Expanding adeno-associated viral capsid engineering to multiple variable regions for diversified tropism

Thesis by

David Gerald Goertsen

In Partial Fulfillment of the Requirements for
the degree of
Doctor of Philosophy

CALIFORNIA INSTITUTE OF
TECHNOLOGY
Pasadena, California

2023
(Defended April 21, 2023)

© 2023

David Gerald Goertsen
ORCID: 0000-0001-7138-1697

ACKNOWLEDGEMENTS

I would like to acknowledge the many outstanding people who have mentored and supported me throughout my scientific career. First and foremost, I want to thank my graduate advisor, Viviana Gradinaru. Under your mentorship, I have been able to become a confident and capable scientist and engineer. You gave me the freedom and platform to explore what I was interested in, build what I thought was valuable, and share my science.

I want to thank my committee members: Professor Kaihang Wang, Professor Dave Van Valen, and Professor Justin Bois. Kaihang, your enthusiasm for science is contagious and I've always been inspired after our conversations. Dave, I was able to learn so much from your teaching and value your support as a committee chair. Justin, I was fortunate to have many opportunities to learn from you. I consider you a second mentor and you've encouraged me to develop and trust my biological intuition, to think generatively, and to appreciate the profound beauty of a well-done experiment.

I want to thank my undergraduate mentors, Professor Vikramaditya Yadav, Eric Jervis, Professor Charles Haynes, Professor Madjid Mohseni, and James Wells. Your mentorship and encouragement made me realize that graduate school was an option in my career and one that I could succeed at.

I have been very fortunate to have been mentored by some incredible people in the Gradinaru lab. Within that group, I specifically want to thank Máté Borsos, Nicholas Flytzanis, Nick Goeden, Catherine Oikonomou, and Elisha Mackey. Máté, you are hardworking and a brilliant biologist, I was lucky to learn from you and appreciate all the support you offered me. Nick Goeden, you welcomed me into the lab, encouraged me to do worthwhile projects, and gave me the template for independence. Nicholas Flytzanis, you always took the time to give me valuable and insightful advice. Catherine Oikonomou, my manuscripts and chapters are much better because of you. Elisha Mackey, thanks for doing a lot of behind the scenes work to make the lab run and for always helping me solve any problems that arose. To everyone else that the took time and effort to teach me, thank you.

Thank you to the peers that I had the fortune of working alongside in the Gradinaru Lab: Acacia Mayfield, Miggy Chuapoco, Xinhong Chen, Gerry Coughlin, Xiaozhe Ding, David Brown, Tatyana Dobreva, Umesh Padia, Michael Altermatt, Rosemary Challis, Min Jee Jang, Sripriya Kumar, Seongmin Jang, Anat Kahan, Timothy Shay, Zhe Qu, Jimin Park, Sayan Dutta, Cynthia Arokiaraj, Ryan Cho, Elliot Robinson, Mengying Zhang, Jonathan Hoang, Changfan Lin, Cameron Jackson, Nikhila Swarna, Karan Mahe, Alex Chung, Tyler Brittain, Yaping, Lei, Erin Sullivan, Damien Wolfe, Nathan Appling, Shinae Park, Nastya Grebin, Alex Wang, and Sujay Champati. You all made the lab an exciting place to do science. I also want to thank the rotation students I worked with for being patient with me as I learned how to mentor and for helping me deepen my own scientific understanding in the process: Rahma Elsiesy, Luis Caldera, and Ange-Celia Priso Fils.

I want to thank the friends I made during graduate school. I deeply appreciate each and every one of you. Kadina Johnston, since day one of bioengineering, I've been lucky to have you as one of my closest friends. Many of the times I laughed the hardest and my best memories during graduate school include you (bees and delayed calendar invites). Thanks for always being down to hang out or talk things through. Drew Honson, we connect on almost everything that I care about. I hope we can share our favorite horror movies, play Dungeons & Dragons, and solve riddles older than the Sphinx far into the future. Charles Guan (BIG CG), you were my first friend at Caltech, and I'm thankful that we were able to keep that going for 5 years! It was fun having you as a housemate, adventuring together, and experiencing many milestones with you. Acacia Mayfield, you're incredibly genuine, kind, and focused, on top of being an incredible scientist and (future) physician. I'm sorry I was rarely hungry before noon to have lunch together. Miggy Chuapoco, you're a template for work-life balance and a great example of how to live an adventurous life during graduate school. Shirin Shivaei, we made it through some tough classes together and you always made them look easy. Renée Wang, your joy and passion for equality are contagious. Patrick Almhjell, you're extremely dedicated and it's hard not to mirror that. Justin Lee, I always have the time of my life with you (even if that's rare these days).

To my friends across the world, I'll see you soon: Marcus Kliewer, Jeffrey Boschman, Nathan Chan, Melissa Beaulac, Marc Apduhan, Faye Cuadra, Shirley Zhang, Chris Wong, Mitch Syberg-Olsen, Andrew Jansen, Bernardo Espinosa, Alex Dounce, and Emman Villamejor.

Thank you to my parents, Gerald and Janine. Your support has mattered immensely to me through my academic journey, allowing me to take risks and explore my passions. Your hard work and principles were templates that I could directly follow towards my own goals to find success. I appreciate everything you do for me.

Thank you to my sisters Amy and Sophie. You're both tenacious, conscientious, persevering, and passionate. I'm proud of you both and thankful that we've been able to closer from afar. Brett and Chris, thank you for taking good care of my sisters and parents and for being the brothers I never had. I love you all and couldn't ask for a better people to call my family.

Mary Ippolito. It has been both thrilling and rewarding to complete our graduate careers alongside one another. You bring so much joy to my life. You have an incredible capacity to support everyone around you, to always say something thoughtful, to be kind in any situation, and to make life fun amongst all of that (by definition, as a lilyrib). Our time together is the exception to moderation that you always insist upon. I can't wait for our next journey together. Thanks for being you.

ABSTRACT

Adeno-associated virus research is critical for the advancement of gene therapy and treatment of myriad debilitating genetic disorders. Targeted delivery of genetic components to a tissue or cell population remains a bottleneck for gene therapy, but the selection of AAV capsids through directed evolution can yield vectors that target desired tissues or cells. This thesis details the engineering of the AAV capsid to acquire desired tropism, namely reduction in liver transduction or increased transduction of the lung. Chapter I chronicles the history of AAV engineering, provides useful information about the AAV capsid proteins, and describes how AAV has been engineered for altered tropism in works preceding this thesis. Chapter II describes the development of AAV9.452sub.LUNG1, an AAV variant that is enriched in the lung of mice after systemic injection. Chapter III details the engineering of variants with attenuated tropism in the liver while maintaining previously acquired brain transduction after systemic injection. Two of these variants, AAV.CAP-B10 and AAV.CAP-B22, display similar tropism in the marmoset after systemic injection. Chapter IV describes the parallel engineering of prominent variable regions of the AAV capsid. Overall, the work presented in this thesis expands the toolbox available for gene therapy and represents an advancement of methods for AAV capsid engineering.

PUBLISHED CONTENT AND CONTRIBUTIONS

1. **Goertsen, D.**, Goeden, N, Flytzanis, N. C., Gradinaru, V. Targeting the lung epithelium after intravenous delivery by directed evolution of underexplored sites on the AAV capsid. *Molecular Therapy - Methods & Clinical Development.* **26**, 331-342 (2022). doi:10.1016/j.omtm.2022.07.010

D.G. analyzed all data and prepared all figures with input from V.G. N.C.F and N.G. designed and performed the variant selection experiments, D.G. characterized the variants and built the image processing pipeline. D.G. wrote the manuscript with input from all authors. V.G. supervised all aspects of the work.

† denotes equal contribution

2. **Goertsen, D.**,† Flytzanis, N. C.,† Goeden, N,† Chuapoco, M. R., Cummins, A., Chen, Y., Fan, Y., Zhang, Q., Sharma, J., Duan, Y., Wang, L., Feng, G., Chen, Y., Ip, N. Y., Pickel, J., Gradinaru, V. AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset. *Nature Neuroscience.* **25**, 106–115 (2022). doi:10.1038/s41593-021-00969-4

D.G. coordinated marmomoset experiments across all research groups, analyzed marmoset tissue data, prepared all figures and data for publication, and wrote the manuscript with input from all authors.

PUBLISHED CONTENT NOT INCLUDED IN THESIS

3. Challis, R. C., Ravindra Kumar S., Chen, X., **Goertsen, D.**, Coughlin, G. M., Hori A. M., Chuapoco, M. R., Otis, T. S., Miles, T. F., Gradinaru, V. Adeno-Associated Virus Toolkit to Target Diverse Brain Cells. *Annu Rev Neurosci*. **45**, 447-469 (2022). doi: 10.1146/annurev-neuro-111020-100834

D.G. built comparison workflow for capsid sequence identity and wrote sections of the manuscript related to AAV capsid delivery to non-human primates.

4. Chen, X., Wolfe, D. A., Bindu, D. S., Zhang, M., Taskin, N., **Goertsen, D.**, Shay, T. F., Sullivan, E., Sheng-Fu, H., Ravindra Kumar, S., Arokiaraj, C. M., Plattner, V., Campos, L. J., Mich, J., Monet, D., Ngo, V., Ding, X., Omstead, V., Weed, N., Bishaw, Y., Gore, B., Lein, E. S., Akrami, A., Miller, C., Levi, B. P., Keller, A., Ting, J. T., Fox, A. S., Eroglu, C., Gradinaru, V. Function gene delivery to and across the brain vasculature of systemic AAVs with endothelial-specific tropism to rodents and broad tropism in primates. *BioRxiv.* (2023) doi:10.1101/2023.01.12.523844

D.G. conducted viral production and purification and prepared sequencing samples for *in vitro* pooled quantification experiments of AAV transduction in macaque and human brain tissue slices.

5. Shay, T. F., Sullivan, E., Ding, X., Chen, X., Ravindra Kumar, S., **Goertsen, D.**, Brown, D., Vielmetter, J., Borsos, M., Lam, A. W., Gradinaru, V. Primate-conserved Carbonic Anhydrase IV and murine-restricted Ly6c1 are new targets for crossing the blood-brain barrier. *BioRxiv.* (2023) doi:10.1101/2023.01.12.523632

D.G. developed and implemented the *in vitro* transduction quantification and plotting pipeline, performed data analysis, and prepared *in vitro* transduction quantification plots.

TABLE OF CONTENTS

Acknowledgements iii
Abstract v
Published content and contributions vi
Published content not included in thesis vii
Table of contents viii
List of figures and tables xi
Abbreviations xiii
Chapter I: Introduction 1
1.1 Overview 1
1.2 Enriching AAV expression in the lung 3
1.3 Attenuating AAV expression in the liver while maintaining brain transduction in mouse and marmoset 4
1.4 Parallel engineering of three variable regions of the AAV capsid 6
Bibliography for Chapter I 7
Chapter II: Targeting the lung epithelium after intravenous delivery by directed evolution of underexplored sites on the AAV capsid 11
Abstract 12
2.1 Introduction 13
2.2 Engineering amino acids 452-458 of AAV9 capsid proteins for lung targeting 15
2.3 A robust imaging processing pipeline to quantify transduction across whole tissues 18
2.4 AAV9.452sub.LUNG1 yields increased transgene expression in mouse lung tissue after IV delivery 20
2.5 Discussion 23
Bibliography for Chapter II 25
Appendix A: Supplementary information for Chapter II 32
A.1 Figures and tables 32
A.2 Experimental procedures 32
A.2.1 Plasmids 32
A.2.2 Viral Production 33
A.2.3 Animals 34
A.2.4 DNA/RNA recovery and sequencing 34
A.2.5 NGS data alignment and processing 36
A.2.6 Tissue preparation and immunofluorescence 36
A.2.7 Imaging and Quantification 37
A.2.8 Statistics 38
A.2.9 Data Availability 39

Bibliography for Appendix A 39
Chapter III: AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset 41
Abstract 42
3.1 Introduction 43
3.2 Engineering AAV capsids at the three-fold point of symmetry 45
3.3 Capsid engineering refines expression patterns in mice 48
3.4 AAV.CAP-B10 yields CNS-specific transgene expression in mice 49
3.5 AAV.CAP-B10 brain transgene expression is neuronal specific 50
3.6 Engineered variants maintain robust tropism in marmosets 52
3.7 Discussion 58
Bibliography for Chapter III 60
Appendix B: Supplementary information for Chapter III 64
B.1 Extended data figures 64
B.1 Extended data tables 71
B.3 Methods 72
B.3.1 General experimental methods 72
B.3.2 Viral production 73
B.3.3 Animals 73
B.3.4 DNA/RNA recovery and sequencing 76
B.3.5 NGS data alignment and processing 79
B.3.6 Tissue preparation, immunohistochemistry and immunofluorescence 80
B.3.7 Imaging and quantification 81
B.3.8 Statistics and reproducibility 83
B.3.9 Data availability 84
B.3.10 Code availability 85
Bibliography for Appendix B 85
Chapter IV: Simultaneous engineering of multiple variable regions of the adeno-associated virus capsid 86
Abstract 86
4.1 Introduction 87
4.2 Increasing the production scale of diverse libraries 89
4.3 Validation of short promoter element for DNA based capsid selection 91
4.4 Generation of diverse libraries at multiple sites 93
4.5 Discussion 98
Bibliography for Chapter IV 99
Appendix C: Supplementary information for Chapter IV 102
C.1 Figures and tables 102
C.2 Experimental procedures 109

C.2.1 Plasmids and cloning 109
C.2.2 Substitution library DNA production 110
C.2.3 Viral Production 113
C.2.4 Animals 114
C.2.5 DNA/RNA recovery and sequencing 114
C.2.5 NGS data alignment and processing 117
C.2.6 Data Availability 117
Bibliography for Appendix C 117
CONCLUSION 119

LIST OF FIGURES AND TABLES

Table Number *Page*
B-1 Systemic marmoset injections of a CNS targeting engineered viral pool 71
B-1 Systemic marmoset injections of CNS targeting engineered viral variants 71
C-1 Complementary insert oligonucleotides 109
C-2 Cloning primers 110
C-3 Single site substitution library generation primers 111
C-4 Multi-site substitution library generation primers 112
C-5 Transfection quantities per plate for library production comparison 113
C-6 Transfection quantities per plate for single site substitution libraries 113
C-7 Next generation sequencing primers 115
C-8 Read depth for Cre-dependent and T7 amplification experiments 116

Figure Number *Page*
1-1 Thesis overview 3
2-1 Capsid engineering and characterization of AAV9.452sub.LUNG1 17
2-2 Image processing pipeline for quantification of NLS-GFP expression in whole tissue sections 20
2-3 Characterization of AAV9.452sub.LUNG1 in lung tissue 22
2-4 Comparison of transgene expression efficiency across tissues after delivery with AAV5, AAV9, or AAV9.452sub.LUNG1 23
A-1 Comparison of median cell brightness in transgene expressing cells across tissues after delivery with AAV5, AAV9, or AAV9.452sub.LUNG1 32
3-1 Capsid engineering locations and CAP-B library characterization 48
3-2 AAV.CAP-B10 tissue expression profile is biased toward the brain, with a significant decrease in liver targeting 52
3-3 Characterization of pooled capsid transgene expression in NHPs 55
3-4 Characterization of single-variant expression after delivery with each of AAV9, AAV-PHP.eB, AAV.CAP-B10 and AAV.CAP-B22 in marmosets. 58
B-1 Amino-acid contributions across the substitution library 64

B-2 Quantification of decreased targeting from non-CNS organs in mice 66
B-3 Brain regions selected for cell-type quantification, depicted on brain expressing NLS-GFP after systemic delivery with AAV-PHP.eB 67
B-4 Quantification of Purkinje cell expression 68
B-5 Astrocyte specificity of AAV9, AAV.CAP-B10, and AAV.CAP-B22 characterized through immunostaining for the HA tag in conjunction with S100β 70
4-1 Comparison between T7 promoter and Cre-dependent amplification 90
4-2 AAV9 viral capsid diversity and substitution sites 93
4-3 Generating diversity across multiple variable regions of AAV9 95
4-4 Combining diversity from multiple variable regions of AAV9 97
C-1 Virus production of linearized library 102
C-2 Enrichment correlation between brain samples 103
C-3 Enrichment correlation between liver samples 104
C-4 Enrichment correlation between brain samples, excluding variants with zero measurements in either sample 105
C-5 Enrichment correlation between liver samples, excluding variants with zero measurements in either sample 106
C-6 Hamming distance of substitution libraries in multiple variable regions 107
C-7 Hamming distance of 2-site substitution libraries 107
C-8 Hamming distance of 3-site substitution libraries 108

ABBREVIATIONS

AA	amino acids
AAV	adeno-associated virus
ABCA3	ATP-binding cassette sub-family A member 3
ACE2	angiotensin-converting enzyme 2
ATI	alveolar epithelial type I
ATII	alveolar epithelial type II
A.U.	arbitrary units
BBB	blood-brain barrier
bp	base pair
CAG	cytomegalovirus early enhancer, chicken beta-actin promoter, rabbit beta-globin splice acceptor
CF	cystic fibrosis
cDNA	complementary DNA
CHAT	choline acetyltransferase, cholinergic neuronal marker
CNS	central nervous system
CREATE	Cre recombination–based AAV targeted evolution
C57/Bl6	C57 black 6 laboratory mouse strain
d	day(s)
DAPI	4′,6-diamidino-2-phenylindole, nuclear label
DNA	deoxyribonucleic acid
DRG	dorsal root ganglion
E. coli	*Escherichia coli*
ECDF	empirical cumulative density function
EGFP	enhanced green fluorescent protein
FASTQ	human-readable file format that stores the nucleotide base sequences
FXN	frataxin
GAPDH	glyceraldehyde 3-phosphate dehydrogenase, ubiquitously expressed
GFAP	glial fibrillary acidic protein, astrocyte marker
GFP	green fluorescent protein
h	hour(s)
HA	human influenza hemagglutinin
HEK 293	human embryonic kidney cells
hSyn	human synapsin, neuronal marker
IF	immunofluorescence
IPF	idiopathic pulmonary fibrosis
ITR	inverted terminal repeat
LiCl	lithium chloride
M-CREATE	Cre recombination–based AAV targeted evolution
mRNA	messenger ribonucleic acid
MPV	marmoset pool virus

MSV	marmoset single virus
NGS	next generation sequencing
NLS	nuclear localization signal
NHP	non-human primate
NS	not significant
PBS	phosphate buffer saline
PCR	polymerase chain reaction
PEG	polyethylene glycol
PEI	polyethylenimine
PFA	paraformaldehyde
P	probability of observing the test statistic being at least as extreme as what was measured if the null hypothesis is true
proSPC	prosurfactant protein C
RCA	rolling circle amplification
SE	standard error
Ss	single stranded
SP-B	surfactant protein B
SP-C	surfactant protein C
TelN	TelN proteolomerase
TelRL	TelN proteolomerase recognition site
TH	tyrosine hydroxylase, noradrenergic and dopaminergic neuronal marker
VGAT	vesicular GABA transporter, GABAergic neuronal marker
VP	viral protein
VR	variable region
WT	wild type

Chapter I

INTRODUCTION

1.1 Overview

Adeno-associated virus (AAV) are an exciting gene therapy and research vector due to their natural non-pathogenicity[1], amenability to engineering[2,3], and availability of safe and simple modes of administration[4]. Applications of AAVs in gene therapy and basic research have made them a subject of engineering efforts to tailor capsid variants to specific functions. While those efforts have expanded their application as a research tool, several limitations have impeded further success of these methods. These limitations include the following: a lack of specificity of viral capsids to specific cell-types, tissues, or organs[4,5], accumulation in immunogenic organs like the liver[6,7], failure to translate improvements to new model organisms[8], and inefficiencies of AAV library manufacturing[9]. Engineering the AAV capsid and adapting current library production methods may enable us to address these limitations and improve vector engineering for gene therapy applications.

AAV is a *Dependoparvovirus*, a genus of viruses that require co-infection by another virus to replicate and spread in a natural context[10]. *Dependoparvovirus* are more broadly classified as *Parvoviridae*, a family of viruses that have single stranded DNA genomes containing components that code replication, capsid, packaging, and egress proteins, which are flanked by telomeres which form hairpin structures important for replication[11]. If the capsid and replication components for AAV are supplied in trans, any genetic material can be packaged into the relevant AAV capsid provided that the sequence length is less than 4.7 kilobases and the flanking inverted terminal repeats are included[4]. This technique is the mechanism through which gene delivery and gene therapy can be accomplished using AAV.

The three capsid proteins of AAV are produced from a single gene by mRNA splicing and non-canonical start codons[11]. Prominent surface protrusions of the capsid occur at each threefold axis of symmetry, which are composed of two separate viral protein (VP)

monomers[12]. These surface exposed loops have been implicated in receptor binding interactions, indicating their importance for infectivity[13,14]. Natural AAV serotypes are non-specific, leading to transgene expression across a variety of tissues and cell-types after systemic administration[15]. While targeting of specific cell populations is possible with the use of regulatory elements, cell-type specific enhancers[16–19], promoters[20,21], or miRNA target sites[22,23], the limited packaging capacity of AAV capsids provide an incentive to engineering the capsid for cell-type and tissue specificity[24].

Different AAV serotypes can be compared to determine which regions of the AAV capsid are variable[25], a sign that these positions are tolerable to mutation. The variability of the surface exposed loops at the three-fold axis of symmetry is drastic between AAV serotypes[26], indicating their tractability to engineering. Introducing diversity into these variable regions (VR) of the AAV capsid has been pursued through a variety of protein engineering methods. Natural serotypes can be shuffled to create recombinant AAV containing a distinct combination of sequences[27–30]. Error prone PCR can be used on a capsid sequence to introduce mutations stochastically throughout a specific region[31–33]. Finally, amino acid sequences can be inserted or substituted into variable regions of the capsid[34–37]. Each of these techniques can be used to create a capsid library, which can then be selected for a desired phenotype. Due to surface prominence and the role of this location in heparin sulfate proteoglycan binding[38,39], amino acid insertion has typically been performed near AA588 of the capsid viral proteins[40] (VR VIII). Other surface prominent locations, such as AA452-458 (located within VR IV) and AA492-498 (located within VR V), are much less frequently mutated for selection[40], despite implication as a binding epitope for the immune system[41] and prominence of their surface protrusions[26]. Engineering at the AA588-589 site, in combination with Cre-recombinase selection *in vivo*, previously yielded AAV-PHP.B which can cross the blood-brain barrier (BBB) to transduce brain cells[34]. Mutational scanning a 3-mer along this insertion, yielding AAV-PHP.eB, further improved this phenotype to produce a powerful delivery tool for the CNS[42]. Peptide insertion within VR VIII has repeatedly yielded valuable vectors for gene therapy research[34–37].

The work in this thesis focuses on expanding engineering efforts beyond peptide insertion in VR VIII of the AAV capsid. In Chapter II and Chapter III of this thesis, variants derived from substitutions libraries between AA452 and AA458 demonstrate enriched expression in the lung or attenuation in the liver while maintaining expression in the brain (**Figure 1-1**). In Chapter IV of this thesis, VR IV and VR VIII, with and without mutations in VR V, are engineered in parallel to explore an expanded protein landscape.

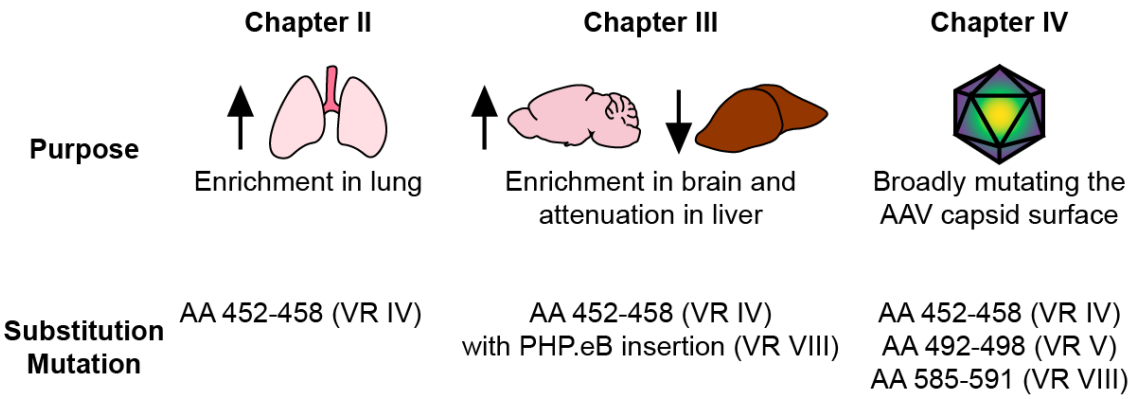

Figure 1-1. Thesis overview. The experimental aim of each thesis chapter is briefly described along with the corresponding mutations implemented in the AAV9 capsid.

1.2 Enriching AAV expression in the lung

Development of vectors capable of efficiently transducing the brain has been the primary focus of AAV capsid engineering due to the inefficiency of natural serotypes in this tissue[15], the invasiveness of direct injection[43–45], and the prevalence and severity of neurological genetic disorders[46,47]. In other organs and tissues, gene therapy efforts using AAV have continued to rely on natural serotypes. Clinical trials using AAVs have primarily focused on the CNS[34,36,37,42], eye[48,49], liver[50,51], and muscle[35,52,53], while other organs and tissues are drastically underrepresented[5]. This disparity occurs despite the prevalence and potential tractability of the genetic disorders in other tissues[54,55]. Clinical trials aiming to treat cystic fibrosis in the lung[54], for example, demonstrated safety of AAV administration, but displayed limited efficacy. The development of engineered variants enriched in lung could address this limitation.

AAV selections have typically investigated 7-mer peptide insertions within VR VIII of the AAV capsid. In Chapter II, I investigate whether other variable regions of AAV9, namely AA452-AA458 in VR IV, are amenable to substitution (**Figure 1-1**). Substitution libraries at this site proved tolerable and two rounds of *in vivo* selection for lung enrichment in mice yielded variant AAV9.452sub.LUNG1.

One barrier to robust capsid characterization was the lack of high throughput quantification methods. To characterize the *in vivo* transgene expression profile of this variant, I developed an image processing workflow to quantify the number of nuclei displaying fluorescence in whole tissue images. With this workflow, I was able to characterize the prevalence of our new variant across the major organs in comparison to natural serotypes AAV9 and AAV5. AAV9.452sub.LUNG1 displayed transgene expression in 18-fold more cells than AAV9 and 60-fold more cells than AAV5, while displaying comparable transgene expression to AAV9 in other tissues. The improved transgene expression extends to Alveolar Type II pneumocytes, an important cell type in disorders of the lung, expanding the gene therapy toolbox at the pre-clinical stage[56].

This work demonstrates two key things about capsid engineering: substitution can be used in place of insertion for AAV engineering to derive new tropism and engineering alternative sites can yield vectors with specialized functions.

1.3 Attenuating AAV expression in the liver while maintaining brain transduction in mouse and marmoset

While engineered variants AAV-PHP.B and AAV-PHP.eB have proven valuable tools for neuroscience research, the phenotype of crossing the BBB and transduction of the brain seems to be restricted to C57/BL6 mice and has not translated to non-human primates[8,57], an important criterion for the translation of these vectors to clinical applications. Additionally, these variants accumulate in the liver[15,58], which can promote an immune response and cause complications with systemic delivery.

Iterative engineering of AAV capsid variants, where mutations were scanned across previously selected residues, has been done to enhance PHP.B tropism in the brain, resulting in the variant AAV-PHP.eB. However, separate variable regions have not been engineered together to enhance or refine tropism. Due to proximity of AA452-458 to the AAV-PHP.eB insertion in the folded structure of the capsid, we questioned whether mutations in this region could facilitate further refinement of the tropism to minimize capsid accumulation in the liver or acquire tropism in new species.

In Chapter III, I clarify whether iteratively engineering surface exposed variable regions of the viral capsid could selectively alter tropism away from the liver while maintaining blood-brain barrier crossing and brain transduction (**Figure 1-1**). We created a substitution library between AA452-458 alongside the 7-mer insertion of AAV-PHP.eB between AA588 and AA589, and using the M-CREATE protocol[37], we performed differential selection in multiple Cre-recombinase expressing mouse lines to identify variants capable of brain transduction while being targeted away from peripheral organs, primarily the liver. From this selection, variants were individually validated to cross the blood brain barrier and transduce the brain while being targeting away from the liver in mice. Six variants displayed brain transduction and attenuation in the liver after systemic injection in C57/Bl6 mice.

Two variants, AAV.CAP-B10 and AAV.CAP-B22, displayed elevated RNA expression in the brain of marmosets in a pooled injection and were selected for individual characterization in marmoset. After systemic injection, each variant displays elevated transgene expression in the brain compared to AAV9 and AAV-PHP.eB, each with a distinct tropism: transgene expression after delivery with AAV.CAP-B10 was more prominent in neurons, while AAV.CAP-B22 displayed a broader tropism. Notably, transgene expression in the liver was dramatically reduced after delivery with AAV.CAP-B10. Broad and robust expression of the marmoset nervous system was attained after delivery with AAV.CAP-B10, representing an important step forward in the development of gene therapy vectors targeting the brain after systemic delivery in non-human primates.

1.4 Parallel engineering of multiple variable regions of the AAV capsid

Advances leveraging amino acid insertion to acquire new tropism has yielded exciting vectors for gene therapy and neuroscience research. However, the structural flexibility of the insertion residues may not be best suited for all receptor chemistries. Additionally, persisting residues from parental capsid sequence can cause off-target tropism, as seen in Chapter II, or neutralization by the immune system. Iterative engineering allows tropism to be refined, as seen in Chapter III, but this limits the explorable protein landscape and creates dependency on previous mutations. In Chapter IV, I parallelize the engineering of prominent residues from two or three variable regions to broadly engineer the AAV capsid surface. In addition, I characterize library production at increased transfection scales and validate a short promoter element for DNA-based selections without the requirement for Cre-transgenics.

Efforts to expand library capabilities beyond 7-mers have been impeded by production limitations. Selection of capsid libraries requires a connection between phenotype and genotype[59]. During capsid library production, concerns of cross-packaging, where an incorrect genome is incorrectly packaged due to co-transfection or serial infection, have limited transfection inputs, despite literature supporting higher transfection rates[60,61]. To investigate these concerns, production of the AAV.CAP-B library (built for the selections performed in Chapter III) was explored. These experiments indicate that capsid switching occurs in transfection inputs across several orders of magnitude and suggest that larger transfection inputs, necessary for poorly producing substitution libraries, can be used.

Viral genome based capsid selections have previously relied on transgenic animals[34,37], while recent mRNA-based selections are vulnerable to promoter bias and transcriptional noise. To avoid the dependency upon transgenic animals, amplification using a T7 promoter was validated to detect viral transduction without the necessity for Cre-transgenic animals. Additionally, 38 capsid variants were discovered to be enriched in the brain across all samples, without being enriched in the liver. DNA based selections without the necessity of Cre-recombinase transgenics could facilitate directed evolution selections of capsid libraries in new contexts (i.e. mouse strains or species where Cre-transgenics aren't available).

Additionally, combination of T7 with developed mRNA-based selections[35,36] could increase our confidence in selection data.

By implementing screening of capsid libraries for production, each of three surface exposed variable regions (VR VI, VR V, and VR VIII) of the three-fold axis were be mutated prior to selection to enable a wider search of the protein landscape. Pre-selected libraries at each site were then be combined to yield capsids with all prominent, surface-exposed residues mutated. Substitution libraries containing 7, 14, and 21 amino acid substitutions respectively were produced, and we discovered that our parallel strategy yielded diverse libraries. Mutating the AAV capsid to this extent in VR IV, VR V, and VR VIII has not been previously achieved, to our knowledge.

The development of these methods and tools could be transformative for capsid engineering through directed evolution. The ability to broadly engineer the AAV capsid across variable regions and expand the available protein landscape prior to *in vivo* selections without requiring Cre-transgenic animals could unlock the discovery of new AAV vectors with cross-species tropism or novel receptor binding partners.

Chapter I Bibliography

1. Daya, S. & Berns, K. I. Gene therapy using adeno-associated virus vectors. *Clin. Microbiol. Rev.* **21**, 583–593 (2008).
2. Koerber, J. T., Maheshri, N., Kaspar, B. K. & Schaffer, D. V. Construction of diverse adeno-associated viral libraries for directed evolution of enhanced gene delivery vehicles. **1**, 701–706 (2006).
3. Grimm, D. & Büning, H. Small But Increasingly Mighty: Latest Advances in AAV Vector Research, Design, and Evolution. *Hum. Gene Ther.* **28**, 1075–1086 (2017).
4. Bedbrook, C. N., Deverman, B. E. & Gradinaru, V. Viral Strategies for Targeting the Central and Peripheral Nervous Systems. *Annu. Rev. Neurosci.* **41**, 323–348 (2018).
5. Kuzmin, D. A. *et al.* The clinical landscape for AAV gene therapies. *Nat. Rev. Drug Discov.* **20**, 173–174 (2021).
6. Ronzitti, G., Gross, D.-A. & Mingozzi, F. Human Immune Responses to Adeno-Associated Virus (AAV) Vectors. *Front. Immunol.* **11**, 670 (2020).
7. Verdera, H. C., Kuranda, K. & Mingozzi, F. AAV Vector Immunogenicity in Humans: A Long Journey to Successful Gene Transfer. *Mol. Ther.* **28**, 723–746 (2020).

8. Hordeaux, J. *et al.* The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice. *Molecular Therapy* **26**, 664–668 (2018).
9. Sharon, D. & Kamen, A. Advancements in the design and scalable production of viral gene transfer vectors. *Biotechnol. Bioeng.* **115**, 25–40 (2018).
10. Atchison, R. W., Casto, B. C. & Hammon, W. M. Adenovirus-Associated Defective Virus Particles. *Science (80-.).* **149**, 754 LP – 755 (1965).
11. Naso, M. F. *et al.* Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs* **31**, 317–334 (2017).
12. Xie, Q. *et al.* The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc. Natl. Acad. Sci.* **99**, 10405 LP – 10410 (2002).
13. Michelfelder, S. *et al.* Peptide Ligands Incorporated into the Threefold Spike Capsid Domain to Re-Direct Gene Transduction of AAV8 and AAV9 In Vivo. *PLoS One* **6**, e23101 (2011).
14. Girod, A. *et al.* Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2. *Nat. Med.* **5**, 1052–1056 (1999).
15. Srivastava, A. In vivo tissue-tropism of adeno-associated viral vectors. *Curr. Opin. Virol.* **21**, 75–80 (2016).
16. Allen, W. E. *et al.* Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex. *Neuron* **94**, 891-907.e6 (2017).
17. Shima, Y. *et al.* A Mammalian enhancer trap resource for discovering and manipulating neuronal cell types. *Elife* **5**, e13503 (2016).
18. Graybuck, L. T. *et al.* Enhancer viruses for combinatorial cell-subclass-specific labeling. *Neuron* **109**, 1449-1464.e13 (2021).
19. Mich, J. K. *et al.* Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. *Cell Rep.* **34**, 108754 (2021).
20. Gray, S. J. *et al.* Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors. *Hum. Gene Ther.* **22**, 1143–1153 (2011).
21. de Leeuw, C. N. *et al.* Targeted CNS Delivery Using Human MiniPromoters and Demonstrated Compatibility with Adeno-Associated Viral Vectors. *Mol. Ther. Methods Clin. Dev.* **1**, 5 (2014).
22. Keaveney, M. K. *et al.* A MicroRNA-Based Gene-Targeting Tool for Virally Labeling Interneurons in the Rodent Cortex. *Cell Rep.* **24**, 294–303 (2018).
23. Hordeaux, J. *et al.* MicroRNA-mediated inhibition of transgene expression reduces dorsal root ganglion toxicity by AAV vectors in primates. *Sci. Transl. Med.* **12**, eaba9188 (2020).
24. Duan, D., Yue, Y. & Engelhardt, J. F. Expanding AAV Packaging Capacity with Trans-splicing or Overlapping Vectors: A Quantitative Comparison. *Mol. Ther.* **4**, 383–391 (2001).
25. Agbandje-McKenna, M. & Kleinschmidt, J. *AAV Capsid Structure and Cell Interactions*. (2011). doi:10.1007/978-1-61779-370-7_3
26. DiMattia, M. A. *et al.* Structural Insight into the Unique Properties of Adeno-Associated Virus Serotype 9. *J. Virol.* **86**, 6947–6958 (2012).
27. Yang, L. *et al.* A myocardium tropic adeno-associated virus (AAV) evolved by

DNA shuffling and in vivo selection. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 3946–3951 (2009).

28. Li, W. *et al.* Generation of novel AAV variants by directed evolution for improved CFTR delivery to human ciliated airway epithelium. *Mol. Ther.* **17**, 2067–2077 (2009).
29. Lisowski, L. *et al.* Selection and evaluation of clinically relevant AAV variants in a xenograft liver model. *Nature* **506**, 382–386 (2014).
30. Choudhury, S. R. *et al.* In Vivo Selection Yields AAV-B1 Capsid for Central Nervous System and Muscle Gene Therapy. *Mol. Ther.* **24**, 1247–1257 (2016).
31. Sallach, J. *et al.* Tropism-modified AAV Vectors Overcome Barriers to Successful Cutaneous Therapy. *Mol. Ther.* **22**, 929–939 (2014).
32. Perabo, L. *et al.* Combinatorial engineering of a gene therapy vector: directed evolution of adeno-associated virus. *J. Gene Med.* **8**, 155–162 (2006).
33. Maheshri, N., Koerber, J. T., Kaspar, B. K. & Schaffer, D. V. Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. *Nat. Biotechnol.* **24**, 198 (2006).
34. Deverman, B. E. *et al.* Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. *Nat. Biotechnol.* **34**, 204–209 (2016).
35. Tabebordbar, M. *et al.* Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. *Cell* **184**, 4919-4938.e22 (2021).
36. Nonnenmacher, M. *et al.* Rapid evolution of blood-brain-barrier-penetrating AAV capsids by RNA-driven biopanning. *Mol. Ther. - Methods Clin. Dev.* **20**, 366–378 (2021).
37. Kumar, S. *et al.* Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. *Nat. Methods* **17**, 541–550 (2020).
38. Kern, A. *et al.* Identification of a Heparin-Binding Motif on Adeno-Associated Virus Type 2 Capsids. *J. Virol.* **77**, 11072–11081 (2003).
39. Naumer, M. *et al.* Development and validation of novel AAV2 random libraries displaying peptides of diverse lengths and at diverse capsid positions. *Hum. Gene Ther.* **23**, 492–507 (2012).
40. Büning, H. & Srivastava, A. Capsid Modifications for Targeting and Improving the Efficacy of AAV Vectors. *Mol. Ther. Methods Clin. Dev.* **12**, 248–265 (2019).
41. Gurda, B. L. *et al.* Capsid Antibodies to Different Adeno-Associated Virus Serotypes Bind Common Regions. *J. Virol.* **87**, 9111–9124 (2013).
42. Chan, K. Y. *et al.* Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. *Nat. Neurosci.* **20**, 1172–1179 (2017).
43. Gray, S. J., Woodard, K. T. & Samulski, R. J. Viral vectors and delivery strategies for CNS gene therapy. *Ther. Deliv.* **1**, 517–534 (2010).
44. Golebiowski, D. *et al.* Direct Intracranial Injection of AAVrh8 Encoding Monkey $β$-N-Acetylhexosaminidase Causes Neurotoxicity in the Primate Brain. *Hum. Gene Ther.* **28**, 510–522 (2017).
45. Hocquemiller, M., Giersch, L., Audrain, M., Parker, S. & Cartier, N. Adeno-Associated Virus-Based Gene Therapy for CNS Diseases. *Hum. Gene Ther.* **27**,

478–496 (2016).
46. Combs, B., Kneynsberg, A. & Kanaan, N. M. Gene therapy models of alzheimer's disease and other dementias. in *Methods in Molecular Biology* **1382**, 339–366 (2016).
47. Deverman, B. E., Ravina, B. M., Bankiewicz, K. S., Paul, S. M. & Sah, D. W. Y. Gene therapy for neurological disorders: progress and prospects. *Nat. Rev. Drug Discov.* **17**, 641 (2018).
48. Öztürk, B. E. *et al.* scAAVengr, a transcriptome-based pipeline for quantitative ranking of engineered AAVs with single-cell resolution. *Elife* **10**, e64175 (2021).
49. Byrne, L. C. *et al.* In vivo-directed evolution of adeno-associated virus in the primate retina. *JCI Insight* **5**, 1–12 (2020).
50. Paulk, N. K. *et al.* Bioengineered AAV Capsids with Combined High Human Liver Transduction In Vivo and Unique Humoral Seroreactivity. *Mol. Ther.* **26**, 289–303 (2018).
51. Qian, R., Xiao, B., Li, J. & Xiao, X. Directed Evolution of AAV Serotype 5 for Increased Hepatocyte Transduction and Retained Low Humoral Seroreactivity. *Mol. Ther. - Methods Clin. Dev.* **20**, 122–132 (2021).
52. Sarcar, S. *et al.* Next-generation muscle-directed gene therapy by in silico vector design. *Nat. Commun.* **10**, 1–16 (2019).
53. Pulicherla, N. *et al.* Engineering liver-detargeted AAV9 vectors for cardiac and musculoskeletal gene transfer. *Mol. Ther.* **19**, 1070–1078 (2011).
54. Griesenbach, U., Geddes, D. M. & Alton, E. W. F. W. Gene therapy for cystic fibrosis: An example for lung gene therapy. *Gene Ther.* **11**, S43–S50 (2004).
55. Whitsett, J. A., Wert, S. E. & Weaver, T. E. Alveolar surfactant homeostasis and the pathogenesis of pulmonary disease. *Annu. Rev. Med.* **61**, 105–119 (2010).
56. Meyer-Berg, H. *et al.* Identification of AAV serotypes for lung gene therapy in human embryonic stem cell-derived lung organoids. *Stem Cell Res. Ther.* **11**, 1–6 (2020).
57. Goertsen, D. *et al.* AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset. *Nat. Neurosci.* **25**, 106–115 (2022).
58. Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV Serotypes 1–9 Mediated Gene Expression and Tropism in Mice After Systemic Injection. *Mol. Ther.* **16**, 1073–1080 (2008).
59. Doi, N. & Yanagawa, H. Genotype-phenotype linkage for directed evolution and screening of combinatorial protein libraries. *Comb. Chem. High Throughput Screen.* **4**, 497–509 (2001).
60. Schmit, P. F. *et al.* Cross-Packaging and Capsid Mosaic Formation in Multiplexed AAV Libraries. *Mol. Ther. Methods Clin. Dev.* **17**, 107–121 (2020).
61. Nonnenmacher, M., Van Bakel, H., Hajjar, R. J. & Weber, T. High capsid-genome correlation facilitates creation of AAV libraries for directed evolution. *Mol. Ther.* **23**, 675–682 (2015).

Chapter II

TARGETING THE LUNG EPITHELIUM AFTER INTRAVENOUS DELIVERY BY DIRECTED EVOLUTION OF UNDEREXPLORED SITES ON THE AAV CAPSID

Material from this chapter appears in: "**Goertsen, D.**, Goeden, N, Flytzanis, N. C., Gradinaru, V. Targeting the lung epithelium after intravenous delivery by directed evolution of underexplored sites on the AAV capsid. *Molecular Therapy - Methods & Clinical Development.* **26**, 331-342 (2022). doi:10.1016/j.omtm.2022.07.010"

D.G. analyzed all data and prepared all figures with input from V. Gradinaru. N.C.F and N.G. designed and performed the variant selection experiments. D.G. characterized the variants and built the image processing pipeline. D.G. and V.G. wrote the manuscript with input from all authors. V.G. supervised all aspects of the work.

ABSTRACT

Advances in adeno-associated virus (AAV) engineering have provided exciting new tools for research and potential solutions for gene therapy. However, the lung has not received the same tailored engineering as other major targets of debilitating genetic disorders. To address this, here we engineered the surface-exposed residues AA 452-458 of AAV9 capsid proteins at the three-fold axis of symmetry and employed a Cre-transgenic-based screening platform to identify AAV capsids targeted to the lung after intravenous delivery in mice. Using a custom image processing pipeline to quantify transgene expression across whole tissue images, we found that one engineered variant, AAV9.452sub.LUNG1, displays dramatically improved transgene expression in lung tissue after systemic delivery in mice. This improved transduction extends to alveolar epithelial type II cells, expanding the toolbox for gene therapy research for diseases specific to the lung.

2.1 Introduction

Efforts to treat genetic disorders have been aided by the development of new techniques and technologies for gene therapy. In these efforts, adeno-associated virus (AAV) has been a preferred gene therapy vector due to its non-pathogenicity,[1] stable expression *in vivo*,[2] and amenability to genome and capsid engineering.[3] The first use of AAV for gene therapy was approved in 2012 to treat lipoprotein lipase deficiency.[4] Since then, clinical trials using AAVs have overwhelmingly focused on the central nervous system (CNS), eye, liver, and muscle,[5,6] with other organs like lung drastically underrepresented.[5] This disparity occurs despite the prevalence of lung-specific genetic disorders and their potential tractability to gene therapy.[7] Clinical trials to treat cystic fibrosis (CF), for example, demonstrated the safety of AAV administration, but had limited efficacy.[7] Engineered variants with higher transduction efficiency in lung[8] could help address this limitation.

Alveolar epithelial type II (ATII) cells are of particular importance in a wide array of genetic lung conditions and lung infections due to their role in host defense and repair of respiratory tissues.[9] Impairment of ATII cells is shown to contribute to idiopathic pulmonary fibrosis (IPF),[10,11] a debilitating chronic lung disease of which current therapies only slow disease progression.[12,13] ATII cells are also a primary target for infection by SARS-CoV-2,[14,15] as they express the angiotensin converting enzyme 2 (ACE2) receptor, and their infection and subsequent cell death is associated with cases of severe COVID-19 illness.[9,14,15] Additionally, mutations in genes producing the SP-B,[16] SP-C,[17,18] and ABCA3[19,20] proteins specific to ATII cells are associated with acute respiratory failure and interstitial lung diseases[21,22] and frequently lead to respiratory failure and death in affected patients.[23] While good murine models are available for SP-B,[16] SP-C,[17,18] and ABCA3 deficiency,[19,20] the study and treatment of other lung diseases, like IPF, are hindered by a lack of suitable animal models.[13] Preclinical model development, research, and treatment options for these diseases could benefit from additional tailored gene therapy vectors specifically targeting ATII cells.

Most gene therapy efforts using AAV rely on naturally occurring serotypes with broad, overlapping tropism,[24] which can lead to decreased efficiency, safety, and applicability due

to off-target transduction and immune response.[25] To address this, substantial work has been dedicated towards developing methods to obtain AAV variants with enhanced tissue tropism and improved efficacy.[26–32] Similar to the focus of clinical trials, development of AAV variants has been devoted to the CNS,[26,31,32] liver,[33–35] eye,[36,37] and muscle.[27,38,39] The lung has received much less engineering focus,[5] with, to our knowledge, one variant displaying improved efficacy in the mouse lung (AAV6.2FF),[40,41] one transducing the endothelium of the pulmonary vasculature in the mouse lung (AAV2-ESGHGYF),[42] one efficiently transducing human airway epithelia in culture (AAV2.5T),[43] and one transducing pig airway epithelia *in vivo* (AAV2H22).[44,45] Deriving specialized variants based on different serotypes, as described herein using AAV9, could help expand treatment options, evade host immune response,[46,47] and/or target distinct cell populations.

Variable regions in surface-exposed loops at the 3-fold axis of symmetry are a locus for binding interactions with cell-surface receptors.[48–50] Usually, the AAV capsid is engineered by point mutation[51–53] or by peptide insertion within variable region VIII,[54] between amino acids (AA) 588 and 589 of AAV capsid proteins.[26,27,30,55–58] This variable region is easily amenable to peptide display[56,59] and it is responsible for heparan sulfate binding in AAV2.[50] However, in AAV9 the surface-exposed loop containing AA455 is the furthest protruding[54] and amino acid substitution at this site in conjunction with 588 insertion has been previously shown to yield productive variants displaying modified tropism.[60] Introducing diversity into this less frequently explored region of the AAV capsid proteins could promote new receptor binding interactions, facilitating the discovery of vectors capable of elevated transduction across different targets.

Currently, validation and characterization of engineered AAV variants across tissues is a bottleneck in the AAV discovery process.[61] The development of tools to consistently characterize variants across tissues will be important for validation of unique variants derived from different screens. Here, we address this bottleneck by developing an image processing pipeline to compare transgene expression from AAV variants in whole tissue images. Using this screening tool, the CREATE method[32] of positive selection, and a substitution library

based on 7-mer substitution at the underexplored 452-458 site, we derive an AAV9 variant, AAV9.452sub.LUNG1, which displays markedly improved transduction of the mouse lung, and ATII cells, after systemic injection.

2.2 Engineering amino acids 452-458 of AAV9 capsid proteins for lung targeting

Due to the prominence of the surface-exposed loop containing AA455 (**Figure 2-1A**), we theorized that the substitution of amino acids in this region could redirect viral tropism and facilitate cell receptor interactions in the lung. With this aim, we engineered and performed two rounds of selection with a 7-amino-acid substitution library in the AA455 loop, between AA452-458 (**Figure 2-1A**) in AAV9.

To derive AAV variants specific to the lung after systemic administration, we applied the CREATE method developed by Deverman and colleagues.[32] We first generated a library of AAV capsid sequences, theoretically containing all 1.28 billion 7-mer substitutions, and produced the corresponding viruses in HEK293 cells that package a replication-incompetent version of the corresponding viral genome with a polyadenylation sequence flanked by Cre/Lox sites (**Figure 2-1B**). This viral library was then intravenously injected into transgenic mouse lines (hSyn-Cre, GFAP-Cre, TH-Cre, VGAT-Cre, CHAT-Cre, Tek-Cre) which express Cre recombinase throughout the body. Variants capable of successfully transducing Cre+ cells have their genome flipped, enabling the recovery of that sequence. Two weeks post-injection, we isolated viral genomes from adipose tissue, brain, dorsal root ganglion, heart, liver, lung, kidney, quadricep muscle, ovary, pancreas, spine, spleen, and stomach. We then sequenced the variants collected from each tissue using next-generation sequencing (NGS) and calculated their enrichment, defined as the relative abundance of the sequence found within a specific tissue divided by the relative abundance of that sequence within the injected viral library.

From a first round of selection, we selected the most enriched amino acid variants from each tissue and transgenic line for a second round of selection, yielding a subset of 5,840 amino acid variants. This collection of variants was synthesized, along with codon replicates to determine the reproducibility of each variant. The variants from each tissue and transgenic

line were synthesized and produced together to avoid propagating bias due to variability in the first round of selection[31] and to simplify the selection workflow. Since high levels of Tek gene expression have been previously reported in the lung relative to other tissues,[62–66] we focused our second round of selection in Tek-Cre transgenic mice. After injecting the viral pool into Tek-Cre transgenic mice and allowing transduction and expression to occur over two weeks, we collected viral DNA from the lung, sequenced it by NGS, and ranked variants by their enrichment.

After two rounds of capsid selection, only 426 variants showed enrichment of both replicates in the lung of both animals. In the enrichment ranking, one variant clearly outperformed all other variants (**Figure 2-1C**): the amino acid motif KDNTPGR, which we call AAV9.452sub.LUNG1 and which was enriched nearly an order of magnitude more than any other variant. Retrospective analysis of the first round of selection revealed that this variant was originally included due to enrichment in the spine of hSyn-Cre animals. Additionally, the variant was present at low levels in the liver of one hSyn-Cre animal and one TH-Cre animal but was not identified in the lung of either Tek-Cre animal after the first round. These results support performing multiple rounds of selection to reduce noise and permit variants with altered tropism to become enriched within tissues of interest, as described previously.[31,32]

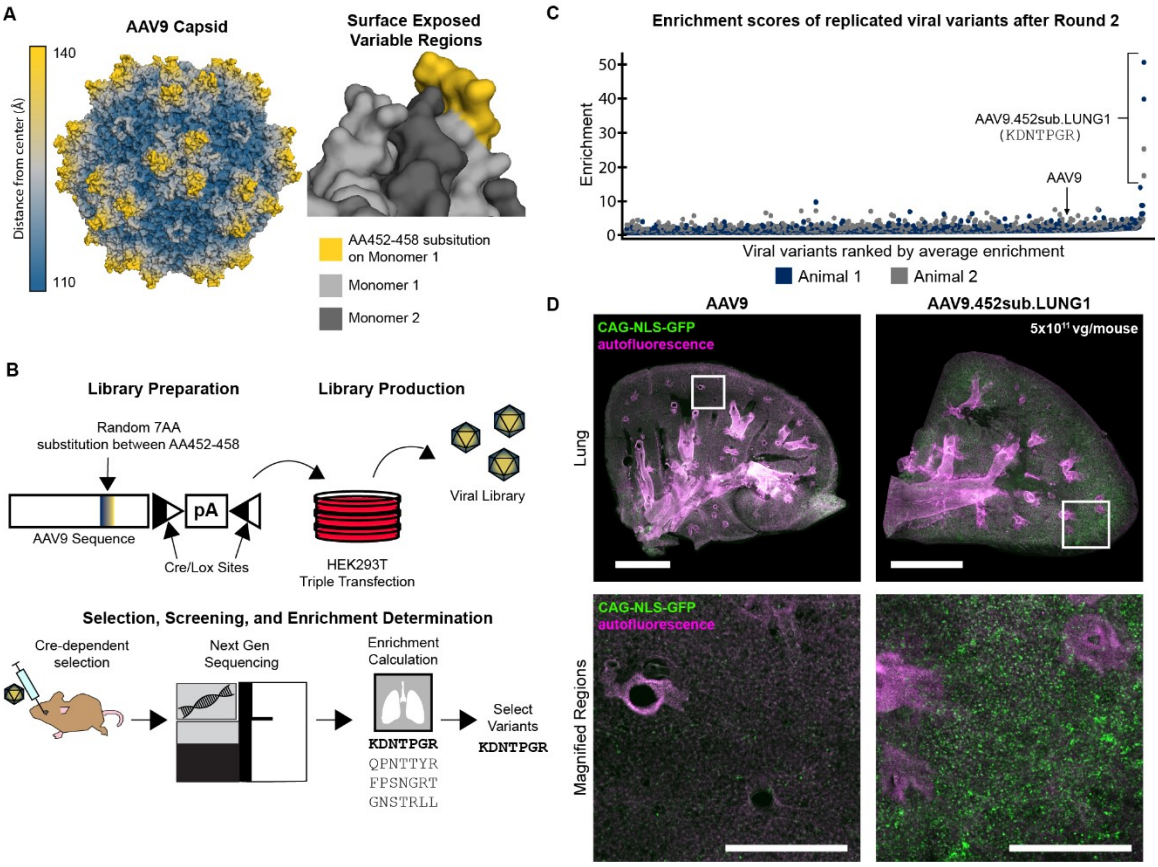

Figure 2-1. Capsid engineering and characterization of AAV9.452sub.LUNG1. (A) Left, AAV9 capsid surface model illustrates the location of the protruding loop structures, shown in yellow. Right, zoomed-in view shows the AA replaced by substitution between AA452-458. (B) To generate capsids specific to the mouse lung, a library of variants was created by mutating amino acids AA452-458 in the AAV9 backbone. Cre-lox sites were inserted into the viral genome to enable detection of genomes that reach Cre-recombinase containing cells. Variants were systemically administered via retro-orbital injection into Cre mouse lines. After two weeks of expression, relevant tissues were harvested and viral DNA isolated and sequenced. Top performing variants were subjected to a second round of selection. (C) Positively enriched variants in the lung after two rounds of selection *in vivo* were ranked by their average enrichment across codon and animal replicates. (D) NLS-GFP was packaged into the AAV9.452sub.LUNG1 variant under control of the ubiquitous CAG promoter and intravenously administered to mice at a dose of 5×10^{11} vg/animal. Transgene expression was assayed by fluorescence throughout lung. Comparison to the parent AAV9 (left) confirms the tropism measured by NGS. Scale bars in whole tissue images are 2 mm. Scale bars in magnifications are 500 µm.

2.3 A robust imaging processing pipeline to quantify transduction across whole tissues

During initial efforts to characterize this variant, we realized that robust high-throughput methods were lacking for quantification of fluorescent cells in tissue images.[67] To address this, we built a simple image processing workflow using open-source Python packages to quantify nuclear-localized (NLS-)GFP signal of whole tissue images acquired by fluorescent imaging (**Figure 2-2A, B**). Natural autofluorescence in tissues can make it hard to discriminate real signal from noise. To combat this, we also acquired images in a channel corresponding to wavelengths (565 nm) distinct from the wavelengths (495 nm) associated with our signal to allow autofluorescence subtraction. Since we were using nuclear localized fluorescence, we created a mask of nuclei by passing a ~6 μm circular template over the autofluorescence subtracted image and thresholding regions for how well they matched the template. We also smoothed the autofluorescence subtracted image by applying a gaussian blur subtraction and used this smoothed image to identify bright regions of the image by applying a histogram-based intensity threshold. We finally compared these two maps to identify regions of the image that were both bright and circular, which we counted to determine the number of NLS-GFP+ cells. Dividing this count by the total tissue area provides a measure of transgene expression per tissue area. Segmenting the initial image using the identified cell labels allows quantification of the brightness of identified cells.

We found that this pipeline could be applied to a variety of whole tissue sections with minimal parameter adjustment, improving the consistency of quantification across tissues (**Figure 2-2C**). Applying this workflow allowed single-cell resolution biodistribution profiling over 690 mm^3 of tissue volume, comprising eight tissues collected from eighteen mice, six from each of the three AAV variants.

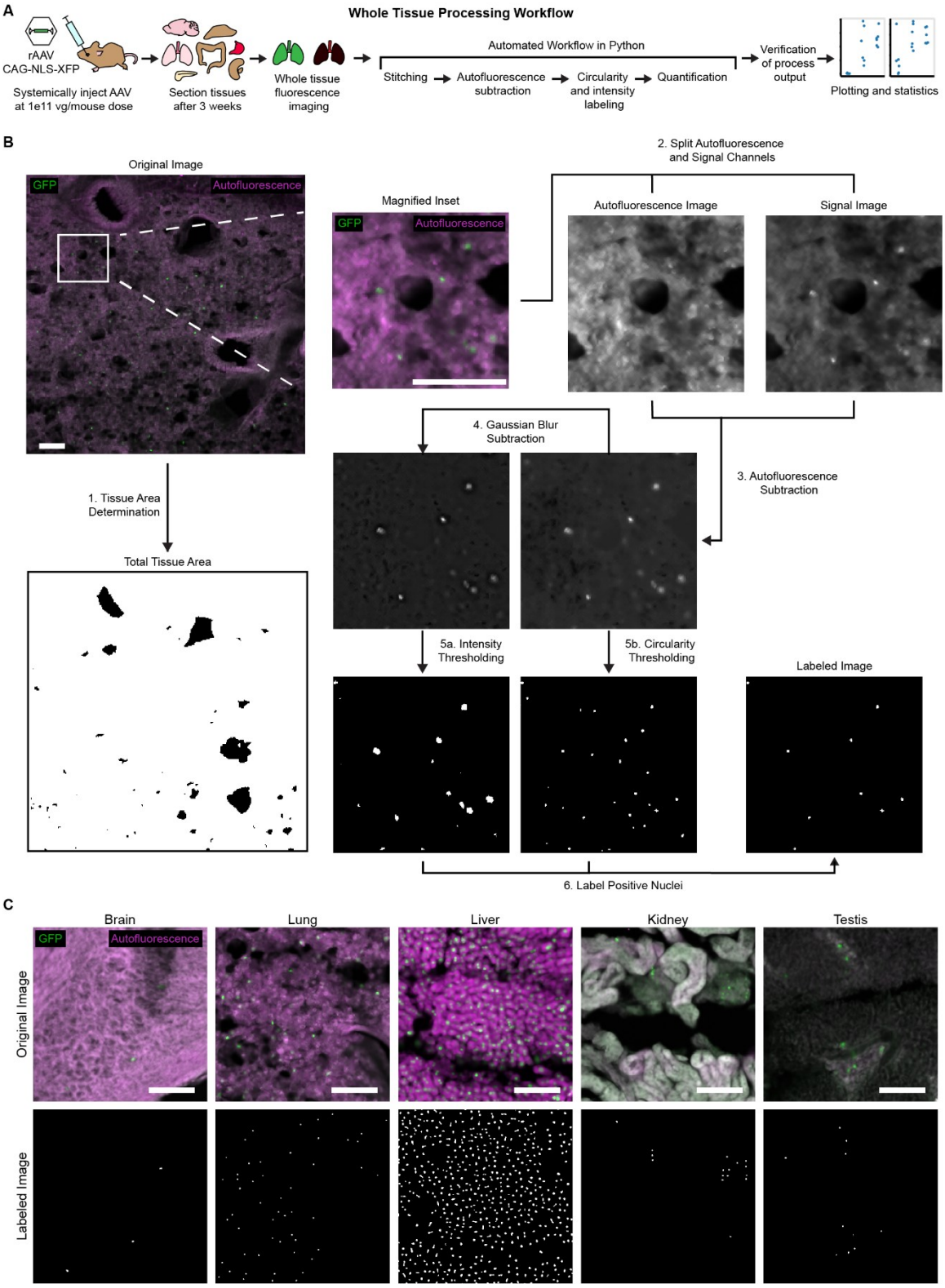
A
Whole Tissue Processing Workflow
rAAV
CAG-NLS-XFP
Systemically inject AAV at 1e11 vg/mouse dose
Section tissues after 3 weeks
Whole tissue fluorescence imaging
Automated Workflow in Python
Stitching
Autofluorescence subtraction
Circularity and intensity labeling
Quantification
Verification of process output
Plotting and statistics
B
Original Image
GFP
Autofluorescence
Magnified Inset
2. Split Autofluorescence and Signal Channels
Autofluorescence Image
Signal Image
4. Gaussian Blur Subtraction
3. Autofluorescence Subtraction
1. Tissue Area Determination
Total Tissue Area
5a. Intensity Thresholding
5b. Circularity Thresholding
Labeled Image
6. Label Positive Nuclei
C
Brain
Lung
Liver
Kidney
Testis
Original Image
Labeled Image

Figure 2-2. Image processing pipeline for quantification of NLS-GFP expression in whole tissue sections (A) Overview of the characterization experimental workflow. First, AAVs packaging nuclear localized GFP under regulation of the ubiquitous CAG promoter were systemically administered via retro-orbital injection into mice at a dose of $1x10^{11}$ viral genomes per animal. After three weeks of expression *in vivo*, tissues were collected, sectioned and imaged by fluorescence microscopy, including a non-GFP channel for autofluorescence subtraction. Whole tissue images were processed using the developed pipeline. (B) An overview of the processing steps applied to obtain cell counts from tissue images. Although only a single image is displayed here, the pipeline is run over whole tissue images. First, histogram-based thresholding determines the total tissue area. Next, autofluorescence and signal images are compared to generate a subtracted image. The circularity of signal is determined by comparison to a template to generate an image mask of circular regions. Noise is reduced by subtracting Gaussian blur and the second derivative of the histogram is used to determine an intensity threshold, which is applied to generate a mask of bright regions of the image. This is combined with the circular region mask to obtain a mask of cells positive for fluorescent nuclear signal. Scale bars, 100 µm. (C) Identification of nuclear localized fluorescent signal is shown for brain, lung, liver, kidney, and testis, with minimal parameter adjustment between tissues. Scale bars, 100 µm.

2.4 AAV9.452sub.LUNG1 yields increased transgene expression in mouse lung tissue after IV delivery

To validate the sequencing data from the Cre-recombinase selections, we produced both AAV9 and AAV9.452sub.LUNG1 packaging NLS-GFP under regulation of the strong, ubiquitous CAG promoter and intravenously injected each into mice at a dose of 5×10^{11} viral genomes per animal to test the transduction efficacy in lung tissue after systemic injection. Tissue was collected after two weeks of expression and compared to AAV9 (**Figure 2-1D**). The protein expression we observed in the lung was consistent with the NGS enrichments, with many more GFP positive cells with AAV9.452sub.LUNG1 than with parental AAV9.

To further characterize AAV9.452sub.LUNG1 (**Figure 2-2A**), we compared the transgene expression efficiency to AAV9 and AAV5, a natural serotype previously identified as promising for lung delivery.[68] We produced each virus packaging a single stranded genome expressing CAG-NLS-GFP. These variants were intravenously injected into mice at a dose of 1×10^{11} viral genomes per animal and tissue was collected and sectioned after three weeks of expression. A lower dose was selected for characterization to ensure that transduction did

not become saturated for quantification. Whole tissue sections from the lung, liver, brain, kidney, pancreas, spleen, intestine, and testis were imaged and compared between the three serotypes.

Qualitatively, we saw a substantial increase in AAV9.452sub.LUNG1 transgene expression in the lung compared to AAV9 and AAV5 (**Figure 2-3A**). Quantification of GFP+ nuclei per tissue area (**Figure 2-3B**) showed that AAV9.452sub.LUNG1 displays much higher expression in the lung than either AAV9 (18-fold) or AAV5 (60-fold). Both AAV9 and AAV9.452sub.LUNG1 display increased brightness relative to AAV5, with comparable brightness between AAV9 and AAV9.452sub.LUNG1 populations. We measured transgene expression in alveolar type I (ATI) and alveolar type II (ATII) cells by co-staining with α-podoplanin or α-prosurfactant protein C (proSPC), respectively (**Figure 2-3C**). We observed very little co-localization between NLS-GFP expression and ATI cell markers, while much of the transgene expression was co-localized with ATII cells (**Figure 2-3C**). Quantification of transgene expression efficiency in ATII cells revealed that AAV9.452sub.LUNG1 transduced ~30-fold more ATII cells than AAV9. Quantifying the fraction of GFP+ nuclei co-localized with α-proSPC indicated that the relative fraction of ATII cells expressing transgene was comparable between AAV5, AAV9, and AAV9.452sub.LUNG1. This suggests that the increased ATII transgene expression is not due to increased cell type specificity but is instead a result of increased overall transgene expression efficiency across the whole lung.

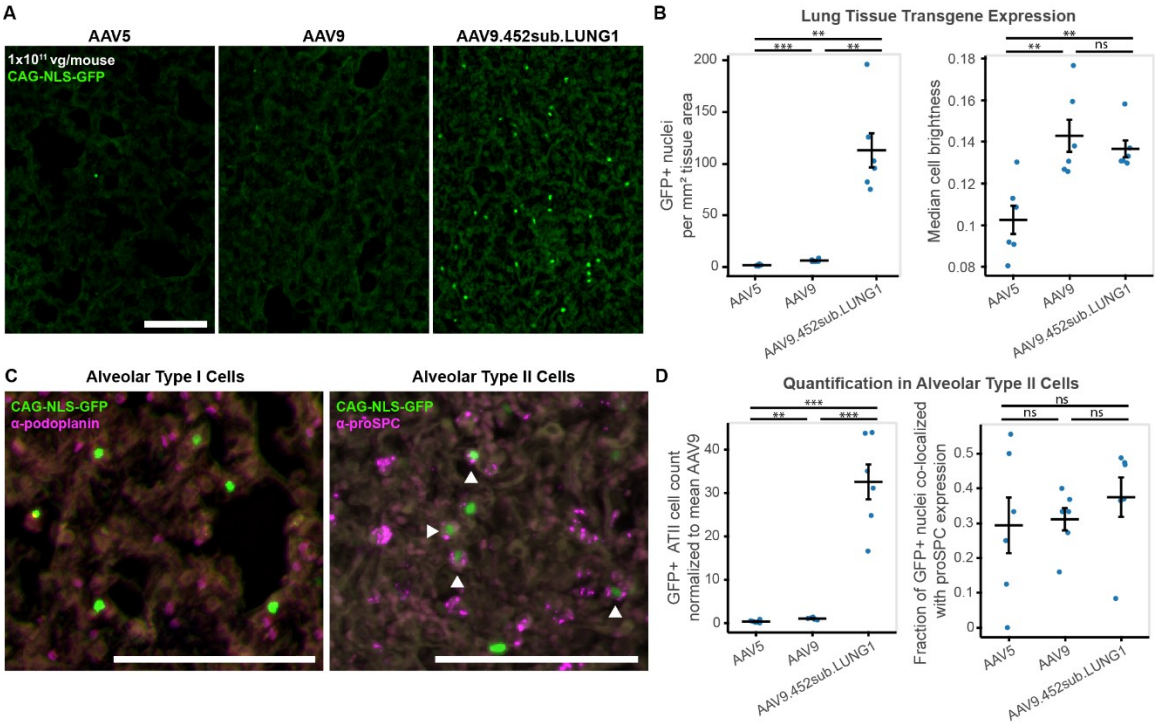

Figure 2-3. Characterization of AAV9.452sub.LUNG1 in lung tissue (A) Qualitative comparison of transgene expression in the lung shows that there are many more cells expressing transgene after delivery with AAV9.452sub.LUNG1 than AAV5 or AAV9. Scale bar, 100 µm. (B) Quantification of NLS-GFP expression indicate that AAV9.452sub.LUNG1 displays ~18 and ~60-fold increased transgene expression efficiency compared to AAV9 (P = 0.002) and AAV5 (P = 0.002), respectively. Comparing natural serotypes AAV9 and AAV5, AAV9 yielded GFP expression in ~3-fold more cells than AAV5 (P = 0.00004). AAV9 and AAV9.452sub.LUNG1 exhibited comparable brightness per cell (P = 0.5), while both were brighter on average than AAV5 (AAV9: P = 0.004; AAV9.452sub.LUNG1: P = 0.005). (C) Cell-type specific antibodies show the specificity of AAV9.452sub.LUNG1. Minimal colocalization was observed with alveolar type I cells, while there was clear colocalization with alveolar type II cells (identified with arrowheads). Scale bars are 100 µm. (D) Quantification of co-localization between NLS-GFP and α-proSPC staining indicates that AAV9.452sub.LUNG1 transduced ~30 fold more ATII cells than AAV9 (P = 0.0007) and ~95-fold more than AAV5 (P = 0.0008). This does not appear to be due to increased specificity towards ATII cells, as the relative number of GFP+ cells that co-localized with α-proSPC was unchanged compared to AAV9 (P = 0.4) and AAV5 (P = 0.5). n = 6 mice per group, mean ± s.e.m. Statistical significance was determined using two-sided Welch's t-tests.

We also measured the number of GFP+ nuclei per tissue area and brightness per cell in the brain, liver, kidney, pancreas, spleen, intestine, and testis (**Figure 2-4** and **Appendix A**, **Figure A-1**). Quantification indicated that transgene expression of AAV9.452sub.LUNG1

was comparable to AAV9 except in the two tissues with the lowest transduction frequency: pancreas and brain. AAV9.452sub.LUNG1 transgene expression in the pancreas was significantly lower than AAV9, while in the brain it was significantly higher.

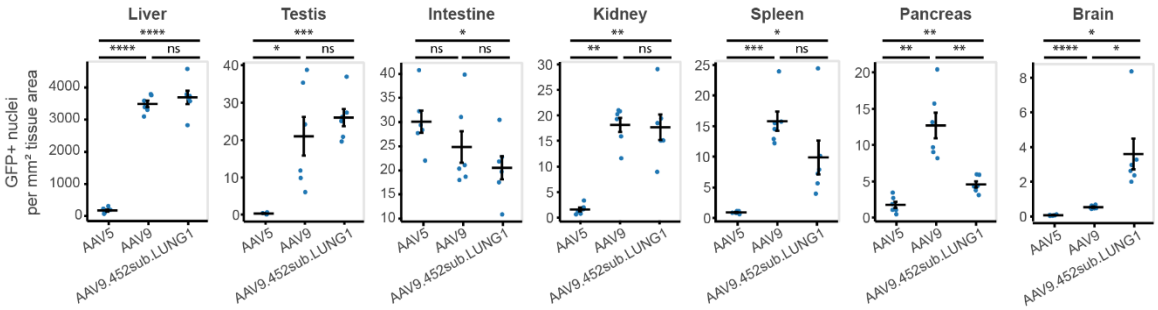

Figure 2-4. Comparison of transgene expression efficiency across tissues after delivery with AAV5, AAV9, or AAV9.452sub.LUNG1 Expression of NLS-GFP was measured across the liver, testis, intestine, kidney, spleen, pancreas, and brain, with tissues ordered from highest to lowest expression efficiency. Quantification indicates that AAV9.452sub.LUNG1 was not enriched in other tissues in comparison to AAV9, except in the brain where it expressed transgene in ~7-fold more cells (P = 0.02). In the pancreas, AAV9.452sub.LUNG1 expressed GFP in fewer cells than AAV9 (P = 0.007). In all other tissues measured, there was no significant difference in the number of cells expressing GFP between AAV9 and AAV9.452sub.LUNG1 (Liver: P = 0.5; Testis: P = 0.4; Intestine: P = 0.4; Kidney: P = 0.9; Spleen: P = 0.1). Comparing natural serotypes AAV9 and AAV5, AAV9 yielded GFP expression in more cells than AAV5 in all tissues (Liver: P < 0.00001; Testis: P = 0.01; Kidney: P = 0.00006; Spleen: P = 0.0003; Pancreas: P = 0.002; Brain: P = 0.00004) except intestine, where there was no significant difference (P = 0.3). n = 6 mice per group, mean ± s.e.m. Statistical significance was determined using two-sided Welch's t-tests. See also **Appendix A**, **Figure A-1**.

2.5 Discussion

The lung is a critical target for gene therapy. However, very few clinical trials aimed at treating diseases in this organ using AAV have been conducted,[5,6] likely due to the lack of specialized gene therapy vectors targeting this tissue. The power of directed evolution for the development of effective AAV variants with diverse tropism has been well demonstrated in other tissues.[26,27,30,32] Here, we extend and adapt these methods to the lung by substitution in non-traditional sites of the AAV capsid to discover AAV9.452sub.LUNG1, a variant capable of improving transduction efficiency in the lung by ~18-fold and expressing transgene in ~30-fold more ATII cells after systemic delivery compared to AAV9.

We selected this variant in a Tek-Cre mouse line. Although high Tek gene expression levels have been measured in the mouse lung,[62–66] recent single cell transcriptomic analysis[69] did not identify Tek expression in the ATII cell population. The improved transduction of this cell type that we observed could be due to non-specific expression of Cre-recombinase in the mouse lung. Expression profiling in the adult respiratory system of the Cre-transgenic line[70] indicates that Cre-recombinase is abundant in the alveolus, suggesting that selection of ATII transduction was likely an unintended result of performing selections in the Tek-Cre transgenic line.

Compared with its overall improved transduction in the lung, AAV9.452sub.LUNG1 exhibited similar biodistribution to AAV9 in other tissues. The significant transgene expression in the liver mediated by AAV9 has been associated with toxicity[71–73] and should be noted for future work with this vector. Natural serotypes with tropism towards the lung after system administration, such as AAV4 or AAV6,[74] could be considered as alternatives during capsid evaluation.

The improved transduction yielded in the lung by AAV9.452sub.LUNG1 could enable the development of therapeutic options for lung-specific genetic disorders which currently lack effective treatments, specifically diseases where ATII cells play a central role, such as idiopathic pulmonary fibrosis (IPF) or acute respiratory distress syndrome. Animal models which more closely resemble the progression of IPF in humans are needed,[13] and the efficient transduction of ATII cells using this variant could enable their production. For example, induction of pulmonary fibrosis in mice through ATII targeting and expression of Fas-associated protein with death domain[75,76] could provide a useful pre-clinical model. However, we note that AAV9.452sub.LUNG1 was both derived and characterized in C57BL/6J mice, and transduction in other strains and species has not yet been measured. In addition, vector biodistribution has only been tested after systemic administration, so for developing disease models, alternative modalities of delivery such as intranasal[77,78] and intratracheal[79] administration could reduce off-target accumulation of viral particles in the liver or the CNS.

During the course of characterizing this AAV variant, we ran into the bottleneck of quantifying transgene expression in whole tissue images. Manual counting is subject to bias and using overall fluorescence as a proxy for the number of transduced cells can lead to data misinterpretation. These concerns led us to develop a Python-based image processing pipeline that quantifies the number, and brightness, of cells displaying nuclear localized fluorescence per tissue area. This pipeline can be applied across tissue types with minimal adjustment, allowing unbiased and high-throughput comparison of variants. We make our code publicly available in the hope that it can aid other researchers in characterizing variants in whole tissue. Using this method, we were able to characterize the transgene expression in whole tissue images across many major organs.

Other promising, previously-engineered AAV variants with lung tropism are derived from different serotypes than AAV9.452sub.LUNG1 (AAV2-ESGHGYF,[42] AAV2.5T,[43] and AAV2H22[44] from AAV2 , and AAV6.2FF[40] from AAV6). The distinct antigenic footprints of these serotypes could facilitate sequential administration to avoid host immune response.[47] A wide array of capsid options is particularly valuable when a treatment plan requires different interventions at different time-points. The mutations in these engineered capsids are also distinct and could display further efficacy when combined, especially since one of the mutations in AAV6.2FF was previously shown to modestly improve lung targeting in AAV9.[51] Future iterations of engineering with the KDNTPGR substitution could also investigate natural serotypes AAV4 and AAV6 as templates, potentially heightening their natural tropism towards lung.[74] Ultimately, AAV9.452sub.LUNG1 expands the toolbox for gene delivery to the lung and displays promise in enabling lung-specific premedical research to inform gene therapy.

Chapter II Bibliography

1. Daya, S. & Berns, K. I. Gene therapy using adeno-associated virus vectors. Clin. Microbiol. Rev. 21, 583–593 (2008).
2. Rivière, C., Danos, O. & Douar, A. M. Long-term expression and repeated administration of AAV type 1, 2 and 5 vectors in skeletal muscle of immunocompetent adult mice. Gene Ther. 13, 1300–1308 (2006).
3. Naso, M. F., Tomkowicz, B., Perry 3rd, W. L., Strohl, W. R., Perry, W. L. & Strohl,

W. R. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. BioDrugs 31, 317–334 (2017).
4. Gaudet, D., Méthot, J. & Kastelein, J. Gene therapy for lipoprotein lipase deficiency. Curr. Opin. Lipidol. 23, 310–320 (2012).
5. Kuzmin, D. A., Shutova, M. V., Johnston, N. R., Smith, O. P., Fedorin, V. V., Kukushkin, Y. S., van der Loo, J. C. M. & Johnstone, E. C. The clinical landscape for AAV gene therapies. Nat. Rev. Drug Discov. 20, 173–174 (2021).
6. Ginn, S. L., Amaya, A. K., Alexander, I. E., Edelstein, M. & Abedi, M. R. Gene therapy clinical trials worldwide to 2017: An update. J. Gene Med. 20, e3015 (2018).
7. Griesenbach, U., Geddes, D. M. & Alton, E. W. F. W. Gene therapy for cystic fibrosis: An example for lung gene therapy. Gene Ther. 11, S43–S50 (2004).
8. Liqun Wang, R., McLaughlin, T., Cossette, T., Tang, Q., Foust, K., Campbell-Thompson, M., Martino, A., Cruz, P., Loiler, S., Mueller, C., et al. Recombinant AAV serotype and capsid mutant comparison for pulmonary gene transfer of α-1-antitrypsin using invasive and noninvasive delivery. Mol. Ther. 17, 81–87 (2009).
9. Ruaro, B., Salton, F., Braga, L., Wade, B., Confalonieri, P., Volpe, M. C., Baratella, E., Maiocchi, S. & Confalonieri, M. The History and Mystery of Alveolar Epithelial Type II Cells: Focus on Their Physiologic and Pathologic Role in Lung. Int. J. Mol. Sci. 22, 2566 (2021).
10. Parimon, T., Yao, C., Stripp, B. R., Noble, P. W. & Chen, P. Alveolar Epithelial Type II Cells as Drivers of Lung Fibrosis in Idiopathic Pulmonary Fibrosis. Int. J. Mol. Sci. 21, 2269 (2020).
11. Barbas-Filho, J. V, Ferreira, M. A., Sesso, A., Kairalla, R. A., Carvalho, C. R. R. & Capelozzi, V. L. Evidence of type II pneumocyte apoptosis in the pathogenesis of idiopathic pulmonary fibrosis (IFP)/usual interstitial pneumonia (UIP). J. Clin. Pathol. 54, 132–138 (2001).
12. Bonella, F., Stowasser, S. & Wollin, L. Idiopathic pulmonary fibrosis: current treatment options and critical appraisal of nintedanib. Drug Des. Devel. Ther. 9, 6407–6419 (2015).
13. Tashiro, J., Rubio, G. A., Limper, A. H., Williams, K., Elliot, S. J., Ninou, I., Aidinis, V., Tzouvelekis, A. & Glassberg, M. K. Exploring Animal Models That Resemble Idiopathic Pulmonary Fibrosis. Front. Med. 4, 118 (2017).
14. Morris, G., Bortolasci, C. C., Puri, B. K., Olive, L., Marx, W., O'Neil, A., Athan, E., Carvalho, A. F., Maes, M., Walder, K., et al. The pathophysiology of SARS-CoV-2: A suggested model and therapeutic approach. Life Sci. 258, 118166 (2020).
15. Carcaterra, M. & Caruso, C. Alveolar epithelial cell type II as main target of SARS-CoV-2 virus and COVID-19 development via NF-Kb pathway deregulation: A physio-pathological theory. Med. Hypotheses 146, 110412 (2021).
16. Clark, J. C., Wert, S. E., Bachurski, C. J., Stahlman, M. T., Stripp, B. R., Weaver, T. E. & Whitsett, J. A. Targeted disruption of the surfactant protein B gene disrupts surfactant homeostasis, causing respiratory failure in newborn mice. Proc. Natl. Acad. Sci. 92, 7794–7798 (1995).
17. Glasser, S. W., Burhans, M. S., Korfhagen, T. R., Na, C. L., Sly, P. D., Ross, G. F.,

Ikegami, M. & Whitsett, J. A. Altered stability of pulmonary surfactant in SP-C-deficient mice. Proc. Natl. Acad. Sci. U. S. A. 98, 6366–6371 (2001).
18. Glasser, S. W., Witt, T. L., Senft, A. P., Baatz, J. E., Folger, D., Maxfield, M. D., Akinbi, H. T., Newton, D. A., Prows, D. R. & Korfhagen, T. R. Surfactant protein C-deficient mice are susceptible to respiratory syncytial virus infection. Am. J. Physiol. Lung Cell. Mol. Physiol. 297, L64–L72 (2009).
19. Hammel, M., Michel, G., Hoefer, C., Klaften, M., Müller-Höcker, J., de Angelis, M. H. & Holzinger, A. Targeted inactivation of the murine Abca3 gene leads to respiratory failure in newborns with defective lamellar bodies. Biochem. Biophys. Res. Commun. 359, 947–951 (2007).
20. Fitzgerald, M. L., Xavier, R., Haley, K. J., Welti, R., Goss, J. L., Brown, C. E., Zhuang, D. Z., Bell, S. A., Lu, N., Mckee, M., et al. ABCA3 inactivation in mice causes respiratory failure, loss of pulmonary surfactant, and depletion of lung phosphatidylglycerol. J. Lipid Res. 48, 621–632 (2007).
21. Whitsett, J. A. & Weaver, T. E. Hydrophobic Surfactant Proteins in Lung Function and Disease. N. Engl. J. Med. 347, 2141–2148 (2002).
22. Whitsett, J. A., Wert, S. E. & Weaver, T. E. Alveolar surfactant homeostasis and the pathogenesis of pulmonary disease. Annu. Rev. Med. 61, 105–119 (2010).
23. Hamvas, A., Nogee, L. M., Mallory, J., Spray, T. L., Huddleston, C. B., August, A., Dehner, L. P., DeMello, D. E., Moxley, M., Nelson, R., et al. Lung transplantation for treatment of infants with surfactant protein B deficiency. J. Pediatr. 130, 231–239 (1997).
24. Srivastava, A. In vivo tissue-tropism of adeno-associated viral vectors. Curr. Opin. Virol. 21, 75–80 (2016).
25. Verdera, H. C., Kuranda, K. & Mingozzi, F. AAV Vector Immunogenicity in Humans: A Long Journey to Successful Gene Transfer. Mol. Ther. 28, 723–746 (2020).
26. Chan, K. Y., Jang, M. J., Yoo, B. B., Greenbaum, A., Ravi, N., Wu, W.-L. L., Sánchez-Guardado, L., Lois, C., Mazmanian, S. K., Deverman, B. E., et al. Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. Nat. Neurosci. 20, 1172–1179 (2017).
27. Tabebordbar, M., Lagerborg, K. A., Stanton, A., King, E. M., Ye, S., Tellez, L., Krunnfusz, A., Tavakoli, S., Widrick, J. J., Messemer, K. A., et al. Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. Cell 184, 4919-4938.e22 (2021).
28. Maheshri, N., Koerber, J. T., Kaspar, B. K. & Schaffer, D. V. Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. Nat. Biotechnol. 24, 198 (2006).
29. Davis, A. S., Federici, T., Ray, W. C., Boulis, N. M., O'Connor, D., Clark, K. R. & Bartlett, J. S. Rational design and engineering of a modified adeno-associated virus (AAV1)-based vector system for enhanced retrograde gene delivery. Neurosurgery 76, 216–225 (2015).
30. Nonnenmacher, M., Wang, W., Child, M. A., Ren, X.-Q., Huang, C., Ren, A. Z., Tocci, J., Chen, Q., Bittner, K., Tyson, K., et al. Rapid evolution of blood-brain-

barrier-penetrating AAV capsids by RNA-driven biopanning. Mol. Ther. - Methods Clin. Dev. 20, 366–378 (2021).
31. Kumar, S., Miles, T. T. F., Chen, X., Brown, D., Dobreva, T., Huang, Q., Ding, X., Luo, Y., Einarsson, P. P. H., Greenbaum, A., et al. Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. Nat. Methods 17, 541–550 (2020).
32. Deverman, B. E., Pravdo, P. L., Simpson, B. P., Kumar, S. R., Chan, K. Y., Banerjee, A., Wu, W., Yang, B., Huber, N., Pasca, S. P., et al. Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat. Biotechnol. 34, 204–209 (2016).
33. Qian, R., Xiao, B., Li, J. & Xiao, X. Directed Evolution of AAV Serotype 5 for Increased Hepatocyte Transduction and Retained Low Humoral Seroreactivity. Mol. Ther. - Methods Clin. Dev. 20, 122–132 (2021).
34. Paulk, N. K., Pekrun, K., Zhu, E., Nygaard, S., Li, B., Xu, J., Chu, K., Leborgne, C., Dane, A. P., Haft, A., et al. Bioengineered AAV Capsids with Combined High Human Liver Transduction In Vivo and Unique Humoral Seroreactivity. Mol. Ther. 26, 289–303 (2018).
35. Biswas, M., Marsic, D., Li, N., Zou, C., Gonzalez-Aseguinolaza, G., Zolotukhin, I., Kumar, S. R. P., Rana, J., Butterfield, J. S. S., Kondratov, O., et al. Engineering and In Vitro Selection of a Novel AAV3B Variant with High Hepatocyte Tropism and Reduced Seroreactivity. Mol. Ther. - Methods Clin. Dev. 19, 347–361 (2020).
36. Öztürk, B. E., Johnson, M. E., Kleyman, M., Turunç, S., He, J., Jabalameli, S., Xi, Z., Visel, M., Dufour, V. L., Iwabe, S., et al. scAAVengr, a transcriptome-based pipeline for quantitative ranking of engineered AAVs with single-cell resolution. Elife 10, e64175 (2021).
37. Byrne, L. C., Day, T. P., Visel, M., Strazzeri, J. A., Fortuny, C., Dalkara, D., Merigan, W. H., Schaffer, D. V. & Flannery, J. G. In vivo-directed evolution of adeno-associated virus in the primate retina. JCI Insight 5, 1–12 (2020).
38. Sarcar, S., Tulalamba, W., Rincon, M. Y., Tipanee, J., Pham, H. Q., Evens, H., Boon, D., Samara-Kuko, E., Keyaerts, M., Loperfido, M., et al. Next-generation muscle-directed gene therapy by in silico vector design. Nat. Commun. 10, 1–16 (2019).
39. Pulicherla, N., Shen, S., Yadav, S., Debbink, K., Govindasamy, L., Agbandje-McKenna, M. & Asokan, A. Engineering liver-detargeted AAV9 vectors for cardiac and musculoskeletal gene transfer. Mol. Ther. 19, 1070–1078 (2011).
40. van Lieshout, L. P., Domm, J. M., Rindler, T. N., Frost, K. L., Sorensen, D. L., Medina, S. J., Booth, S. A., Bridges, J. P. & Wootton, S. K. A Novel Triple-Mutant AAV6 Capsid Induces Rapid and Potent Transgene Expression in the Muscle and Respiratory Tract of Mice. Mol. Ther. - Methods Clin. Dev. 9, 323–329 (2018).
41. Kang, M. H., van Lieshout, L. P., Xu, L., Domm, J. M., Vadivel, A., Renesme, L., Mühlfeld, C., Hurskainen, M., Mižíková, I., Pei, Y., et al. A lung tropic AAV vector improves survival in a mouse model of surfactant B deficiency. Nat. Commun. 11, 3929 (2020).
42. Körbelin, J., Sieber, T., Michelfelder, S., Lunding, L., Spies, E., Hunger, A., Alawi,

M., Rapti, K., Indenbirken, D., Müller, O. J., et al. Pulmonary Targeting of Adeno-associated Viral Vectors by Next-generation Sequencing-guided Screening of Random Capsid Displayed Peptide Libraries. Mol. Ther. 24, 1050–1061 (2016).

43. Excoffon, K. J. D. A., Koerber, J. T., Dickey, D. D., Murtha, M., Keshavjee, S., Kaspar, B. K., Zabner, J. & Schaffer, D. V. Directed evolution of adeno-associated virus to an infectious respiratory virus. Proc. Natl. Acad. Sci. 106, 3865–3870 (2009).
44. Steines, B., Dickey, D. D., Bergen, J., Excoffon, K. J., Weinstein, J. R., Li, X., Yan, Z., Abou Alaiwa, M. H., Shah, V. S., Bouzek, D. C., et al. CFTR gene transfer with AAV improves early cystic fibrosis pig phenotypes. JCI insight 1, e88728–e88728 (2016).
45. Cooney, A. L., Thornell, I. M., Singh, B. K., Shah, V. S., Stoltz, D. A., McCray, P. B. J., Zabner, J. & Sinn, P. L. A Novel AAV-mediated Gene Delivery System Corrects CFTR Function in Pigs. Am. J. Respir. Cell Mol. Biol. 61, 747–754 (2019).
46. Bočkor, L., Bortolussi, G., Iaconcig, A., Chiaruttini, G., Tiribelli, C., Giacca, M., Benvenuti, F., Zentilin, L. & Muro, A. F. Repeated AAV-mediated gene transfer by serotype switching enables long-lasting therapeutic levels of hUgt1a1 enzyme in a mouse model of Crigler–Najjar Syndrome Type I. Gene Ther. 24, 649–660 (2017).
47. Calcedo, R. & Wilson, J. M. Humoral Immune Response to AAV. Front. Immunol. 4, 1–7 (2013).
48. Raupp, C., Naumer, M., Müller, O. J., Gurda, B. L., Agbandje-McKenna, M. & Kleinschmidt, J. A. The Threefold Protrusions of Adeno-Associated Virus Type 8 Are Involved in Cell Surface Targeting as Well as Postattachment Processing. J. Virol. 86, 9396–9408 (2012).
49. Shen, S., Bryant, K. D., Brown, S. M., Randell, S. H. & Asokan, A. Terminal N-Linked Galactose Is the Primary Receptor for Adeno-associated Virus 9. J. Biol. Chem. 286, 13532–13540 (2011).
50. Kern, A., Schmidt, K., Leder, C., Muller, O. J., Wobus, C. E., Bettinger, K., Von der Lieth, C. W., King, J. A. & Kleinschmidt, J. A. Identification of a Heparin-Binding Motif on Adeno-Associated Virus Type 2 Capsids. J. Virol. 77, 11072–11081 (2003).
51. Martini, S. V, Silva, A. L., Ferreira, D., Rabelo, R., Ornellas, F. M., Gomes, K., Rocco, P. R. M., Petrs-Silva, H. & Morales, M. M. Tyrosine Mutation in AAV9 Capsid Improves Gene Transfer to the Mouse Lung. Cell. Physiol. Biochem. Int. J. Exp. Cell. Physiol. Biochem. Pharmacol. 39, 544–553 (2016).
52. Zhong, L., Li, B., Mah, C. S., Govindasamy, L., Agbandje-McKenna, M., Cooper, M., Herzog, R. W., Zolotukhin, I., Warrington, K. H., Weigel-Van Aken, K. A., et al. Next generation of adeno-associated virus 2 vectors: Point mutations in tyrosines lead to high-efficiency transduction at lower doses. Proc. Natl. Acad. Sci. 105, 7827–7832 (2008).
53. Mao, Y., Wang, X., Yan, R., Hu, W., Li, A., Wang, S. & Li, H. Single point mutation in adeno-associated viral vectors -DJ capsid leads to improvement for gene delivery in vivo. BMC Biotechnol. 16, 1 (2016).
54. DiMattia, M. A., Nam, H.-J., Van Vliet, K., Mitchell, M., Bennett, A., Gurda, B. L.,

McKenna, R., Olson, N. H., Sinkovits, R. S., Potter, M., et al. Structural Insight into the Unique Properties of Adeno-Associated Virus Serotype 9. J. Virol. 86, 6947–6958 (2012).

55. Perabo, L., Büning, H., Kofler, D. M., Ried, M. U., Girod, A., Wendtner, C. M., Enssle, J. & Hallek, M. *In vitro* selection of viral vectors with modified tropism: the adeno-associated virus display. Mol. Ther. 8, 151–157 (2003).
56. Girod, A., Ried, M., Wobus, C., Lahm, H., Leike, K., Kleinschmidt, J., Deléage, G. & Hallek, M. Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2. Nat. Med. 5, 1052–1056 (1999).
57. Müller, O. J., Kaul, F., Weitzman, M. D., Pasqualini, R., Arap, W., Kleinschmidt, J. A. & Trepel, M. Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. Nat. Biotechnol. 21, 1040–1046 (2003).
58. Yu, C.-Y., Yuan, Z., Cao, Z., Wang, B., Qiao, C., Li, J. & Xiao, X. A muscle-targeting peptide displayed on AAV2 improves muscle tropism on systemic delivery. Gene Ther. 16, 953–962 (2009).
59. Michelfelder, S., Varadi, K., Raupp, C., Hunger, A., Körbelin, J., Pahrmann, C., Schrepfer, S., Müller, O. J., Kleinschmidt, J. A. & Trepel, M. Peptide Ligands Incorporated into the Threefold Spike Capsid Domain to Re-Direct Gene Transduction of AAV8 and AAV9 In Vivo. PLoS One 6, e23101 (2011).
60. Goertsen, D., Flytzanis, N. C., Goeden, N., Chuapoco, M. R., Cummins, A., Chen, Y., Fan, Y., Zhang, Q., Sharma, J., Duan, Y., et al. AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset. Nat. Neurosci. 25, 106–115 (2022).
61. Brown, D., Altermatt, M., Dobreva, T., Chen, S., Wang, A., Thomson, M. & Gradinaru, V. Deep Parallel Characterization of AAV Tropism and AAV-Mediated Transcriptional Changes via Single-Cell RNA Sequencing. Front. Immunol. 12, 730825 (2021).
62. Iwama, A., Hamaguchi, I., Hashiyama, M., Murayama, Y., Yasunaga, K. & Suda, T. Molecular Cloning and Characterization of Mouse TIE and TEK Receptor Tyrosine Kinase Genes and Their Expression in Hematopoietic Stem Cells. Biochem. Biophys. Res. Commun. 195, 301–309 (1993).
63. Taichman, D. B., Schachtner, S. K., Li, Y., Puri, M. C., Bernstein, A. & Scott Baldwin, H. A unique pattern of Tie1 expression in the developing murine lung. Exp. Lung Res. 29, 113–122 (2003).
64. Keane, T. M., Goodstadt, L., Danecek, P., White, M. A., Wong, K., Yalcin, B., Heger, A., Agam, A., Slater, G., Goodson, M., et al. Mouse genomic variation and its effect on phenotypes and gene regulation. Nature 477, 289–294 (2011).
65. Huntley, M. A., Lou, M., Goldstein, L. D., Lawrence, M., Dijkgraaf, G. J. P., Kaminker, J. S. & Gentleman, R. Complex regulation of ADAR-mediated RNA-editing across tissues. BMC Genomics 17, 61 (2016).
66. Grzenda, A., Shannon, J., Fisher, J. & Arkovitz, M. S. Timing and expression of the angiopoietin-1–Tie-2 pathway in murine lung development and congenital diaphragmatic hernia. Dis. Model. Mech. 6, 106–114 (2013).
67. Alves, S., Bode, J., Bemelmans, A.-P., von Kalle, C., Cartier, N. & Tews, B.

Ultramicroscopy as a novel tool to unravel the tropism of AAV gene therapy vectors in the brain. Sci. Rep. 6, 28272 (2016).

68. Gruntman, A. M., Mueller, C., Flotte, T. R. & Gao, G. Gene Transfer in the Lung Using Recombinant Adeno-Associated Virus. Curr. Protoc. Microbiol. 26, Unit14D.2.1-Unit14D.2.17 (2012).
69. Hurskainen, M., Mižíková, I., Cook, D. P., Andersson, N., Cyr-Depauw, C., Lesage, F., Helle, E., Renesme, L., Jankov, R. P., Heikinheimo, M., et al. Single cell transcriptomic analysis of murine lung development on hyperoxia-induced damage. Nat. Commun. 12, 1565 (2021).
70. Kisanuki, Y. Y., Hammer, R. E., Miyazaki, J., Williams, S. C., Richardson, J. A. & Yanagisawa, M. Tie2-Cre Transgenic Mice: A New Model for Endothelial Cell-Lineage Analysis in Vivo. Dev. Biol. 230, 230–242 (2001).
71. Hinderer, C., Katz, N., Buza, E. L., Dyer, C., Goode, T., Bell, P., Richman, L. K. & Wilson, J. M. Severe Toxicity in Nonhuman Primates and Piglets Following High-Dose Intravenous Administration of an Adeno-Associated Virus Vector Expressing Human SMN. Hum. Gene Ther. 29, 285–298 (2018).
72. Kubes, P. & Jenne, C. Immune Responses in the Liver. Annu. Rev. Immunol. 36, 247–277 (2018).
73. Bieghs, V. & Trautwein, C. The innate immune response during liver inflammation and metabolic disease. Trends Immunol. 34, 446–452 (2013).
74. Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV Serotypes 1–9 Mediated Gene Expression and Tropism in Mice After Systemic Injection. Mol. Ther. 16, 1073–1080 (2008).
75. Maeyama, T., Kuwano, K., Kawasaki, M., Kunitake, R., Hagimoto, N., Matsuba, T., Yoshimi, M., Inoshima, I., Yoshida, K. & Hara, N. Upregulation of Fas-signalling molecules in lung epithelial cells from patients with idiopathic pulmonary fibrosis. Eur. Respir. J. 17, 180–189 (2001).
76. Sisson, T. H., Mendez, M., Choi, K., Subbotina, N., Courey, A., Cunningham, A., Dave, A., Engelhardt, J. F., Liu, X., White, E. S., et al. Targeted Injury of Type II Alveolar Epithelial Cells Induces Pulmonary Fibrosis. Am. J. Respir. Crit. Care Med. 181, 254–263 (2010).
77. Belur, L. R., Temme, A., Podetz-Pedersen, K. M., Riedl, M., Vulchanova, L., Robinson, N., Hanson, L. R., Kozarsky, K. F., Orchard, P. J., Frey 2nd, W. H., et al. Intranasal Adeno-Associated Virus Mediated Gene Delivery and Expression of Human Iduronidase in the Central Nervous System: A Noninvasive and Effective Approach for Prevention of Neurologic Disease in Mucopolysaccharidosis Type I. Hum. Gene Ther. 28, 576–587 (2017).
78. Gadenstaetter, A. J., Schmutzler, L., Grimm, D. & Landegger, L. D. Intranasal application of adeno-associated viruses: a systematic review. Transl. Res. (2022) doi:https://doi.org/10.1016/j.trsl.2022.05.002.
79. Santry, L. A., Ingrao, J. C., Yu, D. L., de Jong, J. G., van Lieshout, L. P., Wood, G. A. & Wootton, S. K. AAV vector distribution in the mouse respiratory tract following four different methods of administration. BMC Biotechnol. 17, 43 (2017).

Appendix A

SUPPLEMENTARY INFORMATION FOR CHAPTER II

A.1 Figures

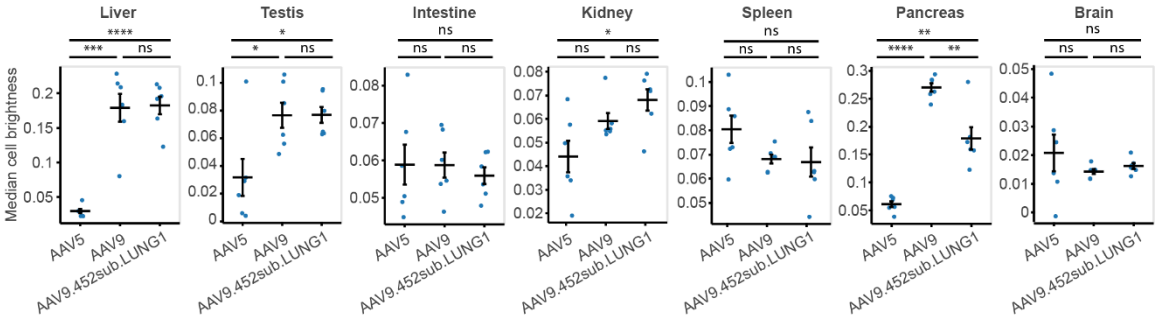

Figure A-1. Comparison of median cell brightness in transgene expressing cells across tissues after delivery with AAV5, AAV9, or AAV9.452sub.LUNG1 There was no statistically significant difference in median cell brightness between AAV9 and AAV9.452sub.LUNG1 in all tissues (Brain: P = 0.2; Liver: P = 0.9; Kidney: P = 0.2; Intestine: P = 0.5; Spleen: P = 0.9; Testis: P = 1) except pancreas, where AAV9 was brighter per cell (P = 0.007). AAV9.452sub.LUNG1 expressed transgene in more cells than AAV5 in all tissues (Brain: P = 0.01; Liver: P = 0.00002; Kidney: P = 0.001; Pancreas: P = 0.002; Spleen: P = 0.03; Testis: P = 0.0002) except the intestine, where it expressed transgene in fewer cells (P = 0.03). There was no statistically significant difference in median cell brightness between AAV5 and AAV9.452sub.LUNG1 in brain, intestine, and spleen (Brain: P = 0.5; Intestine: P = 0.7; Spleen: P = 0.2). In the liver, kidney, pancreas, and testis, the median cell brightness after delivery with AAV9.452sub.LUNG1 was increased over AAV5 (Liver: P = 0.00006; Kidney: P = 0.02; Pancreas: P = 0.002; Testis: P = 0.02). Comparing natural serotypes AAV5 and AAV9, GFP expression after delivery with AAV9 was brighter than AAV5 in liver, pancreas, and testis (Liver: P = 0.00095; Pancreas: P < 0.00001; Testis: P = 0.01) but not significantly different in brain, kidney, intestine, or spleen (Brain: P = 0.4; Kidney: P = 0.1; Intestine: P = 1; Spleen: P = 0.1). n = 6 mice per group, mean ± s.e.m. Statistical significance was determined using two-sided Welch's t-tests.

A.2 Experimental procedures

A.2.1 Plasmids

The first-round viral DNA library was generated by amplification of a section of the AAV9 capsid genome between amino acids 450-599 using NNK degenerate primers (Integrated DNA Technologies, Inc., IDT) to substitute amino acids 452-458 with all possible variations. The resulting library inserts were then introduced into the rAAV-ΔCap-in-cis-Lox plasmid

via Gibson assembly as previously described.[1] The resulting capsid DNA library, rAAV-Cap-in-cis-Lox, contained a theoretical diversity of ~1.28 billion variants at the amino acid level. The second-round viral DNA library was generated similarly to the first round, but instead of NNK degenerate primers at the 452-458 location, a synthesized oligo pool (Twist Bioscience) was used to generate only selected variants. This second-round DNA library contained a diversity of 5840 variants at the amino acid level.

The AAV2/9 REP-AAP-ΔCap plasmid transfected into HEK293T cells for library viral production was modified from the AAV2/9 REP-AAP plasmid previously used[1] by deletion of the amino acids between 450-592. This modification prevents production of a wild-type AAV9 capsid during viral library production after a plausible recombination event between this plasmid co-transfected with rAAV-ΔCap-in-cis-Lox containing the library inserts.

One rAAV genome was used in this study, pAAV-CAG-NLS-GFP (Addgene #104061), which utilizes an ssAAV genome containing the fluorescent protein EGFP flanked by two nuclear localization sites, PKKKRKV, under control of the CAG promoter.

A.2.2 Viral Production

Recombinant AAVs were generated according to established protocols.[2] Briefly, HEK293T cells (ATCC) were triple transfected using polyethylenimine (PEI); virus was collected after 120 hours from both cell lysates and media and purified over iodixanol (Optiprep, Sigma). A modified protocol was used for transfection and purification of viral libraries. First, to prevent mosaic capsid formation, only 10 ng of rAAV-Cap-in-cis-Lox library DNA was transfected (per 150 mm plate) to decrease the likelihood of multiple library DNAs entering the same cell. Second, virus was collected after 60 hours, instead of 120 hours, to limit secondary transduction of producer cells. Finally, instead of PEG precipitation of the viral particles from the media, as performed in the standard protocol, media was concentrated >60-fold for loading onto iodixanol.

A.2.3 Animals

All rodent procedures were approved by the Institutional Animal Use and Care Committee (IACUC) of the California Institute of Technology. Transgenic animals, expressing Cre under the control of various cell-type-specific promoters, and C57Bl/6J WT mice (000664) were purchased from the Jackson Laboratory (JAX). Transgenic mice included Tek-Cre (8863),[3] Syn1-Cre (3966),[4] GFAP-Cre (012886),[5] TH-Cre (008601),[6] CHAT-Cre (028861),[7] and VGAT-Cre (028862).[8] For the first round of viral library selection, we used one male and one female mouse from each transgenic line (aged 8-12 weeks), as well as a single male C57Bl/6J mouse. For the second round of selection, we used one male and female mouse from the Tek-Cre line (aged 8-12 weeks), as well as a single male C57Bl/6J mouse. For validation of individual viral variants, male C57Bl/6J mice aged 6-8 weeks were used. Mice were housed under standard conditions between 71 and 75 °F, 30% - 70% humidity, and light cycle of 13 hours on & 11 hours off. Intravenous administration of rAAV vectors was performed via injection into the retro-orbital sinus.

A.2.4 DNA/RNA recovery and sequencing

Round 1 and round 2 viral libraries were injected into C57Bl/6J and Cre-transgenic animals at a dose of 8×10^{10} vg/animal and rAAV genomes were recovered two weeks post injection, as described in the M-CREATE protocol.[9] Cre-transgenic lines used for round 1 were Syn1-Cre, GFAP-Cre, TH-Cre, VGAT-Cre, CHAT-Cre, and Tek-Cre, selected to cover a wide range of cell types across tissues. For the second round of selection, only Tek-Cre animals were used. To determine the number of variants included in round 2, 0.01 times the enrichment of the top variant in each tissue was set as a threshold and variants above that threshold were included. Mice were euthanized, and most major organs were recovered, snap frozen on dry ice, and placed into long-term storage at −80 °C. Tissues collected included: brain, spinal cord, dorsal root ganglia (DRGs), liver, lungs, heart, stomach, intestines, kidneys, spleen, pancreas, testes, skeletal muscle, and adipose tissue. 100 mg of each tissue (~250 mg for brain hemispheres, <100 mg for DRGs) was homogenized in Trizol (Life Technologies, 15596) using a BeadBug (Benchmark Scientific, D1036) and viral DNA was isolated according to the manufacturer's recommended protocol. Recovered viral DNA was

treated with RNase, underwent restriction digestion with SmaI (located within the ITRs) to improve later rAAV genome recovery by PCR, and purified with a Zymo DNA Clean and Concentrator kit (D4033). Viral genomes flipped by Cre-recombinase in select transgenic lines (or pre-flipped in WT animals) were selectively recovered using the following primers: 5'-CTTCCAGTTCAGCTACGAGTTTGAGAAC-3' and 5'-CAAGTAAAACCTCTACAAATGTGGTAAAATCG-3', after 25 cycles of 98 °C for 10 s, 60 °C for 15 s, and 72 °C for 40 s, using Q5 DNA polymerase in five 25 µl reactions with 50% of the total extracted viral DNA as a template.

After Zymo DNA purification, samples from the WT C57Bl/6J animals were serially diluted from 1:10 – 1:10,000 and each dilution further amplified around the library variable region. This amplification was done using primers: 5'-ACGCTCTTCCGATCTAATACTTGTACTATCTCTCTAGAACTATT-3' and 5'-TGTGCTCTTCCGATCTCACACTGAATTTTAGCGTTTG-3' and 10 cycles of 98 °C for 10 s, 61°C for 15 s, and 72 °C for 20 s, to recover 73 bp of viral genome around and including the 21 bp variable region and add adapters for Illumina next-generation sequencing. After PCR cleanup, these products were further amplified using NEBNext Dual Index Primers for Illumina sequencing (New England Biolabs, E7600), after 10 cycles of 98 °C for 10 s, 60 °C for 15 s, and 72°C for 20 s. The amplification products were run on a 2% low-melting-point agarose gel (ThermoFisher Scientific, 16520050) for better separation and recovery of the 210 bp band. The dilution series was analyzed for each WT tissue and the highest concentration dilution which resulted in no product was chosen for further amplification of the viral DNA from the transgenic animal tissues. This process was performed to differentiate between viral genomes flipped prior to packaging or due to Cre in the animal. Pre-flipped viral genomes should be avoided to minimize false positives in the NGS sequencing results.

All Cre-flipped viral genomes from transgenic animal tissues were similarly amplified (using the dilutions that do not produce pre-flipped viral genomes) to add Illumina sequencing adapters and subsequently for index labeling. The amplified products now containing unique

indices for each tissue from each animal were run on a low-melting-point agarose gel and the correct bands extracted and purified with a Zymoclean Gel DNA Recovery kit.

Packaged viral library DNA was isolated from the injected viral library by digestion of the viral capsid and purification of the contained ssDNA. These viral genomes were amplified by two PCR amplification steps, like the viral DNA extracted from tissue, to add Illumina adapters and then indices and extracted and purified after gel electrophoresis. This viral library DNA, along with the viral DNA extracted from tissue, was sent for deep sequencing using an Illumina HiSeq 2500 System (Millard and Muriel Jacobs Genetics and Genomics Laboratory, Caltech).

A.2.5 NGS data alignment and processing

Raw fastq files from NGS runs were processed with CREATE data analysis code (available on Github at: https://github.com/GradinaruLab/mCREATE) that align the data to an AAV9 template DNA fragment containing the 21 bp diversified region between AA 452-458, for the two rounds of AAV evolution/selection. The pipeline to process these datasets involved filtering to remove low-quality reads, utilizing a quality score for each sequence, and eliminating bias from PCR-induced mutations or high GC-content. The filtered dataset was then aligned by a perfect string match algorithm and trimmed to improve the alignment quality. For the AAV engineering, read counts for each sequence were pulled out and displayed along with their enrichment score, defined as the relative abundance of the sequence found within the specific tissue over the relative abundance of that sequence within the injected viral library.

A.2.6 Tissue preparation and immunofluorescence

Mice were euthanized with Euthasol and transcardially perfused with ice-cold 1× PBS and then freshly prepared, ice-cold 4% paraformaldehyde (PFA) in 1× PBS. All organs were excised and post-fixed in 4% PFA at 4 °C for 48 hours and then sectioned at 50 µm with a vibratome. Immunofluorescence (IF) was performed on floating sections with primary and secondary antibodies in PBS containing 10% donkey serum and 0.1% Triton X-100. Primary antibodies used were Syrian hamster anti-podoplanin (1:200, Abcam, ab11936) and rabbit

anti-Prosurfactant Protein C (1:200, Sigma-Aldrich, AB3786). Primary antibody incubations were performed for 16–20 hours at room temperature. Sections being stained with anti-podoplanin were washed prior to incubation with secondary antibody AffiniPure Goat Anti-Syrian Hamster IgG (1:200, Jackson ImmunoResearch Laboratories, Inc., 107-005-142) for 6-hours at room temperature. This was followed by incubation with Alexa-647 conjugated donkey anti-goat antibody (1:200, ThermoFisher Scientific, A32849) for 6–8 hours at room temperature. Sections being stained with anti-Prosurfactant Protein C were washed and incubated with secondary Alexa-647 conjugated anti-rabbit FAB fragment antibody (1:200, Jackson ImmunoResearch Laboratories, Inc., 711-607-003) for 6–8 hours at room temperature. Stained sections were then mounted with ProLong Diamond Antifade Mountant (ThermoFisher Scientific, P36970).

A.2.7 Imaging and Quantification

Initial comparisons of transgene expression after delivery with AAV9.452sub.LUNG1 or AAV9 at a dose of 5 x 10^{11} viral genomes per animal (**Figure 2-1**) were imaged on a Zeiss LSM 880 confocal microscope using a Fluar 5× 0.25 M27 objective, with matched laser powers, gains, and gamma across all samples of the same tissue. The acquired images were processed in Zen Black 2.3 SP1 (Zeiss).

Comparisons of AAV9.452sub.LUNG1 with AAV9 and AAV5 at the lower dose of 1 x 10^{11} viral genomes per animal were imaged on a Keyence BZ-X all-in-one fluorescence microscope at 48-bit resolution with the following objectives: PlanApo-λ 20x/0/75 (1 mm working distance) or PlanApo-λ 10x/0.45 (4 mm working distance). The filters used were BZ-X Filters OP-87763 for NLS-GFP, OP-87764 for autofluorescence, and OP-87766 for imaging tissues stained with 647 antibodies. In each tissue, exposure settings and changes to gamma or contrast were maintained across images.

All image processing was performed using our custom Python image processing pipeline, available at https://doi.org/10.5281/zenodo.7908822. In brief, images are first stitched together using a custom algorithm, keeping channels distinct. An Otsu threshold with 1024 bins is used on the autofluorescence image to determine the area within the image that is

classified as tissue. The mean intensity of both the signal and the autofluorescence images are determined, and the ratio of these means is used to adjust the autofluorescence image prior to subtraction, to ensure that the average image intensities are matched prior to adjustment. On the product of subtraction, template matching to a circular template is then used to determine regions that appear nuclear using skimage.feature.match_template, outputting a binary mask of all the regions matching the circular template. Also on the product of subtraction, skimage.filters.gaussian is applied (sigma = 5, truncation = 2) and subtracted from the image to remove large features from the background. From the background subtracted image, the second derivative of the histogram is then taken, and a predetermined value of the second derivative is used to acquire an intensity threshold for the image. This threshold is applied to generate a binary mask of bright regions prior to application of skimage.morphology.binary_closing to fill in partially identified intense regions. Both the circular template mask and the intensity thresholded mask are multiplied together to make a mask of colocalized regions. Finally, objects fewer than 10 pixels are removed to minimize noise. skimage.measure.label is used to determine the number of cells in this image. Custom code was used to calculate the brightness of cells within the image.

A.2.8 Statistics

Custom Python code was used for statistical analysis and data representation. Populations were compared using two-sided Welch's t-test, implemented using the stats module from the SciPy package. This module was implemented to compare each set of samples without assuming equal variance as: stats.ttest_ind(first population, second population, equal_var = False), where the first and second populations were the data being compared. Unless otherwise noted, all experimental groups were n = 6, determined using preliminary data and experimental power analysis. For the statistical analysis in mice and related graphs, a single data point was defined as two tissue sections per animal. For each section, the whole tissue was imaged and processed. For all statistical analyses, significance is represented as: *$P \leq 0.05$; **$P \leq 0.01$; ***$P \leq 0.001$; ****$P \leq 0.0001$; n.s., $P \geq 0.05$.

A.2.9 Data Availability

The NGS datasets for the capsid selection reported in this article are available under Sequence Read Archive accession code PRJNA821550. The vector plasmid for AAV9.452sub.LUNG1 was deposited in Addgene for distribution (184592). The AAV9.452sub.LUNG1 has been deposited on GenBank under Accession Number ON959566. All other constructs and tools are available through the Beckman Institute CLOVER Center (https://clover.caltech.edu/). Python code used for this analysis is available at https://doi.org/10.5281/zenodo.7908822.

Bibliography for Appendix A

1. Deverman, B. E., Pravdo, P. L., Simpson, B. P., Kumar, S. R., Chan, K. Y., Banerjee, A., Wu, W., Yang, B., Huber, N., Pasca, S. P., et al. Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat. Biotechnol. 34, 204–209 (2016).
2. Challis, R. C., Kumar, S. R., Chan, K. Y., Challis, C., Beadle, K., Jang, M. J., Kim, H. M., Rajendran, P. S., Tompkins, J. D., Shivkumar, K., et al. Systemic AAV vectors for widespread and targeted gene delivery in rodents. Nat. Protoc. 14, 379–414 (2019).
3. Kisanuki, Y. Y., Hammer, R. E., Miyazaki, J., Williams, S. C., Richardson, J. A. & Yanagisawa, M. Tie2-Cre Transgenic Mice: A New Model for Endothelial Cell-Lineage Analysis in Vivo. Dev. Biol. 230, 230–242 (2001).
4. Zhu, Y., Romero, M. I., Ghosh, P., Ye, Z., Charnay, P., Rushing, E. J., Marth, J. D. & Parada, L. F. Ablation of NF1 function in neurons induces abnormal development of cerebral cortex and reactive gliosis in the brain. Genes Dev. 15, 859–876 (2001).
5. Garcia, A. D. R., Doan, N. B., Imura, T., Bush, T. G. & Sofroniew, M. V. GFAP-expressing progenitors are the principal source of constitutive neurogenesis in adult mouse forebrain. Nat. Neurosci. 7, 1233–1241 (2004).
6. Savitt, J. M., Jang, S. S., Mu, W., Dawson, V. L. & Dawson, T. M. Bcl-x is required for proper development of the mouse: Substantia nigra. J. Neurosci. 25, 6721–6728 (2005).
7. Rossi, J., Balthasar, N., Olson, D., Scott, M., Berglund, E., Lee, C. E., Choi, M. J., Lauzon, D., Lowell, B. B. & Elmquist, J. K. Melanocortin-4 receptors expressed by cholinergic neurons regulate energy balance and glucose homeostasis. Cell Metab. 13, 195–204 (2011).
8. Vong, L., Ye, C., Yang, Z., Choi, B., Chua, S. & Lowell, B. B. Leptin Action on GABAergic Neurons Prevents Obesity and Reduces Inhibitory Tone to POMC Neurons. Neuron 71, 142–154 (2011).
9. Kumar, S., Miles, T. T. F., Chen, X., Brown, D., Dobreva, T., Huang, Q., Ding, X., Luo, Y., Einarsson, P. P. H., Greenbaum, A., et al. Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. Nat. Methods

17, 541–550 (2020).

Chapter III

AAV CAPSID VARIANTS WITH BRAIN-WIDE TRANSGENE EXPRESSION AND DECREASED LIVER TARGETING AFTER INTRAVENOUS DELIVERY IN MOUSE AND MARMOSET

Goertsen, D. †, Flytzanis, N. C. †, Goeden, N†, Chuapoco, M. R., Cummins, A., Chen, Y., Fan, Y., Zhang, Q., Sharma, J., Duan, Y., Wang, L., Feng, G., Chen, Y., Ip, N. Y., Pickel, J., Gradinaru, V. AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset. Nature Neuroscience. 25, 106–115 (2022). doi:10.1038/s41593-021-00969-4

† denotes equal contribution

D.G., N.C.F., and N.G. analyzed all data and prepared all figures, with input from V. Gradinaru. N.C.F and N.G. designed and performed the variant selection experiments, characterized the variants in mice, and prepared the virus for the pooled marmoset experiments. J.P and A.C. performed National Institutes of Health (NIH) marmoset experiments and analyzed the associated data. J.P. supervised the NIH experiments and helped prepare the associated figure. During revision, the Massachusetts Institute of Technology (MIT), the Shenzhen Institute of Advanced Technology (SIAT) and the Hong Kong University of Science and Technology (HKUST) provided additional marmoset data on controls (AAV-PHP.eB) and to increase 'n' for all cohorts. D.G. and M.C. prepared the virus for single-variant characterization and performed marmoset single-variant characterization tissue analysis of MIT and NIH marmoset tissues. Q.Z. and J.S. assisted with the MIT AAV injections, animal perfusion, and tissue sample collection, under the supervision of G.F. Yijing Chen and Y.F. performed single-variant characterization of the SIAT animals and analyzed the associated data, under the supervision of Yu Chen. L.W. supervised the marmoset breeding at the SIAT. Y.D. coordinated with D.G. for experiments at the SIAT, under the supervision of N.Y.I. N.C.F., N.G., D.G., and V.G. wrote the manuscript, with input from all authors. V.G. supervised all aspects of the work.

ABSTRACT

Genetic intervention is increasingly being explored as a therapeutic option for debilitating disorders of the central nervous system. The safety and efficacy of gene therapies rely upon expressing a transgene in affected cells while minimizing off-target expression. Here we show organ-specific targeting of adeno-associated virus (AAV) capsids after intravenous delivery, which we achieved by employing a Cre-transgenic-based screening platform and sequential engineering of AAV-PHP.eB between the surface-exposed AA452 and AA458. From this selection, we identified capsid variants that were enriched in the brain and targeted away from the liver in C57BL/6J mice. This tropism extends to marmoset (*Callithrix jacchus*), enabling robust, non-invasive gene delivery to the marmoset brain after intravenous administration. Notably, the capsids identified result in distinct transgene expression profiles within the brain, with one exhibiting high specificity to neurons. The ability to cross the blood–brain barrier with neuronal specificity in rodents and non-human primates enables new avenues for basic research and therapeutic possibilities unattainable with naturally occurring serotypes.

3.1 Introduction

The wide array of debilitating disorders affecting the central nervous system (CNS) make it a primary target for the application of novel therapeutic modalities, as evidenced by the rapid development and implementation of gene therapies targeting the brain. AAVs are increasingly used as delivery vehicles owing to their strong clinical safety record, low pathogenicity and stable, non-integrating expression *in vivo*[1]. AAVs were first approved for gene therapy in humans in 2012 to treat lipoprotein lipase deficiency[2] and have, more recently, been approved to treat spinal muscular atrophy[3], retinal dystrophy and hemophilia[4,5]. Many more advanced-stage clinical trials using AAVs are underway[6]. Currently, most AAV-based gene therapies rely on naturally occurring serotypes with highly overlapping tropisms[7], limiting the applicability, efficacy, and safety of novel gene therapies. At the same time, the inability to broadly and efficiently target many therapeutically relevant cell populations within organs that have traditionally been refractory to AAV delivery has hindered novel exploratory and therapeutic efforts. These constraints, which are especially prevalent for the CNS, motivated us to enhance AAV efficiency and specificity through directed evolution.

Naturally occurring AAVs have evolved to broadly infect cells[8], which is a desirable characteristic for the survival of the virus but undesirable for targeting specific cell types. This limitation has typically been addressed by injecting viral vectors directly into the area of interest in the CNS[9], through either intracranial or intrathecal injection. Although direct injection is a valuable technique for targeting focal cell populations, it requires considerable surgical expertise, and this approach cannot address applications that require broad and uniform area coverage (for example, all cortex or all striatum) or applications that are surgically difficult to access owing to the invasiveness of direct injections (for example, cerebellum and dorsal raphe)[9,10,11]. Systemic administration into the bloodstream is an appealing solution to the limitations of direct injections, as gene therapy vectors can be non-invasively delivered throughout the body. However, naturally occurring AAV serotypes tend to target non-CNS tissues, notably the liver, at high levels[7,8]. The liver is an immunologically active organ, with large populations of phagocytic cells that play a critical role in immune

activation[12,13]. Viral targeting to these tissues can trigger immune response, such as liver toxicity[14,15], reducing the safety and limiting the efficacy of systemic injection. Naturally occurring serotypes also have severely limited transduction efficiency in the brain owing to the stringency of the blood–brain barrier, consequently requiring the production of large, high-quality vector titers. Enabling and refining both direct and systemic injections will best equip the community with flexible options to pursue a broader range of neuroscience applications.

To target a precise anatomical region and/or cell type in conjunction with systemic injection, specificity can be obtained or refined through the inclusion of cell-type-specific promoters[16,17,18], enhancer elements[19,20,21,22,23] and microRNA target sites[24,25] into AAV viral genomes, or viral capsids can be engineered to alter their tissue tropism. Previously, we harnessed the power and specificity of Cre-transgenic mice to apply increased selective pressure to viral engineering, leading to variants AAV-PHP.B and AAV-PHP.eB, which cross the blood–brain barrier after intravenous administration and broadly transduce cells throughout the CNS[22,26]. These variants transformed the way that the CNS is studied in rodents, opening up completely novel approaches for measuring and affecting brain activity, in both exploratory and translational contexts[18,20,27,28]. However, these variants also transduce cells in the liver and other off-target organs. Therefore, the development of AAV variants with decreased liver targeting is particularly important to avoid strong, systemic immune responses[29,30,31].

To achieve specificity with the viral capsid, it is imperative that one applies both positive and negative selective pressure to engineered capsid libraries. The M-CREATE method[32] applies next-generation sequencing (NGS) of synthetic libraries with built-in controls to screen viral variants across multiple Cre-transgenic lines for both positive and negative features. Using the M-CREATE method, we selected a panel of novel variants that are both highly enriched in the CNS and targeted away from varying peripheral organs. Transgene expression after delivery with AAV.CAP-B10, described herein, was found to be highly specific for neurons in the CNS, significantly decreased in all peripheral organs assayed and targeted away from

the liver in mice. Notably, although AAV-PHP.B failed to translate to non-human primates (NHPs)[33, 34], here we show robust transgene expression after intravenous administration of newly engineered variants in the adult marmoset CNS with minimal liver expression compared to AAV9 and AAV-PHP.eB.

3.2 Engineering AAV capsids at the three-fold point of symmetry

The most commonly altered position within the AAV capsid is the surface-exposed loop containing amino acid (AA) 588, because it is the site of heparan sulfate binding in AAV2 (ref. [35]) and is amenable to peptide display[36,37]. The only known receptors for AAV9 are N-linked terminal galactose[38] and the AAV receptor[39,40], but the possibility of co-receptors is still unexplored. Binding interactions with cell surface receptors occur near the three-fold axis of symmetry of the viral capsid (**Figure 3-1a**) where the surface-exposed loop containing AA455 of AAV9 is the farthest protruding[41]. Having already engineered AAV-PHP.B and AAV-PHP.eB at the AA588 loop for enhanced CNS transduction[22,26], and with the additional goal of decreasing viral transduction of peripheral organs, we theorized that introducing diversity into the AA455 loop would enhance the interaction with existing mutations of the AA588 loop and refine transduction. Thus, we engineered and performed two rounds of selection with a 7-AA substitution library of the AA455 loop, between AA452 and AA458 (**Figure 3-1a**) in AAV-PHP.eB.

We applied the M-CREATE method developed by Kumar et al.[32] to identify AAV variants with desired tropism after systemic administration. Initially, we generated a library of AAV capsid sequences, theoretically containing 1.28 billion 7-mer AA substitutions, and produced the corresponding viruses in HEK293 cells that package a replication-incompetent version of their own genome with a polyadenylation sequence flanked by Cre-Lox sites (**Figure 3-1b**). This viral library is then injected into transgenic animals expressing Cre recombinase in specific cell populations. Variants that successfully transduce Cre+ cells have their genome flipped, and the corresponding sequences are recovered. To exert simultaneous positive and negative selective pressure, we performed parallel selections of this library in multiple transgenic lines expressing Cre recombinase in different cell populations. The differential

expression of Cre in the transgenic lines allows for shifts in the tropism of each viral variant to be compared between tissue types, enabling the recovery of unique sequences with a desired transduction profile after only two rounds of selection (for example, strong CNS and weak liver transduction).

We injected viral libraries into four separate Cre-transgenic mouse lines to select viral variants specific to the CNS: Tek-Cre for endothelial cells throughout the body; hSyn1-Cre for neurons of the CNS and peripheral nervous system; GFAP-Cre for astrocytes; and TH-Cre for midbrain cells. Using the described Cre-dependent method, we recovered viral DNA from all tissues across each mouse line and identified 434,770 capsid variants capable of tissue transduction through NGS. Afterwards, we ranked variant sequences from each tissue by their enrichment score, defined as the relative abundance of the sequence found within the specific tissue over the relative abundance of that sequence within the injected viral library. After ranking by enrichment, we selected the top 10–20% of variants with desired profiles from each tissue for the next round of selection.

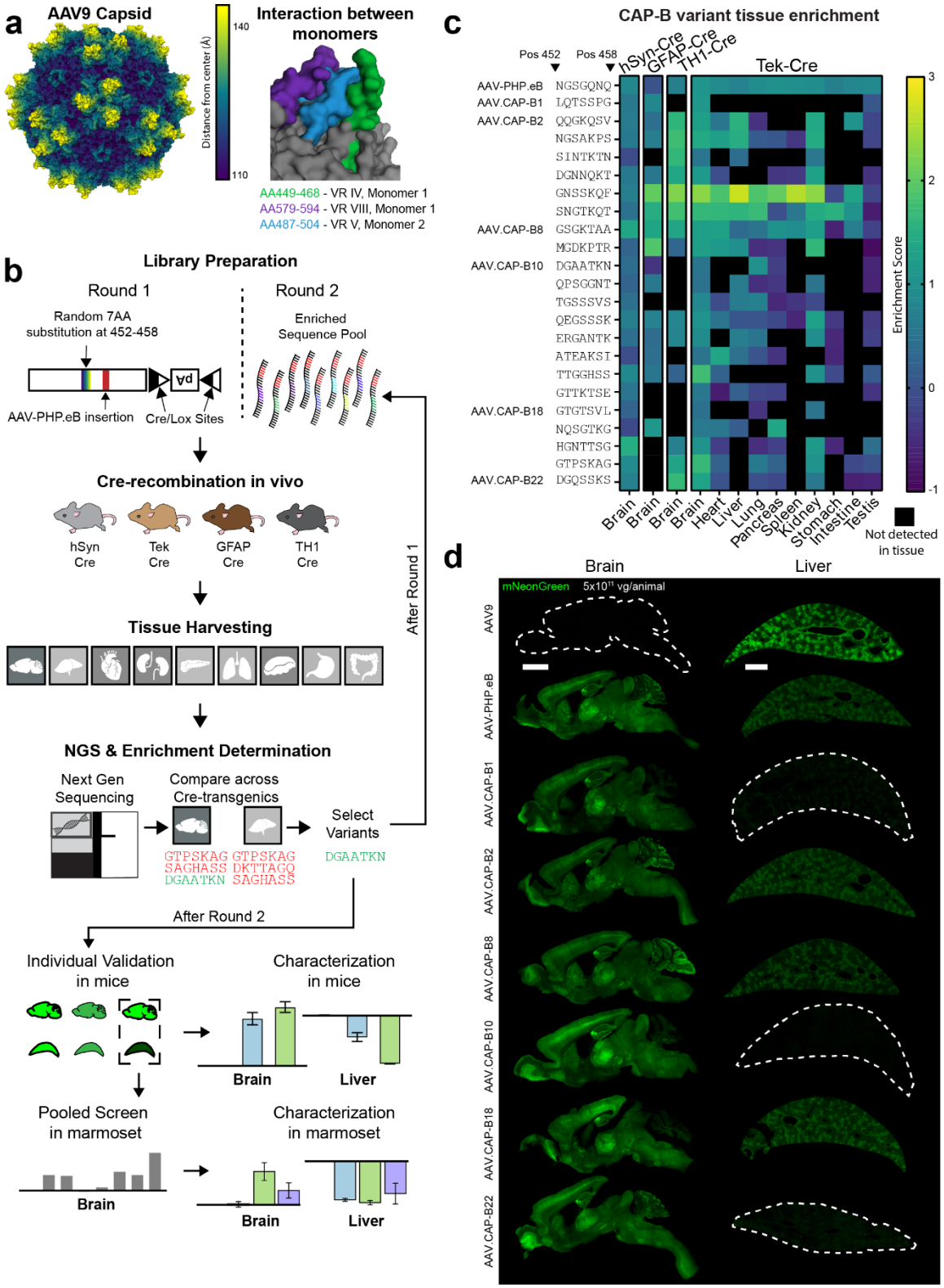
a
AAV9 Capsid
Distance from center (Å)
140
110
Interaction between monomers
AA449-468 - VR IV, Monomer 1
AA579-594 - VR VIII, Monomer 1
AA487-504 - VR V, Monomer 2
b
Library Preparation
Round 1
Round 2
Random 7AA substitution at 452-458
Enriched Sequence Pool
pA
AAV-PHP.eB insertion
Cre/Lox Sites
Cre-recombination in vivo
hSyn Cre
Tek Cre
GFAP Cre
TH1 Cre
Tissue Harvesting
NGS & Enrichment Determination
Next Gen Sequencing
Compare across Cre-transgenics
Select Variants
GTPSKAG SAGHASS DGAATKN
GTPSKAG DKTTAGQ SAGHASS
DGAATKN
After Round 1
After Round 2
Individual Validation in mice
Characterization in mice
Brain
Liver
Pooled Screen in marmoset
Characterization in marmoset
Brain
Brain
Liver
c
CAP-B variant tissue enrichment
Pos 452
Pos 458
hSyn-Cre
GFAP-Cre
TH1-Cre
Tek-Cre
AAV-PHP.eB NGSGQNQ
AAV.CAP-B1 LQTSSPG
AAV.CAP-B2 QQGKQSV
NGSAKPS
SINTKTN
DGNNQKT
GNSSKQF
SNGTKQT
AAV.CAP-B8 GSGKTAA
MGDKPTR
AAV.CAP-B10 DGAATKN
QPSGGNT
TGSSSVS
QEGSSSK
ERGANTK
ATEAKSI
TTGGHSS
GTTKTSE
AAV.CAP-B18 GTGTSVL
NQSGTKG
HGNTTSG
GTPSKAG
AAV.CAP-B22 DGQSSKS
Brain
Brain
Brain
Brain
Heart
Liver
Lung
Pancreas
Spleen
Kidney
Stomach
Intestine
Testis
Enrichment Score
3
2
1
0
-1
Not detected in tissue
d
Brain
Liver
mNeonGreen 5x10^11 vg/animal
AAV9
AAV.PHP.eB
AAV.CAP-B1
AAV.CAP-B2
AAV.CAP-B8
AAV.CAP-B10
AAV.CAP-B18
AAV.CAP-B22

Figure 3-1. Capsid engineering locations and CAP-B library characterization. a, Left, AAV9 capsid surface model illustrates the location of the protruding loop structures, in yellow. Right, capsid zoom-in shows the spike created by the AA579–AA594 and the AA449–AA468 variable regions of one AAV9 monomer interacting with the AA487–AA504 variable region of a second monomer. **b**, Experimental workflow. First, a library of variants is designed by mutating AAs between the AA452 and AA458 sites on an AAV-PHP.eB backbone. Cre-Lox sites are inserted into the viral genome to enable detection of genomes that reach Cre-recombinase-containing cells. This viral library is used to create viral capsids containing their corresponding genome. Next, variants are injected into transgenic mouse lines, where Cre-recombination occurs after transduction. After 3 weeks of expression, relevant tissues are harvested from each transgenic line, and the DNA is isolated, amplified and prepared for sequencing. After NGS, variants are ranked based on enrichment in select tissues, and the top 10–20% performers repeat the selection process above. Top enriched variants with unique tissue tropism profiles after two rounds of selection are validated individually for transgene expression by *in vivo* screening in mice. Variants that display similar brain, but reduced liver, transgene expression are selected for pooled marmoset testing. **c**, Heat map plotting the enrichment scores of a subset of the 22 variants identified after two rounds of selection *in vivo* across Cre-transgenic lines shows enrichment and specificity for the CNS. The sequences for the 7-AA substitutions present in each of the variants show divergence from the parent AAV-PHP.eB sequence (see also **Appendix B, Figure B-1**). **d**, mNeonGreen was packaged in each variant under control of the ubiquitous CAG promoter and intravenously administered to mice at a dose of 5×10^{11} viral genomes per animal. Transgene expression was assayed by mNeonGreen fluorescence throughout brain and liver. The expression from ssAAV9, ssAAV-PHP.eB, ssAAV.CAP-B1, ssAAV.CAP-B2, ssAAV.CAP-B8, ssAAV.CAP-B10, ssAAV.CAP-B18 and ssAAV.CAP-B22 packaging CAG-mNeonGreen as assessed after 2 weeks. Direct comparison of the expression profiles in brain and liver of the top-performing variants shows a correlation between validated tropisms and those predicted by the NGS data. Scale bars, 2 mm.

3.3 Capsid engineering refines expression patterns in mice

We used the enriched sequences from the first round of selection to synthesize a library for the second round of selection. The second-round library contained codon sequences of 82,710 unique variants and duplicate, codon-modified versions of each variant as a replicate. Using the same selection methods and transgenic lines as round 1, we determined the enrichment of each variant across tissues and transgenic lines. The second round of selection reduced the number of enriched variants by two orders of magnitude, and we identified 39,034 sequences that result in CNS enrichment and decreased targeting of the liver. Using these sequences, we differentiated a subset of AAs in each position that contributed to this

tropism profile (**Appendix B, Figure B-1**). We ranked sequences by their specificity to CNS tissues and compared against the enrichment in peripheral tissues across the different transgenic lines, selecting 22 sequences with diverse sequence identity (**Figure 3-1c**) for further testing *in vivo*.

To validate the enrichment profile obtained from sequencing, we individually screened each of these 22 capsid variants for their tissue tropism after systemic injection in mice. We produced capsid variants, packaging a mNeonGreen fluorescent reporter under regulation of the ubiquitous CAG promoter, and administered each variant to wild-type (WT) mice by intravenous injection at a dose of 5×10^{11} viral genomes per animal. After 2 weeks of expression, we imaged brains and livers and quantified fluorescence to assess mNeonGreen expression. Six variants displayed similar or higher fluorescence in the brain to AAV-PHP.eB as well as reduced liver expression in mice (**Figure 3-1d**): AAV.CAP-B1, AAV.CAP-B2, AAV.CAP-B8, AAV.CAP-B10, AAV.CAP-B18, and AAV.CAP-B22. Owing to the negligible liver expression of AAV.CAP-B10, we characterized this variant further in mice, whereas we selected all six variants for pooled testing in marmosets.

3.4 AAV.CAP-B10 yields CNS-specific transgene expression in mice

To characterize the performance of AAV.CAP-B10 in comparison to AAV9 and AAV-PHP.eB, we packaged a nuclear-localized enhanced green fluorescent protein (EGFP) under regulation of a ubiquitously expressed CAG promoter, enabling quantification across cell types throughout the body. The variants were injected into mice at a dose of 1×10^{11} viral genomes per animal, and tissue was collected after 3 weeks of expression. Detailed cell counts of overall EGFP expression throughout the body showed that expression after delivery with AAV.CAP-B10 is highly specific to the CNS. In the brain, delivery with AAV.CAP-B10 yielded a similar number of EGFP-expressing cells with a similar level of expression per cell to AAV-PHP.eB, whereas both AAV.CAP-B10 and AAV-PHP.eB had many more cells expressing EGFP than AAV9 (**Figure 3-2a**). In the spinal cord, 40% fewer cells displayed EGFP expression after delivery with AAV.CAP-B10 than with AAV-PHP.eB but approximately 16-fold more than AAV9 (**Appendix B, Figure B-2**).

Conversely, the number of cells expressing EGFP after delivery with AAV.CAP-B10 was significantly reduced in the liver compared to both AAV-PHP.eB (~50-fold) and AAV9 (>100-fold). EGFP expression after delivery with AAV.CAP-B10, but not AAV-PHP.eB, was significantly dimmer per cell than AAV9 in the liver (~tenfold) (**Figure 3-2b**). This trend was maintained in the other peripheral organs, with significantly fewer cells displaying EGFP expression after delivery with AAV.CAP-B10 than AAV9 (**Appendix B, Figure B-2**).

3.5 AAV.CAP-B10 brain transgene expression is neuronal specific

To further characterize transgene expression in the CNS after delivery with AAV.CAP-B10 compared to AAV-PHP.eB, we co-stained for neurons (αNeuN), astrocytes (αS100), oligodendrocytes (αOlig2) and Purkinje cells (αCalbindin) and quantified the efficiency of EGFP expression after delivery with each capsid in each cell type across various brain regions (**Figure 3-2c–h** and **Appendix B, Figure B-3**). Whereas neurons displayed EGFP expression with similar efficiencies after delivery by AAV.CAP-B10 or AAV-PHP.eB across brain regions (**Figure 3-2c,d**), roughly four- to five-fold fewer astrocytes and oligodendrocytes display EGFP expression after delivery by AAV.CAP-B10 compared to AAV-PHP.eB (**Figure 3-2e–h**). This indicates that the AAV.CAP-B10 mutation confers a bias for neurons over other cell types, an interesting deviation from AAV9, which mostly targets astrocytes in the brain[42,43]. A noteworthy indication from the NGS data for AAV.CAP-B10 was this variant's decreased presence in the cerebellum, which was also evidenced by a decrease in the overall cerebellar fluorescence in the initial characterization (**Figure 3-1d**). When comparing the cerebellar expression after delivery with AAV.CAP-B10 to that of AAV-PHP.eB, we found a significant, roughly four-fold decrease in the number of Purkinje cells displaying EGFP expression (**Appendix B, Figure B-4**).

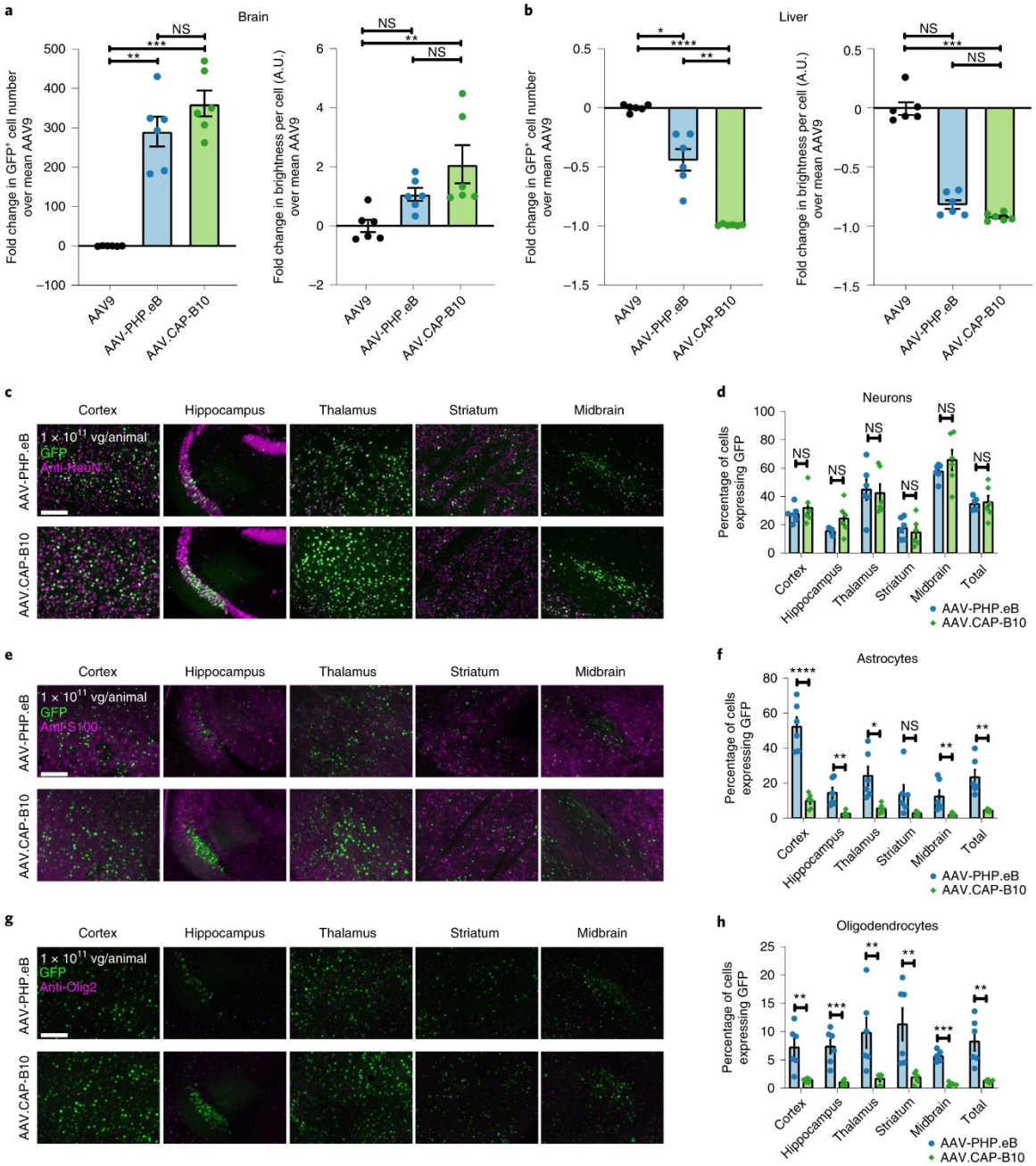

Figure 3-2. AAV.CAP-B10 tissue expression profile is biased toward the brain, with a significant decrease in liver targeting. a, ssAAV9, ssAAV-PHP.eB, and ssAAV.CAP-B10, packaging a nuclear-localized GFP under the control of the CAG promoter, were intravenously injected into male adult mice at 1×10^{11} viral genomes per mouse. GFP fluorescence was assessed after three weeks of expression. Quantification of the total number of cells expressing GFP in the brain ($P = 0.0016$ (AAV9 versus AAV-PHP.eB), $P = 0.0003$ (AAV9 versus AAV.CAP-B10), and $P = 0.4345$ (AAV-PHP.eB versus AAV.CAP-B10)) and the average brightness per cell ($P = 0.06$ (AAV9 versus AAV-PHP.eB), $P = 0.0043$ (AAV9 versus AAV.CAP-B10), and $P > 0.999$ (AAV-PHP.eB versus AAV.CAP-B10))

show an increase in transgene expression efficiency in the brain of AAV.CAP-B10 similar to AAV-PHP.eB. **b**, Quantification of the total percentage of cells expressing GFP in the liver (P = 0.0127 (AAV9 versus AAV-PHP.eB), P < 0.0001 (AAV9 versus AAV.CAP-B10), and P = 0.0047 (AAV-PHP.eB versus AAV.CAP-B10)) and average brightness per cell (P = 0.080 (AAV9 versus AAV-PHP.eB), P = 0.0009 (AAV9 versus AAV.CAP-B10), and P = 0.48 (AAV-PHP.eB versus AAV.CAP-B10)) show an iterative decrease in transgene expression efficiency from AAV-PHP.eB to AAV.CAP-B10. **a**, **b**, n = 6 mice per group, mean ± s.e. Statistical significance was determined using Brown–Forsythe and Welch ANOVA tests with the Dunnett T3 correction for multiple comparisons for transgene expression and the Kruskal–Wallis test with Dunn's correction for multiple comparisons for brightness. **c**, **d**, Within the brain, AAV.CAP-B10 is biased toward neurons (P = 0.44 (cortex), P = 0.21 (hippocampus), P = 0.81 (thalamus), P = 0.56 (striatum), P = 0.31 (midbrain) and P = 0.82 (total)), stained with αNeuN (Abcam, 177487). Expression and percentage of neurons expressing transgene. **e**, **f**, Expression and percentage of astrocytes expressing transgene (P = 0.00002 (cortex), P = 0.0034 (hippocampus), P = 0.017 (thalamus), P = 0.057 (striatum), P = 0.0070 (midbrain), and P = 0.0012 (total)), stained with αS100 (Abcam, 868). **g**, **h**, Expression and percentage of oligodendrocytes expressing transgene (P = 0.0039 (cortex), P = 0.00046 (hippocampus), P = 0.0047 (thalamus), P = 0.0085 (striatum), P = 0.00019 (midbrain), and P = 0.0018 (total)), stained with αOlig2 (Abcam, 109186). **c**–**h**, n = 6 mice per group, except for hippocampal NeuN+ counts of AAV.CAP-B10, where n = 3, mean ± s.e. Statistical significance was determined using two-sided Welch's t-tests. The contribution of cells from different classification to overall EGFP expression was measured and indicates a shift toward neuronal specificity of AAV.CAP-B10 compared to AAV-PHP.eB (see also **Appendix B, Figure B-2,3,4**). Scale bars, 200 µm. A.U., arbitray units; NS, not significant; vg, viral genomes.

3.6 Engineered variants maintain robust tropism in marmosets

Of primary concern for the therapeutic applicability of variants engineered in rodents is how well their transgene expression profile translate to NHPs. Therefore, we sought to characterize the marmoset CNS transgene expression profile of a subset of the variants validated in mice, along with AAV9 and AAV-PHP.eB as controls. We produced a pool of eight viruses—AAV9, AAV-PHP.eB, AAV.CAP-B1, AAV.CAP-B2, AAV.CAP-B8, AAV.CAP-B10, AAV.CAP-B18, and AAV.CAP-B22—each packaging an HA-tagged frataxin (FXN) with a unique molecular barcode under control of the ubiquitous CAG promoter. We opted to use FXN because it is an endogenous protein expressed throughout the body, and previous efforts to characterize NHP transgene expression after delivery with naturally occurring and engineered serotypes have resulted in deleterious effects for the host, potentially due to the packaging of exogenous transgenes such as GFP[29,44,45]. We also

included a separate 12-base RNA barcode into each genome to differentiate the contribution of each virus from the rest after using NGS. The eight viruses were pooled at equal ratios and intravenously injected into two adult marmosets at doses of 1.2×10^{14} (marmoset pool virus 1 (MPV1)) and 7×10^{13} (MPV2) viral genomes per kg (**Appendix B, Table B-1**). After 6 weeks of expression, during which no adverse health effects were observed, the brains and livers were recovered, and sections were taken for RNA sequencing and immunohistochemistry.

Staining for the HA tag revealed that robust and broad expression was achieved in the adult marmoset brain, with strong expression of the viral pool observed in cortical, subcortical and cerebellar layers throughout (**Figure 3-3a,b**). In the liver, moderate FXN expression was observed after staining for the HA tag (**Figure 3-3c**). From multiple slices per animal distributed throughout the brain and liver, we extracted RNA, performed NGS and quantified the relative expression levels of each of the barcoded viruses. AAV.CAP-B10 recapitulated the trends observed in mice (albeit at lower levels), with a ~six-fold increase in RNA in the brain and a ~five-fold decrease in RNA in the liver compared to AAV9 (**Figure 3-3d,e**). AAV.CAP-B22 exhibited the largest increase in RNA levels in the brain, with a greater than 12-fold increase in the brain compared to AAV9, but relatively similar levels in the liver (**Figure 3-3d,e**). Based on their unique expression profiles in the pooled screen, we selected AAV.CAP-B10 and AAV.CAP-B22 for individual characterization of their CNS transgene expression profile in adult marmosets compared to AAV9 and AAV-PHP.eB.

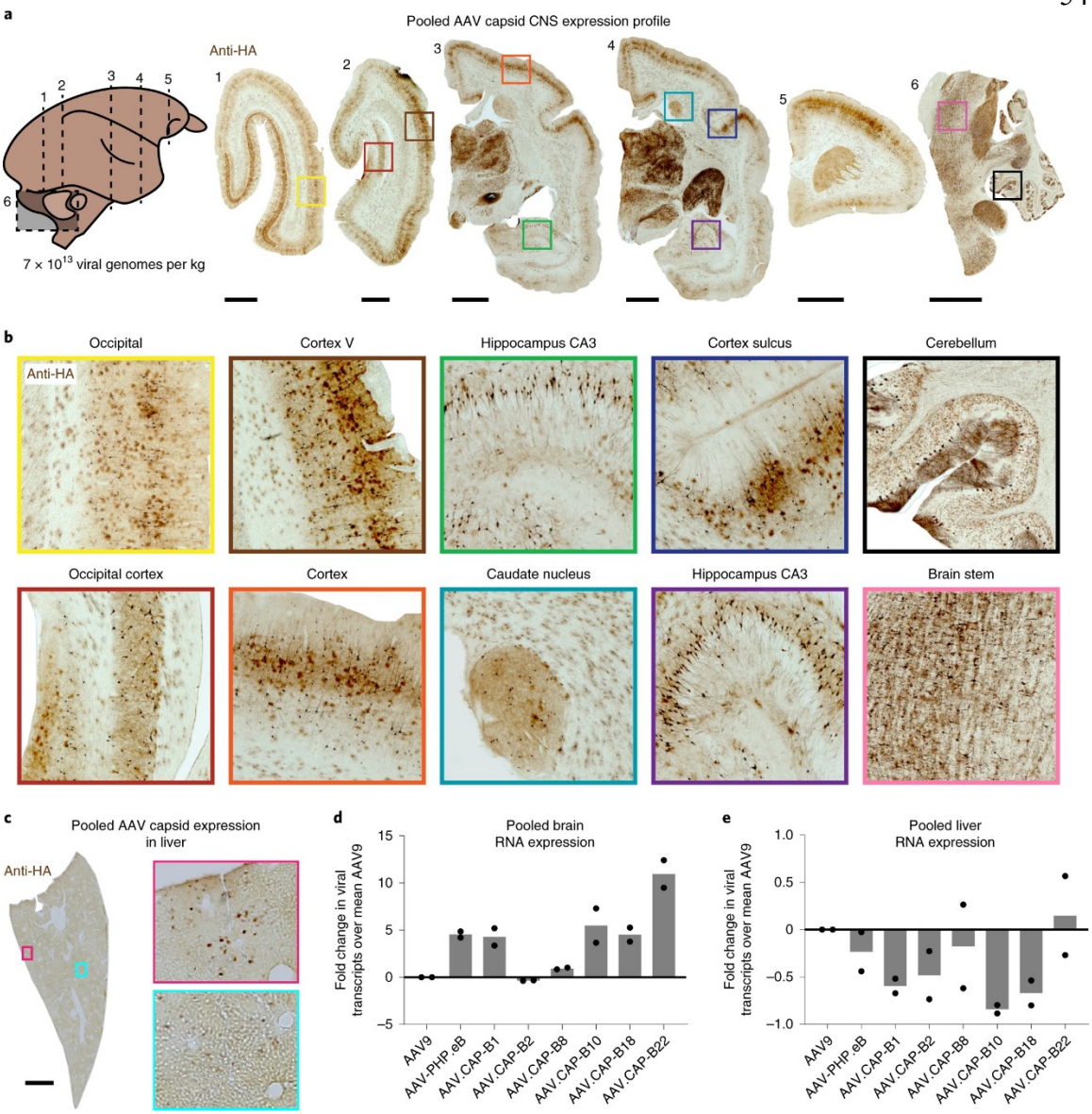

Figure 3-3. Characterization of pooled capsid transgene expression in NHPs. a, AAV.CAP-B1, AAV.CAP-B2, AAV.CAP-B8, AAV.CAP-B10, AAV.CAP-B18, and AAV.CAP-B22, along with AAV9 and AAV-PHP.eB as controls, packaging a human FXN fused to an HA tag under the control of the ubiquitous CAG promoter, were pooled and intravenously injected into two adult marmosets at doses of 1.2×10^{14} (MPV1) and 7×10^{13} (MPV2) viral genomes per kg total. Six sections distributed throughout the brain of one marmoset (MPV2) showed robust expression after immunostaining for the HA tag (Cell Signaling Technologies, C29F). Scale bars, 2 mm. **b**, Magnified frames from **a** for a variety of cortical and sub-cortical regions. **c**, Liver section taken from one marmoset (MPV2) shows minimal expression after immunostaining for the HA tag (Cell Signaling Technologies, C29F). **d**, NGS quantification within brain tissue of the unique RNA barcode associated with each virus shows a substantial increase for several variants, including a more than 12-fold increase in RNA levels of AAV.CAP-B22 and a five-fold increase for AAV.CAP-B10

compared to AAV9. **e**, NGS quantification of pooled variant injections, showing relative targeting away from the liver for the different variants. AAV.CAP-B22 contributes similar RNA levels as AAV9, but AAV.CAP-B10 contributes more than five-fold less. n = 2 marmosets per group. Scale bars, 2 mm.

We intravenously injected individual adult marmosets with AAV9, AAV-PHP.eB, AAV.CAP-B10, or AAV.CAP-B22, each packaging an HA-tagged FXN under the control of the ubiquitous CAG promoter, at a dose of 7×10^{13} viral genomes per kg (**Appendix B, Table B-2**). The brains were recovered, and sections were taken for sequencing and immunofluorescence between 34 d and 42 d after intravenous injection. To characterize the CNS transgene expression after delivery with the four variants, we stained sections for the presence of the HA tag. In stained sections, both AAV.CAP-B10 and AAV.CAP-B22 conditions displayed a marked increase in FXN expression efficiency within cortical regions in comparison to AAV9 and AAV-PHP.eB (**Figure 3-4a**). Each of AAV-PHP.eB, AAV.CAP-B10, and AAV.CAP-B22 displayed reduction in liver FXN expression relative to AAV9 (**Figure 3-4c**).

After delivery with AAV.CAP-B22, FXN expression was observed across broad cell types in the marmoset brain. Qualitatively, many more astrocytes (measured by co-staining with S100β) displayed FXN expression compared to delivery with AAV9 or AAV.CAP-B10 (**Appendix B, Figure B-5**). The FXN expression levels of AAV.CAP-B22 had high variability among animals in both the brain and liver, and, therefore, we observed no significant difference in the neuronal FXN expression after delivery with AAV.CAP-B22 relative to AAV9 or AAV-PHP.eB.

The neuronal specificity of AAV.CAP-B10 in mice was recapitulated in marmoset brains, where HA-positive cells were highly correlated with NeuN staining across brain regions. In the marmoset cortex, AAV.CAP-B10 displays a statistically significant ~four-fold increase in HA-positive neurons over AAV9 and AAV-PHP.eB (**Figure 3-4b**). Notably, this increased protein expression in cortical neurons after delivery with AAV.CAP-B10 occurs despite no significant difference between the bulk viral genome and transcript measurements of AAV.CAP-B10 and AAV9 or AAV-PHP.eB (**Appendix B, Figure B-6b**), further

highlighting tropism differences between AAV9 and AAV.CAP-B10. The decreased liver targeting of AAV.CAP-B10 in mice was also recapitulated in the marmoset, where there were ~17-fold fewer cells expressing the transgene compared to an AAV9 control (**Figure 3-4d**). Viral genome and transcript measurements corroborate the protein expression data in the liver (**Appendix B, Figure B-6a**). We observed broad and robust transgene expression after delivery with AAV.CAP-B10 across cortical, sub-cortical, and cerebellar regions as well as in the spinal column and the dorsal root ganglia (DRG) (**Figure 3-4e,f**).

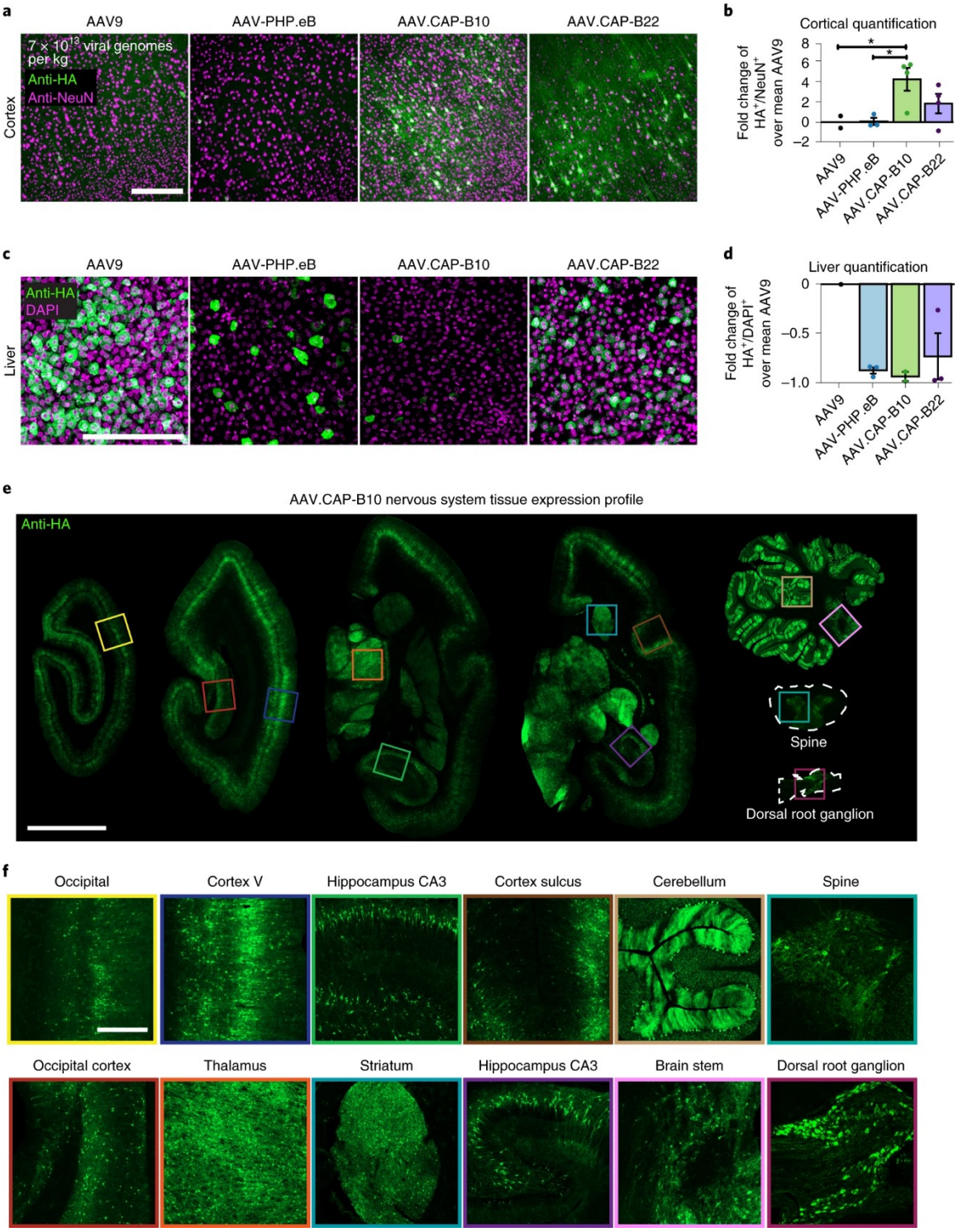

Figure 3-4. Characterization of single-variant expression after delivery with each of AAV9, AAV-PHP.eB, AAV.CAP-B10, and AAV.CAP-B22 in marmosets. Human FXN fused to an HA tag is packaged in each variant under control of the ubiquitous CAG promoter. Marmosets were injected at a dose of 7×10^{13} viral genomes per kg (see also

Appendix B, Figure B-5,6). **a**, Cortical expression is compared for AAV9, AAV-PHP.eB, AAV.CAP-B10, and AAV.CAP-B22 by immunostaining for the HA tag (Roche, 3F10) on the FXN transgene in conjunction with NeuN (Abcam, 177487). A qualitative increase in transgene expression efficiency for AAV.CAP-B10 and AAV.CAP-B22 is observed in comparison to AAV9 and AAV-PHP.eB. Displayed sections are taken from a similar plane and cortical region. Scale bar, 200 µm. **b**, Similar regions across all layers of the marmoset cortex are quantified, where AAV.CAP-B10 shows ~four-fold increase in HA-positive neurons over AAV9 (P = 0.0299) and AAV-PHP.eB (P = 0.0295), respectively. Differences among other groups are not significant. n = 2 for AAV9, n = 3 for AAV-PHP.eB, n = 4 for AAV.CAP-B10, and n = 4 for AAV.CAP-B22, mean ± s.e. Significance is determined by two-tailed Welch's t-test. **c**, Liver expression is compared for AAV9, AAV-PHP.eB, AAV.CAP-B10, and AAV.CAP-B22 by immunostaining for the HA tag (Roche, 3F10) on the FXN transgene in conjunction with DAPI. Scale bar, 200 µm. **d**, In the liver, we quantified the fraction of HA-positive cells out of the total nuclei stained, where AAV.CAP-B10 had expression in ~17-fold fewer cells than AAV9. No significant differences were observed among AAV-PHP.eB, AAV.CAP-B10, or AAV.CAP-B22. n = 1 for AAV9, n = 3 for AAV-PHP.eB, n = 2 for AAV.CAP-B10, and n = 3 for AAV.CAP-B22, mean ± s.e. **e**, Five sections distributed throughout the brain, spinal cord and DRG of the marmoset show robust expression after immunostaining for the HA tag (Roche, 3F10). Scale bar, 5 mm. **f**, Magnified frames from **e** display expression across cortical, sub-cortical, cerebellar, spinal column, and DRG regions. Scale bar, 500 µm.

3.7 Discussion

The power of directed evolution and AAV engineering to confer novel tropisms and tissue specificity has broadened potential research applications and enabled new therapeutic approaches in the CNS[46,47,48,49,50,51]. Our results show that introducing diversity at multiple locations on the capsid surface into native or previously engineered variants can provide useful features, such as increased transgene expression efficiency and tissue or neuronal specificity in WT mice and marmosets. This finding has broad implications for the field of AAV engineering, and the variants discovered using the M-CREATE method[22,26,32] have the potential to provide targeted gene delivery in WT animals.

Novel variants from this library achieved overall similar levels of transgene expression throughout the mouse brain as the previously engineered parent AAV (AAV-PHP.eB) but with striking deviations in cell type specificity and targeting or, notably, targeting away from other organs. Across the two engineering efforts, iterative modifications made to adjacent loops first conferred a new phenotype—the ability to cross the blood–brain barrier

efficiently—and then refined the phenotype away from other organs or toward specific cell types. Most notably from this dataset, AAV.CAP-B10 is targeted away from the entire periphery and exhibits specificity for neurons over other cell types in the CNS. Of the 82,710 variants that comprised our second round of selection, roughly 39,000 exhibited positive enrichment in the brain and negative enrichment in the liver. Of those, the variants tested *in vivo* in mice showed expression patterns that correlated with their NGS enrichments and rankings. Furthermore, the presence of additional sequences that were positively enriched in the brain and negatively enriched in peripheral organs indicates the richness of this dataset and engineering location, with many more potentially interesting variants to be identified and characterized. Together, these results indicate that future engineering efforts can also take a stepwise approach toward attaining specificity for certain targets, with each engineering round refining and enhancing novel tropisms identified in the previous round. Notably, this implies that AAV.CAP-B10 might be further refined in its specificity to neuronal sub-classes[52,53] in the CNS or away from the DRG, which has been associated with toxicity in NHP studies[54].

The therapeutic use of engineered AAVs for treating disease has increased exponentially in recent years, including the use of systemically administered AAVs for CNS gene therapies. The brain is a prime target for gene therapy of a wide array of diseases, such as Huntington's disease, Parkinson's disease and Friedreich's ataxia[55,56,57], and, by developing gene therapy vectors for systemic administration, we can enable the wide coverage and non-invasive delivery conducive to effective treatment of these diseases. To minimize side effects typically associated with systemic delivery, gene therapy vectors should be engineered for specificity toward the therapeutic target of interest in conjunction with decreased off-target expression. Basic research in NHP brains, compared to rodent research, has been hampered by few options for genetic access. Gene delivery vectors could bridge this gap and enable eventual use in gene therapy applications. However, the translation of engineered variants from mice to NHPs has been challenging. Two generations of blood–brain-barrier-crossing engineered AAV variants (AAV-PHP.B and AAV-PHP.eB) have failed[33] to increase transgene expression in the brain after systemic injection in NHPs. Therefore, variants engineered in

this study, including AAV.CAP-B10 and AAV.CAP-B22, represent notable progress toward evolving variants that are capable of efficiently, specifically, and safely delivering gene therapies to the CNS in NHPs and, ultimately, in humans. Robust transgene expression in the CNS, decreased liver targeting and neuronal bias of AAV.CAP-B10 observed after two rounds of selection in mice were recapitulated in marmosets, indicating that M-CREATE in rodents can be effectively used as a screening platform before direct screening or validation in NHPs, thus greatly increasing the throughput of the engineering process. Together, these results constitute an important step forward toward achieving the goal of engineered AAV vectors that can be used to broadly deliver gene therapies to the CNS in humans.

Chapter III Bibliography

1. Naso, M. F. et al. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. BioDrugs 31, 317–334 (2017).
2. Gaudet, D., Méthot, J. & Kastelein, J. Gene therapy for lipoprotein lipase deficiency. Curr. Opin. Lipidol. 23, 310–320 (2012).
3. Al-Zaidy, S. A. & Mendell, J. R. From Clinical Trials to Clinical Practice: Practical Considerations for Gene Replacement Therapy in SMA Type 1. Pediatr. Neurol. 100, 3–11 (2019).
4. Ameri, H. Prospect of retinal gene therapy following commercialization of voretigene neparvovec-rzyl for retinal dystrophy mediated by RPE65 mutation. J. Curr. Ophthalmol. 30, 1–2 (2018).
5. Evens, H., Chuah, M. K. & VandenDriessche, T. Haemophilia gene therapy: From trailblazer to gamechanger. Haemophilia 24, 50–59 (2018).
6. Ginn, S. L., Amaya, A. K., Alexander, I. E., Edelstein, M. & Abedi, M. R. Gene therapy clinical trials worldwide to 2017: An update. J. Gene Med. 20, e3015 (2018).
7. Srivastava, A. In vivo tissue-tropism of adeno-associated viral vectors. Curr. Opin. Virol. 21, 75–80 (2016).
8. Zincarelli, C., Soltys, S., Rengo, G. & Rabinowitz, J. E. Analysis of AAV Serotypes 1–9 Mediated Gene Expression and Tropism in Mice After Systemic Injection. Mol. Ther. 16, 1073–1080 (2008).
9. Gray, S. J., Woodard, K. T. & Samulski, R. J. Viral vectors and delivery strategies for CNS gene therapy. Ther. Deliv. 1, 517–534 (2010).
10. Golebiowski, D. et al. Direct Intracranial Injection of AAVrh8 Encoding Monkey β-N-Acetylhexosaminidase Causes Neurotoxicity in the Primate Brain. Hum. Gene Ther. 28, 510–522 (2017).
11. Hocquemiller, M., Giersch, L., Audrain, M., Parker, S. & Cartier, N. Adeno-Associated Virus-Based Gene Therapy for CNS Diseases. Hum. Gene Ther. 27,

478–496 (2016).
12. Kubes, P. & Jenne, C. Immune Responses in the Liver. Annu. Rev. Immunol. 36, 247–277 (2018).
13. Gao, B. Basic liver immunology. Cell. Mol. Immunol. 13, 265–266 (2016).
14. Ronzitti, G., Gross, D.-A. & Mingozzi, F. Human Immune Responses to Adeno-Associated Virus (AAV) Vectors. Front. Immunol. 11, 670 (2020).
15. Verdera, H. C., Kuranda, K. & Mingozzi, F. AAV Vector Immunogenicity in Humans: A Long Journey to Successful Gene Transfer. Mol. Ther. 28, 723–746 (2020).
16. Gray, S. J. et al. Optimizing promoters for recombinant adeno-associated virus-mediated gene expression in the peripheral and central nervous system using self-complementary vectors. Hum. Gene Ther. 22, 1143–1153 (2011).
17. de Leeuw, C. N. et al. Targeted CNS Delivery Using Human MiniPromoters and Demonstrated Compatibility with Adeno-Associated Viral Vectors. Mol. Ther. Methods Clin. Dev. 1, 5 (2014).
18. Allen, W. E. et al. Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex. Neuron 94, 891-907.e6 (2017).
19. Shima, Y. et al. A Mammalian enhancer trap resource for discovering and manipulating neuronal cell types. Elife 5, e13503 (2016).
20. Graybuck, L. T. et al. Enhancer viruses for combinatorial cell-subclass-specific labeling. Neuron 109, 1449-1464.e13 (2021).
21. Mich, J. K. et al. Functional enhancer elements drive subclass-selective expression from mouse to primate neocortex. Cell Rep. 34, 108754 (2021).
22. Chan, K. Y. et al. Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. Nat. Neurosci. 20, 1172–1179 (2017).
23. Vormstein-Schneider, D. et al. Viral manipulation of functionally distinct interneurons in mice, non-human primates and humans. Nat. Neurosci. 23, 1629–1636 (2020).
24. Keaveney, M. K. et al. A MicroRNA-Based Gene-Targeting Tool for Virally Labeling Interneurons in the Rodent Cortex. Cell Rep. 24, 294–303 (2018).
25. Hordeaux, J. et al. MicroRNA-mediated inhibition of transgene expression reduces dorsal root ganglion toxicity by AAV vectors in primates. Sci. Transl. Med. 12, eaba9188 (2020).
26. Deverman, B. E. et al. Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat. Biotechnol. 34, 204–209 (2016).
27. Zelikowsky, M. et al. The Neuropeptide Tac2 Controls a Distributed Brain State Induced by Chronic Social Isolation Stress. Cell 173, 1265-1279.e19 (2018).
28. Zeng, J. et al. TRIM9-Mediated Resolution of Neuroinflammation Confers Neuroprotection upon Ischemic Stroke in Mice. Cell Rep. 27, 549-560.e6 (2019).
29. Gao, G. et al. Adeno-Associated Virus-Mediated Gene Transfer to Nonhuman Primate Liver Can Elicit Destructive Transgene-Specific T Cell Responses. Hum. Gene Ther. 20, 930–942 (2009).
30. Mingozzi, F. & High, K. A. Immune responses to AAV vectors: overcoming barriers to successful gene therapy. Blood 122, 23–36 (2013).

31. Calcedo, R., Chichester, J. A. & Wilson, J. M. Assessment of Humoral, Innate, and T-Cell Immune Responses to Adeno-Associated Virus Vectors. Hum. Gene Ther. Methods 29, 86–95 (2018).
32. Kumar, S. et al. Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. Nat. Methods 17, 541–550 (2020).
33. Matsuzaki, Y. et al. Intravenous administration of the adeno-associated virus-PHP.B capsid fails to upregulate transduction efficiency in the marmoset brain. Neurosci. Lett. 665, 182–188 (2018).
34. Hordeaux, J. et al. The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice. Molecular Therapy vol. 26 664–668 (2018).
35. Kern, A. et al. Identification of a Heparin-Binding Motif on Adeno-Associated Virus Type 2 Capsids. J. Virol. 77, 11072–11081 (2003).
36. Girod, A. et al. Genetic capsid modifications allow efficient re-targeting of adeno-associated virus type 2. Nat. Med. 5, 1052–1056 (1999).
37. Michelfelder, S. et al. Peptide Ligands Incorporated into the Threefold Spike Capsid Domain to Re-Direct Gene Transduction of AAV8 and AAV9 In Vivo. PLoS One 6, e23101 (2011).
38. Shen, S., Bryant, K. D., Brown, S. M., Randell, S. H. & Asokan, A. Terminal N-Linked Galactose Is the Primary Receptor for Adeno-associated Virus 9. J. Biol. Chem. 286, 13532–13540 (2011).
39. Summerford, C., Johnson, J. S. & Samulski, R. J. AAVR: A Multi-Serotype Receptor for AAV. Mol. Ther. 24, 663–666 (2016).
40. Zhang, R. R. et al. Adeno-associated virus 2 bound to its cellular receptor AAVR. Nat. Microbiol. 4, 675–682 (2019).
41. DiMattia, M. A. et al. Structural Insight into the Unique Properties of Adeno-Associated Virus Serotype 9. J. Virol. 86, 6947–6958 (2012).
42. Foust, K. D. et al. Intravascular AAV9 preferentially targets neonatal neurons and adult astrocytes. Nat. Biotechnol. 27, 59–65 (2009).
43. Samaranch, L. et al. Adeno-Associated Virus Serotype 9 Transduction in the Central Nervous System of Nonhuman Primates. Hum. Gene Ther. 23, 382–389 (2012).
44. Flotte, T. R. & Büning, H. Severe Toxicity in Nonhuman Primates and Piglets with Systemic High-Dose Administration of Adeno-Associated Virus Serotype 9–Like Vectors: Putting Patients First. Hum. Gene Ther. 29, 283–284 (2018).
45. Unzu, C. et al. Transient and intensive pharmacological immunosuppression fails to improve AAV-based liver gene transfer in non-human primates. J. Transl. Med. 10, 122 (2012).
46. Giannelli, S. G. et al. Cas9/sgRNA selective targeting of the P23H Rhodopsin mutant allele for treating retinitis pigmentosa by intravitreal AAV9.PHP.B-based delivery. Hum. Mol. Genet. 27, 761–779 (2018).
47. Schubert, R. et al. Virus stamping for targeted single-cell infection in vitro and in vivo. Nat. Biotechnol. 36, 81–88 (2017).
48. Morabito, G. et al. AAV-PHP.B-Mediated Global-Scale Expression in the Mouse Nervous System Enables GBA1 Gene Therapy for Wide Protection from Synucleinopathy. Mol. Ther. 25, 2727–2742 (2017).

49. Hillier, D. et al. Causal evidence for retina-dependent and -independent visual motion computations in mouse cortex. Nat. Neurosci. 20, 960–968 (2017).
50. Dayton, R. D., Grames, M. S. & Klein, R. L. More expansive gene transfer to the rat CNS: AAV PHP.EB vector dose–response and comparison to AAV PHP.B. Gene Ther. 25, 392–400 (2018).
51. Rincon, M. Y. et al. Widespread transduction of astrocytes and neurons in the mouse central nervous system after systemic delivery of a self-complementary AAV-PHP.B vector. Gene Ther. 25, 83–92 (2018).
52. Yuste, R. et al. A community-based transcriptomics classification and nomenclature of neocortical cell types. Nat. Neurosci. 23, 1456–1468 (2020).
53. Krienen, F. M. et al. Innovations present in the primate interneuron repertoire. Nature 586, 262–269 (2020).
54. Hinderer, C. et al. Severe Toxicity in Nonhuman Primates and Piglets Following High-Dose Intravenous Administration of an Adeno-Associated Virus Vector Expressing Human SMN. Hum. Gene Ther. 29, 285–298 (2018).
55. Coune, P. G., Schneider, B. L. & Aebischer, P. Parkinson's Disease: Gene Therapies. Cold Spring Harb. Perspect. Med. 2, a009431--a009431 (2012).
56. Piguet, F. et al. Rapid and Complete Reversal of Sensory Ataxia by Gene Therapy in a Novel Model of Friedreich Ataxia. Mol. Ther. 26, 1940–1952 (2018).
57. Combs, B., Kneynsberg, A. & Kanaan, N. M. Gene therapy models of alzheimer's disease and other dementias. in Methods in Molecular Biology vol. 1382 339–366 (2016).

Appendix B

SUPPLEMENTARY INFORMATION FOR CHAPTER III

B.1 Extended data figures

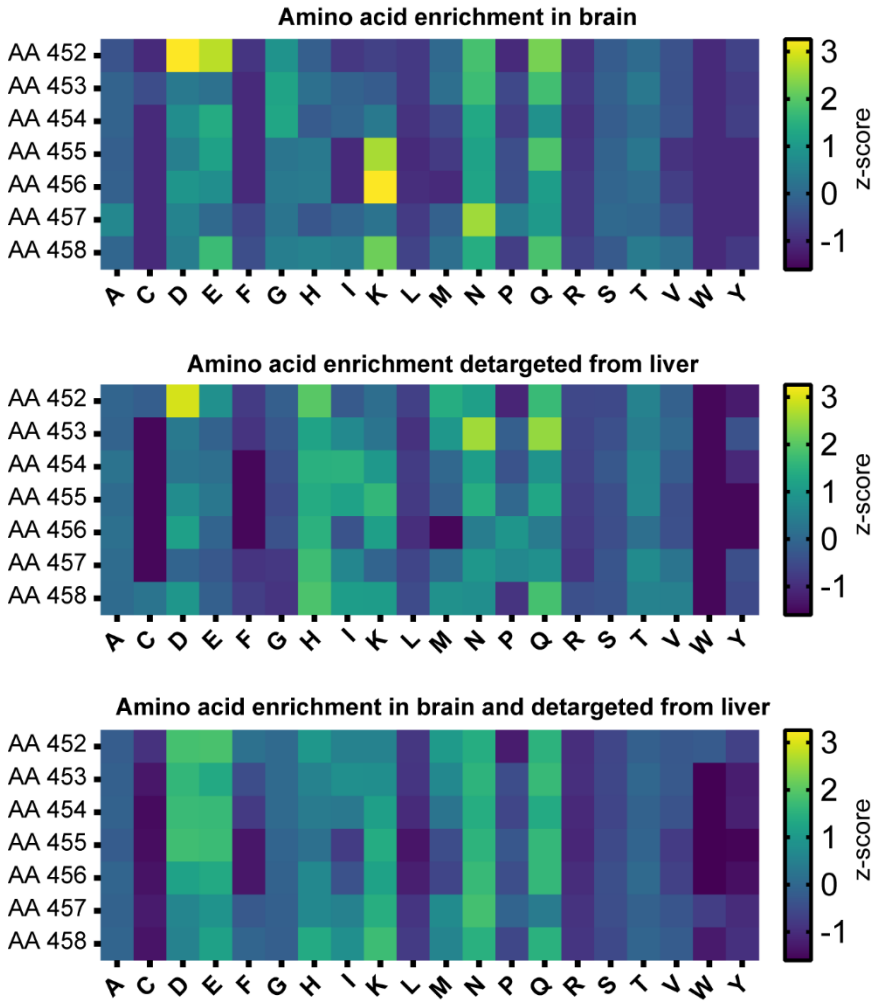

Figure B-1. Amino-acid contributions across the substitution library. Amino-acid contributions across the 7-mer substitution to variants enriched in the brain and targeted away from the liver. The 1k variants with the highest enrichment in the brain of hSyn-Cre animals, the 1k variants with the lowest enrichment in the liver of Tek-Cre animals, and all 39,034 variants with positive enrichment in the brain and negative enrichment in the liver from round 2 were analyzed. Plotted is the z-score of all amino acids at each position. Amino acids found at a higher prevalence than average are displayed in yellow and those found at a lower prevalence are displayed in purple

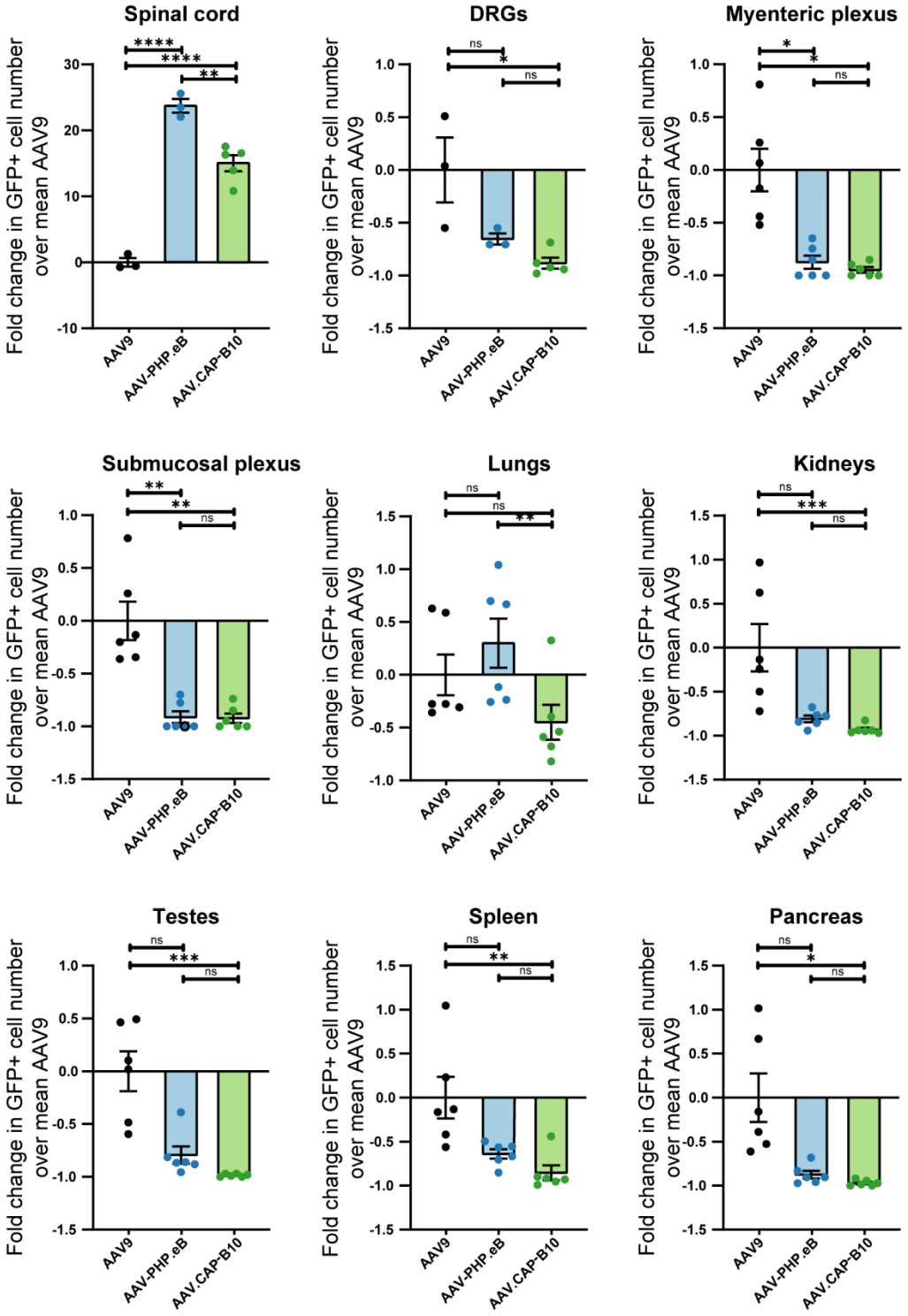
Spinal cord
DRGs
Myenteric plexus
Submucosal plexus
Lungs
Kidneys
Testes
Spleen
Pancreas
Fold change in GFP+ cell number over mean AAV9
AAV9
AAV-PHP.eB
AAV.CAP-B10
****
**
*
ns
***

Figure B-2. Quantification of decreased targeting from non-CNS organs in mice. ssAAV9:CAG-NLSx2-EGFP, ssAAV-PHP.eB:CAG-NLSx2-EGFP, and ssAAV.CAP-B10:CAG-NLSx2-EGFP were intravenously injected into male adult mice at 1×10^{11} vg/mouse. GFP fluorescence was assessed after three weeks of expression. The number of cells expressing GFP after delivery with AAV.CAP-B10 in the spinal cord is significantly increased relative to AAV9 and AAV-PHP.eB ($p = <0.0001$ [AAV9 vs AAV-PHP.eB], < 0.0001 [AAV9 vs AAV.CAP-B10], 0.0016 [AAV-PHP.eB vs. AAV.CAP-B10]). The number of cells expressing GFP after delivery with AAV.CAP-B10 in the DRGs is significantly decreased relative to AAV9 and not significantly different from AAV-PHP.eB ($p = 0.6488$ [AAV9 vs AAV-PHP.eB], 0.0229 [AAV9 vs AAV.CAP-B10], 0.5956 [AAV-PHP.eB vs. AAV.CAP-B10]). In the myenteric and submucosal plexi of the intestines, the number of cells expressing GFP after delivery with AAV.CAP-B10 is significantly decreased relative to AAV9 and not significantly different relative to AAV-PHP.eB (Myenteric Plexi: $p = 0.0165$ [AAV9 vs AAV-PHP.eB], 0.0147 [AAV9 vs AAV.CAP-B10], 0.6461 [AAV-PHP.eB vs. AAV.CAP-B10]; Submucosal Plexi: $p = 0.0066$ [AAV9 vs AAV-PHP.eB], 0.0095 [AAV9 vs AAV.CAP-B10], >0.9999 [AAV-PHP.eB vs. AAV.CAP-B10]). In the lungs, the number of cells expressing GFP after delivery with AAV.CAP-B10 is significantly decreased relative to AAV-PHP.eB and not significantly different from AAV9 ($p = 0.5286$ [AAV9 vs AAV-PHP.eB], 0.2805 [AAV9 vs AAV.CAP-B10], 0.0073 [AAV-PHP.eB vs. AAV.CAP-B10]). In the kidneys, spleen, pancreas, and testes, the number of cells expressing GFP after delivery with AAV.CAP-B10 transduction efficiency is significantly decreased relative to AAV9 and non-significantly changed relative to AAV-PHP.eB (Kidneys: $p = 0.1118$ [AAV9 vs AAV-PHP.eB], 0.0010 [AAV9 vs AAV.CAP-B10], 0.3894 [AAV-PHP.eB vs. AAV.CAP-B10]; Spleen: $p = 0.1363$ [AAV9 vs AAV-PHP.eB], 0.0024 [AAV9 vs AAV.CAP-B10], 0.5293 [AAV-PHP.eB vs. AAV.CAP-B10]; Pancreas: $p = 0.0677$ [AAV9 vs AAV-PHP.eB], 0.0452 [AAV9 vs AAV.CAP-B10], 0.2085 [AAV-PHP.eB vs. AAV.CAP-B10]; Testes: $p = 0.2497$ [AAV9 vs AAV-PHP.eB], 0.0005 [AAV9 vs AAV.CAP-B10], 0.1191 [AAV-PHP.eB vs. AAV.CAP-B10]). $n = 6$ mice per group except for the spinal cord and DRGs, which had $n = 3$ mice for the AAV9 and AAV-PHP.eB groups and $n = 5$ mice for the AAV.CAP-B10 group; mean ± SE; significance was determined using Ordinary one-way ANOVA with Tukey's correction for spinal cord, Brown-Forsythe and Welch one-way ANOVA tests with Dunnett's T3 correction for myenteric plexus and pancreas, Kruskal-Wallis one-way ANOVA test with Dunn's correction for DRGs, submucosal plexus, lungs, kidneys, spleen, and testes.

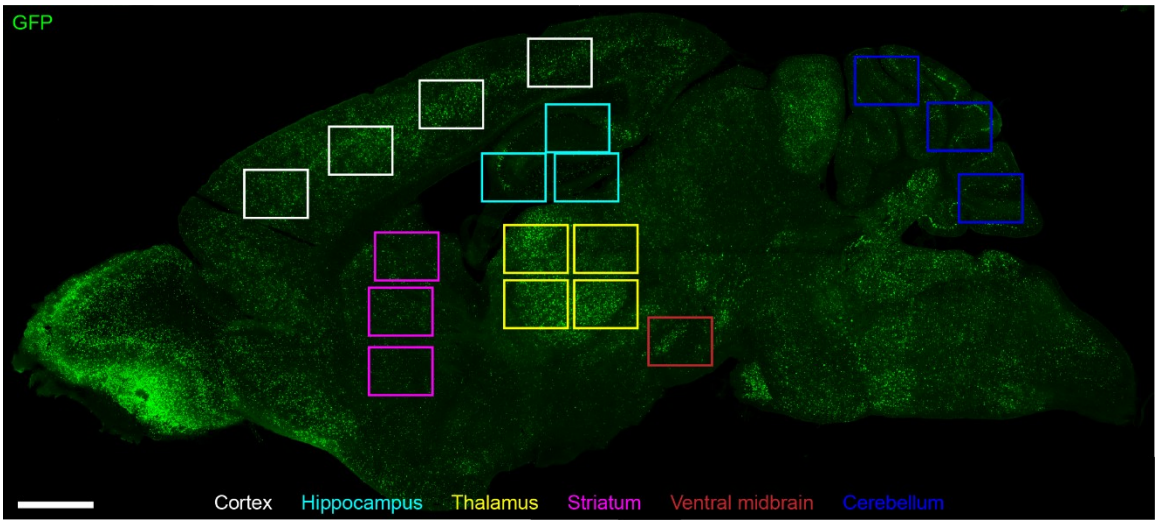

Figure B-3. Brain regions selected for cell-type quantification, depicted on brain expressing NLS-GFP after systemic delivery with AAV-PHP.eB. Four areas of cortex (white), three areas of hippocampus (teal), four areas of thalamus (yellow), three areas of striatum (purple), one area of ventral midbrain (red), and three areas of cerebellum (dark blue) were selected from each brain section quantified. Each section was located roughly 1200 µm from the midline for consistency across animals. Scale bar is 1mm.

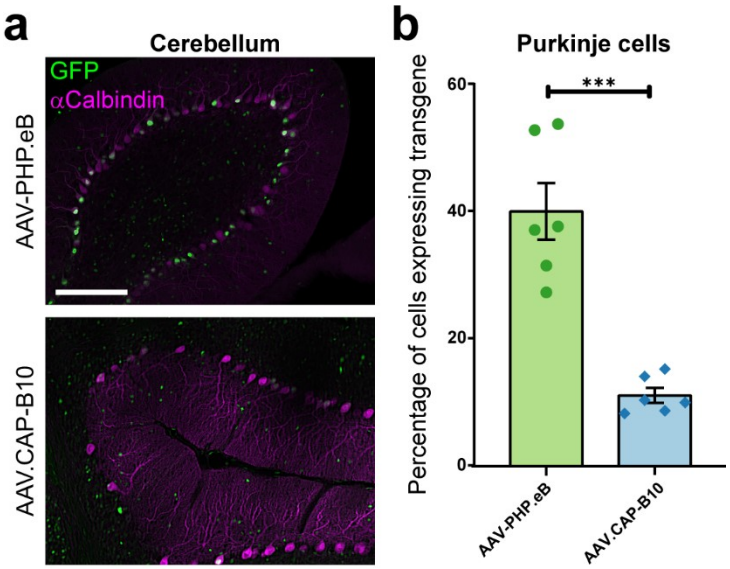

Figure B-4. Quantification of Purkinje cell expression. AAV.CAP-B10 is significantly targeted away from Purkinje cells in the cerebellum. ssAAV-PHP.eB:CAG-NLSx2-EGFP and ssAAV.CAP-B10:CAG-NLSx2-EGFP were intravenously injected into male adult mice at 1×10^{11} vg/mouse. GFP fluorescence was assessed after three weeks of expression. **a, b,** Quantification of the percentage of Purkinje cells displaying GFP expression in the cerebellum shows significantly fewer cells express EGFP after delivery by AAV.CAP-B10 than by AAV-PHP.eB (P = 0.0009). n = 6 mice per group, mean ± SE, two-tailed Welch's t test. Scale bar is 200 µm.

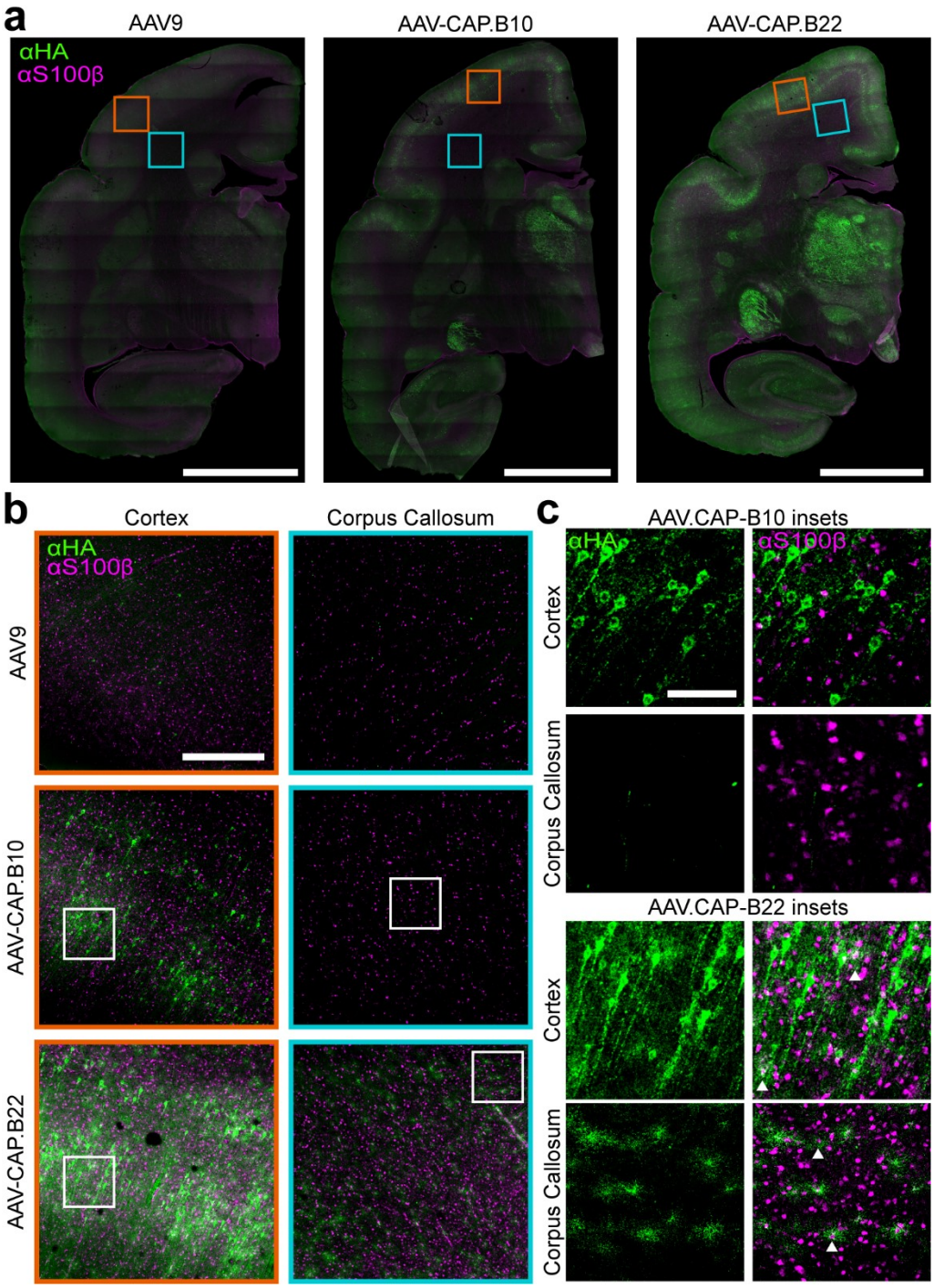
a
AAV9
AAV-CAP.B10
AAV-CAP.B22
αHA
αS100β
b
Cortex
Corpus Callosum
AAV9
AAV-CAP.B10
AAV-CAP.B22
c
AAV.CAP-B10 insets
AAV.CAP-B22 insets
Cortex
Corpus Callosum

Figure B-5. Astrocyte specificity of AAV9, AAV.CAP-B10, and AAV.CAP-B22 characterized through immunostaining for the HA tag in conjunction with S100β. Astrocyte specificity of AAV9, AAV.CAP-B10, and AAV.CAP-B22 characterized through immunostaining for the HA tag in conjunction with S100β. **a**, Coronal sections from marmoset brains after injection of individual variants, showing the locations of the areas selected for magnification. Scale bar is 5mm. **b**, The number of astrocytes expressing FXN was compared in magnified frames taken from the cortex and corpus callosum of animals injected with single variants. AAV9 and AAV.CAP-B10 display very little FXN expression in astrocytes, whereas AAV.CAP-B22 displays FXN expression in astrocytes in both the cortex and the corpus callosum. Scale bar is 500 µm. **c**, Further magnification indicates colocalization (examples marked with arrows) of HA and S100β immunostaining in both the cortex and corpus callosum of marmosets for AAV.CAP-B10 and AAV.CAP-B22. Scale bar is 100 µm.

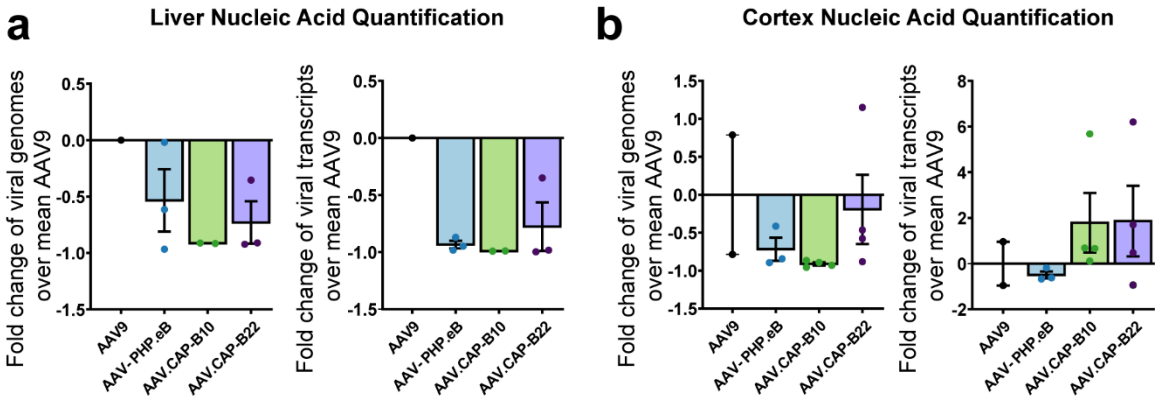

Figure B-6. Viral genome and RNA transcript measurements in cortex and liver tissue after single variant injection in marmosets. a, Viral genomes and RNA expression measurements in the liver corroborate protein expression data, indicating that AAV.CAP-B10 and AAV.CAP-B22 have decreased transduction and expression in the liver relative to AAV9. n = 1 for AAV9, 3 for AAV-PHP.eB, 2 for AAV.CAP-B10, 3 for AAV.CAP-B22. mean ± SE. **b**, Viral genome and RNA expression measurements in the cortex indicate no significant difference between experimental conditions. n = 2 for AAV9, 3 for AAV-PHP.eB, 4 for AAV.CAP-B10, 4 for AAV.CAP-B22, mean ± SE, two-tailed Welch's t-test. Values for both viral genomes and RNA transcripts were normalized by internal GAPDH control.

B.2 Extended data tables

Table B-1. Systemic marmoset injections of a CNS targeting engineered viral pool†.

MARMOSET ID#	AGE (YEARS)	SEX	WEIGHT AT INJECTION (KG)	DOSE (VG/KG)	EXPRESSION (DAYS)
MPV1	7.6	male	0.386	1.2×10^{14}	42
MPV2	11.5	male	0.375	7×10^{13}	42

Table B-2. Systemic marmoset injections of CNS targeting engineered viral variants††.

MARMOSET ID#	LOCATION	AGE (YEARS)	SEX	WEIGHT INJECTION (KG)	VARIANT INJECTED	DOSE (VG/KG)	LENGTH (DAYS)
MSV 1	NIH	11.5	female	0.300	AAV9	7×10^{13}	34
MSV 2	NIH	6.9	female	0.450	AAV.CAP-B10	7×10^{13}	34
MSV 3	NIH	12.1	male	0.360	AAV.CAP-B10	7×10^{13}	34
MSV 4	NIH	14.0	male	0.400	AAV.CAP-B22	7×10^{13}	34
MSV 5	SIAT	1.7	male	0.370	AAV9	7×10^{13}	36
MSV 6	SIAT	2.0	male	0.377	AAV.CAP-B10	7×10^{13}	40
MSV 7	SIAT	2.0	male	0.364	AAV.CAP-B22	7×10^{13}	29*
MSV 8	MIT	2.1	male	0.314	AAV-PHP.eB	7×10^{13}	42
MSV 9	MIT	2.8	female	0.356	AAV.CAP-B10	7×10^{13}	42
MSV 10	MIT	2.9	male	0.358	AAV.CAP-B22	7×10^{13}	42
MSV 11	NIH	6.5	male	0.504	AAV-PHP.eB	7×10^{13}	44
MSV 12	NIH	3.5	male	0.465	AAV-PHP.eB	7×10^{13}	48
MSV 13	NIH	10.0	female	0.320	AAV.CAP-B22	7×10^{13}	48

† MPV stands for Marmoset Pool Virus

†† MSV stands for Marmoset Single Virus

*Tissue collected after early death unrelated to the virus

B.3 Methods

B.3.1 Plasmids

The first-round viral DNA library was generated by amplification of a section of the AAV-PHP.eB capsid genome between AA450 and AA599 using NNK degenerate primers (Integrated DNA Technologies) to substitute AA452–AA458 with all possible variations. The resulting library inserts were then introduced into the rAAV-ΔCap-in-cis-Lox plasmid via Gibson assembly as previously described[1]. The resulting capsid DNA library, rAAV-Cap-in-cis-Lox, contained a diversity of ~1.28 billion variants at the AA level. The second-round viral DNA library was generated similarly to the first round, but, instead of NNK degenerate primers at the AA452–AA458 location, a synthesized oligo pool (Twist Bioscience) was used to generate only selected variants. This second-round DNA library contained a diversity of ~82,000 variants at the AA level.

The AAV2/9 REP-AAP-ΔCap plasmid transfected into HEK293T cells for library viral production was modified from the AAV2/9 REP-AAP plasmid previously used[26] by deletion of the AAs between 450 and 592. This modification prevents production of a WT AAV9 capsid during viral library production after a plausible recombination event between this plasmid co-transfected with rAAV-ΔCap-in-cis-Lox containing the library inserts.

Three rAAV genomes were used in this study. The first, pAAV-CAG-mNeonGreen (Addgene, 99134), uses a single-stranded (ss) rAAV genome containing the fluorescent protein mNeonGreen under control of the ubiquitous CMV-β-actin-intron-α-globin hybrid promoter (CAG). The second, pAAV-CAG-NLS-GFP (Addgene, 104061), uses an ssAAV genome containing the fluorescent protein EGFP flanked by two nuclear localization sites, PKKKRKV, under control of the CAG promoter. The third, pAAV-CAG-FXN-HA, uses an ssAAV genome containing an HA-tagged FXN protein under control of the CAG promoter

and harboring a unique 12-bp sequence in the 3′ untranslated region to differentiate different capsids packaging the same construct.

B.3.2 Viral production

rAAVs were generated according to established protocols[2]. In brief, HEK293T cells (American Type Culture Collection) were triple transfected using polyethylenimine; virus was collected after 120 h from both cell lysates and media and purified over iodixanol (OptiPrep, Sigma-Aldrich). The isolated variants investigated *in vivo* (AAV.CAP-B1, AAV.CAP-B2, AAV.CAP-B8, AAV.CAP-B10, AAV.CAP-B18, and AAV.CAP-B22) have similar production titer to AAV9, with normal titers around $1 \pm 0.7 \times 1012$ viral genomes per 15-cm dish.

A modified protocol was used for transfection and purification of viral libraries. First, to prevent mosaic capsid formation, only 10 ng of rAAV-Cap-in-cis-Lox library DNA was transfected (per 150-mm plate) to decrease the likelihood of multiple library DNAs entering the same cell. Second, virus was collected after 60 h, instead of 120 h, to limit secondary transduction of producer cells. Finally, instead of polyethylene glycol precipitation of the viral particles from the media, as performed in the standard protocol, media were concentrated more than 60-fold for loading onto iodixanol.

B.3.3 Animals

All rodent procedures were approved by the Institutional Animal Use and Care Committee of the California Institute of Technology. Transgenic animals, expressing Cre under the control of various cell-type-specific promoters, and C57Bl/6J WT mice (000664) were purchased from Jackson Laboratory. Transgenic mice included Syn1-Cre (3966), GFAP-Cre (012886), Tek-Cre (8863), and TH-Cre (008601). Mice were housed under standard conditions between 71 °F and 75 °F, in 30–70% humidity and on a light cycle of 13 h on and 11 h off. For round 1 and round 2 selections from the viral library, we used one male and one female mouse from each transgenic line (aged 8–12 weeks), as well as a single male C57Bl/6J mouse. For validation of individual viral variants, male C57Bl/6J mice aged 6–8

weeks were used. Intravenous administration of rAAV vectors was performed via injection into the retro-orbital sinus.

Marmoset (C. jacchus) procedures for MPV1 and MPV2 and for marmoset single variant 1–4 (MSV1–4) and MPV11–13 were approved by the Animal Care and Use Committee of the National Institutes of Mental Health (NIMH). MPV1, MPV2, MSV1–4, and MSV11–13 were born and raised in NIMH colonies and housed in family groups under standard conditions of 27 °C and 50% humidity. They were fed ad libitum and received enrichment as part of the primate enrichment program for NHPs at the National Institutes of Health (NIH). For AAV infusions, animals were screened for endogenous neutralizing antibodies. None of the animals that were screened showed any detectible blocking reaction at 1:5 dilution of serum (Penn Vector Core, University of Pennsylvania). They were then housed individually for several days and acclimated to a new room before injections. Two animals were used for the pooled injection study, both males, aged 7.6 (MPV1) and 11.5 (MPV2) years (**Appendix B**, **Table B-1**). Five animals were injected with single variants for characterization, but only four were usable (**Appendix B**, **Table B-2**), as one animal (AAV.CAP-B22 injected, 6.9 years, female, 0.475 kg) was found dead (27 d after injection), and, at necropsy, the pathology report indicated chronic nephritis unrelated to the virus. The day before infusion, the animals' food was removed. Animals were anesthetized with isoflurane in oxygen, the skin over the femoral vein was shaved and sanitized with an isopropanol scrub, and the virus (**Appendix B**, **Table B-1, 2**) was infused over several minutes. Anesthesia was withdrawn, and the animals were monitored until they became active, upon which they were returned to their cages. Activity and behavior were closely monitored over the next 3 d, with daily observations thereafter.

Marmoset procedures for MSV8–10 were approved by the Committee on Animal Care of the Massachusetts Institute of Technology (MIT), and all experiments were performed in accordance with the relevant guidelines and regulations. Marmosets were born and raised in an MIT facility accredited by the Association for Assessment and Accreditation of Laboratory Animal Care. Marmosets were housed in social groups under standard conditions

of 23.3 ± 1.1 °C, 50% ± 20% humidity and a 12-h light/dark cycle. They were fed ad libitum with standard diet as well as fruits, vegetables and various protein sources. Periodic neutralizing antibody testing of animals in the facility did not reveal significant levels of neutralizing antibodies against AAV9. Each of the animals was injected with single variants for characterization. The day before infusion, the animals' food was removed. Animals were sedated by alfaxalone; the skin over the cephalic vein was shaved and sanitized with an isopropanol scrub; and the virus (**Appendix B**, **Table B-2**) was infused through a 24-gauge catheter over several minutes. After the viral infusion was completed, animals were recovered on a warm water blanket (38 °C) until they regained normal motor functions. Then, animals were returned to their cages and monitored closely for normal behavior over the next 4 d, followed by daily observations thereafter.

Marmoset procedures for MSV5–7 were approved by the Institutional Animal Care and Use Committee of Shenzhen Institute of Advanced Technology (SIAT), Chinese Academy of Sciences. Marmosets were born and raised in SIAT colonies and housed in family groups under standard conditions of 22 ± 1 °C and 40–70% relative humidity. The marmoset breeding and housing facilities are accredited by the Association for Assessment and Accreditation of Laboratory Animal Care. Although animals were not screened for endogenous neutralizing antibodies, the animals were born and raised in the animal facility, and the housing environment for each animal was clean and isolated to prevent bacterial and viral infection. Therefore, the possibility of the animal carrying neutralizing antibodies for AAV virus is low. Before injection, marmosets were separated from family groups, housed two animals per each room for several days and acclimated to a new room before injections. Each of the animals was injected with single variants for characterization, but one animal (AAV.CAP-B22 injected, 2 years, male, 0.364 kg) was found dead (29 d after injection); at necropsy, the pathology report indicated that the death was unrelated to the virus. The day before infusion, the animals' food was removed. Animals were anesthetized with isoflurane in air, the skin over the saphenous vein was shaved and sanitized with an ethanol scrub, and the virus (**Appendix B**, **Table B-2**) was infused over several minutes. Anesthesia was withdrawn, and the animals were monitored until they became active, upon which they were

returned to their cages. Activity and behavior were closely monitored over the next 3 d, followed by daily observations thereafter.

B.3.4 DNA/RNA recovery and sequencing

Round 1 and round 2 viral libraries were injected into C57Bl/6J and Cre-transgenic animals (Syn1-Cre, GFAP-Cre, Tek-Cre and TH-Cre) at a dose of 8×10^{10} viral genomes per animal, and rAAV genomes were recovered 2 weeks after injection, as described in the M-CREATE protocol[3]. To determine the number of variants included in round 2, 0.01 times the enrichment of the top variant in each tissue was set as a threshold, and variants above that threshold were included. Mice were euthanized, and most major organs were recovered, snap-frozen on dry ice, and placed into long-term storage at −80 ºC. Tissues collected included brain, spinal cord, DRG, liver, lungs, heart, stomach, intestines, kidneys, spleen, pancreas, testes, skeletal muscle, and adipose tissue. Then, 100 mg of each tissue (~250 mg for brain hemispheres and <100 mg for DRG) was homogenized in TRIzol (Life Technologies, 15596) using a BeadBug (Benchmark Scientific, D1036), and viral DNA was isolated according to the manufacturer's recommended protocol. Recovered viral DNA was treated with RNase, underwent restriction digestion with SmaI (found within the inverted terminal repeats) to improve later rAAV genome recovery by polymerase chain reaction (PCR) and purified with a Zymo DNA Clean and Concentrator kit (D4033). Viral genomes flipped by Cre recombinase in select transgenic lines (or pre-flipped in WT animals) were selectively recovered using the following primers: 5′-CTTCCAGTTCAGCTACGAGTTTGAGAAC-3′ and 5′-CAAGTAAAACCTCTACAAATGTGGTAAAATCG-3′, after 25 cycles of 98 °C for 10 s, 60 °C for 15 s and 72 °C for 40 s, using Q5 DNA polymerase in five 25-µl reactions with 50% of the total extracted viral DNA as a template.

After Zymo DNA purification, samples from the WT C57Bl/6J animals were serially diluted from 1:10 to 1:10,000, and each dilution was further amplified around the library variable region. This amplification was done using the following primers: 5′-ACGCTCTTCCGATCTAATACTTGTACTATCTCTCTAGAACTATT-3′ and 5′-

TGTGCTCTTCCGATCTCACACTGAATTTTAGCGTTTG-3′, after ten cycles of 98 °C for 10 s, 61 °C for 15 s, and 72 °C for 20 s, to recover 73 bp of viral genome around and including the 21-bp variable region and add adapters for Illumina NGS. After PCR cleanup, these products were further amplified using NEBNext Dual Index Primers for Illumina sequencing (New England Biolabs, E7600), after ten cycles of 98 °C for 10 s, 60 °C for 15 s, and 72 °C for 20 s. The amplification products were run on a 2% low-melting-point agarose gel (Thermo Fisher Scientific, 16520050) for better separation and recovery of the 210-bp band. The dilution series was analyzed for each WT tissue and the highest concentration that resulted in no product from WT tissue on the gel was chosen for the amplification of the viral DNA from the transgenic animal tissues. This process was performed to differentiate between viral genomes flipped before packaging or due to Cre in the animal. Pre-flipped viral genomes should be avoided to minimize false positives in the NGS results.

All Cre-flipped viral genomes from transgenic animal tissues were similarly amplified (using the dilutions that do not produce pre-flipped viral genomes) to add Illumina sequencing adapters and subsequently for index labeling. The amplified products now containing unique indices for each tissue from each animal were run on a low-melting-point agarose gel, and the correct bands were extracted and purified with a Zymoclean Gel DNA Recovery kit.

Packaged viral library DNA was isolated from the injected viral library by digestion of the viral capsid and purification of the contained ssDNA. These viral genomes were amplified by two PCR amplification steps, like the viral DNA extracted from tissue, to add Illumina adapters and then indices. Correct bands were extracted and purified after gel electrophoresis. This viral library DNA, along with the viral DNA extracted from tissue, was sent for deep sequencing using an Illumina HiSeq 2500 System (Millard and Muriel Jacobs Genetics and Genomics Laboratory, California Institute of Technology).

A pool of eight viruses (AAV9, AAV-PHP.eB, AAV.CAP-B1, AAV.CAP-B2, AAV.CAP-B8, AAV.CAP-B10, AAV.CAP-B18, and AAV.CAP-B22) packaging CAG-FXN-HA with unique 12-bp barcodes were injected into two adult marmosets (**Table B-1**). After 6 weeks, animals were euthanized, and brain and liver were recovered and snap-frozen. Then, 1-mm

coronal sections from each tissue (4 mm for the brain) were homogenized in TRIzol (Life Technologies, 15596) using a BeadBug (Benchmark Scientific, D1036), and total RNA was recovered according to the manufacturer's recommended protocol. Recovered RNA was treated with DNase, and cDNA was generated from the mRNA using SuperScript III (Thermo Fisher Scientific, 18080093) and oligo(dT) primers according to the manufacturer's recommended protocol. Barcoded FXN transcripts were recovered from the resulting cDNA library, as well as the injected pool, using the following primers: 5′-TGGACCTAAGCGTTATGACTGGAC-3′ and 5′-GGAGCAACATAGTTAAGAATACCAGTCAATC-3′, after 25 cycles of 98 °C for 10 s, 63 °C for 15 s and 72 °C for 20 s, using Q5 DNA polymerase in five reactions using 50 ng of cDNA or viral DNA each as a template. After Zymo DNA purification, samples were diluted 1:100 and further amplified around the barcode region using the following primers: 5′-ACGCTCTTCCGATCTTGTTCCAGATTACGCTTGAG-3′ and 5′-TGTGCTCTTCCGATCTTGTAATCCAGAGGTTGATTATCG-3′, after ten cycles of 98 °C for 10 s, 55 °C for 15 s and 72 °C for 20 s. After PCR cleanup, these products were further amplified using NEBNext Dual Index Primers for Illumina sequencing (New England Biolabs, E7600), after ten cycles of 98 °C for 10 s, 60 °C for 15 s and 72 °C for 20 s. The amplification products were run on a 2% low-melting-point agarose gel (Thermo Fisher Scientific, 16520050) for better separation and recovery of the 210-bp band. All indexed samples were sent for deep sequencing as before.

Individual variants (AAV9, AAV-PHP.eB, AAV.CAP-B10 and AAV.CAP-B22) packaging CAG-FXN-HA with unique 12-bp barcodes were injected into 13 individual adult marmosets (**Table B-2**). After between 5 and 6 weeks, animals were euthanized, and brain and liver were recovered and snap-frozen. For DNA extraction from these tissues, 25-mg sections from the cortex and the liver were processed using a QIAamp DNA Mini Kit (Qiagen, 51304) to obtain purified viral and genomic DNA from the samples. For RNA transcript extration, 100-mg sections from the cortex and the liver (taken from consistent sections of tissue across animals) were homogenized in TRIzol (Life Technologies, 15596) using a BeadBug (Benchmark Scientific, D1036), and total RNA was recovered according

to the manufacturer's recommended protocol. From purified RNA, cDNA was generated with SuperScript IV VILO MasterMix (Thermo Fisher Scientific, 11766050) and oligo (dT) primers according to the manufacturer's recommended protocol (including the DNase step).

Viral genome and RNA transcript copy numbers were determined through qPCR as described59 using FXN-HA-specific primers: 5′-GACCTAAGCGTTATGACTGG-3′ and 5′-AATCTGGAACATCGTATGGG-3′. Within each sample, the viral genome or transcript copy number was normalized on a per-cell basis by quantifying GAPDH transcripts in each sample using the GAPDH-specific primers: 5′-TGTTCCAGTATGATTCCACC-3′ and 5′-GATGACCCTTTTGGCTCC-3′. DNA and RNA in tissues from animals MSV5, MSV6, and MSV7 were analyzed at the SIAT, whereas MPV1–2, MSV1–4, and MSV8–13 were analyzed at the California Institute of Technology.

B.3.5 NGS data alignment and processing

Raw FASTQ files from NGS runs were processed with M-CREATE data analysis code (available on GitHub at https://github.com/GradinaruLab/mCREATE) that align the data to an AAV9 template DNA fragment containing the 21-bp diversified region between AA452 and AA458, for the two rounds of AAV evolution/selection, or to an FXN-HA template containing the 12-bp unique barcode, for the marmoset virus pool. The pipeline to process these datasets involved filtering to remove low-quality reads, using a quality score for each sequence, and eliminating bias from PCR-induced mutations or high GC content. The filtered dataset was then aligned by a perfect string match algorithm and trimmed to improve the alignment quality. For the AAV engineering, read counts for each sequence were pulled out and displayed along with their enrichment score, defined as the relative abundance of the sequence found within the specific tissue over the relative abundance of that sequence within the injected viral library. For the pooled barcodes, read counts for each sequence were pulled out and normalized to the respective contribution of that

barcode to the initial, injected pooled virus to account for small inequalities in the amount of each member of the pool that was injected into the marmosets.

B.3.6 Tissue preparation, immunohistochemistry and immunofluorescence

Mice were euthanized with Euthasol and transcardially perfused with ice-cold 1× PBS and then freshly prepared, ice-cold 4% paraformaldehyde (PFA) in 1× PBS. All organs were excised and post-fixed in 4% PFA at 4 °C for 48 h and then sectioned at 50 µm with a vibratome. Immunofluorescence was performed on floating sections with primary and secondary antibodies in PBS containing 10% donkey serum and 0.1% Triton X-100. Primary antibodies used were rabbit anti-NeuN (1:200, Abcam, 177487), rabbit anti-S100 (1:200, Abcam, 868), rabbit anti-Olig2 (1:200, Abcam, 109186) and rabbit anti-Calbindin (1:200, Abcam, 25085). Primary antibody incubations were performed for 16–20 h at room temperature. The sections were then washed and incubated with secondary Alexa Fluor 647-conjugated anti-rabbit FAB fragment antibody (1:200, Jackson ImmunoResearch, 711-607-003) for 6–8 h at room temperature. For nuclear staining, floating sections were incubated in PBS containing 0.2% Triton X-100 and DAPI (1:1,000, Sigma-Aldrich, 10236276001) for 6–8 h and then washed. Stained sections were then mounted with ProLong Diamond Antifade Mountant (Thermo Fisher Scientific, P36970).

Marmosets were euthanized (Euthanasia, VetOne) and perfused with 1× PBS. One hemisphere of the brain (cut into coronal blocks) and the liver were flash-frozen in 2-methylbutane (Sigma-Aldrich, M32631) chilled with dry ice. The other hemisphere and organs were removed and post-fixed with 4% PFA at 4 °C for 48 h. These organs were then cryoprotected using 10% glycerol followed by 20% glycerol and flash-frozen in 2-methylbutane chilled with dry ice. The blocks of tissue were sectioned on an AO sliding microtome, except for spinal cord and DRG, which were cryosectioned. Then, 50-µm slices (20-µm for spinal cord and DRG) were collected in PBS, and immunohistochemistry or immunofluorescence was performed on floating sections.

To visualize cells expressing the HA-tagged FXN from the variant pool, slices were incubated overnight at room temperature with a rabbit anti-HA primary antibody (1:200, Cell

Signaling Technologies, C29F4). After primary incubation, sections were washed in PBS and then incubated with a biotinylated goat anti-rabbit secondary antibody (1:200, Vector Laboratories, BA1000) for 1 h at room temperature. Sections were again washed in PBS and incubated for 2 h in ABC Elite (Vector Laboratories, PK6100) as outlined by the supplier. The ABC peroxidase complex was visualized using 3,3′-diaminobenzidine tetrahydrochloride hydrate (Sigma-Aldrich, D5637) for 5 min at room temperature. Sections were then mounted for visualization.

To visualize cells expressing the HA-tagged FXN for individual variant injections, immunofluorescence staining was performed on floating sections with primary and secondary antibodies in PBS containing 10% donkey serum and 0.1% Triton X-100. Primary antibodies used were rat anti-HA (1:200, Roche, 3F10), rabbit anti-NeuN (1:200, Abcam, 177487) and rabbit anti-S100 beta (1:200, Abcam, 52642). Primary antibody incubations were performed for 16–20 h at room temperature. The sections were then washed and incubated with secondary anti-rat Alexa Fluor 488 (1:200, Thermo Fisher Scientific, A-21208) and anti-rabbit Alexa Fluor 647 (1:200, Thermo Fisher Scientific, A32795). Stained sections were then washed with PBS and mounted with ProLong Diamond Antifade Mountant.

B.3.7 Imaging and quantification

All CAG-mNeonGreen-expressing tissues were imaged on a Zeiss LSM 880 confocal microscope using a Fluar ×5 0.25 M27 objective, with matched laser powers, gains, and gamma across all samples of the same tissue. The acquired images were processed in Zen Black 2.3 SP1 (Zeiss).

All CAG-NLS-GFP-expressing tissues were imaged on a Keyence BZ-X all-in-one fluorescence microscope at 48-bit resolution with the following objectives: PlanApo-λ ×20/0.75 (1 mm working distance) or PlanApo-λ ×10/0.45 (4 mm working distance). For co-localization of GFP expression to antibody staining, in some cases the exposure time for the green (GFP) channel was adjusted to facilitate imaging of high- and low-expressing cells while avoiding oversaturation. In all cases in which fluorescence intensity was compared

between samples, exposure settings and changes to gamma or contrast were maintained across images. To minimize bias, multiple fields of view per brain region and peripheral organ were acquired for each sample. For brain regions, the fields of view were matched between samples and chosen based on the antibody staining rather than GFP signal. For peripheral tissues, fields of view were chosen based on the DAPI or antibody staining to preclude observer bias.

Marmoset tissue sections transduced with the variant pool were examined and imaged on a Zeiss AxioImager Z1 with an Axiocam 506 color camera. Acquired images were processed in Zen Blue 2 (Zeiss).

Marmoset tissues transduced with individual variants were imaged on a Zeiss LSM 880 confocal microscope. Tissues from animals MSV5, MSV6, and MSV7 were imaged at the SIAT, whereas all other marmosets were imaged at the California Institute of Technology. Whole tissue sections were imaged using a Fluor ×5 0.25 M27 objective with a gallium arsenide phosphide photomultiplier tube detector at a pixel size of 1.25 µm × 1.25 µm. Images for cortex quantification were imaged with a LD-LCI Plan-Apochromat ×10/0.45 M27 objective with a photomultiplier tube detector at a pixel size of 0.42 µm × 0.42 µm. Images for liver quantification were taken with an LD-LCI Plan-Apochromat ×10/0.45 M27 objective with a photomultiplier tube detector at a pixel size of 0.42 µm × 0.42 µm or an LD-LCI Plan-Apochromat ×20/0.8 M27 with a gallium arsenide phosphide photomultiplier tube detector at a pixel size of 0.156 µm × 0.156 µm (MSV5, MSV6, and MSV7). Imaging at each magnification was performed with matched laser powers, gains, and gamma across all samples of the same tissue imaged in the same location, whereas images taken in different locations were matched by maximizing the range indicator. The acquired images were processed in Zen Black 2.3 SP1 (Zeiss). Quantification of transgene expression in marmoset cortex was performed through manual cell counting of co-localized HA+ cells with NeuN+ staining in maximum intensity projection images. Quantification of transgene expression in the marmoset liver was performed through manual cell counting of co-localized HA+ cells with DAPI staining in maximum intensity projection images.

All image processing was performed with the Keyence BZ-X Analyzer (version 1.4.0.1). Co-localization between the GFP signal and antibody or DAPI staining was performed using the Keyence BZ-X Analyzer with the hybrid cell count automated plugin. Automated counts were validated and routinely monitored by comparison with manual hand counts and found to be below the margin of error for manual counts.

To compare total cell counts and fluorescence intensity throughout the brain between samples, an entire sagittal section located 1,200 μM from the midline was imaged using matched exposure conditions with the Keyence BZ-X automated XY stitching module. Stitched images were then deconstructed in the Keyence BZ-X Analyzer suite and run through the hybrid cell count automated plugin to count the total number of cells in the entire sagittal section. Average fluorescence intensity was calculated by creating a mask of all GFP-positive cells throughout the sagittal section and measuring the integrated pixel intensity of that mask. The total integrated pixel intensity was divided by the total cell count to obtain the fluorescence intensity per cell. In all cases where direct comparisons of fluorescence intensity were made, exposure settings and post-processing contrast adjustments were matched between samples.

B.2.8 Statistics and reproducibility

For the initial characterization of brain and liver expression in mice (**Figure 3-1d**), the experiment was repeated with n = 3. For the statistical analysis in mice and related graphs (**Figure 3-2** and **Figure B-2, 3, 4**), a single data point was defined as two tissue sections per animal, with multiple technical replicates per section when possible. Technical replicates were defined as multiple fields of view per section, with the following numbers for each region or tissue of interest: cerebellum = 3, cortex = 4, hippocampus = 3, midbrain = 1, striatum = 3, thalamus = 4, liver = 4, spleen = 2, testis = 2, kidney = 2, lung = 2, spine = 1, DRG = 1, and whole sagittal = 1. Unless otherwise noted, all experimental groups were n = 6, determined using preliminary data and experimental power analysis. Normality was tested to ensure that the data matched the assumptions of the statistical tests used.

For the initial pooled experiments in marmosets (**Figure 3-3**), the experiment was repeated in two separate animals (**Table B-1**). For the statistical analysis in marmoset and related graphs (**Figure 3-4** and **Figure B-5**, **6**), a single data point was defined as a single tissue section from a single animal. The data were collected across four separate cohorts in three separate locations (**Table B-2**), and the trend was recapitulated in each. For representative brain images and quantification (**Figure 3-4a**,**b** and **Figure B-6b**), the experiment was repeated n times, where n = 2 (AAV9), n = 3 (AAV.PHP.eB), n = 4 (AAV.CAP-B10), and n = 4 (AAV.CAP-B22), except for astrocyte staining (**Figure B-5**), which was repeated n = 2 (AAV9), n = 3 (AAV.CAP-B10) and n = 2 (AAV.CAP-B22). For representative liver images and quantification (**Figure 3-4c,d** and **Figure B-6a**), the experiment was repeated n times, where n = 1 (AAV9), n = 3 (AAV.PHP.eB), n = 2 (AAV.CAP-B10), and n = 3 (AAV.CAP-B22). Global brain analysis of AAV.CAP-B10 nervous system expression (**Figure 3-4e,f**) was performed with n = 2. No statistical methods were used to predetermine sample sizes, but our sample sizes were similar to those reported in our previous publications[4,5]. Data distribution was assumed to be normal, but this was not formally tested.

No data were excluded from analysis. Allocation of organisms and samples to separate groups was random, and animals were allocated to experimental conditions based on availability in each cohort. The investigators were not blinded to allocation during experiments and outcome assessment.

Microsoft Excel for Microsoft 365 (version 2107) and GraphPad Prism 8 for Windows (version 8.4.3 (686)) were used for statistical analysis and data representation. For all statistical analyses, significance is represented as $*P \leq 0.05$, $**P \leq 0.01$, $***P \leq 0.001$, and $****P \leq 0.0001$; not significant, $P \geq 0.05$.

B.3.9 Data availability

The NGS datasets for capsid selection and marmoset pooled screening that are reported in this article are available under Sequence Read Archive accession code PRJNA769435. The following vector plasmids were deposited in Addgene for distribution, and viruses might be available for commonly packaged genomes (http://www.addgene.org): AAV.CAP-B10

(Addgene, 175004) and AAV.CAP-B22 (Addgene, 175005). All other constructs and tools will be available through the Beckman Institute CLOVER Center (https://clover.caltech.edu/). The data that support the findings of this study are available from the corresponding author upon reasonable request.

B.3.10 Code availability

The codes used for M-CREATE data analysis were published previously[33] and are available on GitHub: https://github.com/GradinaruLab/mCREATE.

Bibliography for Appendix B

1. Deverman, B. E. et al. Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat. Biotechnol. 34, 204–209 (2016).
2. Challis, R. C. et al. Systemic AAV vectors for widespread and targeted gene delivery in rodents. Nat. Protoc. 14, 379–414 (2019).
3. Kumar, S. et al. Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. Nat. Methods 17, 541–550 (2020).
4. Matsuzaki, Y. et al. Intravenous administration of the adeno-associated virus-PHP.B capsid fails to upregulate transduction efficiency in the marmoset brain. Neurosci. Lett. 665, 182–188 (2018).
5. Hinderer, C. et al. Severe Toxicity in Nonhuman Primates and Piglets Following High-Dose Intravenous Administration of an Adeno-Associated Virus Vector Expressing Human SMN. Hum. Gene Ther. 29, 285–298 (2018).

Chapter IV

SIMULTANEOUS ENGINEERING OF MULTIPLE VARIABLE REGIONS OF THE ADENO-ASSOCIATED VIRAL CAPSID

ABSTRACT

Engineered adeno-associated virus (AAV) capsids are traditionally produced through *in vivo* selection of standard 7-mer insertions between AA588 and AA589 of AAV viral proteins. However, these capsid variants frequently transduce off-target tissues, potentially due to the continued presence of parental capsid remaining on prominent surface-exposed regions or the flexibility of the inserted residues. Efforts to substitute these residues and broadly mutate the capsid surface beyond 7-mers have been limited due to combinatorial issues during production. By implementing early screening of capsid libraries for production, each of the surface-exposed variable regions of the three-fold axis can be diversified prior to selection. Pre-selected libraries at each site can then be combined to yield capsids with all prominent, surface-exposed residues mutated. In conjunction with improved transfection and screening methods, the selection capabilities of AAV can be expanded, revealing a broader protein landscape and potentially unlocking previously unattainable interactions with cell-surface receptors for the AAV capsid.

4.1 Introduction

Adeno-associated capsid engineering has yielded exciting vectors for gene therapy and neuroscience research. Capsid engineering has primarily occurred in variable region (VR) VIII using 7-mer AA insertions, due to production tolerance and previous success[1–4]. Through 7-mer amino acid insertion at VR VIII and DNA-based selection aided by Cre-recombinase transgenics, variant AAV.PHP-B enables blood-brain barrier crossing and broad tropism in the brain after systemic delivery[2]. However, this variant (along with the further engineered PHP.eB[1]) fails to translate between mouse strains[5] or between species[6]. These variants also transduce the liver, which can promote immune response and impair clinical efficacy and safety of these variants[7–11]. With updated methodology implementing differential[12] or RNA-based selections[4], 7-mer amino acid insertions at VR VIII repeatedly yield variants that translate between mouse strains[4,12].

Examination of the cryo-EM structure of PHP.eB, containing a 7-mer insertion between AA588 and AA589, indicates high structural flexibility of this insertion[13]. However, the structural flexibility of the insertion residues may not be best suited for all receptor chemistries. Indeed, across separate selections and methods variants repeatedly emerge which target a limited panel of receptors[14,15]. Through implementation of substitution libraries, we could instead leverage the natural structure of the AAV capsid to access receptor pairings potentially inaccessible to flexible insertions.

Variable regions located within the three-fold axis of symmetry of the AAV capsid interact closely and are implicated in receptor binding[16]. Persisting residues from the parental capsid sequence in these variable regions can cause off-target tropism[17] or neutralization by the immune system[18] due to pre-existing antibodies against common serotypes[19]. Substitution libraries between AA452-458 in this region have been implemented to select for increased transduction[17] or cross-species compatibility[20]. Iterative engineering of these residues allows tropism to be refined[21], but this limits the explorable protein landscape and creates dependency on previous mutations. Parallel engineering of multiple variable regions has not been explored, likely due to the combinatorial limitations during production. This limitation

is based on the exponentially increasing combinatorial complexity as the number of mutations increases[22,*]. As a result, capsid variants containing 21 AA mutations within the 3-fold axis of symmetry have not been previously produced, to our knowledge. Machine learning models trained to increase AAV capsid diversity (across select residues) were limited to generating capsids with 19 AA substitutions spread throughout the capsid structure[23].

To simultaneously engineer prominent residues within each of the surface-exposed variable regions of the three-fold axis, we screened capsid libraries within each single variable region for production. Pre-selected libraries at each site were then combined to yield capsids with all prominent, surface-exposed residues mutated. We discovered that our parallel strategy yielded diverse capsid libraries across multiple variable regions of the AAV capsid. Expanding the sequence space for directed evolution through engineering diversity at these variable regions could enable the discovery of AAV vectors with desirable tissue tropism, cell-type specificity, attenuated transduction in off-target organs, or neutralizing antibody evasion.

Viral genome based capsid selections have previously relied on transgenic animals[1,2,12], while recent mRNA-based selections are vulnerable to promoter bias and transcriptional noise. DNA based selections without the necessity of Cre-recombinase transgenics could facilitate directed evolution selections of capsid libraries in new contexts (i.e. mouse strains or species where Cre-transgenics aren't available). Additionally, combination of T7 with developed mRNA-based selections[3,4] could increase our confidence in selection data. To avoid the dependency upon transgenic animals for future selections, amplification using a T7 promoter was validated to detect viral transduction without the necessity for Cre-transgenic animals.

---

[*] The engineerable landscape of AAV has historically been limited to 7-mer AA mutations due to an inability to transfect library complexities beyond this. This limitation is based on the exponentially increasing combinatorial complexity as the number of mutations increase. Within a 7-mer library manufactured with NNK degenerate primers, 3.44×10^{10} nucleotide variants are present, corresponding to 262 ng of DNA 7411 nucleotides in length. Extension of the mutation region by an additional amino acid, an 8-mer, increases the variant diversity to 1.79×10^{12}, corresponding to 13.6 µg of DNA.

4.2 Increasing the production scale of diverse libraries

To generate diverse input libraries, an amplicon containing a variable sequence is typically combined with a vector containing the components for AAV production. While this protocol can yield high-quality diverse libraries, it scales poorly for large productions. To increase the output quantity from this process, rolling circle amplification and digestion with TelN proteolomerase has been applied previously[4]. To confirm that viral library amplified with RCA and linearized by TelN is capable of viral production, library vector producing AAV9 and containing one or two TelN cut sites was compared to an uncut and unamplified plasmid control (**Appendix C**, **Figure C-1A**). The loss in production was least for the single TelN cut linear DNA compared to the intact plasmid control. Therefore, a single TelN cut site was included in the AAV capsid library plasmid.

Efforts to expand library capabilities beyond 7-mers have been impeded by production limitations. Selection of capsid libraries requires a connection between phenotype and genotype[24]. During capsid library production, concerns of cross-packaging, where an incorrect genome is packaged due to co-transfection or infection with produced viral particles, have limited transfection inputs[1,2,12,17,21], despite literature supporting higher transfection rates[25,26]. To maintain connection between phenotype and genotype during selection and avoid capsid switching, 10 ng transfections are routinely performed. Based on the mechanism of polyethylenimine (PEI), which forms polyplexes with many DNA plasmids[27], it is possible that capsid switching is an innate property of PEI transfection rather than a variable of transfection rate. To test this, we produced a sequence pool containing 82,710 unique capsid variants (each with a codon replicate)[21]. In this production, we included an additional sequence containing stop codons at a frequency ~10 fold higher than other variants in the pool. This variant pool was transfected using PEI over several orders of magnitude of input, purified, and sequenced to determine whether capsid switching increased as a function of transfection rate. Across transfection inputs, increasing the amount of input library commensurately increased the production of the variant pool (**Figure 4-1A, Figure C-1B**), indicating the utility of increasing transfection quantity of viral libraries.

The virus libraries were sequenced to determine whether stop codon integration into viral capsids increased as a function of transfection rate. Across transfections scales for linear libraries, stop codons measured in viral genomes did not appear to correlate with production. Plasmid transfection at 10 ng, serving as a positive control, also produced sequences containing stop codons, although at a lower relative frequency compared to all transfected linear DNA (**Appendix C**, **Figure C-1C**). These experiments indicate that the transfection rate can be increased for library preparations, corroborating previous literature[4,25,26,28].

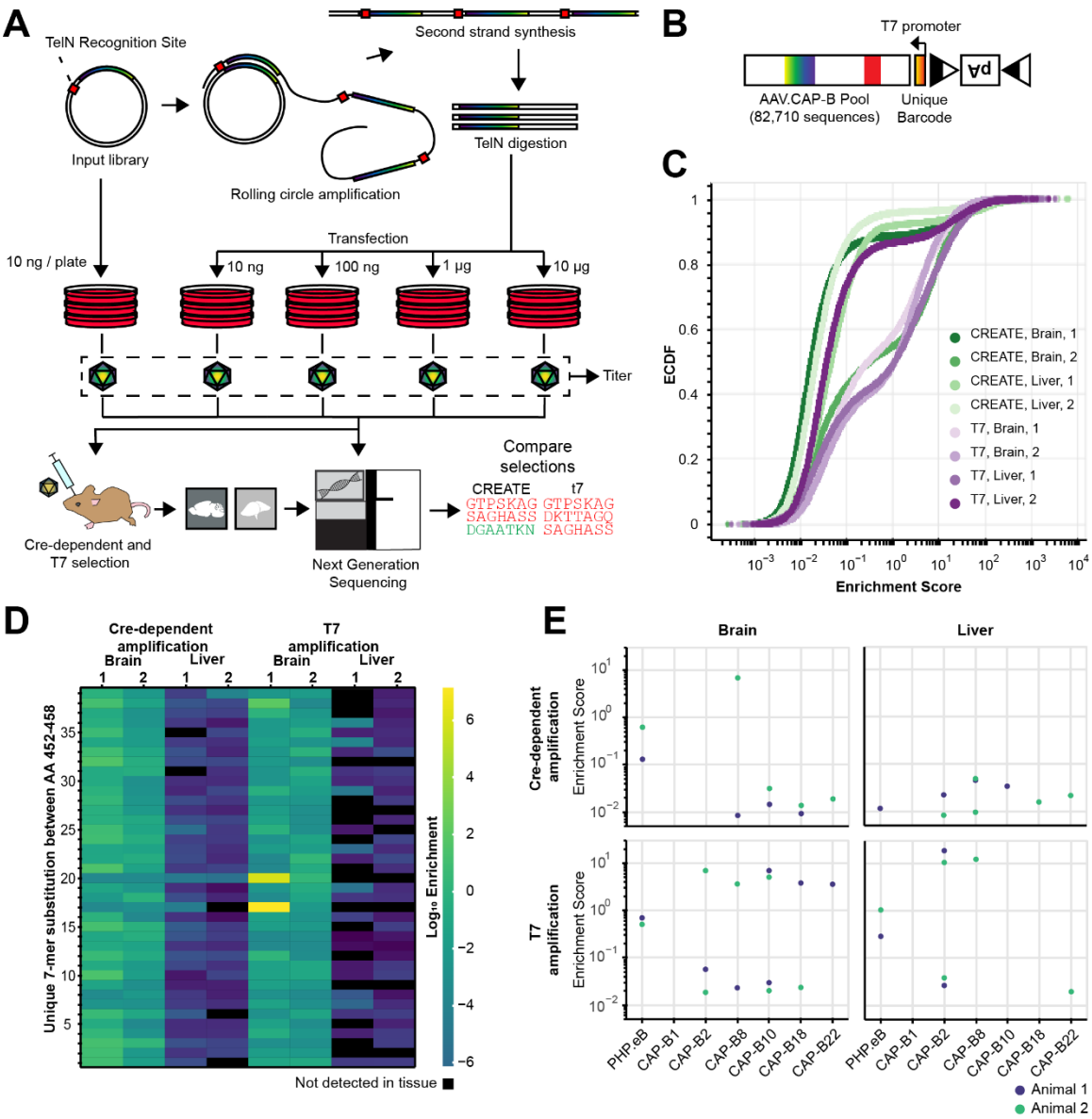

Figure 4-1. Comparison between T7 promoter and Cre-dependent amplification A

diverse library was amplified through rolling circle amplification and digested with TelN prior to transfection. Direct transfection was used as a positive control. Viral libraries from different transfection scales were titered (see **Appendix C**, **Figure C-1**) and combined prior to systemic injection into Cre-recombinase transgenic mice. Brain and liver were collected and sequenced, in addition to the input viral pool. (B) The CAP-B round 2 library was used for production along with a unique barcode to quantify capsid switching (see **Appendix C, Figure C-1**), while a T7 promoter sequence was included to investigate potential Cre-recombinase-free DNA measurement. (C) The empirical cumulative density function of enrichment scores across both brain and liver of two animal replicates are compared between measurement methods. Cre-recombinase-dependent amplified samples are shown in green, while T7 amplified samples are shown in purple. (D) Variants that were enriched across all brain samples and enriched in no liver samples are identified and compared across samples by $\log_{10}$ enrichment score (see **Appendix C**, **Figures C-2, C-3, C-4, C-5**). (E) Comparison of Cre-dependent amplification and T7 amplification by the detection of previously characterized variants which display brain tropism and liver attenuation compared to AAV9.

4.3 Validation of short promoter element for DNA based capsid selection

To select capsid variants enriched in a specific tissue, historically Cre-recombinase-based selections were used[1,2,12]. However, Cre-recombinase transgenics restrict the species and strains for capsid selections. Recently, RNA-based selections have emerged as an alternative measurement technique, yielding valuable AAV capsid variants to target the CNS or muscle in rodents[3,4]. However, promoter choice can bias selections to specific tissues or require use of that promoter for applications. Therefore, validation of a small promoter element for DNA selection in conjunction with RNA selections could allow increased confidence during capsid selection. With this aim, the T7 promoter element[29] was cloned into the library plasmid upstream of the polyA sequence, flanked by the lox71 and lox66 sites typically used for CREATE selections[2].

The previously prepared viral libraries containing 82,710 capsid variants were produced containing this element (**Figure 4-1B**) and systemically injected into male and female hSyn and Tek Cre-recombinase transgenic mice (**Figure 4-1A**) to determine whether recovery of viral genomes through T7 amplification could identify enriched capsid variants. Two weeks after injection, the brain and liver of each animal was collected, and viral genomes were isolated from tissue. From these viral genomes, the CREATE protocol[2] and T7 amplification protocol[29] were performed in parallel and prepared for deep sequencing (**Appendix C.2.5**).

Comparable read depths were obtained through next-generation sequencing of these samples (**Table C-8**). We quantified variant enrichment, defined as the relative abundance of a sequence in tissue over the relative abundance of a sequence in the input library, and found that T7 amplification was able to produce highly diverse libraries, in both the total unique sequences obtained (**Table C-8**) and their enrichment scores (**Figure 4-1C**). Examination of the enrichment scores between tissues, biological replicates, and amplification mode indicate that a variant library of this size (165,420 sequences) results in many variants present in one sample (with varying degrees of enrichment) and absent in the other (**Appendix C**, **Figures C-2, C-3**). This result indicates that, for a variant library of similar magnitude, DNA-based selections are not inherently reproducible. When variants that are absent in either sample are excluded and capsid variants are compared for their enrichment in each sample, four classifications emerge (**Appendix C**, **Figures C-4, C-5**). Variants either fail to be enriched in either sample, are enriched in one of the two samples, or are enriched in both samples. There does not seem to be any systemic error induced using T7 amplification instead of CREATE, based on comparison of enrichment between selection methods, tissues, and replicates. From these selections, we identified 38 capsid variants which were enriched in all brain samples, across both selection methods, and not enriched in any liver samples (**Figure 4-1D**).

This capsid library had previously been used to derive six variants which maintain brain tropism and acquire liver attenuation: AAV.CAP-B1, AAV.CAP-B2, AAV.CAP-B8, AAV.CAP-B10, AAV.CAP-B18, and AAV.CAP-B22. In addition, AAV-PHP.eB was present in the variant library as an internal control. Therefore, measuring these known variants through the different selection methods can indicate the utility of T7 amplification for DNA-based selection. Using known capsid variants as benchmarks, we can see that T7 amplification performs well compared to Cre-dependent amplification (**Figure 4-1E**). These variants were more frequently enriched in the brain with T7 amplification compared to Cre-dependent amplification. Additionally, absence in the liver was more frequently observed with T7 amplification. In exception to this are AAV.CAP-B2 and AAV.CAP-B8, which were enriched in the liver using T7 amplification.

4.4 Generation of diverse libraries at multiple sites

Due to the variability in the surface-exposed regions in the three-fold axis of symmetry, we wanted to extensively mutate this region through AA substitution, with the hope that an expanded protein landscape could enable discovery of AAV vectors with new receptor binding capability, specific tropism, or the ability to evade neutralizing antibodies. Rather than expand the number of AAs mutated at once, we attempted parallel mutations of 7-mer AA substitutions between AA452-458 (located within VR IV), AA492-498 (located within VR V), and AA585-591 (located within VR VIII) separately. These three 7-mer mutations would broadly cover the three-fold axis of symmetry of the AAV9 capsid surface (**Figure 4-2**).

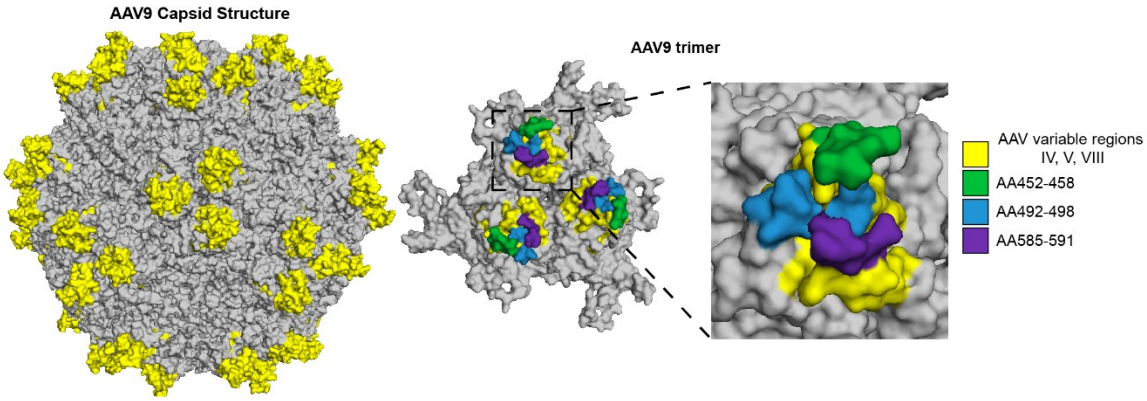

Figure 4-2. AAV9 viral capsid diversity and substitution sites Viral capsid diversity of AAV9 in the three-fold axis of symmetry is highlighted in yellow across the surface of the viral capsid. On an individual trimer, the coverage of three separate substitutions are shown to display the coverage attainable through multi-site engineering.

We generated 7-mer substitutions within each variable region in parallel (**Figure 4-3A**). The empirical cumulative density function of read counts in both the input DNA libraries (**Figure 4-3B**) and the produced viral capsid libraries (**Figure 4-3C**) indicated diversity, as more than half the sequences are present as a single read count. Substitution libraries between AA492-498 display a relative drop in diversity compared to the other two substitutions, due to decreased efficiency when the number of inserts is increased during DNA assembly (see **Appendix C.2.2** and **Table C-3**).

The Hamming distance, a quantification of the AA sequence differences between each capsid variant and the wild-type AAV, was calculated within each substitution library (**Appendix C**, **Figure C-6**). In the input DNA libraries (**Appendix C**, **Figure C-6A**), each of substitution libraries are distinct from AAV9: more than 70% of the sequences are distinct in all seven residues, while an additional ~25% of the library is distinct at six residues. This diversity is maintained through viral production (**Appendix C**, **Figure C-6B**), as the viral genomes produced from each substitution library display a similar quantity of mutations.

In the input DNA libraries, codon bias is observed across each site (**Figure 4-3D**). Between AA452-458 and AA492-498, amino acids serine, phenylalanine, and leucine are overrepresented while tryptophan, glutamine, lysine, and glutamic acid are underrepresented. Between AA585-591, amino acids serine, threonine, asparagine, and leucine are overrepresented, while tryptophan and glutamic acid are underrepresented. A likely cause for the difference between input library diversity is the use of NNM for the degenerate primers for AA585-591. After production, the libraries become more heterogenous across the individual residues and less tolerable residues, cysteine and tryptophan, as well as stop codons, become less prevalent (**Figure 4-3E**). Some codon bias is maintained after production: serine and leucine maintain prevalence in the AA452-458 and AA492-498 substitutions, while serine, threonine, asparagine, and leucine maintain prevalence in the AA585-591 substitution. Some enrichment of residues occurs during single site production (i.e. aspartic acid increases in prevalence in all libraries, while asparagine becomes more prevalent between AA452-458 and AA492-498).

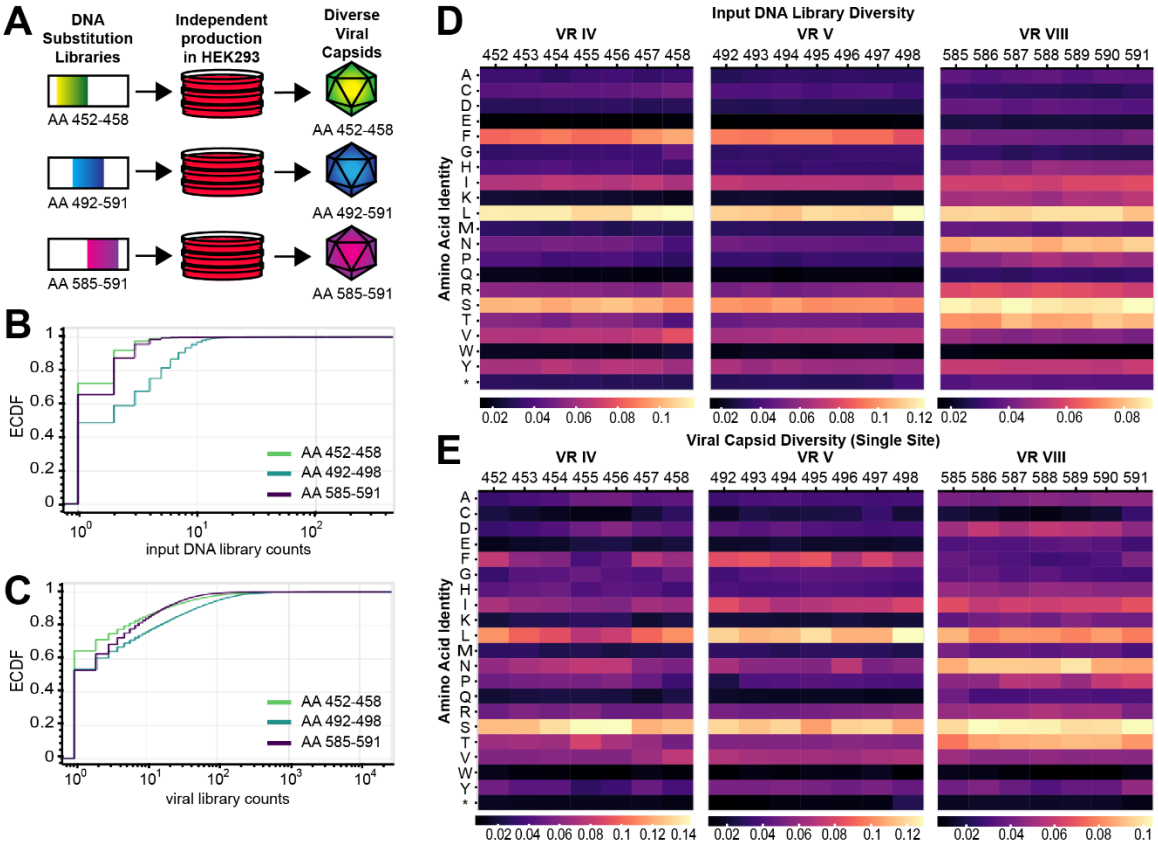

Figure 4-3. Generating diversity across multiple variable regions of AAV9 (A) Parallel production strategy for each substitution library. (B) Read counts for each individual library were sub-sampled for 500,000 sequences and the empirical cumulative density function compared for the input DNA of each substitution library (see **Appendix C**, **Figure C-6**). (C) Read counts for each individual library were sub-sampled for 500,000 sequences and the empirical cumulative density function compared for the viral genomes of each substitution library (see **Appendix C**, **Figure C-6**). (D) Amino acid residue frequency at each position, within each substitution region, are displayed for the input DNA of each substitution library. (E) Amino acid residue frequency at each position, within each substitution region, are displayed for the viral genomes produced in each substitution library.

Capsid variants capable of production in individual substitution libraries were then purified, isolated, amplified, and recombined together to generate DNA libraries diverse between either two (only AA452-458 and AA585-591) or three sites (**Figure 4-4A**). The DNA libraries resulting from recombination were diverse (**Figure 4-4B**) and do not seem to acquire additional changes in the amino acid profile across positions (**Figure 4-4C, D**), reinforcing the use of RCA for high yield production of diverse libraries. These variant

libraries were then produced, purified, isolated, and sequenced to determine the extent of diversity and bias at each site.

The Hamming distance was also calculated for each multi-site substitution library (**Appendix C**, **Figures C-7, C-8**). Both the input DNA libraries (**Appendix C**, **Figures C-7A, C-8A**) and the viral genomes contain substantial diversity (**Appendix C**, **Figures C-7B, C-8B**). The two-site DNA library (**Appendix C**, **Figures C-7A**) is distinct from AAV9 at all 14 positions in ~40% of sequences, while another ~40% is distinct in 13 positions. The three-site DNA library (**Appendix C**, **Figures C-8A**) is distinct from AAV9 at all 21 positions in ~20% of sequences, while another ~60% is distinct in 19 or 20 amino acid residues. This diversity is maintained through viral production (**Appendix C**, **Figures C-7B, C-8B**), as the viral genomes produced from both the two-site and the three-site libraries display a similar quantity of mutations to the corresponding input DNA.

Libraries resulting from recombination of single sites display diversity across amino acid position and site (**Figure 4-4E, F**). Certain residue bias seems to be maintained from the codon library: serine remains prevalent between AA452-458 and AA492-498, while asparagine, serine, and threonine all maintain prevalence between AA585-591. Clear deviations from the input diversity also occur: aspartic acid becomes more prevalent across variable regions, while phenylalanine and leucine become much less prevalent between AA 452-458 and AA 492-498. This seems to indicate that both tolerance and input bias seem to influence prevalence through multiple rounds.

Additionally, individual residues seem to be acquiring bias after recombination. For example, production selects for glutamine in AA585 and proline or alanine in AA591 of the two-site library, where these residues are less prevalent in other positions. In the three-site library, asparagine is selected at AA496 and glutamine is selected at AA585. There are also residues that are attenuated across libraries, i.e. leucine is not tolerated at AA455, while proline is not tolerated at AA492.

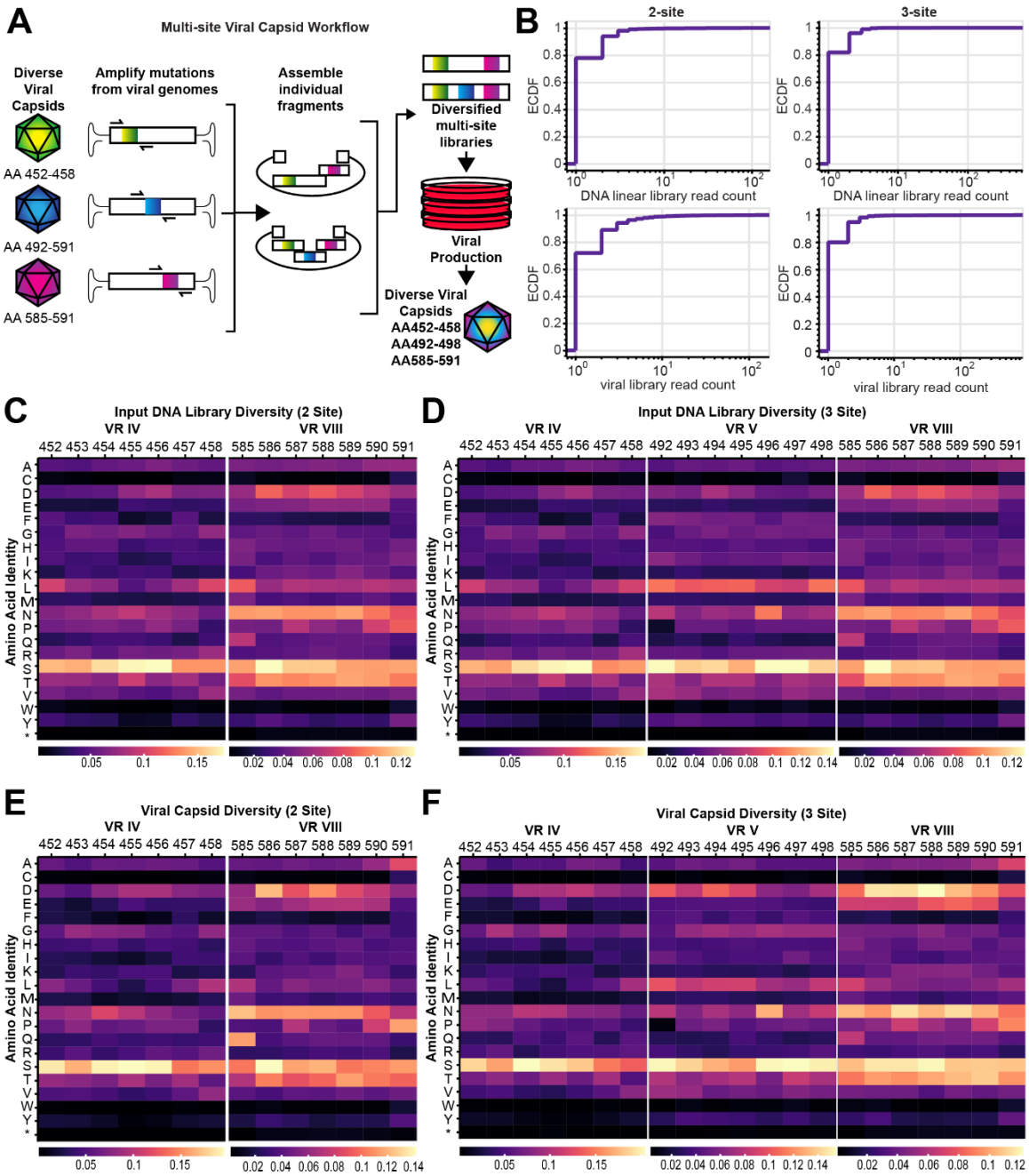

Figure 4-4. Combining diversity from multiple variable regions of AAV9 (A) Combination of substitution libraries strategy. (B) Read counts were sub-sampled for 10,000 sequences and the empirical cumulative density function compared for the input DNA and viral genomes of each multi-site substitution library (see **Appendix C**, **Figures C-7, C-8**). (C, D) Amino acid residue frequency at each position, within each substitution region, are displayed for the input DNA of the two-site (C) and three-site (D) substitution libraries. (E, F) Amino acid residue frequency at each position, within each substitution region, are displayed for the viral genomes produced in both the two-site (E) and three-site (F) substitution libraries.

4.5 Discussion

AAV capsid mutation and directed evolution through DNA-based selections has yielded transformative variants for neuroscience[1,2,4] and pre-clinical research for gene therapy[3]. In this work, we iterate upon these methods to repurpose a viral promoter element, T7, for DNA recovery from tissue. This alternative amplification method resulted in a similar quantity of variants after tissue recovery and similarly diverse profiles of variant enrichment. In these experiments, the irreproducibility of both DNA-based selections on 165,420 sequences was an unexpected result. These results indicate much smaller libraries should be included for Round 2 selections, although the exact depth should be determined through further experimentation. Based on the performance of known capsid variants, T7 amplification is a suitable candidate for DNA-based selections without requiring Cre-transgenic strains. Investigation of variants that were enriched across all brain samples for both methods and not enriched in any liver samples identified 38 capsid variants needing further validation.

In these experiments, we corroborate previous literature[25,26] increasing the transfection scale for increased library yield. While the exact extent of capsid switching is unclear, the data suggests that capsid switching occurs at similar rates across several orders of magnitude of transfection. Including higher quantities of stop codons in the input pool and obtaining higher read counts for these libraries could elucidate the exact extent of capsid switching. Insertions or substitutions at another site would be required for this experiment, as the barcoded amplicon containing AA452-458 was greater than 900bp and resulted in poor clustering and is prone to template switching during library preparation.

In this work, we engineered multiple variable regions in parallel and in different configurations to generate substitution libraries displaying diversity across 7, 14, and 21 amino acids respectively. These results confirm reports in the literature[4] that rolling circle amplification can be used to scale high quality DNA libraries prior to transfection. Diversity in these libraries acquired bias due to tolerance during production and from initial codon prevalence. Specific positions within this library also displayed bias, indicating interactivity between variable regions. Further examination of the sequence libraries for frequently

emerging combinations of amino acids could improve our understanding about interactions between variable regions and inform future viral library productions.

The significant extent of simultaneous capsid mutations in the two-site and three-site libraries expands the protein landscape substantially over conventional 7-mer insertions, while localizing the mutations in surface prominent regions. This improves upon previous attempts which implement machine-guided design and were limited to 19 AA mutations spread across the capsid[23]. We hope that these methods can be extended such that a broader protein landscape is accessible prior to *in vivo* variant selections, potentially unlocking previously inaccessible receptor binding, new tropism, cell type specificity, or immune escape.

Chapter IV Bibliography

1. Chan, K. Y. et al. Engineered AAVs for efficient noninvasive gene delivery to the central and peripheral nervous systems. Nat. Neurosci. 20, 1172–1179 (2017).
2. Deverman, B. E. et al. Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat. Biotechnol. 34, 204–209 (2016).
3. Tabebordbar, M. et al. Directed evolution of a family of AAV capsid variants enabling potent muscle-directed gene delivery across species. Cell 184, 4919-4938.e22 (2021).
4. Nonnenmacher, M. et al. Rapid evolution of blood-brain-barrier-penetrating AAV capsids by RNA-driven biopanning. Mol. Ther. - Methods Clin. Dev. 20, 366–378 (2021).
5. Hordeaux, J. et al. The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice. Molecular Therapy 26, 664–668 (2018).
6. Matsuzaki, Y. et al. Intravenous administration of the adeno-associated virus-PHP.B capsid fails to upregulate transduction efficiency in the marmoset brain. Neurosci. Lett. 665, 182–188 (2018).
7. Mingozzi, F. & High, K. A. Overcoming the Host Immune Response to Adeno-Associated Virus Gene Delivery Vectors: The Race Between Clearance, Tolerance, Neutralization, and Escape. Annu. Rev. Virol. 4, 511–534 (2017).
8. Vandamme, C., Adjali, O. & Mingozzi, F. Unraveling the Complex Story of Immune Responses to AAV Vectors Trial After Trial. Hum. Gene Ther. 28, 1061–1074 (2017).
9. Calcedo, R., Chichester, J. A. & Wilson, J. M. Assessment of Humoral, Innate, and T-Cell Immune Responses to Adeno-Associated Virus Vectors. Hum. Gene Ther. Methods 29, 86–95 (2018).
10. Gao, G. et al. Adeno-Associated Virus-Mediated Gene Transfer to Nonhuman Primate Liver Can Elicit Destructive Transgene-Specific T Cell Responses. Hum. Gene Ther. 20, 930–942 (2009).

11. Hinderer, C. et al. Severe Toxicity in Nonhuman Primates and Piglets Following High-Dose Intravenous Administration of an Adeno-Associated Virus Vector Expressing Human SMN. Hum. Gene Ther. 29, 285–298 (2018).
12. Kumar, S. et al. Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. Nat. Methods 17, 541–550 (2020).
13. Jang, S., Shen, H. K., Ding, X., Miles, T. F. & Gradinaru, V. Structural basis of receptor usage by the engineered capsid AAV-PHP.eB. Mol. Ther. Methods Clin. Dev. 26, 343–354 (2022).
14. Shay, T. F. et al. Primate-conserved Carbonic Anhydrase IV and murine-restricted Ly6c1 are new targets for crossing the blood-brain barrier. bioRxiv 2023.01.12.523632 (2023). doi:10.1101/2023.01.12.523632
15. Huang, Q. et al. Targeting AAV vectors to the CNS via de novo engineered capsid-receptor interactions. bioRxiv 2022.10.31.514553 (2022). doi:10.1101/2022.10.31.514553
16. Kern, A. et al. Identification of a Heparin-Binding Motif on Adeno-Associated Virus Type 2 Capsids. J. Virol. 77, 11072–11081 (2003).
17. Goertsen, D., Goeden, N., Flytzanis, N. C. & Gradinaru, V. Targeting the lung epithelium after intravenous delivery by directed evolution of underexplored sites on the AAV capsid. Mol. Ther. - Methods Clin. Dev. 26, 331–342 (2022).
18. Gurda, B. L. et al. Capsid Antibodies to Different Adeno-Associated Virus Serotypes Bind Common Regions. J. Virol. 87, 9111–9124 (2013).
19. Louis Jeune, V., Joergensen, J. A., Hajjar, R. J. & Weber, T. Pre-existing anti-adeno-associated virus antibodies as a challenge in AAV gene therapy. Hum. Gene Ther. Methods 24, 59–67 (2013).
20. Gonzalez, T. J. et al. Cross-species evolution of a highly potent AAV variant for therapeutic gene transfer and genome editing. Nat. Commun. 13, 5947 (2022).
21. Goertsen, D. et al. AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset. Nat. Neurosci. 25, 106–115 (2022).
22. Nov, Y. When second best is good enough: another probabilistic look at saturation mutagenesis. Appl. Environ. Microbiol. 78, 258–262 (2012).
23. Bryant, D. H. et al. Deep diversification of an AAV capsid protein by machine learning. Nat. Biotechnol. 39, 691–696 (2021).
24. Doi, N. & Yanagawa, H. Genotype-phenotype linkage for directed evolution and screening of combinatorial protein libraries. Comb. Chem. High Throughput Screen. 4, 497–509 (2001).
25. Schmit, P. F. et al. Cross-Packaging and Capsid Mosaic Formation in Multiplexed AAV Libraries. Mol. Ther. Methods Clin. Dev. 17, 107–121 (2020).
26. Nonnenmacher, M., Van Bakel, H., Hajjar, R. J. & Weber, T. High capsid-genome correlation facilitates creation of AAV libraries for directed evolution. Mol. Ther. 23, 675–682 (2015).
27. Sabin, J., Alatorre-Meda, M., Miñones, J., Domínguez-Arca, V. & Prieto, G. New insights on the mechanism of polyethylenimine transfection and their implications on gene therapy and DNA vaccines. Colloids Surfaces B Biointerfaces 210, 112219

(2022).
28. Nyberg, W. A. et al. An evolved AAV variant enables efficient genetic engineering of murine T cells. Cell 186, 446-460.e19 (2023).
29. Askary, A. et al. In situ readout of DNA barcodes and single base edits facilitated by in vitro transcription. Nat. Biotechnol. 38, 66–75 (2020).

Appendix C

SUPPLEMENTARY INFORMATION FOR CHAPTER IV

C.1 Figures

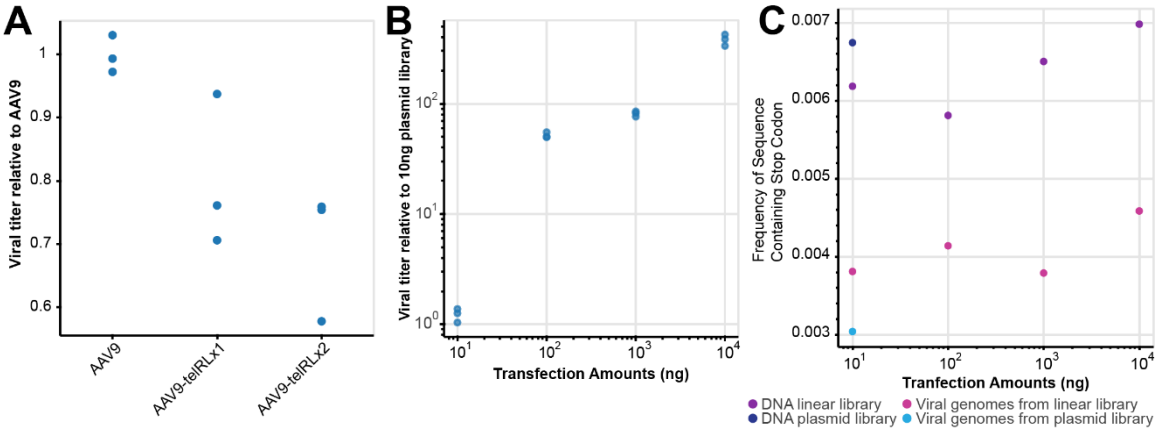

Figure C-1. Virus production of linearized library (A) Viral production yield was compared between undigested plasmid and plasmid linearized with TelN protelomerase. A measurable decrease was observed between linearized plasmids and circular control, and a single TelRL cut site had higher production than two TelRL cut sites. (B) Transfections of input across several orders of magnitude were performed, and viral product was purified and titered. Production was compared relative to standard 10 ng plasmid control. (C) Stop codon frequency was measured in viral reads across transfection scales. Variable rates of stop codon frequency were observed across scales.

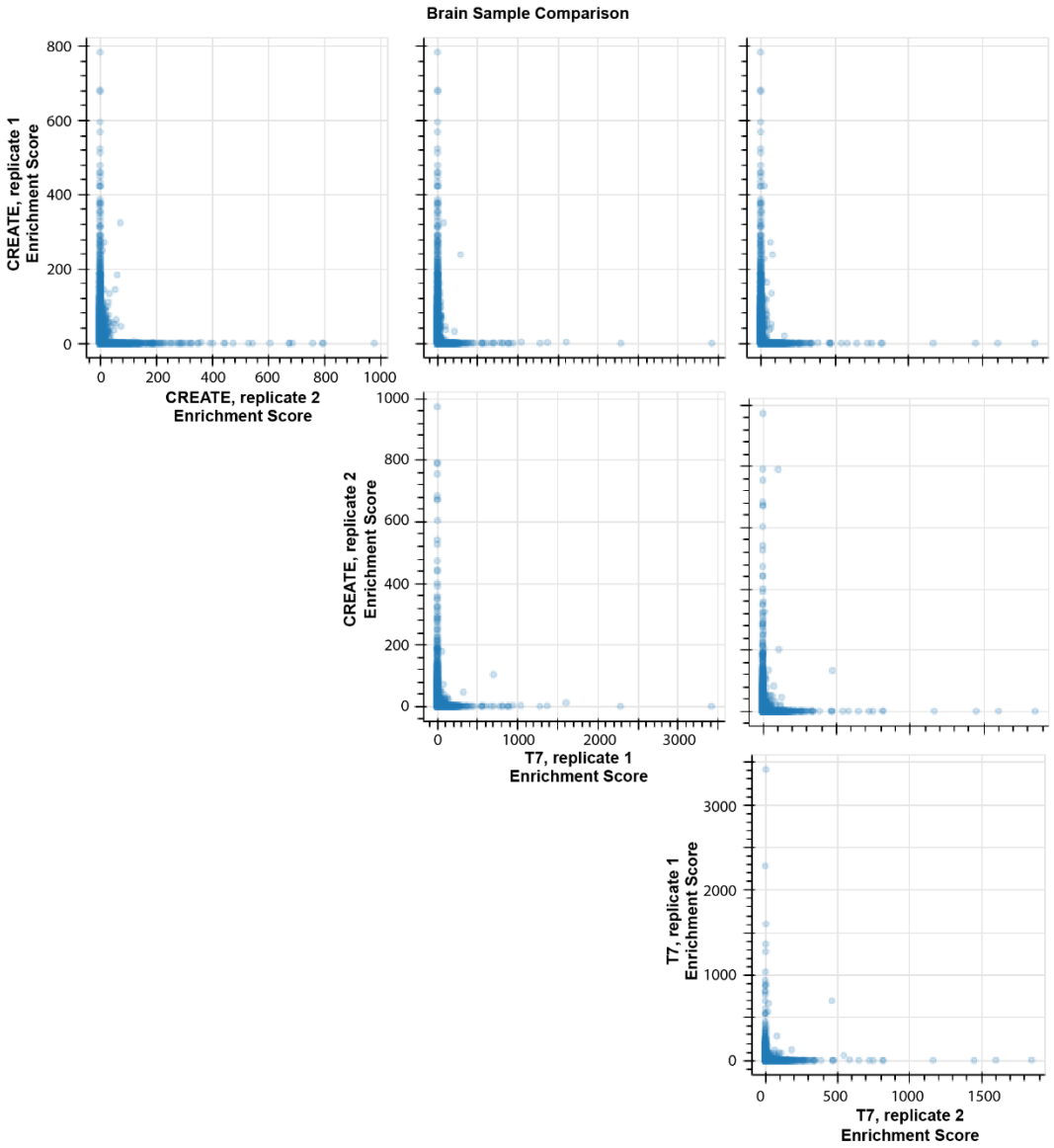

Figure C-2. Enrichment correlation between brain samples Enrichment is compared in the brain using two extraction methods, Cre-recombinase-dependent amplification and T7 amplification, and two animal replicates.

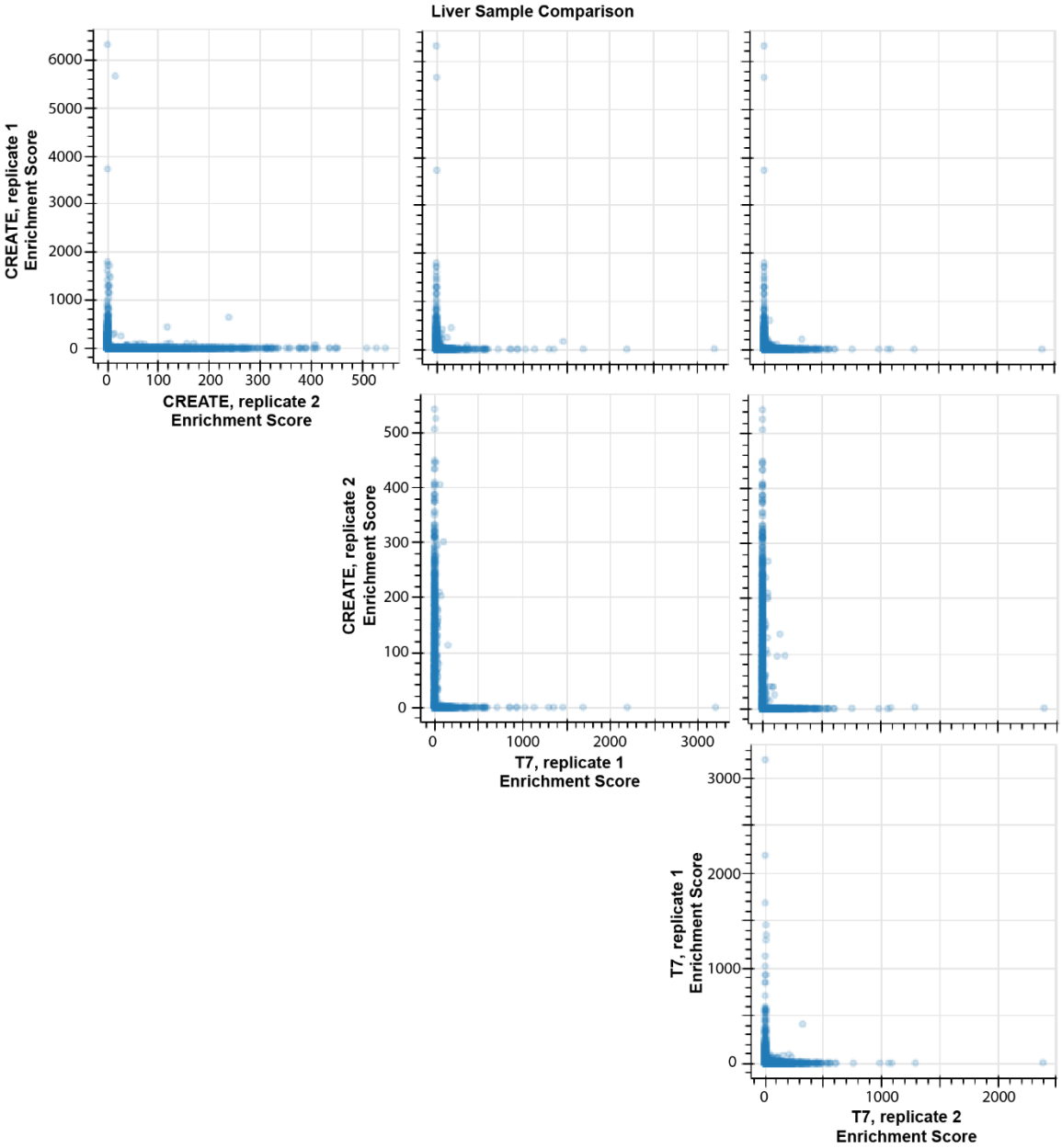

Figure C-3. Enrichment correlation between liver samples Enrichment is compared in the liver using two extraction methods, Cre-recombinase-dependent amplification and T7 amplification, and two animal replicates.

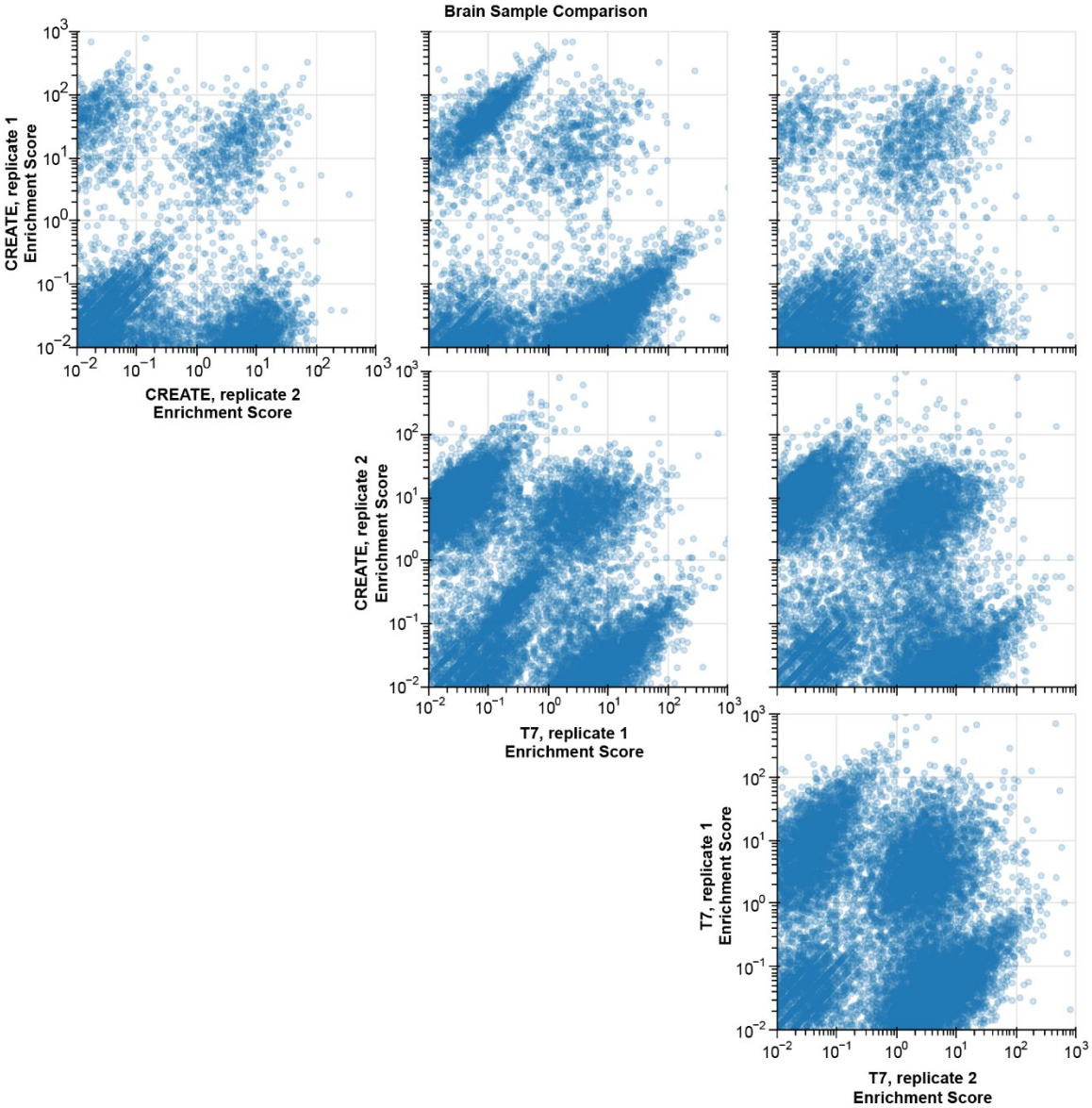

Figure C-4. Enrichment correlation between brain samples, excluding variants with zero measurements in either sample Variants with positive enrichment are compared in the brain using two extraction methods, Cre-recombinase-dependent amplification and T7 amplification, and two animal replicates.

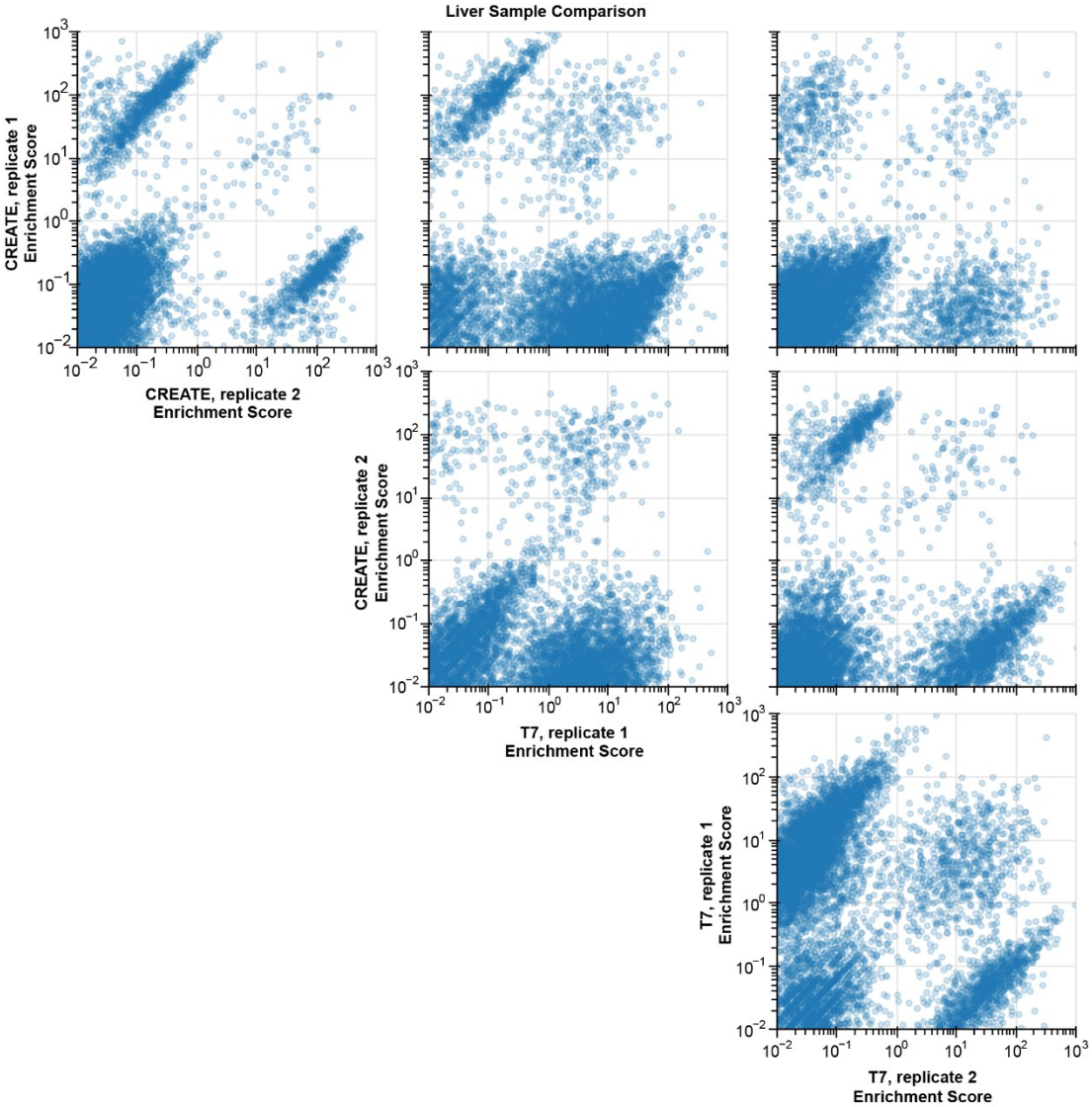

Figure C-5. Enrichment correlation between liver samples, excluding variants with zero measurements in either sample Variants with positive enrichment are compared in the liver using two extraction methods, Cre-recombinase-dependent amplification and T7 amplification, and two animal replicates.

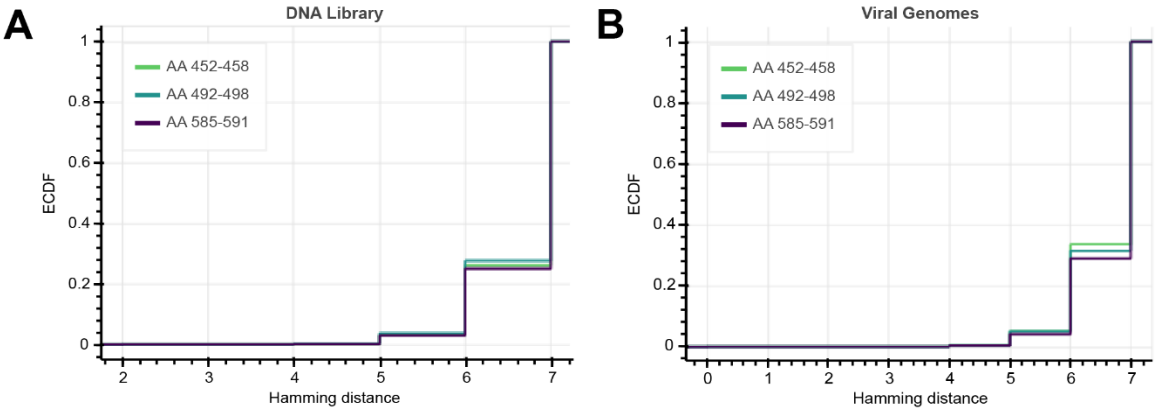

Figure C-6. Hamming distance of substitution libraries in multiple variable regions (A) The number of amino acid differences compared to wild type AAV9 (Hamming distance) in the input DNA libraries was calculated. The sequence measurements were sampled 500,000 times and the empirical cumulative density function was compared between substitution libraries. (B) The number of amino acid differences compared to wild type AAV9 (Hamming distance) in the viral genome libraries was calculated. The sequence measurements were sampled 500,000 times and the empirical cumulative density function was compared between substitution libraries.

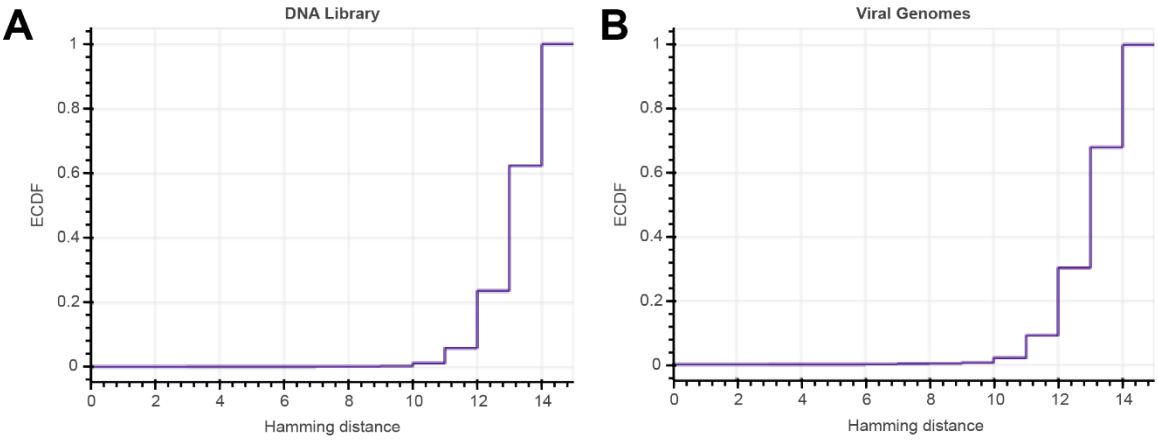

Figure C-7. Hamming distance of 2-site substitution libraries (A) The number of amino acid differences compared to wild type AAV9 (Hamming distance) in the input DNA libraries was calculated. The sequence measurements were sampled 500,000 times and the empirical cumulative density function is examined for the libraries containing diversity between AA 452-458 and AA 585-591. (B) The number of amino acid differences compared to wild type AAV9 (Hamming distance) in the viral genome libraries was calculated. The sequence measurements were sampled 500,000 times and the empirical cumulative density function is examined for the libraries containing diversity between AA 452-458 and AA 585-591.

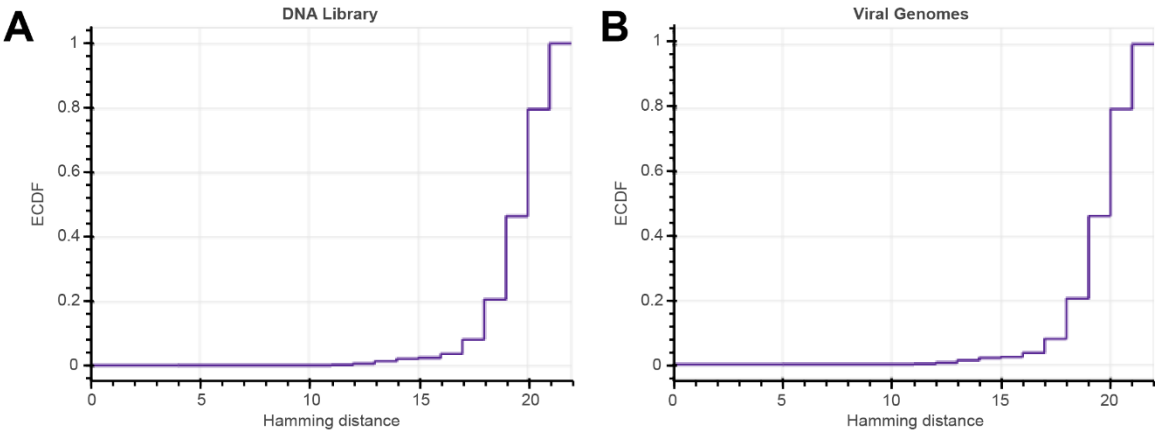

Figure C-8. Hamming distance of 3-site substitution libraries (A) The number of amino acid differences compared to wild type AAV9 (Hamming distance) in the input DNA libraries was calculated. The sequence measurements were sampled 500,000 times and the empirical cumulative density function is examined for the libraries containing diversity between AA 452-458, AA 492-498, and AA 585-591. (B) The number of amino acid differences compared to wild type AAV9 (Hamming distance) in the viral genome libraries was calculated. The sequence measurements were sampled 500,000 times and the empirical cumulative density function is examined for the libraries containing diversity between AA 452-458, AA 492-498, and AA 585-591.

C.2 Experimental procedures

C.2.1 Plasmids and cloning

To include T7 promoter in the library plasmid for Cre-recombinase-independent amplification of viral genomes from tissue, oligonucleotide pair 1 (**Table C-1**) was phosphorylated and annealed together to create the T7 promoter as an insert. 2 µL of each oligonucleotide was combined with 13 µL IDT duplex buffer, 2 µL T4 DNA Ligase Buffer, and 1 µL T4 Polynucleotide Kinase. The temperature was held at 30 °C for 30 minutes, 95 °C for 5 minutes, and subsequently decreased to 25C at 0.1 °C/s. In parallel, 2.5 µg of pAAV-Cap-in-cis-Lox, described previously[1,2], was simultaneously dephosphorylated and digested, in a reaction with 5 µL CutSmart (New England Biolabs, B6004), 2.5 µL Sal-HF (New England Biolabs, BR3138), 2.5 µL Quick CIP (New England Biolabs, M0525), and 37.5 µL nuclease-free water. The reaction was performed for 4 hours at 37 °C prior to purification (Zymo Research, D4007) from a 1% agarose gel. The digested vector was ligated together with T7 oligonucleotide insert according to manufacturer's T4 DNA Ligase protocol (New England Biolabs, M0202), transformed into *E.coli* (New England Biolabs, C3040), and colonies were picked and sequence confirmed to yield pAAV-Cap-in-cis-T7-Lox.

Table C-1. Complementary insert oligonucleotides

Oligonucleotide Pair	Forward primer (5' to 3')	Reverse primer (5' to 3')
1	TCGACCCTATAGTGAGTCGTATTA	TCGATAATACGACTCACTATAGG
2	CATGTTATCAGCACACAATTGCCCATTATACGCGCGTA TAATGGACTATTGTGTGCTGATAA	CATGTTATCAGCACACAATAGTCCATTATACGCGCGTATAAT GGGCAATTGTGTGCTGATAA
3	GCGCCTATCAGCACACAATTGCCCATTATACGCGCGTA TAATGGACTATTGTGTGCTGATAG	GCGCCTATCAGCACACAATAGTCCATTATACGCGCGTATAAT GGGCAATTGTGTGCTGATAG

Next, TelRL sites were cloned into the rAAV-Cap-in-cis-T7-Lox plasmid. First, oligonucleotide pairs 2 and 3 were annealed together to create the TelRL inserts: 5 µL of each primer was combined in 40 µL IDT duplex buffer. The temperature was held at 95 °C

for 5 minutes and subsequently decreased to 25 °C at 0.1 °C/s. In parallel, pAAV-CAG-eGFP-W3SL was digested with PciI, chosen due to the proximity to the AAV ITR, according to the manufacturer's protocol (New England Biolabs, R0655) and purified (Zymo Research, D4033). The DNA product was ligated together with the insert from oligonucleotide pair 2 using T4 DNA ligase according to the manufacturer's protocol (New England Biolabs, M0202), transformed into E. coli (New England Biolabs, C3040), and grown overnight at 30 °C. Bacterial colonies were picked, and sequence confirmed to obtain pAAV-CAG-eGFP-W3SL-TelRL×1. This construct was digested with KasI, chosen due to the proximity to the other AAV ITR, and purified (Zymo Research, D4007). The DNA product was ligated together with insert from oligonucleotide pair 3 and isolated, similar to oligonucleotide pair 2, to obtain pAAV-CAG-eGFP-W3SL-TelRL×2. Both pAAV-CAG-eGFP-W3SL-TelRL×1 and pAAV-CAG-eGFP-W3SL-TelRL×2 were amplified using primer pair 1 (**Table C-2**) in 25 cycles of 98 °C for 10 s, 55 °C for 15 s, and 72 °C for 120 s, using Q5 DNA polymerase. In parallel, rAAV-Cap-in-cis-Lox plasmid was amplified using primer pair 2, in 25 cycles of 98°C for 10 s, 55 °C for 15 s, and 72 °C for 180 s. Products from each PCR were run on an 0.8% agarose gel, their size confirmed, and purified (Zymo Research, D4007). Amplicon containing Cap-in-cis-Lox was ligated together with amplicons containing ITR with either TelRL×1 or TelRL×2 sites using T4 DNA ligase, according to manufacturer's protocol (New England Biolabs, M0202), generating pAAV-Cap-in-cis-T7-Lox-TelRL×1 and pAAV-Cap-in-cis- T7-Lox-TelRL×2.

Table C-2. Cloning primers

Primer Pair	Forward primer (5' to 3')	Reverse primer (5' to 3')
1	GGGAGGTAAGCTTAACAAGGTACCTTGGGCCGC	CTAGTCACGCGTCAATGTCCTCGAGGCGGCCG
2	GGTTCACGCGTGTGGCCTCCGCGCCG	TTCCTAAGCTTGATATCATAACTTCGTATAGCATAC

C.2.2 Substitution library DNA production

Plasmid rAAV-Cap-in-cis-Lox-TelRLx1 was digested with XbaI (New England Biolabs, R0145) and AgeI-HF (New England Biolabs, R3552) and purified from an agarose gel. In

parallel, viral DNA libraries at each variable region were generated by amplification of AAV9 capsid genome with NNK degenerate primers (**Table C-3**). To generate mutations between AA 452-458 (VP1 numbering), amplification was performed with primer pair 3. To generate mutations between 492-498 amplification was performed with both primer pairs 4 and 5. To generate mutations between 585-591, amplification was performed with primer pair 6. Each amplified region was cut and purified (Zymo Research, D4007) from a 2% low melting point agarose gel. Digested rAAV-Cap-in-cis-T7-Lox-TelRLx1 was combined with the variable inserts using HiFi DNA Assembly (New England Biolabs, E2621). For this reaction, 0.26 pmol of vector and 0.52 pmol of each insert was used (for substitution between AA 492-498, both inserts were included at this amount) in a 130 µL reaction at 50 °C for 15 minutes. Potential contaminating undigested parent DNA was removed from the DNA assembly reaction by incubation with Plasmid-Safe DNase using the manufacturer's protocol (Lucigen, E3101K) for 30 minutes at 37 °C. The resulting plasmid was purified using Zymo DNA Clean and Concentrator kit (D4033).

Table C-3. Single site substitution library generation primers

Primer Pair	AA Substitution	Fragment	Forward primer (5' to 3')	Reverse primer (5' to 3')
3	452-458	1	CTGGACCGACTAATGAATCCACTCATCGAC CAATACTTGTACTATCTCTCTAGAACTATTN NKNNKNNKNNKNNKNNKNNKCAAACGCTA AAATTCAGTG	GTCCTGCCAAACCATACCCGG
4	492-498	1	CCCAGCTACCGACAACAACGTGTCTCAACC NNKNNKNNKNNKNNKNNKNNKAGCGAATT TGCTTGGCCTGGAGCTTC	GTCCTGCCAAACCATACCCGG
5		2	CTGGACCGACTAATGAATCCACTCATCGAC CAATACTTGTACTATCTCTCTAGAAC	GGTTGAGACACGTTGTTGTCGGTAGCTGGG
6	585-591	1	CTGGACCGACTAATGAATCCACTCATCGAC CAATACTTGTACTATCTCTCTAGAAC	AAGTATTCCTTGGTTTTGAACCCAACCGGTCTGM NNMNNMNNMNNMNNMNNMNNGTGGTTTGTGG CCACTTGTCCATAGGACTC

The plasmid product was added as template to overnight RCA reactions, scaled to 800 µL, according to the manufacturer's protocol (Molecular Cloning Laboratories, PPK-100). Heat inactivation was not performed to avoid melting complementary library DNA. The input mass varied per substitution library: 584 ng for AA452-458, 304 ng for AA492-498, 484 ng for AA585-591. The product DNA was precipitated by ethanol precipitation, as follows. 10% v/v 3M sodium acetate (ThermoFisher Scientific, R1181) and 300% v/v 100% ethanol (pre-

chilled) were added to the reaction mixture. After incubation at -20 °C for 1 hour, samples were centrifuged at >16,000 g for 30 minutes at 4 °C. The supernatant was discarded and 500 µL 70% ethanol was added to each sample prior to centrifugation at >16,000g for 5 minutes at 4 °C. Washing with ethanol was repeated prior to drying the pellet at room temperature. The DNA pellet was then resuspended in nuclease-free water. 60 µg (~500 fmol of TelRL sites) of product DNA was digested in 500 uL TelN reactions, which were performed at 30 °C for one hour. The reaction product was purified using a Zymo DNA Clean and Concentrator kit (D4033) and 10 µL was run on an agarose gel to confirm digestion. This product is linear, closed-end DNA ready for transfection.

To amplify mutations from the substitution virus library productions, viral genomes isolated from individual substitution libraries were amplified using the primer pairs in **Table C-4**. Each amplified region was cut and purified (Zymo Research, D4007) from a 2% low melting point gel.

Table C-4. Multi-site substitution library generation primers

Primer Pair	AA Substitutions	Fragment	Forward primer (5' to 3')	Reverse primer (5' to 3')
7	452-458 & 585-591	1	CTTCCAGTTCAGCTACGAGTTTGAGAAC	CATGACTTTGTCCGCATCCACGTTGTCTC
8		2	GAGACAACGTGGATGCGGACAAAGTCATG	TGTTGAAGGCCGTTGGAGG
9	452-458 & 492-498 & 585-591	1	CTTCCAGTTCAGCTACGAGTTTGAGAAC	GGTTGAGACACGTTGTTGTCGGTAGCTGGG
10		2	CCCAGCTACCGACAACAACGTGTCTCAACC	CATGACTTTGTCCGCATCCACGTTGTCTC
11		3	GAGACAACGTGGATGCGGACAAAGTCATG	TGTTGAAGGCCGTTGGAGG

Digested rAAV-Cap-in-cis-Lox-TelRLx1 was combined with the variable inserts using HiFi DNA Assembly (New England Biolabs, E2621). For this reaction, 0.22 pmol of vector and 1.1 pmol of each insert was used in a 360 µL reaction at 50 °C for 15 minutes or 60 minutes, for two or three fragments, respectively. High quantities of transfection-ready DNA were prepared similar to single site substitution libraries.

C.2.3 Viral Production

Recombinant AAVs were generated according to established protocols[3]. Briefly, HEK293T cells (ATCC) were triple transfected using polyethylenimine (PEI) and virus was collected from both cell lysates and media and purified over iodixanol (Optiprep, Sigma). Virus was collected after 60 hours, instead of 120 hours, to limit secondary transduction of producer cells. For comparison of viral production, rAAV-Cap-in-cis-Lox library DNA was transfected at a total plasmid mass of 40 µg per 0.709 g of PEI max. Plasmid ratios are displayed in **Table C-5**.

Table C-5. Transfection quantities per plate for library production comparison

Library Format	Library DNA mass (µg)	Rep-ΔCap mass (µg)	pHelper mass (µg)	Total DNA mass (µg)	PEI max mass (µg)
Plasmid DNA	0.01	15.97	23.95	40	0.709
Linear DNA	0.01	15.97	23.95	40	0.709
Linear DNA	0.1	15.93	23.9	40	0.709
Linear DNA	1	15.57	23.36	40	0.709
Linear DNA	10	12	18	40	0.709

For virus production from substitution libraries, rAAV-Cap-in-cis-Lox library DNA was transfected according to **Table C-6**. Multi-site libraries were produced after recombination of single site libraries as described in **Appendix C.2.2**.

Table C-6. Transfection quantities per plate for single site substitution libraries

Substitution Library	Library DNA mass (µg)	Rep-ΔCap mass (µg)	pHelper mass (µg)	Total DNA mass (µg)	PEI max mass (µg)
AA 452-458	1	15.57	23.36	40	0.709
AA 492-498	1	15.57	23.36	40	0.709
AA 585-591	1	15.57	23.36	40	0.709
AA 452-458 & AA 585-591	0.1	15.93	23.9	40	0.709
AA 452-458 & AA 492-498 & AA 585-591	0.1	15.93	23.9	40	0.709

C.2.4 Animals

All rodent procedures were approved by the Institutional Animal Care and Use Committee (IACUC) of the California Institute of Technology. Transgenic animals, expressing Cre under the control of various cell-type-specific promoters, and C57BL/6J WT mice (000664) were purchased from the Jackson Laboratory (JAX). Transgenic mice included Tek-Cre (8863),[3] and Syn1-Cre (3966)[4]. For the CREATE and T7 selections of CAP-B library, we used one male and one female mouse from each transgenic line (aged 8-12 weeks), as well as a single male C57BL/6J mouse. Mice were housed under standard conditions between 71 and 75 °F, 30% - 70% humidity, and light cycle of 13 hours on & 11 hours off. Intravenous administration of rAAV vectors was performed via injection into the retro-orbital sinus.

C.2.5 DNA/RNA recovery and sequencing

CAP-B viral libraries[4] were injected into C57BL/6J and Cre-transgenic animals at a dose of 7×10^{10} vg/animal and tissues were recovered two weeks post injection. Cre-transgenic lines used were Syn1-Cre and Tek-Cre, selected for comparison to previously generated libraries. Mice were euthanized, brain and liver were recovered, snap frozen on dry ice, and placed into long-term storage at −80 °C. 100 mg of each tissue was homogenized in TRIzol (Life Technologies, 15596) using a BeadBug (Benchmark Scientific, D1036) and viral DNA was isolated according to the manufacturer's recommended protocol. Recovered viral DNA was treated with RNase, underwent restriction digestion with SmaI (located within the ITRs) to improve later rAAV genome recovery by PCR, and purified with a Zymo DNA Clean and Concentrator kit (D4033). Samples were then divided in two for processing via the CREATE or T7 methods.

One half of the sample was processed using the CREATE methodology. Viral genomes flipped by Cre-recombinase in select transgenic lines (or pre-flipped in WT animals) were selectively recovered using primer pair 12 in **Table C-7**. After 25 cycles of 98°C for 10 s, 60°C for 15 s, and 72°C for 40 s, using Q5 DNA polymerase in five 25 µl reactions with 50% of the total extracted viral DNA as a template. After Zymo DNA purification, samples from the WT C57BL/6J animals were serially diluted from 1:10 – 1:10,000 and each dilution

further amplified around the library variable region. The dilution series was analyzed for each WT tissue and the highest concentration dilution which resulted in no product was chosen for further amplification of the viral DNA from the transgenic animal tissues. This process was performed to differentiate between viral genomes flipped prior to packaging or due to Cre in the animal. Pre-flipped viral genomes should be avoided to minimize false positives in the NGS sequencing results.

The other half of the sample was processed through T7 transcription. Complementary RNA was generated through transcription using T7 RNA polymerase according to the manufacturer's protocol (MEGAscript Transcription Kit; Invitrogen) at 37 °C for 3 hours, using 8 µL of the sample as DNA template in a 20 µL reaction. Viral DNA was degraded in the product through TURBO DNAse incubation for 15 minutes at 37°C, prior to LiCl precipitation and resuspension in 20 µL nuclease-free water. Using 16 µL of the resulting product, cDNA was generated through reverse transcriptase (SuperScript™ IV VILO™ Master Mix) according to the manufacturer's protocol. Samples were diluted 1:10 and further amplified around the library variable region.

Amplification was done from the previous products using primer pair 13 from **Table C-7**, 10 cycles of 98 °C for 10 s, 61 °C for 15 s, and 72 °C for 20 s, to recover 73 bp of viral genome around and including the 21 bp variable region and add adapters for Illumina next-generation sequencing. After PCR cleanup, these products were further amplified using NEBNext Dual Index Primers for Illumina sequencing (New England Biolabs, E7600), with 10 cycles of 98 °C for 10 s, 60 °C for 15 s, and 72 °C for 20 s. The amplified products now containing unique indices for each method, tissue, animal, and replicate were run on a 2% low-melting-point agarose gel (ThermoFisher Scientific, 16520050) for better separation and recovery of the 210 bp band, which was extracted and purified (Zymo Research, D4007).

Table C-7. Next-generation sequencing primers

Primer Pair	Forward primer (5' to 3')	Reverse primer (5' to 3')
12	CTTCCAGTTCAGCTACGAGTTTGAGAAC	CAAGTAAAACCTCTACAAATGTGGTAAAATCG
13	ACGCTCTTCCGATCTAATACTTGTACTATCTCTCTAGAACTATT	TGTGCTCTTCCGATCTCACACTGAATTTTAGCGTTTG
14	CACGACGCTCTTCCGATCTCAACAACGTGTCTCAACC	TGTGCTCTTCCGATCTGGCCAAGCAAATTCGC

15	CACGACGCTCTTCCGATCTAGTCCTATGGACAAGTGGCCACA	TGTGCTCTTCCGATCTTCCTTGGTTTTGAACCCAACCG
16	ACGCTCTTCCGATCTAATACTTGTACTATCTCTCTAGAACTATT	TGTGCTCTTCCGATCTTCCTTGGTTTTGAACCCAACCG

Packaged viral library DNA was isolated from the injected viral library by digestion of the viral capsid and purification of the contained ssDNA. These viral genomes were amplified by two PCR amplification steps, like the viral DNA extracted from tissue, to add Illumina adapters and then indices and extracted and purified (Zymo Research, D4007).

Libraries variable at multiple sites were isolated for sequencing by amplifying viral genomes directly with the following primer pairs: primer pair 13 for AA 452-458, primer pair 14 for AA 492-498, primer pair 15 for AA 585-591, and primer pair 16 for multiple variable regions. Each of these PCRs was performed with 10 cycles of 98°C for 10 s, 61°C for 15 s, and 72°C for 30s. PCR cleanup was performed, and NEBNext Dual index primers were added for Illumina sequencing (New England Biolabs, E7600). This amplification (207 bp for AA 452-458, 194 bp for AA 492-498, 214 bp for AA 585-591, or 613 bp for multiple variable regions) was performed with 10 cycles of 98°C for 10 s, 60°C for 15 s, and 72°C for 30 s. The reactions were run on a 2% low-melting-point agarose gel (ThermoFisher Scientific, 16520050), and extracted and purified (Zymo Research, D4007).

All viral library DNA, including the samples extracted from tissue, were sequenced using Illumina MiSeq with MiSeq Reagent Kit V3 (Illumina, MS-102-3001) or sent for deep sequencing using an Illumina NextSeq2000 System (Millard and Muriel Jacobs Genetics and Genomics Laboratory, Caltech). The read depth for Cre-dependent and T7 amplified samples was similar, and the order of magnitude of capsid variants isolated from tissue was similar between samples (**Table C-8**).

Table C-8. Read depth for Cre-dependent and T7 amplification experiments

Amplification Method	Tissue	Replicate	Read Depth ($\times 10^6$ bp)	Unique Sequences (bp)
N/A	Input Pool	N/A	29.7	162651
CREATE	Brain	1	26.6	27439
CREATE	Brain	2	26.3	25063

CREATE	Liver	1	23.8	22170
CREATE	Liver	2	22.6	29504
T7	Brain	1	20.3	30973
T7	Brain	2	21.4	35101
T7	Liver	1	23.7	20437
T7	Liver	2	20.4	25987

C.2.5 NGS data alignment and processing

Raw fastq files from NGS runs were processed with custom data analysis code, available at https://doi.org/10.5281/zenodo.7909202, that isolate and quantifies the 21 bp diversified regions between AA 452-458, AA 492-298, or AA 585-591, depending on the libraries being examined. For comparison of Cre-dependent and T7 amplification, read counts for each sequence were pulled out and displayed along with their enrichment score, defined as the relative abundance of the sequence found within the specific tissue over the relative abundance of that sequence within the injected viral library. To identify sequences enriched in the brain and attenuated in the liver across methods, the sequences enriched in liver were subtracted from the intersection of sequences enriched in brain.

C.2.6 Data Availability

The code to analyze NGS datasets from FASTQ files is available at https://doi.org/10.5281/zenodo.7909202. Sequencing data can be obtained upon reasonable request.

Bibliography for Appendix C

1. Deverman, B. E., Pravdo, P. L., Simpson, B. P., Kumar, S. R., Chan, K. Y., Banerjee, A., Wu, W., Yang, B., Huber, N., Pasca, S. P., et al. Cre-dependent selection yields AAV variants for widespread gene transfer to the adult brain. Nat. Biotechnol. 34, 204–209 (2016).
2. Kumar, S. *et al.* Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. *Nat. Methods* **17**, 541–550 (2020).

3. Challis, R. C., Kumar, S. R., Chan, K. Y., Challis, C., Beadle, K., Jang, M. J., Kim, H. M., Rajendran, P. S., Tompkins, J. D., Shivkumar, K., et al. Systemic AAV vectors for widespread and targeted gene delivery in rodents. Nat. Protoc. 14, 379–414 (2019).
4. Goertsen, D. *et al.* AAV capsid variants with brain-wide transgene expression and decreased liver targeting after intravenous delivery in mouse and marmoset. *Nat. Neurosci.* **25**, 106–115 (2022).

CONCLUSION

As limitations of natural serotypes manifest during gene therapy research, engineering AAV capsid proteins for specific tropism becomes increasingly desirable. The tools and methods presented in this thesis address limitations of natural serotypes and demonstrate the modularity and flexibility of viral capsid engineering. Historically, capsid engineering has been limited to insertions within variable region VIII. In this thesis, capsid engineering is extended to substitutions across variable regions VR IV, VR V, and VR VIII. Extensive substitution mutations within these variable regions are tolerated in the AAV9 capsid, indicating the amenability of the capsid to mutation. This flexibility enables deep diversification of the AAV capsid.

These diverse substitution libraries can be selected to enable new cell-surface receptor interactions or specific tropism. In Chapter II, directed evolution selections of substitution mutations in VR IV yielded capsid variant AAV9.452sub.LUNG1, which targets the lung approximately 18-fold more than parent AAV9. This variant also transduces therapeutically relevant alveolar type II epithelial cells. These results address a lack of AAV variant development for the lung.

Substitution mutations in variable region IV are also tolerated with an accompanying insertion in VR VIII, revealing the modularity of the AAV capsid variable regions. The interactivity between variable regions enables an iterative engineering strategy. Previously engineered capsids can be further diversified at an alternative site and selected through directed evolution to shift tropism towards or away from cell populations. This was demonstrated in Chapter III, where substitution mutations were created within VR IV of a previously engineered capsid variant, AAV-PHP.eB. Implementing selections for maintenance in the brain and attenuation in the liver yielded capsids which cross the blood-brain barrier to target the brain while displaying liver attenuation after systemic injection. Two capsids from this selection translated to marmoset, an important model organism for neuroscience research. These results are meaningful for preclinical gene therapy

development, where liver toxicity or limited efficiency in non-human primates has repeatedly hindered translation to the clinic.

Through parallel engineering and recombination of substitution libraries within multiple variable regions, the surface protrusions of AAV can be broadly covered with substitution mutations. With the methods developed and described in Chapter IV, extensive diversification of the AAV surface protrusions is possible: capsid variants containing diversity across 7, 14, and 21 amino acid positions were produced and characterized. Mutation to this extent has not been previously achieved within the surface protrusion of the AAV capsid. An expanded protein landscape could enable new cell-receptor interactions or specific tropism.

The techniques developed in this thesis allow us to overcome limitations of previously engineered variants and natural serotypes. By diversification of the AAV capsid surface through substitution mutations, capsids can be developed with specific and targeted tropism profiles. Therapeutically relevant organs like the brain or the lung can be targeted through systemic injection while attenuating targeting to immunogenic organs like the liver. These advancements move us closer to the development of engineered AAV vectors targeting specific cell-populations to safely deliver gene therapy to humans.