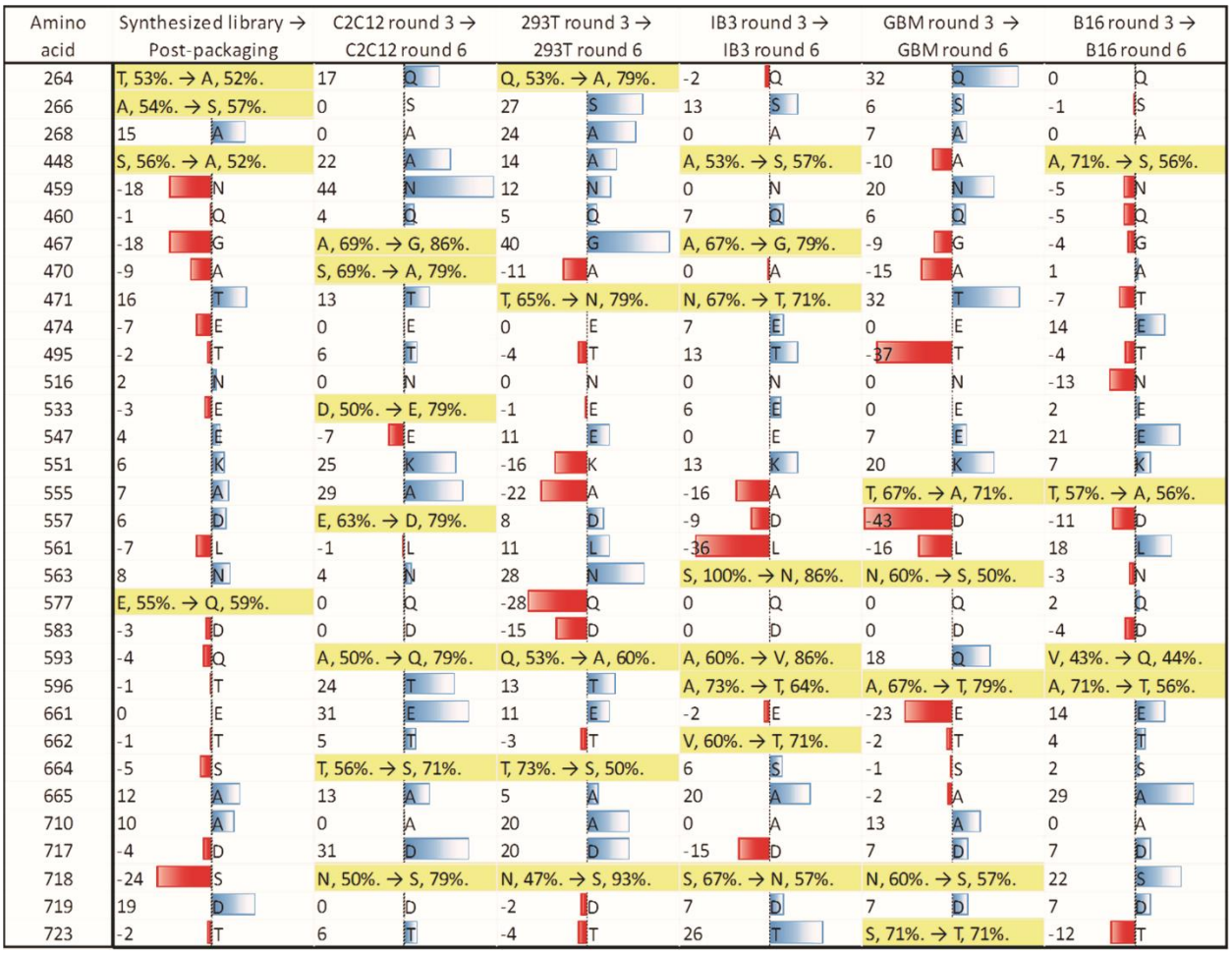


Figure B.4. Dominant amino acids at variable positions after three rounds of selection. A heat map was generated based on the frequency of the most common amino acid at each position in the different libraries. The dominant amino acid and frequency at each position were determined based on sequencing results from individual clones (n =14).

Amino acid	C2C12 round 3	293T round 3	IB3 round 3	GBM round 3	B16 round 3
264	Q, 69%.	Q, 53%.	A, 73%.	A, 47%.	Q, 50%.
266	S, 100%.	S, 59%.	S, 87%.	S, 80%.	A, 57%.
268	S, 100%.	S, 76%.	S, 100%.	S, 93%.	S, 100%.
448	S, 56%.	S, 50%.	A, 53%.	A, 67%.	A, 71%.
459	N, 56%.	T, 88%.	T, 100%.	T, 80%.	T, 93%.
460	R, 81%.	R, 88%.	R, 93%.	R, 87%.	R, 93%.
467	A, 69%.	A, 53%.	A, 67%.	A, 73%.	A, 79%.
470	S, 69%.	S, 88%.	S, 93%.	S, 93%.	S, 92%.
471	N, 88%.	T, 65%.	N, 67%.	N, 53%.	N, 57%.
474	A, 100%.	A, 100%.	A, 93%.	A, 100%.	A, 86%.
495	S, 94%.	S, 71%.	S, 87%.	S, 87%.	S, 86%.
516	D, 100%.	D, 100%.	D, 100%.	D, 100%.	D, 100%.
533	D, 50%.	D, 94%.	D, 80%.	D, 100%.	D, 86%.
547	Q, 100%.	Q, 82%.	Q, 100%.	Q, 93%.	Q, 79%.
551	A, 75%.	A, 82%.	A, 87%.	A, 80%.	A, 93%.
555	T, 50%.	A, 82%.	A, 73%.	T, 67%.	T, 57%.
557	E, 63%.	E, 59%.	E, 73%.	E, 100%.	E, 86%.
561	M, 94%.	M, 82%.	M, 100%.	M, 73%.	M, 57%.
563	S, 75%.	S, 59%.	S, 100%.	N, 60%.	S, 71%.
577	Q, 100%.	Q, 88%.	Q, 100%.	Q, 100%.	Q, 86%.
583	S, 100%.	S, 88%.	S, 100%.	S, 100%.	S, 86%.
593	A, 50%.	Q, 53%.	A, 60%.	A, 53%.	V, 43%.
596	T, 69%.	A, 67%.	A, 73%.	A, 67%.	A, 71%.
661	A, 69%.	A, 53%.	A, 67%.	A, 80%.	A, 86%.
662	V, 88%.	V, 60%.	V, 60%.	V, 67%.	V, 64%.
664	T, 56%.	T, 73%.	T, 87%.	T, 87%.	T, 86%.
665	P, 88%.	P, 73%.	P, 73%.	P, 73%.	P, 71%.
710	T, 100%.	T, 80%.	T, 100%.	T, 87%.	T, 100%.
717	N, 69%.	N, 80%.	N, 93%.	N, 71%.	N, 93%.
718	N, 50%.	N, 47%.	S, 67%.	N, 60%.	S, 71%.
719	E, 100%.	E, 67%.	E, 93%.	E, 93%.	E, 93%.
723	S, 94%.	T, 62%.	S, 60%.	S, 71%.	S, 93%.

Figure B.5. Change in amino acid frequency at variable positions between rounds three and six of selection. The percent change in amino acid frequency between the third and sixth round of selection on each cell line was calculated. If the identity of the dominant amino acid did not change, the increase (blue) or decrease (red) in frequency is displayed. If selection resulted in a change in amino acid identity at that position, the new amino acid and frequency is shown (yellow).



Amino acid	Synthesized library → Post-packaging	C2C12 round 3 → C2C12 round 6	293T round 3 → 293T round 6	IB3 round 3 → IB3 round 6	GBM round 3 → GBM round 6	B16 round 3 → B16 round 6
264	T, 53%. → A, 52%.	17 Q	Q, 53%. → A, 79%.	-2 Q	32 Q	0 Q
266	A, 54%. → S, 57%.	0 S	27 S	13 S	6 S	-1 S
268	15 A	0 A	24 A	0 A	7 A	0 A
448	S, 56%. → A, 52%.	22 A	14 A	A, 53%. → S, 57%.	-10 A	A, 71%. → S, 56%.
459	-18 N	44 N	12 N	0 N	20 N	-5 N
460	-1 Q	4 Q	5 Q	7 Q	6 Q	-5 Q
467	-18 G	A, 69%. → G, 86%.	40 G	A, 67%. → G, 79%.	-9 G	-4 G
470	-9 A	S, 69%. → A, 79%.	-11 A	0 A	-15 A	1 A
471	16 T	13 T	T, 65%. → N, 79%.	N, 67%. → T, 71%.	32 T	-7 T
474	-7 E	0 E	0 E	7 E	0 E	14 E
495	-2 T	6 T	-4 T	13 T	-37 T	-4 T
516	2 N	0 N	0 N	0 N	0 N	-13 N
533	-3 E	D, 50%. → E, 79%.	-1 E	6 E	0 E	2 E
547	4 E	-7 E	11 E	0 E	7 E	21 E
551	6 K	25 K	-16 K	13 K	20 K	7 K
555	7 A	29 A	-22 A	-16 A	T, 67%. → A, 71%.	T, 57%. → A, 56%.
557	6 D	E, 63%. → D, 79%.	8 D	-9 D	-43 D	-11 D
561	-7 L	-1 L	11 L	-36 L	-16 L	18 L
563	8 N	4 N	28 N	S, 100%. → N, 86%.	N, 60%. → S, 50%.	-3 N
577	E, 55%. → Q, 59%.	0 Q	-28 Q	0 Q	0 Q	2 Q
583	-3 D	0 D	-15 D	0 D	0 D	-4 D
593	-4 Q	A, 50%. → Q, 79%.	Q, 53%. → A, 60%.	A, 60%. → V, 86%.	18 Q	V, 43%. → Q, 44%.
596	-1 T	24 T	13 T	A, 73%. → T, 64%.	A, 67%. → T, 79%.	A, 71%. → T, 56%.
661	0 E	31 E	11 E	-2 E	-23 E	14 E
662	-1 T	5 T	-3 T	V, 60%. → T, 71%.	-2 T	4 T
664	-5 S	T, 56%. → S, 71%.	T, 73%. → S, 50%.	6 S	-1 S	2 S
665	12 A	13 A	5 A	20 A	-2 A	29 A
710	10 A	0 A	20 A	0 A	13 A	0 A
717	-4 D	31 D	20 D	-15 D	7 D	7 D
718	-24 S	N, 50%. → S, 79%.	N, 47%. → S, 93%.	S, 67%. → N, 57%.	N, 60%. → S, 57%.	22 S
719	19 D	0 D	-2 D	7 D	7 D	7 D
723	-2 T	6 T	-4 T	26 T	S, 71%. → T, 71%.	-12 T

Figure B.6. Glycan dependency of ancestral libraries and select ancestral variants. a) The transduction efficiency of ancestral libraries (after six rounds of selection) and select AAV variants C4, C7, and G4 carrying self-complimentary CMV-GFP was quantified by flow cytometry 72 hours after infection. For the libraries, infections were carried out at a genomic MOI of 2,000 (Pro5, Lec1, Lec2) and 50,000 (CHO-K1, pgsA). For select clones from the Round 6 C2C12 (C4, C7) and GBM (G4) libraries, infections were carried out at a genomic MOI of 500 (Pro5, Lec1, Lec2) and 13,000 (CHO-K1, pgsA) to ensure an adequate number of GFP positive cells for analysis. The CHO-K1/pgsA comparison examines heparan sulfate proteoglycan dependence, while Pro5/Lec1 and Pro5/Lec2 probe sialic acid dependence. Data are presented as mean ± SEM, $n = 3$. AL, ancestral library.

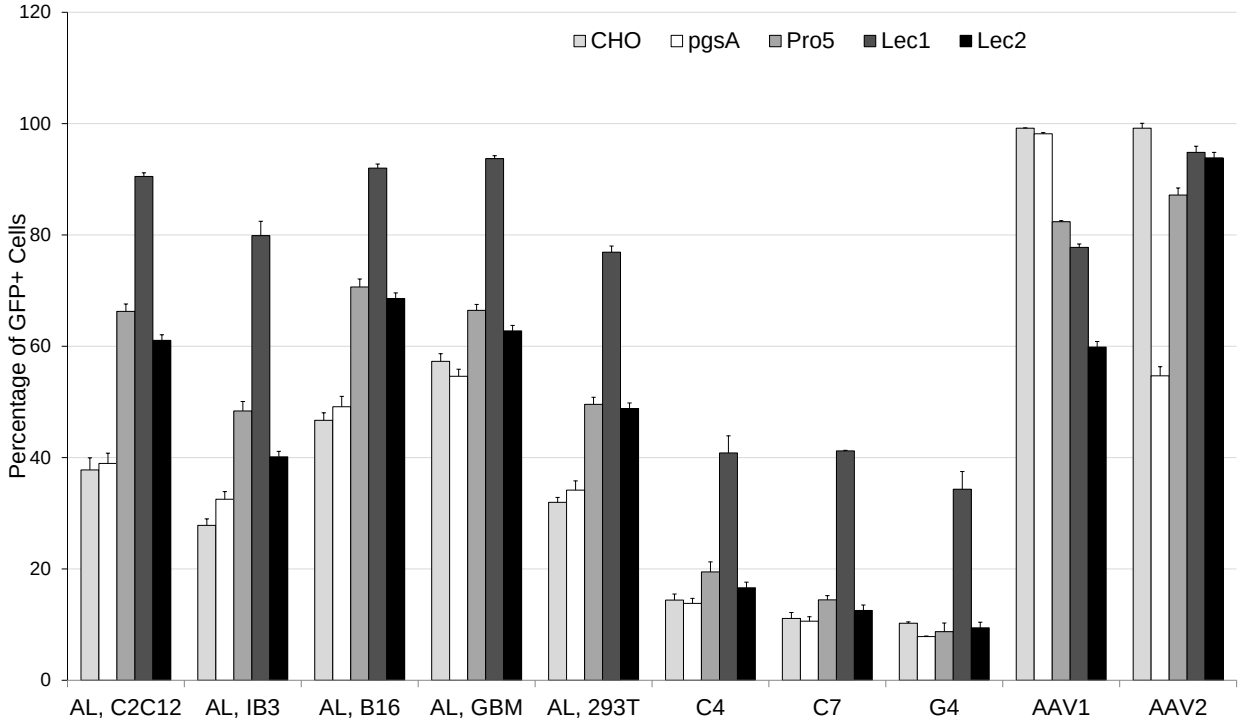


Figure B.7. Ancestral AAV libraries are neutralized by human intravenous immunoglobulin (IVIG) in vitro. Recombinant round 6 ancestral AAV libraries and AAV1 were packaged with a self-complimentary CMV-GFP cassette, incubated for one hour at 37°C with serial dilutions of heat-inactivated IVIG, then used to infect HEK293T cells at a genomic MOI of 2,000 (*n*=3). The fraction of GFP expressing cells was quantified by flow cytometry 72 hours later. Data are presented as mean ± SEM, *n* = 3. AL, ancestral library.

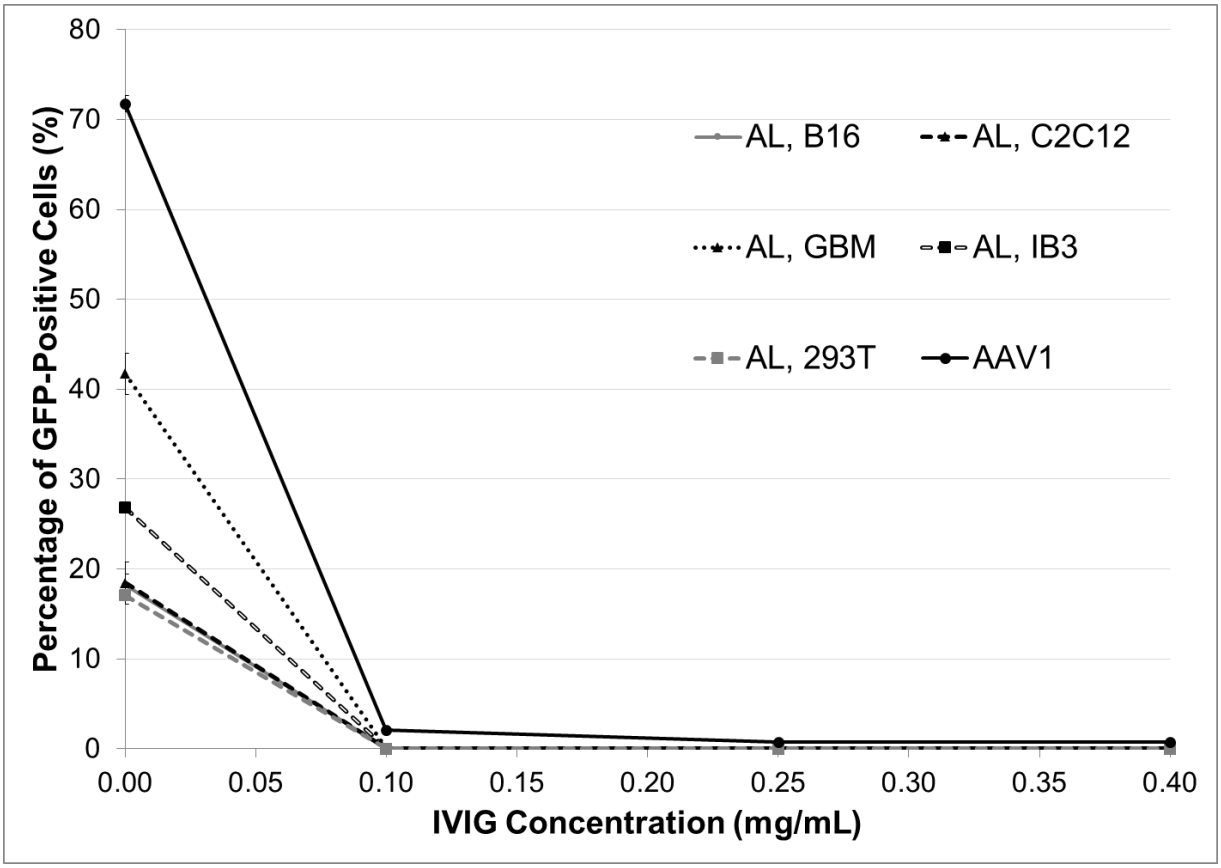


Table B.1. Selection stringency applied in ancestral AAV library selections.

Round of Selection	Genomic Multiplicity of Infection
1	5,000
2	500
3	250
4	250
5	50
6	25

Table B.2. Identities of the 32 variable amino acids present in the candidate ancestral clones evaluated in vivo.

Amino Acid	Ancestral AAV Clone		
	C4	**C7**	**G4**
264	A	Q	A
266	S	S	S
268	S	S	S
448	A	S	A
459	N	N	T
460	R	R	R
467	G	G	G
470	S	A	S
471	N	N	N
474	A	A	A
495	S	S	T
516	D	D	D
533	D	E	D
547	E	Q	Q
551	A	A	A
555	A	T	A
557	E	D	D
561	L	M	I
563	N	S	N
577	Q	Q	Q
583	S	S	S
593	A	Q	A
596	A	T	T
661	A	A	T
662	T	V	V
664	T	S	S
665	P	P	P
710	T	T	T
717	N	N	N
718	N	S	S
719	E	E	E
723	S	S	T

Appendix C: Application of CRISPR/Cas9 synergistic activation mediator and GeCKO v2 genome wide screens to identify endogenous gene targets that increase AAV packaging yields

The HEK293T cell lines described in this work were kindly generated by Prajit Limsirichai

C.1 Background

Higher rAAV production yields are needed to translate AAV gene therapies to larger patient populations. To address this need, modified HEK293T packaging cell lines that support increased virus yields could be generated. CRISPR/Cas9 technologies that facilitate genome-wide functional screens of all human coding isoforms have recently been developed. These include the human CRISPR activation pooled library (Synergistic activation mediator or SAM) and the human CRISPR knockout pooled library (GeCKO v2). The synergistic activation mediator library targets 23,430 genes with 70,290 unique guide sequences [1], while the GeCKO v2 library targets 19,052 genes with 122,417 unique guide sequences and 1,000 control non-targeting guide sequences [2]. Here, I present progress in utilizing SAM and GeCKO v2 genome wide screens to identify endogenous genes that can be activated or knocked out to increase AAV yield in HEK293T cells.

C.2 Development of tools for rAAV CRISPR/Cas9 library selections

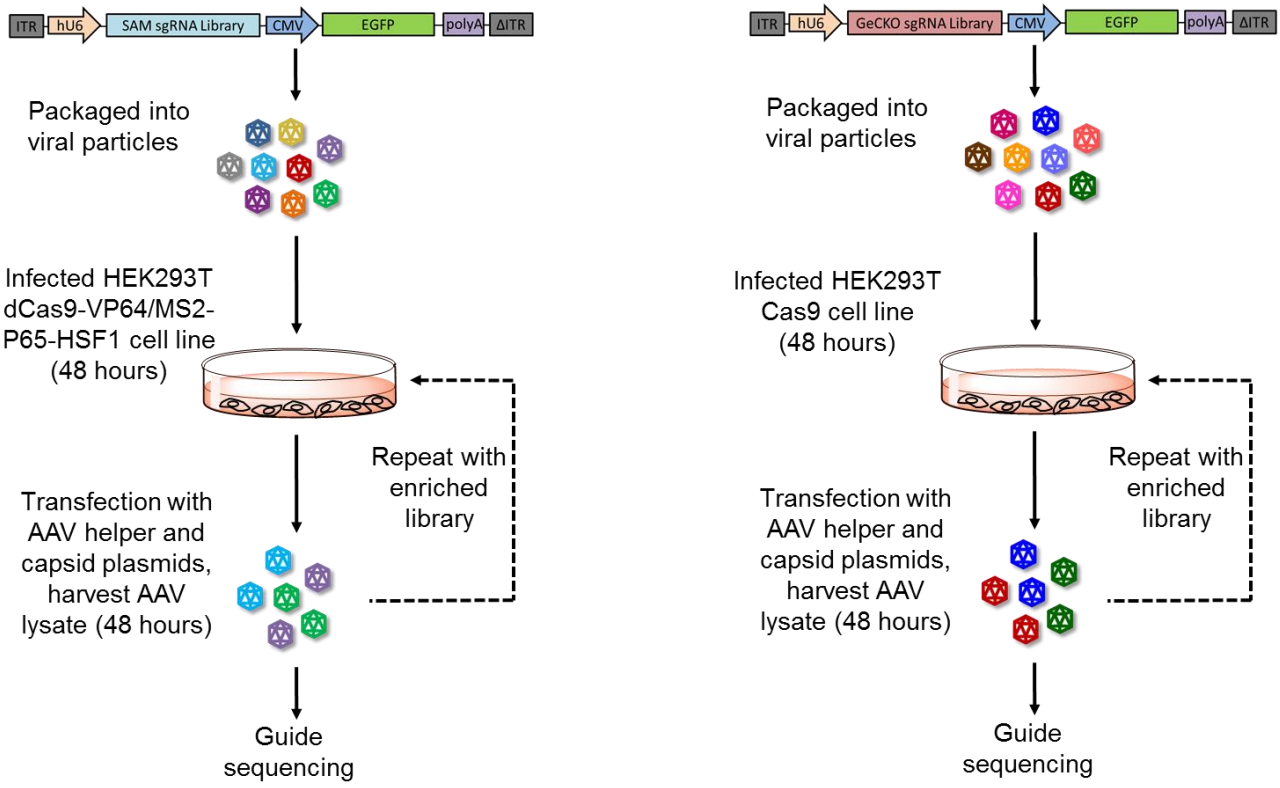
We first generated a HEK293T cell line that constitutively expresses a nucleolytically active Cas9 by infection with a lentivirus vector packaged with lentiCas9-Blast (Addgene Plasmid #52962) and selection using blastomycin. We next created a HEK293T cell line that constitutively expresses both a deactivated Cas9-VP64 fusion and the activation helper proteins MS2-P65-HSF1 by infection with lentivirus vectors packaged with lenti dCAS-VP64_Blast (Addgene Plasmid #61425) or lenti MS2-P65-HSF1_Hygro (Addgene Plasmid #61426) and selection using both hygromycin and blastomycin.

After creating the cell lines for selections, I next cloned the hU6 promoter and sgRNAs from the SAM plasmid library of lenti sgRNA(MS2)_zeo (Addgene Plasmid #61427) into a self-complementary rAAV backbone to generate the hU6-SAMsgRNA-CMV-GFP rAAV library. Electroporation of the library ligation into *E. coli* resulted in a library diversity (5×10^6 colonies) exceeding the size of the sgRNA library. The same procedure was repeated to clone the sgRNAs of the GeCKO v2 library into a self-complementary rAAV backbone to form the hU6-GeCKO v2 sgRNA-CMV-GFP rAAV library. These SAM and GeCKO v2 rAAV libraries express both the single-guide RNAs for gene targeting (from the hU6 promoter) and a GFP reporter (from CMV) used to monitor rAAV infection.

C.3 In vitro selections for single-guide RNA sequences that activate or knock out genes to increase AAV packaging yields

A schematic of the selection strategy employed is presented in Figure C.1.

Figure C.1. Selection strategy for SAM and GeCKO v2 genome wide screens to increase AAV packaging yield.



The selection strategy was implemented as follows:

1) The HEK293T cell lines expressing nucleolytically active Cas9 (for GeCKO v2 selections) or dCas9-VP64 and MS2-P65-HSF1 (for SAM selections) were cultured in blastomycin or blastomycin plus hygromycin respectively. Note that these cells have already been selected for lentivirus integration events, so it is no longer strictly necessary to supplement with the selection antibiotics. However, the antibiotics can periodically be re-introduced to ensure that the line has not been contaminated by wild type HEK293T, for example due to transcriptional silencing of the lentiviral vector expression constructs.
2) rAAV2 was packaged containing the sgRNA libraries using standard rAAV packaging techniques (triple transfection of pXX2-AAV2, pHelper, hU6-GeCKO v2 sgRNA-CMV-GFP or hU6-SAM sgRNA-CMV-GFP library) and titered by qPCR.
3) To initiate the first round of selection, cell lines were infected at a range of MOIs (25-2000) in 24-well plate format (500,000 cells). SAM and GeCKO v2 selections were conducted separately.

4) The infected cell lines were incubated for 48 hours to ensure onset of sgRNA expression and activation or knockout of the genes specified by the sgRNA. GFP reporter expression was observed to verify onset of gene expression and transduction efficiency.
5) An MOI of 500 was selected to process for round 2 based on its transduction efficiency that was high enough to ensure amplification of capsids but low enough to reduce the number of sgRNA per cell. Samples were transfected with pHelper and pXX2-AAV2 containing *rep* and *cap* to package rAAVs containing the sgRNA transgene cassette for the next round of selection. Crucially, sgRNAs that support higher packaging yield should produce a greater viral burst, leading to enrichment of those sgRNAs through rounds of selection.

Through four rounds of selection with the sgRNA GeCKO v2 library, and five rounds with the SAM library, I observed decreasing levels of GFP infection. This result implies that rAAV packaging from a relatively small number of cells (500,000), coupled with reliance on transfection to rescue episomal viral genomes, is not efficient. Although low efficiency effectively leads to a reduction in MOI in each round, thereby increasing the selection stringency as is typically desired, I was concerned that these outcomes would bottleneck the library selection and generate stochastic outcomes and/or false positives. To mitigate this risk, I extracted genomic DNA from the cells after rounds 4 and 5 respectively, PCR amplified the sgRNA library, and re-cloned it into the rAAV backbone. This library was then packaged using standard techniques as in round 1 to ensure a high virus titer for infection in rounds 5 and 6 respectively. Three additional rounds of selection were conducted, after which samples from the plasmid library, packaged AAV library (prior to selection), rounds 4 and 7 of the GeCKO v2 selection, and rounds 5 and 8 of the SAM selection were deep sequenced to identify the sgRNA sequences that were most enriched after selection.

The resulting deep sequencing reads were trimmed, and the sgRNA sequences were extracted. The frequency of each sgRNA at each round was calculated as well as the fold-change between the packaged library and the final round of selection. The gene identity and function of the most enriched sgRNAs from each library were annotated and are presented in Tables C.1 and C.2.

Table C.1. Gene identity, function, and fold change of the most enriched sgRNAs from the GeCKO v2 selection.

Gene Annotation	Fold Increase Over Packaged Library	Gene Function
Interferon-induced protein with tetratricopeptide repeats 5	1289	Interferon-induced RNA-binding protein that specifically binds single-stranded RNA bearing a 5'-triphosphate group (PPP-RNA), thereby acting as a sensor of viral single-stranded RNAs. Single-stranded PPP-RNAs, which lack 2'-O-methylation of the 5' cap and bear a 5'-triphosphate group instead, are specific from viruses, providing a molecular signature to distinguish between self and non-self mRNAs by the host during viral infection.
hsa-mir-4770	1090	Unknown function
Vacuolar protein sorting-associated protein VTA1 homolog	629	Involved in the endosomal multivesicular bodies (MVB) pathway. MVBs contain intraluminal vesicles (ILVs) that are generated by invagination and scission from the limiting membrane of the endosome and mostly are delivered to lysosomes enabling degradation of membrane proteins, such as stimulated growth factor receptors, lysosomal enzymes and lipids.
Dermatan-sulfate epimerase	847	This protein is involved in the pathway for heparan sulfate biosynthesis
Endoplasmic reticulum lectin 1	2222	Probable lectin that binds selectively to improperly folded lumenal proteins. May function in endoplasmic reticulum quality control and endoplasmic reticulum-associated degradation (ERAD) of both non-glycosylated proteins and glycoproteins.
hsa-mir-5581	535	Predicted targets: MED13. The mediator complex is recruited by transcriptional activators or nuclear receptors to induce gene expression, possibly by interacting with RNA polymerase II and promoting the formation of a transcriptional pre-initiation complex.
Kinesin-like protein KIF23	1350	Component of the centralspindlin complex that serves as a microtubule-dependent and Rho-mediated signaling required for the myosin contractile ring formation during the cell cycle cytokinesis. Essential for cytokinesis in Rho-mediated signaling. Required for the localization of ECT2 to the central spindle.
Ubiquitin carboxyl-terminal hydrolase 17-like protein 22	82	Deubiquitinating enzyme that removes conjugated ubiquitin from specific proteins to regulate different cellular processes that may include cell proliferation, progression through the cell cycle, apoptosis, cell migration, and the cellular response to viral infection
Caspase recruitment domain-containing protein 8	1259	Inhibits NF-kappa-B activation. May participate in a regulatory mechanism that coordinates cellular responses controlled by NF-kappa-B transcription factor.
C5orf52	1084	Unknown function
Cyclin-H	540	Involved in cell cycle control and in RNA transcription by RNA polymerase II. Its expression and activity are constant throughout the cell cycle
Zinc finger protein 860	1075	May be involved in transcriptional regulation
Uncharacterized protein C20orf202	152	Unknown function
CD63 antigen	916	Plays a role in intracellular vesicular transport processes
26S protease regulatory subunit 7	320	The 26S protease is involved in the ATP-dependent degradation of ubiquitinated proteins. The regulatory (or ATPase) complex confers ATP dependency and substrate specificity to the 26S complex.
Tensin-3	285	May play a role in actin remodeling. Involved in the dissociation of the integrin-tensin-actin complex. EGF activates TNS4 and down-regulates TNS3 which results in capping the tail of ITGB1. Seems to be involved in mammary cell migration.
Tripartite motif-containing protein 3	250	Probably involved in vesicular trafficking via its association with the CART complex. The CART complex is necessary for efficient transferrin receptor recycling but not for EGFR degradation.
Transmembrane protein 106B	526	Involved in dendrite morphogenesis and maintenance by regulating lysosomal trafficking via its interaction with MAP6. May act by inhibiting retrograde transport of lysosomes along dendrites.
RANBP2-like and GRIP domain-containing protein 5/6	497	Protein targeting to Golgi
Dynamin-2	407	Microtubule-associated force-producing protein involved in producing microtubule bundles and able to bind and hydrolyze GTP. Plays an important role in vesicular trafficking processes, in particular endocytosis.
Protein N-terminal asparagine amidohydrolase	198	Side-chain deamidation of N-terminal asparagine residues to aspartate. Required for the ubiquitin-dependent turnover of intracellular proteins that initiate with Met-Asn.

Table C.2. Gene identity, function, and fold change of the most enriched sgRNAs from the SAM selection.

Gene Annotation	Fold Increase Over Packaged Library	Gene Function
Interferon-induced protein with tetratricopeptide repeats 5	1289	Interferon-induced RNA-binding protein that specifically binds single-stranded RNA bearing a 5'-triphosphate group (PPP-RNA), thereby acting as a sensor of viral single-stranded RNAs. Single-stranded PPP-RNAs, which lack 2'-O-methylation of the 5' cap and bear a 5'-triphosphate group instead, are specific from viruses, providing a molecular signature to distinguish between self and non-self mRNAs by the host during viral infection.
hsa-mir-4770	1090	Unknown function
Vacuolar protein sorting-associated protein VTA1 homolog	629	Involved in the endosomal multivesicular bodies (MVB) pathway. MVBs contain intraluminal vesicles (ILVs) that are generated by invagination and scission from the limiting membrane of the endosome and mostly are delivered to lysosomes enabling degradation of membrane proteins, such as stimulated growth factor receptors, lysosomal enzymes and lipids.
Dermatan-sulfate epimerase	847	This protein is involved in the pathway for heparan sulfate biosynthesis
Endoplasmic reticulum lectin 1	2222	Probable lectin that binds selectively to improperly folded lumenal proteins. May function in endoplasmic reticulum quality control and endoplasmic reticulum-associated degradation (ERAD) of both non-glycosylated proteins and glycoproteins.
hsa-mir-5581	535	Predicted targets: MED13. The mediator complex is recruited by transcriptional activators or nuclear receptors to induce gene expression, possibly by interacting with RNA polymerase II and promoting the formation of a transcriptional pre-initiation complex.
Kinesin-like protein KIF23	1350	Component of the centralspindlin complex that serves as a microtubule-dependent and Rho-mediated signaling required for the myosin contractile ring formation during the cell cycle cytokinesis. Essential for cytokinesis in Rho-mediated signaling. Required for the localization of ECT2 to the central spindle.
Ubiquitin carboxyl-terminal hydrolase 17-like protein 22	82	Deubiquitinating enzyme that removes conjugated ubiquitin from specific proteins to regulate different cellular processes that may include cell proliferation, progression through the cell cycle, apoptosis, cell migration, and the cellular response to viral infection
Caspase recruitment domain-containing protein 8	1259	Inhibits NF-kappa-B activation. May participate in a regulatory mechanism that coordinates cellular responses controlled by NF-kappa-B transcription factor.
C5orf52	1084	Unknown function
Cyclin-H	540	Involved in cell cycle control and in RNA transcription by RNA polymerase II. Its expression and activity are constant throughout the cell cycle
Zinc finger protein 860	1075	May be involved in transcriptional regulation
Uncharacterized protein C20orf202	152	Unknown function
CD63 antigen	916	Plays a role in intracellular vesicular transport processes
26S protease regulatory subunit 7	320	The 26S protease is involved in the ATP-dependent degradation of ubiquitinated proteins. The regulatory (or ATPase) complex confers ATP dependency and substrate specificity to the 26S complex.
Tensin-3	285	May play a role in actin remodeling. Involved in the dissociation of the integrin-tensin-actin complex. EGF activates TNS4 and down-regulates TNS3 which results in capping the tail of ITGB1. Seems to be involved in mammary cell migration.
Tripartite motif-containing protein 3	250	Probably involved in vesicular trafficking via its association with the CART complex. The CART complex is necessary for efficient transferrin receptor recycling but not for EGFR degradation.
Transmembrane protein 106B	526	Involved in dendrite morphogenesis and maintenance by regulating lysosomal trafficking via its interaction with MAP6. May act by inhibiting retrograde transport of lysosomes along dendrites.
RANBP2-like and GRIP domain-containing protein 5/6	497	Protein targeting to Golgi
Dynamin-2	407	Microtubule-associated force-producing protein involved in producing microtubule bundles and able to bind and hydrolyze GTP. Plays an important role in vesicular trafficking processes, in particular endocytosis.
Protein N-terminal asparagine amidohydrolase	198	Side-chain deamidation of N-terminal asparagine residues to aspartate. Required for the ubiquitin-dependent turnover of intracellular proteins that initiate with Met-Asn.

C.4 Quantification of improved AAV yields after gene activation or knockdown using selected sgRNA sequences

After analysis of gene function and fold-enrichment, 24 of the most intriguing sgRNAs from the GeCKO v2 library were cloned into pSPgRNA (Addgene Plasmid #47108) and 24 sgRNAs from the SAM library were cloned into lenti sgRNA(MS2)_zeo (Addgene Plasmid #61427). These plasmids drive expression of the sgRNA from a hU6 promoter but do not contain AAV ITRs.

To evaluate the impact of each sgRNA on AAV packaging, the following experiment was conducted.

1) The HEK293T cell lines expressing nucleolytically active Cas9 (for GeCKO v2 selections) or dCas9-VP64 and MS2-P65-HSF1 (for SAM selections) were seeded one day prior to transfection at 125,000 cells per well in a 24-well plate.
2) 24 hours after cell seeding, 0.03 µg of pcDNA 3.1 CMV GFP transfection control and 0.48 µg of the sgRNA expression plasmid were transfected using PEI.
3) 48 hours later, the media was changed and a cocktail of 0.01 µg pcDNA 3.1 CMV mCherry transfection control, 0.15 µg pAAV-CAG-Luciferase, 0.16 µg pXX2 AAV2, and 0.19 µg pHelper were transfected using PEI.
4) 48 hours after the second transfection, cells were harvested by trypsinization and an aliquot of each well was taken for analysis by flow cytometry to determine transfection efficiency using the reporter transfection controls. The remainder of the samples were harvested per the usual AAV harvest protocol.
5) 10 µL of the rAAV-CAG-Luciferase virus lysate was used to infect 15,000 HEK293T cells in 96-well plate format in triplicate and the resulting luciferase expression was assayed using a high-throughput plate reader. Here, the luciferase expression from rAAV2 was used as a proxy for packaging efficiency given that qPCR is not sensitive enough to detect low virus yield from 24-well plate format, or high enough in throughput to evaluate all of the sgRNAs tested.

Variability in the transfection efficiency and luciferase assay in this experiment format ultimately led to high levels of noise across experimental replicates that obfuscated any significant differences between sgRNAs. Next steps would be to take a few of the most promising sgRNAs from the library, scale up packaging in 15 cm plate format, and evaluate improvement in packaging yield by qPCR relative to wild type HEK293T cells and the control sgRNA targeted to the Rosa locus.

Additional future directions include combination of validated hits from both selections for potential cooperative increases in yield, confirmation that results are generalizable to all AAV serotypes, and studies of the mechanism of increased packaging efficiency.

1. Konermann, S., et al., *Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex.* Nature, 2015. **517**(7536): p. 583-8.
2. Sanjana, N.E., O. Shalem, and F. Zhang, *Improved vectors and genome-wide libraries for CRISPR screening.* Nat Methods, 2014. **11**(8): p. 783-4.

Appendix D: Attachment of proteins to the AAV capsid using the SpyTag-SpyCatcher covalent interaction

This work was conducted in collaboration with Leah Byrne and Tim Day of the Flannery Lab

In this appendix I outline the progress we have made in developing an AAV SpyTag-SpyCatcher platform for attaching proteins to the capsid surface. SpyTag and SpyCatcher form a covalent isopeptide bond in a wide range of reaction conditions [1]. The same reaction can be achieved with SpyTag and KTag in the presence of SpyLigase [2]. I evaluated insertion of SpyTag or KTag into positions 453 or 588 on the AAV2 capsid. Importantly, insertion into VP3 ensures that each capsid will contain the peptide since VP3 is necessary for capsid formation. Alternative strategies to incorporate the peptide into the N-terminus of VP2 do not ensure incorporation since VP2 is not necessary for capsid formation.

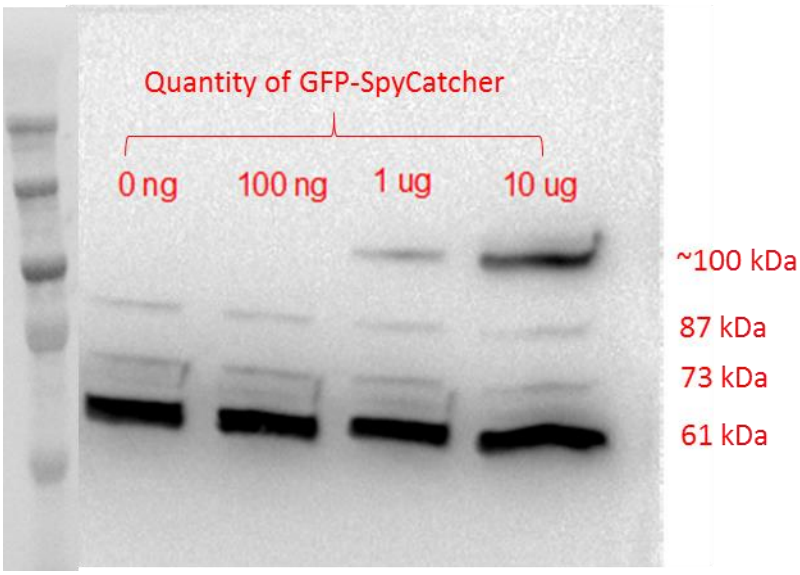
Table D.1. rAAV packaging yields with various SpyTag or KTag insertion sites.

Recombinant AAV packaged with sc-CAG-mCherry-bgH cassette	Number of 15 cm plates transfected	Genomic titer vg/mL	Total Viral Genomes
rAAV2 SpyTag inserted at AA 453 with no linker sequences	4	3.55E+10	5.33E+10
rAAV2 LA SpyTag A inserted at AA 453	4	1.08E+09	1.62E+09
rAAV2 TG SpyTag GLS inserted at AA 453	4	6.72E+10	1.01E+11
rAAV2 TG KTag GLS inserted at AA 453	4	2.22E+10	3.33E+10
rAAV2 TG SpyTag GLS inserted at AA 588	4	6.19E+10	9.28E+10
rAAV2 TG KTag GLS inserted at AA 588	4	8.01E+09	1.20E+10
Wild type rAAV2	4	8.54E+10	1.28E+11
Replication competent libraries packaged at low cap MOI			
rAAV2 SpyTag library inserted at AA 453 with a library of linker sequences	3	4.40E+10	6.60E+10
rAAV2 KTag library inserted at AA 453 with a library of linker sequences	3	1.71E+11	2.57E+11
rAAV2 SpyTag library inserted at AA 588 with a library of linker sequences	3	6.35E+10	9.53E+10
rAAV2 KTag library inserted at AA 588 with a library of linker sequences	3	1.84E+11	2.76E+11

rAAV packaging yields varied for each peptide insertion site but were generally within 10-fold of wild type rAAV2. All rAAVs packaged with sc-CAG-mCherry were used to infect 293T cells at a genomic MOI of 2,000. Each capsid variant successfully transduced cells resulting in mCherry expression despite insertion of the SpyTag or KTag peptide into the capsid. However, co-transfection of a pBluescript plasmid driving expression of GFP-SpyTag, GFP-SpyCatcher, GFP-KTag, or SpyLigase as appropriate using a CAG promoter did not support attachment of GFP to the capsid visible by microscopy after AAV infection or by western blot.

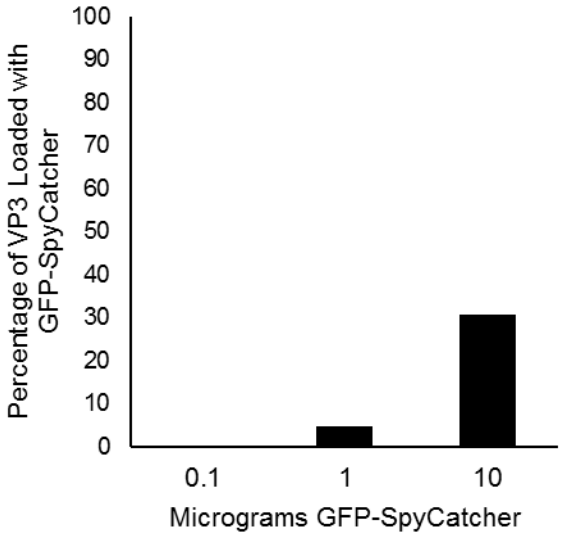
I next tested separate purification of rAAV2-588-TG_SpyTag_GLS and superfolder 6-His-GFP-SpyCatcher followed by *in vitro* mixing in an eppendorf tube. I generated a pET14b-6His-sfGFP-SpyCatcher construct and purified superfolder 6His-GFP-SpyCatcher from Rosetta2 BL21 *E.coli*. 6His-GFP-SpyCatcher was dialyzed into PBS with 10% glycerol. rAAV2-588-TG_SpyTag_GLS sc-CAG-mCherry was packaged at 5.21×10^{12} viral genomes/mL in 250 microliters PBS-0.001% Tween 20 from ten 150 mm tissue culture plates. The rAAV sample was diluted to 2.5×10^{9} viral genomes per microliter in PBS/0.001% Tween 20. 20 uL of diluted rAAV was mixed with 1 uL of diluted 6His-GFP-SpyCatcher (100 nanograms, 1 microgram, or 10 micrograms) or PBS control in a low protein binding tube. The mixture was incubated at room temperature for 25 minutes and then overnight at 4 °C. The entire 21 uL sample volume was loaded for western blot analysis with anti-AAV VP1/VP2/VP3 clone B2 antibody.

Figure D.1. Western blot analysis of VP3-SpyTag isopeptide bond formation with GFP-SpyCatcher. The protein ladder is the Precision Plus Protein WesternC Standard.



An increase in VP3 protein size was observed, indicating isopeptide bond formation with the 40 kDa 6His-GFP-SpyCatcher. Incubation with increasing amounts of 6His-GFP-SpyCatcher yielded a higher fraction of GFP tagged VP3 proteins (Figure D2).

Figure D.2. Quantification of western blot results.



Incubation time and/or the ratio of virus to GFP-SpyCatcher could be tuned to increase or decrease the fraction of VP3 tagged with GFP. Next steps would also include infectious assays to determine the impact of GFP addition on virus infectivity. Other AAV variants that incorporate SpyTag at

amino acid 453 should also be tested to determine if GFP-SpyCatcher can be loaded onto sites that do not block virus affinity for heparan sulfate proteoglycans.

Additionally, anticipating that rationally designed linker sequences flanking the SpyTag or KTag peptide insertions may not optimally present the peptide for isopeptide bond formation, I generated a library of linker sequences flanking the SpyTag or KTag insertion in AAV2 at amino acid 453 or 588. These four libraries contain either 2 and 1, 2 and 2, or 2 and 3 amino acids (NNK) 5' and 3' of the peptide insertion. This library was packaged as replication-competent AAV in HEK293T with transfection of low numbers of *cap* genes per cell to limit cross-packaging and/or mosaic particles. The replication-competent virus titers of these libraries are presented in Table D.l. PCR amplification of successfully packaged capsid genomes from this pre-packaged library may enable identification of linkers that support high titer packaging of AAV SpyTag capsids. Alternatively, the 6-His tag on the GFP SpyCatcher could be used to pull-down capsids from the library that optimally present SpyTag to promote isopeptide bond formation with GFP-SpyCatcher.

1. Reddington, S.C. and M. Howarth, *Secrets of a covalent interaction for biomaterials and biotechnology: SpyTag and SpyCatcher.* Curr Opin Chem Biol, 2015. **29**: p. 94-9.
2. Fierer, J.O., G. Veggiani, and M. Howarth, *SpyLigase peptide-peptide ligation polymerizes affibodies to enhance magnetic cancer cell capture.* Proc Natl Acad Sci U S A, 2014. **111**(13): p. E1176-81.