

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Directed Evolution for the Discovery of Novel AAV Variants and Improving Library Characterization

Permalink

<https://escholarship.org/uc/item/88c0c2md>

Author

Carneiro, Ana

Publication Date

2024

Peer reviewed|Thesis/dissertation

Directed Evolution for the Discovery of Novel AAV Variants and Improving Library
Characterization

By

Ana Denise Carneiro

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 Karthik Shekhar

Professor Britt A. Glaunsinger

Summer 2024

Directed Evolution for the Discovery of Novel AAV Variants and Improving Library
Characterization

Copyright 2024
by
Ana Denise Carneiro

Abstract

Directed Evolution for the Discovery of Novel AAV Variants and Improving Library Characterization

by

Ana Denise Carneiro

Doctor of Philosophy in Chemical Engineering

University of California, Berkeley

Professor David V. Schaffer, Chair

Adeno-associated viruses (AAVs) are a powerful tool in gene therapy, offering a safe and highly effective method for delivering therapeutic genes to treat human disease. AAV-based gene delivery vehicles, or vectors, have shown success in both preclinical studies and clinical trials, with six therapies having received U.S. FDA approvals. Several properties of AAVs make them an advantageous vector for delivering DNA for clinical gene therapy: their minimal lack of integration into the host genome, lack of pathogenicity, and ability to promote high expression of the delivered therapeutic. However, issues still arise over the efficiency and specificity of AAV gene delivery, as well as possible immune responses. Naturally-occurring AAVs have the ability to transduce a wide range of tissue and cell types, reducing delivery efficacy to the intended tissue, requiring higher dosages to be administered, and increasing the potential for off-target effects. In addition, the presence of neutralizing antibodies against AAVs significantly limits the effectiveness of any AAV-based treatment or re-administration of the therapeutic. To overcome this, engineering the AAV capsid, or virus shell, is necessary to improve both its delivery efficiency to target cell types and prevent recognition of our body's neutralizing antibodies.

Several strategies have been implemented to engineer the AAV capsid and thus alter its properties and function. Directed evolution is a technique that was designed to mimic natural evolution by subjecting a large library of AAV variants to iterative rounds of selection against a selective pressure to isolate variants with the desired properties. Directed evolution strategies have been applied to the AAV *cap* gene, which encodes the proteins that comprise the capsid, to generate millions of novel AAV capsids and identify variants with enhanced transduction towards specific cell types, evade neutralizing antibodies, and/or improve packaging capabilities. In Chapter 2, we use the process of directed evolution to identify novel AAV variants that successfully transduce wild-type Schwann cells in human or mouse. The ability to deliver therapeutics to Schwann cells provides an opportunity to

preventatively and actively treat genetic diseases of the peripheral nervous system, including Neurofibromatosis Type I (NF1) and Charcot-Marie Tooth disease. Through this work, we have subjected millions of AAV variants to iterative rounds of selection in human and mouse Schwann cells to identify variants with improved transduction towards human or mouse Schwann cells. We have identified two groups of capsids: human Schwann cell-tropic capsids with increased resistance to antibody neutralization, and mouse Schwann cell-tropic capsids that also demonstrate decreased localization to the liver compared to wild-type AAVs.

While directed evolution strategies offer a high-throughput method of screening millions of AAV variants for enhanced properties, there is a significant lack of knowledge about the composition of many AAV DNA libraries due to sequencing limitations. The ability to fully sequence these DNA libraries offers insight into the evolutionary process and allows for earlier detection of successful variants. However, complete characterization is hindered by current sequencing methodologies. Short-read sequencing, or next generation sequencing (NGS), is limited to 300-500bp reads, unable to sequence individual variants whose mutations span the entire 2.2kb cap open reading frame. Long-read sequencing methodologies suffer from high error rates and limits on the number of variants that can be sequenced. In Chapter 3, we develop a novel synthetic, long-read sequencing methodology, called Barcode-Labeled Short Reads (BLaSR), to overcome the major barriers that current technologies encounter. We harness the random mechanism of the Tn5 transposon (also known as a ‘jumping gene’) to insert random barcodes into a diverse AAV plasmid library such that each AAV variant contains insertions with a unique, identifiable barcode. This allows for the generation of barcode-linked short reads that can be sequenced with NGS technologies, so that sequencing reads with the same barcode can be traced to the same AAV variant and thus assembled into the full AAV variant *cap* sequence. We have engineered a robust pipeline to ensure highly efficient barcoded-Tn5 insertion into the AAV genome and have characterized BLaSR according to error rate, read depth, and Tn5 insertion bias. We also applied BLaSR to a diverse, randomly shuffled AAV capsid library and estimate the full assembly of thousands of variants. Further work is ongoing to completely assemble and characterize the sequenced shuffled library, as well as sequence the shuffled library after a round of selection in adult human brain tissue to identify enriched variants.

To my family and loved ones

Contents

Contents	ii
List of Figures	iv
List of Tables	v
1 Engineering Novel Adeno-Associated Viruses (AAVs) for Improved Delivery in the Nervous System	1
1.1 Chapter Abstract	1
1.2 Introduction	1
1.3 Barriers to Efficient Transduction in the Nervous System	4
1.4 Engineering the AAV Capsid	6
1.5 Future Perspectives	9
1.6 Conclusion	10
2 Directed Evolution of Novel AAV Capsids for Enhanced Delivery to Mouse and Human Schwann Cells	12
2.1 Chapter Abstract	12
2.2 Introduction	12
2.3 Results	14
2.4 Discussion	21
2.5 Materials and Methods	24
2.6 Acknowledgements	28
3 Long-read sequencing of Diverse DNA libraries through Barcode-Labeled Short Reads (BLaSR)	29
3.1 Chapter Abstract	29
3.2 Introduction	30
3.3 Results	31
3.4 Discussion and Future Steps	36
3.5 Materials and Methods	39
3.6 Acknowledgements	42

Bibliography	43
Appendices	56
A Supplementary Information for Chapter 2	57
B Supplementary Information for Chapter 3	63

List of Figures

1.1	Recombinant AAV design and production	2
1.2	AAVs in clinical trials	3
1.3	Engineering of AAV capsids	7
2.1	Overview of directed evolution strategy	15
2.2	Evolution of hSC-tropic capsids	16
2.3	Characterization of hSC-tropic capsids	18
2.4	Evolution of mSC-tropic capsids	20
2.5	Characterization of mSC-tropic capsids	21
3.1	Overview of BLaSR sequencing methodology	32
3.2	Characterization of BLaSR	34
3.3	Sequencing a shuffled AAV library with BLaSR	37
A.1	Natural serotype transduction in hSCHWs	59
A.2	Next-generation sequencing (NGS) analysis of variant enrichment in mouse Schwann cells	60
A.3	Mapping of single amino acid point mutations in C1 capsid	61
A.4	7mer variant validation in hSCHWs	62
B.1	Development of BLaSR experimental pipeline	64
B.2	Restriction enzyme digest to assess insertion bias	65
B.3	Assessment of Tn5 insertion bias	65
B.4	Application of BLaSR in directed evolution	66

List of Tables

A.1	Primers used in study	58
B.1	Primers used in study	63

Acknowledgments

This PhD would not have been possible without the support of so many people in my life. While this thesis shows the work I've completed these past six years, I could not have done it without the academic, emotional, and mental support I've received. First and foremost, I'd like to thank my PhD advisor Professor David Schaffer for his academic support and mentorship. With his expansive knowledge about the gene therapy (and in general, scientific) landscape, Dave has always provided insightful and cutting-edge ideas and feedback, ensuring our work can have a novel and positive impact on human health. He has fostered a kind, collaborative, and fun lab environment that has helped establish a supportive lab culture. Thank you to Wanichaya Noiwangklang and Ruby Nelson for being phenomenal lab managers who have held our lab together with last-minute orders and lab maintenance, scavenged for whatever reagent or equipment we need for our experiments, and kept the lab running. I would also like to thank Professors Britt Glaunsinger and Karthik Shekhar for serving on my thesis committee and providing valuable feedback during my qualification exam and thesis review.

My colleagues in the Schaffer Lab have been instrumental to my time at UC Berkeley, and I am lucky to have been able to work with such intelligent, collaborative, and fun people. I appreciate all the guidance and mentorship I received from many of the older graduate students when I joined the lab, including Chris Barnes, Kaz Lewis, Andy Lopez, Rocio Sampayo, Eric Qiao, Sabrina Sun, and Christina Fuentes. A huge thank you to my mentor, Bonnie Zhu (now Professor Zhu!), who joined the lab at the same time as me and worked alongside me as we both started our own gene therapy projects. We learned about AAV library generation, packaging, and selection together, and as we later undertook other projects, she continued to provide valuable scientific advice and became my go-to person for any questions. I value her mentorship and friendship, and I am so excited to see the work that comes out of her lab. I was also fortunate to join the lab the same year as Hunter Johnson and Adam Schieferecke, two of the most hard-working and intelligent scientists I know. I am thankful for our years together in Schaffer lab and our lab shenanigans, and I always look forward to our annual photo together.

Another thank you to Sarah Yang and Hyuncheol Lee for their help with my mice studies. Animal work was not one of my favorite things to do as part of my work, but Sarah and Hyuncheol made it easier. Sarah was extremely patient when training me in animal procedures and helped me to become more comfortable around the mice. Hyuncheol was a star and helped me with most of my animal injections, demonstrating his impressive veterinarian knowledge, steady hand, persistence, and ability to tackle the smallest of veins. I would also like to thank Li Lin for his seemingly endless knowledge about everything. From our time together in the Yuan Lab at Purdue to the Schaffer Lab at Berkeley, Li was always willing to answer any question and brainstorm solutions. And finally, thank you to everyone else in the Schaffer lab. I have truly enjoyed working with everyone. I have been so fortunate to have such kind, intelligent, creative, fun-loving, and supportive colleagues – you all made the lab a great place to work.

Outside of lab, there were many people at UC Berkeley who shaped my time here. First, thank you to my housemates, past and present: Natalie Lefton, Samantha Crowe, Emily Carvalho, Monica Gurung, and Kalena Yearwood. From roommate dinners, nights on the deck, kitchen discussions, and a Covid lockdown, they made Rose St house a home. Thank you to Alison Lui for being someone I could always talk to – our weekly coffee chats encouraged me whenever I got stuck in a rut. I'd also like to thank Zach K., Branden, Johnny, Alec, Zach H., Neel, Eric, Clay, Nick, and the rest of the Class of 2018 for their support and friendship. Our annual cohort trips were always something I looked forward to, and I couldn't have asked for a better group to go through grad school with.

Thank you to Natalie Goh and Julie Fornaciari for their emotional and mental support – their wisdom, advice, and friendship were invaluable to my PhD journey. In addition, thank you to all my friends in Berkeley for their support, including Lorena Grundy, Rachael Morton, Bridgette LaFaye, Darby Hickson, Morgan Seidler, Erika Ding, Emily Abdo, Anna Weldy, Elizabeth Voke, and many others.

I also wanted to acknowledge the people of Bay Area Scientists Inspiring Students (BASIS), Graduate Women in Engineering (GWE) at UC Berkeley, and ComSciCon. Teaching science lessons to third grade students, organizing events for fellow graduate students, and getting to dive into science communication were some of my favorite parts of my PhD and a nice break from lab work.

I would also like to thank my friends many miles away: Priyanka Santhapuram, Mahera Husain, Megan Skiba, Elyse Kizior, Maya Jain, Rutva Shah, Hamna Qureshi, Nithya Sridhar, Claire Benedict, Marie Leeson, and Edwin Jacobsen – thank you for the many FaceTimes, visits, and encouragement throughout the years.

Thank you Puneeth for your constant support, alleviating my stress, and cooking me many dinners so I wouldn't have to.

Finally, and most importantly, I would not be here without the support of my family. My extended family has consistently shown their love and support for me from across the world. Thank you Nikhil for your humor and your rundowns on the latest sports or entertainment news. And lastly, thank you to my parents for their emotional, mental, and financial support throughout my entire life. Always ready to pick up my phone call at any hour, talk me through my anxious thoughts, cheer on my successes, and give me encouraging pep talks, my parents have been my rock.

Chapter 1

Engineering Novel Adeno-Associated Viruses (AAVs) for Improved Delivery in the Nervous System

1.1 Chapter Abstract

Harnessing adeno-associated virus (AAV) vectors for therapeutic gene delivery has emerged as a progressively promising strategy to treat disorders of both the central nervous system (CNS) and peripheral nervous system (PNS), with many ongoing clinical trials. However, unique physiological and molecular characteristics of the CNS and PNS pose obstacles to efficient vector delivery, ranging from the blood brain barrier to the diverse nature of nervous system disorders. It has become necessary to engineer novel AAV capsids to overcome these ongoing challenges and maximize transgene delivery. This article discusses strategies for innovative AAV capsid development, highlighting recent advances. Notably, advances in next generation sequencing and machine learning have sparked new approaches for capsid engineering and understanding. Furthermore, we outline future directions and additional challenges in AAV-mediated gene therapy in the CNS and PNS.

1.2 Introduction

Gene therapy, the delivery of genetic material into target cells to replace defective genes or otherwise ameliorate a disease condition, can be a powerful and long-lasting approach to treating or preventing disease compared to traditional treatments such as therapeutic proteins that typically require long-term, repeated dosing. Over the past decade, adeno-associated virus (AAV) has emerged as an effective and predominantly safe viral vehicle for delivering DNA in clinical gene therapy. Its ability to mediate effective transgene expression, lack of pathogenicity, and minimal integration into the host genome make AAV an attractive option for gene manipulation both *in vivo* and *in vitro*. In addition, six recent approvals of

AAV-based gene therapies by the US Food and Drug Administration (FDA) have highlighted its broad potential (U.S. FDA).

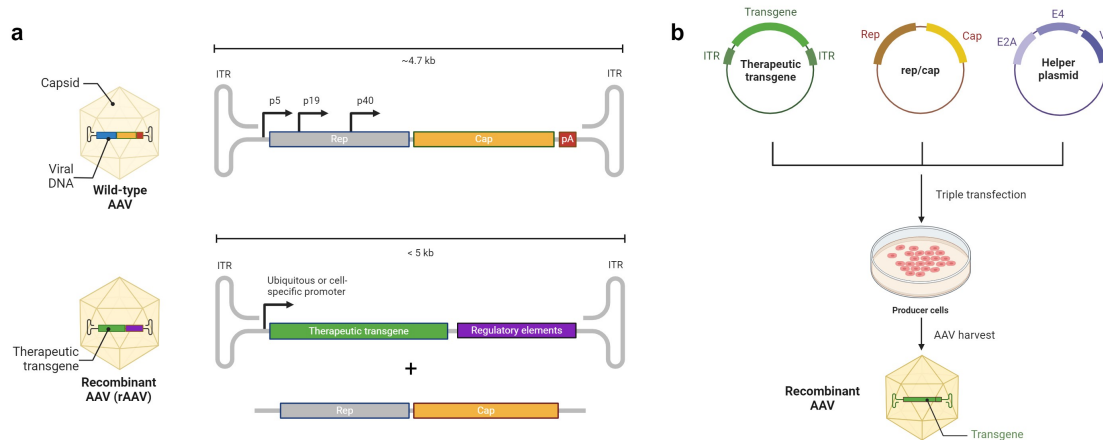


Figure 1.1: Recombinant AAV design and production. (a) Wild-type AAVs are converted to gene delivery vehicles by removing the *rep* and *cap* genes and instead inserting a therapeutic gene of interest, along with its necessary regulatory elements. (b) Recombinant AAVs are manufactured by triple transfection of a producer cell line. Viral particles are then harvested from the producer cells and purified to generate AAV particles containing the gene of interest. Created with BioRender.com.

Originally discovered as a contaminant of adenovirus stocks, AAV is a non-enveloped member of the parvovirus family with a 4.8-kb single-stranded DNA genome[1]. It is dependent upon co-infection with a helper virus, such as adenovirus or herpesvirus, to replicate. The AAV genome is comprised of two open reading frames (ORFs): *rep* and *cap*, flanked by two inverted terminal repeats (ITRs). The *rep* ORF encodes four proteins – Rep78, Rep68, Rep52, and Rep40 – which are involved in replication of the viral genome and assembly of viral particles. The *cap* ORF encodes the three capsid subunits – VP1, VP2, and VP3 – that form the outer shell that encapsulates the genome [1]. In addition, assembly-activated protein (AAP), membrane-associated accessory protein (MAAP), and X are proteins generated from alternative translational start sites within *cap* and assist with viral assembly[2, 3].

Wild-type AAVs have been converted into therapeutic delivery vehicles by replacing the viral genome with a therapeutic gene of interest flanked by the two ITRs to form a recombinant AAV (rAAV) genome, which must be less than 5 kilobases in length due to capsid size constraints (**Figure 1.1a**). When packaging rAAV particles, the *rep* and *cap* genes, along with necessary helper virus genes, are supplied *in trans*[2] to assemble complete virions (**Figure 1.1b**).

The nervous system is an important target for clinical gene therapy, in part due to severe unmet medical need. Within the central nervous system (CNS), neurons, microglia, and oligodendrocytes are important gene delivery targets that play major functional roles including learning, memory, regulation of movement, blood-brain barrier support, neuronal main-

tenance, and immune responses. Analogously in the PNS, Schwann cells, satellite cells, and enteric glia are the primary glial cells that mediate periphery nerve health and function[4].

Given the importance of the central and peripheral nervous systems to organismal function and the broad range of disorders involving nervous system dysfunction, they are important gene therapy targets. There have been over 60 AAV-based therapies in clinical trials for the nervous system, with major disease types highlighted in (Figure 1.2). Importantly, the first two FDA-approved AAV gene therapies target the central nervous system (Luxturna for the retina and Zolgensma for the spinal cord)[5].

DISEASE	TRANSGENE	CAPSID	ROUTE OF ADMINISTRATION	SPONSOR	PHASES	CLINICAL TRIAL NUMBER
Alzheimer's Disease	Brain-derived neurotrophic factor (BDNF)	AAV2	IP	Mark Tuszynski	I	NCT05040217
	Nerve growth factor (NGF)	AAV2	IC	Sangamo Therapeutics	II	NCT00876863
Canavan Disease	Aspartoacylase (ASPA)	AAV9	IV	Aspa Therapeutics	I/II	NCT04998396
Huntington's Disease	microRNA targeting huntingtin (miHTT)	AAV5	IP	UniQure Biopharma B.V.	I/II	NCT05243017
	Cholesterol 24-hydroxylase (CH24H)	AAVrh.10	IC	Brainvectis	I/II	NCT05541627
Parkinson's Disease	Neurturin (NTN)	AAV2	IP	Sangamo Therapeutics	I/II	NCT00985517
	Aromatic l-amino acid decarboxylase (AADC)	AAV2	IP	Genzyme, a Sanofi Company	I	NCT00229736
	Glial cell line-derived neurotrophic factor (GDNF)	AAV2	IP	National Institute of Neurological Disorders and Stroke (NINDS)	I	NCT01621581
Neuromuscular Disorders						
Duchenne Muscular Dystrophy	Micro-dystrophin	AAV2.5	IM	Nationwide Children's Hospital	I	NCT00428935
	Micro-dystrophin	AAV9	IV	Pfizer	I	NCT03362502
	Micro-dystrophin	AAVrh74	IV	West China Hospital	I	NCT06114056
Spinal Muscular Atrophy	Survival motor neuron 1 (SMN1)	AAV9	IV	United BioSource, LLC		NCT03955679
	Survival motor neuron 1 (SMN1)	AAV9	IV, IT	Novartis Pharmaceuticals	III	NCT05335876
Retinal Disorders						
Leber Congenital Amaurosis	Retinoid isomerohydrolase (RPE65)	AAV2	Subretinal	Applied Genetic Technologies Corp	I/II	NCT00749957
	Retinoid isomerohydrolase (RPE65)	AAV2/5	Subretinal	MeiraGTx UK II Ltd	I/II	NCT02781480
	Retinoid isomerohydrolase (RPE65)	AAV2/5	Intraocular	MeiraGTx UK II Ltd		NCT02946879
Achromatopsia	Cone photoreceptor cyclic nucleotide-gated B3 (CNGB3)	AAV2/8	Subretinal	MeiraGTx UK II Ltd	I/II	NCT03001310
	Cone photoreceptor cyclic nucleotide-gated A3 (CNGA3)	AAV2/8	Subretinal	MeiraGTx UK II Ltd	I/II	NCT03758404
Lysosomal Storage Diseases (LSDs)						
Gaucher Disease I	Glucosylceramidase beta 1 (GBA1)	AAV9	IV	Prevail Therapeutics	I/II	NCT05487599
	GCCase-85	AAVS3	IV	Freeline Therapeutics	I	NCT05324943
Gaucher Disease II	Glucosylceramidase beta 1 (GBA1)	AAV9	ICV	Xinhua Hospital, Shanghai Jiao Tong University School of Medicine	EARLY I	NCT06272149
Pompe Disease	Alpha glucosidase (GAA)	AAV8	IV	Asklepios Biopharmaceutical, Inc.	I/II	NCT03533673
	Alpha glucosidase (GAA)	AAV1	IM	University of Florida	I/II	NCT00976352
	Alpha glucosidase (GAA)	AAV9	IV	GeneCradle Inc	I/II	NCT05793307

Figure 1.2: Selection of AAVs in current clinical trials in the nervous system. Data from clinicaltrials.gov. IP = intraparenchymal, IC = intracerebral, IM = intramuscular, IV = intravenous, IT = intrathecal, ICV = intracerebroventricular.

However, therapeutic development for most nervous system diseases, especially in the PNS, is hindered by ineffective delivery. For instance, spinocerebellar ataxias (SCAs) and Rett syndrome are attractive gene therapy targets in the CNS but have not yet seen clinical successes due in large part to challenges associated with efficient delivery to the large affected structures[6, 7]. In the PNS, Schwann cells play a vital role in peripheral nerve regeneration

and are involved in several PNS-associated diseases, including neurofibromatosis, Charcot-Marie Tooth, and multiple sclerosis. However, no AAV-based vector has been shown to effectively transduce this cell type to date. This review will describe the current state of AAV-based gene therapy in the central and peripheral nervous systems while discussing the major hurdles AAVs will have to overcome.

1.3 Barriers to Efficient Transduction in the Nervous System

In general, numerous obstacles limit efficient transduction of the central and peripheral nervous system, including immune responses, physical tissue barriers, and cellular trafficking.

Immune barriers

The host immune system can significantly impede effective delivery and transgene expression, presenting one of the most critical challenges to AAV-based gene delivery. The presence of pre-existing neutralizing antibodies (NABs) against the AAV capsid in blood and other fluids bind and neutralize AAVs, preventing them from reaching their targets. Thus, any intravenously (IV) administered AAV, such as for attempted targeting of the PNS or blood brain barrier (BBB) crossing, must resist NAB neutralization. With up to 70% of the human population having anti-AAV antibodies due to natural AAV exposure[8], this immunological memory is a significant challenge. In addition, exceedingly high level NABs that develop following administration of AAV therapeutics can significantly limit the effectiveness of re-administration of the therapeutic, should it be necessary[9].

Routes of administration and physical barriers

Anatomical barriers pose hurdles for AAV delivery, and an optimal route of administration must be strategically chosen to reduce these depending on the nature and location of the disorder as well as the properties of a given AAV. For example, lysosomal storage disorders result in the accumulation of undegraded substrates throughout the body, including the CNS and PNS, leading to widespread pathology and necessitating widespread delivery of a potential therapeutic[10]. Neurodegenerative diseases such as Parkinson's, Alzheimer's, and Huntington's disease affect more specific regions of the brain, arguing for more targeted administration[11]. Similarly, neurosensory disorders, such as Leber Congenital Amaurosis (LCA) that affect the retina, entail highly localized delivery, enabling a higher therapeutic concentration in the delivered area and avoiding widespread biodistribution. Many diseases of the peripheral nervous system, like peripheral neuropathy and neurofibromatosis, affect cells throughout the body or are difficult to surgically access, likely warranting systemic delivery[12].

Intravenous administration offers the potential for widespread CNS and/or PNS delivery, though the blood-brain barrier (BBB) in the CNS and the blood-nerve barrier (BNB) in the PNS are major challenges. These barriers are comprised of endothelial cells with specialized tight junctions to protect neural tissue by restricting the passage of substances within the barrier[13, 14]. AAV9 and AAVrh.10 has been shown to cross the BBB in neonates, though crossing the mature BBB remains a major hurdle[15, 16]. Systemic administration of AAV vectors to reach the PNS has seen moderate success[17, 18] due to the fact that the BNB is known mediate higher levels of transcytosis, i.e. to be leakier, than the BBB[19, 20]. Direct injection of AAV vectors into the brain or PNS via intraparenchymal or peripheral nerve injection offers a potential solution to avoid these endothelial barriers and effectively transduce cells in a given region with relatively low dosages[21, 22, 23]; however, these localized routes of administration are not possible for cellular targets that are surgically difficult to access or occupy a large tissue volume. Administration of therapeutics into the CSF via intrathecal, intracisternal, or intracerebroventricular injections is an alternative for more widespread CNS delivery, including the spinal cord, and has demonstrated effectiveness in treating diseases such as SMA and ALS[24, 25]. While intra-CSF routes have demonstrated high levels of delivery in the spinal cord, transduction efficiency remains low in the adult brain due to limited diffusion beyond the ventricular lining[26, 27].

Cell-specificity and intracellular trafficking

Targeting and efficiently infecting specific cells within the diverse environment of the nervous system is yet another challenge for many CNS and PNS disorders. For example, Parkinson's disease is associated with a loss of dopamine-producing neurons in the substantia nigra, multiple sclerosis is an autoimmune attack against CNS glial cells, and neurofibromatosis is associated with tumors of Schwann cell-origin in the PNS[11, 28]. Naturally-occurring AAVs have limited overall delivery efficiency and are unable to achieve cell-specific transduction. The latter can lead to off-target toxicity, reduced efficacy, and/or the need for higher dosages. Natural AAVs utilize different cellular surface receptors to bind to target cells, including heparan sulfate proteoglycan (AAV2, 3B, 6, 13), sialic acid (AAV1, 4, 5, 6), and galactose (AAV9), influencing the types of cells to which they can bind[29]. While this primary glycan binding strongly influences viral tropism, internalization of the AAV is thought to depend on the transmembrane protein AAVR. First discovered by Pillay et al., AAVR binding has been shown to be essential for the internalization for many AAV serotypes[30]. Once internalized within the cell, AAVs face additional hurdles in endosomal escape, nuclear localization, uncoating of the capsid, and second-strand synthesis of the single-stranded DNA genome[31]. Hindrances in any one of these steps can substantially compromise downstream gene expression.

1.4 Engineering the AAV Capsid

Overcoming the immune, physical, and transport barriers to AAV transduction is in many cases critical for therapeutic benefit, and engineering the capsid is an attractive avenue due to its role in cell-specificity, immune recognition, and overall transduction efficiency. Increasingly utilized methods for engineering the capsid include rational design, directed evolution, and in silico design (**Figure 1.3**).

Rational Design

In rational design, mechanistic knowledge of viral interactions with the target cell or tissue type is harnessed to strategically design capsid variants. Such techniques include site-directed mutagenesis of specific amino acid residues and the insertion of rationally chosen targeting peptide sequences within the *cap* gene. For example, Zhong et al. discovered that phosphorylation of AAV2 vectors by protein tyrosine kinases resulted in ubiquitin-mediated degradation, preventing vectors from reaching the nucleus[32]. They additionally found that mutating surface-exposed tyrosines to non-phosphorylatable residues led to an increase in transduction efficiency[33]. Wang et al. used strategic point mutations to modify the three-fold axis of AAV9, which has been shown to play a role in receptor binding and blood clearance of rAAV vectors. They generated AAV9.HR, an AAV9-based vector with H527Y and R533S double mutations at the three-fold axis, and demonstrated its capacity for CNS transduction with PNS-detargeting properties in neonatal mice[34].

Strategic insertions of peptides and ligands into the AAV capsid also led to the identification of novel CNS-tropic capsids. Girod et al. first demonstrated peptide insertion to retarget AAV2 by inserting a 14-amino-acid peptide into six different putative loops of the capsid[35]. One of these variants successfully transduced an otherwise AAV2-resistant cell line. Yao et al. displayed a cell-penetrating peptide on the AAV9 capsid, and the resulting variant AAV.CPP.16 demonstrated enhanced transduction in both mouse and NHP brain tissue[36].

Directed Evolution

Rational design strategies, however, are limited by an incomplete mechanistic understanding of the many complex ways in which AAV capsid elements interact with tissues and cells to mediate various steps of the viral transduction pathway (e.g. biodistribution, tissue spread, vector internalization, intracellular trafficking, genome expression). Directed evolution, a process designed to mimic natural evolution, addresses this limitation by creating a diverse library of viral variants and subjecting it to iterative rounds against a selective pressure to isolate optimal performers. A key advantage of directed evolution is that a mechanistic understanding of the underlying delivery barriers is not necessary, yet molecular investigation of the resulting variants can enhance mechanistic knowledge, akin to forward genetic screens in model organisms.

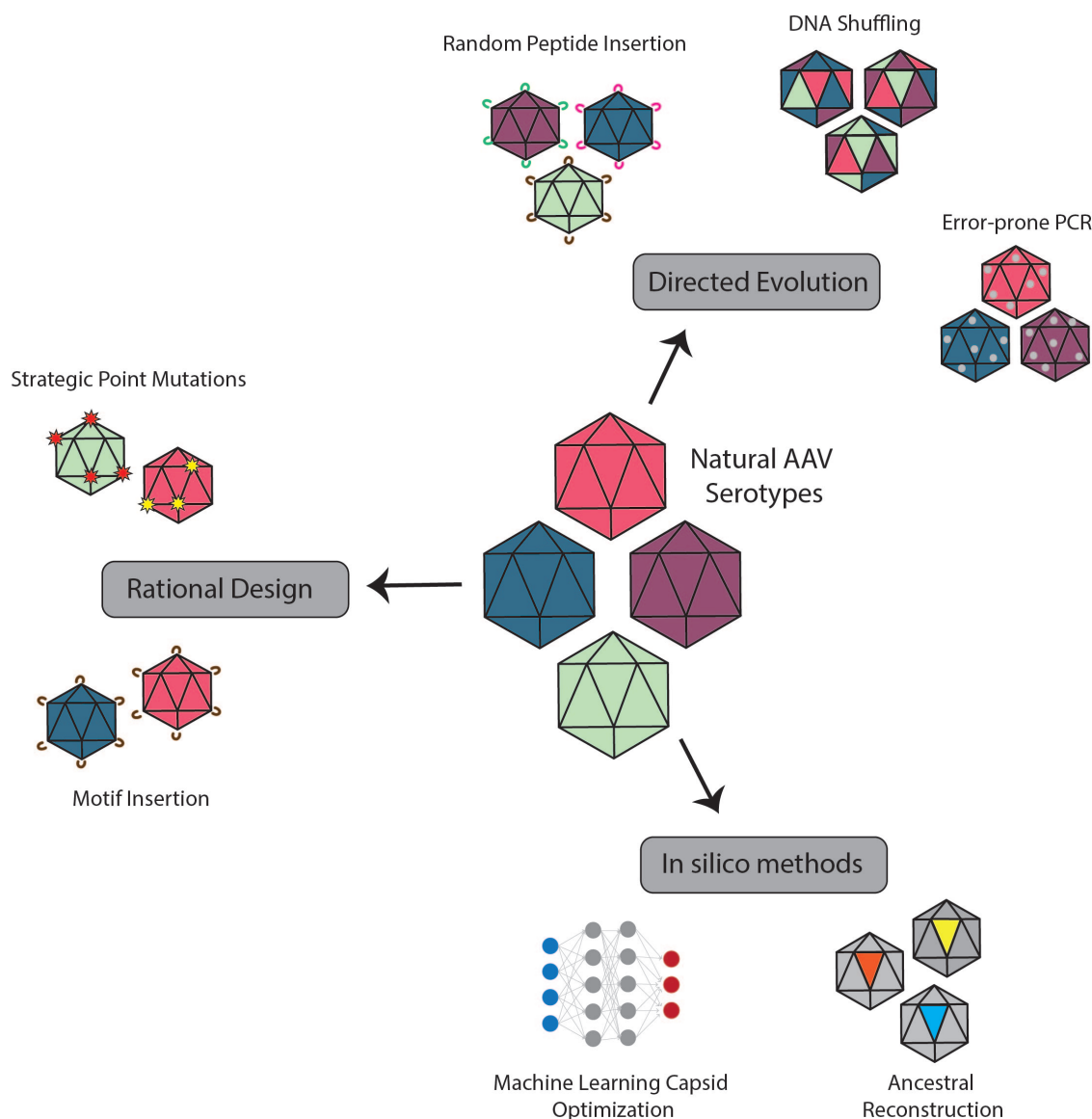


Figure 1.3: Engineering of AAV capsids. Rational engineering strategies use biological knowledge to guide modifications of the AAV capsid through methods such as motif insertions or strategic point mutations. Directed evolution techniques generate highly diverse libraries of AAV variants, which are then subjected to a selective pressure to identify novel variants with enhanced properties. *In silico* strategies use computational methods to develop improved AAV libraries as a better starting point for variant selection.

Directed evolution was first applied to AAV by Maheshri et al., work in which an AAV2 *cap* library was generated through error-prone PCR and selected for the capacity to mediate gene delivery in the presence of neutralizing antibodies or for altered binding to the AAV2 primary receptor heparan sulfate[37]. This study resulted in the isolation of several variants that exhibited reduced antibody neutralization *in vitro* and *in vivo*. Additional techniques

for capsid diversification have been developed, including DNA shuffling[38, 39] and random peptide insertions[40, 41], enabling the identification of novel AAV capsid variants with enhanced transduction towards cells in the nervous system. *In vivo* directed evolution studies offer the advantage of identifying more clinically-relevant variants. Dalkara et al. identified 7m8 through an *in vivo*-directed evolution screen in mice designed to identify variants that could efficiently transduce the retina via a non-invasive, intravitreal injection in the eye[42], an approach that has since been extended to dog[43] and NHP[40]. Another study evolved rAAV2-retro, an AAV variant that mediates retrograde access to projection neurons at comparable efficiencies as synthetic tracer dyes[44]. For CNS diseases associated with large volumes of neuronal tissue, rAAV2-retro offers a potential increase in therapeutic distribution within the tissue.

Additional techniques were developed to further enhance the selection process, such as barcoding AAV genomes to analyze *cap* variants in a high-throughput manner[45, 46] and implementing Cre-lox systems to select for cell-specific, nuclear localizing variants *in vivo*[41, 47, 48, 49]. Recently, Chuapoco et al. identified AAV.CAP-Mac after screening AAV9-based peptide insertion libraries in marmosets and macaques[50]. AAV.CAP-Mac demonstrated higher delivery to the brain in marmosets, rhesus macaques, and green monkeys after systemic administration, demonstrating the ability to cross the BBB in NHPs. In addition, directed evolution studies have recently shifted towards RNA-based selections to enrich for variants that are able to successfully deliver as well as transcribe their cargo[46, 51, 52]. For instance, Stanton et al. identified novel proline arginine loop (PAL) capsid variants through an RNA-based screen in juvenile cynomolgus macaques that demonstrated higher transduction efficiencies in the macaque cerebrum compared to AAV9[53].

In silico-guided capsid optimization

Computational methods have also been used to generate novel AAV libraries, addressing gaps in both rational design and traditional directed evolution diversification strategies. SCHEMA is a computational algorithm that predicts optimal recombination junctures in proteins without disruption of the three-dimensional structure, facilitating the development of random chimeras of natural AAV serotypes[54]. Ojala et al. applied SCHEMA to natural AAVs to generate a new chimeric capsid library. The variant SCH9 was selected from this library for high levels of transduction in subventricular neural stem cells of mice[49]. Another computer-guided method of developing new capsids is through ancestral capsid reconstruction[55]. Through the reconstruction of phylogenetic trees to predict putative ancestral sequences, multiple studies have created ancestrally-based AAV capsids libraries, which resulted in the identification of thermostable variants (Anc80L65 and Anc80) with improved tropism towards various tissue types[55, 56].

Enabled by next generation sequencing, machine learning approaches are increasingly being applied to large datasets connecting viral sequence to function to train algorithms that predict the impact of capsid modifications on viral behavior. Ogden et al. comprehensively characterized individual mutations within the AAV2 capsid. Specifically, they

generated a single amino acid mutation AAV2 *cap* library that encompassed all single amino acid substitutions, insertions, and deletions at all 735 positions to generate an AAV2 capsid fitness landscape. Their results revealed many interesting findings, including the benefit of negatively-charged amino acids at external projections, and the generally deleterious property of positive charges across *cap*[57]. This work also revealed a novel open reading frame, membrane-associated accessory protein (MAAP). A subsequent study applied deep learning to a 28-amino acid region of the AAV2 capsid to generate highly diverse capsid proteins that were still able to package a DNA genome. Through their machine learning models, a library of over 200,000 variants was designed, 57,000 of which retained the ability to package[58]. Huang et al., who had previously discovered a novel BBB-crossing mechanism of the AAV.PHP.B viral family via interactions with the mouse CNS endothelium surface receptor LY6A[59], utilized protein target-binding assays to screen AAV capsid libraries for the ability to bind LY6A and an additional surface protein LY6C1. ML-guided approaches then utilized these datasets to further explore the capsid sequence space and generate additional BBB-crossing sequences[60].

Machine learning can also be used to design ‘more packageable’ AAV libraries to provide a better starting point for selections that are broadly applicable to various selective pressures. Danqing et al. developed a machine learning algorithm to design an AAV5-based peptide insertion library that optimizes the ability to package while maintaining high levels of diversity[61]. The newly designed library was then subjected to a round of selection in primary human brain tissue, which identified top variants that both packaged at high titers and demonstrated high glial transduction in human brain slices. The integration of machine learning with AAV capsid engineering represents a transformative step towards more effective, safe, and tailored gene therapies for a variety of diseases.

1.5 Future Perspectives

Motivated by considerable unmet medical need, the central nervous system remains one of the most pursued therapeutic areas in AAV clinical trials[62], with approximately 60 AAV-CNS therapies advanced to the clinic to date[63]. Notably, only few clinical trials have addressed PNS-related diseases (clinicaltrials.gov). Improved capsids are still needed for the central and especially for the peripheral nervous system. For example, Schwann cells are an important therapeutic target due to their vital role in peripheral nerve health and survival, and the ability to target these cells will offer avenues for treatment of several related diseases including neurofibromatosis, multiple sclerosis, and Charcot-Marie-Tooth disease. As the PNS neurological network extends to every corner of the body to relay commands and ensure the brain can respond to any external and internal stimulus, delivery is challenging compared to the relatively compartmentalized CNS. PNS-associated therapeutics will need to widely target the entire system, likely via systemic administration. Exposure to neutralizing antibodies would thus pose an additional challenge. Several initiatives are underway to create novel AAVs to overcome these PNS challenges.

Machine learning-guided approaches hold great potential for novel AAV library generation and analysis. One challenge is that work to date has primarily focused on capsid regions that are short enough to fit within a next-generation sequencing read, excluding libraries whose diversity spans the entire capsid such as shuffled or ancestral libraries. Long-read sequencing offers the potential to cover more expansive regions of the *cap* gene, but at substantially lower read depth and relatively high error rates[64]. Improvements in highly accurate, high-throughput, long-read sequencing methodologies can allow ML-guided strategies to be applied to a wider range of library diversification techniques.

In addition to engineering the capsid, AAV vector manufacturing is another major concern for diseases requiring high dosages of AAV, though studies have engineered producer cells or other genomic elements to improve manufacturing titers of AAV[65, 66]. Barnes et al. tackled this issue by conducting genome-wide screens in HEK293 producer cells to identify genetic factors that increased AAV production[65]. Focusing on the AAV genome itself, Schieferecke et al. engineered the AAV-associated protein MAAP, which has been shown to play a role in AAV secretion[66]. After rounds of packaging selection, the evolved bulk MAAP variant library demonstrated increases in both overall titer and ratio of secreted vector genomes. They also identified a MAAP variant, MAAP-SL08, that demonstrated increased transgene packaging in AAV2, AAV6, and AAV9[66].

Another important factor in AAV gene therapy success is post-delivery toxicity. Studies have also shown serious adverse events may occur or worsen due to immunogenicity of AAV vectors, their occurrence often increasing with higher dosages. Atypical hemolytic uremic syndrome (aHUS), defined by thrombocytopenia, hemolytic anemia, and acute kidney injury, is linked to higher dosages of AAV due to complement activation[67]. In addition, a recent clinical trial using AAV-based delivery of the human MTM1 gene to treat X-linked myotubular myopathy resulted in four deaths due to hepatotoxicity[68]. There has also been evidence of dorsal root ganglion (DRG) inflammation and toxicity associated with direct delivery into the cerebral spinal fluid (CSF) or high dosages in the blood in non-human primates (NHPs)[69]. These studies highlight the importance of AAV targeted delivery, which can lead to reduced dosages and a higher safety profile of AAV vectors.

1.6 Conclusion

Substantial progress has been made in the development of novel AAV capsids for enhanced transduction in the central and peripheral nervous systems. These capsids are able to bypass anatomical barriers and/or mediate efficient cellular transduction to enable more effective gene expression. While additional development is needed in vector delivery, advances towards these targets offer extensive therapeutic potential for the treatment of genetic CNS- and PNS-associated diseases. Other important considerations in addition to delivery include manufacturing costs, transgene expression durability, genome optimization, and immune clearance. Furthermore, cross-species differences between animal models and humans significantly hinder translation of preclinical studies to clinical development. These challenges

highlight the need for further exploration of these areas and addressing them will enhance the therapeutic potential of AAV-based gene therapies in the CNS and PNS.

Chapter 2

Directed Evolution of Novel AAV Capsids for Enhanced Delivery to Mouse and Human Schwann Cells

2.1 Chapter Abstract

Effective delivery to target cell types remains a major obstacle in the field of gene therapy. Recombinant adeno-associated viruses (AAVs) have proven to be effective delivery vectors for the treatment of several disease types yet have not demonstrated efficient delivery to many cell types, including cells of the peripheral nervous system. Schwann cells, essential for myelination, nerve regeneration, and general nerve maintenance, are strategic gene therapy targets for treating peripheral neuropathies but present unique challenges for gene delivery due to their protective barriers and extensive localization throughout the body. To address this issue, we engineered novel AAV variants with enhanced transduction efficiency towards Schwann cells. Using directed evolution, we identified two groups of capsids with significantly improved delivery to Schwann cells in mice or human. In addition to higher transduction efficiencies, the human-tropic capsids also demonstrated higher resistance to antibody neutralization compared to natural serotypes. *In vivo* studies of the mouse-tropic capsids confirm higher specificity, improved packaging titers, and decreased localization to the liver. These engineered AAVs can provide a valuable tool for studying Schwann cells, offering new avenues for the treatment of peripheral nerve disorders.

2.2 Introduction

Gene therapy offers the potential for prolonged, and even permanent, therapeutic efficacy through the delivery of genetic material to cells to treat human disease. A primary challenge in the field, however, is the effective delivery of the therapeutic to the intended cell type and region of the body. Recombinant adeno-associated viruses (AAVs) are the most broadly-

used delivery vectors *in vivo*[3] and have already seen success in the treatment of a variety of disorders, including neurological, ocular, and hematological diseases[70]. Their lack of pathogenicity, minimal integration into the host genome, and long-term transgene expression make them advantageous delivery vehicles over other viral vectors[71]. Over 100 AAV-based therapies have advanced to clinical trials, and six to date have received US Federal Drug Administration (FDA) approvals (fda.gov). Given this track record, AAV-based gene therapy holds great potential for the delivery of therapeutics to a wide range of tissue types, including cells of the peripheral nervous system (PNS).

Schwann cells are the primary glial cells of the PNS and are crucial for the survival and function of peripheral nerves. They myelinate neurons (depending on their subtype), help guide developing or regenerating neurons to their target site, and are involved in the conduction of action potentials along the axon[72]. Dysregulation of Schwann cell function has been associated with several pathologies, including Charcot-Marie Tooth disease, neurofibromatosis, and multiple sclerosis[73, 74]. For example, neurofibromatosis type 1 (NF1), which results from mutations in the *NF1* gene, often involves the formation of Schwann-cell derived neurofibromas, benign tumors that form along peripheral nerves. Due to their proximity to neurons, they can cause disfigurement, cognitive learning deficits, and loss of motor function, and they additionally have the potential to develop into malignant peripheral nerve sheath tumors (MPNSTs)[75, 76]. Thus, the ability to deliver therapeutics to Schwann cells offers the potential to preventatively and actively treat such conditions. However, given the dispersed nature of the PNS, there is a need for widespread, yet targeted, delivery throughout the body. AAV-mediated delivery of genetic material via systemic administration provides a potential treatment for such disorders.

Natural serotypes of AAV have previously demonstrated moderate efficacy in transducing Schwann cells *in vitro* and *in vivo*[77, 78, 79, 18]. In particular, AAV1, 2, and 6 have been shown to be the best natural serotypes in targeting Schwann cells across rat, mouse, and human cultures and nerve segments[77, 78, 79]. However, naturally-occurring AAVs exhibit broad tropism, resulting in low on-target and considerable off-target delivery, and suffer from high levels of neutralization by pre-existing neutralizing antibodies (NABs)[80]. For example, natural serotypes are predominantly retained in the liver after intravenous delivery, diminishing therapeutic delivery to the target cell and increasing the likelihood of hepatotoxicity[67]. Thus, any therapeutic candidate must be able to efficiently target Schwann cells while avoiding neutralization and minimizing localization to the liver to achieve both efficacy and safety.

In this study, our objective was to engineer AAV capsids with improved delivery to Schwann cells, in both mouse to harness animal models of human disease and human for clinical translation. Accordingly, we conducted alternating rounds of evolution in *in vitro* human Schwann cell (hSC) cultures and stringent *in vivo* mouse selections to identify novel AAV variants. Surprisingly, our evolved variants demonstrated improved transduction towards one species, but not both. We identified two groups of capsids: hSC-tropic capsids C1 and C6, and mouse Schwann cell (mSC)-tropic capsids SCHW1 and SCHW2. C1 and C6 demonstrate high transduction efficiencies of hSCs and display high resistance to human

antibody neutralization. SCHW1 and SCHW2 exhibit higher transduction efficiency and specificity of delivery to mSCs compared to natural serotypes. These variants also show significantly improved packaging titers and lower biodistribution to liver, contributing to better manufacturing potential and safety profiles. While our results suggest that the selective pressures exerted by infection in each species were distinct enough to evolve variants towards one species or the other, these groups of AAV capsids offer new tools for the development of treatments for peripheral neuropathies across different model systems.

2.3 Results

Iterative *in vivo* and *in vitro* selections reveal two hSC-tropic AAV variants

We applied directed evolution to diverse AAV capsid libraries to select for capsids with enhanced Schwann cell transduction. Seven previously-designed AAV capsid libraries were selected for evolution: 1) an AAV1 seven-amino acid (7mer) peptide insertion at amino acid 590, 2) an AAV2 7mer peptide insertion at amino acid 588[81], 3) an AAV5 7mer peptide insertion at amino acid 575[61], 4) a SCHEMA AAV library[49], 5) a chimeric shuffled library comprised of AAV1, 2, 4, 5, 6, 8, and 9[39], 6) an error-prone library based on a shuffled variant, and 7) an ancestrally-reconstructed library[55]. These libraries were cloned into the pSub2FloxCap backbone, described previously[49], in which mutant sites loxP66 and loxP71 flank the *cap* gene to favor unidirectional inversion of the *cap* (**Figure 2.1a**). Libraries were then packaged in HEK293T cells via triple-transfection, purified, and mixed at an equimolar ratio for iterative selection.

To select for novel AAV capsids with enhanced transduction in mouse and human Schwann cells, alternating rounds of *in vivo* and *in vitro* selection were performed (**Figure 2.1b**). *In vivo* rounds of selection were conducted in C57BL/6 mice expressing Cre recombinase downstream of the Schwann cell-expressed desert hedgehog regulatory locus (DhhCre). Variants that successfully transduced Schwann cells *in vivo* undergo inversion of their *cap* sequences, allowing for specific PCR amplification of Schwann cell-transducing variants, as described in analogous systems[48, 49]. Selections in DhhCre mice were performed by systemically injecting AAV via the tail-vein at 1×10^{11} - 1×10^{12} viral genomes (vg)/mouse in eight-week-old mice to identify variants that can transduce Schwann cells while overcoming physical and immunological barriers. For our human Schwann cell-based system, we utilized the human telomerase reverse transcriptase (hTERT)-immortalized Schwann cell line hTERT ipn02.3 2 λ , generated via the exogenous expression of both hTERT and murine cyclin-dependent kinase (mCdk4) from patient-derived wild-type Schwann cells[82, 83, 84, 85].

Three rounds of selection were conducted initially (**Figure 2.2a**). The combined AAV library was first injected systemically into DhhCre mice ($n = 5$). Two weeks later, AAV genomic DNA was extracted from the sciatic nerve using a modified Hirt extraction protocol. To prevent too stringent of a selection pressure initially, we amplified all *cap* genes (both

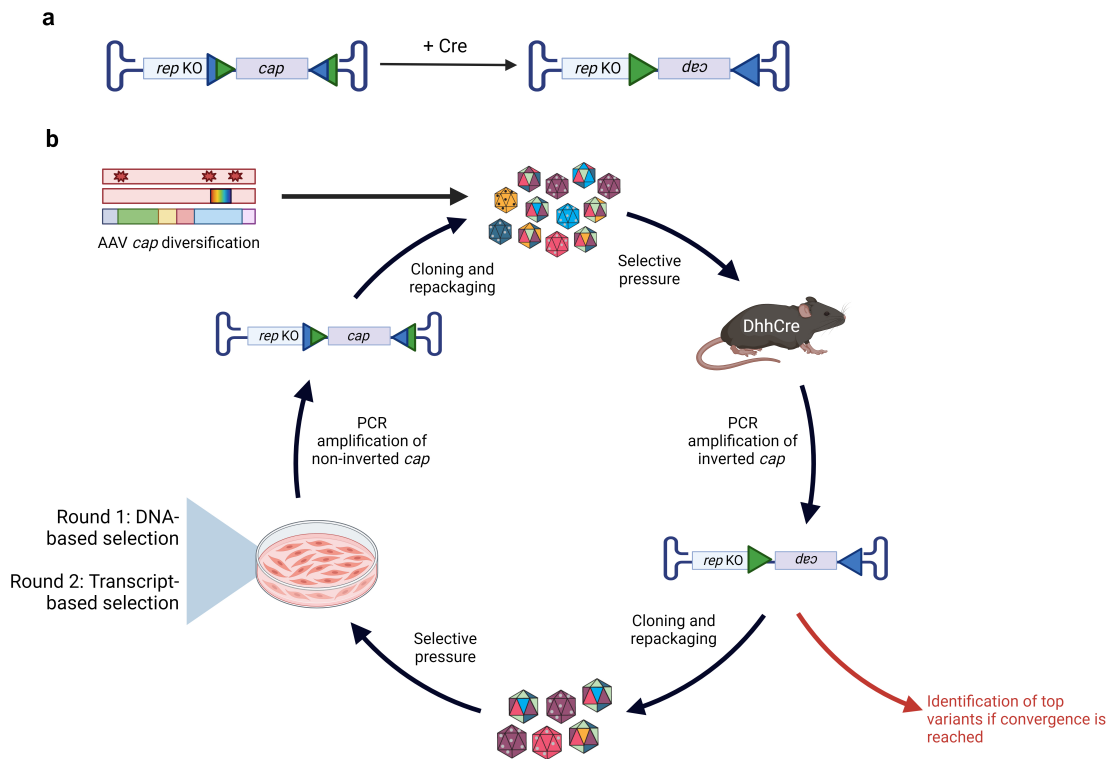


Figure 2.1: Overview of directed evolution strategy. (a) Schematic depicting positioning of the mutant loxP sites in the pSub2FloxCap library plasmid. The location of the lox66 (5' loxP site) disables rep. In the presence of Cre recombinase, the *cap* gene is inverted. (b) Diagram exhibiting the selection strategy used to identify Schwann-cell tropic capsids. Alternating rounds of selection in DhhCre mice and Schwann cells (hSCs) were conducted until convergence of the libraries was achieved.

inverted and non-inverted) from the mouse sciatic nerve tissue, and the resulting amplicons were cloned back into the pSub2FloxCap backbone and repackaged. Upon sampling with Sanger sequencing, most libraries were still represented in the selected library.

The second round of selection was then performed in hSCs. hSCs were transduced at an MOI of 5000 and incubated at 37°C. After three days, viral genomic DNA was harvested, and AAV *cap* genes were amplified and subsequently cloned back into the pSub2FloxCap backbone and repackaged. Sampling with Sanger sequencing revealed approximately 40% of the selected variants were from either the shuffled or SCHEMA libraries.

The final round of selection was performed *in vivo* ($n = 5$). AAV DNA was extracted from the sciatic nerve after two weeks, and only Schwann cell-transducing variants (i.e., inverted *cap*) were PCR amplified. After these three alternating rounds *in vivo* and *in vitro* selection, Sanger sequencing analysis of 72 clones showed convergence of the library to three variants: C1, C5, and C6. Variants C1 and C5 were from the randomly shuffled library, whereas variant C6 was a SCHEMA clone comprised of AAV2, AAV8, and AAV9 blocks.

Before characterization assays were conducted, we first assessed the packaging titer of our variants. Manufacturability and scalability are important factors that contribute to clinical success; thus, we wanted to ensure the packaging fitness of our variants was not impacted in the directed evolution process. Upon packaging in HEK293T cells, we quantified viral packaging yields. C5 packaged substantially better than both AAV1 and AAV2, while C1 and C6 demonstrated a three-fold increase in titer over AAV2 and were comparable to AAV1, highlighting the ability for scale-up production of all three vectors (**Figure 2.2b**).

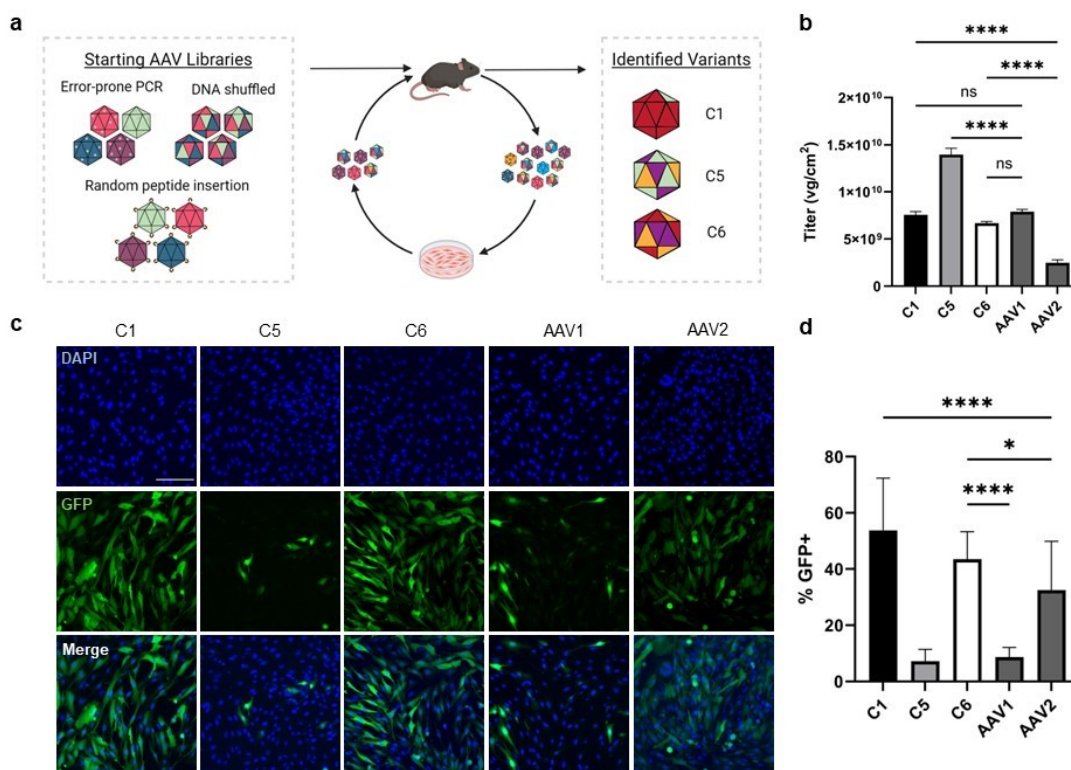


Figure 2.2: Evolution of hSC-tropic capsids. (a) Diagram representing the evolution process to show identification of variants C1, C5, and C6. (b) Packaging titers of C1, C5, and C6 compared to AAV1 and AAV2. Data are normalized to the cell culture plate surface area and presented as mean $\pm$ SD, $n = 3$. (c) Representative images of the transduction of C1, C5, and C6 in hSCs at an MOI of 5000. Cells were stained for GFP (green) and DAPI (blue). Scale bar represents 100 μ m. (d) Quantification of transduction efficiency as the percentage of GFP+ cells. Data are presented as mean $\pm$ SD, $n = 3$. Statistical analyses were done with a one-way ANOVA with Tukey's multiple comparisons test ($*p < 0.05$, $****p < 0.0001$).

These variants were next analyzed for transduction of hSCs, in comparison to AAV1 and AAV2, using a self-complementary CAG-GFP cassette (scCAG-GFP). Based on previous work[77, 79] and our preliminary screen of natural serotypes on hSCs (**Figure A.1**), AAV1 and AAV2 were chosen as controls for validation studies. hSCs were transduced with each variant at an MOI of 5000, fixed, and stained for GFP and DAPI three days later. Variants

C1 and C6 demonstrated a higher percentage of transduced human Schwann cells compared to AAV1 and AAV2, while variant C5 performed similarly to AAV1 (**Figure 2.2c-d**).

Given the improvement in transduction efficiency in hSCs, we next assessed efficacy in mice; however, these variants unexpectedly performed only comparably to the wild-type AAVs. Specifically, validation was performed *in vivo* using next-generation sequencing (NGS) to quantify enrichment of the variants in mouse sciatic nerve. Unique 16-bp DNA barcodes were cloned downstream of GFP in the plasmid pAAV-FLEX-GFP, which contains a doubly-fluxed inverted GFP gene. In the presence of Cre recombinase, the inverted GFP gene is reversed and is thus expressed under the CAG promoter. C1, C5, and C6, as well as AAV1 and AAV2 as controls, were packaged with the barcoded pAAV-FLEX-GFP cassette, linking each variant with a unique barcode (**Figure A.2a**). Barcoded AAVs were mixed at an equimolar ratio and intravenously administered to DhhCre mice at 9×10^{11} vg/mouse. Viral DNA was harvested from sciatic nerves two weeks after administration. Schwann cell-transducing barcodes were then amplified from the recovered DNA with primers specific for reverted GFP sequences and prepped for NGS using nested PCRs. Enrichment scores for each AAV were calculated, and analysis of the deep sequencing data revealed variant C5 and AAV1 demonstrated positive enrichment in mSCs, though AAV1 appeared to be the most positively enriched. Variant C1 showed slight positive enrichment, while variant C6 and AAV2 were significantly de-enriched (**Figure A.2b**).

Characterization of variants C1 and C6

The results from the variant transduction profiles *in vitro* and *in vivo* prompted further evaluation of the hSC-tropic capsids C1 and C6. C1 is a variant originating from the randomly shuffled library, comprised of only AAV2 and AAV6. The first 183 amino acids, belonging to the VP1/VP2 unique region, align to AAV6, while the rest of the capsid aligns to AAV2 (**Figure 2.3a-b**). Interestingly, three point mutations were also present in C1: Q598L, I648V, V708I (mapping of these single amino acid mutations can be seen in **Figure A.3**). Variant C6 is derived from the SCHEMA AAV library (chimera AAV2, 9, 9, 8, 2, 2, 8, 9) (**Figure 2.3a**). Assembly of the three-dimensional structure of C6 indicates a dominance of AAV2 at the three-fold protrusions and AAV9 at both the five-fold and two-fold axes of symmetry (**Figure 2.3b**).

Neutralization by pre-existing neutralizing antibodies (NABs) is another major obstacle that AAV-based delivery methods must overcome to effectively deliver its cargo. To assess the neutralizing properties of our novel human-tropic capsids, we quantified the ability of our variants to avoid neutralization by human intravenous immune globulin (IVIG), or pooled antibodies. C1 and C6, along with AAV1 and AAV2, were packaged with scCAG-GFP, incubated with serial dilutions of IVIG, and then used to transduce a permissive HEK cell line, 2v6.11. We quantified both percentage of GFP-positive cells transduced and total intensity to fully capture antibody evasion (**Figure 2.3c-d**). Comparing IC₅₀ concentrations, C1 demonstrated similar neutralization resistance as AAV1 and a seven-fold increase in re-

sistance compared to AAV2. Most notably, variant C6 showed a 16-fold improvement over AAV2 and a two-fold improvement over AAV1.

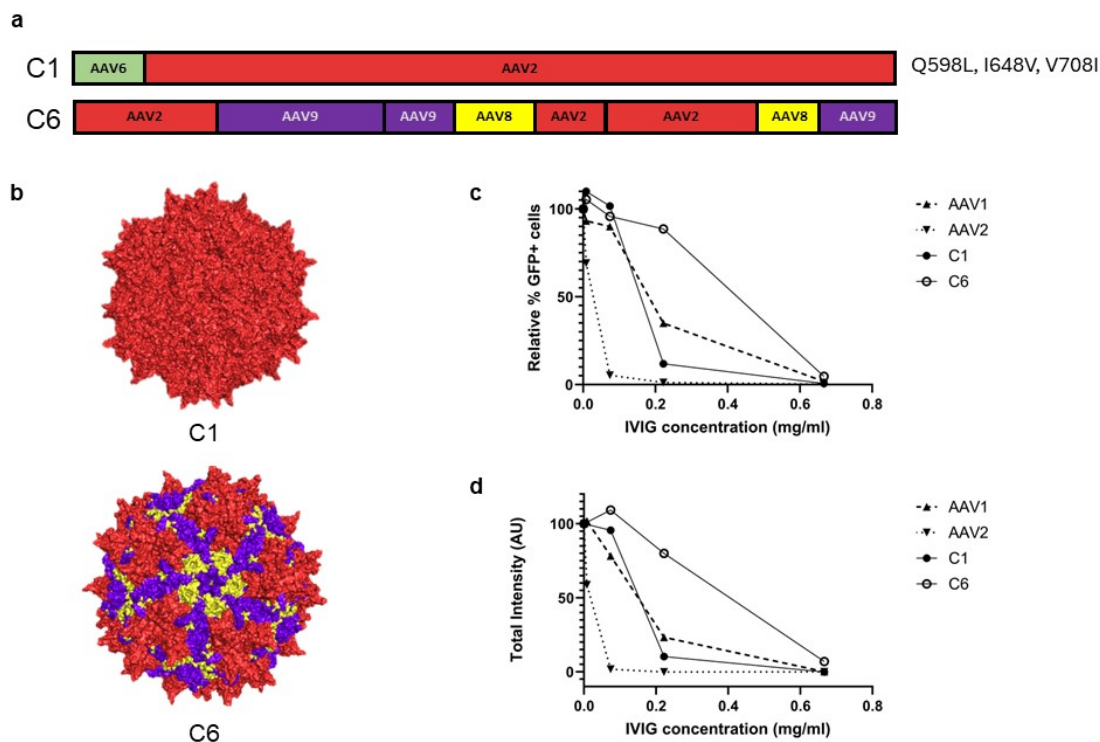


Figure 2.3: Characterization of hSC-tropic capsids. (a) Block schematic indicating the composition of each variant with the corresponding parental serotype. (b) Full capsid assembly of each variant, mapped onto AAV2 (PDB: 1LP3). AAV2 = red, AAV8 = yellow, AAV9 = purple). (c and d) Neutralization of C1, C6, AAV1, and AAV2 after incubation with serial dilutions of human IVIG and transduction of 2v6.11 cells ($n = 3$). Neutralization was quantified by (c) the number of GFP+ cells and (d) total fluorescence intensity normalized to untreated AAV.

Re-diversification and additional evolution yield new mSC-tropic capsids SCHW1 and SCHW2

The evolved variants C1 and C6, while demonstrating enhanced transduction of hSCs compared to AAV1 and AAV2, still lack efficient delivery to Schwann cells *in vivo*. These capsids have human clinical relevance for potential therapeutic development; however, we wanted to assess whether it was possible to derive variants with both human tropism *in vitro* and mouse infectivity *in vivo*. Evolved variants with the ability to effectively transduce SCs

in both species would yield insights into determinants of species selective tropism, as well as enable the use of mouse efficacy models for pre-clinical development. Accordingly, we re-diversified our selected variants for additional rounds of evolution to enhance delivery to murine Schwann cells *in vivo* while ideally maintaining their delivery capabilities to hSCs.

Random 7mer peptides libraries, flanked by linkers, were inserted into the C1 and C6 variants to establish new capsid libraries. An LA-(NNK)₇-A peptide library was inserted at amino acid position 586 of the C1 capsid and amino acid 588 of the C6 capsid (VP1 numbering) (**Figure 2.4a**). C5 was also included in the re-diversification process due to its positive enrichment in mSCs (**Figure A.2b**), in which a TG-(NNK)₇-GLS peptide library was inserted at amino acid 590 (**Figure 2.4a**). To enable both Cre-based and transcription-based selections for this round of evolution, the three peptide insertion libraries were then cloned into the pSub2-CMV-FloxCap backbone, which has a CMV promoter driving expression of the floxed *cap*. Estimated diversity of these new libraries ranged $5 \times 10^6 - 1 \times 10^7$ unique variants and were combined at an equimolar ratio. The AAV5 7mer library[61] was also cloned into the pSub2-CMV-FloxCap backbone and spiked in to increase diversity of the starting library.

Our libraries were then subjected to two rounds of selection: one *in vitro* transcript-based selection followed by a round *in vivo* (**Figure 2.4b**). hSCs were transduced by the re-diversified library at an MOI of 15,000. RNA was harvested after three days and reverse transcribed to cDNA. *Cap* genes were PCR amplified and recloned into the pSub2FloxCap backbone. The selected library was then injected intravenously into DhhCre mice ($n = 5$), and after two weeks, inverted *cap* genes were PCR amplified from the harvested sciatic nerve gDNA. Because double-stranded DNA is required for successfully Cre recombination (necessitating successful viral trafficking), RNA recovery *in vivo* was not needed to ensure isolation of capsids that had undergone nuclear localization and second strand synthesis. Sanger sequencing of 58 clones revealed convergence of the library to two variants: SCHW1 and SCHW2. SCHW1 is a C1-based variant with the insertion LA-DSNRSAM-A (**Figure 2.4c**), and SCHW2 is an AAV5-based variant with insertion TG-DTNSGIV-GLS (**Figure 2.4d**).

SCHW1 and SCHW2 were first characterized for packaging efficiency. SCHW1 packaged four-fold more efficiently than AAV1 and 23-fold more efficiently than AAV2, whereas SCHW2 showed a 2.8-fold higher degree of packaging compared to AAV1 and a 13-fold increase compared to AAV2 (**Figure 2.5a**).

To determine whether SCHW1 and SCHW2 retained the capacity to transduce human Schwann cells, we transduced hSCs with our variants and AAV1 at varying MOIs. Contrary to expectation, transduction efficiencies were lower than AAV1 (**Figure A.4**).

We then validated the transduction efficiency of these two variants *in vivo*. Variants were packaged with scCAG-GFP and injected systemically at 5×10^{11} vg/mouse into wild-type C57BL/6J mice. AAV1 was used as the control based on our previous *in vivo* validation (**Figure A.2b**). Sciatic nerves were harvested after 3.5 weeks, fixed, cryosectioned, and stained for GFP, S100 (a Schwann cell marker), and DAPI (**Figure 2.5b**). Transduction efficiency was calculated by the percentage of Schwann cell-stained area transduced by

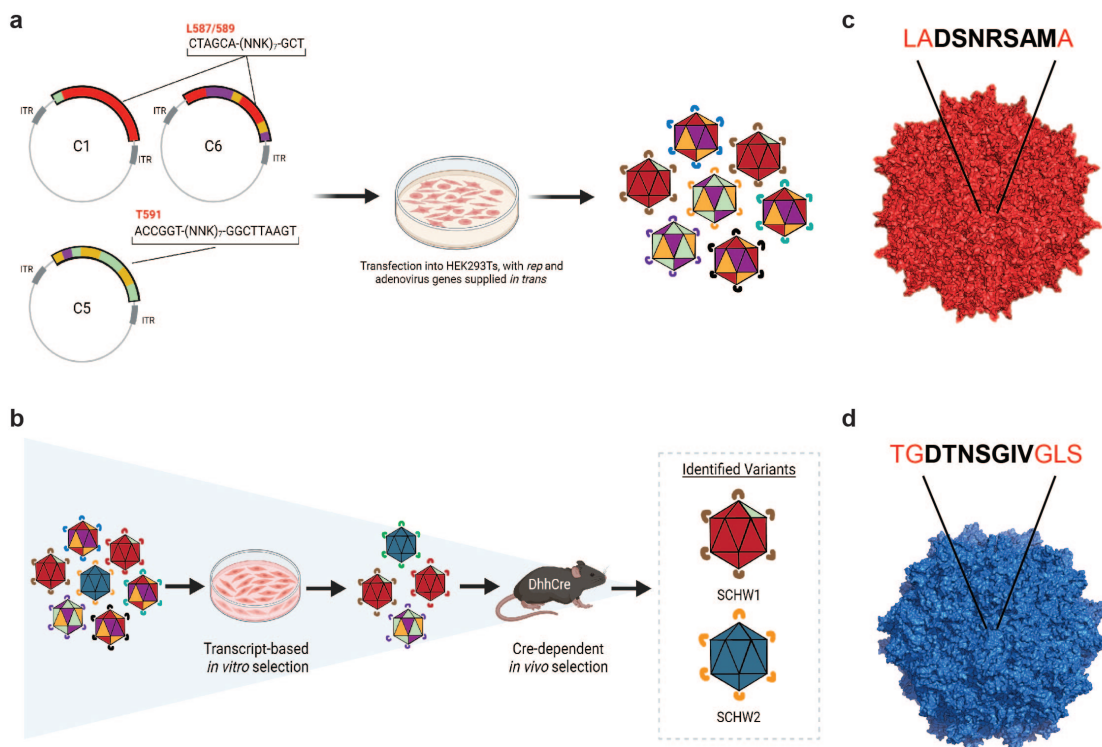


Figure 2.4: Evolution of mSC-tropic capsids. (a) Schematic of 7mer capsid library design and viral production to generate re-diversified library. (b) Diagram representing the evolution process to show identification of variants SCHW1 and SCHW2. (c) Three-dimensional structure of SCHW1 with peptide insertion sequence indicated. Structure was mapped onto AAV2 (PDB: 1LP3). (d) Three-dimensional structure of SCHW2 with peptide insertion sequence indicated (PDB: 7PK3). Graphics created with BioRender.com and PyMol.

AAV. SCHW1 demonstrated higher transduction of Schwann cells compared to AAV1, while SCHW2 showed comparable efficiencies to AAV1 (**Figure 2.5c**). Schwann cell specificity, quantified by the percentage of AAV-transduced area stained by S100, was also calculated. Interestingly, even though our evolution process did not select for Schwann cell specificity, both SCHW1 and SCHW2 exhibited significantly improved specificity towards Schwann cells *in vivo*, with SCHW2 demonstrating a 2.5-fold increase in specificity relative to AAV1 (**Figure 2.5d**).

Hepatotoxicity is an adverse event associated with high dose intravenous AAV delivery. Liver detargeting, which can accompany enhanced targeting of a desired cell target, is a beneficial property that can enable lower dosages and reduce the risk of hepatotoxicity. Following intravenous administration, we quantified SCHW1 and SCHW2 viral DNA in the liver after 3.5 weeks. SCHW1 demonstrated a 2.5-fold lower transduction and SCHW2 a 57-fold reduction in liver transduction (**Figure 2.5e**). Thus, the mouse-tropic variants SCHW1 and SCHW2 demonstrate improvements in Schwann cell transduction, specificity, packaging, and liver detargeting, though with decreased hSC transduction *in vitro*.

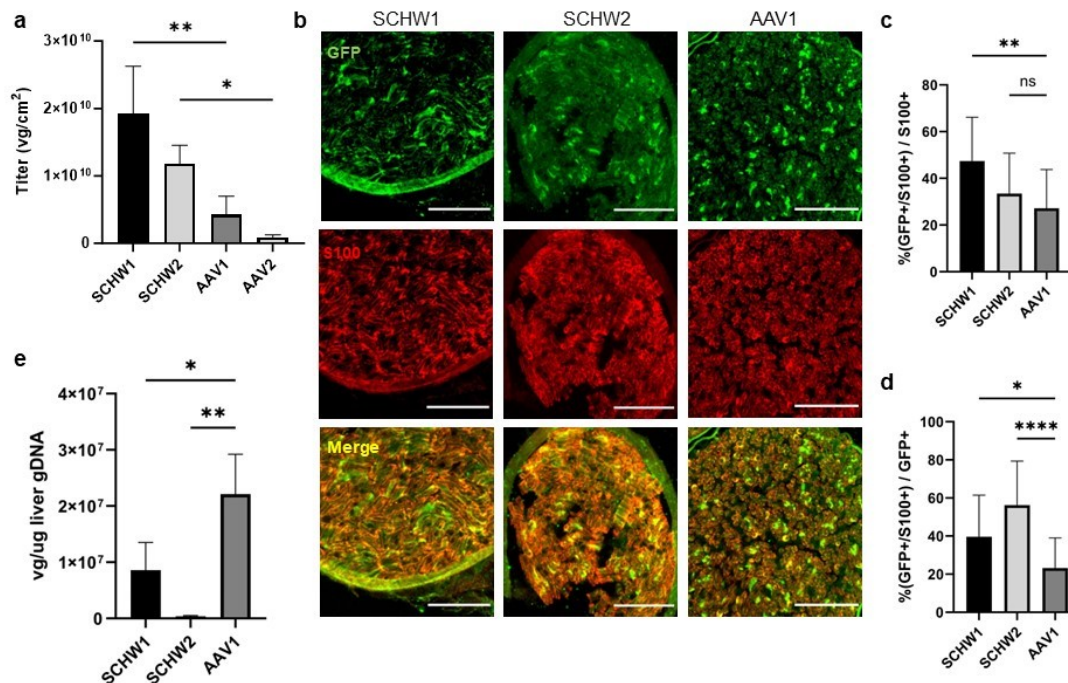


Figure 2.5: Characterization of mSC-tropic capsids. (a) Packaging titers of SCHW1 and SCHW2 compared to AAV1 and AAV2. Data are normalized to the cell culture plate surface area and presented as mean $\pm$ SD, $n = 3$. (b) Representative images of the transduction of SCHW1 and SCHW2 in mouse sciatic nerve. Cryosections were stained for GFP (green) and S100 (red). Scale bars indicate 100 μ m. (c and d) Quantification of c) transduction efficiency and d) Schwann cell specificity in mouse sciatic nerves compared to AAV1. Data are presented as mean $\pm$ SD, $n = 3$. (e) Biodistribution of each variant to the liver, quantified by number of viral copies normalized to total gDNA concentration extracted. Data are presented as mean $\pm$ SD, $n = 3$. All statistical analyses were done with a one-way ANOVA with Tukey's multiple comparisons test (* $p < 0.05$, ** $p < 0.01$, **** $p < 0.0001$).

2.4 Discussion

There have been hundreds of clinical trials investigating AAV-based gene therapies towards various diseases, including neurological, ocular, cardiovascular, and metabolic diseases. While showing great potential for treating a wide range of diseases, effective AAV-based delivery has seen little success in the peripheral nervous system, specifically towards Schwann cells. Schwann cells are associated with several PNS-associated diseases and represent an important target for AAV gene therapy.

In this work, we evolved highly-diverse AAV libraries to identify novel capsids with enhanced Schwann cell transduction. Alternating rounds of mouse *in vivo* and human *in vitro* selection were conducted to enrich for capsids with improved delivery capabilities in these two species. A Cre-dependent selection strategy was used during the *in vivo* rounds to isolate Schwann cell-transducing variants. Because double-stranded genome synthesis is required for

Cre recombination to occur, utilizing this stringent strategy ensures the isolated Cre-inverted variants have successfully overcome intracellular trafficking hurdles to the nucleus, necessary for transgene expression. Human telomerase reverse transcriptase (hTERT)-immortalized Schwann cells were used for the *in vitro* rounds of selection. This cell line has been shown to remain physiologically similar to the original primary Schwann cell, allowing for the selection of human Schwann cell-targeting variants[84].

After three alternating rounds of selection, variants C1 and C6 were identified and shown to demonstrate efficient transduction of hSCs and high resistance to human IVIG neutralization compared to natural serotypes. Because pure hSC cultures were used in the selection process, it would be beneficial to evaluate the transduction efficiency of C1 and C6 in other PNS cell types, such as human fibroblasts, sensory neurons, and mast cells, for future study. In addition, assessment of transduction in human systems that recapitulate the three-dimensional environment of the peripheral nervous system, such as human nerve segments, will be critical for evaluating the delivery efficacy in human Schwann cells. While more research is needed in evaluating these variants in human tissue, the significant improvements in IVIG resistance make C1 and C6 attractive vectors for further development.

It is interesting to note the high degree of similarity between the sequences of C1 and AAV2 yet varying phenotypes. C1 differs from AAV2 by the first 183 amino acids, which align to AAV6. This region is within the VP1/VP2u region, implying a role in endosomal escape and nuclear localization[86] and consequently, differences in transduction efficiency. C1 also differs from AAV2 by three single amino acid mutations within the VP3 region: Q598L, I648V, and V708I. A study investigating human polyclonal antibody neutralization of the AAV9 capsid visualized prevalent Fab binding to the 2-fold region of the AAV9 capsid using cryo-EM[87]. Mapping of the three amino acid point mutations of C1 (**Figure A.3**) shows mutations in the 2-fold and 3-fold axes, suggesting these mutations may play a role in antibody resistance, given the conserved general surface features of AAV capsids. In addition, both C1 and C6 retain sites for heparan sulfate proteoglycan (HSPG) binding, but C6, despite having blocks of AAV9, does not have the required N470 or W503 for galactose binding, suggesting that both variants rely on HSPG for cellular entry.

Unexpectedly, despite rounds of selection in both human *in vitro* and mouse *in vivo* models, C1 and C6 did not demonstrate improved transduction in our mouse *in vivo* model, suggesting key species differences in our model systems. Though C1 and C6 demonstrated high levels of transduction in hSCs, we re-diversified our variants with the insertion of a random peptide library and performed additional rounds of selection to improve cross-species translatability. Translatability between murine and human systems offers both clinical relevance for therapeutic development, while also allowing for further preclinical validation of potential therapeutics in relevant animal models. For this round of evolution, we implemented a transcript-based selection strategy in our *in vitro* rounds to select for capsids that successfully express their genetic cargo. After two rounds of selection, variants SCHW1 and SCHW2 were identified and showed significant improvements in transduction and specificity of Schwann cells within the mouse sciatic nerve over AAV1 after systemic administration. They also retained their ability to transduce hSCs, though at lower efficiencies compared to

wild-type AAVs, again raising the question of species differences.

Despite selection by infectivity of both hSCs *in vitro* and mSCs *in vivo*, both rounds of directed evolution yielded variants with enhanced transduction towards one system, but not both. It appears that selective pressures in each model system are significantly different and achieving efficacy in both systems is difficult. Several species-related factors may play a role in this differing behavior, including cell surface receptors, intracellular trafficking, and transcriptional and epigenetic regulation. Differences in transduction efficiencies of wild-type AAVs in rat and human SC cultures and nerve segments suggests differing cell surface and biological properties between human and rodent SCs[79]. It has also been shown that terminal Schwann cells at the neuromuscular junction differ in both morphology and synaptic relationships between mouse and human[88]. In addition, nerve regeneration is more accelerated in mice compared to human in response to peripheral nerve damage, due to faster SC injury responses in mice[89]. However, we cannot exclude the possibility that comparing *in vitro* vs *in vivo* validations may contribute to the difference in behavior, not solely due to biological species differences[79].

While the conducted rounds of selection applied a strong positive selective pressure for improved transduction of Schwann cells, assessment of off-targeting variants was not included during the evolution process. It is interesting to note how our evolved variants SCHW1 and SCHW2 demonstrate significantly higher specificity towards Schwann cell *in vivo* than AAV1 and AAV2. While Schwann cell-specific promoters can be used to improve cell-specific expression of a therapeutic transgene, these promoters are often long, minimizing space for the transgene, and not as strong as ubiquitous promoters like CAG and CMV[90]. Thus, it is advantageous for AAV capsids to exhibit target cell specificity to minimize off-target effects and lower the dosages necessary to achieve a therapeutic effect.

Another advantage of our *in vivo* variants is decreased targeting of the liver. A major barrier in the efficacy of AAV-based gene therapy is its safety profile. Toxicities of the liver have been reported due to high dosages of systemically-administered AAV. In a study for X-linked myotubular myopathy (XLMTM), a fatal neuromuscular condition involving mutations in the *MTM1* gene, three patients died several weeks after being administered a high AAV dosage of 3.5×10^{14} vg/kg due to severe hepatotoxicity. All three patients had pre-existing liver conditions and a lack of response to immunosuppressants[67]. Even after lowering the dosage to 1.3×10^{14} vg/kg, a fourth patient died, showing cholestatic liver failure[68]. Thus, being able to administer lower dosages of AAV and diminish transduction of the liver may alleviate potential liver-related adverse events. SCHW1 and SCHW2 both demonstrate liver detargeting properties, with SCHW2 showing a 57-fold decrease in liver transduction. In combination with the increased transduction efficiency and specificity towards Schwann cells *in vivo*, SCHW1 and SCHW2 may require lower dosages for potential therapeutic benefits, increasing the safety profile of these vectors.

During the course of this study, a similar study also evolved novel variants for enhanced transduction, instead using primary human Schwann cell cultures for rounds of selection, reiterating the need for better delivery vehicles towards the PNS[91]. Their AAV2-peptide insertion variants show high delivery efficacy to various human Schwann cell cultures and

human nerve segments when administered locally, showing great potential for therapeutic development. However, delivery efficiency diminished and tropism appeared to shift when injected into the sciatic nerve of mice. This highlights the difficulty of cross-species translatability and emphasizes the need for relevant model systems. Validation in mouse sciatic nerve also revealed lower overall transduction of nerve tissue, with one variant shifting tropism towards axons. In addition, many PNS-associated diseases, like neurofibromatosis type 1, are spread widely throughout the body and necessitate widespread therapeutic delivery. Our variants SCHW1 and SCHW2 demonstrate efficient *in vivo* delivery when administered systemically, overcoming immune and physical barriers that vectors administered locally may not be able to overcome.

In summary, we have generated a toolbox of novel AAV vectors for the development of *in vivo* and human-relevant gene therapy strategies targeting Schwann cell-associated diseases. While we had anticipated identifying effective delivery vectors for both mouse and human Schwann cell delivery, cross-species differences created trade-offs in delivery between the two model systems, a hurdle other studies have noted[92, 93, 52]. Instead, we engineered two groups of variants: C1 and C6 with enhanced hSC transduction and high resistance to neutralization by human IVIG over AAV1 and AAV2, and SCHW1 and SCHW2 with high levels of transduction and specificity towards mSCs in the sciatic nerve while minimizing transduction of liver tissue. In addition, all variants packaged at similar or increased titers compared to parental serotypes, highlighting the manufacturability of these vectors and increasing their translational value. Given the many improvements of these vectors over natural AAV serotypes, these engineered variants offer new avenues for the study and development of Schwann cell-associated therapies.

2.5 Materials and Methods

Cell culture conditions

HEK293T cells (ATCC, CRL-3216) were grown in DMEM (Corning), supplemented with 10% fetal bovine serum (Invitrogen) and 1% Anti-antibiotics (Invitrogen) at 37°C and 5% CO₂. hTERT ipn02.3 2λ (ATCC, CRL-3392) were used as our hSCs for *in vitro* rounds of selections and validation assays. These cells were originally isolated from a female patient with no disease indication and immortalized through exogenous expression of human telomerase reverse transcriptase (hTERT) and murine cyclin-dependent kinase (mCdk4). hSCs were cultured in DMEM (Corning), supplemented with 10% fetal bovine serum (Invitrogen) and 1X GlutaMAX (Thermo Fisher Scientific). The permissive 2v6.11 cell line (Cytion, 305147), derived from HEK293, allows for the inducible expression of the E4 34K protein, enabling higher assay sensitivities when evaluating neutralization resistance. 2v6.11 cells were grown in EMEM (ATCC) supplemented with 10% fetal bovine serum (Invitrogen).

AAV vector production

A range of previously generated plasmid libraries, including libraries generated through capsid shuffling, peptide insertions, and ancestral reconstruction, were packaged in HEK293T cells. AAV vectors were produced by triple transient transfection with polyethylenimine (PEI) in 15cm plates of HEK293T cells. For library production, 70ng library plasmid, 5ug pRepHelper, 13.5ug pHelper, and 9ug pBluescript were added per 15cm plate. For recombinant vector production, 6ug transgene plasmid, 12ug pHelper, and 10ug rep-cap plasmid were added per 15cm plate. Seventy-two hours after transfection, viral particles were harvested, purified by iodixanol density centrifugation, and buffer exchanged into PBS with 0.001% Tween 20 by Amicon filtration. DNase-resistant viral genomic titers were measured by real-time qPCR using a Bio-Rad iCycler (Bio-Rad, Hercules, CA, USA).

In vivo selections

DhhCre mice (JAX 012929) were a generous gift from David Kirsch's lab at Duke University. Animal protocols were approved by the Animal Care and Use Committee at the University of California, Berkeley. For rounds of *in vivo* selection, eight-week-old DhhCre mice were injected via tail-vein with 1×10^{11} - 1×10^{12} vg/mouse. Mice were sacrificed two weeks after injection, and sciatic nerves were harvested. Sciatic nerves were then sliced into small pieces using a razor blade and enzymatically digested in Hirt lysis buffer with proteinase K (New England Biolabs) and RNase A (Thermo Fisher) at 55°C for 3-4 hours. Viral DNA was then extracted as described previously[94]. For PCR amplification of both inverted and non-inverted *cap*, primers Cap_longFWD and Cap_longRev were used (**Table A.1**). For amplification of inverted *cap* only, a nested PCR was performed after amplification of all *cap* using primers Cap_InvF3 and Cap_Rev2 (**Table A.1**). *Cap* genes were then digested with *HindIII* and *NotI* (New England Biolabs) and cloned into pSub2FloxCap for additional rounds of selection or pXX2 for validation assays.

In vitro selections in hSCs

For *in vitro* selections, hSCs were transduced with AAV libraries at an MOI of 5000. After three days, cells were rinsed with PBS (Gibco), dissociated with trypsin (Corning), pelleted at 200g for 4 min, and either viral DNA or RNA was extracted. Viral DNA was harvested by Hirt extraction described previously, and *cap* genes were amplified with primers Cap_NSF4 and Cap_Rev4. RNA was extracted from the cell pellet with the RNA Plus Mini Kit (Qiagen, 74134) according to manufacturer's protocol. cDNA was synthesized using iScript™ Select cDNA Synthesis Kit (Bio-Rad, 1708898), and *cap* genes were amplified using primers Cap_cDNA_FWD and Cap_cDNA_REV (**Table A.1**). *Cap* genes amplified from either viral DNA or RNA extraction were then digested with *HindIII* and *NotI* and cloned back into pSub2FloxCap.

Construction of GFP-barcoded variants and sequencing analysis

Unique 16-bp DNA barcodes were cloned downstream of the doubly-floxed GFP in the plasmid pAAV-FLEX-GFP (Addgene, 28304). Individual variants were packaged with their corresponding GFP barcode, mixed at an equimolar ratio, and systemically injected into DhhCre mice at 9×10^{11} vg/ mouse ($n = 5$). Two weeks after injection, viral DNA was extracted from sciatic nerves and barcodes from reverted GFP genes were amplified. The pre-selected (before injection) and post-selected (after injection) barcode pools were submitted for sequencing on the MiSeq v3 (UC Berkeley Genomics Sequencing Laboratory) after library prep. After filtering out low-quality reads, barcode counts for pool were quantified, and enrichment scores for each AAV were calculated by normalizing the percentage of each barcode after tissue harvest to the percentage of each barcode in the pre-selected library.

In vivo variant characterization and immunohistochemistry

For individual variant validation, variants packaged with scCAG-GFP were injected in wild-type C57BL/6J mice (Jackson Laboratories 000664) via tail-vein injection. After three weeks, mice were sacrificed, and harvested sciatic nerves were fixed in 10% neutral-buffered formalin (Sigma-Aldrich) overnight, followed by a 30 min incubation in dH₂O and storage in 70% ethanol. Livers were also harvested from each mouse for biodistribution analysis.

To quantify GFP expression in the sciatic tissue, nerves were cryoprotected in 10% and 30% w/v sucrose gradients in PBS before being frozen in optimal cutting temperature (OCT) compound (Tissue-Tek; Sakura Finetek USA) overnight. Frozen nerves were then cryosectioned into 12um slices and mounted onto glass coverslips (VWR). Before staining, tissue was permeabilized with 0.5% Triton-X in PBS for 10 minutes and blocked with 5% donkey serum (Sigma-Aldrich) for one hour at room temperature. Primary antibody staining was done at 4°C overnight using the following antibodies: rabbit anti-S100 beta (Abcam, ab52642) and goat anti-GFP (Abcam, ab6673). After washing three times with PBS, tissue was then stained with the corresponding secondary antibodies for one hour and DAPI for 10 minutes at room temperature. Slides were mounted with ProLong™ Gold Antifade Mountant (Thermo Fisher) and imaged on the BioTek Lionheart FX (Agilent).

In vitro characterization in hSCs and immunocytochemistry

For individual variant characterization *in vitro*, AAVs packaged with GFP were added to hSCs at an MOI of 5000 unless otherwise specified. Three days after transduction, cells were fixed in 4% paraformaldehyde (Thermo Fisher) in PBS for 20 minutes at room temperature. After washing three times in PBS, cells were permeabilized in 0.5% Triton-X in PBS for 10 minutes and blocked in 5% donkey serum (Sigma-Aldrich) for one hour at room temperature. Primary antibody staining was performed at 4°C overnight with rabbit anti-GFP (Abcam, ab6556). Cells were then washed three times with PBS and stained with a secondary anti-

body for one hour and DAPI for 10 minutes at room temperature. Images were taken on the ImageXpress Micro (IXM) imaging system.

Design of chimeric-based peptide insertion library

For the generation of our peptide insertion libraries from chimeric variants C1, C5, and C6, a stop codon was first inserted within each capsid sequence first to ensure that any variant without a 7mer insertion would not be able to successfully package. The stop codon (nucleotide sequence TGA) was inserted after amino acid G586 of C1, G588 of C6, and D590 of C5 using the Q5 Site-Directed Mutagenesis Kit (New England Biolabs). A variable 7mer insertion region flanked by amino acid linkers (LA-A for C1 and C6 and TG-GLS for C5) was inserted into the capsid using splice overlap extension PCR. Briefly, left and right fragments were amplified using primer pair seq_fwd/seq_X_rev and 7mer_X_fwd/7mer_rev (Invitrogen) respectively, where X represents the corresponding linker. The random 7mer region was incorporated into the 5' end of the right fragment. A second PCR using equimolar amounts of the purified left and right fragments (250 ng total DNA) and primer pair 7mer_fwd/7mer_rev (Invitrogen) was then run to generate full length *cap* genes with the variable peptide region (**Table A.1**). The generated library was digested with *HindIII* and *NotI* (New England Biolabs), ligated into the pSub2FloxCap backbone, and electroporated into electrocompetent *E. coli* (Thermo Fisher Scientific, 18290015) for plasmid library production.

Neutralization assay

To quantify antibody evasion, 2v6.11 cells were seeded into a 96-well plate (Corning). After 16 hours, the media was replaced with fresh media supplemented with ponasterone A (Invitrogen, H10101) at a final concentration of 1 μ g/ml to induce the E4 gene. C1, C6, AAV1, and AAV2 were packaged with GFP and incubated with serial dilutions of human IVIG (Gamaguard) for two hours at 37°C. The neutralized samples were then added to the 2v6.11 cells at an MOI of 10,000 ($n = 3$), 24 hours after supplementation with ponasterone A. Cells were fixed after three days, stained for GFP and DAPI, and imaged as previously mentioned. The percentage of cells transduced and total intensity were measured and normalized to GFP expression in the absence of IVIG.

Biodistribution assay

Three weeks after systemic injection of variants packaged with scCAG-GFP into C57BL/6J mice, livers were harvested and sliced into small pieces using a razor blade. gDNA was collected using the DNeasy Blood and Tissue Kit (Qiagen, 69504) following manufacturer's instructions. Vector copy number was then quantified using real-time qPCR using a Bio-Rad iCycler (Bio-Rad, Hercules, CA, USA) with primers GFP_FWD and GFP_REV (**Table A.1**) and normalized to the extracted concentration.

2.6 Acknowledgements

Thank you to Dr. Hyuncheol Lee for performing the mouse tail-vein injections. All images were taken at the QB3 Cell and Tissue Analysis Facility at University of California, Berkeley (UC Berkeley), which provided access to the BioTek Lionheart FX (Agilent) and ImageXpress Micro (IXM) imaging systems. We thank Mary West for microscope training and support. We also thank Dr. Lixia Luo in Prof. David Kirsch's lab at Duke University for providing the DhhCre mice and assisting with animal protocols. Thank you to Dr. Madeleine Sitton in Prof. Charles Gersbach's lab at Duke University and Dr. Kecheng Lei in Prof. Nicholas Boulis's lab at Emory University for experimental collaboration. This work was supported by a National Science Foundation Graduate Research Fellowship Program (NSF GRFP) fellowship and the Gilbert Family Foundation Gene Therapy Initiative.

Chapter 3

Long-read sequencing of Diverse DNA libraries through Barcode-Labeled Short Reads (BLaSR)

3.1 Chapter Abstract

A major shortcoming of short-read sequencing technologies – the inability to sequence long stretches of DNA – has been addressed by newer, long-read technologies, such as Oxford Nanopore and PacBio SMRT Sequencing. These sequencers can characterize longer read lengths ranging from a few thousand to several thousands of base pairs. However, limitations still exist. Long-read sequencing techniques, particularly single-molecule long-read technologies, suffer from high error rates, ranging up to 10%, which cannot capture single-nucleotide variations in highly diverse DNA libraries. In addition, current methodologies are often labor-intensive and require large amounts of input material. There is a need for accurate profiling of highly-diverse DNA libraries with minimal error that existing long-read sequencing technologies cannot achieve. Here, we demonstrate a new synthetic long-read sequencing methodology we term Barcode-Labeled Short Reads (BLaSR) that can sequence a diverse AAV *cap* library using a randomly-barcoded Tn5 transposon system. By leveraging transposition within a bacterial system, we harness the Tn5 transposon to insert random barcodes throughout the AAV *cap* library such that each *cap* variant contains insertions with a unique, identifiable barcode. This allows for the generation of barcode-linked next-generation sequencing short reads in which reads with the same barcode can be traced to the same AAV variant and thus be assembled into the longer *cap* variant sequence. We have engineered a robust pipeline to ensure highly efficient barcoded-Tn5 insertion into the AAV genome and have assembled thousands of 2.2kb AAV *cap* variants with an average 20x read depth with a 0.1% error rate, with a path towards considerably larger numbers of sequenced *cap* variants per run through further optimizations.

3.2 Introduction

Recombinant adeno-associated viruses (AAVs) are an effective gene delivery vehicle and offer several advantages over other viral and non-viral vectors for clinical gene therapy. Their ability to promote high expression of the therapeutic cargo and lack of pathogenicity make AAVs an attractive option[70]. In addition, AAV therapies have achieved success in numerous clinical trials, with six to date having received US Food and Drug Administration (FDA) approvals and others in preclinical and clinical development (fda.gov).

Despite the benefits of AAV-based vectors for gene therapy, it has become crucial to modify naturally-occurring AAVs to enhance their efficacy for most tissue and disease targets. Natural AAVs exhibit broad tissue tropism, contributing to diminished delivery to target cell types and potential off-target effects[95]. Neutralization of the AAV upon administration is also another hurdle. With 30-60% of the human population previously exposed to naturally occurring AAVs, neutralizing antibodies (NABs) against AAV can substantially reduce the efficacy of AAV-based treatments and greatly hinder the potential for re-administration[96]. To address this issue, it is essential to engineer the AAV capsid to enhance its delivery efficiency and specificity while also evading the body's immune recognition.

Directed evolution is a high-throughput approach for creating and identifying novel AAV capsid variants with enhanced properties, such as infectivity, targeted delivery, tissue spread, or immune evasion[40, 50, 48, 37]. Mimicking natural evolution, AAV capsid directed evolution is a process that involves genetically diversifying the AAV *cap* gene to generate a large, highly-diverse library of mutant AAV capsids, which are then subjected to iterative rounds of a selective pressure to identify variants with improved properties. The ability to deeply sequence these libraries offers insight into the evolutionary process, as well enable earlier identification of effective variants; however, accurate characterization of these diversified AAV libraries is not possible due to limitations in current sequencing technologies.

The integration of short read next-generation sequencing (NGS) into directed evolution processes has allowed for data-guided engineering of AAVs, but with a major restriction[97]. NGS allows for comprehensive sequencing of highly-diverse libraries with a relatively low error rate and can capture variant mutations in which mutations are restricted to a specific region or within the read length of NGS sequencers. However, a major issue with short-read technologies is that libraries that have been diversified via methods such as error-prone PCR, random shuffling, or ancestral reconstruction[37, 49, 39, 55], in which diversity is spread through the full *cap* open reading frame, are unable to be correctly sequenced as individual clones. Long read sequencing offers this potential, but at substantially lower read depth and high error rates (1-10%). Significant improvements have been made in lowering error rates[98]; still, large amounts of input DNA are required, data yield per run is lower than short-read technologies, and costs per data remain higher than short-read methodologies[99].

Here, we have made progress in addressing the limitations of both long- and short-read sequencing methodologies to profile a diverse AAV *cap* (~2.2kb) library. We have developed a synthetic long-read methodology Barcode-Labeled Short Reads (BLaSR) that aims to address the pitfalls of existing sequencing methodologies and have applied it here

to mutagenized adeno-associated virus (AAV) *cap* libraries that have been diversified across the entire *cap* open reading frame. Moreover, the use of BLaSR is not limited to AAV *cap* libraries; rather, this long-read sequencing methodology can be applied to other diverse DNA libraries that cannot otherwise be accommodated by next-generation sequencing.

3.3 Results

Overview of BLaSR

Tn5 transposons are DNA elements that can change their genomic integration positions, or “jump” from one place in the genome to another, via a “cut and paste” mechanism[100, 101]. They are commonly used in molecular biology and genetics for their ability to randomly insert into DNA, unlike other transposons, such as the mariner transposon that inserts preferentially at TA dinucleotides[102], which makes them valuable tools for mutagenesis, mutant fitness assays, and the creation of insertion libraries[103, 104, 105, 106, 107, 108]. Additionally, Tn5 transposons are integral in DNA library preparation for sequencing purposes, where their transposase activity can simultaneously fragment DNA and attach sequencing adapters in a highly efficient and streamlined process, known as tagmentation[109]. Conventional tagmentation involves incubation of the Tn5 transposase with each adapter set (i.e the DNA to be “inserted”) and the target DNA to be tagmented in a sample well. However, for every unique DNA sequence to be inserted, a separate reaction well is required. A crucial aspect of BLaSR is the establishment of a link between a unique Tn5 insertion and an AAV variant, which is nearly impossible to scale up for highly-diverse libraries that contain millions of variants with current tagmentation strategies. Here, we utilize bacterial cells as the “reaction wells” to tagment AAV plasmid variants. A Tn5 transposon-containing plasmid, carrying short, random barcodes, is co-transformed into *E. coli* with an AAV plasmid library, such that cells receive one copy of a barcoded Tn5 plasmid and one copy of an AAV variant. During the culture period, the target AAV plasmid replicates inside the bacterium, and when the bacterium divides, it gives many copies of the same AAV plasmid. Simultaneously, Tn5 insertion of the barcode is happening into the AAV plasmids, at different sites for each plasmid, so that the end result is an originally-clonal population of AAV plasmids with Tn5 insertions randomly distributed throughout them. The AAV variant is then sequenced using next-generation sequencing, beginning from the inserted barcode. Thus, we harness random Tn5 insertions to barcode Illumina short reads so that each short read containing the same barcode can be traced back to the same AAV variant and assembled to form the full *cap* sequence.

The BLaSR methodology is summarized in **Figure 3.1a**. Briefly, a library of Tn5 transposon-containing plasmids with random barcode sequences were co-transformed into *E. coli* with an AAV *cap* plasmid library to generate barcoded AAV plasmids. Co-transformation was optimized to maximize the number of bacterial cells that received a copy of each plasmid while minimizing the chance of a bacterial cell receiving two plasmids from the same library

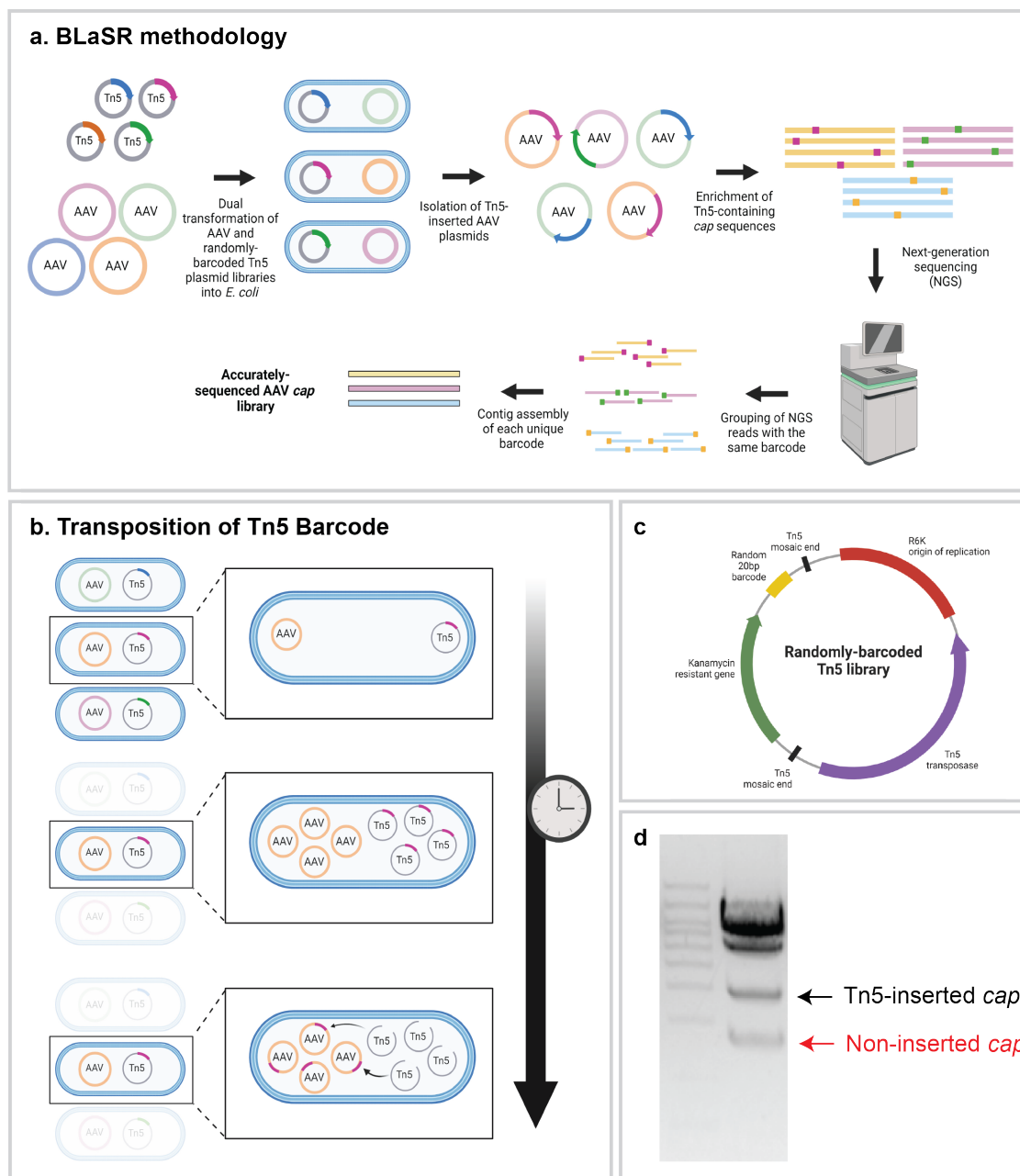


Figure 3.1: Overview of BLaSR sequencing methodology. (a) Schematic for BLaSR process for the assembly of diverse AAV *cap* sequences. (b) Representation of Tn5 transposition into the AAV genome. (c) Design of randomly-barcoded Tn5 plasmid library. (d) Restriction enzyme digest confirming Tn5 insertion into the AAV *cap* gene. Graphics created with BioRender.com.

(Figure B.1). The transformed bacteria were then cultured in dual-resistant media, allowing for replication and transposition of Tn5 barcodes into the AAV plasmid (Figure 3.1b).

Barcoded AAV *cap* genes were selectively extracted and sequenced using next-generation sequencing. Short reads were then bioinformatically grouped according to barcode and assembled to generate full-length *cap* sequences.

Generation of a barcoded AAV *cap* library using randomly-barcoded transposons

The BLaSR transposon plasmid library was modeled after the plasmid pBAM1[110]. In short, the backbone contains an R6K gamma origin of replication and a hyperactive Tn5 transposase gene (E54K, M56A, and L372P modifications to the wild-type sequence) upstream of the transposable element. The transposable element contains a kanamycin selection marker (KAN) to allow for dual selection with the ampicillin-resistant AAV plasmid library, flanked by two 19-bp hyperactive inverted terminal repeat regions, termed mosaic ends (MEs). Unique 20-bp barcodes were cloned downstream of KAN and within the MEs (**Figure 3.1c**).

For validation of insertion, BLaSR sequencing was first applied to the wild-type AAV2 *cap* sequence. The AAV2 *cap* sequence was cloned into the pSub2 backbone, which contains a pBR322 origin of replication and an ampicillin resistance marker (AMP), allowing for co-replication with a Tn5 plasmid within the same bacterial cell. After two rounds of transformation to enrich for Tn5-inserted AAV plasmids, Tn5 transposition into the AAV *cap* was validated by digesting the Tn5-inserted AAV plasmid library with two restriction enzymes *HindIII* and *NotI* that flank the *cap* gene and analyzing the fragments on a gel (**Figure 3.1d**). Additional experimental optimizations were performed to maximize the number of Tn5 insertions into the AAV genome and characterize locations of Tn5 insertions to develop the complete experimental pipeline (**Figure B.1**).

Characterization of BLaSR applied to AAV *cap*

Before applying BLaSR to a diverse AAV *cap* library, we first investigated the nature of Tn5 insertion sites into AAV *cap*. In order for BLaSR to be most effective, the Tn5 barcode should insert in an unbiased manner and ideally uniformly across the AAV capsid. To assess this, we applied BLaSR to the wild type AAV2 capsid gene, a known sequence, and mapped all reads to the AAV2 *cap* sequence to identify the exact locations of Tn5 insertions with high precision. BLaSR utilizes paired-end Illumina sequencing to sequence both ends of Tn5-inserted *cap* amplicons (average amplicon length of 500bp), with Read 2 containing the Tn5 barcode and Read 1 containing mostly AAV DNA (**Figure 3.2a**). Mapping of Read 2 sequences (containing the Tn5 barcode insertion location) can pinpoint the insertion sites. With the heterogeneity of NGS amplicon length, Read 1 provides additional insight into sequencing coverage of *cap* (**Figure 3.2a**). Because BLaSR requires paired-end sequencing, both Read 1 and Read 2 (excluding the Tn5-associated region) were mapped to AAV2. Interestingly, a bias in coverage was noted at both the 5' and 3' ends of *cap*, though a relatively unbiased coverage throughout the rest of *cap* (**Figure 3.2b-c**).

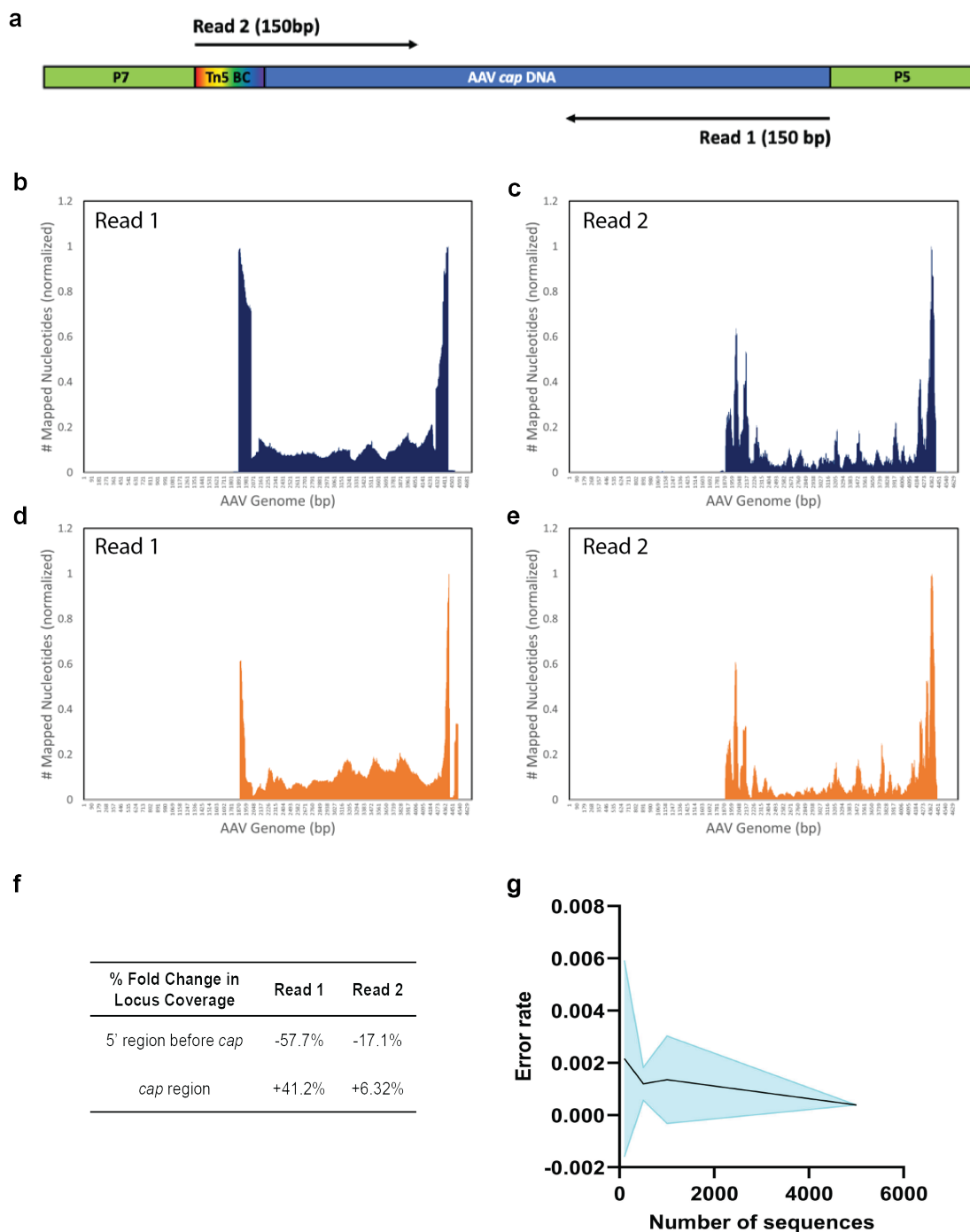


Figure 3.2: Characterization of BLaSR. (a) Schematic representing the NGS amplicon library. (b) NGS analysis of Read 1 coverage across AAV template before optimization of Tn5 bias. (c) NGS analysis of Read 2 coverage across AAV template before optimization of Tn5 bias. (d) NGS analysis of Read 1 coverage across AAV template after optimization of Tn5 bias. (e) NGS analysis of Read 2 coverage across AAV template after optimization of Tn5. (f) Locus sequence coverage of 5' region before *cap* and *cap* region before and after optimization of Tn5 bias. (g) Error rate of the BLaSR methodology, as applied to known sequence AAV2 *cap*.

To further assess bias, a restriction enzyme digest was performed. After extracting the Tn5-inserted *cap* fragments from the Tn5-inserted AAV plasmid library by restriction enzyme digest, a Tn5-specific restriction enzyme (*SpeI*) was then used to digest the *cap*-containing fragments. If the Tn5-barcode had in fact inserted uniformly, we would expect to see a smear after gel electrophoresis. Instead, our results correlated to our NGS results and revealed two dominant bands in addition to a smear, suggesting an insertion bias for the end regions of *cap* (**Figure B.2a-b**). The smear was gel-extracted, excluding the dominant bands, and submitted for NGS. Upon analysis of the mapped reads, a decreased coverage in the 5' end by 57% in Read 1 and 20% in Read 2 and increased coverage in the remainder of *cap* was observed, suggesting removal of the dominant bands improved coverage of the desired *cap* region, though bias towards the end regions was still observed (**Figure 3.2d-f**). It is also known that the Tn5 transposition creates a 9-bp duplicate of the target DNA insertion sequence[111, 112], so sequence logos, used to show motifs or patterns in biological sequences, were constructed before and after removal of the dominant bands (**Figure B.3**). Interestingly, there appears to be a slight preference for G/C bases at the first and last position of the 9bp target sequence, corresponding to previous studies[113].

Upon further analysis, it was noticed that if, in fact, there was a bias of insertion at both ends of *cap*, removal of the dominant bands after the *SpeI* digestion might leave behind the smaller (>300 bp), corresponding fragments that may still contribute to bias in coverage and thus less efficient coverage of the *cap* (**Figure B.2b - highlighted in green**). It was observed upon assessment of the reads that mapped to either end of *cap* that before removal of the dominant bands, reads that aligned to either end of *cap* were of a mix of either orientation (FWD and REV). However, upon removal of the dominant bands, reads that aligned to the 5' end were entirely of the reverse (3' to 5') orientation, while reads that mapped to the 3' end were mostly of the forward (5' to 3') orientation, hinting that not all of the insertion bias had been experimentally removed. Thus, a DNA cleanup step that removed fragments under 300bp after library preparation was implemented before Illumina sequencing to further reduce bias (data not shown).

Another important aspect of BLaSR to characterize is error rate, a major limitation in current single-molecule long-read sequencing methodologies. Due to the hurdle of long-read sequencing, however, the shuffled AAV *cap* library has yet to be fully sequenced, and thus lacks a benchmark for comparison. Accordingly, we applied BLaSR to wild-type AAV2, as mentioned previously, with Illumina's MiSeq Nano, which outputs approximately 1×10^6 reads. To quantify the error rate, we randomly sampled varying numbers of sequences and assembled contigs, which were then aligned to AAV2 *cap*. The number of point mutations was counted and then normalized to the length of the assembled contig to calculate error rate. We plotted the error rate versus the number of sequences used to assemble the contig (**Figure 3.2g**). Our calculated error rate ranged from 0.1% with 1000 sequences to 0.2% with 100 sequences, superior to existing long-read sequencing technologies.

Sequencing of a shuffled AAV capsid library using BLaSR

To assess efficacy of this synthetic long-read methodology, we next applied BLaSR to a diverse, chimeric, randomly-shuffled capsid library, described previously[39]. We applied BLaSR to approximately 45ng AAV DNA (about 5×10^9 AAV plasmids), used a modified GUIDE-seq protocol for NGS library preparation of the barcoded *cap* library, removed sequences under 300bp, and sequenced on the NovaSeq6000 platform.

An overview of data analysis is provided in **Figure 3.3a**. For analysis of the sequenced data, reads were first filtered by base pair quality scores and length, with about 2 billion reads remaining post-filtering. After filtering, we first assessed the number of unique Tn5 barcodes. Given that the number of unique barcodes is the theoretical number of variants that can be sequenced, it is imperative that this number is greater than the number of variants we aim to sequence. We calculated 252 million unique barcodes, which is 10 to 100-fold higher than the maximum size of the shuffled library, based on the number of independent bacterial clones from generation of this library. We next sought to quantify the number of reads associated with a given Tn5 barcode (**Figure 3.3b-c**). Barcodes had varying numbers of reads, with 2.3×10^4 barcodes having more than 5000 sequences and about 3×10^5 barcodes containing 1000 or more sequences.

With this information, we next aimed to estimate the number of sequences needed to assemble a full *cap* sequence. For this, we randomly sampled 30 barcodes with varying numbers of associated sequences. Using Geneious Prime’s *de novo* assembly program, we manually assembled individual contigs and plotted the length of the assembled contig against the number of sequences used to assemble that contig (**Figure 3.3d**). From this, we estimated that approximately 3000 sequences are needed to assemble the full *cap* sequence, implying that BLaSR can sequence up to 7.1×10^4 *cap* sequences based on **Figures 3.3b-c**. However, further analysis is ongoing to comprehensively quantify the number of sequences that can be assembled (*see Future Directions*).

In addition, we also estimated the read depth of our assembled contigs. Read depth is a crucial parameter in long-read sequencing, and high read depth allows for consensus building and thus higher sequencing accuracy. It also improves the quality of the assembled sequences by ensuring that all regions of the *cap* sequence are sufficiently covered. From our sequenced shuffled library, we randomly selected three barcodes that had successfully assembled the full capsid sequence with 2000-3000 sequences and plotted the read coverage (**Figure 3.3e**). Read depth averaged approximately 20x over these variants.

3.4 Discussion and Future Steps

Long-read sequencing of diverse AAV *cap* libraries is an invaluable tool to characterize libraries that have yet to be fully sequenced - libraries in which mutations span the entire length of the *cap* gene - due to read length limitations, error rates, and read numbers of existing technologies. The ability to fully sequence diverse AAV *cap* libraries provides infor-

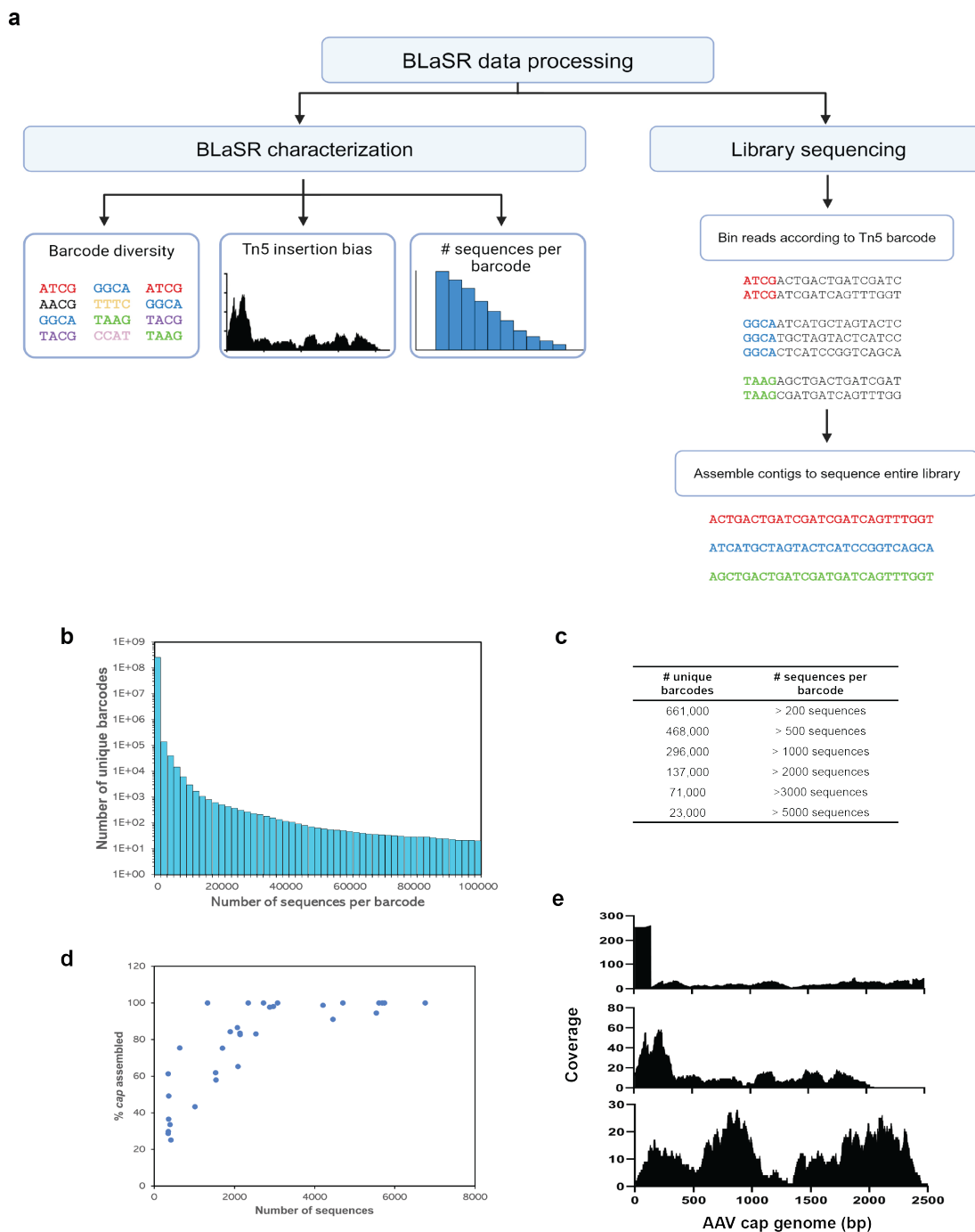


Figure 3.3: Sequencing a shuffled AAV library with BLaSR. (a) Overview of BLaSR data processing pipeline and analyses. Graphic created with BioRender.com. (b) Analysis of number of unique barcodes identified versus the number of sequences linked with each unique barcode. (c) Number of unique barcodes that have a particular number of sequences per barcode. (d) A subset of AAV assembled *cap* contigs, indicating the number of sequences used to generate the assembly versus the percentage of the full *cap* sequence assembled. (e) Assembly of a small subset of AAV *cap* variants, demonstrating read depth across the AAV *cap* sequence.

mation into the diversity of the library, as well as insight into the selection process of directed evolution. By providing a comprehensive view of the *cap* gene variants, long-read sequencing facilitates the development of more effective AAV vectors for gene therapy, ensuring that the full spectrum of viral diversity and evolution is captured.

To that end, we have developed a synthetic long-read sequencing methodology termed BLaSR to enable full characterization of AAV *cap* libraries and move towards addressing the limitations of current sequencing technologies. Despite its read length limitations, Illumina is superior to long-read sequencing in cost, run time, amount of input material needed, and read accuracy (especially with strong read depth)[99]; thus, we aimed to harness Illumina’s technology to create a synthetic long-read sequencing method by inserting a barcode into each short read. Using a randomly-barcode Tn5 transposon, we tagged AAV *cap* variants and generated an NGS-compatible library such that each short read contained a 20bp barcode that could be mapped back to an AAV clone, allowing for the assembly of complete *cap* sequences. By first applying BLaSR to wild-type AAV2, we were able to quantify insertion bias and error rate of this methodology. Error rates were calculated to be 0.1-0.2%, depending on the number of sequences per barcode, which is significantly improved over Oxford Nanopore and marginally improved over PacBio’s Sequel II system[99]. In addition, while the Tn5 insertion demonstrated preference of insertion at the 5’ and 3’ ends of *cap*, we were able to reduce insertion bias at these regions and saw approximately 20x coverage of the AAV *cap*.

An imperative consideration in the development of BLaSR is the co-transformation of the barcoded-Tn5 library and AAV library into *E. coli* for transposition. For BLaSR to be effective, it is crucial that we maximize the number of cells that receive one copy of each plasmid to maximize the number of Tn5 barcodes inserted into the *cap* gene. By strategically designing the barcoded Tn5 plasmid to contain a different origin of replication (R6K) and selection marker (KAN) than our AAV library, our two plasmids can successfully co-replicate in a singular bacterial cell. Varying experimental conditions were also tested to maximize Tn5 insertion into the AAV plasmid. However, it is important to also minimize the number of bacterial cells that uptake more than one copy of either plasmid. Two Tn5 barcodes in the same cell as one AAV variant splits the number of Tn5 insertions - and thus sequencing reads - of the variant among two different barcodes, which may lead to not enough transpositions to sequence the full *cap*. Two AAV variants transformed into a bacterial cell with one Tn5 plasmid would cause two AAV variants to contain the same Tn5 barcode, leading to possible incorrect assembly of the AAV variant. Because of this issue, measures were incorporated to minimize this occurrence. First, low amounts of DNA (15ng total DNA per transformation) were used to co-transform the two libraries into *E. coli*. The probability of three plasmids transforming the same cell with this amount of DNA via electroporation is highly statistically unlikely. Tomoiaga et al. showed that electroporation of DH5alpha cells with three plasmids, at varying amounts of total DNA and differing in just their fluorescent marker, resulted in about zero percent of colonies being transformed with all three plasmids[114]. In addition, plasmids with the same origin of replication (e.g two AAV plasmids or two Tn5 plasmids) compete for the same replication machinery if taken up by the same cell, leading to one

plasmid “outcompeting” the other in replication[115, 116]. Upon co-transformation of the two plasmid libraries, we sampled individual bacterial colonies and used Sanger sequencing to assess whether individual cells had taken up two copies of either plasmid. Results indicated no or minimal occurrence of triple transformations based on the output of a singular sequence and base-pair quality of the read (data not shown).

Further work is needed and currently ongoing to comprehensively sequence the shuffled AAV library. While we have estimates of the number of variants that can be fully assembled, we are developing a pipeline to assemble contigs of every variant to provide insight into the composition of our shuffled library. In addition, we plan to utilize existing long-read methodologies, such as PacBio’s Sequel II system (8 million reads), for comparison. It will be interesting to compare the number of sequences, error rate, ease of assembly, and cost of both these sequencing technologies to understand the advantages and disadvantages of each.

We have also applied BLaSR to the shuffled library that has undergone a round of selection in adult human brain tissue to demonstrate its utility in direction evolution selections. AAV targeting of central nervous system (CNS) cell types offers great potential for treating CNS diseases, including Huntington’s and Parkinson’s disease[63]. Our shuffled AAV *cap* library was packaged and administered to three brain slices at an MOI of 1000. Three days later, gDNA was harvested from the bulk tissue, and *cap* was PCR amplified and cloned back into the pSub2 backbone. The selected library was sequenced with BLaSR via the NextSeq 2000, and the number of sequences per barcode were quantified (**Figure B.4**). Analysis is ongoing to fully sequence these variants and calculate the enrichment score for each variant to identify AAV capsids with improved delivery to adult human brain tissue.

While additional analysis is needed to fully characterize BLaSR’s capabilities, we are optimistic in its ability to accurately characterize AAV libraries that have yet to be comprehensively sequenced. BLaSR is a novel synthetic long-read sequencing methodology that may provide a useful alternative to capturing/characterizing diverse AAV libraries and has the potential to be used for other DNA libraries that may not be able to be sequenced using existing technologies.

3.5 Materials and Methods

Generation of randomly-barcoded Tn5 transposon vector

Full Tn5 transposon vector, without any barcodes, was synthesized by GenScript, modeled after the plasmid pBAM1[110]. The synthesized plasmid contained an R6K gamma origin of replication, a kanamycin resistance marker, two 19-bp mosaic ends, and the Tn5 transposase gene. Barcodes were generated via PCR using the primers Tn5_BC_FWD and Tn5_BC_REV (**Table B.1**). Barcodes were then subcloned into the synthesized Tn5 vector using the restriction sites *PpuMI* and *EcoRI*. The ligated reaction was electroporated into TransforMax™ EC100D™ pir-116 Electrocompetent *E. coli* (Lucigen) and midipreped (Invitrogen), resulting in the randomly-barcoded Tn5 vector library.

Strains and standard growth conditions

TransforMax™ EC100D™ pir-116 Electrocompetent *E. coli* (Lucigen) were cultured at 37°C and 225 rpm. ElectroMAX™ DH10B electrocompetent cells (Invitrogen) were also cultured 37°C and 225 rpm. Overnight cultures refer to a 14-hour incubation.

Transposon mutant AAV capsid library construction (BLaSR Methodology)

AAV and Tn5 plasmids were electroporated into TransforMax™ EC100D™ pir-116 Electrocompetent *E. coli* at 2.5 kV, 200Ω, 25 uF using a BioRad GenePulser II electroporator. A total of 15ng was electroporated, at a ratio of 10:1 AAV:Tn5. Transformed bacteria were then cultured overnight for 14 hours in LB media supplemented with kanamycin and ampicillin to enrich for cells that received both types of plasmids. After 14 hours, the culture was diluted 1:3 and cultured for another six hours at the same conditions. The midprepped library was then electroporated again into DH10B electrocompetent cells at 80ng per reaction and cultured in LB media supplemented with kanamycin and ampicillin for 14 hours to enrich for AAV plasmids with the Tn5 insertion. DH10B cells do not support the replication of plasmids containing an R6K origin of replication, so non-transposed Tn5 plasmids would not propagate through this transformation. In addition, any uninserted AAV plasmid would not propagate due to the lack of kanamycin resistance.

BLaSR NGS library preparation

From the barcoded AAV library, Tn5-inserted *cap* genes were first extracted by digestion with *HindIII* and *NotI*. The library was then prepped according to the GUIDE-seq protocol, described previously[117]. Briefly, the Tn5-inserted *cap* library was sheared to an average length of 500bp using the Covaris S2 sonicator. After a cleanup with AMPure XP SPRI beads (Beckman Coulter), Y-adapters were blunt-end ligated to the sheared library after end repair and A-tailing. A nested PCR was then performed to specifically amplify fragments with the Tn5-barcode and add the Illumina adapters. The prepped library was then sequenced on either the MiSeq Nano, NovaSeq 6000, or NextSeq 2000 (2 x 150bp paired end reads).

BLaSR data analysis

Raw data was first filtered based on base pair quality scores and PCR duplicates were then removed. With an average amplicon length of 500bp, a majority of the read pairs did not overlap in sequence; however, a small percentage of Read 1 sequences did read into the Tn5-associated region. To improve accuracy of the contig assemble pipeline, all Tn5-associated regions from Read 1 were subsequently trimmed. Tn5 barcodes were then split from the AAV-associated sequence of Read 2. All AAV sequences linked to the same Tn5 barcode

were then binned into one fasta file, and contig assembly was done manually with Geneious Prime. An automated contig assembly pipeline is currently being developed.

AAV vector production

The shuffled AAV *cap* plasmid library[39] was packaged in HEK293T cells. AAV vectors were produced by triple transient transfection with polyethylenimine (PEI) in 15cm plates of HEK293T cells. For library production, 70ng library plasmid, 5ug pRepHelper, 13.5ug pHelper, and 9ug pBluescript were added per 15cm plate. Seventy-two hours after transfection, viral particles were harvested, purified by iodixanol density centrifugation, and buffer exchanged into PBS with 0.001% Tween 20 by Amicon filtration. DNase-resistant viral genomic titers were measured by real-time qPCR using a Bio-Rad iCycler (Bio-Rad, Hercules, CA, USA).

Brain selection

Adult brain tissue was obtained, sliced into 300um sections, and cultured by Prof. Tom Nowakowski's lab at University of California, San Francisco (UCSF). The packaged shuffled AAV library was added dropwise to the culture media at an MOI of 1000 and incubated at 37°C for 72 hours. Tissue was then flash frozen for storage and transportation. Genomic DNA was then extracted using the DNeasy Blood and Tissue Kit (Qiagen, 69504) following manufacturer's instructions. *Cap* genes were then PCR amplified using primers Cap_NSF4 and Cap_Rev4 (**Table B.1**), digested with *HindIII* and *NotI* (New England Biolabs), and cloned into pSub2 for BLaSR sequencing. Post-selection library was sequenced on the NextSeq 2000 (2 x 150bp paired end reads).

Insertion bias and error rate

To quantify both insertion bias and error rate of this methodology, BLaSR was applied to the wild-type AAV2 genome and sequenced on the MiSeq Nano (2 x 150bp paired end reads). For assessing Tn5 insertion locations, the raw sequencing data was filtered based on base pair-quality score thresholding, and PCR duplicates were removed. For Read 1, Tn5-associated regions were removed in the case of shorter amplicons. For Read 2, AAV DNA immediately downstream of the Tn5 insertion was extracted. Both trimmed Read 1 and Read 2 sequences were then aligned to the wild-type genome template using Rsubread and visualized in Geneious Prime.

To quantify the error rate, raw data was processed similarly to analysis of insertion locations. Reads were filtered based on quality score, PCR duplicates were removed, and both Read 1 and Read 2 were trimmed of Tn5-associated sequences. Due to the limited number of reads (~one million), barcodes were not included in this analysis. Groups of sequences were randomly selected, ranging from 100 to 5000 sequences. Contigs were assembled and aligned to the wild-type AAV2 genome sequence. The number of point mutations was measured,

and error rate was calculated by normalizing the number of point mutations to the length of the assembled contig.

3.6 Acknowledgements

Thank you to Professor Danqing (Bonnie) Zhu, who assisted with initial BLaSR methodology development and is performing the complete high-throughput contig assembly and sequence analysis of the AAV capsid libraries. We'd also like to thank Li Lin for his assistance in data analysis. Thanks to Professor Tom Nowakowski's lab at UCSF for assisting with the AAV selection in adult brain tissue. Specifically, thanks to Julio Rivera de Jesus, Anthony Lee, and Albert Wang for procuring, culturing, and transducing the adult brain tissue. Library preparation (DNA shearing, bead cleanup, and library quantification) and sequencing was done by the Functional Genomics Lab (FGL) and the Vincent J. Coates Genomics Sequencing Lab (GSL) at University of California, Berkeley.

Bibliography

- [1] Michael F. Naso, Brian Tomkowicz, William L. Perry, and William R. Strohl. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy, aug 2017.
- [2] Chengwen Li and R. Jude Samulski. Engineering adeno-associated virus vectors for gene therapy, apr 2020.
- [3] Dan Wang, Phillip W.L. Tai, and Guangping Gao. Adeno-associated virus vector as a platform for gene therapy delivery, may 2019.
- [4] Simon J. O’Carroll, William H. Cook, and Deborah Young. AAV Targeting of Glial Cell Types in the Central and Peripheral Nervous System and Relevance to Human Gene Therapy. *Frontiers in Molecular Neuroscience*, 13:618020, jan 2021.
- [5] Marco Ledri, Andreas T. Sørensen, Merab Kokaia, David P.D. Woldbye, and Casper R. Gøtzsche. Editorial: Gene Therapy in the CNS – Progress and Prospects for Novel Therapies. *Frontiers in Molecular Neuroscience*, 14:778134, oct 2021.
- [6] Shaila D Ghanekar, Sheng-Han Kuo, Joseph S Staffetti, and Theresa A Zesiewicz. Current and emerging treatment modalities for spinocerebellar ataxias.
- [7] Samantha Powers, Shibi Likhite, Kamal K. Gadalla, Carlos J. Miranda, Amy J. Hufenberger, Cassandra Dennys, Kevin D. Foust, Pablo Morales, Christopher R. Pierson, Federica Rinaldi, Stephanie Perry, Brad Bolon, Nicolas Wein, Stuart Cobb, Brian K. Kaspar, and Kathrin C. Meyer. Novel MECP2 gene therapy is effective in a multicenter study using two mouse models of Rett syndrome and is safe in non-human primates. *Molecular Therapy*, 31(9):2767–2782, sep 2023.
- [8] Sylvie Boutin, Virginie Monteilhet, Philippe Veron, Christian Leborgne, Olivier Benveniste, Marie Françoise Montus, and Carole Masurier. Prevalence of serum IgG and neutralizing factors against adeno-associated virus (AAV) types 1, 2, 5, 6, 8, and 9 in the healthy population: implications for gene therapy using AAV vectors. *Human gene therapy*, 21(6):704–712, jun 2010.
- [9] B. L. Gurda, M. A. DiMattia, E. B. Miller, A. Bennett, R. McKenna, W. S. Weichert, C. D. Nelson, W.-j. Chen, N. Muzyczka, N. H. Olson, R. S. Sinkovits, J. A.

- Chiorini, S. Zolotutkhin, O. G. Kozyreva, R. J. Samulski, T. S. Baker, C. R. Parrish, and M. Agbandje-McKenna. Capsid Antibodies to Different Adeno-Associated Virus Serotypes Bind Common Regions. *Journal of Virology*, 87(16):9111–9124, aug 2013.
- [10] Bethan J. Critchley, H. Bobby Gaspar, and Sara Benedetti. Targeting the central nervous system in lysosomal storage diseases: Strategies to deliver therapeutics across the blood-brain barrier. *Molecular Therapy*, 31(3):657–675, mar 2023.
- [11] Rashad Hussain, Hira Zubair, Sarah Pursell, and Muhammad Shahab. Neurodegenerative Diseases: Regenerative Mechanisms and Novel Therapeutic Approaches. *Brain Sciences*, 8(9), sep 2018.
- [12] Olivia J. Yang, Gabriella L. Robilotto, Firoj Alom, Karla Alemán, Karthik Devulapally, Abigail Morris, and Aaron D. Mickle. Evaluating the transduction efficiency of systemically delivered AAV vectors in the rat nervous system. *Frontiers in Neuroscience*, 17:1001007, jan 2023.
- [13] Kelly A. Langert and Eric M. Brey. Strategies for targeted delivery to the peripheral nerve. *Frontiers in Neuroscience*, 12(NOV):419793, nov 2018.
- [14] Liza Malong, Ilaria Napoli, Giulia Casal, Adrienne Flanagan, Hien T Zhao, and Alison C Lloyd Correspondence. Characterization of the structure and control of the blood-nerve barrier identifies avenues for therapeutic delivery. 2023.
- [15] Kevin D. Foust, Emily Nurre, Chrystal L. Montgomery, Anna Hernandez, Curtis M. Chan, and Brian K. Kaspar. Intravascular AAV9 preferentially targets neonatal neurons and adult astrocytes. *Nature Biotechnology* 2008 27:1, 27(1):59–65, dec 2008.
- [16] Yannick Tanguy, Maria G. Biferi, Aurore Besse, Stephanie Astord, Mathilde Cohen-Tannoudji, Thibaut Marais, and Martine Barkats. Systemic AAVrh10 provides higher transgene expression than AAV9 in the brain and the spinal cord of neonatal mice. *Frontiers in Molecular Neuroscience*, 8(JULY):151552, jul 2015.
- [17] Sara E. Gombash, Christopher J. Cowley, Julie A. Fitzgerald, Jodie C.E. Hall, Christian Mueller, Fedias L. Christofi, and Kevin D. Foust. Intravenous AAV9 efficiently transduces myenteric neurons in neonate and juvenile mice. *Frontiers in Molecular Neuroscience*, 7(OCT), oct 2014.
- [18] A. Kagiava, J. Richter, C. Tryfonos, M. Leal-Julià, I. Sargiannidou, C. Christodoulou, A. Bosch, and K. A. Kleopa. Efficacy of AAV serotypes to target Schwann cells after intrathecal and intravenous delivery. *Scientific Reports* 2021 11:1, 11(1):1–14, dec 2021.
- [19] G. Allt and J. G. Lawrenson. The blood–nerve barrier: enzymes, transporters and receptors—a comparison with the blood–brain barrier. *Brain Research Bulletin*, 52(1):1–12, may 2000.

- [20] Kishena C. Wadhvani and Stanley I. Rapoport. Transport properties of vertebrate blood-nerve barrier: Comparison with blood-brain barrier. *Progress in Neurobiology*, 43(3):235–279, jun 1994.
- [21] Zhilong Chen, Guoqing Fan, Anan Li, Jing Yuan, and Tonghui Xu. rAAV2-Retro Enables Extensive and High-Efficient Transduction of Lower Motor Neurons following Intramuscular Injection. *Molecular Therapy. Methods Clinical Development*, 17:21, jun 2020.
- [22] Zhong Wang, Tong Zhu, Chunping Qiao, Liqiao Zhou, Bing Wang, Jian Zhang, Chunlian Chen, Juan Li, and Xiao Xiao. Adeno-associated virus serotype 8 efficiently delivers genes to muscle and heart. *Nature biotechnology*, 23(3):321–328, mar 2005.
- [23] Hongwei Yu, Gregory Fischer, Lejla Ferhatovic, Fan Fan, Alan R. Light, Dorothee Weihrauch, Damir Sapunar, Hiroyuki Nakai, Frank Park, and Quinn H. Hogan. Intraganglionic AAV6 Results in Efficient and Long-Term Gene Transfer to Peripheral Sensory Nervous System in Adult Rats. *PLoS ONE*, 8(4):61266, apr 2013.
- [24] Wei Cheng, Jing Huang, Xiu Qing Fu, Wei Ya Tian, Peng Ming Zeng, Yang Li, and Zhen Ge Luo. Intrathecal delivery of AAV-NDNF ameliorates disease progression of ALS mice. *Molecular Therapy*, 31(11):3277–3289, nov 2023.
- [25] F. E.V. Scheijmans, I. Cuppen, M. M. Zwartkruis, I. Signoria, C. van Ekris, F. Asselman, R. I. Wadman, E. F. Knol, W. L. van der Pol, and E. J.N. Groen. Inflammatory markers in cerebrospinal fluid of paediatric spinal muscular atrophy patients receiving nusinersen treatment. *European Journal of Paediatric Neurology*, 42:34–41, jan 2023.
- [26] Sara Marcó, Virginia Haurigot, Maria Luisa Jaén, Albert Ribera, Víctor Sánchez, Maria Molas, Miguel Garcia, Xavier León, Carles Roca, Xavier Sánchez, Joan Bertolin, Jennifer Pérez, Gemma Elias, Marc Navarro, Ana Carretero, Martí Pumarola, Anna Andaluz, Yvonne Espada, Sonia Añor, and Fatima Bosch. Seven-year follow-up of durability and safety of AAV CNS gene therapy for a lysosomal storage disorder in a large animal. *Molecular Therapy. Methods Clinical Development*, 23:370, dec 2021.
- [27] Kai Zhou, Jinming Han, Yafeng Wang, Yaodong Zhang, and Changlian Zhu. Routes of administration for adeno-associated viruses carrying gene therapies for brain diseases. *Frontiers in Molecular Neuroscience*, 15:988914, oct 2022.
- [28] Yuan Zhu, Pritam Ghosh, Patrick Charnay, Dennis K. Burns, and Luis F. Parada. Neurofibromas in NF1: Schwann Cell Origin and Role of Tumor Environment. *Science (New York, N.Y.)*, 296(5569):920, may 2002.
- [29] Michael J. Castle, Heikki T. Turunen, Luk H. Vandenberghe, and John H. Wolfe. Controlling AAV Tropism in the Nervous System with Natural and Engineered Capsids. *Methods in molecular biology (Clifton, N.J.)*, 1382:133, 2016.

- [30] S. Pillay, N. L. Meyer, A. S. Puschnik, O. Davulcu, J. Diep, Y. Ishikawa, L. T. Jae, J. E. Wosen, C. M. Nagamine, M. S. Chapman, and J. E. Carette. An essential receptor for adeno-associated virus infection. *Nature*, 530(7588):108–112, feb 2016.
- [31] Jalish M. Riyad and Thomas Weber. Intracellular trafficking of adeno-associated virus (AAV) vectors: challenges and future directions. *Gene Therapy* 2021 28:12, 28(12):683–696, mar 2021.
- [32] Li Zhong, Baozheng Li, Giridhararao Jayandharan, Cathryn S. Mah, Lakshmanan Govindasamy, Mavis Agbandje-McKenna, Roland W. Herzog, Kirsten A. Weigel-Van Aken, Jacqueline A. Hobbs, Sergei Zolotukhin, Nicholas Muzyczka, and Arun Srivastava. Tyrosine phosphorylation of AAV2 vectors and its consequences on viral intracellular trafficking and transgene expression. *Virology*, 381(2):194, nov 2008.
- [33] Li Zhong, Baozheng Li, Cathryn S. Mah, Lakshmanan Govindasamy, Mavis Agbandje-McKenna, Mario Cooper, Roland W. Herzog, Irene Zolotukhin, Kenneth H. Warrington, Kirsten A. Weigel-Van Aken, Jacqueline A. Hobbs, Sergei Zolotukhin, Nicholas Muzyczka, and Arun Srivastava. Next generation of adeno-associated virus 2 vectors: Point mutations in tyrosines lead to high-efficiency transduction at lower doses. *Proceedings of the National Academy of Sciences of the United States of America*, 105(22):7827, jun 2008.
- [34] Dan Wang, Shaoyong Li, Dominic J. Gessler, Jun Xie, Li Zhong, Jia Li, Karen Tran, Kim Van Vliet, Lingzhi Ren, Qin Su, Ran He, Jason E. Goetzmann, Terence R. Flotte, Mavis Agbandje-McKenna, and Guangping Gao. A Rationally Engineered Capsid Variant of AAV9 for Systemic CNS-Directed and Peripheral Tissue-Detargeted Gene Delivery in Neonates. *Molecular therapy. Methods clinical development*, 9:234–246, jun 2018.
- [35] Anne Girod, Martin Ried, Christiane Wobus, Harald Lahm, Kristin Leike, Jürgen Kleinschmidt, Gilbert Deléage, and Michael Hallek. Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2. *Nature medicine*, 5(9):1052–1056, sep 1999.
- [36] Yizheng Yao, Jun Wang, Yi Liu, Yuan Qu, Kaikai Wang, Yang Zhang, Yuxin Chang, Zhi Yang, Jie Wan, Junfeng Liu, Hiroshi Nakashima, Sean E. Lawler, E. Antonio Chiocca, Choi Fong Cho, and Fengfeng Bei. Variants of the adeno-associated virus serotype 9 with enhanced penetration of the blood–brain barrier in rodents and primates. *Nature Biomedical Engineering* 2022 6:11, 6(11):1257–1271, oct 2022.
- [37] Narendra Maheshri, James T Koerber, Brian K Kaspar, and David V Schaffer. Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. 2006.

- [38] Sourav R. Choudhury, Zachary Fitzpatrick, Anne F. Harris, Stacy A. Maitland, Jennifer S. Ferreira, Yuanfan Zhang, Shan Ma, Rohit B. Sharma, Heather L. Gray-Edwards, Jacob A. Johnson, Aime K. Johnson, Laura C. Alonso, Claudio Punzo, Kathryn R. Wagner, Casey A. Maguire, Robert M. Kotin, Douglas R. Martin, and Miguel Sena-Esteves. In vivo selection yields AAV-B1 capsid for central nervous system and muscle gene therapy. *Molecular Therapy*, 24(7):1247–1257, aug 2016.
- [39] James T Koerber, Jae-Hyung Jang, and David V Schaffer. DNA Shuffling of Adeno-associated Virus Yields Functionally Diverse Viral Progeny. *Molecular Therapy*, 16(10):1703–1709, 2008.
- [40] Leah C. Byrne, Timothy P. Day, Meike Visel, Jennifer A. Strazzeri, Cécile Fortuny, Deniz Dalkara, William H. Merigan, David V. Schaffer, and John G. Flannery. In vivo-directed evolution of adeno-associated virus in the primate retina. *JCI Insight*, 5(10), may 2020.
- [41] Ken Y. Chan, Min J. Jang, Bryan B. Yoo, Alon Greenbaum, Namita Ravi, Wei Li Wu, Luis Sánchez-Guardado, Carlos Lois, Sarkis K. Mazmanian, Benjamin E. Deverman, and Viviana Gradinaru. Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. *Nature neuroscience*, 20(8):1172–1179, aug 2017.
- [42] Deniz Dalkara, Leah C. Byrne, Ryan R. Klimczak, Meike Visel, Lu Yin, William H. Merigan, John G. Flannery, and David V. Schaffer. In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous. *Science translational medicine*, 5(189), jun 2013.
- [43] Bilge E. Öztürk, Molly E. Johnson, Michael Kleymann, Serhan Turunç, Jing He, Sara Jabalameli, Zhouhuan Xi, Meike Visel, Valérie L. Dufour, Simone Iwabe, Felipe Pompeo Marinho, Gustavo D. Aguirre, José Alain Sahel, David V. Schaffer, Andreas R. Pfennig, John G. Flannery, William A. Beltran, William R. Stauffer, and Leah C. Byrne. Scaavengr, a transcriptome-based pipeline for quantitative ranking of engineered aavs with single-cell resolution. *eLife*, 10, oct 2021.
- [44] D. Gowanlock R. Tervo, Bum Yeol Hwang, Sarada Viswanathan, Thomas Gaj, Maria Lavzin, Kimberly D. Ritola, Sarah Lindo, Susan Michael, Elena Kuleshova, David Ojala, Cheng Chiu Huang, Charles R. Gerfen, Jackie Schiller, Joshua T. Dudman, Adam W. Hantman, Loren L. Looger, David V. Schaffer, and Alla Y. Karpova. A Designer AAV Variant Permits Efficient Retrograde Access to Projection Neurons. *Neuron*, 92(2):372–382, oct 2016.
- [45] Kei Adachi, Tatsuji Enoki, Yasuhiro Kawano, Michael Veraz, and Hiroyuki Nakai. Drawing a high-resolution functional map of adeno-associated virus capsid by massively parallel sequencing. *Nature Communications* 2014 5:1, 5(1):1–14, jan 2014.

- [46] Marcus Davidsson, Gang Wang, Patrick Aldrin-Kirk, Tiago Cardoso, Sara Nolbrant, Morgan Hartnor, Janitha Mudannayake, Malin Parmar, and Tomas Björklund. A systematic capsid evolution approach performed in vivo for the design of AAV vectors with tailored properties and tropism. *Proceedings of the National Academy of Sciences of the United States of America*, 116(52):27053–27062, dec 2019.
- [47] Xinhong Chen, Sripriya Ravindra Kumar, Cameron D Adams, Andrew S Fox, Isaac M Chiu, Viviana Gradinaru Correspondence, Daping Yang, Tongtong Wang, Damien A Wolfe, Cynthia M Arokiaraj, Victoria Ngo, Lillian J Campos, Jessica A Griffiths, Takako Ichiki, Sarkis K Mazmanian, Peregrine B Osborne, Janet R Keast, Cory T Miller, and Viviana Gradinaru. Engineered AAVs for non-invasive gene delivery to rodent and non-human primate nervous systems II NeuroResource Engineered AAVs for non-invasive gene delivery to rodent and non-human primate nervous systems. *Neuron*, 110:2242–2257.e6, 2022.
- [48] Benjamin E. Deverman, Piers L. Pravdo, Bryan P. Simpson, Sripriya Ravindra Kumar, Ken Y. Chan, Abhik Banerjee, Wei Li Wu, Bin Yang, Nina Huber, Sergiu P. Pasca, and Viviana Gradinaru. Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nature Biotechnology*, 34(2):204–209, feb 2016.
- [49] David S. Ojala, Sabrina Sun, Jorge L. Santiago-Ortiz, Mikhail G. Shapiro, Philip A. Romero, and David V. Schaffer. In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the SVZ. *Molecular Therapy*, 26(1):304–319, jan 2018.
- [50] Miguel R. Chuapoco, Nicholas C. Flytzanis, Nick Goeden, J. Christopher Octeau, Kristina M. Roxas, Ken Y. Chan, Jon Scherrer, Janet Winchester, Roy J. Blackburn, Lillian J. Campos, Kwun Nok Mimi Man, Junqing Sun, Xinhong Chen, Arthur Lefevre, Vikram Pal Singh, Cynthia M. Arokiaraj, Timothy F. Shay, Julia Vendemiatti, Min J. Jang, John K. Mich, Yemeserach Bishaw, Bryan B. Gore, Victoria Omstead, Naz Taskin, Natalie Weed, Boaz P. Levi, Jonathan T. Ting, Cory T. Miller, Benjamin E. Deverman, James Pickel, Lin Tian, Andrew S. Fox, and Viviana Gradinaru. Adeno-associated viral vectors for functional intravenous gene transfer throughout the non-human primate brain. *Nature Nanotechnology 2023 18:10*, 18(10):1241–1251, jul 2023.
- [51] Mathieu Nonnenmacher, Wei Wang, Matthew A. Child, Xiao Qin Ren, Carol Huang, Amy Zhen Ren, Jenna Tocci, Qingmin Chen, Kelsey Bittner, Katherine Tyson, Nilesh Pande, Charlotte Hiu Yan Chung, Steven M. Paul, and Jay Hou. Rapid evolution of blood-brain-barrier-penetrating AAV capsids by RNA-driven biopanning. *Molecular Therapy. Methods Clinical Development*, 20:366, mar 2021.
- [52] Mohammadsharif Tabebordbar, Kim A. Lagerborg, Alexandra Stanton, Emily M. King, Simon Ye, Liana Tellez, Allison Krunnusz, Sahar Tavakoli, Jeffrey J. Widrick, Kathleen A. Messemer, Emily C. Troiano, Behzad Moghadaszadeh, Bryan L. Peacker,

- Krystynne A. Leacock, Naftali Horwitz, Alan H. Beggs, Amy J. Wagers, Pardis C. Sabeti, Emily M. King, Simon Ye, Liana Tellez, Allison Krunnusz, Sahar Tavakoli, Jeffrey J. Widrick, Kathleen A. Messemmer, Emily C. Troiano, Behzad Moghadaszadeh, Bryan L. Peacker, Krystynne A. Leacock, and Naftali Horwitz. Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. *Cell*, 184(19), sep 2021.
- [53] Alexandra C. Stanton, Kim A. Lagerborg, Liana Tellez, Allison Krunnusz, Emily M. King, Simon Ye, Isaac H. Solomon, Mohammadsharif Tabebordbar, and Pardis C. Sabeti. Systemic administration of novel engineered AAV capsids facilitates enhanced transgene expression in the macaque CNS. *Med*, 4(1):31–50.e8, jan 2023.
- [54] Christopher A Voigt, Carlos Martinez, Zhen-Gang Wang, Stephen L Mayo, and Frances H Arnold. Protein building blocks preserved by recombination. 2002.
- [55] J. Santiago-Ortiz, D. S. Ojala, O. Westesson, J. R. Weinstein, S. Y. Wong, A. Stein-sapir, S. Kumar, I. Holmes, and D. V. Schaffer. AAV ancestral reconstruction library enables selection of broadly infectious viral variants. *Gene Therapy*, 22(12):934–946, dec 2015.
- [56] Eric Zinn, Simon Pacouret, Vadim Khaychuk, Heikki T. Turunen, Livia S. Carvalho, Eva Andres-Mateos, Samiksha Shah, Rajani Shelke, Anna C. Maurer, Eva Maurer, Ru Xiao, and Luk H. Vandenberghe. In silico reconstruction of the viral evolutionary lineage yields a potent gene therapy vector. *Cell reports*, 12(6):1056, aug 2015.
- [57] Pierce J. Ogden, Eric D. Kelsic, Sam Sinai, and George M. Church. Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. *Science*, 366(6469):1139–1143, nov 2019.
- [58] Drew H. Bryant, Ali Bashir, Sam Sinai, Nina K. Jain, Pierce J. Ogden, Patrick F. Riley, George M. Church, Lucy J. Colwell, and Eric D. Kelsic. Deep diversification of an AAV capsid protein by machine learning. *Nature biotechnology*, 39(6):691–696, jun 2021.
- [59] Qin Huang, Ken Y. Chan, Isabelle G. Tobey, Yujia Alina Chan, Tim Poterba, Christine L. Boutros, Alejandro B. Balazs, Richard Daneman, Jonathan M. Bloom, Cotton Seed, and Benjamin E. Deverman. Delivering genes across the blood-brain barrier: LY6A, a novel cellular receptor for AAV-PHP.B capsids. *PLOS ONE*, 14(11):e0225206, nov 2019.
- [60] Qin Huang, Albert T. Chen, Ken Y. Chan, Hikari Sorensen, Andrew J. Barry, Bahar Azari, Qingxia Zheng, Thomas Beddow, Binhui Zhao, Isabelle G. Tobey, Cynthia Moncada-Reid, Fatma Elzahraa Eid, Christopher J. Walkey, M. Cecilia Ljungberg,

- William R. Lagor, Jason D. Heaney, Yujia A. Chan, and Benjamin E. Deverman. Targeting AAV vectors to the central nervous system by engineering capsid–receptor interactions that enable crossing of the blood–brain barrier. *PLOS Biology*, 21(7):e3002112, jul 2023.
- [61] Danqing Zhu, David H. Brookes, Akosua Busia, Ana Carneiro, Clara Fannjiang, Galina Popova, David Shin, Kevin C. Donohue, Li F. Lin, Zachary M. Miller, Evan R. Williams, Edward F. Chang, Tomasz J. Nowakowski, Jennifer Listgarten, and David V. Schaffer. Optimal trade-off control in machine learning–based library design, with application to adeno-associated virus (AAV) for gene therapy. *Science Advances*, 10(4):3786, jan 2024.
- [62] Hau Kiu Edna Au, Mark Isalan, and Michal Mielcarek. Gene Therapy Advances: A Meta-Analysis of AAV Usage in Clinical Settings. *Frontiers in Medicine*, 8:809118, feb 2022.
- [63] Lin Kang, Shilin Jin, Jiayi Wang, Zhongyue Lv, Chengqi Xin, Chengcheng Tan, Mengke Zhao, Liang Wang, and Jing Liu. AAV vectors applied to the treatment of CNS disorders: Clinical status and challenges. *Journal of Controlled Release*, 355:458–473, mar 2023.
- [64] Haowen Zhang, Chirag Jain, and Srinivas Aluru. A comprehensive evaluation of long read error correction methods. *BMC Genomics*, 21(6):1–15, dec 2020.
- [65] Christopher R. Barnes, Hyuncheol Lee, David S. Ojala, Kazuomori K. Lewis, Prajit Limsirichai, and David V. Schaffer. Genome-wide activation screens to increase adeno-associated virus production. *Molecular therapy. Nucleic acids*, 26:94–103, dec 2021.
- [66] Adam J. Schieferecke, Hyuncheol Lee, Aleysha Chen, Vindhya Kilaru, Justin Krish Williams, and David V. Schaffer. Evolving membrane-associated accessory protein variants for improved adeno-associated virus production. *Molecular Therapy*, dec 2023.
- [67] Takashi Kei Kishimoto and Richard Jude Samulski. Addressing high dose AAV toxicity – ‘one and done’ or ‘slower and lower’? *Expert Opinion on Biological Therapy*, 22(9):1067–1071, sep 2022.
- [68] Perry B. Shieh, Nancy L. Kuntz, James J. Dowling, Wolfgang Müller-Felber, Carsten G. Bönnemann, Andreea M. Seferian, Laurent Servais, Barbara K. Smith, Francesco Muntoni, Astrid Blaschek, A. Reghan Foley, Dimah N. Saade, Sarah Neuhaus, Lindsay N. Alfano, Alan H. Beggs, Ana Buj-Bello, Martin K. Childers, Tina Duong, Robert J. Graham, Minal Jain, Julie Coats, Vicky MacBean, Emma S. James, Jun Lee, Fulvio Mavilio, Weston Miller, Fatbardha Varfaj, Michael Murtagh, Cong Han, Mojtaba Noursalehi, Michael W. Lawlor, Suyash Prasad, and Salvador Rico. Safety and efficacy of gene replacement therapy for X-linked myotubular myopathy

- (ASPIRO): a multinational, open-label, dose-escalation trial. *The Lancet Neurology*, 22(12):1125–1139, dec 2023.
- [69] Juliette Hordeaux, Elizabeth L. Buza, Brianne Jeffrey, Chunjuan Song, Tahsin Jahan, Yuan Yuan, Yanqing Zhu, Peter Bell, Mingyao Li, Jessica A. Chichester, Roberto Calcedo, and James M. Wilson. MicroRNA-mediated inhibition of transgene expression reduces dorsal root ganglion toxicity by AAV vectors in primates. *Science Translational Medicine*, 12(569):9188, nov 2020.
- [70] Jiang Hui Wang, Dominic J. Gessler, Wei Zhan, Thomas L. Gallagher, and Guangping Gao. Adeno-associated virus as a delivery vector for gene therapy of human diseases. *Signal Transduction and Targeted Therapy* 2024 9:1, 9(1):1–33, apr 2024.
- [71] Matthew J. During. Adeno-associated virus as a gene delivery system. *Advanced Drug Delivery Reviews*, 27(1):83–94, aug 1997.
- [72] Kanav Bhatheja and Jeffrey Field. Schwann cells: Origins and role in axonal maintenance and regeneration. *The International Journal of Biochemistry Cell Biology*, 38(12):1995–1999, jan 2006.
- [73] Khidhir Kamil, Muhammad Dain Yazid, Ruszymah Bt Hj Idrus, Srijit Das, and Jaya Kumar. Peripheral Demyelinating Diseases: From Biology to Translational Medicine. *Frontiers in Neurology*, 10:433827, mar 2019.
- [74] Sophia H. Zhang, Galina V. Shurin, Hasan Khosravi, Rashek Kazi, Oleg Kruglov, Michael R. Shurin, and Yuri L. Bunimovich. Immunomodulation by Schwann cells in disease. *Cancer Immunology, Immunotherapy : CII*, 69(2):245, feb 2020.
- [75] Leah J. Kershner, Kwangmin Choi, Jianqiang Wu, Xiyuan Zhang, Melissa Perrino, Nathan Salomonis, Jack F. Shern, and Nancy Ratner. Multiple Nf1 Schwann cell populations reprogram the plexiform neurofibroma tumor microenvironment. *JCI Insight*, 7(18), sep 2022.
- [76] Namrata G.R. Raut, Laura A. Maile, Leila M. Oswalt, Irati Mitxelena, Aaditya Adlakha, Kourtney L. Sprague, Ashley R. Rupert, Lane Bokros, Megan C. Hofmann, Jennifer Patrissi-Cram, Tilat A. Rizvi, Luis F. Queme, Kwangmin Choi, Nancy Ratner, and Michael P. Jankowski. Schwann cells modulate nociception in neurofibromatosis 1. *JCI Insight*, 9(2), jan 2024.
- [77] Ren Yuan Bai, Dominic Esposito, Ada J. Tam, Frank McCormick, Gregory J. Riggins, D. Wade Clapp, and Verena Staedtke. Feasibility of Using NF1-GRD and AAV for Gene Replacement Therapy in NF1-Associated Tumors. *Gene therapy*, 26(6):277, jun 2019.

- [78] J. Homs, L. Ariza, G. Pagès, E. Udina, X. Navarro, M. Chillón, and A. Bosch. Schwann cell targeting via intrasciatic injection of AAV8 as gene therapy strategy for peripheral nerve regeneration. *Gene Therapy* 2011 18:6, 18(6):622–630, feb 2011.
- [79] S. A. Hoyng, F. De Winter, S. Gnani, L. Van Egmond, C. L. Attwell, M. R. Tannemaat, J. Verhaagen, and M. J.A. Malessy. Gene delivery to rat and human Schwann cells and nerve segments: a comparison of AAV 1–9 and lentiviral vectors. *Gene Therapy* 2015 22:10, 22(10):767–780, may 2015.
- [80] Jerry R. Mendell, Anne M. Connolly, Kelly J. Lehman, Danielle A. Griffin, Sohrab Z. Khan, Sachi D. Dharia, Lucía Quintana-Gallardo, and Louise R. Rodino-Klapac. Testing preexisting antibodies prior to AAV gene transfer therapy: rationale, lessons and future considerations. *Molecular Therapy. Methods Clinical Development*, 25:74, jun 2022.
- [81] Oliver J. Müller, Felix Kaul, Matthew D. Weitzman, Renata Pasqualini, Wadih Arap, Jürgen A. Kleinschmidt, and Martin Trepel. Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. *Nature Biotechnology*, 21(9):1040 – 1046, sep 2003.
- [82] Yi Hui Gu, Xi Wei Cui, Jie Yi Ren, Man Mei Long, Wei Wang, Cheng Jiang Wei, Rehanguli Aimaier, Yue Hua Li, Man Hon Chung, Bin Gu, Qing Feng Li, and Zhi Chao Wang. Selection of internal references for RT-qPCR assays in Neurofibromatosis type 1 (NF1) related Schwann cell lines. *PLOS ONE*, 16(2):e0241821, feb 2021.
- [83] Marianne F. James, Elizabeth Stivison, Roberta Beauchamp, Sangyeul Han, Hua Li, Margaret R. Wallace, James F. Gusella, Anat O. Stemmer-Rachamimov, and Vijaya Ramesh. Regulation of mTOR complex 2 signaling in neurofibromatosis 2-deficient target cell types. *Molecular cancer research : MCR*, 10(5):649–659, may 2012.
- [84] Hua Li, Lung Ji Chang, Debbie R. Neubauer, David F. Muir, and Margaret R. Wallace. immortalization of human normal and NF1 neurofibroma Schwann cells. *Laboratory Investigation*, 96(10):1105–1115, oct 2016.
- [85] Shanshan Ma, Zihao Mi, Zhenzhen Wang, Lele Sun, Tingting Liu, Peidian Shi, Chuan Wang, Xiaotong Xue, Wenjie Chen, Zhe Wang, Yueqian Yu, Yuan Zhang, Fangfang Bao, Na Wang, Honglei Wang, Qianqian Xia, Hong Liu, Yonghu Sun, and Furen Zhang. Single-cell sequencing analysis reveals development and differentiation trajectory of Schwann cells manipulated by *M. leprae*. *PLOS Neglected Tropical Diseases*, 17(7), jul 2023.
- [86] Mavis Agbandje-McKenna and Jürgen Kleinschmidt. AAV Capsid Structure and Cell Interactions. *Methods in Molecular Biology*, 807:47–92, 2011.

- [87] Grant J. Logan, Mario Mietzsch, Neeta Khandekar, Arlene D'Silva, Daniel Anderson, Mawj Mandwie, Jane Hsi, Austin R. Nelson, Paul Chipman, Jennifer Jackson, Peter Schofield, Daniel Christ, Christopher C. Goodnow, Joanne H. Reed, Michelle A. Farrar, Robert McKenna, and Ian E. Alexander. Structural and functional characterization of capsid binding by anti-AAV9 monoclonal antibodies from infants after SMA gene therapy. *Molecular Therapy*, 31(7):1979–1993, jul 2023.
- [88] Abrar Alhindi, Ines Boehm, Rachael O. Forsythe, Janice Miller, Richard J.E. Skipworth, Hamish Simpson, Ross A. Jones, and Thomas H. Gillingwater. Terminal Schwann cells at the human neuromuscular junction. *Brain Communications*, 3(2), apr 2021.
- [89] Sofia Meyer zu Reckendorf, Christine Brand, Maria T. Pedro, Jutta Hegler, Corinna S. Schilling, Raissa Lerner, Laura Bindila, Gregor Antoniadis, and Bernd Knöll. Lipid metabolism adaptations are reduced in human compared to murine Schwann cells following injury. *Nature Communications* 2020 11:1, 11(1):1–15, may 2020.
- [90] Sherif G. Ahmed, Farnaz Hadaegh, and Gary J. Brenner. Developing myelin specific promoters for schwannoma gene therapy. *Journal of neuroscience methods*, 323:77, jul 2019.
- [91] Matthieu Drouyer, Tak Ho Chu, Elodie Labit, Florencia Haase, Renina Gale Navarro, Deborah Nazareth, Nicole Rosin, Jessica Merjane, Suzanne Scott, Marti Cabanes-Creus, Adrian Westhaus, Erhua Zhu, Rajiv Midha, Ian E. Alexander, Jeff Biernaskie, Samantha L. Ginn, and Leszek Lisowski. Novel AAV variants with improved tropism for human Schwann cells. *Molecular Therapy Methods and Clinical Development*, 32(2):101234, jun 2024.
- [92] Juliette Hordeaux, Qiang Wang, Nathan Katz, Elizabeth L. Buza, Peter Bell, and James M. Wilson. The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice. *Molecular Therapy*, 26(3):664, mar 2018.
- [93] Sophie N. Mathiesen, Jasmine L. Lock, Lucia Schoderboeck, Wickliffe C. Abraham, and Stephanie M. Hughes. CNS Transduction Benefits of AAV-PHP.eB over AAV9 Are Dependent on Administration Route and Mouse Strain. *Molecular Therapy. Methods Clinical Development*, 19:447, dec 2020.
- [94] Uri Arad. Modified Hirt Procedure for Rapid Purification of Extrachromosomal DNA from Mammalian Cells. *BioTechniques*, 24(5):760–762, 1998.
- [95] Shaza S. Issa, Alisa A. Shaimardanova, Valeriya V. Solovyeva, and Albert A. Rizvanov. Various AAV Serotypes and Their Applications in Gene Therapy: An Overview. *Cells* 2023, Vol. 12, Page 785, 12(5):785, mar 2023.

- [96] Vedell Louis Jeune, Jakob A. Joergensen, Roger J. Hajjar, and Thomas Weber. Pre-existing anti-adenoviral virus antibodies as a challenge in AAV gene therapy. *Human gene therapy methods*, 24(2):59–67, apr 2013.
- [97] Jingxuan Guo, Li F. Lin, Sydney V. Oraskovich, Julio A. Rivera de Jesús, Jennifer Listgarten, and David V. Schaffer. Computationally guided AAV engineering for enhanced gene delivery. *Trends in Biochemical Sciences*, 49(5):457–469, may 2024.
- [98] Josephine B. Oehler, Helen Wright, Zornitza Stark, Andrew J. Mallett, and Ulf Schmitz. The application of long-read sequencing in clinical settings. *Human Genomics*, 17(1):1–13, dec 2023.
- [99] Daniel P. Agostinho, Yilei Fu, Vipin K. Menon, Ginger A. Metcalf, Todd J. Treangen, and Fritz J. Sedlazeck. Unveiling microbial diversity: harnessing long-read sequencing technology. *Nature Methods* 2024 21:6, 21(6):954–966, apr 2024.
- [100] Niannian Li, Kairang Jin, Yanmin Bai, Haifeng Fu, Lin Liu, and Bin Liu. Tn5 Transposase Applied in Genomics Research. *International Journal of Molecular Sciences*, 21(21):1–15, nov 2020.
- [101] Mindy Steiniger-White, Ivan Rayment, and William S. Reznikoff. Structure/function insights into Tn5 transposition. *Current Opinion in Structural Biology*, 14(1):50–57, 2004.
- [102] Surya Tripathi, Carlos Geert Pieter Voogdt, Stefan Oliver Bassler, Mary Anderson, Po Hsun Huang, Nazgul Sakenova, Tümay Capraz, Sunit Jain, Alexandra Koumoutsis, Afonso Martins Bravo, Valentine Trotter, Michael Zimmerman, Justin L. Sonnenburg, Cullen Buie, Athanasios Typas, Adam M. Deutschbauer, Anthony L. Shiver, and Kerwyn Casey Huang. Randomly barcoded transposon mutant libraries for gut commensals I: Strategies for efficient library construction. *Cell Reports*, 43(1):113517, jan 2024.
- [103] Samuel R. Fels, Grant M. Zane, Sean M. Blake, and Judy D. Wall. Rapid transposon liquid enrichment sequencing (TnLE-seq) for gene fitness evaluation in underdeveloped bacterial systems. *Applied and Environmental Microbiology*, 79(23):7510–7517, dec 2013.
- [104] Jeffrey D. Gawronski, Sandy M.S. Wong, Georgia Giannoukos, Doyle V. Ward, and Brian J. Akerley. Tracking insertion mutants within libraries by deep sequencing and a genome-wide screen for *Haemophilus* genes required in the lung. *Proceedings of the National Academy of Sciences of the United States of America*, 106(38):16422–16427, sep 2009.
- [105] Andrew L. Goodman, Meng Wu, and Jeffrey I. Gordon. Identifying microbial fitness determinants by insertion sequencing using genome-wide transposon mutant libraries. *Nature Protocols* 2011 6:12, 6(12):1969–1980, nov 2011.

- [106] Gemma C. Langridge, Minh Duy Phan, Daniel J. Turner, Timothy T. Perkins, Leopold Parts, Jana Haase, Ian Charles, Duncan J. Maskell, Sarah E. Peters, Gordon Dougan, John Wain, Julian Parkhill, and A. Keith Turner. Simultaneous assay of every Salmonella Typhi gene using one million transposon mutants. *Genome Research*, 19(12):2308–2316, dec 2009.
- [107] Tim van Opijnen, Kip L. Bodi, and Andrew Camilli. Tn-seq: high-throughput parallel sequencing for fitness and genetic interaction studies in microorganisms. *Nature Methods* 2009 6:10, 6(10):767–772, sep 2009.
- [108] Kelly M. Wetmore, Morgan N. Price, Robert J. Waters, Jacob S. Lamson, Jennifer He, Cindi A. Hoover, Matthew J. Blow, James Bristow, Gareth Butland, Adam P. Arkin, and Adam Deutschbauer. Rapid quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. *mBio*, 6(3):1–15, may 2015.
- [109] Andrew C. Adey. Tagmentation-based single-cell genomics. *Genome Research*, 31(10):1693, oct 2021.
- [110] Esteban Martínez-García, Belén Calles, Miguel Arévalo-Rodríguez, and Víctor De Lorenzo. PBAM1: An all-synthetic genetic tool for analysis and construction of complex bacterial phenotypes. *BMC Microbiology*, 11(1):1–13, feb 2011.
- [111] William S. Reznikoff. Transposon Tn5. *Annual review of genetics*, 42:269–286, 2008.
- [112] H. Schaller. The intergenic region and the origins for filamentous phage DNA replication. *Cold Spring Harbor symposia on quantitative biology*, 43 Pt 1(1):401–408, 1979.
- [113] Houyu Zhang, Ting Lu, Shan Liu, Jianyu Yang, Guohuan Sun, Tao Cheng, Jin Xu, Fangyao Chen, and Kuangyu Yen. Comprehensive understanding of Tn5 insertion preference improves transcription regulatory element identification. *NAR Genomics and Bioinformatics*, 3(4), dec 2021.
- [114] Delia Tomoiaga, Jaclyn Bubnell, Liam Herndon, and Paul Feinstein. High rates of plasmid cotransformation in *E. coli* overturn the clonality myth and reveal colony development. *Scientific Reports* 2022 12:1, 12(1):1–15, jul 2022.
- [115] K. Nordstrom and S. J. Austin. Mechanisms that contribute to the stable segregation of plasmids. *Annual review of genetics*, 23:37–69, 1989.
- [116] Richard P Novick. Plasmid incompatibility. *Microbiological Reviews*, 51(4):381–395, dec 1987.
- [117] Shengdar Q. Tsai, Zongli Zheng, Nhu T. Nguyen, Matthew Liebers, Ved V. Topkar, Vishal Thapar, Nicolas Wyvekens, Cyd Khayter, A. John Iafrate, Long P. Le, Martin J. Aryee, and J. Keith Joung. GUIDE-seq enables genome-wide profiling of off-target

cleavage by CRISPR-Cas nucleases. *Nature Biotechnology* 2014 33:2, 33(2):187–197, dec 2014.

Appendix A

Supplementary Information for Chapter 2

Table A.1: Primers used in study

Primer Name	Primer Sequence (5'to 3')
Cap_longFWD	CGCAGGAGCAGAACAAAGAGAATC
Cap_longREV	ACCGATACCACACGAACAGGTG
Cap_InvF3	AGAGACCAAAGTTCAACTGAAACG
Cap_Rev2	TACCACACGAACAGGTGTAGTT
Cap_NSF4	GCGCAGCCATAGATCTATAACTTC
Cap_Rev4	TAGTTGCCAATGACGACCACA
Cap_cDNA_FWD	GAACGGTACGAAGCTTCGATCA
Cap_cDNA_REV	GATTAACAAGCGGCCGCA
GFP_FWD	AGCAGCACGACTTCTTCAAGTCC
GFP_REV	TGTAGTTGTACTCCAGCTTGTGCC
seq_fwd	GGTGGAGCATGAATTCTACGTC
seq_LAA_rev	GCCTCTCTGGAGGTTGGTAGATAC
seq_TGGLS_rev	GTCTGTGCTGCTGCTCTGGAA
7mer_LAA_fwd	GTATCTACCAACCTCCAGAGAGGCCTAGCANNKNNKNNK NNKNNKNNKNNKGCTAACAGACAAGCAGC
7mer_TGGLS_fwd	TTCCAGAGCAGCAGCACAGACACCGGTNNKNNKNNK NNKNNKNNKNNKGGCTTAAGTCCTGCGACCGGAGAT
7mer_rev	TCAGTTGAACTTTGGTCTCTGCG
7mer_fwd	GGTACGAAGCTTCGATCAACTACG

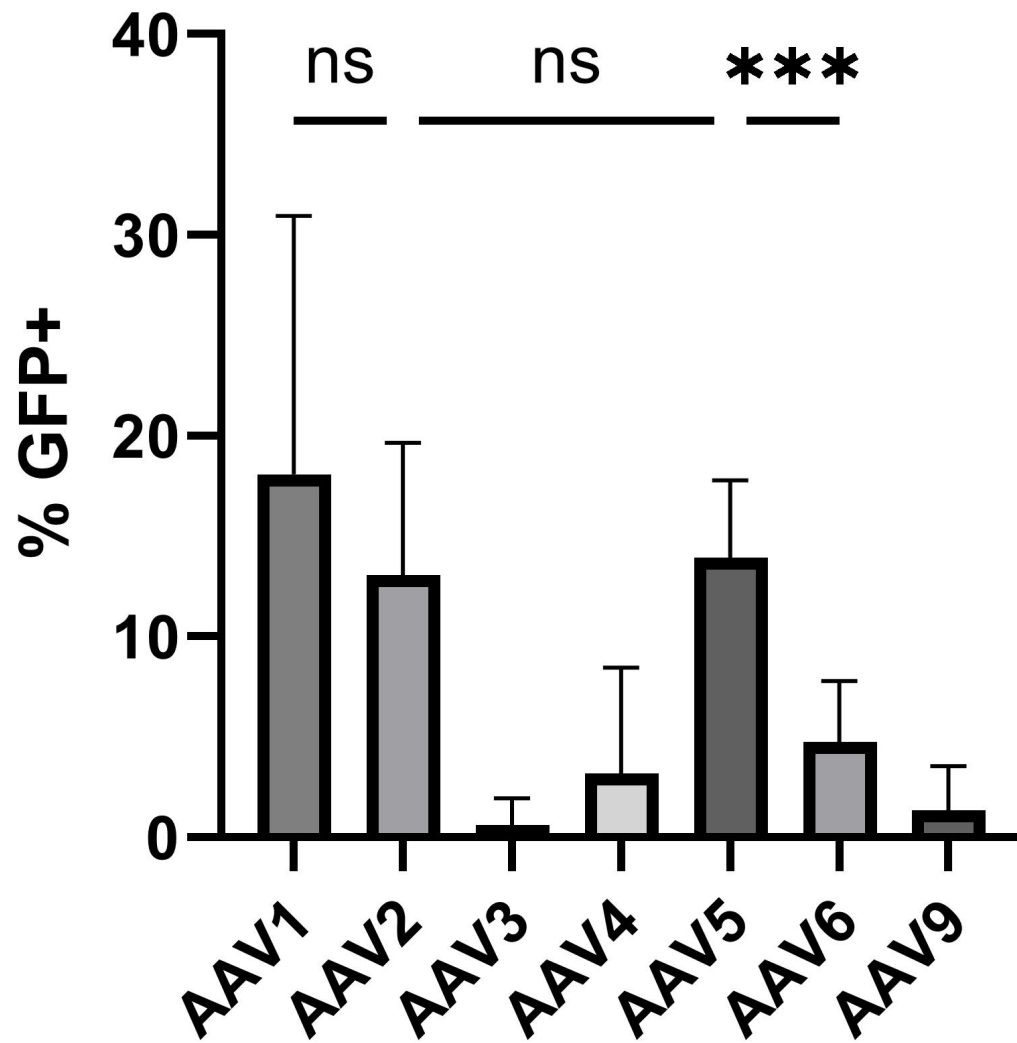


Figure A.1: Natural serotype transduction in hSCHWs. hSCHWs were transduced with each natural serotype packaged with scCAG-GFP at an MOI of 5000. After three days, cells were fixed, stained for GFP and DAPI, and imaged. Data are presented as mean $\pm$ SD, $n=3$. All statistical analyses were done with a one-way ANOVA with Tukey's multiple comparisons test (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.0001$).

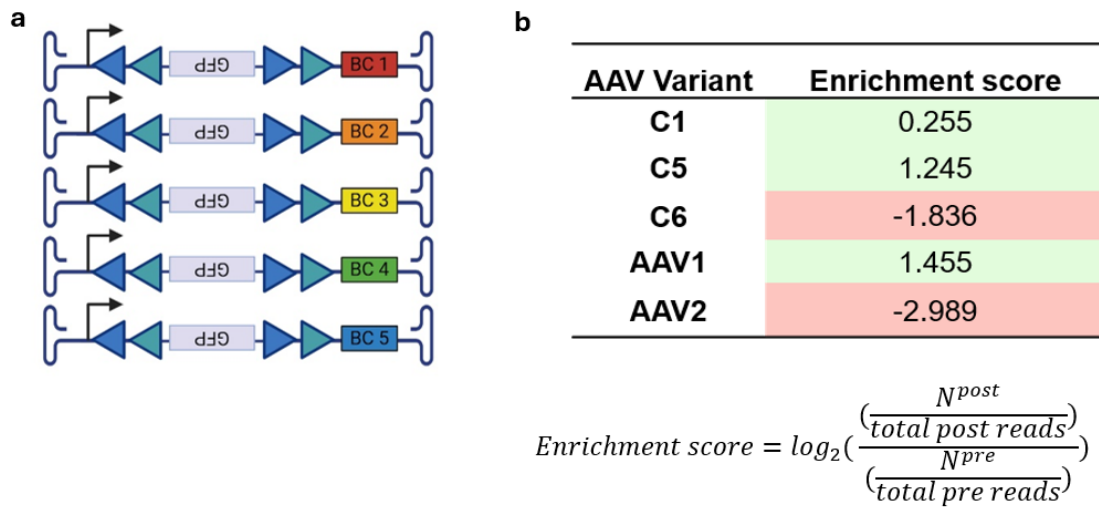


Figure A.2: Next-generation sequencing (NGS) analysis of variant enrichment in mouse Schwann cells. a) Schematic demonstrating the barcoded pAAV-FLEX-GFP plasmids and b) quantification of enrichment of each AAV in mouse Schwann cells. Enrichment was calculated as shown, where N = number of reads, pre = equimolar library pool before injection, post = harvested library after injection.

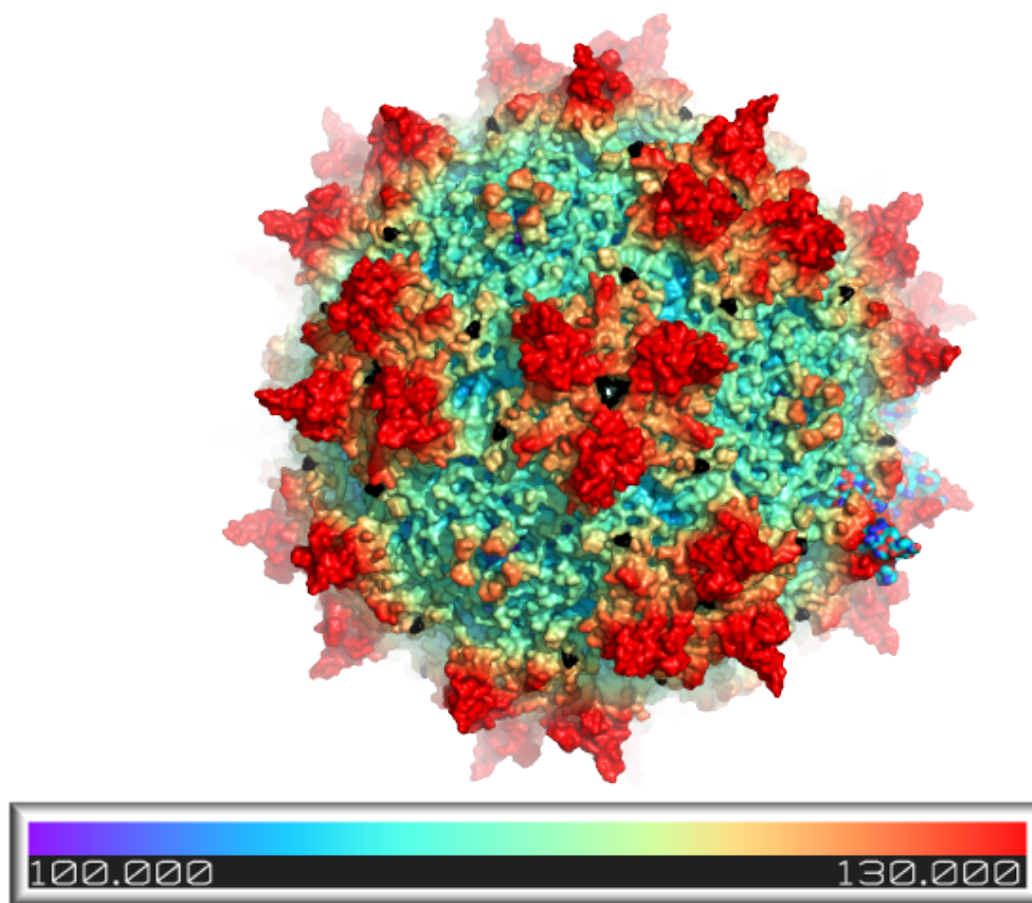


Figure A.3: Mapping of single amino acid point mutations in C1 capsid. The structure of the C1 capsid was modeled onto the AAV2 capsid (PDB: 1LP3). Distance from the center of the viral capsid is indicated by the scale bar below (Angstroms). The three point amino acid mutations are colored black.

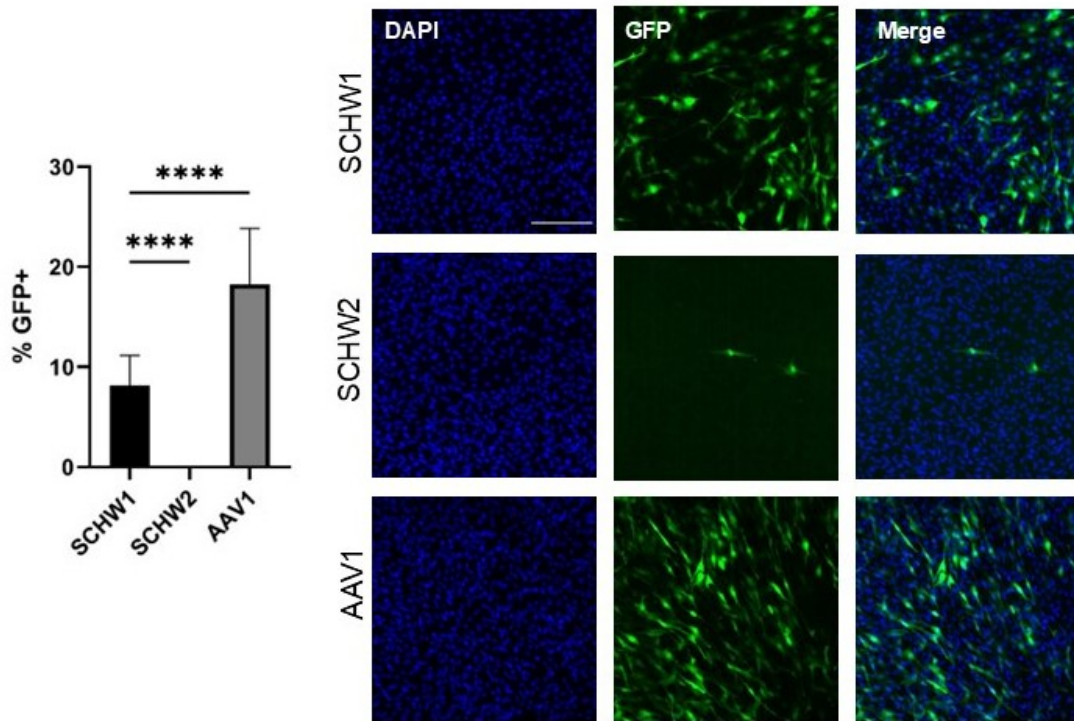


Figure A.4: 7mer variant validation in hSCHWs. hSCHWs were transduced with SCHW1, SCHW2, and AAV1 at an MOI of 50,000. After three days, cells were fixed, stained for GFP and DAPI, and imaged. Scale bar represents 100um. Data are presented as mean $\pm$ SD, n= 3. All statistical analyses were done with a one-way ANOVA with Tukey's multiple comparisons test (*p < 0.05, **p<0.01, ****p<0.0001).

Appendix B

Supplementary Information for Chapter 3

Table B.1: Primers used in study

Primer Name	Primer Sequence (5' to 3')
Tn5_BC_FWD	AATTGGAAGCTTTAGAGGTCCCC
Tn5_BC_REV	ATCCCGGAATTCCGGGTCGACCTGCAGCGTACGNNNNNNNNNNNNNNNNNNNN AGAGACCTCGTGGACATCTTCGAGCTCAAGACTGCGT
Cap_NSF4	GCGCAGCCATAGATCTATAACTTC
Cap_Rev4	TAGTTGCCAATGACGACCACA

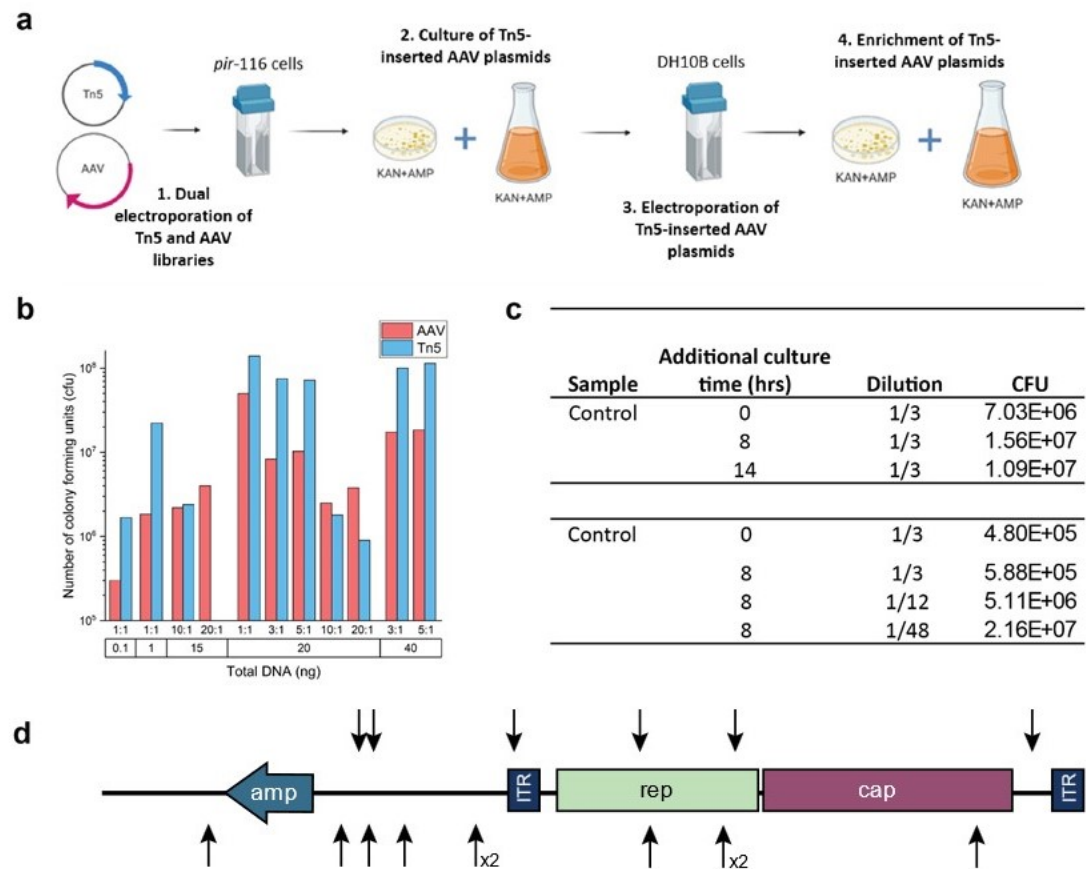


Figure B.1: Development of BLASR experimental pipeline. (a) Schematic depicting BLASR experimental pipeline. Barcoded-Tn5 and AAV plasmid libraries are electroporated into pir-116 electrocompetent *E. coli* and culture in dual-resistance LB media for 14 hours. Afterward, the culture is diluted and cultured for an additional 6 hours, and Tn5-inserted AAV plasmids are harvested. Tn5-inserted AAV plasmids are then further enriched by an additional electroporation into DH10B electrocompetent *E. coli* and culture in dual-resistance LB media. Graphic created with BioRender.com. (b) Transformation efficiencies of the dual electroporation of barcoded-Tn5 and AAV plasmid libraries at varying ratios and total amounts of DNA. (c) Optimization of dilution and additional culture time after the initial 14-hour culture to maximize the number of Tn5-inserted AAV plasmids. (d) Location of Tn5 insertions within the AAV plasmid after sampling 15 individual colonies after the second round of electroporation.

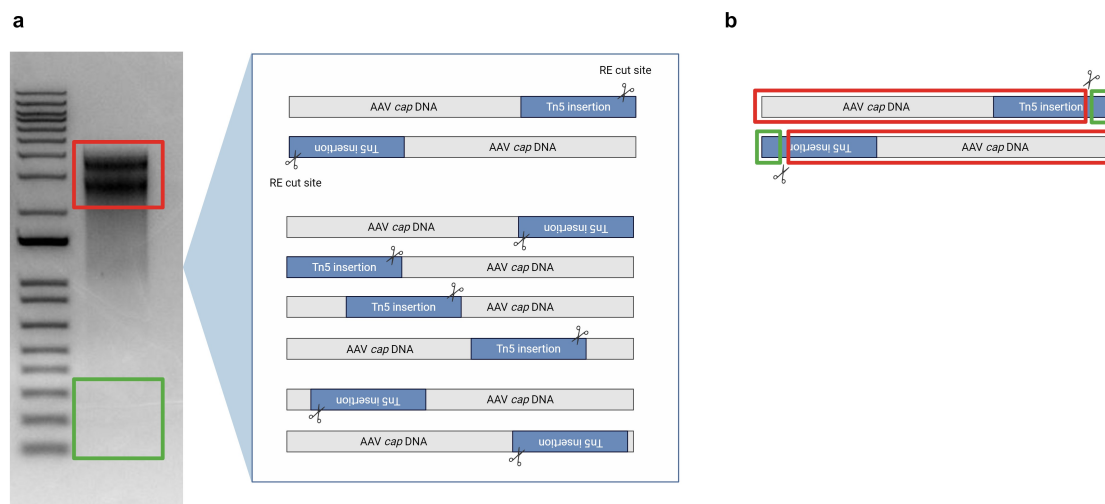


Figure B.2: Restriction enzyme digest to assess insertion bias. (a) Tn5-inserted cap sequences were digested with SpeI and run on an agarose gel. Schematic depicting example digestions of individual sequences. (b) Diagram showing suggested origin of dominant bands and potential insertion bias. While removal of the dominant bands removes biased insertion towards the ends of the cap, the “left-over” fragments under 300bp must also be removed to reduce insertion bias. Graphics created with BioRender.com.

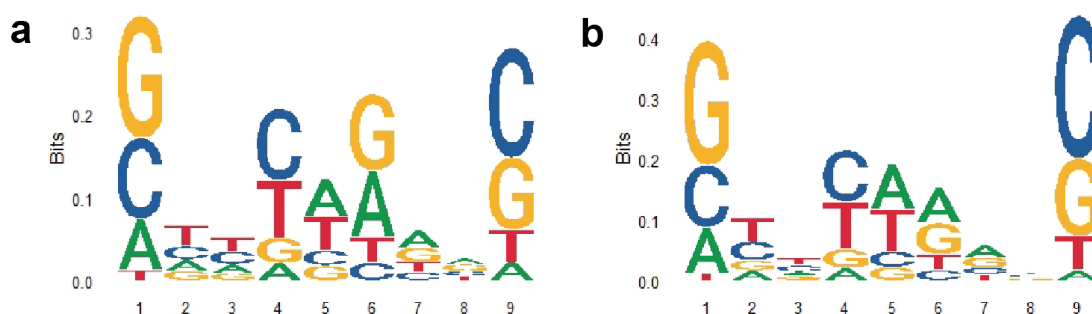


Figure B.3: Assessment of Tn5 insertion bias. (a) Sequence logo profiling the 9-bp Tn5 insertion site before optimization of Tn5 bias. (b) Sequence logo profiling the 9-bp Tn5 insertion site after optimization of Tn5 bias.

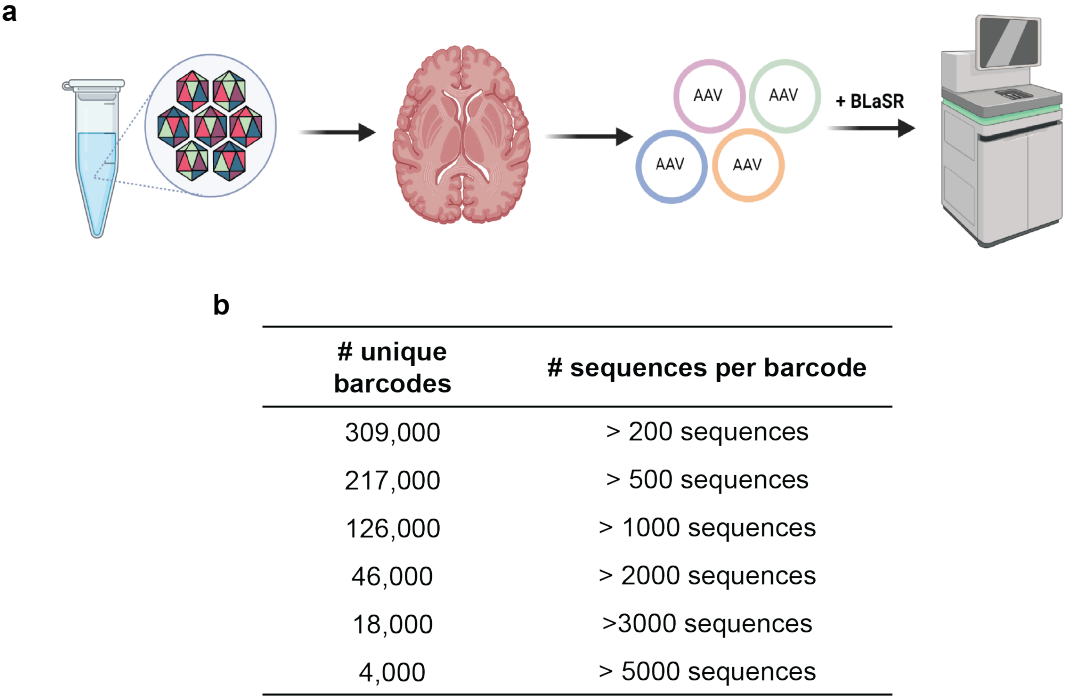


Figure B.4: Application of BLaSR in directed evolution. (a) Schematic depicting selection strategy of a diverse shuffled AAV cap library and application of BLaSR sequencing to the selected library. Graphic created with BioRender.com. (b) Number of unique barcodes that have a particular number of sequences per barcode for the selected library.