Engineering vectors for non-invasive gene delivery to the central nervous system using Multiplexed-CREATE

Thesis by

Sripriya Ravindra Kumar

In Partial Fulfillment of the Requirements for

the degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2020

(Defended May 5, 2020)

Sripriya Ravindra Kumar
ORCID: 0000-0001-6033-7631

TO

My Dad, Mom, Sister, Brother-in-law, Niece, Friends.

And

My Beloved Nephew: Sri Praganesh

as I write this in your loving memory.

ACKNOWLEDGEMENTS

My thesis would have been impossible without my mentor, Dr. Viviana Gradinaru. I am extremely fortunate to have met her early on in my career. She has been an amazing mentor to me till date. Every day she inspires me to do more than what I think I am capable of, and it helped me a lot to grow under her guidance. She is the best mentor I could have ever asked for. She is like a "Superwoman" to me who I admire everyday on her skills to actively balance an outstanding career that involves a lot of hard work, every single day, while also maintaining all her other roles in her personal life as a wife, mother and daughter. Her wisdom and foresightedness have helped me a lot in shaping my professional as well as personal life. She has been extremely kind and supportive as a friend and as a mentor in the most difficult times in my personal life. I am very grateful to her for her support. She stands out as someone who deeply cares about the health and wellbeing of people around her, which I think is a rare virtue in today's fast-paced productivity-focused world. She created an extremely diverse scientific environment where over the years I had pursued projects that were in very different scientific disciplines, and that motivated me to stay in her lab to pursue my graduate study. The lab meetings would look more or less like a department seminar as everyone dwelled on very different areas of science. Over the years, I felt there was a lot of learning on a day-to-day basis, and I can't imagine any other lab where I could have gained this knowledge. This knowledge rich environment helped me grow to become independent as well as to actively collaborate with other scientists in lab. I admire Dr. Gradinaru for being extremely kind and generous. I can't think of a mentor who would just give away a planery talk at a prestigious conference, let alone to a graduate student. I am very grateful to her for believing in me, allowing me to take up different responsibilities over the years, and for giving me all the wonderful opportunities to grow intellectually, professionally and personally. I hope I can manage to live up to her expectations.

I'm extremely grateful to Dr. Benjamin Deverman, who introduced me to the science of vector engineering, and who closely mentored and trained me during his time at Caltech. This thesis would have been impossible without his mentorship. I got to know him a few months before I entered the graduate program and it was around that time when I decided what I absolutely wanted to do for my doctoral thesis. He taught me the importance of asking the right questions that can have a direct impact in the scientific community, and society at large. It was an absolute pleasure to work under his mentorship, and I look up to him for several reasons: his integrity, his patience to pursue challenges, and his foresightedness, and also for being a thorough, smart, and hardworking scientist. Above all, what I admire most about him is that he is such a warm and fun-loving person in the workspace and outside, and that it is an absolute pleasure to have him as a mentor and as a friend. Even outside lab, we shared memorable times exploring food in LA, his house parties, and an uphill biking adventure where he almost got me killed on what was supposedly a fun 9-mile-ride for a cause. Thank you for everything, Ben.

This work would have been impossible without my colleagues in Dr. Gradinaru's lab. First and foremost, I would like to thank Xinhong Chen, for being an amazing and cheerful

lab mate, and a friend who has been part of my every day activities in the lab. I admire him for his selflessness, energy, motivation, hard work and his positive attitude, and it is an absolute pleasure to work with him. He knew how to keep me on my toes. Outside lab, we also shared some great memorable times given our love for food and boba tea. I would like to thank Dr. Tim Miles for being a great lab mate and a friend as I actively collaborate with him on our projects, and in the process I learnt a lot. He has been extremely helpful in times of need, and it's been an absolute pleasure to brainstorm ideas with him, and I'm very thankful to all those intellectual conversations that I've had with him in the recent year which helped me to shape my thesis. I am very thankful to him for all the times in lab and outside that we shared during my graduate school. I would also like to thank Dr. Ken Chan for being a great friend and colleague during his time at Caltech. He kept the environment light and fun, and he trained me on several techniques, and I am very thankful for his time and effort. We also shared a lot of good memories outside lab as we explored the food and the city, and I feel very fortunate to have him as my friend till date. I am also very fortunate and thankful to have met Dr. Jennifer Treweek, the queen of prankster in lab, and we shared many great memories in lab and outside, as she continued to remind me of the importance to never grow old to have fun, and it definitely cheered me up to get through some of the dark, non-productive days. She is someone I look up to for her voracious appetite for knowledge and thoroughness; and I'm very thankful to her for teaching me the science and techniques in lab. I'm very thankful to Dr. Ryan Cho for being a great colleague and friend, and who also introduced me to Korean food and movies, and we had a wonderful time both in lab and outside. I'm very thankful to Dr. Alon Greenbaum who is absolutely wonderful to work with, and apart from being a great coworker, I'm fortunate to have his friendship and advice in shaping my professional and personal life. I am also very thankful to Dr. Claire Bedbrook, Dr. Collin Challis, Dr. Rose Challis, Dr. Min Jee Jang, Keith Beadle, Xiaozhe Ding, Dr. Qin Huang, Yaping Lei, Dr. Zhe Qu, Elisha Mackey, Dr. Anat Kahan, Michael Altermatt, Gerry Coughlin, Dr. Elliot Robinson, Damien Wolfe, Erin Sullivan, Yicheng Lei, and Dr. Nicholas Flytzanis who I had the opportunity to work and grow through the collaborations. They were wonderful co-workers, and above all were a great pal in the lab and outside. They brought a healthy work culture and fun environment that made my years at Caltech very pleasant and memorable.

I would also like to thank my committee members: Dr. Carlos Lois, Dr. Mitchell Guttman, and Dr. Rebecca Voorhees in addition to my advisor, Dr. Gradinaru for their time and advice on my thesis, and am deeply honored that they agreed to be on my committee.

I would also like to thank other faculty members at Caltech who I've had the opportunity to interact and be mentored. I'm extremely grateful to Dr. Dianne K. Newman who I have had the wonderful opportunity to be mentored on a collaborative project with Dr. Gradinaru lab. She is someone who inspired me with her intellect, humility, and kindness, and I was deeply honored to work under her mentorship. I'm extremely grateful to Dr. Mikhail G. Shapiro, who gave me the wonderful opportunity to rotate in his lab during my 1st year of graduate school, and it was an absolute pleasure to work under his mentorship.

This work wouldn't have been possible if not for being at an outstanding place like Caltech. I started my journey at Caltech in late 2012 as a Scientific Researcher in Dr. Viviana Gradinaru's lab and since then I've had the great privilege to engage with the brilliant scientists in this community who inspired me every day not just to become a better scientist but also remain humbled. Without any doubts, I can say that the valuable lessons learnt during my Caltech years are going to help me navigate through the rest of my scientific career and life in general. I am very grateful to the Graduate School office, International Student office, and Graduate Student Housing for being extremely kind and helpful in times of dire needs and am very thankful for all their generous support, time, and efforts till date. I'm also thankful to the Caltech research facilities, especially Millard and Muriel Jacobs Genetics and Genomics Laboratory, Dr. Igor Antoshechkin and Caltech Biological Imaging facility, and Dr. Andres Collazo and the Office of Laboratory Animal Resources (OLAR) for their time and service; their contributions were pivotal to my thesis.

I'm very grateful to Dr. Nagendranath Reddy for offering generous awards to female graduate students in Biology, and I'm deeply honored to be a recipient of his prestigious fellowship.

I would also like to take this opportunity to thank some of my previous mentors who had a significant role to play towards my scientific training. I'm very grateful to Dr. Baskar Gurunathan, my research advisor from my undergraduate years who gave me the first opportunity to do science. Those early years of training were very helpful to motivate me towards a higher education in biology. I'm very grateful and fortunate to have met Dr. James A. Imlay, my research advisor for my Master's thesis, early on in my life, who first trained me to become an independent researcher. I was extremely lucky to work under his mentorship, and the lessons learnt back then helped me in my Caltech years, and will continue to help me throughout my scientific career and life in general.

My thesis wouldn't have been possible without my family's support system. I'm very fortunate to have grown up in a family where I was encouraged to pursue my dreams. My dad, Ravindra Kumar Subramani, mom, Chitra Ravindra Kumar, and sister, Srividya Srinivasan, have supported me with my goals throughout, and without their constant motivation and their presence, I wouldn't be where I'm right now. And I hope that someday I can live up to their expectations.

I would like to use this opportunity to acknowledge the significant people in my life who were with me in my highs and lows in the past years. Dr. Anupama Lakshmanan, my friend and ex-roommate, was an absolute pleasure to be with and she made my years at Caltech memorable and cheerful as we enjoyed exploring the city, food and entertainment together. I'm very fortunate to have such a caring friend and as I write this she checks in to make sure I am on track with my thesis. Aarathi Shanmugam, my friend since my undergraduate years, has been with me on every step ever since and am very fortunate to have her in my life. Anand Chandrasekaran, my best friend since the time I moved to US, and he made sure I was de-stressed from time to time to keep me going with my thesis. Dr. Poorna Subramaniam, who I have had the pleasure to know since the last couple of years, has been

a great friend to me, and I look up to her for her timely advice. Sumana Sundaramurthy, my friend since undergraduate, is definitely one of those rare people in my life who shares a similar path and she has been a great support to me till date. Sanjana Pandian and Sujatha Venkatasubbu were the two friends from middle school who had a great role to play in my early years as we grew up together in a healthy companionship, and will always be grateful to have met them in my life. Pavithra Sellaperumal and Ramya Balaji, who I have known since my undergraduate years, have been wonderful friends till date, and I'm grateful for their friendship. Last but not the least, I'm grateful to Srikant Subramaniam who I've had the pleasure to know in the past months, and has been a great support to me till date.

ABSTRACT

Viruses are widely modified and used as gene delivery vectors for various applications in science and therapeutics. To this end, my thesis focuses on modifying the recombinant adeno-associated viral (rAAV) vectors that are identified as a safer choice for cargo delivery compared to other known viral vectors. They are widely used in the scientific communities, have seen promising outcomes in gene therapy clinical trials, and as of today have three products approved to use in humans. However, the natural repertoire of rAAVs have broad tropism when delivered systemically, and there is room for further improvement on the efficiency and specificity, especially for gene delivery in the central nervous system (CNS). The prior work done in Dr. Gradinaru lab addresses the issue by using a directed evolution approach called CREATE, Cre recombination-based AAV targeted evolution, to identify AAV-PHP.B and AAV-PHP.eB capsids, which broadly transduce the CNS (Deverman et al, 2016; Chan et al, 2017). CREATE selects for functional lox-flipped viral DNA that crosses the blood-brain barrier (BBB) and successfully transduces a specific nerve cell-type expressing Cre, thereby applying a strong selection pressure. However, the method is limited by its ability to identify a handful of enriched variants, and may also be prone to false positives resulting from experimental biases. The effort to fully understand the selection landscape, and to select for capsids that are not just efficient towards a cell-type but also specific towards it, led to the development of Multiplexed-CREATE (M-CREATE). M-CREATE allows parallel positive selections across different cell-types of interest, enables post-hoc negative selections across off-targets using a next-generation sequencing (NGS) based capsid recovery, and retains the principles of Cre-dependent functional recovery from CREATE. The method has a synthetic library generation approach to minimize biases within selection rounds, a variant replicate feature to identify the signal versus noise within a biological system, and an analysis pipeline to group families of enriched variants based on amino acid motifs, all of which together increases the confidence in the outcome and the throughput from a single experiment. Selections across brain endothelial cells, neurons, and astrocytes yielded several AAV-PHP.B-like variants that broadly transduce the CNS, AAV-PHP.V variants that can efficiently transduce the vascular cells forming the BBB, a AAV-

PHP.N variant that transduces neurons with greater specificity, and AAV-PHP.C variants that cross the BBB without murine strain specificity across tested strains. The AAV-PHP.C variants have different amino acid motifs compared to the AAV-PHP.Bs that have been previously shown to have limited CNS transduction across some mouse strains due to its interaction with the strain specific host cell surface receptor, ly6a, a homolog of which is not found in humans. (Hordeaux et al, 2018, Hordeaux et al, 2019; Huang et al, 2019; Batista et al, 2019) Therefore AAV-PHP.Cs offer some hope towards translation across other species. In summary, the M-CREATE methodology turns out to be a high-confidence, robust selection platform to yield several novel viral capsids for use in neuroscience and potential gene therapy related applications.

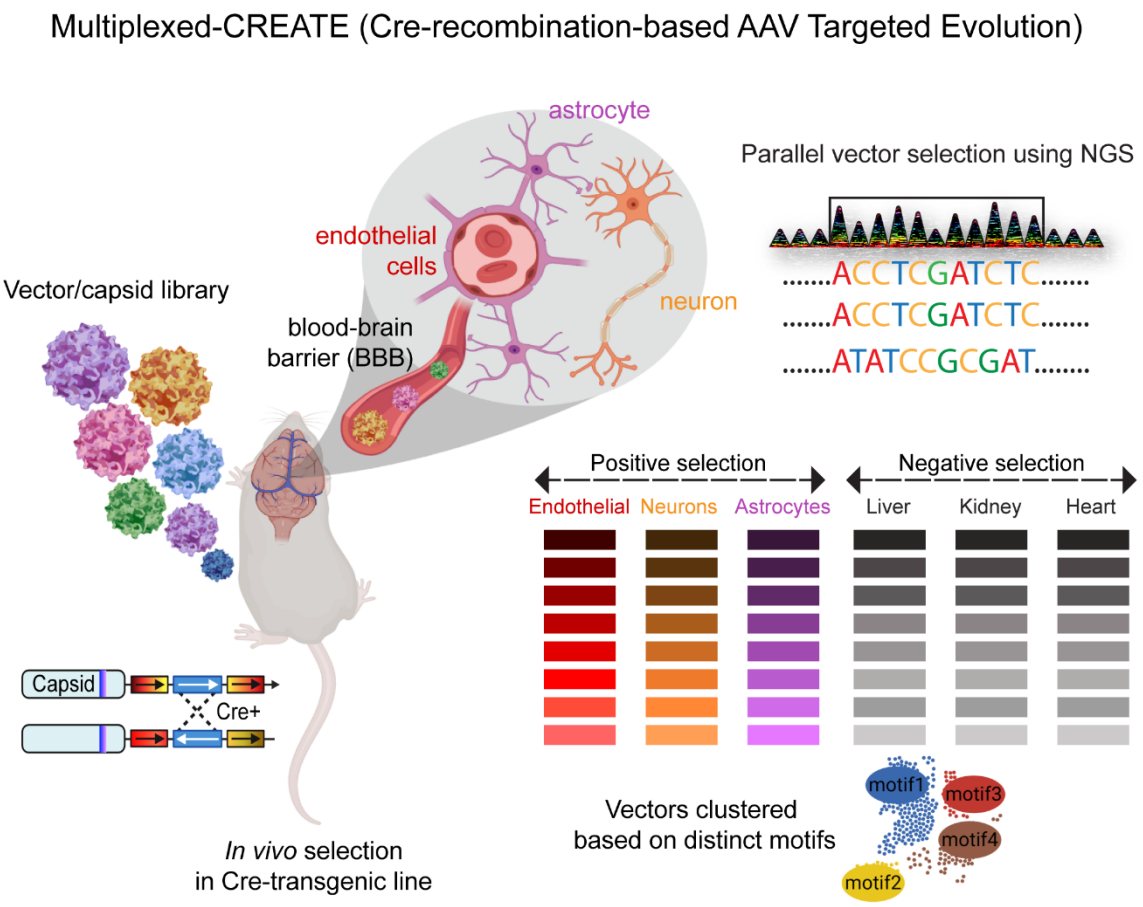


Figure 1.1: Concept of Multiplexed-CREATE

M-CREATE is a high-confidence, multiplexed, *in vivo* selection platform that yields vectors (i.e., viral capsids) with desired tropisms. M-CREATE identifies positively enriched vectors from on-target tissues or cell types that are also negatively enriched across off-targets through next-generation sequencing (NGS)-based vector recovery. The selected vector libraries are subjected to clustering based on the shared mutation patterns (or motifs) to identify distinct families of vectors, thereby generating multiple candidates to address gene delivery challenges.

PUBLISHED CONTENT AND CONTRIBUTIONS

- **Ravindra Kumar, S.,** Miles, T. F., Chen, X., Brown, D., Dobreva, T., Huang, Q., Ding, X., Luo, Y., Einarsson, P.H., Greenbaum, A., Jang, J.J., Deverman, B. E., Gradinaru, V. Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. *Nature Methods (2020).* https://doi.org/10.1038/s41592-020-0799-7
  S.R.K designed and performed the experiments, data analysis, and prepared the manuscript.

- Challis, R. C.*****, **Ravindra Kumar, S.***, Chan, K. Y., Challis, C., Beadle K., Jang, M. J., Kim H. M., Rajendran, P. S., Tompkins, J. D., Shivkumar, K., Deverman, B. E., Gradinaru, V. Systemic AAV vectors for widespread and targeted gene delivery in rodents. *Nature Protocol* 14**,** 379–414 (2019). https://doi.org/10.1038/s41596-018-0097-3
  **equal contribution. S.R.K designed and performed the experiments, data analysis, and prepared the manuscript.*

- Bourdeau, R. W., Lee-Gosselin, A., Lakshmanan, A., Farhadi, A., **Ravindra Kumar, S.**, Nety, S. P., & Shapiro, M. G. Acoustic reporter genes for noninvasive imaging of microorganisms in mammalian hosts. *Nature* 553, 86–90 (2018). https://doi.org/10.1038/nature25021
  S.R.K performed experiments related to characterization of Salmonella.

- Deverman, B.E., Pravdo, P.L., Simpson, B.P., **Ravindra Kumar, S.**, Chan, K.Y., Banerjee, A., Wu, W.L., Yang, B., Huber, N., Pasca, S.P., Gradinaru, V. Cre-dependent Capsid Selection Yields AAVs for Global Gene Transfer to the Adult Brain. *Nature Biotechnology* 34, 204–209 (2016). https://doi.org/10.1038/nbt.3440
  S.R.K. performed experiments and data analysis related to AAV variant characterization.

PUBLISHED CONFERENCE ABSTRACTS

- **Ravindra Kumar, S.#,** Miles, T. F., Chen, X., Brown, D., Dobreva, T., Huang, Q., Ding, X., Luo, Y., Einarsson, P.H., Greenbaum, A., Jang, J.J., Deverman, B. E., Gradinaru, V. Evolution and Investigation of Engineered AAV Capsids Exhibiting Enhanced Transduction of the Central Nervous System with or without Murine Strain Specificity. *Molecular Therapy, in press* (2020).
  #Presenting author, oral. Received "Meritorious Abstract Travel Award" by American Society of Gene and Cell Therapy (ASGCT). S.R.K designed and performed the experiments, and data analysis.

- Chen, X.[#], **Ravindra Kumar, S.,** Wolfe, D., Gradinaru, V. Targeting the Rodent Peripheral Nervous System Efficiently and with Greater Specificity through Intravenous Delivery of AAV Capsids Evolved by Multiplexed-CREATE. *Molecular Therapy, in press* (2020).
  #Presenting author, poster. S.R.K designed and performed the experiments, and data analysis.

- Padia, U.[#], Brown, D., Ding, X., Chen, X., **Ravindra Kumar, S.,** Gradinaru, V. Cloud-based Software for NGS Data Management and Analysis for Directed Evolution of Peptide-Based Delivery Vectors. *Molecular Therapy, in press* (2020).
  #Presenting author, poster. S.R.K contributed towards the study design.

- Brown, D.[#], Altermatt, M., Dobreva, T., Park, J.H., **Ravindra Kumar, S.**, Chen, X., Coughlin, G.M., Pool, A.H., Thomson, M., Gradinaru, V. A Computational and Experimental Platform for Detecting Full Transcriptome Cell Type Tropism of Lowly Expressed Barcoded and Pooled AAV Variants via Single-Cell RNA Sequencing. *Molecular Therapy, in press* (2020).
  #Presenting author, poster. S.R.K contributed towards the study design.

- **Ravindra Kumar, S[#].**, Chen, X., Deverman, B.E., Brown, D., Dobreva, T., Huang, Q., Ding, X., Luo, Y., Einarsson, P.H., Goeden,N., Flytzanis, N., Greenbaum, A., Gradinaru, V. Multiplexed-CREATE Selection Yields AAV Vectors Targeting Different Cell Types of the Central Nervous System Following Systemic Delivery. *Molecular Therapy* 27 (4), Art. No. 99 (2019). https://doi.org/10.1016/j.ymthe.2019.04.004
  #Presenting author, oral. S.R.K designed and performed the experiments, and data analysis.

- Ding, X[#], **Kumar, S.**, Malyutin, A., Gradinaru, V. Structural Studies of Engineered Adeno-Associated Virus Capsids that Cross Blood-Brain Barrier Efficiently. *Protein Science* 28 (S1), 116-117 (2019). https://doi.org/10.1002/pro.3710
  #Presenting author, poster. S.R.K prepared the viral vectors for experiments.

- **Ravindra Kumar, S[#].**, Chan, K., Jang, M.J., Huang, Q., Brown, B., Dobreva, T., Kim H.M., Luo, Y., Hurt, R.C., Chen, X., Deverman, B.E., Gradinaru, V. Developing AAV Vectors for More Efficient and Selective Gene Expression in Specific Cell Types of the Nervous System Following Systemic Delivery. *Molecular Therapy* 26 (5), 321 (2018). https://doi.org/10.1016/j.ymthe.2018.05.001
  #Presenting author, poster. Received "Outstanding Poster Presentation Award" by American Society of Gene and Cell Therapy (ASGCT). S.R.K designed and performed the experiments, and data analysis.

- Deverman, B.E[#]., Pravdo, P.L., Simpson, B.P., **Ravindra Kumar, S.**, Luo, Y., Chan, K.Y., Banerjee, A., Wu, W.L., Yang, B., Huber, N., Pasca, S.P., Gradinaru, V. Using

Cre-Dependent In Vivo Selection to Identify AAV Variants That Enable Efficient and Widespread Gene Transfer to the Adult Central Nervous System. *Molecular Therapy* 24, S99 (2016). https://doi.org/10.1016/S1525-0016(16)33061-1
#Presenting author, poster. S.R.K. performed experiments and data analysis related to AAV variant characterization.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS AND TABLES

Chapter 1

BACKGROUND AND MOTIVATION

Section 1.5 of this chapter has been adapted from:

Deverman, B.E., Pravdo, P.L., Simpson, B.P., **Ravindra Kumar, S.**, Chan, K.Y., Banerjee, A., Wu, W.L., Yang, B., Huber, N., Pasca, S.P., Gradinaru, V. Cre-dependent Capsid Selection Yields AAVs for Global Gene Transfer to the Adult Brain. *Nature Biotechnology* 34, 204–209 (2016). https://doi.org/10.1038/nbt.3440

1.1. VECTORS FOR GENE DELIVERY

One of the fundamental needs in biology are tools to study cells or associated pathologies, and eventually design therapies to treat the impaired cells. At a cellular level, all of the above involve an exogenous supply of genes encoding proteins of interest. Molecular tools than can package a desired gene and deliver it to a cell are called vectors or vehicles. Vectors can turn out to be the most powerful tool for a biologist if we can design them to precisely target a cell-type of interest efficiently, and can become a vital resource across various disciplines, such as in basic and biomedical sciences and in clinics.

The gene delivery vectors can be broadly classified based on the source as non-viral and viral vectors.

Non-viral vectors

The non-viral delivery systems comprise chemical and biochemical vectors such as lipid-based vectors, peptide-based vectors, natural and synthetic polymers, calcium phosphates, and metal nanoparticles[1–11]. These non-viral vectors are attractive for their biosafety aspect given their synthetic nature, and thereby their inability to elicit strong

immune response in the host. However, these non-viral vectors are limited in their applications due to lower transfection efficiency, and have been designed only for fewer target cell population given the time, cost, and labor involved in engineering such vectors[12–15].

Viral vectors

Viruses that are known to infect mammalian cells are naturally existing vectors, and they are subjected to engineering by making them replication deficient, stripping them of their toxic genome, and replacing it with our gene of interest to serve as a safe gene delivery vector (alias recombinant viral vector). Some viruses have been recognized and engineered for this purpose including; retroviruses, lentivirus, adenovirus (AD), adeno-associated virus (AAV), herpes simplex virus (HSV), rabies, and baculovirus. However, depending on the viral vectors, they have their own pros and cons including tropisms, transduction efficiency and specificity, stability of transgene expression, cytotoxicity, and immune response, all of which influence their potential applications in science and therapies[16–19]. Some notable vectors are highlighted below.

Retroviruses

Retroviruses are RNA viruses, and were the first to be recognized and engineered as a viral vector in the 1980s[20], and they eventually entered the first gene therapy clinical trial in the 1990s for adenosine deaminase deficiency (ADA)[21]. However, the use of retroviral vectors observed a major setback due to its innate ability to randomly integrate into host genome, which may at times lead to insertional oncogenesis[22–25]. Hence, the issues associated with the use of these vectors limit their applications[26–30].

Lentivirus

Lentivirus is a RNA retrovirus, and the recombinant viral vectors[31–34] stands out for its ability to transduce both dividing and non-dividing cells[35–37]. Given its broad tropism, these recombinant viral vectors have gained a lot of attention in both basic science applications[38–

[42] and in gene therapy[43–45]. However, even with the continuous engineering efforts to reduce the random integration events into host genome, the risk is not eliminated[46].

Adenovirus (AD)

Adenoviruses are non-enveloped, icosahedral, double-stranded DNA viruses that are widely used as gene delivery vectors[47–49], and also as recombinant viral vaccines[50–52]. ADs are attractive for their ability to exist as episomal DNA (extra-chromosomal DNA that does not integrate with host DNA), to package large genomes[53] and modified tropisms through engineering[54,55]. However, one of the major drawbacks for gene therapy application is their ability to induce strong host innate immune responses[56,57], such as the one seen in the 18-year-old Jesse Gelsinger case, where a fatal inflammatory response in days after administering AD packaged with ornithine-transcarbamylase led to his death[58,59].

1.2. ADENO-ASSOCIATED VIRAL VECTORS FOR GENE DELIVERY

Adeno-associated virus (AAV)

AAVs are 25 nm, non-enveloped, icosahedral (formed from 60 monomers), DNA parvoviruses with a single-stranded 4.7 kb genome[60–65]. AAVs are replication incompetent, and depend on a second virus for infection such as AD or HSV[66]. The genome carries two genes: *rep* and *cap*, comprising the Rep proteins (Rep78, Rep68, Rep52, Rep40) that coordinate AAV replication[67], a capsid assembly-activating protein (AAP)[68] and a membrane associated accessory protein (MAAP)[69] from alternate open reading frames, and capsid proteins (VP1, VP2, VP3) that form the capsid coat[70]. These genes are flanked by inverted terminal repeats (ITRs) which are packaging signals for the genome and are involved in 2nd strand DNA synthesis in the host cell, and also has a terminal resolution site to enable site specific integration into human chromosome 19 (AAVS1)[71–74].

Advantages of using rAAVs as vectors for gene delivery

What makes rAAVs appealing for *in vivo* gene delivery applications is their latent (or harmless) infection without the helper virus[66,75,76], and most of the viral genome can be stripped off to insert the cargo of interest as the only viral sequences required within the rAAV genome are the ITRs (145 bases on either end of the genome)[77–79]. The vectors are non-pathogenic, and have been shown to exhibit low immunogenic response. Their long-term persistence as episomal DNA in transduced cells lowers the risk of random chromosomal integration events that are otherwise observed with retroviruses. All of the above features make them a safer choice of gene delivery vector over other viruses. Above all their ability to transduce both dividing and non-dividing cells makes them widely usable for various applications[61,78,80–85].

Current AAV vector applications

Vectors can find their applications across various disciplines in science.

Basic Science

There has been more acceptance in using the rAAV viral vectors in laboratories all over the world to address a myriad of science questions that requires delivery of genes to a specific target of interest. The scientific needs involve delivering a specific or multiple genes under-investigation which includes an endogenous gene, an actuator, a sensor, or a reporter[86]. This could be easily achieved with rAAV viral vectors[87]. rAAVs have been used in applications like cell circuits tracing[88], probing cellular activity using techniques like optogenetics or chemogenetics[89], building disease models[90], etc. The current alternative experimental setup using transgenic animals carrying a modified gene has limitations such as the cost and time involved in generation and maintenance of transgenic lines. The growing needs to cover the diverse cell populations or pathological conditions under investigation limits its applications[91–93]. Compared to transgenics, the gene delivery vectors are easier and faster to implement; they could be used to deliver broadly across various targets under investigation, higher tissue or cell-type specificity based on the mode of vector delivery, the

nature of the cargo, or by the use of engineered vectors; and they allow time-controllable gene delivery in a host, which is crucial for studying pathologies[87,93].

Biomedical Science

The use of AAV viral vectors for drug discovery *in vivo* can speed up progress in these areas of research, thereby increasing the potential to identify new therapeutic targets, and therefore newer therapies[94–98].

Gene therapy

Gene therapy is a methodology that uses identified therapeutic genes to be delivered to the diseased cells to repair the loss of function. rAAVs, in addition to their predominant use in science, are promising vectors in gene therapies too for treating monogenetic disorders[99–101]. The first long term persistence of gene expression in mammalian brain by rAAV based gene delivery led to a great start in 1990s[102], which then quickly followed with the first human gene therapy trial for cystic fibrosis (CF)[103]. In 2008, rAAV was shown to have promising outcome to treat leber congenital amarosus (LCA)[104–106].

Three rAAV based gene therapies have been approved till date to use in clinics. In 2012, the first gene replacement therapy using rAAV based drug named Glybera, was approved by the European Regulatory Commission (ERC) to treat lipoprotein lipase deficiency (LPLD)[107]. In 2017, Food and Drug Administration (FDA) approved the first rAAV based gene therapy in US, named Luxturna (voretigene neparvovec-rzyl) to correct RPE65 mutation-associated inherited retinal dystrophy[108–110]. In 2019, FDA approved Zolgensma (onasemnogene abeparvovec) to treat spinal muscular atrophy (SMA)[111,112]. In addition to these approved therapies, currently ~14% of the gene therapies that have entered the clinical trials uses rAAVs, which is at 2nd position following ~27% using vaccinia virus. Other trials comprise ~11% using ADs or retroviruses, ~6% using LVs, ~4% using naked or plasmid DNA, and ~1% using HSV.

Current limitations of AAVs

While the properties of AAVs certainly account for their use in research labs and clinics, their use needs a significant improvement on the transduction efficiency and target specificity[113]. There are currently 12 known serotypes of AAVs with sequence similarity between 50 - 99% that are known to exhibit different but overlapping tropisms[80]. These tropisms are often attributed to their uptake by specific cell surface receptors. For instance, AAV2 binds to heparin sulfate receptor[114] and AAV9 binds to galactose receptor[115]. In addition to these major receptors, several have identified co-receptors which contribute to their cellular uptake[116,117]. While these receptors contribute to some differences in the tropism among serotypes, these receptors are broadly expressed across different tissues. In other words, the AAVs are naturally evolved to have broad tropisms, and this needs to be addressed if we want to use them as safe vehicles for gene delivery to a specific target. In addition to the serotype specific receptors and co-receptors, the AAVs are also known to have a universal receptor AAVR, and this could be crucial for host cell infectivity[118–123].

Given the presence of multiple receptors and incomplete understanding of the biology, alternate approaches have been taken to tackle the issue on off-target transduction such as by performing localized injection to deliver the cargo[124], or cargo engineering to restrict expression to cell populations of interest, or by coupling injection of AAVs carrying transgene that is controlled by a recombinase enzyme such as Cre into Cre-transgenic animals to get specificity in transgene expression[93]. Having said that, these approaches aren't always feasible and are often very invasive. Alternative strategies to achieve highly targeted gene delivery and protein expression are desired in areas that are anatomically hard to access for AAV injection, in cell-types with unknown specific regulatory elements, and in cell-types with lack of availability of cell-type specific transgenic lines.

1.3 ENGINEERING AAV CARGOS FOR EFFICIENT AND TARGETED PROTEIN EXPRESSION

AAVs with improved transduction efficiencies and improved cell-type specificity, i.e., with minimal off-target transduction via systemic injection would open up many potential

applications in science to begin with and may eventually find their purpose in gene therapies too.

These are currently addressed by various modifications done to the cargo or the rAAV genome[125–127]. To improve transduction efficiency, strong ubiquitous promoters or enhancers have been engineered to fit into the rAAV genome size limit[95,128,129]. ITRs have also been mutated to engineer to a self-complementary AAV (scAAV) version which can deliver the genome as a double stranded DNA (dsDNA) instead of a ssDNA but at half the genome size (2.35 kb)[130]. This scAAV version alleviates the rate limiting step of the second strand synthesis in the host cell, thereby shown to improve the transduction efficiency.

To restrict expression of delivered cargo to the cell-types of interest, regulatory elements are engineered to fit the needs. Cell-type specific promoters or enhancers can regulate expression in a defined cell population[129,131,132]. Inducible promoters (such as tetracycline inducible promoter) can control the dose or time of expression[133]. Other regulatory elements such as microRNA targeting signal are identified across cell-types of interest and can selectively suppress expression in a cell type where the recognized miRNA can act on it[134–137].

While the cargo engineering is a welcome addition to improve the efficiency and specificity of the vector delivered gene expression, these strategies do not completely alleviate the problem when a non-invasive systemic delivery is preferred where the vectors still deliver genes across different organs, thereby turning on the immune response in the host cell[138]. Also, these regulatory elements can have varying degrees of specificity or efficiency across cell-types. Hence it is only seen as an added advantage to the existing vectors, and not a definitive solution to the problem of broad tropism.

1.4 ENGINEERING AAV CAPSIDS TO MODIFY THE VECTOR TROPISM

Engineering AAV capsids to overcome the current limitations has been of great interest to the field. For instance, neuron-specific AAVs will be useful for neural circuit

mapping studies[139], and the organ specific ones, say heart, lung and kidney, will find broad applications to study pathologies and to develop disease models[140–142]. Broadly, the availability of these specific variants can facilitate non-invasive delivery of vectors, which will simplify the whole procedure involved in gene delivery both in science and in clinics.

Several groups have attempted to address this problem using a rational or semi-rational approach[143–148] or directed evolution approach[113,149–153]. Progress with the rational design has been slow due to our limited understanding of the capsid structure-function relationship. In light of this, directed evolution seems to have more promise for vector engineering because it requires little to no knowledge of a receptor target or mechanism of entry[93,154].

1.5 IN VIVO SELECTION OF AAV CAPSIDS USING CREATE

Recombinant AAVs are the preferred vehicles for many *in vivo* gene transfer applications across cell populations; however, applications involving gene transfer to the central nervous system (CNS) are limited due to the lower transduction efficiency and specificity from natural or engineered vectors via systemic delivery across species, and often requires invasive routes of delivery or very high doses of vector[93,143,155,156]. This is because for intravenous delivery, the highly selective blood-brain barrier (BBB) poses a serious challenge, and the cellular heterogeneity of the CNS presents further challenges for gene transfer applications.

Cre recombination-based AAV targeted evolution method (CREATE)

In 2016, Deverman, B., *et al*[157] developed an *in vivo* selection method to provide selective pressure for capsids that cross the BBB and functionally transduce CNS cell types. This method, called CREATE (Cre-recombination based AAV targeted evolution) is a Cre recombination–dependent approach to selectively recover capsids that transduce predefined Cre-expressing target cell populations. CREATE uses an rAAV capsid genome (rAAV-Cap-in-cis-lox) that couples a full-length AAV *cap* gene, controlled by regulatory elements from the AAV *rep* gene with a Cre-invertible switch (***Figure 1.2***). By building capsid libraries

within the rAAV-Cap-in-cis-lox backbone and delivering the virus libraries to animals with Cre expression in a defined cell population, the system enables the selective amplification and recovery of sequences that have transduced the target population (***Figure 1.2***)[157].

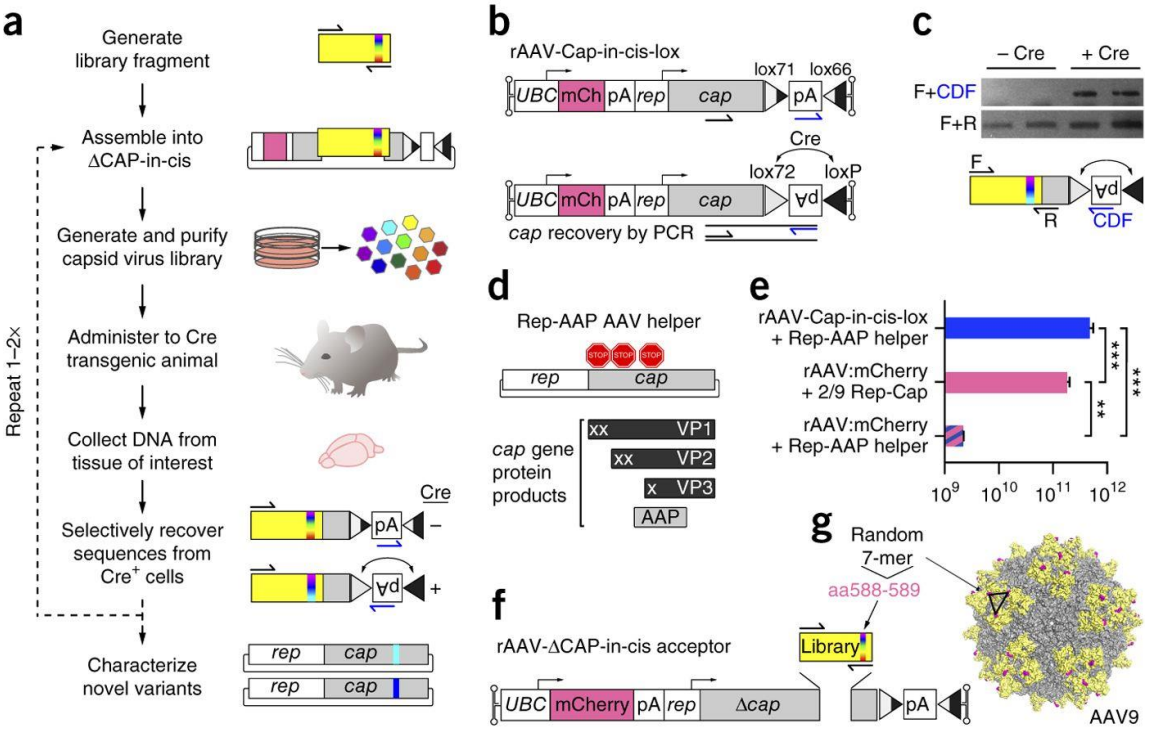


Figure 1.2: Cre-dependent recovery of AAV capsid sequences from transduced target cells.

(a) An overview of the CREATE selection process. PCR is used to introduce diversity (full visual spectrum vertical band) into a capsid gene fragment (yellow). The fragment is cloned into the rAAV genome harboring the remaining capsid gene (gray) and is used to generate a library of virus variants. The library is injected into Cre transgenic animals, and PCR is used to selectively recover capsid sequences from Cre^+ cells. **(b)** The rAAV-Cap-in-cis-lox rAAV genome. Cre inverts the polyadenylation (pA) sequence flanked by the lox71 and lox66 sites. PCR primers (half arrows) are used to selectively amplify Cre-recombined sequences. **(c)** PCR products from Cre recombination–dependent (top) and –independent (bottom) amplification of capsid library sequences recovered from two Cre^+ or Cre^- mice are shown. Schematics (bottom) show the PCR amplification strategies. **(d)** Schematic shows the AAV genes within the Rep-AAP AAV helper plasmid and the proteins encoded by the *cap* gene. Stop codons inserted in the *cap* gene eliminate VP1, VP2 and VP3 capsid protein expression. **(e)** DNase-resistant AAV vector genomes (vg) produced with the split AAV2/9 Rep-AAP and rAAV-Cap-in-cis-lox genome (top) as compared to the vg produced with standard AAV2/9 Rep-Cap helper and rAAV-UBC-mCherry genome (middle) or with the AAV2/9 Rep-AAP and rAAV-UBC-mCherry genome (bottom). $n = 3$ independent trials per group;

mean ± s.d.; $^{**}P < 0.01$, $^{***}P < 0.001$; one-way ANOVA and Tukey's multiple-comparison test. **(f)** Cloning the 7-mer capsid library into the rAAV-ΔCap-in-cis vector. **(g)** The AAV9 surface model shows the location of the 7-mer inserted between amino acids (aa) 588 and 589 (magenta). Sites encoded with the PCR-generated library fragment (aa 450–592) are shown in yellow[157].

Using CREATE, the natural serotype AAV9 that is known to cross the BBB less efficiently was chosen to evolve by targeted insertion of a randomized 7-mer (7-amino acid long peptide) within a surface-exposed site of the capsid (between AA588-589) and then intravenously injected this large library (a theoretical library size of ~1.28 billion variants) in a Cre-transgenic mouse where Cre expression was restricted to the CNS cell-type of interest. GFAP-Cre, an astrocyte specific Cre transgenic line was chosen given the high prevalence of this cell population in the CNS. Two weeks post *in vivo* selection, the mice were sacrificed to collect the brain and spinal cord. The viral DNA was extracted from the tissue, and selectively amplified the viral DNA from the Cre-positive cells using primers that can selectively yield a PCR product of lox-flipped viral DNA. This was then pooled together to make a second viral library for the second round of selection[157].

Post two rounds of selection, the recovered variants were cloned back in *Escherichia coli,* and the highly enriched variants were identified by sanger sequencing. The outcome resulted in a handful of enriched variants, each of which had a unique 7-mer motif. A rAAV capsid variant called AAV-PHP.B expressed the transgene broadly in CNS and the transduction efficiency was at least 40-fold higher compared to the parent (***Figures 1.3***)[157].

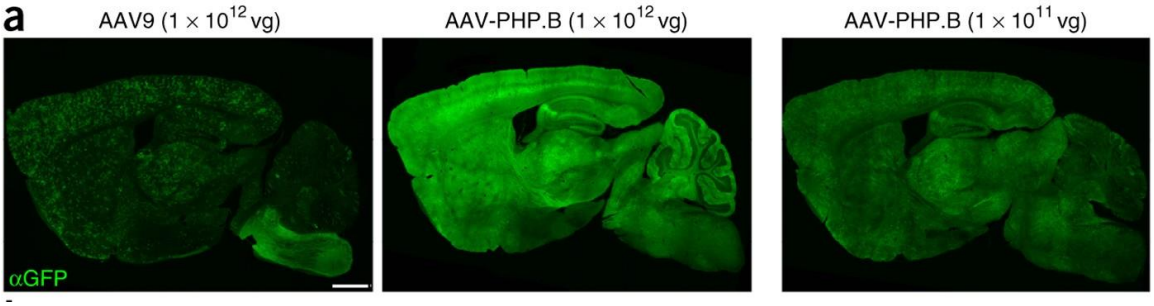


Figure 1.3: AAV-PHP.B mediates efficient gene delivery throughout the CNS after intravenous injection in adult mice.

(a) ssAAV9:CAG-GFP or ssPHP.B:CAG-GFP, at 1×10^{12} or 1×10^{11} vg/mouse (right), was intravenously injected into adult mice. Images show GFP expression 3 weeks after injection. Representative images of GFP IHC in the brains of mice given AAV9 (left) or B (middle and right). Scale bar, 1 mm.

The follow-up work from the Gradinaru lab, by Chan, K., *et al*[133] reported additional AAV variants using CREATE. One new variant AAV-PHP.S provides efficient transduction of sensory neurons and cardiac muscle, and the second reported variant, AAV-PHP.eB, that was further evolved from AAV-PHP.B parent, exhibits broad CNS transduction with higher neuron transduction efficiency compared to AAV-PHP.B.

In summary, these studies support the use of CREATE to evolve AAVs that show greater efficiency for the desired nerve cell types.

1.6 RESEARCH MOTIVATION

Thesis aims

1. To expand the potential use of the engineered PHP vectors through development of a detailed easy-to-follow protocol, and establish strategies to achieve transgene expression in specific cell population through cargo engineering, and study the usage of vectors across mouse strains or species.

AAV-PHP.eB or AAV-PHP.B's ability to broadly transduce different cell-types of CNS has been well received by the scientific community for neuroscience applications[158–163]. However, as every study involves specific cargos to deliver into the desired cell populations, its usage can be broadened by establishing a detailed and optimized protocol to help scientists produce these vectors in a laboratory setting for their own needs. Further, these engineered vectors can further benefit from strategies to refine the delivered transgene expression in a select population under investigation across different host systems.

2. To build a high-confidence in vivo selection platform in order to overcome current limitations of existing selection technology such as CREATE, and thereby speed up the development of the AAV toolkit to target distinct cell populations of the central nervous system via systemic delivery.

Developing highly efficient target specific and safe vehicles are ambitious goals to achieve and any significant success in this direction will benefit the scientific and medical communities. Such selective tropism requires an *in vivo* selection platform where we can perform parallel positive selections across targets of interest, and couple the process with post-hoc negative selections across off-targets. I would like to approach this by building upon the CREATE method that would allow us to find high performing variants in different mouse nerve cell types by positive and negative selection, henceforth referred to Multiplexed-CREATE.

The selected capsids could be supplemented with strategies to further refine the expression to the target cell population such as the one described earlier involving cargo engineering.

Current limitations with CREATE method

The published CREATE system allows us to do positive selection[133,157]. In other words, we can enrich for improved variants with several rounds of selection. While this is optimal to select for high transduction efficiency, this may not be the best strategy if we want to look for other traits such as selective tropism, functional viruses with no tropism, broad tropisms, etc. Also, the CREATE method does not provide us with an understanding of how the enriched variants performed in every step of the selection. This is very useful information for the experimenter to analyze the root cause for enrichment. Were these variants recovered because they highly transduced the cells of interest, or simply because of some other bias introduced by variabilities in virus production, DNA library assembly, or PCR amplification? Since we lacked this first-hand knowledge on the performance of variants in the CREATE method, we often do not have much confidence about which variants to test individually.

Current limitations with the translation of AAV-PHP.B vectors across mouse strains and species

The recent works from different laboratories pointed out the limitations with the use of AAV-PHP.B variants across mouse strains. A mouse strain, BALB/cJ, was first identified to lack the enhanced CNS transduction seen with AAV-PHP.B or eB vectors[164]. Follow-up studies later noticed that the transduction was not consistent across different strains of mice, and broadly classified mouse strains that had PHP.B permissibility to those that did not[165,166]. This suggested that the BBB across strains of mice is not conserved. The differences across these strains led to the identification of the endothelial cell membrane protein receptor, ly6a, to be responsible for the AAV-PHP.B or eB uptake by the endothelial cells that forms the BBB[165,167,168]. This not only limits the use of these vectors across mouse strains but also their ability to translate across species as such the homolog for *ly6a* is not found in humans. Hence to potentially translate the use of vectors across different strains of mouse and across species, we need to select for new variants that may surpass the current limitations on translation.

Chapter 2

CHARACTERIZATION AND PRODUCTION OF SYSTEMIC AAV VECTORS FOR WIDESPREAD AND TARGETED GENE DELIVERY IN RODENTS

Challis, R. C.*****, **Ravindra Kumar, S.***, Chan, K. Y., Challis, C., Beadle K., Jang, M. J., Kim H. M., Rajendran, P. S., Tompkins, J. D., Shivkumar, K., Deverman, B. E., Gradinaru, V. Systemic AAV vectors for widespread and targeted gene delivery in rodents. *Nature Protocol* 14**,** 379–414 (2019). https://doi.org/10.1038/s41596-018-0097-3

**equal contribution.*

2.1 SUMMARY

We recently developed adeno-associated virus (AAV) capsids to facilitate efficient and noninvasive gene transfer to the central and peripheral nervous systems. However, a detailed protocol for generating and systemically delivering novel AAV variants was not previously available. In this protocol, we describe how to produce and intravenously administer AAVs to adult mice to specifically label and/or genetically manipulate cells in the nervous system and organs, including the heart. The procedure comprises three separate stages: AAV production, intravenous delivery, and evaluation of transgene expression. The protocol spans 8 d, excluding the time required to assess gene expression, and can be readily adopted by researchers with basic molecular biology, cell culture, and animal work experience. We provide guidelines for experimental design and choice of the capsid, cargo, and viral dose appropriate for the experimental aims. The procedures outlined here are adaptable to diverse biomedical applications, from anatomical and functional mapping to gene expression, silencing, and editing.

2.2 INTRODUCTION

Recombinant AAVs (rAAVs) are commonly used vehicles for in vivo gene transfer and promising vectors for therapeutic applications[169]. However, AAVs that enable efficient and noninvasive gene delivery to defined cell populations are needed. Current gene delivery methods (e.g., intra-parenchymal surgical injections) are invasive, and alternatives such as intravenous administration require high viral doses and provide relatively inefficient transduction of target cells. We previously developed Cre recombination-based AAV targeted evolution (CREATE) to engineer and screen for AAV capsids that are capable of more efficient gene transfer to specific cell types via the vasculature[133,134,157]. Compared to naturally occurring capsids (e.g., AAV9), the novel AAV-PHP capsids identified by CREATE exhibit markedly improved tropism for cells in the adult mouse central nervous system (CNS), peripheral nervous system (PNS), and visceral organs. In this protocol, we describe how to package genetic cargo into AAV-PHP capsids and intravenously administer AAVs for efficient, noninvasive, and targeted gene delivery at sites throughout the body (***Figure 2.1***).

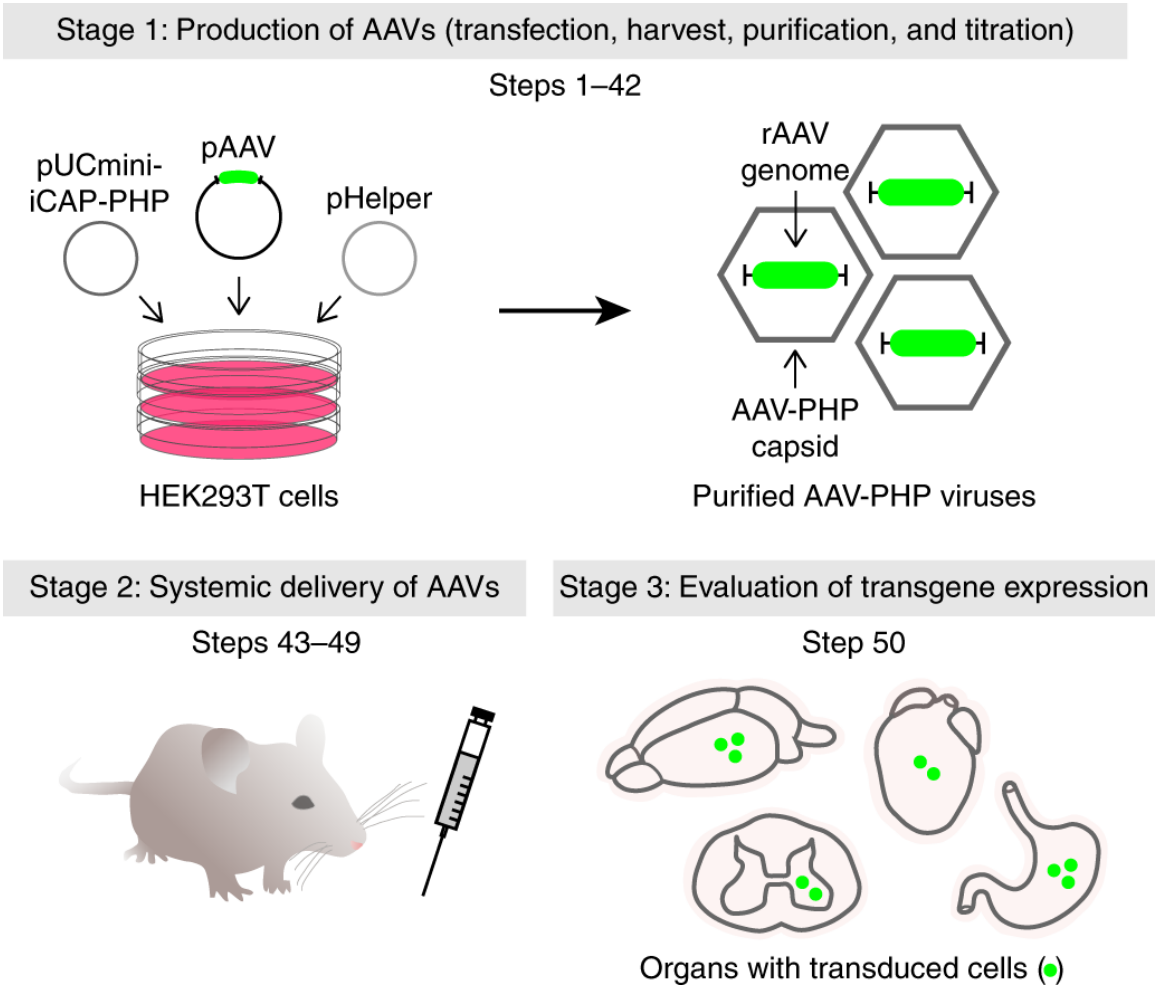


Figure 2.1: Overview of the protocol.

The Procedure comprises three main stages: AAV production (Steps 1–42), intravenous delivery (Steps 43–49), and evaluation of transgene expression (Step 50). The pAAV plasmid contains the rAAV genome (e.g., containing a fluorescent reporter, shown in green) (***Figure 2.6 and Table 2.T1***), which is packaged into an AAV-PHP capsid via triple transient transfection. Systemic administration of AAV-PHP viruses is achieved by retro-orbital injection into wild-type or transgenic mice; transgene expression is evaluated after adequate time has passed for viral transduction and protein expression. AAV-PHP viruses target cells in the CNS (e.g., in the brain and spinal cord) or PNS and visceral organs (e.g., in the heart and gut). Filled green circles represent transduced cells. For illustrative purposes, we use fluorescent labeling as an example of how to assess transgene expression; however, assessment can take other forms (see '*Experimental design*' section for details). See ***Figure 2.7a*** for a time line of the Procedure.

Among our new capsid variants[133,134,157], AAV-PHP.B and the further evolved AAV-PHP.eB efficiently transduce neurons and glia throughout the CNS (***Figure 2.2***); another variant, AAV-PHP.S, displays improved tropism for neurons within the PNS (***Figure 2.3***) and organs, including the gut[133] and heart (***Figure 2.4***). Importantly, these capsids target cell populations that are normally difficult to access because of their location (e.g., sympathetic, nodose, dorsal root, and cardiac ganglia) (***Figures 2.3a–c*** and ***2.4d***) or broad distribution (e.g., throughout the brain or enteric nervous system) (***Figures 2.2*** and ***2.3d***) and can be utilized in several mouse and rat strains (***Figure 2.5***). Together with the capsid, the genetic cargo (or rAAV genome) can be customized to control transgene expression (***Figure 2.6*** and ***Table 2.T1***). The rAAV genome contains the components required for gene expression, including promoters, transgenes, protein trafficking signals, and recombinase-dependent expression schemes. Hence, different capsid–cargo combinations create a versatile AAV toolbox for genetic manipulation of diverse cell populations in wild-type and transgenic animals. Here, we provide researchers, especially those new to working with AAVs or systemic delivery, with resources that will help them utilize AAV-PHP viruses in their own research.

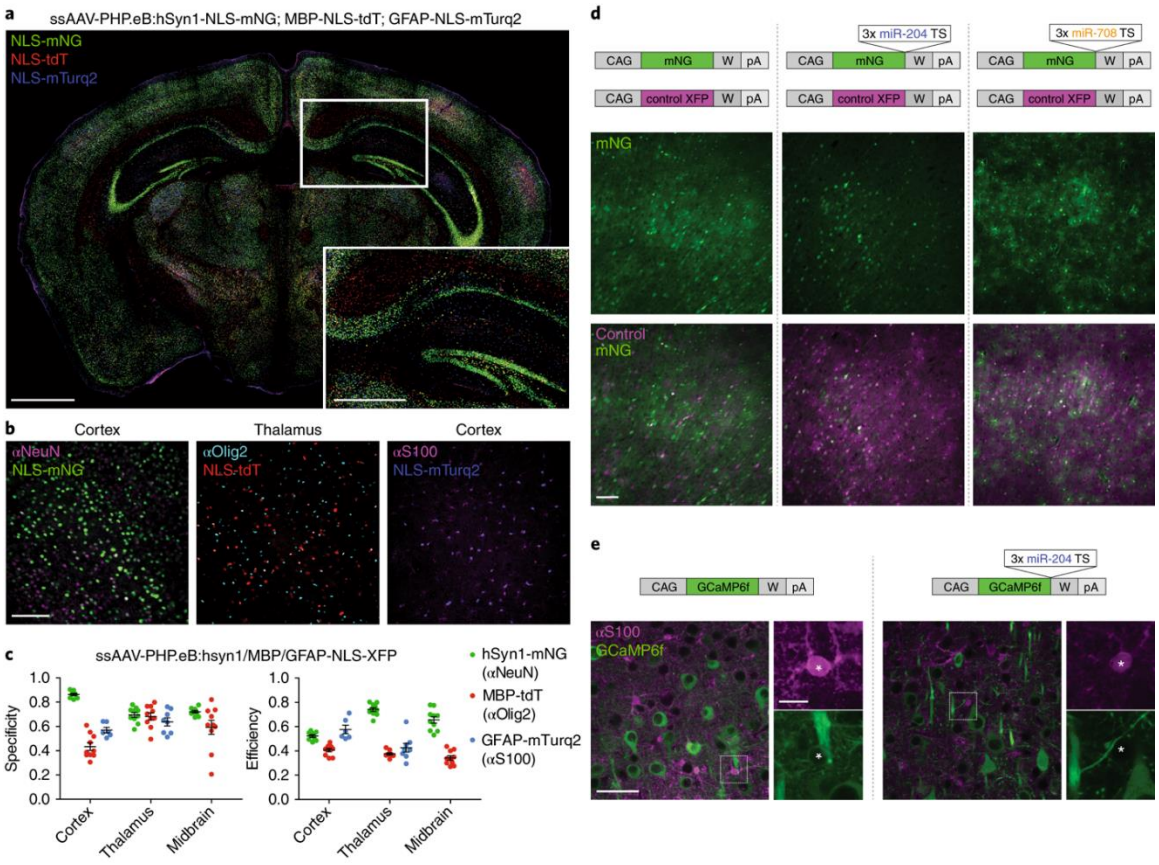


Figure 2.2: AAV-PHP.eB and gene regulatory elements enable cell type–specific gene expression in the brain.

(a–c), We used AAV-PHP.eB to package single-stranded (ss) rAAV genomes that express fluorescent reporters (XFPs), each with two nuclear localization signals (NLS), from cell type–specific promoters. Genomes containing the hSyn1, MBP, or GFAP (GfABC1D) promoters were used to target neurons, oligodendrocytes, or astrocytes, respectively. Viruses were co-delivered by retro-orbital injection to 7-week-old C57BL/6N mice (n = 2) at 3×10^{11} vector genomes (vg)/virus (9×10^{11} vg total). Native fluorescence in coronal brain sections was evaluated 4 weeks later using confocal microscopy. All sections were mounted in Prolong Diamond Antifade before imaging. **(a),** Cell type–specific, nuclear-localized XFPs label distinct cell types throughout the brain. Tile scan of a coronal brain slice, presented as a maximum-intensity projection; inset shows a zoomed-in view of the hippocampus. XFPs were mNeonGreen (mNG; green), tdTomato (tdT; red), and mTurquoise2 (mTurq2; blue). Scale bars, 1 mm and 500 μm (inset). **(b,c),** Antibody staining can be used to determine the specificity and efficiency of cell type–specific promoters. **(b),** Brain sections were stained with NeuN (purple), Olig2 (light blue), and S100 (purple) to mark neurons, oligodendrocyte lineage cells, and a population of glia that consists mainly of astrocytes, respectively. NLS-mNG (green), NLS-tdT (red), and NLS-mTurq2 (dark blue) indicate nuclear-localized XFPs. Images are from a single z plane.

Scale bar, 100 µm. **(c),** AAV-PHP.eB differentially transduces various regions and cell types throughout the brain. 'Specificity' or 'Efficiency' are defined as the ratio of double-labeled cells to the total number of XFP- or antibody-labeled cells, respectively. For image processing, median filtering and background subtraction using morphological opening were first applied to each image to reduce noise and correct imbalanced illumination. Each nucleus expressing XFPs and labeled with antibodies was then segmented by applying a Laplacian of Gaussian filter to the pre-processed images. We considered cells that were both expressing XFPs and labeled with antibodies if the nearest center-to-center distance between blobs (nuclei or cell bodies) in two channels was <7 µm (half of the cell body size). Five images per brain region were analyzed in each mouse; we excluded images with tissue edges because bright edges prevent accurate cell detection. Mean ± s.e.m. is shown. **(d,e),** miRNA target sequences (TS) miR-204-5p or miR-708-5p[135] can be used to achieve expression that is more restricted to neurons or astrocytes, respectively. **(d),** The indicated pairs of vectors were separately packaged into AAV-PHP.eB and co-administered via retro-orbital injection to 6- to 8-week-old C57BL/6J mice (n = 2) at 1×10^{11} vg/virus (2×10^{11} vg total); mNG and control XFP fluorescence were evaluated 3 weeks later using confocal microscopy. The CAG-mNG genome (green) contained no miRNA TS (left) or three tandem copies of miR-204 (middle) or miR-708 (right) TS; the CAG-XFP genome (magenta) contained no miRNA TS and was injected as an internal control. miR-204 reduced expression in cells with the morphology of astrocytes, and miR-708 reduced expression in cells with neuronal morphology. Scale bar, 100 µm. **(e),** ssAAV-PHP.eB:CAG-GCaMP6f-3x-miR122-TS (left) or ssAAV-PHP.eB:CAG-GCaMP6f-3x-miR204-5p-3x-miR122-TS (right) was injected into 6- to 8-week-old C57BL/6J mice (n = 2) at 1×10^{11} vg/mouse; gene expression was evaluated 3 weeks later using confocal microscopy. The miR-204 TS reduced GCaMP6f expression (green) in S100+ glia (magenta) in the cortex. Both vectors contained three tandem copies of miR-122 to reduce expression in hepatocytes[136]. Insets and asterisks highlight representative images of S100+ glia. Scale bars, 50 µm and 10 µm (insets). Refer to ***Table 2.T1*** for details of rAAV genomes. Experiments on vertebrates conformed to all relevant governmental and institutional regulations and were approved by the Institutional Animal Care and Use Committee (IACUC) and the Office of Laboratory Animal Resources at the California Institute of Technology. In our primary publication[133], results were obtained using the C57BL/6J mouse line. pA, polyadenylation signal; W, WPRE.

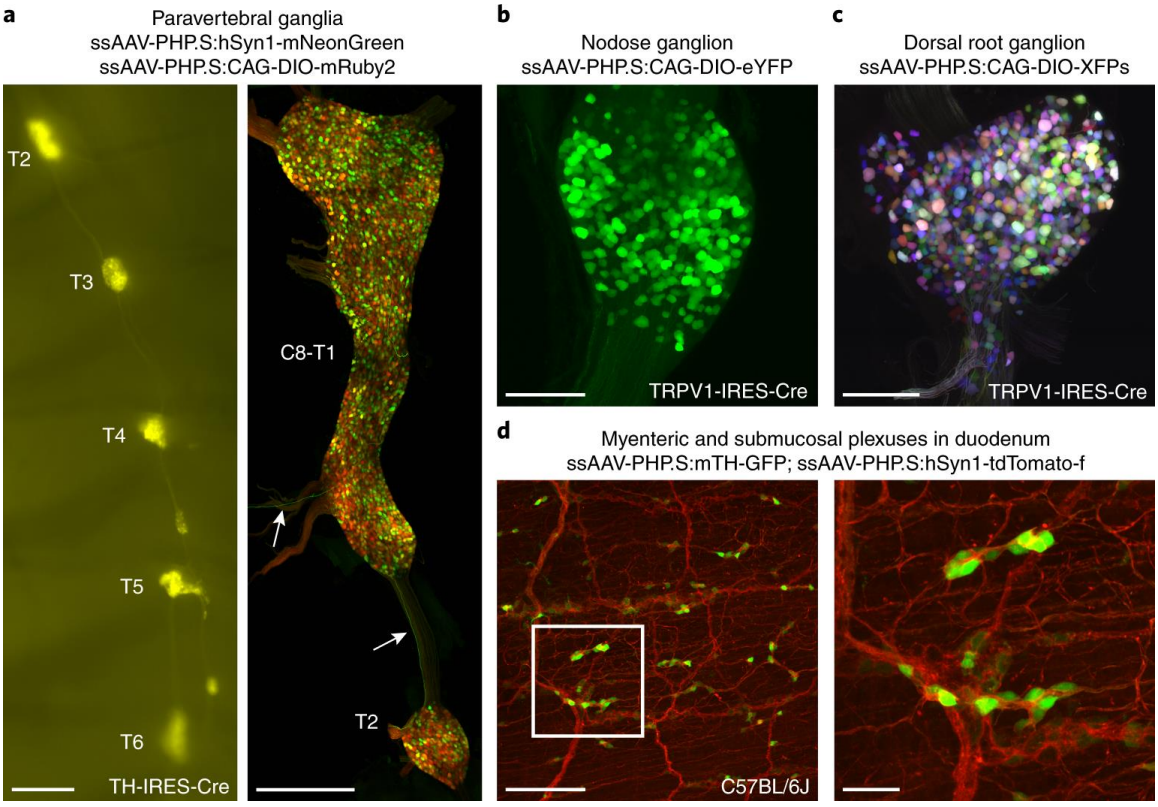


Figure 2.3: AAV-PHP.S transduces neurons throughout the PNS.

We used AAV-PHP.S to package single-stranded (ss) rAAV genomes that express fluorescent reporters from either neuron-specific (e.g., hSyn1 and TH (tyrosine hydroxylase)) or ubiquitous promoters (e.g., CAG). Viruses were delivered by retro-orbital injection to 6- to 8-week-old C57BL/6J or Cre transgenic mice, and transgene expression was evaluated 2–3 weeks later. Whole-mount tissues were optically cleared using either ScaleSQ[170] (**a** (right), **c**, and **d**) or RIMS[171,172] (**b**) and imaged using wide-field or confocal microscopy; confocal images are presented as maximum-intensity projections. **(a)**, ssAAV-PHP.S:hSyn1-mNeonGreen and ssAAV-PHP.S:CAG-DIO-mRuby2 were co-injected into a TH-IRES-Cre mouse at 1×10^{12} vg/virus (2×10^{12} vg total). Native mNeonGreen (green) and mRuby2 (red) fluorescence were assessed 2 weeks later using wide-field (left) or confocal fluorescence microscopy (right). Images are from the second to sixth thoracic (T2–T6) (left) and eighth cervical to second thoracic (C8–T2) (right) paravertebral ganglia, which provide sympathetic innervation to thoracic organs, including the heart. Arrows denote mNeonGreen$^+$ nerve fibers. Scale bars, 1 mm (left) and 500 µm (right). **(b)**, ssAAV-PHP.S:CAG-DIO-eYFP was injected into a TRPV1-IRES-Cre mouse at 1×10^{12} vg; gene expression in a nodose ganglion was evaluated 3 weeks later. Scale bar, 200 µm. **(c)**, A mixture of three separate viruses (ssAAV-PHP.S:CAG-DIO-XFPs) was injected into a TRPV1-IRES-Cre mouse at 1×10^{12} vg/virus (3×10^{12} vg total); gene expression in a dorsal root ganglion was evaluated 2 weeks later. XFPs were mTurquoise2 (blue), mNeonGreen

(green), and mRuby2 (red). Scale bar, 200 µm. **(d)**, ssAAV-PHP.S:mTH-GFP and ssAAV-PHP.S:hSyn1-tdTomato-f (farnesylated) were co-injected into a C57BL/6J mouse at 5×10^{11} vg/virus (1×10^{12} vg total); gene expression in the duodenum was assessed 22 d later. The image stack includes both the myenteric and submucosal plexuses. Inset shows a zoomed-in view of ganglia containing TH^+ cell bodies (green); tdTomato-f (red) labels both thick nerve bundles and individual fibers. Scale bars, 200 µm (left) and 50 µm (right). Refer to ***Table 2.T1*** for details of rAAV genomes. Experiments on vertebrates conformed to all relevant governmental and institutional regulations and were approved by the Institutional Animal Care and Use Committee (IACUC) and the Office of Laboratory Animal Resources at the California Institute of Technology. In our primary publication[133], results were obtained using the ChAT-IRES-Cre driver mouse line.

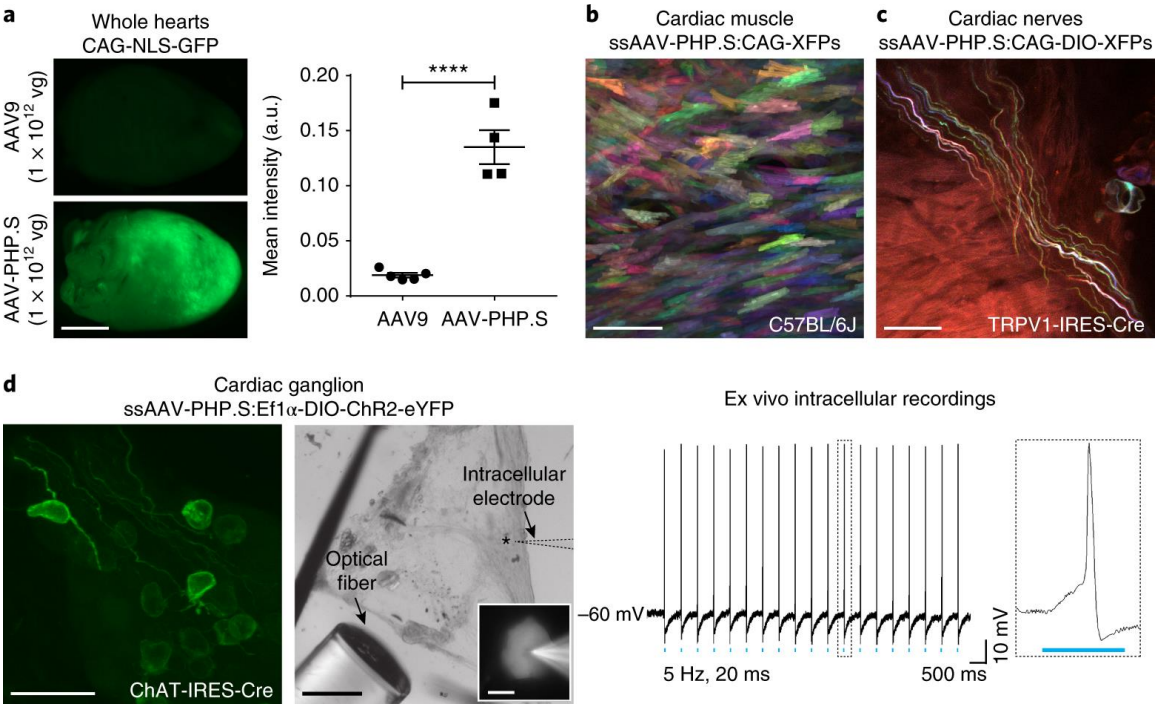


Figure 2.4: AAV-PHP.S for mapping the anatomy and physiology of the heart.

AAV-PHP.S viruses were delivered by retro-orbital injection to 6- to 8-week-old C57BL/6J or Cre transgenic mice. **(a)**, AAV-PHP.S transduces the heart more efficiently than the current standard, AAV9. ssAAV9:CAG-NLS-GFP or ssAAV-PHP.S:CAG-NLS-GFP were injected into C57BL/6J mice at 1×10^{12} vg/mouse. Native GFP fluorescence was assessed in whole-mount hearts 4 weeks later using wide-field fluorescence microscopy (unpaired *t* test, $t_7 = 8.449$, $^{****}P < 0.0001$). For AAV9 and AAV-PHP.S, $n = 5$ and 4 mice, respectively. a.u., arbitrary units. Mean ± s.e.m. is shown. Scale bar, 3 mm. **(b)**, A mixture of three viruses (ssAAV-PHP.S:CAG-XFPs) was injected into a C57BL/6J mouse at 3.3×10^{11} vg/virus (1×10^{12} vg total); gene expression in cardiac muscle was evaluated 11 d later.

Individual cardiomyocytes can be easily distinguished from one another. Scale bar, 200 μm. **(c)**, A mixture of three viruses (ssAAV-PHP.S:CAG-DIO-XFPs) was injected into a TRPV1-IRES-Cre mouse at 1×10^{12} vg/virus (3×10^{12} vg total); gene expression in cardiac nerves was evaluated 2 weeks later. Scale bar, 50 μm. **(d)**, ssAAV-PHP.S:Ef1α-DIO-ChR2-eYFP was injected into ChAT-IRES-Cre mice ($n = 2$) at 1×10^{12} vg; gene expression in a cardiac ganglion was evaluated 3 weeks later (left). *Ex vivo* intracellular recordings were performed after 5 weeks of expression. Differential interference contrast (DIC) image (middle) shows the optical fiber for light delivery and electrode for concurrent intracellular recordings; inset shows a higher-magnification image of a selected cell (asterisk). Cholinergic neurons generated action potentials in response to 473-nm light pulses (5 Hz, 20 ms) (right). Scale bars, 50 μm (left), 300 μm (middle), and 10 μm (inset). Whole-mount tissues in **(b)**, **(c)**, and **(d)** (left) were optically cleared using ScaleSQ[170] and imaged using confocal microscopy; confocal images are presented as maximum-intensity projections. XFPs in **(b)** and **(c)** were mTurquoise2 (blue), mNeonGreen (green), and mRuby2 (red). Refer to ***Table 2.T1*** for details of rAAV genomes. The pAAV-Ef1α-DIO-ChR2-eYFP plasmid was a gift from K. Deisseroth, Stanford University (Addgene, plasmid no. 20298). Experiments on vertebrates conformed to all relevant governmental and institutional regulations and were approved by the Institutional Animal Care and Use Committee (IACUC) and the Office of Laboratory Animal Resources at the California Institute of Technology. In our primary publication[133], results were obtained using the ChAT-IRES-Cre driver mouse line.

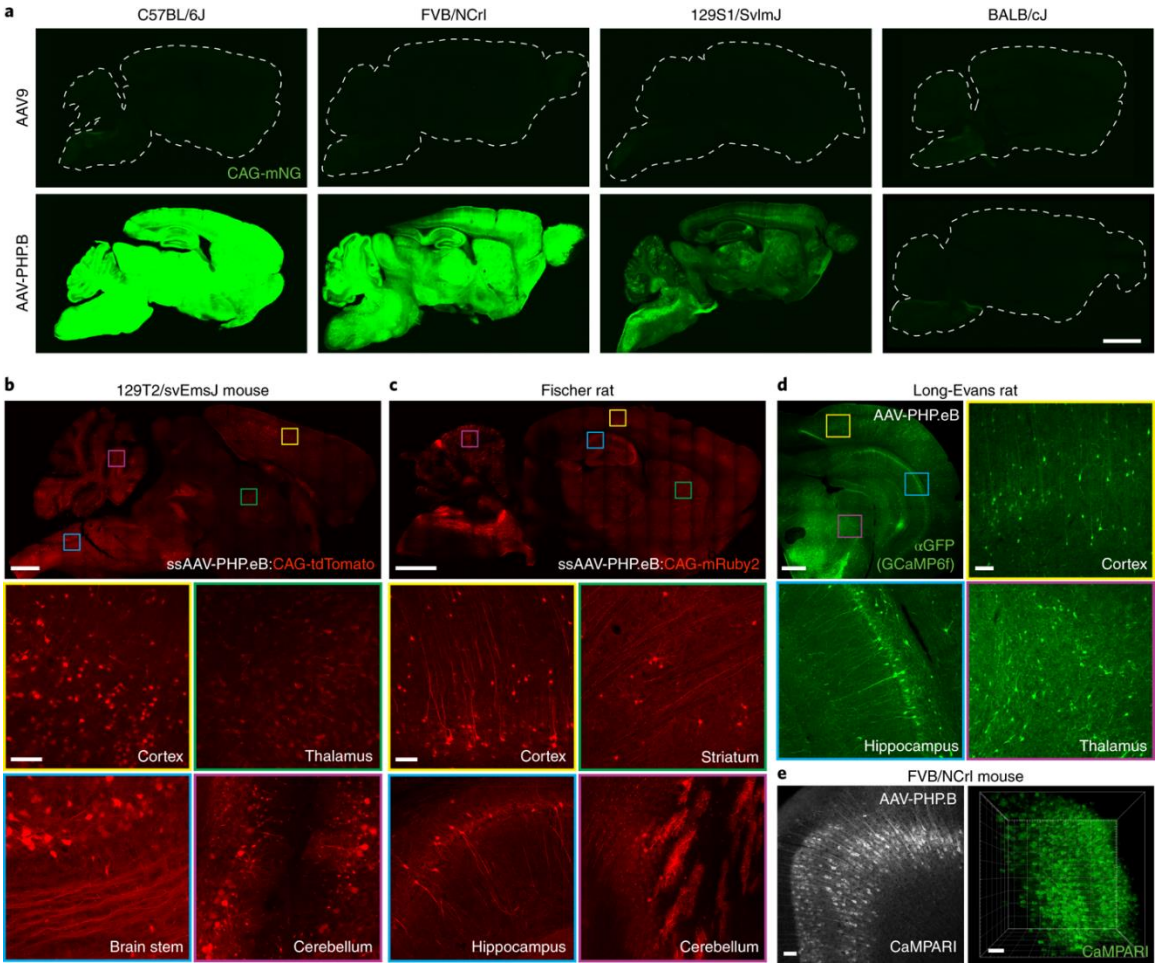


Figure 2.5: AAV-PHP.B and AAV-PHP.eB can be used in several mouse and rat strains.

(a), AAV-PHP.B transduces the brain more efficiently than AAV9 in C57BL/6J, FVB/NCrl, and 129S1/SvImJ mice, but not in BALB/cJ mice. ssAAV9:CAG-mNeonGreen or ssAAV-PHP.B:CAG-mNeonGreen were systemically delivered to 6- to 8-week-old C57BL/6J (n = 1–2 mice per group), FVB/NCrl (n = 2 mice per group), 129S1/SvImJ (n = 2 mice per group), and BALB/cJ mice (n = 2 mice per group) at 1×10^{12} vg/mouse. 3 weeks later, sagittal brain sections were mounted in Vectashield and imaged using confocal microscopy. Imaging and display parameters are matched across all panels. Scale bar, 2 mm. **(b–e)**, Examples of AAV-PHP.B- and AAV-PHP.eB-mediated brain transduction for fluorescent labeling **(b,c)** and calcium imaging **(d,e)** in different mouse and rat strains. Gene expression was evaluated using confocal microscopy. **(b)**, ssAAV-PHP.eB:CAG-tdTomato (Addgene) was delivered by retro-orbital injection to a 10-week-old 129T2/SvEmsJ mouse at 3×10^{11} vg; tdTomato fluorescence (red) was examined 2 weeks later. Scale bars, 1 mm (top) and 100 µm (insets). **(c)**, ssAAV-PHP.eB:CAG-mRuby2 was administered by tail-vein injection to a 6-week-old female Fischer rat at 3×10^{12} vg; 3 weeks later, brain slices were mounted in Prolong Diamond Antifade for imaging. Scale bars, 2 mm (top) and 100 µm (insets). **(d)**, ssAAV-PHP.eB:CMV-hSyn1-GCaMP6f-3x-miR122-TS was delivered by tail-vein

injection to a 4-week-old female Long-Evans rat at 1×10^{13} vg; 3 weeks later, brain slices were stained with a GFP antibody (green) for imaging. Scale bars, 1 mm (top left) and 100 μm (insets). The vector contained three tandem copies of miRNA target sequence (TS) miR-122 (CAAACACCATTGTCACACTCCA) to reduce expression in hepatocytes[136]. Images in **(d)** courtesy of M. Fabiszak/W. Freiwald lab, Rockefeller University. **(e)**, ssAAV-PHP.B:CaMKIIa-CaMPARI (calcium-modulated photoactivatable ratiometric integrator[173]) was administered by retro-orbital injection to a 8-week-old FVB/NCrl mouse at 3×10^{11} vg and cortical expression was assessed 2 weeks later. Images are a 50-μm maximum-intensity projection of the cortex (left) and 500-μm-thick ScaleSQ[170]-cleared 3D volume (right). Scale bars, 100 μm. Experiments on vertebrates conformed to all relevant governmental and institutional regulations and were approved by the Institutional Animal Care and Use Committee (IACUC) and the Office of Laboratory Animal Resources at the California Institute of Technology. In our primary publication[157], results were obtained using the C57BL/6J mouse line. CaMKIIa, calcium/calmodulin-dependent protein kinase type IIa; CMV, cytomegalovirus early enhancer element.

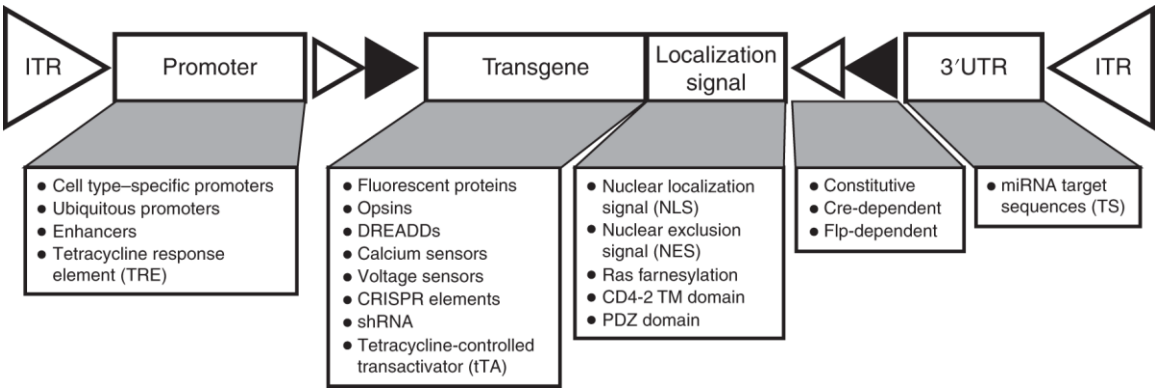


Figure 2.6: A modular AAV toolbox for cell type–specific gene expression.

The rAAV genome, contained in a pAAV plasmid (not shown), consists of an expression cassette flanked by two 145-bp inverted terminal repeats (ITRs); the entire genome, including the ITRs, cannot exceed 4.7–5 kb. The promoters, transgenes, localization signals, and recombination schemes are interchangeable. Gene regulatory elements, such as promoters and microRNA (miRNA) target sequences (TS) (***Figure 2.2d,e***), determine the strength and specificity of transgene expression[125]. Transgenes may be constitutively expressed or flanked by recombination sites for flippase (Flp)- or Cre recombinase (Cre)-dependent expression. In the latter approach, the transgene remains in the double-floxed inverted orientation (DIO); Cre-mediated inversion of the transgene enables cell type–specific expression in transgenic animals (***Figures 2.3a–c*** and ***Figures 2.4c,d***). Localization sequences further restrict gene expression to distinct cellular compartments such as the nucleus (via one or more nuclear localization signals (NLS)) (***Figure 2.2a,b***), cytosol (via a nuclear exclusion signal (NES)[174]), or cell membrane (via farnesylation[175],

the CD4-2[176] transmembrane (TM) targeting domain, or PDZ[177] protein–protein interaction domains) (***Figure 2.3d***). Note that the 3′ UTR contains the woodchuck hepatitis posttranscriptional regulatory element (WPRE) (609 bp) and a polyadenylation signal (e.g., the human growth hormone (hGH) polyA) (479 bp) (not shown), both of which enhance transgene[125]. We recommend that foreign genes be codon-optimized to match the host species to increase expression from the rAAV genome. Use sequence-editing and annotation software to determine the unique attributes of each rAAV genome. In ***Table 2.T1***, we list genomes used here and in our previous work[133,157]; see also Addgene's plasmid repository for pAAVs that may be suitable for different applications. CRISPR, clustered regularly interspaced short palindromic repeats; DREADDs, designer receptors exclusively activated by designer drugs; shRNA, short hairpin RNA.

	Vector name pAAV-	**Expression class**	**Addgene no.**
Tunable expression	TRE[a]-mTurquoise2	tTA-dependent	99113
	TRE-eYFP		104056
	TRE-mRuby2		99114
	TRE-DIO[b]-mTurquoise2	Cre- and tTA-dependent	99115
	TRE-DIO-eYFP		117383
	TRE-DIO-tdTomato		99116
	TRE-DIO-mRuby2		99117
	CAG[c]-tTA[d]	Inducer	99118
	hSyn1[e]-tTA		99119
	ihSyn1[f]-tTA		99120
	ihSyn1-DIO-tTA		99121
Tissue-wide expression	CAG-mTurquoise2	Constitutive	99122
	CAG-eYFP		104055
	CAG-mRuby2		99123
	CAG-NLS[g]-GFP		104061
	CAG-DIO-mTurquoise2	Cre-dependent	104059
	CAG-DIO-eYFP		104052

	Vector name pAAV-	Expression class	Addgene no.
Cell type–specific expression	CAG-DIO-mRuby2	Cell type–specific	104058
	hSyn1-mTurquoise2		99125
	hSyn1-eYFP		117382
	hSyn1-mRuby2		99126
	GFAP[h]-2xNLS-mTurquoise2		104053
	hSyn1-2xNLS-mTurquoise2		118025
	MBP[i]-2xNLS-tdTomato		104054
	mTH[j]-GFP		99128
	hSyn1-tdTomato-f[k]		104060
	GFAP-mKate2.5-f		99129
	mDlx[l]-NLS-mRuby2		99130
	CAG-eYFP-3x-miR204-5p-TS[m]		117380
	CAG-eYFP-3x-miR708-5p-TS[n]		117381
	CAG-GCaMP6f-3x-miR204-5p-3x-miR122-TS[o]		117384

Table 2.T1: pAAV plasmids

A comprehensive list of pAAV plasmids used in this and related work[133,157]. [a]TREpi, second-generation tetracycline-regulated promoter. [b]DIO, double-floxed inverted orientation. [c]CAG, synthetic promoter containing the cytomegalovirus early enhancer element, the promoter, first exon, and first intron of chicken beta-actin gene, and the splice acceptor from the rabbit beta-globin gene. [d]tTA, tetracycline-controlled transactivator. [e]hSyn1, human synapsin I promoter. [f]ihSyn1, inducible intron human synapsin I promoter. [g]NLS, nuclear localization signal. [h]GFAP (GfABC$_1$D), glial fibrillary acidic protein promoter. [i]MBP, myelin basic protein promoter. [j]mTH, mouse tyrosine hydroxylase promoter. [k]f, farnesylation signal from c-Ha-Ras. [l]mDlx, mouse distal-less homeobox promoter. [m]miR-204-5p-TS: AGGCATAGGATGACAAAGGGAA. [n]miR-708-5p-TS: CCCAGCTAGATTGTAAGCTCCTT. [o]miR-122-TS: CAAACACCATTGTCACACTCCA.

Overview of the protocol

We provide an instruction manual for users of AAV-PHP variants. The procedure includes three main stages (***Figure 2.1***): AAV production (Steps 1–42), intravenous delivery (Steps 43–49), and evaluation of transgene expression (Step 50).

The AAV production protocol is adapted from established methods. First, HEK293T cells are transfected with three plasmids[178–180] (Steps 1–3, ***Figure 2.1*** and ***Figure 2.7***): (i) pAAV, which contains the rAAV genome of interest (***Figure 2.6*** and ***Table 2.T1***); (ii) pUCmini-iCAP-PHP, which encodes the viral replication and capsid proteins (***Table 2.T2***); and (iii) pHelper, which encodes adenoviral proteins necessary for replication. Using this triple-transfection approach, a single-stranded rAAV genome is packaged into an AAV-PHP capsid in HEK293T cells. AAV-PHP viruses are then harvested[181] (Steps 4–14), purified[182,183] (Steps 15–31), and titered by quantitative PCR (qPCR)[184] (Steps 32–42) (***Figure 2.7***). Purified viruses are intravenously delivered to mice via retro-orbital injection[185] (Steps 43–49), and gene expression is later assessed using molecular, histological, or functional methods relevant to the experimental aims (Step 50).

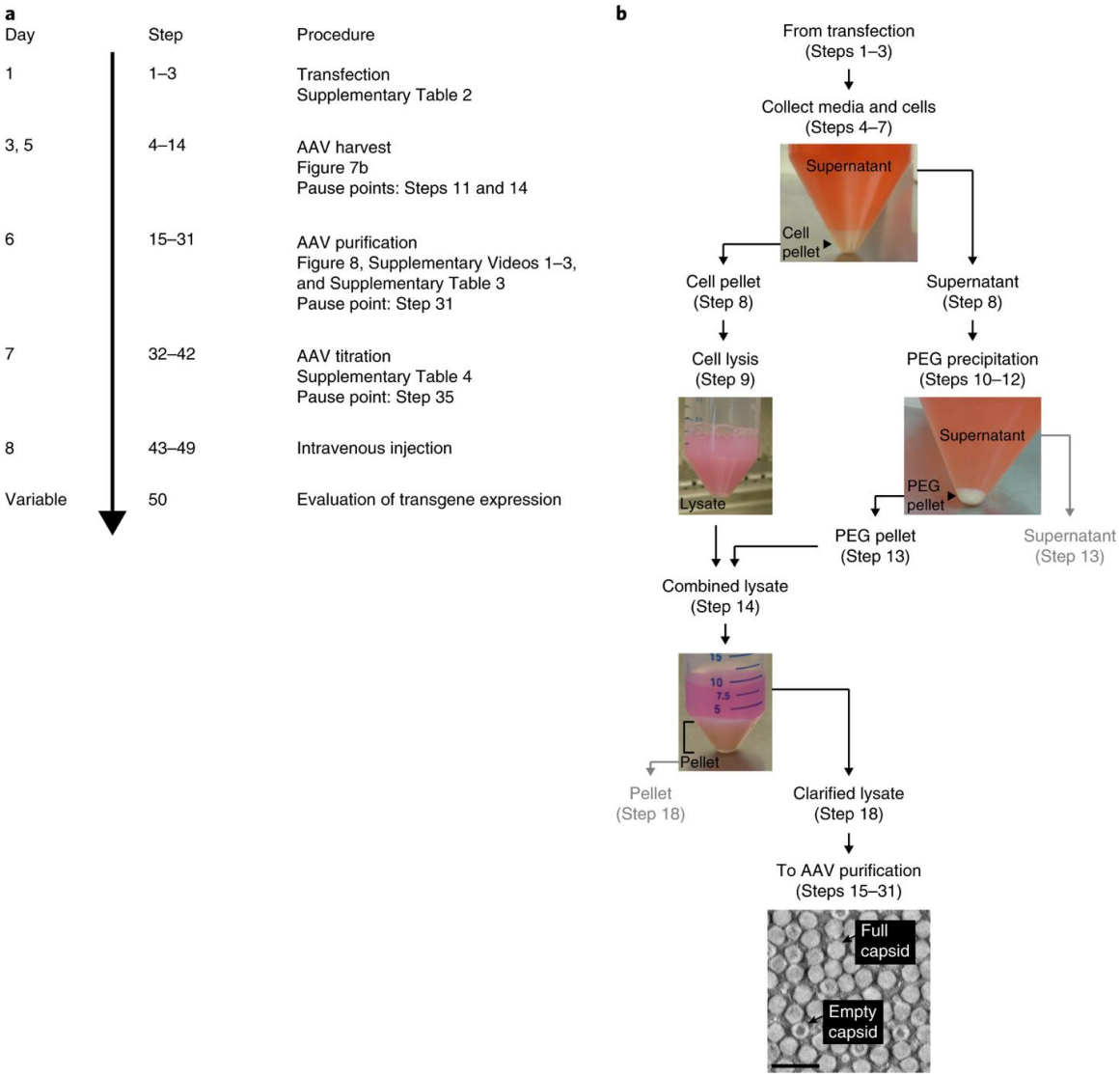


Figure 2.7: Time line and AAV harvest procedure.

(a), Time line of the procedure. The entire protocol spans 8 d, excluding pause points on days 5 (Steps 11 and 14), 6 (Step 31), and 7 (Step 35) and the time required to evaluate transgene expression (Step 50). Days 1–7 (Steps 1–42) constitute the AAV production stage (***Figure 2.1***). **(b)**, Schematic of the AAV harvest procedure, with images corresponding to indicated steps. The iodixanol-based purification protocol does not eliminate empty capsids (i.e., capsids that fail to package an rAAV genome), as determined by negative-staining transmission electron microscopy; empty particles are characterized by an electron-dense core. Scale bar, 50 nm. Gray arrows and text denote steps at which the supernatant and pellet can be bleached and discarded (Steps 13 and 18).

AAV-PHP capsid	Plasmid name	In vivo characteristics	Production efficiency	Addgene no.
AAV-PHP.B	pUCmini-iCAP-PHP.B	Broad CNS transduction	Good	103002
AAV-PHP.B2	pUCmini-iCAP-PHP.B2	Broad CNS transduction	Good	103003
AAV-PHP.B3	pUCmini-iCAP-PHP.B3	Broad CNS transduction	Good	103004
AAV-PHP.eB	pUCmini-iCAP-PHP.eB	Broad CNS transduction	Good	103005
AAV-PHP.S	pUCmini-iCAP-PHP.S	Broad transduction in PNS and visceral organs	Good	103006
AAV-PHP.A[a]	piCAP-PHP.A	Broad astrocyte transduction in CNS	Poor	CLOVER

Table 2.T2: AAV-PHP capsid plasmids

AAV-PHP capsid plasmids have a built-in tTA-TRE-based inducible amplification loop to increase virus production. If the rAAV genome has a tetracycline-regulated element (e.g., TRE), the tTA on the capsid plasmid will drive a high level of expression from the TRE-containing rAAV genome, which may reduce virus production. To increase viral yields, increase the number of dishes per viral prep. [a]Given the poor production efficiency of AAV-PHP.A, and its tendency to aggregate after purification, we suggest using AAV-PHP.eB to target astrocytes. Use an astrocyte promoter, such as GFAP, to drive transgene expression (***Figure 2.2a-c***). To request AAV-PHP.A (listed as CLOVER in the table), visit http://www.clover.caltech.edu/. iCAP, inducible *cap* expression; pUCmini, pUC origin of replication.

This protocol is optimized to produce AAVs at high titer (≥1 × 10^{13} vector genomes (vg)/ml and ≥1 × 10^{12} vg/dish) and with high transduction efficiency *in vivo*[133,157].

Applications of the method

We anticipate that AAV-PHP capsids (***Table 2.T2***) can be used to package rAAV genomes (contained in pAAV plasmids that are available through Addgene and elsewhere) (***Figure 2.6*** and ***Table 2.T1***) to enable a wide range of biomedical applications. Below, we highlight current and potential applications of this method.

Anatomical mapping

Fluorescent reporters are commonly used for cell type–specific mapping and phenotyping[133,186,187] (***Figures 2.2–2.5***). AAV-mediated multicolor labeling (e.g., Brainbow[188]) is especially advantageous for anatomical mapping approaches that require individual cells in the same population to be distinguished from one another. We and others have demonstrated the feasibility of this approach in the brain[133,188] retina[188], heart (***Figure 2.4b,c***), and gut[133], as well as the peripheral ganglia (***Figure 2.3c***). Spectrally distinct labeling is well-suited for studying the organization of cells (e.g., cardiomyocytes (***Figure 2.4b***)) in healthy and diseased tissues and long-range tract tracing of individual fibers through extensive neural networks (e.g., the enteric[133] or cardiac nervous systems (***Figure 2.4c***)).

Functional mapping

AAV-PHP capsids are also relevant for probing cell function. AAV-PHP.B was previously used to target distinct neural circuits throughout the brain for chemogenetic[158,189] and optical imaging applications[190,191]. We predict that AAV-PHP viruses will be beneficial for manipulating neural networks that are typically difficult to access, such as peripheral circuits controlling the heart (***Figure 2.4d***), lungs[192], or gut[193]. AAV-PHP variants could also be utilized to interrogate the function of non-neuronal cell types, including cardiomyocytes[194], pancreatic beta cells[195,196], and hepatocytes[197]. Harnessing AAV-PHP viruses to modulate cell physiology may reveal novel roles for different cells in regulating organ function and/or animal behavior[198].

Gene expression, silencing, and editing

AAV-PHP viruses can be used to test potential therapeutic strategies that would benefit from organ wide or systemic transgene expression[199]. Recently, AAV-PHP.B was used to treat[158] and model[159] neurodegenerative diseases with widespread pathology. Other potential applications include gene editing (e.g., via CRISPR[161,200–202]) or silencing (e.g., via shRNA[203]); importantly, these approaches could be utilized to broadly and noninvasively manipulate cells in both healthy and diseased states for either basic research or therapeutically motivated studies.

AAV capsid engineering

AAV-PHP capsids can be further evolved for more efficient transduction of specific organs and cell types throughout the body. This protocol can be used for AAV engineering applications (e.g., our *in vivo* capsid selection method CREATE[133,157]). Using a modified transfection protocol (Steps 1–3 and online methods in ref.[157]), DNA libraries (generated by diversification of the AAV cap gene) are packaged to produce AAV capsid libraries, which are then harvested (Steps 4–14 and online methods in ref.[157]), purified (Steps 15–31), and titered (Steps 32–42). Libraries are systemically administered to Cre transgenic animals (Steps 43–49) or wild-type animals in which Cre is introduced (e.g., by AAV delivery), and Cre-dependent cap recovery from tissues of interest facilitates further rounds of selection to isolate enriched variants. This protocol can also be used to characterize novel serotypes identified with CREATE or other engineering methods[204].

Limitations of the method

A major limitation of AAV capsids, including AAV-PHP variants, is their relatively small packaging capacity (<5 kb). Some elements of the rAAV genome, such as the woodchuck hepatitis posttranscriptional regulatory element (WPRE), can be truncated[127] or removed[126,205] to accommodate larger genetic components. The development of smaller promoters[128,129] and dual expression systems[206], in which genetic elements are split between

two or more viruses (requiring efficient co-transduction), has also enabled the delivery of larger genomes. Continued development of these approaches will help bypass restrictions on rAAV genome size. Intravenous administration of AAVs also presents unique challenges. For example, systemic transduction may be undesirable for applications in which highly restricted gene expression is vital to the experimental outcome. Possible off-target transduction, due to the broad tropism of AAV-PHP variants and/or lack of compatible cell type–specific promoters, can be reduced by miRNA-mediated gene silencing. Sequences complementary to miRNAs expressed in off-target cell populations can be introduced into the 3′ UTR of the rAAV genome (***Figure 2.6***); this has been shown to reduce off-target transgene expression and better restrict expression to cell types of interest[137,207] (***Figure 2.2d,e***). Another challenge of systemic delivery is that it requires a high viral load, which can elicit an immune response against the capsid and/or transgene and reduce transduction efficiency *in vivo*[208]. Immunogenicity of AAVs may be exacerbated by empty capsid contamination in viral preparations[209,210]. The viral purification protocol (Steps 15–31) reduces, but does not eliminate, empty capsids (***Figure 2.7b***). If this poses a concern for specific applications, viruses can be purified using an alternative approach[181,182,211]. The generation of viruses for systemic administration may impose a financial burden on laboratories due to the doses of virus required. Nevertheless, viral-mediated gene delivery is inexpensive compared to creating and maintaining transgenic animals. Moreover, intravenous injection is faster, less invasive, and less technically demanding than other routes of AAV administration, such as stereotaxic injection, thereby eliminating the need for specialized equipment and survival surgery training.

Experimental design

Before proceeding with the protocol, a number of factors should be considered, namely the expertise and resources available in the lab; the animal model, capsid, and rAAV genome to be used; the dose for intravenous administration; and the method(s) available for assessing transgene expression. Each of these topics is discussed below to guide users in designing their experiments.

Required expertise and resources

This protocol requires that the scientists have basic molecular biology, cell culture, and animal work experience. Users should be approved to handle laboratory animals, human cell lines, and AAVs. A background in molecular cloning is advantageous, although not necessary if relying on available plasmids. In addition to having the above expertise, the labs must be equipped for the molecular and cell culture work relevant to the procedure; we suggest that users read through the entire '*Materials*' and '*Procedure*' sections beforehand to ensure that the required reagents and equipment are available and appropriate safety practices and institutional approvals are in place.

Animal model

This protocol describes the production of AAVs for intravenous delivery to 6- to 8-week-old male and female mice. AAV-PHP viruses have been validated in C57BL/6J mice[133,157,158,160] (***Figures 2.2–2.5***) and numerous Cre driver lines[133,158,189,190], including, but not limited to, TH-IRES-Cre (***Figure 2.3***), TRPV1-IRESCre (***Figures 2.3*** and ***2.4***), and ChAT-IRES-Cre mice[133] (***Figure 2.4***). Intriguingly, AAV-PHP.B demonstrates low transduction throughout the brain when systemically administered to BALB/cJ mice[164] (***Figure 2.5a***). However, the neurotropic properties of AAV-PHP.B are not limited to the C57BL/6J strain in which they were selected. AAV-PHP.B transduces the brain more efficiently than AAV9 in both FVB/NCrl and 129S1/SvImJ mice (***Figure 2.5a***). We also show examples of AAV-PHP.eB transducing neurons in C57BL/6NCrl (***Figure 2.2a–c***) and 129T2/SvEmsJ mice (***Figure 2.5b***), as well as Fischer (***Figure 2.5c***) and Long Evans rats (***Figure 2.5d***). Compared to AAV9 and AAV-PHP.B, AAV-PHP.eB results in more efficient neuronal transduction in Sprague–Dawley rats after either intravenous administration or intra-parenchymal stereotaxic injections[159,162]. We predict that AAV-PHP capsids can be used in multiple species and strains for diverse applications, such as those requiring fluorescent labeling (***Figure 2.5a–c***) and calcium imaging (***Figure 2.5d,e***). We have not compared the transduction efficiencies of AAV9 and AAV-PHP capsids across all rodent strains and species or determined the optimal dose for transducing specific

organs and cell types in different animal models. Users should test these parameters to determine the utility of AAV-PHP variants in their research. See '*Reagents*' for mouse and rat catalog numbers.

Selecting an AAV-PHP capsid

We recommend choosing an AAV-PHP capsid (***Table 2.T2***) based on its tropism and production efficiency. Capsid properties are listed in ***Supplementary Table 2.ST1***; we include species, organs, and cell populations examined to date and note typical viral yields. We anticipate that most researchers will use AAV-PHP.eB (Addgene, plasmid no. 103005) or AAV-PHP.S (Addgene, plasmid no. 103006) in their experiments. AAV-PHP.eB and AAV-PHP.S produce viral yields similar to those of other high-producing naturally occurring serotypes (e.g., AAV9) and enable efficient, noninvasive gene transfer to the CNS or PNS and visceral organs, respectively[133] (***Figures 2.2–2.5***). The earlier capsid variants, which provide widespread CNS transduction, either produce suboptimal yields (AAV-PHP.A)[157] or have since been further evolved for enhanced transduction efficiency *in vivo* (AAV-PHP.B (Addgene, plasmid no. 103002))[133]. We therefore recommend using AAV-PHP.eB for CNS applications, especially when targeting neurons. Note, however, that the chosen capsid will ultimately depend on the experimental circumstances; multiple factors, including species[212], strain[164] (***Figure 2.5***), age[213], gender[214], and health[215], can influence AAV tropism. Testing the AAV-PHP variants in a variety of experimental paradigms will continue to reveal the unique attributes of each capsid and identify those most suitable for different applications.

Selecting an rAAV genome

Users must select an rAAV genome, contained in a pAAV plasmid, to package into the capsid (***Figures 2.1*** and ***2.6***; ***Table 2.T1***). In ***Table 2.T1***, we list the pAAVs used here (***Figures 2.2–2.4***) and in our previous work[133,157]; we direct users to Addgene's plasmid repository for additional pAAVs developed for various applications. Depending on the experimental aims, users can elect to design their own genomes[125] and clone from existing

pAAVs. When customizing plasmids, it is imperative that the rAAV genome, the sequence between and including the two inverted terminal repeats (ITRs), does not exceed 4.7–5 kb (***Figure 2.6***); larger genomes will not be fully packaged into AAV capsids, resulting in truncated genomes and low titers. The ITRs are 145-bp sequences that flank the expression cassette and are required for replication and encapsidation of the viral genome. ITRs are typically derived from the AAV2 genome and must match the serotype of the rep gene contained in the capsid plasmid; pUCmini-iCAP-PHP plasmids contain the AAV2 rep gene and are therefore capable of packaging genomes with AAV2 ITRs (i.e., almost any pAAV available from Addgene). Other genetic components (e.g., promoters, transgenes, localization signals, and recombination schemes) are interchangeable and can be customized for specific applications (***Figure 2.6***).

Dosage for intravenous administration

The optimal dose for intravenous administration to target cell populations must be determined empirically. We encourage users to refer to ***Figures 2.2–2.5*** and related work for suggested AAV-PHP viral doses. AAV-PHP variants have been successfully administered to adult mice[133,157,158,160] (***Figures 2.2–2.5***), neonatal mice[158], and neonatal and adult rats[159,162] **(*Figure 2.5c,d*)** for fluorescent labeling; they have also been used for calcium imaging[190,191] and optogenetic **(*Figure 2.4d*)**, chemogenetic[158,189], and therapeutic applications[158,159]. We typically administer between 1×10^{11} and 5×10^{11} vg of AAV-PHP.eB or between 3×10^{11} and 1×10^{12} vg of AAV-PHP.S to adult mice (6–8 weeks old). However, dosage will vary depending on the target cell population, desired fraction of transduced cells, and expression level per cell. AAVs independently and stochastically transduce cells, typically resulting in multiple genome copies per cell[133]. Therefore, higher doses generally result in strong expression (i.e., high copy number) in a large fraction of cells, whereas lower doses result in weaker expression (i.e., low copy number) in a smaller fraction of cells. To achieve high expression in a sparse subset of cells, users can employ a two-component system in which transgene expression is dependent on co-transduction of an inducer (e.g., a vector expressing Cre[216], Flp[133], or the tetracycline-controlled

transactivator (tTA)[133]); inducers are injected at a lower dose (typically 1×10^9 to 1×10^{11} vg) to limit the fraction of cells with transgene expression. Note that transgenes and gene regulatory elements (e.g., enhancers, promoters, and miRNA target sequences (***Figure 2.2d,e***)) can influence gene expression levels. Therefore, users should assess transgene expression from a series of doses and at several time points after intravenous delivery to determine the optimal experimental conditions.

Evaluation of transgene expression

Following *in vivo* delivery, AAV transduction and transgene expression increase over the course of several weeks. Although expression is evident within days after transduction, it does not reach a steady-state level until at least 3–4 weeks after transduction. Therefore, we suggest waiting for a minimum of 2 weeks before evaluating fluorescent labeling[133,157–159] (***Figures 2.2–2.5***) and at least 3–4 weeks before beginning optogenetic (***Figure 2.4d***), chemogenetic[158,189], and calcium imaging[190,191] experiments. Note that, like other AAVs, AAV-PHP variants are capable of providing long-term transgene expression. AAV-PHP.B-mediated cortical expression of a genetically encoded calcium indicator, GCaMP6s, was reported to last at least 10 weeks post-injection without toxic side effects[191] (i.e., nuclear filling[217]), and we have observed GFP expression throughout the brain >1 year after viral administration (see Supplementary Figure 4 in ref.[157]). However, the time points suggested here are only meant to serve as guidelines; gene expression is contingent on multiple factors, including the animal model, capsid, genome, and dose. The appropriate method(s) for evaluating transgene expression will vary among users and may include functional (e.g., optical imaging[217]), histological[218] (e.g., using endogenous fluorescence, antibodies, or molecular probes), or molecular (e.g., Western blot[219] or qPCR[133]) approaches[220]. To assess transduction efficiency across different organs, users can perform a qPCR-based vector biodistribution assay, in which vector genomes are quantified and normalized to the mouse genome (e.g., a housekeeping gene)[157]. Other approaches typically involve examining fluorescent protein expression in thin or thick (≥100 μm) tissue samples. The CLARITY-based methods such as passive

CLARITY technique (PACT) and perfusion-assisted agent release in situ (PARS)[171] render thick tissues optically transparent while preserving their three-dimensional molecular and cellular architecture, and facilitate deep imaging of large volumes (e.g., using confocal or light-sheet microscopy)[221–223]. Cleared tissues are compatible with endogenous fluorophores, including commonly used markers such as GFP[157,171,218], eYFP[171], and tdTomato[218]. However, some fluorescent signals, such as those from mTurquoise[133], mNeonGreen, and mRuby2, can deteriorate in chemical clearing reagents. To visualize these reporters, we suggest using optical clearing reagents such as refractive index-matching solution (RIMS)[218] or ScaleSQ[170] **(*Figures 2.3a,c,d, 2.4b–d,*** and ***2.5e*)** or commercially available mounting media (Step 50) **(*Figure 2.5a,c*)**. Some fluorescent proteins are sensitive to photobleaching. For example, mRuby2 may bleach over long imaging sessions or at high magnification; tdTomato exhibits similar spectral properties and may be a more suitable alternative, given its photostability[224]. Also, note that auto-fluorescent lipofuscin accumulates in aging post-mitotic tissues (e.g., the brain and heart)[225] and may interfere with examination of transduced cells; in this case, either reduce auto-fluorescence using histological methods[218,226] or, if possible, inject younger adults (≤8 weeks old) and determine the minimum time required for transgene expression.

2.3 MATERIALS

Biological materials

Caution: To address the issue of cell line misidentification and cross-contamination, it is recommended that cell lines be regularly checked to ensure they are authentic and are not infected with mycoplasma.

● Plasmids, supplied as bacterial stabs (Addgene; see ***Table 2.T1*** and ***Table 2.T2*** for plasmids used in this and related work)

Critical: Three plasmids (pAAV, pUCmini-iCAP-PHP, and pHelper) are required for transfection (***Figure 2.1***). The pHelper plasmid is available in Agilent's AAV helper-free kit (Agilent, cat. no. 240071).

● Human embryonic kidney (HEK) cells (293 or 293T; ATCC, cat. no. CRL 1573 or CRL 3216, respectively)

Caution: HEK cells pose a moderate risk to laboratory workers and the surrounding environment and must be handled according to governmental and institutional regulations. Experiments involving HEK cells were performed using biosafety level 2 practices as required by the California Institute of Technology and the US Centers for Disease Control and Prevention.

Critical: HEK293 and HEK293T cells constitutively express two adenoviral genes, E1a and E1b, which are required for AAV production in these cells[180]; we do not recommend using an alternative producer cell line with this protocol.

● Plasmid DNA containing the target sequence to be amplified during AAV titration; used for preparing the DNA standard stock

Critical: The plasmid used to make the DNA standard must contain the same target sequence as the pAAV plasmid used to generate virus. The target sequence must be within the rAAV genome; we typically amplify a portion of the WPRE or hGH polyA (see ***Figure 2.6*** caption for abbreviations and '*Reagents*' for primer sequences).

● Animals to be injected. Wild-type mouse strains used in this work include C57BL/6J (Jackson Laboratory, stock no. 000664), C57BL/6NCrl (Charles River Laboratories, strain code 027), FVB/NCrl (Charles River Laboratories, strain code 207), 129S1/SvImJ (Jackson Laboratory, stock no. 002448), and 129T2/SvEmsJ (Jackson Laboratory, stock no. 002065). Cre driver lines include ChAT-IRES-Cre (Jackson Laboratory, stock no. 028861, heterozygous), TH-IRES-Cre (European Mutant Mouse Archive, stock no. EM00254, heterozygous), and TRPV1-IRES-Cre mice (Jackson Laboratory, stock no.

017769, homozygous). Fischer rats (Charles River Laboratories, strain code 002) and Long-Evans rats (Charles River Laboratories, strain code 006) were used in ***Figure 2.5***. All rats were 4–6 weeks old at the time of AAV administration; mice were 6–10 weeks old. Refer to the '*Experimental design*' section, ***Figure 2.5***, and ***Supplementary Table 2.ST1*** for species and strain considerations.

Caution: Experiments on vertebrates must conform to all relevant governmental and institutional regulations. Animal husbandry and experimental procedures involving mice and rats were approved by the Institutional Animal Care and Use Committee (IACUC) and the Office of Laboratory Animal Resources at the California Institute of Technology.

● For molecular cloning: Recombination-deficient Escherichia coli strains such as NEB stable (New England Biolabs, cat. no. C3040H), Stbl3 (Invitrogen, cat. no. C737303), or SURE 2 competent cells (Agilent, cat. no. 200152)

Reagents

Plasmid DNA preparation

● Agarose (Amresco, cat. no. N605-250G)

● Antibiotics (e.g., carbenicillin disodium salt; Alfa Aesar, cat. no. J61949-06; all plasmids used in this work carry antibiotic resistance genes to ampicillin/carbenicillin)

● DNA ladder (100 bp–10 kb; New England Biolabs, cat. no. N0550S)

● Lysogeny broth (LB; Amresco, cat. no. J106-1KG)

Critical: For large-scale plasmid preparations, such as maxi and giga preps, we typically use Plasmid+ media (Thomson Instrument, cat. no. 446300), an enriched medium formulated to support higher cell densities and plasmid yields, as compared to those of LB.

● LB with agar (Sigma-Aldrich, cat. no. L3147-1KG)

● NucleoBond Xtra Maxi endotoxin-free (EF) plasmid purification kit (Macherey-Nagel, cat. no. 740424.50)

Critical: Triple transient transfection requires large amounts of pUCmini-iCAP-PHP (22.8 µg/dish) and pHelper plasmid DNA (11.4 µg/dish) (***Supplementary Table 2.ST2***, '*Detailed calculations*' sheet); isolating these plasmids may be more convenient with a giga-scale purification kit (NucleoBond PC 10000 EF; Macherey-Nagel, cat. no. 740548). All plasmids should be purified under endotoxin-free conditions. Endotoxin contamination in plasmid preparations can reduce transfection efficiency, and contaminating endotoxins in viral preparations could elicit immune reactions in mammals *in vivo*.

● Restriction enzymes, including SmaI (New England Biolabs, cat. no. R0141S); used for verifying plasmid and ITR integrity

● Sequencing primers (Integrated DNA Technologies); used for verifying plasmid sequence integrity

● SYBR Safe DNA gel stain (Invitrogen, cat. no. S33102)

● Tris-acetate-EDTA (TAE) buffer (50×; Invitrogen, cat. no. B49)

Cell culture

● DMEM (high glucose, GlutaMAX supplement, pyruvate; Gibco, cat. no. 10569-044)

● 70% (vol/vol) Ethanol (prepare from absolute ethanol; J.T. Baker, cat. no. 8025)

Caution: Ethanol is flammable.

● FBS (GE Healthcare, cat. no. SH30070.03)

Critical: Divide into aliquots and store at −20 °C for up to 1 year. Avoid freeze–thaw cycles.

● MEM non-essential amino acids (NEAA) solution (100×; Gibco, cat. no. 11140-050)

● Penicillin–streptomycin (pen–strep; 5,000 U/ml; Gibco, cat. no. 15070-063)

Critical: Divide into aliquots and store at −20 °C for up to 1 year. Avoid freeze–thaw cycles.

● TrypLE Express enzyme (1×; phenol red; Gibco, cat. no. 12605-036)

Transfection

● Polyethylenimine (PEI), linear, molecular weight (MW) 25,000 (Polysciences, cat. no. 23966-1)

Critical: Compared to other commonly used transfection reagents (e.g., Lipofectamine or calcium phosphate), PEI is less expensive, given the scale of transfection, and produces high viral yields ($\geq 1 \times 10^{12}$ vg/dish), which are needed for systemic administration.

● Water for injection (WFI) for cell culture (Gibco, cat. no. A1287304)

● Dulbecco's PBS (DPBS; 1×; no calcium, no magnesium; Gibco, cat. no. 14190-250)

● 1 N Hydrochloric acid (HCl) solution (suitable for cell culture; Sigma-Aldrich, cat. no. H9892)

Caution: HCl is corrosive. Use personal protective equipment.

AAV production

● 10% (vol/vol) Bleach (prepare fresh from concentrated liquid bleach (e.g., Clorox))
Critical: AAV contaminated equipment, surfaces, and labware must be disinfected for 10 min with fresh 10% (vol/vol) bleach; ethanol is not an effective disinfectant against non-enveloped viruses. AAV waste disposal should be conducted according to federal, state, and local regulations.

● Dry ice; optional

● KCl (Macron Fine Chemicals, cat. no. 6858-06)

● MgCl2·6H2O (Macron Fine Chemicals, cat. no. 5958-04)

● Sodium chloride (NaCl; Millipore, cat. no. SX0420-3)

● OptiPrep (60% (wt/vol) iodixanol) density gradient medium (Cosmo Bio USA, cat. no. AXS-1114542-5)

● Phenol red solution (Millipore, cat. no. 1072420100)

● Pluronic F-68 nonionic surfactant (10% (vol/vol) solution; Gibco, cat. no. 24040-032); optional

● Polyethylene glycol (PEG; MW 8,000; Sigma-Aldrich, cat. no. 89510-1KG-F)

● Salt-active nuclease (SAN; 25 U/µl; ArcticZymes, cat. no. 70910-202)

● Tris, ultrapure (MP Biomedicals, cat. no. 819620)

● UltraPure DNase/RNase-free distilled water (Invitrogen, cat. no. 10977-023)

● WFI for cell culture (Gibco, cat. no. A1287304)

● DPBS (1×; no calcium, no magnesium; Gibco, cat. no. 14190-250)

AAV titration

● CaCl2 (anhydrous; J.T. Baker, cat. no. 1311-01)

● DNase I recombinant (RNase-free; 10 U/µl; Roche Diagnostics, cat. no. 4716728001)

● MgCl2·6H2O (Macron Fine Chemicals, cat. no. 5958-04)

● NaCl (Millipore, cat. no. SX0420-3)

● N-lauroylsarcosine sodium salt (Sigma-Aldrich, cat. no. L9150-50G)

● Primers corresponding to the target sequence to be amplified during qPCR (Integrated DNA Technologies) WPRE-forward: GGCTGTTGGGCACTGACAAT WPRE-reverse: CCGAAGGGACGTAGCAGAAG hGH polyA-forward: GTGCCCACCAGCCTTGTC hGH polyA-reverse: TGTCTTCCCAACTTGCCCCTT

Critical: The proximity of the primer binding sites to the ITRs can affect titering results; do not use primers corresponding to the ITRs. Note that titers measured with different primers or across laboratories may not be directly comparable.

● Proteinase K (recombinant, PCR grade; 50 U/ml (2.5 U/mg); Roche Diagnostics, cat. no. 03115828001)

● Qubit dsDNA HS Assay Kit (Invitrogen, cat. no. Q32854)

● ScaI-HF restriction enzyme (New England Biolabs, cat. no. R3122S) or other enzyme that cuts outside of the rAAV genome and within the pAAV backbone

● SYBR Green master mix (Roche Diagnostics, cat. no. 04913850001)

● Tris, ultrapure (MP Biomedicals, cat. no. 819620)

● UltraPure DNase/RNase-free distilled water (Invitrogen, cat. no. 10977-023)

● UltraPure EDTA (0.5 M, pH 8.0; Invitrogen, cat. no. 15575-020)

● UltraPure Tris-HCl (1 M, pH 7.5; Invitrogen, cat. no. 15567-027)

Intravenous (retro-orbital) injection

● 10% (vol/vol) Bleach, prepared fresh, or equivalent disinfectant (e.g., Accel TB surface cleaner; Health Care Logistics, cat. no. 18692)

● Isoflurane, USP (Piramal Critical Care, cat. no. 66794-017-25)

Caution: Isoflurane is a halogenated anesthetic gas associated with adverse health outcomes in humans and must be handled according to governmental and institutional regulations. To reduce the risk of occupational exposure during rodent anesthesia, waste gas was collected in a biosafety cabinet using a charcoal scavenging system as approved by the California Institute of Technology.

● Proparacaine hydrochloride ophthalmic solution, USP (0.5% (wt/vol); Akorn Pharmaceuticals, cat. no. 17478-263-12)

● DPBS (1×; no calcium, no magnesium; Gibco, cat. no. 14190-250)

Equipment

Plasmid DNA preparation equipment

● Centrifuge (Beckman Coulter, model no. Allegra X-15R)

● Gel electrophoresis system (Bio-Rad horizontal electrophoresis system)

● Gel-imaging system (Bio-Rad, Gel Doc EZ system)

● Incubating shaker (Eppendorf, model no. I24)

● Incubator (Thermo Fisher Scientific, Heratherm model) or 37 °C warm room

● Sequence-editing and annotation software (e.g., Lasergene by DNASTAR (https://www.dnastar.com/ software/lasergene/), SnapGene by GSL Biotech

(http://www.snapgene.com/), or Vector NTI by Thermo Fisher Scientific (https://www.thermofisher.com/us/en/home/life-science/cloning/vector-ntisoftware.html))

● Spectrophotometer (Thermo Fisher Scientific, NanoDrop model)

Plasmid DNA preparation supplies

● Petri dishes (100 mm × 15 mm; Corning, cat. no. 351029)

● Test tubes (14 ml; Corning, cat. no. 352059)

● Ultra Yield flasks and AirOtop seals (250 ml; Thomson Instrument Company, cat. nos. 931144 and 899423, respectively); use with Plasmid+ media. Alternatively, use LB and standard Erlenmeyer flasks.

AAV production equipment

● Biological safety cabinet

Caution: HEK293T cells and AAVs are biohazardous materials and must be handled according to governmental and institutional regulations. All experiments involving the aforementioned materials were performed in a Class II biosafety cabinet with annual certification as required by the California Institute of Technology and the US Centers for Disease Control and Prevention.

● Centrifuge that can reach speeds up to 4,000g, refrigerate to 4 °C, and accommodate 250-ml conical centrifuge tubes (Beckman Coulter, model no. Allegra X-15R)

● Fluorescence microscope for cell culture (Zeiss, model no. Axio Vert A1)

● Incubator for cell culture (humidified at 37 °C with 5% CO2; Thermo Fisher Scientific, model no. Heracell 240i)

● Laboratory balance (with a readability of 5–10 mg)

● Support stand with rod and clamp (VWR International, cat. nos. 12985-070, 60079-534, and 89202-624, respectively) (***Figure 2.8f***)

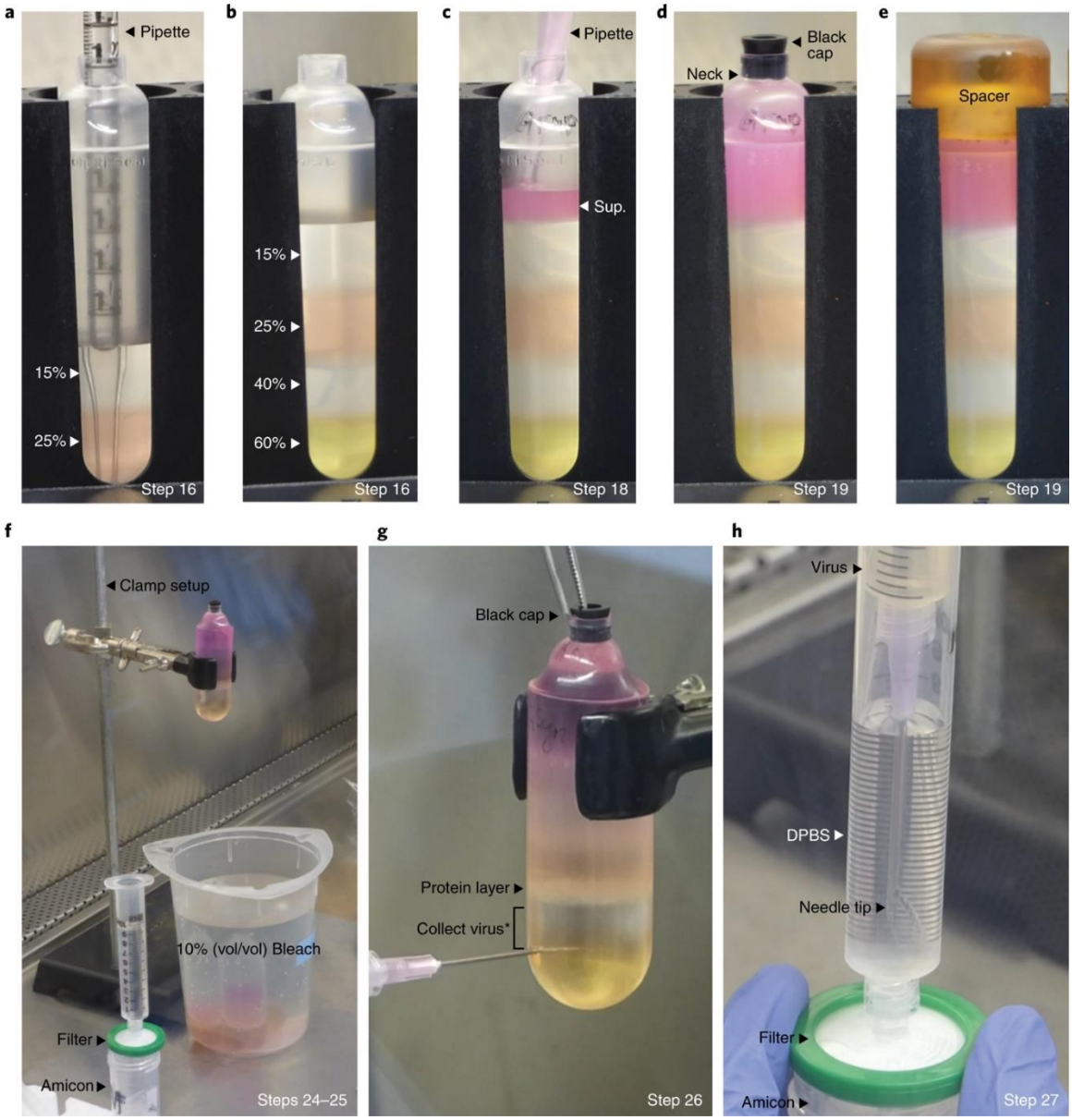


Figure 2.8: AAV purification procedure.

(a,b), In Step 16, pipette the iodixanol density gradients (***Supplementary Video 2.V1***, 0:00–1:45, or ***Supplementary Video 2.V2***, 0:00–1:13). **(a),** Layer the 25% (wt/vol) iodixanol

underneath the 15% layer. **(b),** Add layers of increasing density under the previous layer; the gradients should have a sharp delineation between layers. (**c),** In Step 18, load the supernatant (Sup.) from Step 17 (***Figure 2.7b***) above the 15% layer (***Supplementary Video 2.V1***, 1:46–2:22; the same step is also shown in ***Supplementary Video 2.V2***, 1:14–1:55). **(d,e),** In Step 19, fill each tube up to the neck with SAN digestion buffer and insert a black cap **(d)**; place a spacer on top before weighing the tubes **(e)**. **(f),** After ultracentrifugation (Step 22), secure the tube into the clamp setup above a container of fresh 10% (vol/vol) bleach (Step 24). Allow 10 ml of DPBS to begin dripping through the syringe filter unit into an Amicon filter device (Step 25). **(g)**, In Step 26, collect the virus (***Supplementary Video 2.V3***, 0:00–1:30). Insert the needle ~4 mm below the 40/60% interface (i.e., where the tube just starts to curve). Do not collect virus (asterisk) until the black cap is removed; do not collect from the white protein layer at the 25/40% interface. **(h)**, In Step 27, filter the virus/iodixanol (***Supplementary Video 2.V3***, 1:31–2:32). Inject the virus below the DPBS in the filter-attached syringe barrel before pushing the virus/DPBS through the syringe filter unit and into the Amicon filter device.

● Ultracentrifuge (preparative ultracentrifuge for in vitro diagnostic use; Beckman Coulter, model no. Optima XE-90, with a Type 70Ti fixed-angle rotor)

Caution: During ultracentrifugation, rotors are subjected to enormous forces (350,000g in this protocol). Rotor failure can have catastrophic consequences, including irreparable damage to the centrifuge and laboratory and fatal injuries to personnel. Inspect the rotors for signs of damage or weakness before each use, and always follow the manufacturer's instructions while operating an ultracentrifuge.

● Water bath (Fisher Scientific, Isotemp model)

AAV production supplies

● Amicon Ultra-15 centrifugal filter devices (100-kDa molecular weight cutoff; Millipore, cat. no. UFC910024)

● Barrier pipette tips (low binding, 1,000 µl; Genesee Scientific, cat. no. 23-430)

● Cell scrapers (25-cm handle × 3-cm blade; Corning, cat. no. 353089)

● Centrifuge tube racks for 250-ml tubes, 6-well (Universal Medical, cat. no. HS23224) or empty beakers

● Conical centrifuge tubes (50 ml, 250 ml, and 500 ml (optional); Corning, cat. nos. 352098, 430776, and 431123, respectively)

● Costar Spin-X centrifuge tube filters (Corning, cat. no. 07-200-385); optional

● Empty, sterile media bottles

● OptiSeal tubes (Beckman Coulter, cat. no. 361625); includes black caps

● OptiSeal tube kit (Beckman Coulter, cat. no. 361662); includes a tube rack, spacers, and spacer- and tube-removal tools

● Pipet-Aid XL portable pipetting device (Drummond Scientific, cat. no. 4-000-105) *Critical:* Use a pipetting device with precise control to pour the density gradients (Step 16) and load the virus (Step 18).

● pH indicator strips (Millipore, cat. nos. 109532 and 109584)

● Screw-cap vials (1.6 ml; National Scientific Supply, cat. no. BC16NA-PS)

● Serological pipettes (2 ml, 5 ml, 10 ml, 25 ml, and 50 ml; Corning, cat. no. 356507, and Genesee Scientific, cat. nos. 12-102, 12-104, 12-106, and 12-107, respectively)

Critical: Corning brand 2-ml serological pipettes consistently fit into OptiSeal tubes while pouring the density gradients (Step 16A) and loading the virus (Step 18); other brands should be tested before use.

● Stericup sterile vacuum filtration system (0.22 μm; 1 liter; Millipore, cat. no. S2GPU11RE)

● Sterile bottles (500 ml; VWR International, cat. no. 89166-106)

● Syringes (5 ml and 10 ml; BD, cat. nos. 309646 and 309604, respectively)

● Syringe filter units (0.22 μm; Millipore, cat. no. SLGP033RS)

● Tissue culture dishes (150 mm × 25 mm; Corning, cat. no. 430599)

● Tubing, e.g., polytetrafluoroethylene (PTFE) standard tubing (2 mm i.d. × 3 mm o.d.; Fluorostore) and Tygon tubing (2 mm i.d. × 4 mm o.d.; United States Plastics, cat. no. 57658); optional

Critical: Ensure that the PTFE tubing fits on the tip of a 5-ml serological pipette and into an Optiseal tube before pouring the density gradients (Step 16B). Use the Tygon tubing to secure the PTFE tubing at the pipette tip.

● 16-gauge × 1 1/2 inch needles (BD, cat. no. 305198)

AAV titration equipment

● Centrifuge (Eppendorf, model no. 5418)

● Dry bath and heating blocks (Fisher Scientific, Isotemp models)

● PCR plate spinner (VWR International, cat. no. 89184) or centrifuge equipped with plate adapters

● Quantitative PCR machine (Analytik Jena, model no. qTOWER 2.2)

● Qubit 3.0 fluorometer (Invitrogen, cat. no. Q33216)

AAV titration supplies

● Barrier pipette tips (low binding; 10 μl, 20 μl, 200 μl, and 1,000 μl; Genesee Scientific, cat. nos. 23-401, 23-404, 23-412, and 23-430, respectively)

● DNA Clean & Concentrator kit (Zymo Research, cat. no. D4033 (DCC-25)), for purification of up to 25 µg of the DNA standard

● Microcentrifuge tubes (1.5-ml DNA/RNA LoBind; Eppendorf, cat. no. 86-923)

● Qubit assay tubes (Invitrogen, cat. no. Q32856)

● Sealing film for 96-well PCR plates (Genesee Scientific, cat. no. 12-529)

● Stericup sterile vacuum filtration system (0.22 µm, 250 ml; Millipore, cat. no. SCGPU02RE)

● Sterile bottles (250 ml; VWR International, cat. no. 89166-104)

● 96-well PCR plates (Genesee Scientific, cat. no. 24-310W)

Intravenous (retro-orbital) injection equipment

● Animal anesthesia system (VetEquip, cat. no. 901806, 901807, or 901810)

Critical: Most animal facilities provide anesthesia systems equipped with an induction chamber, isoflurane vaporizer, nose cone, and waste gas scavenging system.

Intravenous (retro-orbital) injection supplies

● Activated charcoal adsorption filters (VetEquip, cat. no. 931401)

● Insulin syringes with permanently attached needles (31 gauge × 5/16 inches; BD, cat. no. 328438)

● Oxygen gas supply (Airgas)

● Screw-cap vials (1.6 ml; National Scientific Supply, cat. no. BC16NA-PS)

Reagent setup

Critical: All reagents used for viral production and administration should be prepared using endotoxin-free materials. Glassware is not endotoxin-free, and autoclaving does not eliminate endotoxins. To minimize contamination, we dissolve chemicals in sterile bottles by shaking and/or heating to mix, use demarcations on bottles to bring solutions to the final volume, and use pH strips rather than a pH meter. When filter-sterilizing solutions, do so in a biosafety cabinet.

Plasmid DNA

Grow bacterial stocks in LB or Plasmid+ media containing the appropriate selective antibiotic; refer to the Addgene catalog for suggested growth conditions. Use a large-scale endotoxin-free plasmid purification kit to isolate plasmids; elute plasmid DNA with the supplied Tris-EDTA (TE) buffer. Measure the DNA purity and concentration using a spectrophotometer and freeze at −20 °C or −80 °C for up to several years.

Critical: Always verify the integrity of purified plasmids by Sanger sequencing (using a DNA sequencing facility) and restriction digestion (https://www.neb.com/tools-a nd-resources) before proceeding with downstream applications. pAAV plasmids contain ITRs ***(Figure 2.6)*** that are prone to recombination in E. coli. pAAVs should be propagated in recombination-deficient strains such as NEB Stable, Stbl3, or SURE 2 competent cells to prevent unwanted recombination. After purification, pAAVs should be digested with SmaI to confirm the presence of ITRs, which are required for replication and encapsidation of the viral genome; use sequence-editing and annotation software to determine the expected band sizes. Note that it is difficult to sequence through the secondary structure of ITRs[227]; avoid ITRs when designing sequencing primers.

Critical: Create bacterial glycerol stocks and store at –80 °C for up to several years.

Cell culture media

Add 25 ml of FBS, 5 ml of NEAA, and 5 ml of pen–strep to a 500-ml bottle of DMEM. Invert to mix and store at 4 °C for up to several months; warm to 37 °C before use. The resulting cell culture media should have a final concentration of 5% (vol/vol) FBS, 1× NEAA, and 50 U/ml pen–strep.

Critical: To quickly expand cells for large viral preps, consider using a final concentration of 10% (vol/vol) FBS in the cell culture media; see guidelines on cell culture below.

Cell culture

Thaw HEK293T cells according to the manufacturer's recommendations. Passage cells using either TrypLE Express enzyme or a standard trypsinization protocol for adherent cultures[228]. Seed cells in 150-mm tissue culture dishes with a final volume of 20 ml of media per dish. Maintain in a cell culture incubator at 37 °C with 5% CO2.

Critical: We suggest passaging cells at a ratio of 1:3 (i.e., divide one dish of cells into three new dishes of cells) every other day when expanding cells for viral production; split cells at a 1:2 ratio (or 6×10^4 cells/cm^2) 24 h before transfection. Use higher split ratios if using 10% (vol/vol) FBS. Always use sterile technique.

Critical: Follow the manufacterer's recommendations to create frozen stocks of HEK cells.

PEI stock solution

Pipette 50 ml of WFI water into a 50-ml conical centrifuge tube for later use. Add 323 mg of PEI to the remaining 950 ml bottle of WFI water and adjust the pH to 2–3 by adding 1 N HCl suitable for cell culture, keeping track of the volume of HCl added. Heat in a 37 °C water bath for several hours (or overnight) and occasionally shake to mix. Once dissolved, add reserved WFI water to a total volume of 1 liter. Filter-sterilize, make aliquots

in 50-ml conical centrifuge tubes, and store at −20 °C for up to 1 year. We routinely freeze–thaw our PEI aliquots.

<u>*Critical:*</u> Both our PEI stock solution recipe and PEI calculations (***Supplementary Table 2.ST2***, '*Detailed calculations*' sheet) are based on ref.[178]. We adjust the pH to 2–3 so that PEI dissolves in water. The designated pH range does not appear to adversely affect cell viability, transfection efficiency, or viral titers. The transfection solution, created by mixing the PEI + DPBS master mix and DNA + DPBS solution (Step 2 and ***Supplementary Table 2.ST2***), has a final pH of 6.5–7.0. To transfect one dish, 2 ml of transfection solution is added to 20 ml of media (Step 2), which further dilutes the PEI.

PEI + DPBS master mix

Thaw PEI in a 37 °C water bath. Bring the PEI to room temperature (RT; 23 °C) and vortex to mix. Add PEI and DPBS to a 50-ml conical centrifuge tube and vortex again to mix. Use ***Supplementary Table 2.ST2*** ('*Transfection calculator*' sheet) to calculate the volumes of PEI (cell I9) and DPBS (cell J9) needed.

<u>*Critical:*</u> Prepare fresh master mix before use.

DNA + DPBS

Bring plasmid DNA to RT and briefly vortex to mix. For each viral prep, add DNA and DPBS to a 50-ml conical centrifuge tube; the solution is vortexed in Step 2. Use ***Supplementary Table 2.ST2*** ('*Transfection calculator*' sheet) to calculate the quantities of DNA (e.g., cells E9, E11, and E13) and DPBS (e.g., cell F9) needed.

<u>*Critical:*</u> Prepare fresh DNA + DPBS solution before use. Re-measure plasmid DNA concentrations immediately before use; multiple freeze–thaw cycles may cause DNA degradation.

SAN digestion buffer

Add 29.22 g of NaCl, 4.85 g of Tris base, and 2.03 g of MgCl2·6H2O to a 1-liter bottle of WFI water and shake to mix. Filter-sterilize and store at RT for up to several months. The resulting SAN digestion buffer should have a final pH of ~10.0 and a final concentration of 500 mM NaCl, 40 mM Tris base, and 10 mM MgCl2.

SAN + SAN digestion buffer

Add 100 U of SAN (4 µl of 25 U/µl SAN) per milliliter of SAN digestion buffer; pipette to mix.

Critical: Prepare fresh solution before use.

40% (wt/vol) PEG stock solution

Decant ~500 ml of WFI water into a 500-ml sterile bottle for later use. Add 146.1 g of NaCl to the remaining 500 ml (in the 1-liter bottle of WFI water) and shake/heat until dissolved. Once completely dissolved, add 400 g of PEG and heat at 37 °C overnight for up to 2 nights. Add reserved WFI water to a total volume of 1 liter. Filter-sterilize and store at RT for up to several months. The resulting stock solution should have a final concentration of 2.5 M NaCl and 40% (wt/vol) PEG.

Critical: Prepare in advance. To expedite the procedure, heat the solution at 65 °C until the PEG is dissolved. The solution will appear turbid, but no flecks of PEG should remain; the mixture will become clear upon cooling.

Critical: Pre-wet the entire filter surface with a minimal volume of water before adding the solution. This solution is extremely viscous and will take 1–2 h to filter.

DPBS + high salt

Add 29.22 g of NaCl, 93.2 mg of KCl, and 101.7 mg of MgCl2·6H2O to a 500-ml bottle of DPBS and shake to mix. Filter-sterilize and store at RT for up to several months. The resulting buffer should have a final concentration of 1 M NaCl, 2.5 mM KCl, and 1 mM MgCl2 (in addition to the salts in the DPBS).

DPBS + low salt

Add 2.92 g of NaCl, 93.2 mg of KCl, and 101.7 mg of MgCl2·6H2O to a 500-ml bottle of DPBS and shake to mix. Filter-sterilize and store at RT for up to several months. The resulting buffer should have a final concentration of 100 mM NaCl, 2.5 mM KCl, and 1 mM MgCl2 (in addition to the salts in the DPBS).

Iodixanol density gradient solutions (15%, 25%, 40%, and 60% (wt/vol) iodixanol)

For each layer, add iodixanol (OptiPrep), DPBS + high salt or DPBS + low salt, and phenol red (if applicable) to a 50-ml conical centrifuge tube. Invert or briefly vortex to mix. Use ***Supplementary Table 2.ST3*** to determine the volumes of each reagent needed. The 25% and 60% layers contain phenol red, which turns the solutions red and yellow, respectively, and facilitates clear demarcation of the gradient boundaries (***Figure 2.8***).

Critical: Prepare fresh solutions on the day of AAV purification. Alternatively, prepare up to 1 d in advance; store at RT and protect from light. Do not pour the density gradients until Step 16.

Critical: In Step 16B, prepare more iodixanol solutions than are needed. For six or fewer gradients, prepare enough of each solution to pour an extra gradient. For eight gradients, prepare enough of each solution to pour two extra gradients. The extra solution is needed to fill the 5-ml pipette and prevent an air bubble from disturbing the gradient when releasing the last of the required volume.

DNase digestion buffer

Use a 50-ml serological pipette to measure 247.5 ml of UltraPure water into a 250-ml sterile bottle. Add 55.5 mg of CaCl2, 2.5 ml of 1 M Tris-HCl, and 508 mg of MgCl2·6H2O and shake to mix. Filter-sterilize and store at RT for up to several months. The resulting buffer should have a final concentration of 2 mM CaCl2, 10 mM Tris-HCl, and 10 mM MgCl2.

DNase I + DNase digestion buffer

Add 50 U of DNase I per milliliter of digestion buffer (a 1:200 dilution of 10 U/µl DNase); pipette to mix.

Critical: Prepare fresh solution before use.

Proteinase K solution

Use a 50-ml serological pipette to measure 250 ml of UltraPure water into a 250-ml sterile bottle. Add 14.61 g of NaCl and shake to mix. Add 2.5 g of N-lauroylsarcosine sodium salt to the mixture and gently swirl to mix; N-lauroylsarcosine sodium salt is a surfactant and will generate bubbles during vigorous mixing. Filter-sterilize and store at RT for up to several months. The resulting solution should have a final concentration of 1 M NaCl and 1% (wt/vol) N-lauroylsarcosine sodium salt.

Proteinase K + proteinase K solution

Add 100 µg of proteinase K per milliliter of solution (a 1:200 dilution of 50 U/ml (2.5 U/mg) proteinase K); pipette to mix.

Critical: Prepare fresh solution before use.

DNA standard stock

Set up a single 50-µl restriction digest reaction; use 60–80 U (3–4 µl) of ScaI (or another suitable enzyme) to linearize 20 µg of the plasmid DNA containing the target sequence. Run a small amount of the reaction on an agarose gel to ensure complete digestion. Purify the reaction using two DNA clean-up columns. Measure the DNA concentration (ng/µl) using a spectrophotometer. Dilute to ~5–10 × 10^9 single-stranded (ss) DNA molecules/µl and use the Qubit assay to verify the concentration (ng/µl). Divide into 20-µl aliquots in DNA/RNA LoBind microcentrifuge tubes and freeze at −20 °C for up to 1 year.

Critical: Before preparing the standard, use sequence-editing and annotation software to confirm that the plasmid contains a single ScaI site in the ampicillin resistance gene.

Critical: Refer to ref.[184] and use ***Supplementary Table 2.ST4*** (cells B7–10) to calculate the number of ssDNA molecules in a given plasmid (cell B13). We typically use linearized pAAV-CAG-eYFP diluted to 10 ng/µl, which corresponds to 6.6×10^9 ssDNA molecules/µl (***Supplementary Table 2.ST4***, '*Example*' sheet).

DNA standard dilutions

Prepare three sets of eight (1:10) serial dilutions of the DNA standard stock. For each set, begin by pipetting 5 µl of the standard into 45 µl of UltraPure water (standard no. 8). Mix by vortexing for 3 s and proceed with the seven remaining dilutions (standard no. 7 to standard no. 1). The final concentrations of the standard dilutions should range from 5–10 × 10^8 (standard no. 8) to 5–10 × 10^1 (standard no. 1) ssDNA molecules/µl.

Critical: Prepare fresh solutions in DNA/RNA LoBind microcentrifuge tubes immediately prior to use; at low concentrations, the linearized DNA is prone to degradation and/or sticking to the walls of the tube[184]. One 20-µl aliquot of the DNA standard stock will provide enough DNA for preparing the dilutions and verifying the concentration via the Qubit assay before qPCR.

qPCR master mix

Prepare a qPCR master mix for the total number of reactions (i.e., wells) needed. One reaction requires 12.5 µl of SYBR Green master mix, 9.5 µl of UltraPure water, and 0.5 µl of each primer (from a 2.5-µM stock concentration), for a total of 23 µl/well. Pipette or vortex for 1–2 s to mix.

Critical: Prepare fresh solution before use.

Equipment setup

Clamp setup for AAV purification

Attach the rod to the support stand. Secure the clamp 25–30 cm above the stand (***Figure 2.8f***).

Anesthesia system

Place the induction chamber, nose cone, and waste gas scavenging system (e.g., activated charcoal adsorption filters) inside a biosafety cabinet. We recommend using an anesthesia system in which the isoflurane vaporizer and oxygen supply remain outside of the cabinet workspace. Connect the associated tubing such that the input is from the vaporizer/oxygen supply and the output is to the charcoal scavenging device[229].

2.4 PROCEDURE

Caution: AAVs are biohazardous materials and must be handled according to governmental and institutional regulations. Experiments involving AAVs were performed using biosafety level 2 practices as required by the California Institute of Technology and the US Centers for Disease Control and Prevention.

Critical: The entire procedure spans 8 d, excluding pause points and the time required to evaluate transgene expression **(*Figure 2.7a*)**. There are no pause points between days 1 and

5, until Step 11; once cells have been transfected, AAVs are harvested on days 3 and 5. Plan accordingly during this time window.

Triple transient transfection of HEK293T cells

● Timing 1–2 h

Critical: For capsids that package well (i.e., AAV9, AAV-PHP.B, AAV-PHP.eB, and AAV-PHP.S), the PEI transfection protocol typically yields ≥1×10^{12} vg per 150-mm dish, as measured post purification[133,157]. Before starting the protocol, determine the number of dishes needed per viral prep and use ***Supplementary Table 2.ST2*** (‘*Transfection calculator*’ sheet) to calculate the quantities of PEI, DPBS, and plasmid DNA required for transfection. Skip to Step 43 if custom AAVs were obtained elsewhere.

1. 24 h before transfection, seed HEK293T cells in 150-mm dishes to attain 80–90% confluency the next day[228]. Incubate the cells in a cell culture incubator at 37 °C with 5% CO2.

2. At the time of transfection, make the PEI + DPBS master mix and the DNA + DPBS solution for each viral prep (*Reagent setup* and ***Supplementary Table 2.ST2***, ‘*Transfection calculator*’ sheet). Using a 5- or 10-ml serological pipette, add the required volume of the PEI + DPBS master mix (e.g., ‘*Transfection calculator*’ cell G9) dropwise to the DNA + DPBS solution (e.g., ‘*Transfection calculator*’ cells E9 + E11 + E13 + F9) while gently vortexing to mix. Cap the tube and thoroughly vortex for 10 s to mix. Allow the mixture to sit at RT for 2–10 min. Add 2 ml of the transfection solution dropwise to each dish and swirl to mix before returning the dishes to the cell culture incubator.

Critical step: We use a pAAV:pUCmini-iCAP-PHP:pHelper plasmid ratio of 1:4:2 based on micrograms of DNA. We use 40 µg of total DNA per 150-mm dish (5.7 µg of pAAV, 22.8 µg of pUCmini-iCAP-PHP, and 11.4 µg of pHelper) (***Supplementary Table 2.ST2***, ‘*Detailed calculations*’ sheet). The plasmid ratio was optimized during the initial development of the AAV production protocol; 1:4:2 gave the best viral yields.

Critical step: The transfection solution will appear slightly cloudy because of the formation of DNA–PEI complexes[178,179]. Allowing the mixture to sit for >10 min may reduce transfection efficiency.

Critical step: Users can opt to run a positive transfection/virus production control (e.g., pAAV-CAG-eYFP); this is especially important if using an untested rAAV genome.

Refer to ***Table 2.T3*** for troubleshooting tips.

3. Change the media 12–24 h post transfection by aspirating the old media in 10% (vol/vol) bleach and replacing it with 20 ml of fresh, warmed media (*Reagent setup*).

Critical step: Do not allow the cells to remain without media for more than a few minutes. To protect the cells from unwanted stress, aspirate the media from five plates at a time and promptly replace it with new media. PEI is moderately cytotoxic[179] and cell death of up to 20% is common[230]. Do not allow the media to remain unchanged for more than 24 h post transfection. Failure to change the media in a timely manner will result in poor cell health and low titers.

Critical step: Beginning 72 h post transfection, examine the cells under a fluorescence microscope to assess fluorescent protein expression, if applicable. Note that expression of the rAAV genome does not necessarily correlate with final viral yield and will depend on the reporter and/or promoter under investigation.

Refer to ***Table 2.T3*** for troubleshooting tips.

AAV harvest

● Timing 5 d

Caution: rAAVs, although replication-incompetent, are potent gene-delivery vehicles and must be handled according to governmental and institutional regulations. The safety of packaged transgenes (e.g., oncogenic genes) should be carefully considered. Perform all

procedures in a certified biosafety cabinet and clean AAV-contaminated equipment, surfaces, and labware with fresh 10% (vol/vol) bleach.

Critical: Carefully label all tubes and replace gloves, pipettes, and cell scrapers between viral preps to avoid cross-contamination. Refer to ***Figure 2.7b*** for a schematic of the AAV harvest procedure.

4. Harvest the cell culture media 72 h (3 d) post transfection. Tilt each dish at a 30° angle and use a 25-ml serological pipette to collect the media. Store in an empty, sterile media bottle or sterile 500-ml bottle at 4 °C until Step 6. Replace the media with 20 ml of fresh, warmed media (*Reagent setup*).

Caution: Tilt dishes away from the front grill of the biosafety cabinet to prevent media from spilling out of the biosafety cabinet.

Critical Step: To avoid cross-contamination, harvest the media from one viral prep at a time.

Critical Step: For AAV-PHP production in HEK293T cells, the media at 72 h post transfection contains ~2×10^{11} vg per dish, or 10–20% of the expected viral yield. Failure to collect and change media at this time point will decrease yields.

Critical Step: If time is limited, media and cells can be harvested together at 72 h or 96 h rather than 120 h (Step 5), but total yields will be reduced.

5. Harvest the media and cells 120 h (5 d) post-transfection. Use a cell scraper to gently scrape the cells in the media. After scraping the first dish, prop it at a 30° angle, using an empty 1.5-ml microcentrifuge tube rack for support. Scrape down the residual cells and media such that they are pooled together. Return the dish lid and scrape the next plate; prop dishes up against one another along the length of the biosafety cabinet until scraping is complete. Use a 25-ml serological pipette to collect the media and cells from each dish;

transfer to a 250-ml conical centrifuge tube. Pool the media and cells from up to 10 dishes in a single tube.

Caution: Scrape the cells with a forward motion (i.e., away from the front grill of the biosafety cabinet) to prevent media and cells from splashing out of the biosafety cabinet. If a spill does occur at this or any other step, immediately cover with paper towels and carefully saturate the towels with fresh 10% (vol/vol) bleach.

Critical Step: To avoid cross-contamination, harvest the media and cells from one viral prep at a time.

Critical Step: For larger viral preps (6–10 dishes), a 250- or 500-ml conical centrifuge tube can be used to harvest the media and cells (Steps 5–9). However, we recommend using two 250-ml tubes in Step 10B because the PEG pellet (Step 12) is difficult to remove from the walls and edges of 500-ml tubes (Step 14).

6. Combine the media collected at 72 h post transfection (Step 4) with the media and cells collected at 120 h post transfection (Step 5). For smaller viral preps (1–5 dishes), use option A. For larger preps (6–10 dishes), use option B.

(A) Harvest from 1–5 dishes

(i) Pour the media collected in Step 4 into the corresponding 250-ml tube of media and cells collected in Step 5.

Critical Step: Save the bottles from Step 4 for Step 8.

(B) Harvest from 6–10 dishes

(i) Pour the media collected in Step 4 into a new 250-ml tube.

Critical Step: Save the bottles from Step 4 for Step 8.

7. Centrifuge the media and cells at 2,000g for 15 min at RT. Ensure that the tube caps are tightly secured. Centrifugation will result in the formation of a cell pellet (***Figure 2.7b***).

8. Pour off the supernatant (i.e., the clarified media) into the corresponding bottle from Step 4. Allow excess media to drip back down onto the beveled edge of the 250-ml tube; remove using a P1000 pipette and add to the supernatant. Store the supernatant at 4 °C until Step 10.

Critical Step: Failure to remove excess media from the pellet will cause several milliliters of media to dilute the SAN digestion buffer in Step 9.

9. Cell pellet resuspension. Prepare 5 ml of SAN + SAN digestion buffer (*Reagent setup*) per viral prep. For smaller viral preps (1–5 dishes), use option A. For larger preps (6–10 dishes), use option B.

(A) Harvest from 1–5 dishes

(i) Use a 5-ml serological pipette to gently resuspend the cell pellet in 5 ml of SAN + SAN digestion buffer; pipette into a 50-ml tube to finish resuspending the pellet (***Figure 2.7b***).

(ii) Incubate in a 37 °C water bath for 1 h and store at 4 °C until Step 14 (up to 1 d).

Critical Step: Be sure to collect the entire pellet, which will stick to the walls and beveled edges of 250-ml tubes. Save the 250-ml tubes for Step 10.

Critical Step: The high salt content of SAN digestion buffer lyses the cells, which release the viral particles and nucleic acids into the solution. Initially, the cell lysate may be viscous and difficult to pipette; SAN will degrade nucleic acids and reduce the viscosity after incubation at 37 °C. The pH of the lysate will decrease to 8–9 or lower during cell lysis, but the lysate should appear pink rather than yellow/orange because of residual phenol red (***Figure 2.7b***). Note that the expression of fluorescent proteins from strong promoters (e.g., CAG) can alter the color of the lysate.

Critical Step: (Optional) Collect a 30-µl sample from the cell lysate for troubleshooting; store at 4 °C for up to 1 week. If the viral yield is lower than expected, the sample can be titered (Steps 32–42) to determine at which stage the virus may have been lost.

Refer to ***Table 2.T3*** for troubleshooting tips.

(B) Harvest from 6–10 dishes

(i) Use a 10-ml serological pipette to partially resuspend the smaller cell pellet in 5 ml of SAN + SAN digestion buffer. Pipette into the second 250-ml tube containing the larger pellet and resuspend together; pipette into a 50-ml tube to finish resuspending the pellet (***Figure 2.7b***).

(ii) Incubate in a 37 °C water bath for 1 h and store at 4 °C until Step 14 (up to 1 d).

Critical Step: Be sure to collect the entire pellet, which will stick to the walls and beveled edges of 250-ml tubes. Save the 250-ml tubes for Step 10.

Critical Step: The high salt content of SAN digestion buffer lyses the cells, which release viral particles and nucleic acids into solution. Initially, the cell lysate may be viscous and difficult to pipette; SAN will degrade nucleic acids and reduce the viscosity after incubation at 37 °C. The pH of the lysate will decrease to 8–9 or lower during cell lysis, but the lysate should appear pink rather than yellow/orange because of residual phenol red (***Figure 2.7b***). Note that expression of fluorescent proteins from strong promoters (e.g., CAG) can alter the color of the lysate.

Critical Step: (Optional) Collect a 30-µl sample from the cell lysate for troubleshooting; store at 4 °C for up to 1 week. If the viral yield is lower than expected, the sample can be titered (Steps 32–42) to determine at which stage the virus may have been lost.

Refer to ***Table 2.T3*** for troubleshooting tips.

10. Retrieve the supernatant collected in Step 8. For smaller viral preps (1–5 dishes), use option A. For larger preps (6–10 dishes), use option B.

(A) Harvest from 1–5 dishes

(i) Pour the supernatant from Step 8 into the corresponding 250-ml tube from Step 9.

Critical Step: (Optional) Collect a 30-µl sample from the media for troubleshooting; store at 4 °C for up to 1 week. If the viral yield is lower than expected, the sample can be titered (Steps 32–42) to determine at which stage the virus may have been lost.

(B) Harvest from 6–10 dishes

(i) Equally divide the supernatant from Step 8 between the two corresponding 250-ml tubes from Step 9.

Critical Step: (Optional) Collect a 30-µl sample from the media for troubleshooting; store at 4 °C for up to 1 week. If the viral yield is lower than expected, the sample can be titered (Steps 32–42) to determine at which stage the virus may have been lost.

11. Use a 25-ml or 50-ml serological pipette to add a 1/5 final volume of 40% (wt/vol) PEG stock solution to the supernatant (i.e., the supernatant should contain a final concentration of 8% (wt/vol) PEG solution). Tighten the cap and thoroughly invert ten times to mix. Incubate on ice for 2 h.

Critical Step: For AAV production in HEK293T cells, the cell culture media contains a large fraction of the expected yield[231]. Failure to PEG-precipitate AAV particles in the media will result in lower viral yields[181].

Pause Point: The PEG–media mixture can be incubated at 4 °C overnight.

12. Centrifuge the PEG–media mixture at 4,000g for 30 min at 4 °C. Centrifugation will result in the formation of a PEG pellet (***Figure 2.7b***).

13. Pour off the supernatant (i.e., the PEG-clarified media) into a used media collection bottle for bleaching. Allow excess media to drip back down onto the beveled edge of the 250-ml tube; aspirate or pipette to remove.

14. PEG pellet resuspension. Prepare 1 ml of SAN + SAN digestion buffer (*Reagent setup*) per viral prep. For smaller viral preps (1–5 dishes), use option A. For larger preps (6–10 dishes), use option B.

(A) Harvest from 1–5 dishes

(i) Use a P1000 pipette to carefully resuspend the PEG pellet in 1 ml of SAN + SAN digestion buffer; pipette into the corresponding lysate from Step 9 (***Figure 2.7b***).

(ii) Incubate in a 37 °C water bath for an additional 30 min.

Critical Step: Resuspending the PEG pellet is difficult and will take ~5 min per pellet. Be sure to collect the entire pellet, some of which will stick to the walls and beveled edges of 250-ml tubes. During resuspension, avoid air bubbles, which can be difficult to collect with a pipette and may disrupt capsid structure. Do not use a serological pipette to resuspend the pellet, which can become lodged within the barrel of the pipette.

Critical Step: (Optional) Collect a 30-µl sample from the PEG pellet resuspension, before adding it to the corresponding lysate, for troubleshooting; store at 4 °C for up to 1 week. If the viral yield is lower than expected, the sample can be titered (Steps 32–42) to determine at which stage the virus may have been lost.

Pause Point: Store the lysate at 4 °C overnight. Alternatively, use a dry ice–ethanol bath to freeze the lysate; store at −20 °C for up to 1 week.

(B) Harvest from 6–10 dishes

(i) Use a P1000 pipette to partially resuspend one of the PEG pellets in 1 ml of SAN + SAN digestion buffer. Pipette into the second 250-ml tube containing the second pellet and

carefully resuspend together; pipette into the corresponding lysate from Step 9 (***Figure 2.7b***).

(ii) Incubate in a 37 °C water bath for an additional 30 min.

Critical Step: Resuspending the PEG pellet is difficult and will take ~5 min per pellet. Be sure to collect the entire pellet, some of which will stick to the walls and beveled edges of 250-ml tubes. During resuspension, avoid air bubbles, which can be difficult to collect with a pipette and may disrupt capsid structure. Do not use a serological pipette to resuspend the pellet, which can become lodged within the barrel of the pipette.

Critical Step: (Optional) Collect a 30-µl sample from the PEG pellet resuspension, before adding it to the corresponding lysate, for troubleshooting; store at 4 °C for up to 1 week. If the viral yield is lower than expected, the sample can be titered (Steps 32–42) to determine at which stage the virus may have been lost.

Pause Point: Store the lysate at 4 °C overnight. Alternatively, use a dry ice–ethanol bath to freeze the lysate; store at −20 °C for up to 1 week.

AAV purification

● Timing 1 d

Critical: One iodixanol density gradient is sufficient to purify virus from up to ten 150-mm dishes. If more than ten dishes per prep are used, divide the lysate into more than one gradient. The AAV purification steps are most easily learned by visualization; refer to ***Figure 2.8*** and ***Supplementary Videos 2.ST1–3*** for details.

15. Determine the number of gradients needed and prepare the iodixanol density gradient solutions (*Reagent setup* and ***Supplementary Table 2.ST3***). Set the OptiSeal tubes in the rack provided in the OptiSeal tube kit; alternatively, use the long edge of a 50-ml tube Styrofoam rack.

Caution: Check the OptiSeal tubes for defects; tubes with dents may collapse during ultracentrifugation.

16. Pour the density gradients (***Figure 2.8a,b*** and ***Supplementary Video 2.V1***, 0:00–1:45, or ***Supplementary Video 2.V2***, 0:00–1:13). Each gradient is composed of the following layers: 6 ml of 15% (wt/vol) iodixanol, 6 ml of 25% (wt/vol) iodixanol, 5 ml of 40% (wt/vol) iodixanol, and 5 ml of 60% (wt/vol) iodixanol (***Supplementary Table 2.ST3***). Pour the layers with a 2- or 5-ml serological pipette. We typically use a 2-ml pipette; using a 5-ml pipette is faster but requires the use of PTFE and Tygon tubing and extra reagents. To load the layers with a 2-ml pipette, choose option A. To load the layers with a 5-ml pipette, choose option B.

(A) Loading with a 2-ml pipette

(i) Begin by pipetting 6 ml (measure to the 3 ml mark twice) of 15% (wt/vol) iodixanol to each tube. Next, add 6 ml of 25% (wt/vol) iodixanol under the 15% layer by lightly touching the pipette tip to the bottom of the tube and slowly releasing the solution (***Figure 2.8a*** and ***Supplementary Video 2.V1***, 0:13–1:29). Continue adding layers of increasing density under the previous layer. The gradients should have a sharp delineation between layers (***Figure 2.8b***).

Critical Step: When loading the 25%, 40%, and 60% layers with a 2-ml pipette, stop releasing the solution and slowly remove the pipette once the iodixanol is ~5 mm from the tip of the pipette (***Supplementary Video 2.V1***, 0:42–0:58 and 1:20–1:25). This will prevent an air bubble from disturbing the gradient. The remaining iodixanol will be released when the pipette is removed from the tube.

Critical Step: Corning brand 2-ml serological pipettes consistently fit into OptiSeal tubes; other brands should be tested before use.

Refer to ***Table 2.T3*** for troubleshooting tips.

(B) Loading with a 5-ml pipette

(i) Attach a piece of tubing (see *Equipment*) to a 5-ml pipette. Begin by pipetting 6 ml of 15% (wt/vol) iodixanol into each tube. Next, add 6 ml of 25% (wt/vol) iodixanol under the 15% layer by lightly touching the tubing to the bottom of the tube and slowly releasing the solution (***Supplementary Video 2.V2***, 0:17–1:13). Continue adding layers of increasing density under the previous layer. The gradients should have a sharp delineation between layers (***Figure 2.8b***).

Critical Step: Fill the 5-ml pipette with more layer solution than is needed (e.g., an extra 1 ml per layer); this will prevent an air bubble from disturbing the gradient when releasing the last of the required volume (***Supplementary Video 2.V2***, 1:09–1:11). Remember to prepare extra solution (*Reagent setup*).

Refer to ***Table 2.T3*** for troubleshooting tips.

17. Centrifuge the lysate from Step 14 at 2,000g for 10 min at RT. Centrifugation will result in the formation of a pellet (***Figure 2.7b***).

18. Use a 2-ml serological pipette to load the supernatant (i.e., the clarified lysate) (~6–7 ml total) from Step 17 above the 15% (wt/vol) iodixanol layer (***Figure 2.8c*** and ***Supplementary Video 2.V1***, 1:46–2:22 or ***Supplementary Video 2.V2***, 1:14–1:55). Touch the pipette tip to the surface of the 15% layer and slowly release the lysate such that a layer forms on top.

Critical Step: Use a pipetting device with precise control. Do not allow the lysate to drip from the pipette tip onto the 15% layer; this will cause the lysate to mix with the gradient. Note that Corning brand 2-ml serological pipettes consistently fit into OptiSeal tubes; other brands should be tested before use.

Critical Step: The pellet may be soft, making it difficult to retrieve all of the supernatant. After loading 6–7 ml of lysate above the 15% layer, spin the lysate for an additional 15

min at 3,000g at RT; use a P200 or P1000 pipette to slowly load the remaining supernatant onto the lysate layer. Discard the pellet in 10% (vol/vol) bleach or a biohazard waste bin.

Critical Step: (Optional) Collect a 30-µl sample from the lysate for troubleshooting; store at 4 °C for up to 1 week. If the viral yield is lower than expected, the sample can be titered (Steps 32–42) to determine at which stage the virus may have been lost.

19. Using a 2-ml serological pipette, fill each tube up to the neck with SAN digestion buffer. Firmly insert a black cap (***Figure 2.8d***) and place a spacer on top (***Figure 2.8e***). Caps and spacers are provided with the OptiSeal tubes and in the OptiSeal tube kit, respectively.

Caution: Overfilling the tube can cause a spill when inserting the black cap. Handling the tubes without caps, or with loosely fitted caps, can also cause spills.

Caution: Avoid air bubbles, which can cause the OptiSeal tubes to collapse during ultracentrifugation.

Critical Step: The black cap should fit right above or touch the lysate.

20. Weigh the tubes with the caps and spacers on. Balance the tubes to within 5–10 mg of each other using SAN digestion buffer. Be sure to adjust the tube weight in the biosafety cabinet; use the tube removal tool provided with the OptiSeal tube kit to remove the black cap and add the appropriate volume of SAN digestion buffer with a P20 or P200 pipette.

Caution: Failure to balance the tubes before ultracentrifugation could result in damaged equipment.

21. Place the ultracentrifuge rotor in the biosafety cabinet. Load the tubes and fasten the lid.

Caution: Ensure that the rotor is in proper working order. This includes checking that the o-rings are intact, as cracked o-rings can cause virus to spill during ultracentrifugation. Also, check that the rotor and tubes are completely dry; moisture between tubes and the tube cavity can cause tubes to collapse. To prevent damage to the rotor, set it on a paper towel so that the over-speed disk at the bottom is not scratched.

22. Carefully transfer the rotor to the ultracentrifuge. Spin the Type 70 Ti rotor at 350,000g (58,400 r.p.m.) for 2 h and 25 min at 18 °C with slow acceleration (no. 3; the instrument will take 3 min to accelerate to 500 r.p.m., followed by maximum acceleration) and deceleration (no. 9; the instrument will decelerate at maximum speed until it reaches 500 r.p.m., then take 6 min to stop) profiles. Alternatively, a Type 60 Ti rotor can be used at 358,000 g (59,000 r.p.m.).

Caution: Always follow the manufacturer's instructions while operating an ultracentrifuge.

23. During ultracentrifugation, gather the supplies and equipment for Steps 24–27. Assemble the clamp setup (*Equipment setup*) and collect one of each of the following per gradient: Amicon Ultra-15 centrifugal filter device, 5-ml syringe, 10-ml syringe, 0.22-µm syringe filter unit, and a 16-gauge needle.

24. After ultracentrifugation, bring the rotor inside the biosafety cabinet and remove the lid. Use the spacer removal tool provided with the OptiSeal tube kit to remove the spacer from the first tube. Next, use the tube removal tool to grip the tube neck. Slowly remove the tube from the rotor and secure it into the clamp setup above a 500-ml or 1-liter beaker containing fresh 10% (vol/vol) bleach (***Figure 2.8f***). Clean the side of the tube with a paper towel or a Kimwipe sprayed with 70% (vol/vol) ethanol.

Caution: The black cap may become dislodged from the tube during removal, increasing the likelihood of a spill. Try replacing the cap before removing the tube from the rotor. Otherwise, replace the cap once the tube is secured in the clamp setup.

Caution: If a tube collapses during ultracentrifugation, take extra care when removing the tube from the rotor. Use fresh 10% (vol/vol) bleach to wipe the tube before proceeding with AAV purification. Viruses purified from collapsed tubes may have lower yields.

Refer to ***Table 2.T3*** for troubleshooting tips.

25. Prepare the supplies for Steps 26 and 27. First, remove and save the plunger from a 10-ml syringe. Attach a 0.22-µm syringe filter unit to the syringe barrel and place it on top of an Amicon filter device. Next, add 10 ml of DPBS to the barrel and allow the solution to begin dripping through the syringe filter unit and into the filter device (***Figure 2.8f***). Last, attach a 16-gauge needle to a 5-ml syringe.

Critical Step: Amicon filter devices contain traces of glycerine. If this interferes with downstream applications, rinse the device with DPBS before use. (Optional) Rinse the filtration membrane of the Amicon filter device by adding 15 ml of DPBS to the top chamber and centrifuging at 3,000 g for 1 min at RT; discard the flow-through. The manufacturer recommends using the device immediately after rinsing.

26. From the tube clamped in Step 24, collect the virus from the 40/60% interface and 40% layer[182,183] (***Figure 2.8g*** and ***Supplementary Video 2.V3***, 0:00–1:30). Hold the top of the OptiSeal tube with your non-dominant hand; use your dominant hand to hold the needle/syringe. Use a forward-twisting motion to insert the needle ~4 mm below the 40/60% interface (i.e., where the tube just starts to curve). Use the tube removal tool in your non-dominant hand to remove the black cap from the tube to provide a hole for air entry. With the needle bevel up, use the needle/syringe to collect 4.0–4.5 ml of virus/ iodixanol from the 40/60% interface and 40% layer. Do not collect from the white protein layer at the 25/40% interface; as this interface is approached, rotate the needle bevel down and continue collecting from the 40% layer. Firmly replace the black cap before removing the needle from the tube.

Caution: Keep your hands out of the path of the needle to prevent accidental exposure to AAVs. Failure to firmly replace the black cap before removing the needle will cause the AAV-contaminated solution to flow out of the needle hole in the tube and potentially onto and out of the biosafety cabinet. Perform this step over a large beaker of fresh 10% (vol/vol) bleach (***Figure 2.8f***).

Critical Step: The virus should concentrate at the 40/60% interface and within the 40% layer10. There will not be a visible virus band, but the phenol red in the 25% and 60% layers helps to better define the 40% cushion.

Critical Step: Before attempting to collect virus from the density gradient, practice on an OptiSeal tube filled with water.

Critical Step: (Optional) Collect a 30-µl sample from the virus/iodixanol for troubleshooting; store at 4 °C for up to 1 week. If the viral yield is lower than expected, the sample can be titered (Steps 32–42) to determine at which stage the virus may have been lost.

Refer to ***Table 2.T3*** for troubleshooting tips.

27. Add the 4.0–4.5 ml of virus/iodixanol to the syringe barrel containing 10 ml of DPBS (prepared in Step 25) (***Figure 2.8h*** and ***Supplementary Video 2.V3***, 1:31–2:06). Layer the virus below the DPBS by placing the needle near the bottom of the barrel and pressing on the plunger. Insert the 10-ml syringe plunger into the barrel and push the virus/DPBS mixture through the syringe filter unit and into the Amicon filter device (***Supplementary Video 2.V3***, 2:07–2:32). Mix well using a P1000 pipette.

Critical Step: This filtration step reduces the likelihood of clogging the filtration membrane in the Amicon filter device. The virus/iodixanol mixture will be difficult to push through the syringe filter unit; DPBS will be easy to push through as it washes the filter.

Critical Step: AAVs adhere to hydrophobic surfaces, including plastics; use low-binding pipette tips (*Reagents*). Pluronic F-68 is a nonionic surfactant that may reduce virus loss associated with sticking to plastics. (Optional) Include 0.001% (vol/vol) Pluronic F-68 in the DPBS for Steps 27–30. 28 Centrifuge the virus/DPBS mixture at 3,000g for 5–8 min at RT, or until the volume of the solution remaining in the top chamber of the Amicon filter device is 500–1,500 µl (>10× concentrated).

Critical Step: This step may take longer because iodixanol initially slows the passage of the solution through the filtration membrane.

29. Discard the flow-through for bleaching. Add 13 ml of DPBS to the virus in the top chamber and use a P1000 pipette to mix.

Critical Step: Remove the filter device, which contains the virus, before discarding the flow through.

30. Centrifuge the virus/DPBS mixture as in Step 28. Wash the virus two more times for a total of four buffer exchanges. During the last spin, retain 300–500 µl of solution in the top chamber.

Critical Step: The third and fourth washes may require only a 2–3-min spin until the desired volume remains in the top chamber.

Critical Step: The volume retained in the top chamber will affect the final virus concentration (vg/ml) (i.e., the lower the volume, the higher the concentration). A final volume of 300–500 µl should work for most applications, assuming a production efficiency of at least 1×10^{12} vg/dish and a dose and injection volume of no more than 1×10^{12} vg and 100 µl, respectively (see '*Experimental design*' section and Step 43 for dose and injection volume recommendations, respectively). For direct injections, a final volume of 200 µl may be optimal. Note that retaining too low a volume may cause the virus to aggregate during storage at 4 °C (see Step 42 for details).

31. Use a P200 pipette to transfer the virus from the top chamber of the Amicon filter device directly to a 1.6-ml screw-cap vial; store at 4 °C.

Critical Step: Amicon filter devices are not sterile. If this is a concern for specific applications, the virus can be filter-sterilized before storage. (Optional) Filter-sterilize the virus. Use a P200 pipette to transfer the virus from the top chamber of the Amicon filter device directly to a Costar Spin-X filter unit within a centrifuge tube. Centrifuge the virus at 3,000 g for 1 min at RT. Discard the filter unit and transfer the purified virus from the centrifuge tube to a 1.6-ml screw-cap vial; store at 4 °C.

Critical Step: The screw-cap vials are not low protein binding; however, they help prevent the formation of aerosols when opening and closing the tubes. We store AAVs in screw-cap vials at 4 °C and typically use them within 3 months, during which time we have not noticed a decrease in titers or transduction efficiency *in vivo*. We have not rigorously tested the effects of long-term storage at −20 °C or −80 °C for systemically delivered viruses.

Refer to ***Table 2.T3*** for troubleshooting tips.

Pause Point: Store the purified virus at 4 °C for up to 3 months.

AAV titration

● Timing 1 d

Critical: The AAV titration procedure below is adapted from ref.[184]. Each virus sample is prepared in triplicate in separate 1.5-ml DNA/RNA LoBind microcentrifuge tubes and later loaded into a 96-well plate for qPCR. All solutions must be accurately pipetted and thoroughly mixed; qPCR is highly sensitive to small changes in DNA concentration.

32. Prepare a plan for the PCR plate setup. Allocate the first 24 wells (A1–B12) for the DNA standards such that standard no. 1 occupies wells A1–A3, standard no. 2 occupies

wells A4–A6, and so on. Use the remaining wells for the virus samples such that the first virus sample occupies wells C1–C3, the second sample occupies wells C4–C6, and so on.

Critical Step: Include DPBS as a negative control and a virus sample with a known concentration as a positive control; prepare the controls with the virus samples in Steps 33–40.

33. Use DNase I to digest DNA that was not packaged into the viral capsid. Prepare DNase I + DNase digestion buffer (*Reagent setup*) and add 100 µl to each 1.5-ml tube. Vortex each virus for 1–2 s to mix; alternatively, use a P200 pipette to mix. Add 2 µl of the virus to each of three tubes. Vortex for 1–2 s to mix and spin down (2,000g, RT, 10 s); incubate in a 37 °C water bath for 1 h.

Critical Step: Do not vortex/pipette the virus vigorously or vortex longer than 1–2 s; exposure to force may disrupt capsid structure.

Critical Step: When dipping the pipette tip into the virus stock, insert the tip just below the surface of the liquid rather than dipping it deep inside. Excess virus carried on the outside of the tip will carry over into the DNase digestion buffer and cause variations in the titer.

Critical Step: Prepare each virus sample in triplicate.

34. Inactivate the DNase. Add 5 µl of EDTA to each tube; vortex for 1–2 s to mix, spin down (2,000 g, RT, 10 s), and incubate in a 70 °C dry bath for 10 min.

Critical Step: DNase must be inactivated or else it will degrade the viral genome when it is released from the viral capsid in Step 35.

35. Use proteinase K to digest the viral capsid and release the viral genome. Prepare proteinase K + proteinase K solution (*Reagent setup*) and add 120 µl to each tube. Vortex for 1–2 s to mix and spin down (2,000 g, RT, 10 s); incubate in a 50 °C dry bath for 2 h.

Pause Point: Samples can be incubated at 50 °C overnight.

36. During the last 20 min of the proteinase K digestion, prepare the DNA standard dilutions (*Reagent setup*) and use the Qubit assay to measure the concentration (ng/µl) of the DNA standard stock.

Critical Step: The concentration of the standard stock solution is used to generate the standard curve after qPCR (***Supplementary Table 2.ST4***, cell B9). To measure the concentration of the standard stock solution, use the Qubit fluorometer, which measures low DNA concentrations with high sensitivity and accuracy.

37. Inactivate the proteinase K. Incubate the tubes in a 95 °C dry bath for 10 min.

Caution: Tube caps may pop open unexpectedly; use safety glasses while removing the tubes from the 95 °C dry bath.

Critical Step: Proteinase K must be inactivated or else it will digest the DNA polymerase during qPCR.

38. Allow the tubes to cool for 5 min. Vortex each sample for 1–2 s to mix and add 3 µl to a new tube containing 897 µl of UltraPure water (a 1:300 dilution). Vortex the diluted samples for 3 s to mix.

39. Prepare the qPCR master mix (*Reagent setup*).

40. Load the PCR plate based on the experimental plan from Step 32. First, pipette 23 µl of qPCR master mix into each designated well. Next, pipette 2 µl of each standard into wells A1–B12. Last, pipette 2 µl of each diluted sample from Step 38 into wells C1 and onward. Seal the plate with sealing film and briefly spin down (500 g, RT, 10 s) in a plate spinner.

41. Place the PCR plate into the qPCR machine. Use the following cycling parameters: Step 1: 95 °C, 10 min Step 2: 95 °C, 15 s Step 3: 60 °C, 20 s Step 4: 60 °C, 40 s Repeat steps 2–4 40×.

42. When the qPCR run is complete, export the cycle threshold (Ct) values to an Excel file. Copy and paste the Ct values into ***Supplementary Table 2.ST4*** (‘*AAV titration calculator*’ sheet) to generate a standard curve and calculate the titer (vg/ml) (cell G27) of each virus; calculate per-plate production (vg/dish) (cell K27) to assess production efficiency. Be sure to enter the appropriate values in cells B7–10 and B18; see ‘*Example*’ sheet.

Critical Step: If the titer is ≥1 × 10^{14} vg/ml, the virus may aggregate during storage at 4 °C. Dilute the virus to between 2 × 10^{13} and 5 × 10^{13} vg/ml with DPBS and re-titer the diluted stock.

Refer to ***Table 2.T3*** for troubleshooting tips.

Intravenous (retro-orbital) injection

● Timing <5 min per mouse, excluding setup and cleanup time

Caution: Follow appropriate institutional and governmental guidelines and regulations for husbandry and handling of laboratory animals. Compared to tail-vein injections, retro-orbital injections require less technical expertise and may cause less distress in mice[185]; however, tail-vein injections appear to result in similar AAV distribution[158,160].

Critical: When possible, verify viral transduction and transgene expression in vitro before systemic administration. Note that co-injecting AAVs with other substances (e.g., dyes) could affect infectivity *in vivo* and should be tested independently.

Critical: Re-titer viruses before injection if more than 1 month has passed since titration; this will ensure that animals are administered the most accurate dose possible.

43. Determine the dose of virus to administer per mouse (see ‘*Experimental design*’ section for recommendations). Divide the dose (vg) by the titer (vg/ml) (***Supplementary Table 2.ST4***, cell G27) to calculate the volume of virus needed to inject one mouse. In a screw-

cap vial, prepare a master mix of virus based on the number of animals to be injected; briefly vortex each virus and master mix for 1–2 s before use. Transport the virus on ice once it is ready for injection.

Caution: Do not inject more than 10% of the mouse blood volume, which corresponds to 150 µl for a 25-g mouse.

Critical Step: Depending on the user, it is easiest to inject 40–80 µl/mouse. If <40 µl/mouse is required, use DPBS or saline to dilute the virus such that a larger volume is injected. If more than 80 µl/mouse is required, it may be more convenient to re-concentrate the virus or perform two separate injections; follow institutional guidelines for multiple eye injections. Virus will be lost in the event of an unsuccessful injection; therefore, prepare more master mix than is required.

Critical Step: To reduce the chance of contaminating the virus stock, avoid using the original virus stock; bring only an aliquot of what is needed for the injections.

Critical Step: Do not store diluted viruses; only dilute what is needed immediately before injection.

Refer to ***Table 2.T3*** for troubleshooting tips.

44. Assemble the anesthesia system[229] (*Equipment setup*) inside the biosafety cabinet.

45. Remove the mouse from its cage and place it in the induction chamber. Anesthetize the mouse with 1–5% isoflurane in oxygen.

Caution: Isoflurane must be handled according to federal, state, and local regulations.

46. While the mouse is being induced, load an insulin syringe with virus. Remove the dead space in the syringe barrel by gently ejecting the virus back into the tube such that air bubbles are expelled. Load the syringe again and repeat the procedure until no bubbles remain in the barrel.

Caution: Introducing air into the vascular system can be fatal.

Critical Step: Introducing air into the virus may cause protein denaturation; perform this step gently and only until no bubbles remain in the syringe barrel.

47. Remove the anesthetized mouse from the induction chamber. Place the animal in a prone position on a small stack of paper towels. Position the mouse such that its head is situated on the same side as the operator's dominant hand. Place the nose cone on the mouse to maintain anesthesia.

48. Use the index finger and thumb of the non-dominant hand to draw back the skin above and below the eye, causing the eye to slightly protrude from the socket[185]. With the dominant hand, insert the needle, bevel down, at a 30–45° angle into the medial canthus and through the conjunctival membrane. The needle should be positioned behind the globe of the eye in the retro-orbital sinus. Slowly release the virus into the sinus and gently remove the needle.

Caution: Assess anesthetic depth by loss of pedal reflex (via toe pinch) before inserting the needle into the retro-orbital sinus. Any movement of the eye or skin when the needle is inserted indicates incorrect needle placement. Keep hands out of the path of the needle to prevent accidental exposure to AAVs. Do not recap needles; discard into an approved biohazardous sharps container immediately after use.

Critical Step: No liquid should leak out of the eye after viral delivery; likewise, little to no bleeding should be observed.

Refer to ***Table 2.T3*** for troubleshooting tips.

49. Following viral injection, apply mild pressure to the eyelid. Apply 1–2 drops of proparacaine to the corneal surface to provide local analgesia. After recovery from anesthesia, place the mouse in a clean cage.

Caution: Monitor the eye daily after injection for 2 d, or according to institutional guidelines.

Evaluation of transgene expression

● Timing variable; see '*Experimental design*' section

Caution: Follow appropriate institutional and governmental guidelines and regulations for husbandry and handling of laboratory animals.

50. To assess endogenous fluorescence in fixed tissue (***Figures 2.2–2.5***), anesthetize and transcardially perfuse[232] the animals after sufficient time has passed for viral transduction and protein expression (see '*Experimental design*' section for recommendations). Cut thin[233] or thick[234] tissue slices and mount them in RIMS[171,218] or a commercially available mounting media such as Prolong Diamond Antifade (Thermo Fisher Scientific, cat. no. P36965) or Vectashield (Vector Laboratories, cat. no. H-1000-10). Alternatively, use PACT- or PARS-based clearing[171,218] of whole organs or animals, respectively, or another clearing method (e.g., ScaleSQ[170]). Ensure that the chosen clearing protocol is compatible with the fluorescent protein(s) under investigation (see '*Experimental design*' and *'Anticipated results'* sections for details). Native fluorescence can be visualized using a fluorescence microscope (e.g., Keyence BZ-X700), confocal microscope (e.g., Zeiss LSM 880), or light-sheet microscope (e.g., custom-made[218] or LaVision BioTec UltraMicroscope II), depending on the tissue volume and desired imaging resolution. For experiments without fluorescent labels, evaluate transgene expression using molecular (e.g., qPCR3 or Western blot[219]), histological[218] (e.g., with antibodies or molecular probes), or functional (e.g., optical imaging[217]) methods relevant to the experimental aims. Regardless of the approach used to evaluate gene expression, cell type–specific promoters should be verified at this stage in the protocol; we typically assess cell morphology and/or use antibody staining to confirm specificity[133] (***Figure 2.2b,c***).

Refer to ***Table 2.T3*** for troubleshooting tips.

2.5 TROUBLESHOOTING

Troubleshooting advice can be found in ***Table 2.T3***.

Table 2.T3: Troubleshooting table

Step	Problem	Possible reason	Solution
2 (Transfection)	Transfection solution is not cloudy	DNA–PEI complexes have not formed	Thoroughly vortex the transfection solution for 10 s and incubate at RT for 2–10 min before use; always use PEI at RT
		Transfection miscalculation	Carefully follow the instructions in the Reagent setup and Supplementary Table 2 ('Transfection calculator' sheet) to prepare the PEI + DPBS master mix and DNA + DPBS solutions
3 (Transfection)	Low or no fluorescent protein expression post transfection	Low DNA purity	Use an endotoxin-free plasmid purification kit to isolate plasmids; assess DNA purity (i.e., 260/280 and 260/230 ratios) before transfection
		Mutations in plasmids	Verify the integrity of pAAV plasmids by sequencing and restriction digestion before transfection
		Poor cell health	Maintain cells in an actively dividing state at recommended ratios (Reagent setup). Ensure cells are not over-confluent at the time of transfection, and change media no more than 24 h post transfection
		Weak fluorescent reporter and/or promoter, or promoter	Include a positive transfection control (e.g., pAAV-CAG-eYFP). Note that some promoters may take 2–3 d to show expression

Step	Problem	Possible reason	Solution
		cannot initiate gene expression in HEK293T cells	
		Transgene expression depends on Flp or Cre recombinase	Include a positive transfection control (see above)
		Transfection miscalculation	Carefully follow the instructions in the Reagent setup and Supplementary Table 2 ('Transfection calculator' sheet) to prepare the PEI + DPBS master mix and DNA + DPBS solutions
9 (AAV harvest)	Cell lysate is not pink	pH of the lysate is too low	Check the pH of the lysate by pipetting 30 µl onto a pH strip; adjust the pH to 8.5 with NaOH suitable for cell culture. In subsequent viral preps, ensure that the pH of SAN digestion buffer is ~10.0; during cell lysis, the pH should drop to 8.5–9.0, which is optimal for SAN digestion
		Fluorescent protein expression from a strong promoter (e.g., CAG)	Expression of blue/green or red fluorescent proteins from strong promoters can cause the lysate to turn yellow or red, respectively; proceed with AAV production
16 (AAV purification)	Density gradients have no clear delineation between iodixanol layers	Layers are mixed	Repour the gradients (Supplementary Video 1, 0:00–1:45, or Supplementary Video 2, 0:00–1:13); gradients should be poured fresh and not allowed to sit for too long

Step	Problem	Possible reason	Solution
24 (AAV purification)		An air bubble was trapped underneath the black cap	Carefully remove the tube from the rotor and wipe it with fresh 10% (vol/vol) bleach before proceeding with AAV purification. In future viral preps, remove air bubbles with a P200 pipette before ultracentrifugation
	Tube collapsed during ultracentrifugation	The rotor and/or OptiSeal tubes were not in proper working order	Carefully remove the tube from the rotor and wipe it with fresh 10% (vol/vol) bleach before proceeding with AAV purification. In future viral preps, check that the rotor and tubes are completely dry; moisture between tubes and the tube cavity can cause tubes to collapse. Also check tubes for dents before pouring the density gradients
	Cannot puncture the OptiSeal tube with the needle	Not enough force is used	Use a forward-twisting motion to insert the needle into the tube (Supplementary Video 3, 0:06–0:21); practice on an OptiSeal tube filled with water
	Two holes were punctured through the OptiSeal tube	Too much force was used	See above. Do not remove the needle; carefully insert a new needle and proceed to collect virus
26 (AAV purification)		Black cap was not removed	Use the tube removal tool to remove the black cap from the tube after inserting the needle but before collecting virus (Fig. 8g and Supplementary Video 3, 0:22–0:30); practice on an OptiSeal tube filled with water
	Cannot collect virus with the needle	Plastic from the tube is lodged inside the needle	Firmly replace the black cap and remove the needle from the tube; insert a new needle into the same hole, remove the black cap, and try collecting virus again

Step	Problem	Possible reason	Solution
31 (AAV purification)	Density gradient flows out of the needle hole in the tube after removal of the needle	Black cap was not firmly replaced	Act quickly; use the beaker of bleach to catch the liquid and firmly replace the black cap to stop the flow. In subsequent viral preps, ensure that the black cap is replaced before removing the needle from the tube (Supplementary Video 3, 1:19–1:30); practice on an OptiSeal tube filled with water
	Purified virus is cloudy	The virus/DPBS mixture was not mixed and contains iodixanol	Repeat the buffer exchanges in Steps 28–30. In future viral preps, thoroughly mix the virus/DPBS mixture using a P1000 pipette in Steps 27, 29, and 30
		Salt, DNA, or viral precipitation	Before titering or injecting the virus, spin down the precipitate at 3,000*g* for 5 min at RT and transfer the supernatant (i.e., the virus) to a new screw-cap vial. We have not noticed a decrease in titer after removing precipitate from our preps; however, it is a good practice to re-titer a virus if precipitate has formed
	Unknown material is suspended in purified virus	Bacterial contamination	Bleach the virus. In future viral preps, filter-sterilize viruses after purification, and only open tubes containing viruses in a biosafety cabinet. During intravenous injections, never use the original virus stock; bring only an aliquot of what is needed for injection
		Carry-over from the filtration membrane of	Filter-sterilize the virus

Step	Problem	Possible reason	Solution
		the Amicon filter device	
	No SYBR signal detected for DNA standards or virus samples	Missing reagents (e.g., primers) in qPCR reaction	Check that all qPCR reagents were added to the master mix and that the DNA standards and virus samples were added to their respective wells
		Degraded reagents	Use fresh, properly stored qPCR reagents
	No SYBR signal detected for virus samples	DNase was not inactivated, resulting in degradation of the viral genome during proteinase K treatment	Repeat the titration procedure; be sure to inactivate DNase with EDTA at 70 °C (Step 34)
42 (AAV titration)		Proteinase K was not inactivated, resulting in degradation of DNA polymerase during qPCR	Repeat the titration procedure; be sure to inactivate Proteinase K at 95 °C (Step 37)
	Triplicates do not have similar Ct values	Inaccurate pipetting and/or inadequate mixing of reagents	Repeat the qPCR; pipette accurately and thoroughly mix all reagents before use
	Standard curve is not linear	Inaccurate pipetting and/or inadequate mixing of reagents while preparing the	Repeat the qPCR; pipette accurately and thoroughly mix all reagents while preparing the DNA standard dilutions. Note that at low concentrations (high Ct values), standard nos. 1 and 2 will deviate from linearity (Supplementary

Step	Problem	Possible reason	Solution
		DNA standard dilutions	Table 4, ‘Example’ sheet). This is normal; the qPCR does not need to be repeated
		DNA standards degraded and/or stuck to the walls of 1.5-ml tubes	Repeat the qPCR; prepare the DNA standard dilutions fresh, immediately before use, and use only DNA/RNA LoBind microcentrifuge tubes. Note that at low concentrations (high Ct values), standard nos. 1 and 2 will deviate from linearity (Supplementary Table 4, ‘Example’ sheet). This is normal; the qPCR does not need to be repeated
	Viral yield is lower than expected (Supplementary Table 4, cell K27)	Transfection, AAV harvest, AAV purification, and/or AAV titration were not successful	Include a positive transfection/virus production control (e.g., pAAV-CAG-eYFP) and a positive titration control. To determine at which stage the virus may have been lost, collect a 30-µl sample from the cell lysate (Step 9), the media before PEG precipitation (Step 10), the PEG pellet resuspension (Step 14), and the lysate before (Step 18) and after iodixanol purification (Step 26). Store samples at 4 °C for up to 1 week for titering (Steps 32–42)
		AAV capsid and/or genome results in poor production efficiency	Scale up viral preps to ensure enough virus is produced for downstream applications
		ITRs underwent recombination	After plasmid purification, but before transfection, digest pAAVs with SmaI to confirm the presence of ITRs, which are required for

Step	Problem	Possible reason	Solution
		during bacterial growth	replication and encapsidation of the viral genome; always propagate pAAVs in recombination-deficient bacterial strains
43 (intravenous injection)	A large volume (e.g., more than 100 µl, depending on user preference) of virus needs to be injected	Virus concentration is too low	Reconcentrate the virus using an Amicon filter device. Add 13 ml of DPBS and the virus to the top chamber of the Amicon filter device and use a P1000 pipette to mix. Centrifuge at 3,000*g* at RT until the desired volume of solution remains in the top chamber
48 (Intravenous injection)	Virus spills out of the eye during injection	Incorrect needle placement	Absorb the spilled virus using a paper towel; disinfect AAV-contaminated surfaces and materials with fresh 10% (vol/vol) bleach or an equivalent disinfectant. Load the same insulin syringe with more virus, position the needle behind the globe of the eye in the retro-orbital sinus, and try the injection again. Practice injections using DPBS or saline until comfortable with the procedure
	Bleeding before, during, or after injection	Incorrect needle placement	Position the needle behind the globe of the eye in the retro-orbital sinus; never puncture the eye itself. Inject the virus slowly; following injection, carefully remove the needle at the same angle at which it was inserted. Practice injections using DPBS or saline until comfortable with the procedure
		Needle is left in the injection site for too long	Once the needle is correctly placed in the eye, immediately inject the virus

Step	Problem	Possible reason	Solution
		Sufficient time has not passed for protein expression	Wait longer for optimal protein expression
		Dose is too low, or dose is too high, causing cell toxicity	Inject multiple animals with a series of doses and sacrifice them at different time points (e.g., weekly) to determine the optimal dose
		Titer is inaccurate	Re-titer viruses before injection if more than 1 month has passed since titration; this will ensure that the animals are administered the most accurate dose possible
50 (Evaluation of transgene expression)	Weak or no transgene expression in the tissue of interest	Virus degraded	See above. Store AAV-PHP viruses at 4 °C for up to 3 months, during which time we have not noticed a decrease in titers or transduction efficiency *in vivo*. Do not store diluted viruses; dilute only what is needed immediately before retro-orbital injection
		Weak or no expression from the AAV genome	Verify the integrity of pAAV plasmids by sequencing and restriction digestion before transfection. If possible, verify viral transduction and transgene expression in vitro before systemic administration
		Poor viral injection	Inject multiple animals to increase the chance of success
		Fluorescent protein and/or signal deteriorates in chemical-clearing reagents	Ensure that the chosen clearing protocol is compatible with the fluorescent protein(s) under investigation (see ‘Experimental design’ and ‘Anticipated results’ sections for details)

Step	Problem	Possible reason	Solution
	Fluorescent signal photobleaches during imaging	Fluorescent protein is sensitive to photobleaching (e.g., during long imaging sessions or at high magnification)	Use a different fluorescent protein with similar spectral properties but higher photostability (e.g., tdTomato rather than mRuby2, or eGFP rather than Emerald)
	Lipofuscin accumulation	Aging tissue	Reduce autofluorescence using histological methods (e.g., Sudan black) or, if possible, inject younger adults (≤8 weeks old) and determine the minimum time required for transgene expression

2.6 TIMING

Refer to ***Figure 2.7a*** for a time line of the Procedure.

Steps 1–3, triple transient transfection of HEK293T cells: 1–2 h

Steps 4–14, AAV harvest: 5 d

Steps 15–31, AAV purification: 1 d

Steps 32–42, AAV titration: 1 d

Steps 43–49, intravenous (retro-orbital) injection: <5 min per mouse, excluding setup and cleanup time Step 50, evaluation of transgene expression: variable.

2.7 ANTICIPATED RESULTS

AAV production

For capsids that package well (i.e., AAV-PHP.B, AAV-PHP.eB, and AAV-PHP.S), the AAV production protocol typically yields ≥1 × 10^{12} vg per 150-mm dish[133,157]. Production efficiency can be determined for each virus in Step 42 (***Supplementary Table 2.ST4***, cell K27). Note that yields may vary from prep to prep and genome to genome. Users can gauge production efficiency for each experiment by running a positive control (e.g., pAAV-CAG-eYFP).

Evaluation of transgene expression

For most applications, users can expect to assess transduced cells beginning 2 or more weeks after intravenous injection (see '*Experimental design*' section for details). The chosen method for evaluating transgene expression will vary from user to user and may involve molecular, histological, and/or functional approaches[220] (Step 50). We typically use fluorescent reporters to assess gene expression in thick (≥100 μm), cleared tissue samples; below, we discuss expected results for the applications presented here (***Figures 2.2–2.5***) and in our previous work[133,157].

Commonly used reporters such as GFP, eYFP, and tdTomato show strong fluorescent labeling in PACT- and PARS-cleared tissues, enabling whole-organ and thick-tissue imaging of transgene expression[157,171,218]. Most markers, including mTurquoise2, mNeonGreen, and mRuby2, can also be detected after mounting labeled tissues in optical clearing reagents such as RIMS[218] or ScaleSQ[170] (***Figures 2.3a,c,d, 2.4b-d*** and ***2.5e***) or commercially available mounting media (***Figure 2.5a,c***). Depending on the rAAV genome, fluorescent proteins can be localized to distinct cellular compartments, including the nucleus (via NLS) (***Figure 2.2a,b***), cytosol (***Figures 2.2d,e, 2.3, 2.4b*** and ***2.5***), and cell membrane (via farnesylation[175] or fusion to a membrane protein such as ChR2) (***Figures 2.3d*** and ***2.4d***).

In summary, we present a comprehensive protocol for the production and administration of AAV-PHP viruses. We have validated the ability of AAV-PHP variants to provide efficient and noninvasive gene delivery to specific cell populations throughout the body. Together, this AAV toolbox equips users with the resources needed for a variety of applications across the biomedical sciences.

2.8 SUPPLEMENTARY INFORMATION

AAV-PHP capsid	Production	Species (strain or line), age injected	Organs/cell populations transduced	References
AAV-PHP.B	Good	Mouse (C57BL/6J, FVB/NCrl, 129S1/SvImJ, and multiple Cre driver lines), ≥6 weeks	Neurons, astrocytes, oligodendrocytes, and endothelial cells across all brain and spinal cord regions examined. Transduction in the retina occurs with variable efficiency across all layers	Refs. 1-6, this work
		Mouse (C57BL/6J), P0-P1	Brain and spinal cord	Ref. 4
		Rat (Sprague-Dawley), P1 and 6 weeks	Neurons throughout the brain and spinal cord	Ref. 7
		Human iPSCs	Neurons and astrocytes in cortical spheroids	Ref. 2
AAV-PHP.eB	Good	Mouse (C57BL/6J, C57BL/6NCrl, and 129T2/SvEmsJ), ≥6 weeks	Similar to AAV-PHP.B but with increased efficiency of neuronal transduction	Ref. 8, this work
		Rat (Fischer), 6 weeks		
		Rat (Long Evans), 4 weeks		
		Rat (Sprague-Dawley), 6 weeks		Ref. 9
AAV-PHP.S	Good	Mouse (C57BL/6J and Cre driver lines), 6-8 weeks	Sensory neurons, peripheral ganglia (sympathetic, nodose, dorsal root, and cardiac ganglia), and the myenteric and submucosal plexuses of the enteric nervous system. Robust transduction of heart muscle, skeletal muscle, and circular and longitudinal muscle of the digestive tract, as well as the liver and lungs	Ref. 8, this work
AAV-PHP.A	Poor	Mouse (C57BL/6J), 6 weeks	Astrocytes throughout the brain	Ref. 2

Supplementary Table 2.ST1: Use of AAV-PHP capsids for efficient transduction across organs and cell populations.

Species/strains, organs, and cell populations examined to-date following intravenous administration of AAV-PHP viruses. To target distinct cell types, use rAAV genomes with

cell type–specific gene regulatory elements and/or Cre- or Flp-dependent recombination schemes (***Figures 2.2-2.4***, ***Figure 2.6*** and ***Table 2.T1***). iPSCs, induced pluripotent stem cells. The refs. 1[190], 2[157], 3[191], 4[158], 5[160], 6[189], 7[159], 8[133], 9[162].

Supplementary Table 2.ST2: Transfection calculator.

This is an interactive calculator and provided as an Excel file (see Step 2 and '*Reagent setup*'). Note that our PEI calculations ('*Detailed calculations*' sheet) are based on ref.[178].

Volume of each layer (ml)	Layer		Number of gradients				
	%	Solution (ml)	1	2	4	6	8
6	15	DPBS + high salt	5.0	9.9	19.8	29.7	39.6
		60% iodixanol	1.7	3.3	6.6	9.9	13.2
6	25	DPBS + low salt	3.9	7.7	15.4	23.1	30.8
		60% iodixanol	2.8	5.5	11.0	16.5	22.0
		Phenol red	0.1	0.1	0.2	0.3	0.4
5	40	DPBS + low salt	1.8	3.7	7.3	11.0	14.7
		60% iodixanol	3.7	7.3	14.7	22.0	29.3
5	60	60% iodixanol	5.5	11.0	22.0	33.0	44.0
		Phenol red	0.1	0.1	0.2	0.3	0.4

Supplementary Table 2.ST3: Pouring the iodixanol density gradients.

Determine the number of gradients needed and prepare the iodixanol density gradient solutions (*Reagent setup*). The 15% layer contains high salt to destabilize ionic interactions between viral particles and cell proteins in the clarified lysate[183]. In Step 16B, prepare more solutions than are needed (*Reagent setup*).

Supplementary Table 2.ST4: Titration calculator.

This is an interactive calculator and provided as an Excel file (see Step 42 and '*Reagent setup*').

Supplementary Video 2.V1: Pouring gradients part 1

Steps 16A and 18: Pouring the density gradients and loading the virus. In Step 16A, use a 2-ml serological pipette to pour the gradients. Next, load the virus (also shown in Step 16B (***Supplementary Video 2.V2***))

Supplementary Video 2.V2: Pouring gradients part 2

Steps 16B and 18: Pouring the density gradients and loading the virus. In Step 16B, use a 5-ml serological pipette to pour the gradients. Next, load the virus (also shown in Step16A (***Supplementary Video 2.V1***))

Supplementary Video 2.V3: Virus collection

Steps 26–27: Collecting the virus

Chapter 3

MULTIPLEXED CRE-DEPENDENT SELECTION (M-CREATE) YIELDS SYSTEMIC AAVS FOR TARGETING DISTINCT BRAIN CELL TYPES

Ravindra Kumar, S., Miles, T. F., Chen, X., Brown, D., Dobreva, T., Huang, Q., Ding, X., Luo, Y., Einarsson, P.H., Greenbaum, A., Jang, J.J., Deverman, B. E., Gradinaru, V. Multiplexed Cre-dependent selection yields systemic AAVs for targeting distinct brain cell types. *Nature Methods (2020).* https://doi.org/10.1038/s41592-020-0799-7

3.1 SUMMARY

Recombinant adeno-associated viruses (rAAVs) are efficient gene delivery vectors via intravenous delivery; however, natural serotypes display a finite set of tropisms. To expand their utility, we evolved AAV capsids to efficiently transduce specific cell types in adult mouse brains. Building upon our Cre-recombination-based AAV targeted evolution (CREATE) platform, we developed Multiplexed-CREATE (M-CREATE) to identify variants of interest in a given selection landscape through multiple positive and negative selection criteria. M-CREATE incorporates next-generation sequencing, synthetic library generation, and a dedicated analysis pipeline. We have identified capsid variants that can transduce the central nervous system broadly, exhibit bias toward vascular cells and astrocytes, target neurons with greater specificity, or cross the blood–brain barrier across diverse murine strains. Collectively, the M-CREATE methodology accelerates the discovery of capsids for use in neuroscience and gene-therapy applications.

3.2 INTRODUCTION

Recombinant adeno-associated viruses (rAAVs) are widely used as gene delivery vectors in scientific research and therapeutic applications due to their ability to transduce both dividing and non-dividing cells, their long-term persistence as episomal DNA in

infected cells, and their low immunogenicity[100,155,235–237]. However, gene delivery by natural AAV serotypes is limited by dose-limiting safety constraints and largely overlapping tropisms. AAV capsids engineered by rational design[145,146,238,239] or directed evolution[69,93,143,149,151,152,204,240–243] have yielded vectors with improved efficiencies for select cell populations[133,147,150,157,187,244,245], yet much work remains to identify a complete toolbox of efficient and specific vectors. Previously, we evolved the AAV-PHP.B and AAV-PHP.eB variants from AAV9 using a selection method called CREATE[157]. This method applies positive selective pressure for capsids capable of infecting a target cell population by pairing a viral genome containing lox sites with *in vivo* selection in transgenic mice expressing Cre in the cell type of interest. This combination allows a Cre–Lox recombination-dependent PCR amplification of only those capsids which successfully deliver their genomes to the nuclei of the target cell type.

To more efficiently expand the AAV toolbox, we developed Multiplexed-CREATE (M-CREATE) (***Figure 3.1a*** and ***Supplementary Figure. 3.S1a,b***), which compares the enrichment profiles of thousands of capsid variants across multiple cell types and organs within a single experiment. This method improves upon its predecessor by capturing the breadth of capsid variants at every stage of the selection process. M-CREATE supports: (1) the calculation of an enrichment score for each variant by using next-generation sequencing (NGS) to correct for biases in viral production prior to selection, (2) reduced propagation of bias in successive rounds of selection through the creation of a post-round one synthetic pool library with equal variant representation, and (3) the reduction of false positives by including codon replicates of each selected variant in the pool. These improvements allow interpretation of variants' relative infection efficiencies across a broad range of enrichments in multiple positive selections and enable post-hoc negative screening by comparing capsid libraries recovered from multiple target cell types or organs. Collectively, these features allow prioritization of capsid variants for validation and characterization.

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

3.3 RESULTS

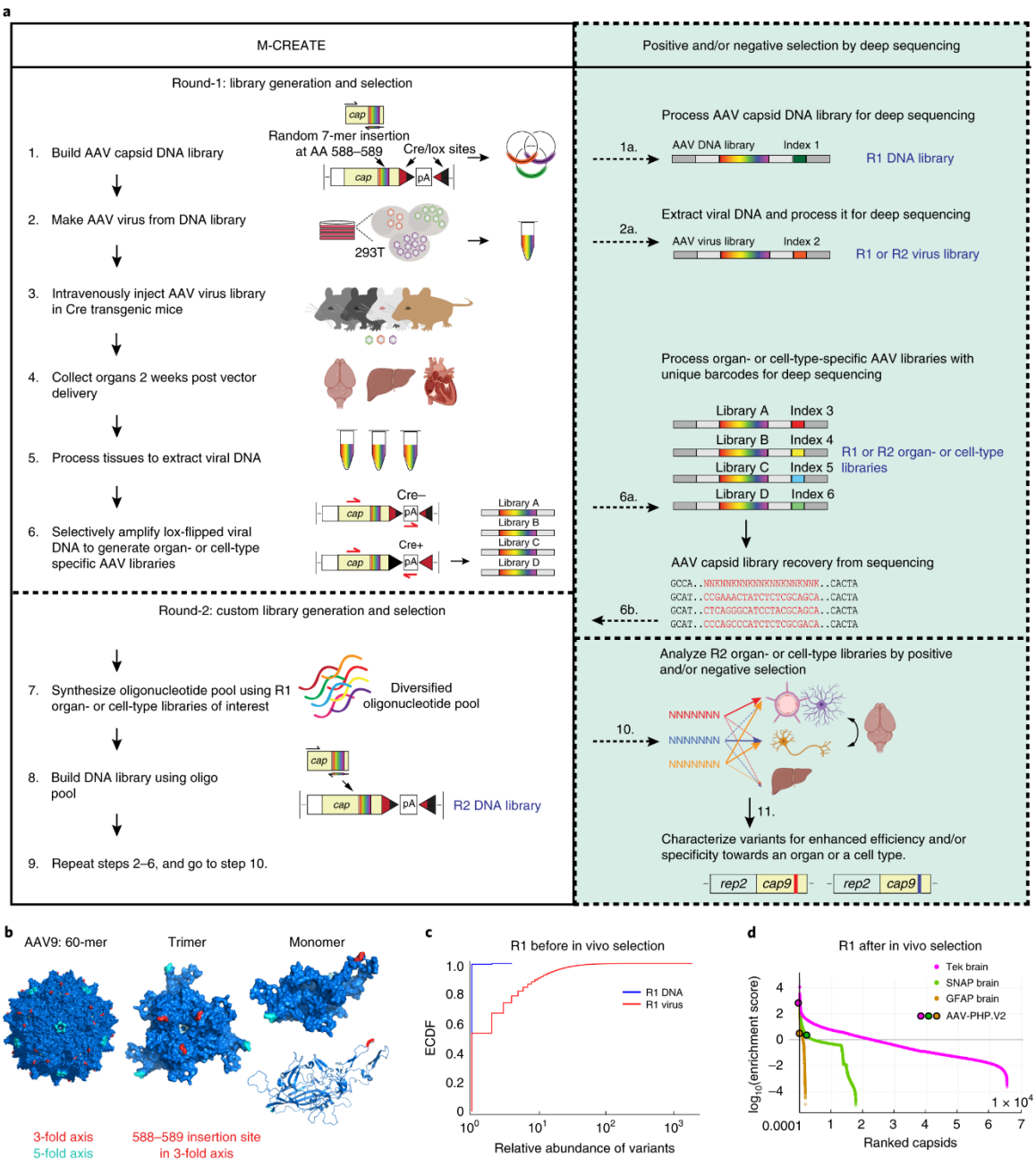


Figure 3.1: Workflow of M-CREATE and analysis of 7-mer-i selection in round-1.

(a), A multiplexed selection approach to identify capsids with specific and broad tropisms. Steps 1–6 describe the workflow in round-1 (R1) selection, steps 7–9 describe round-2 (R2) selection using the synthetic-pool method, steps 1a, 2a and 6a,b show the incorporation of deep sequencing to recover capsids after R1 and R2 selection, and steps 10–11 describe positive and/or negative selection criteria followed by variant characterization. The genes *rep2* and *cap9* in step 11 refers to *rep* from AAV2

and *cap* from AAV9, respectively, and the colored bar within *cap9* represents the targeted mutation. **(b)**, Structural model of the AAV9 capsid (PDB 3UX1) with the insertion site for the *7-mer-i* library highlighted in red in the 60-meric (left), trimeric (middle) and monomeric (right) forms. **(c)**, Empirical cumulative distribution frequency (ECDF) of R1 DNA and virus libraries that were recovered by deep sequencing post Gibson assembly and virus production, respectively. **(d)**, Distributions of variants recovered from three R1 libraries from Tek-Cre, SNAP25-Cre and GFAP-Cre brain tissue ($n = 2$ per Cre line) are shown with capsid libraries, sorted by decreasing order of the enrichment score. The enrichment scores of the AAV-PHP.V2 variant are mapped as well.

Analysis of capsid libraries during round-1 selection

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

During DNA- and virus-library generation there is potential for biased accumulation and over-representation of certain capsid variants, obscuring their true enrichment during *in vivo* selection. These deviations may result from PCR amplification bias in the DNA library or sequence bias in the efficiency of virus production across various steps such as capsid assembly, genome packaging and purification. We investigated this with a *7-mer-i* (i for insertion) library, in which a randomized 7-amino acids (AA) library is inserted between AA 588 and 589 of AAV9 (***Figure 3.1a,b***) in the rAAV-ΔCap9-in-cis-Lox2 plasmid (*Methods* and ***Supplementary Figure 3.S1a***; theoretical library size, 3.4×10^{10} unique nucleotide sequences, and an estimated $\sim 1 \times 10^{8}$ nucleotide sequences upon transfection). We sequenced the libraries after DNA assembly and after virus purification to a depth of 10–20 million (M) reads, which was adequate to capture the bias among variants during virus production (***Figure 3.1c*** and ***Supplementary Figure 3.S1b–d***). The DNA library had a uniform distribution of 9.6 M unique variants within ~10 M total reads (read count (RC) mean = 1.0, s.d. = 0.074), indicating minimal bias. In contrast,

the virus library had 3.6 M unique variants within ~20 M depth (RC mean = 4.59, s.d. = 11.15) indicating enrichment of a subset of variants during viral production. Thus, even permissive sites like 588–589 will impose biological constraints on sampled sequence space.

For *in vivo* selection, we intravenously administered the *7-mer-i* viral library in transgenic mice expressing Cre in astrocytes (GFAP-Cre), neurons (SNAP25-Cre) or endothelial cells (Tek-Cre) at a dose of 2×10^{11} vector genomes (vg) per adult mouse ($n = 2$ mice per Cre transgenic line). Two weeks after intravenous (i.v.) injection, we collected brain, spinal-cord and liver tissues. We extracted the rAAV genomes from tissues and selectively amplified the capsids that transduced Cre-expressing cells (***Supplementary Figure 3.S1e–i***). Upon deep sequencing, we observed ~8×10^4 unique nucleotide variants in brain tissue samples (~48% of which were identified in the sequenced portion of the virus library) and <50 variants in spinal-cord samples across the transgenic lines, and each variant was represented with an enrichment score that reflects the change in relative abundance between the brain and the starting virus library (*Methods* and ***Figure 3.1d***).

Two features of this dataset stand out. First, the variants recovered from brain tissue were disproportionately represented in the sequenced fraction of the viral library, demonstrating how production biases can skew selection results. Second, the distribution of capsid read counts reveals that more than half of the unique variants recovered after selection appear at low read counts. These variants may either have arisen spontaneously from errors during experimental manipulation or retain AAV9's basal levels of central nervous system (CNS) transduction (***Supplementary Figure 3.S1e***).

Unbiased round-2 library design improves the selection outcome

Concerned that the sequence bias during viral production and recovery would propagate across selection rounds despite our post-hoc enrichment scoring, we designed an unbiased library based on the round-1 (R1) output (synthetic pool library) via

oligonucleotide pools. We compared this library with a library PCR amplified directly from the recovered R1 DNA (PCR pool library) (***Figure 3.2a*** and ***Table 3.T1***).

Design Parameters	***Synthetic pool* design**	***PCR pool* design**
Carryover of R1 selection bias among variants	No, likelihood of false positives is low	Yes, potential to minimize by normalization
Carryover of R1 selection induced mutants	No	Yes
Confidence in library performance	High, using alternate codon replicates	Low
Customize library or add internal controls	Yes, in an unbiased manner	Yes, with greater risk of bias
Control library size	Yes, without reducing libraries or pooling	Yes, with libraries reduced for pooling
Cost for R2 library generation	High	Low

Table 3.T1: Comparison between the two methods for R2 selection.

The table summarizes the pros and cons of selection design parameters by the synthetic pool and PCR pool R2 selection methods.

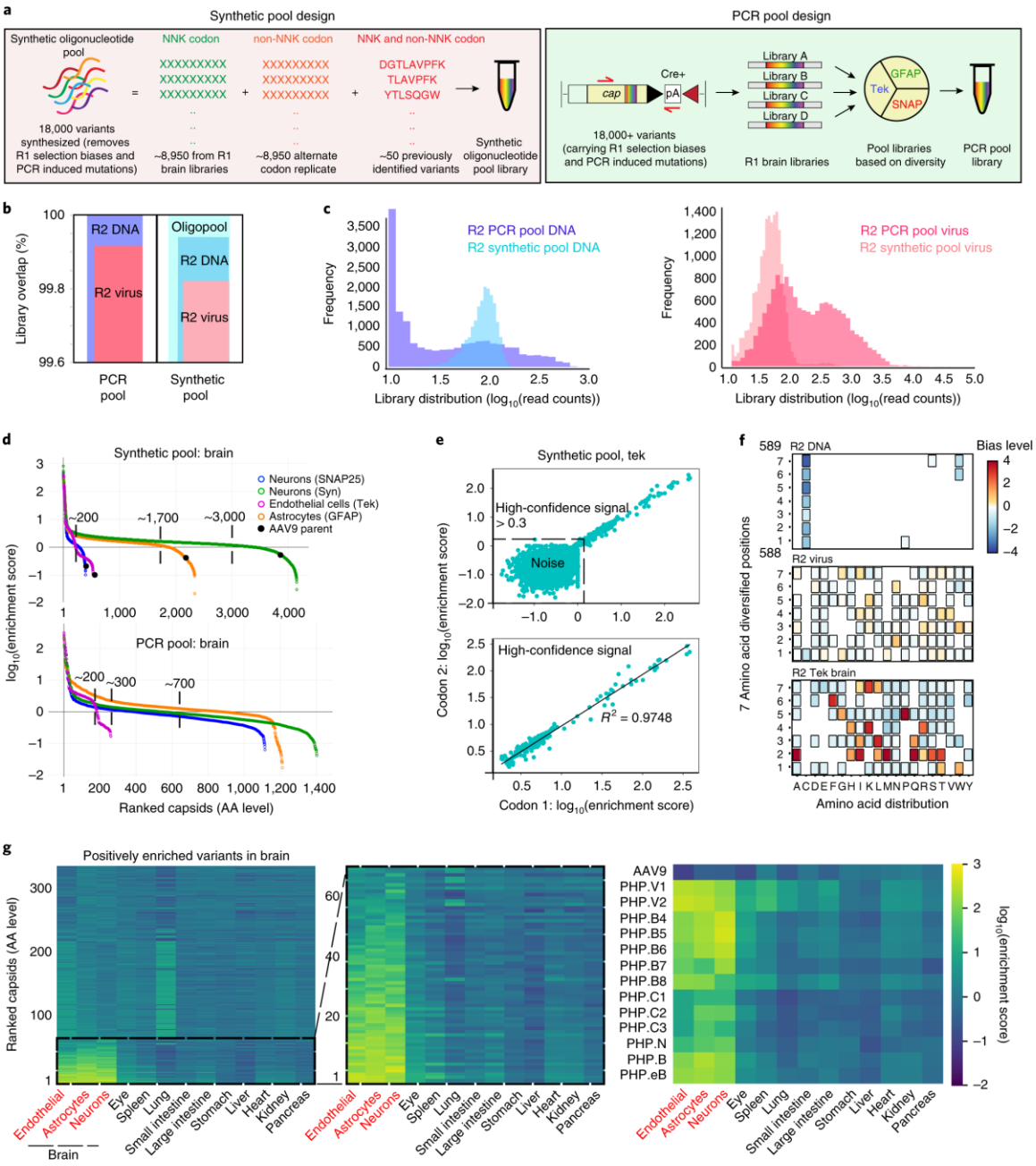


Figure 3.2: Round-2 capsid selections by synthetic pool and PCR pool methods.

(a), Schematic of R2 synthetic pool (left) and PCR pool (right) library design. **(b)**, Overlapping bar chart showing the percentage of library overlap between the mentioned libraries and their theoretical composition. **(c)**, Histograms of DNA and virus libraries from the two methods, where the variants in a library are binned by their read counts (in $\log_{10}$ scale) and the height of the histogram is proportional to their frequency. **(d)**, Distributions of R2 brain libraries from all Cre transgenic lines ($n = 2$ mice per Cre Line,

mean is plotted) and both methods, in which the libraries are sorted in decreasing order of enrichment score ($\log_{10}$ scale). The total number of positively enriched variants from these libraries are highlighted by dotted straight lines and AAV9's relative enrichment is mapped on the synthetic pool plot. **(e)**, Comparison of the enrichment scores ($\log_{10}$ scale) of two alternate codon replicates for 8,462 variants from the Tek-Cre brain library ($n = 2$ mice, mean is plotted). The broken line separates the high-confidence signal (>0.3) from noise. For the high-confidence signal (below), a linear least-squares regression is determined between the two codons and the regression line (best fit). The coefficient of determination r^2 is shown. **(f)**, Heat maps representing the magnitude ($\log_2$(fold change)) of a given amino acid's relative enrichment or depletion at each position given statistical significance is reached (boxed if $P \leq 0.0001$, two-sided, two-proportion *z*-test, *P* values corrected for multiple comparisons using Bonferroni correction). R2 DNA normalized to oligopool (top, ~9,000 sequences), R2 virus normalized to R2 DNA (middle, n = ~9,000 sequences), R2 Tek brain library with enrichment over 0.3 (high-confidence signal) from synthetic pool method normalized to R2 virus (bottom, 154 sequences) are shown ($n = 2$ for brain library, one per mouse; all other libraries, $n = 1$). **(g)**, Heat map of Cre-independent relative enrichment across organs ($n = 2$ mice per Cre line, mean across 6 samples from 3 Cre lines is plotted) for variants enriched in the brain tissue of at least one Cre-dependent synthetic pool selection (red text, $n = 2$ mice per cell-type, mean is plotted) (left). Zoom-in of the most CNS-enriched variants (middle), and of the variants that are characterized in the current study along with spike-in library controls (right) are shown.

The synthetic pool library design comprised: (1) equimolar amounts of ~8,950 capsid variants present at high read counts in at least one of the R1 selections from brain and spinal cord (***Supplementary Figure 3.S1e***); (2) alternative codon replicates of those ~8,950 variants (optimized for mammalian codons) to reduce false positives; and (3) a *spike-in* library of controls (***Supplementary Dataset 3.D1***), resulting in a total library size of 18,000 nucleotide variants.

AAV9 and ~50 additional variants (and their alternative codon duplicates), identified in previously described work[133,157] (includes well characterized variants like AAV-PHP.B or AAV-PHP.eB as well as many variants identified using the previous methodology but uncharacterized *in vivo*) act as internal selection controls and standards for the relative performance of the new variants (see ***Supplementary Dataset 3.D1***). The *PCR pool* library was generated by pooling the amplicons from the R1 brain selections after normalization based on their relative diversity (see *Methods*).

Both round-2 (R2) virus libraries produced a high titer (~6×10^{11} vg per 10 ng of R2 DNA library per 150-mm dish; ***Supplementary Figure 3.S2a***), and ~99% of variants of the R2 DNA were found after viral production (***Figure 3.2b***). However, the distribution of the DNA and virus libraries from both designs differed notably. The PCR pool library carries forward the R1 selection biases (***Figure 3.2c*** and ***Supplementary Figure 3.S2b,c***) where the abundance reflects prior enrichment across tissues in R1 as well as bias from viral production and sample mixing. Comparatively, the synthetic pool DNA library is more evenly distributed, minimizing bias amplification across selection rounds.

For *in vivo* selection, we intravenously administered a dose of 1×10^{12} vg per adult transgenic mouse into three of the previously used Cre lines ($n = 2$ mice per Cre transgenic line, GFAP, SNAP25, Tek), as well as the Syn-Cre line (for neurons). Two weeks after i.v. injection, we extracted, selectively amplified and deep-sequenced rAAV genomes from brain samples (as in R1). The synthetic pool library produced a greater number of enriched capsid variants than the PCR pool brain library (for example, ~1,700 versus ~700 variants per tissue library at the AA level in GFAP-Cre mice) (***Figure 3.2d*** and ***Supplementary Figure 3.S2d***). In the synthetic pool, ~90% of the variants from the spike-in library were enriched (***Supplementary Figure 3.S2d***, middle panel, and ***Supplementary Dataset 3.D1***).

The degree of correlation between variant enrichment scores for PCR and synthetic pool libraries varies in each Cre transgenic line, indicating the presence of noise within experiments (***Supplementary Figure 3.S2e***).

Experimental noise

The degree of enrichment at which correlation breaks down appears to vary with Cre-line. A downside of *PCR pool* is that there is no way to tell whether it or *synthetic pool* is the more ‘true’ enrichment score, or even that there may be cause for concern regarding certain enrichment values. The correlation among positively enriched variants between the two methods were found to improve with the magnitude of positive enrichment. For each

experiment there is a level of enrichment below which the scores become irreproducible, or noisy.

Supplementary Figure 3.S3a demonstrates that neither PCR pool nor Synthetic pool is inherently more 'true' at lower enrichment scores. This is because Synthetic pool methodology with its codon replicates has a self-contained control to determine an enrichment level below which enrichment value has no further predictive power. In the main text we use the term 'noise' to refer to regions of enrichment in a particular experiment below which values lose their reproducibility and predictive power. Being able to experimentally determine enrichment signal above noise allows researchers to focus their attention and data analyses on enrichment levels that are internally reproducible and thereby avoid selecting false positive variants or drawing invalid conclusions.

Thus, if one is interested in only the highest enriched variants for a particular tissue, *PCR pool* design coupled with enrichment normalization to virus library may not drastically differ from *synthetic pool* design over one additional round of selection for a subset of *in vivo* selections (such as Tek-Cre or SNAP-Cre). Without additional validation, however, it is difficult to predict whether a given *in vivo* system will perform akin to Tek-Cre. This becomes critical in a multiplexed selection study where target-specific variants may not garner the highest enrichments in one particular *in vivo* selection.

The synthetic pool's codon-replicate feature addresses this predicament by pinpointing the level of enrichment needed within each selection to rise above noise (***Figure 3.2e*** and ***Supplementary Figure 3.S3a,b***). This is a substantial advantage over the PCR pool design, allowing us to confidently interpret enrichment scores in a given selection.

Analysis of capsid libraries after round-2 selections

Whereas the amino acid distribution of the DNA library closely matched the Oligopool design, virus production selected for a motif within the hepta AA diversified

insertion (between AA 588 and AA 589), with Asn at position 2, β-branched amino acids (I, T, V) at position 4 and positively charged amino acids (K, R) at position 5 (***Figure 3.2f*** and ***Supplementary Figure 3.S3c***). Fitness for BBB crossing resulted in a different pattern. For instance, variants highly enriched after recovery from brain tissue (across all Cre lines) shared preferences for Pro in position 5, and Phe in position 6.

By assessing enrichment score reproducibility within the synthetic pool design, we next determined the brain enriched variants' distribution across peripheral organs (***Figure 3.2g***, left). About 60 variants that are highly enriched in brain are comparatively depleted across other organs (***Figure 3.2g***, middle). Encouraged by the expected behavior of spike-in control variants (AAV9, PHP.B, PHP.eB), we chose eleven additional variants for further validation (***Figure 3.2g***, right), including several that would have been overlooked if the choice had been based on PCR pool or CREATE (***Table 3.T2***).

AAV Variants	***Synthetic pool* enrichment rank**	***PCR pool* enrichment rank**	***PCR pool* read count rank**
PHP.V1	1	4	3
PHP.V2	2	1	1
PHP.B4	4	10	56
PHP.B7	6	13	36
PHP.B8	3	7	23
PHP.C1	13	34	74
PHP.C2	12	20	293
PHP.C3	16	Not recovered	Not recovered

Table 3.T2: Ranking of AAV-PHP capsids across methods.

Ranks of selected variants among all capsids recovered from R2 Tek-Cre selection by *synthetic pool* enrichment score (representing M-CREATE), *PCR pool* enrichment score (representing closer to M-CREATE), or *PCR pool* read counts (representing CREATE), the highest ranks of which starts from 1, and “Not recovered” represent absence of the variant from R2 sequencing data.

We chose these variants due to their enrichments and where they fall in sequence space. We noticed that the enriched variants cluster into distinct families based on sequence similarity. The most enriched variants form a distinct family across selections with a common motif: T in position 1, L in position 2, P in positive 5, F in position 6 and K or L in position 7 (***Figure 3.3a*** and ***Supplementary Figure 3.S3d***). This amino acid pattern closely resembles the TLAVPFK motif in the previously identified variant AAV-PHP.B[157]. Given the sequence similarity among members of this family, we next tested whether selected variants can cross the BBB and target the CNS with similar efficiency and tropism.

The dominance of PHP.B-like motif

The ability to twice recover the AAV-PHP.B sequence family from completely independently constructed and selected libraries confirms that the viral library's sequence space coverage was broad enough to recover a family of variants sharing a common motif. Unlike CREATE which identified only one variant, AAV-PHP.B, M-CREATE yielded a diverse PHP.B-like family that hints toward important chemical features of this motif. The sequence diversity within this family suggests that isolating AAV-PHP.B was not simply good fortune in our prior study (considering a theoretical starting library size of ~1.3 billion), and that this is a dominant family for this particular experiment.

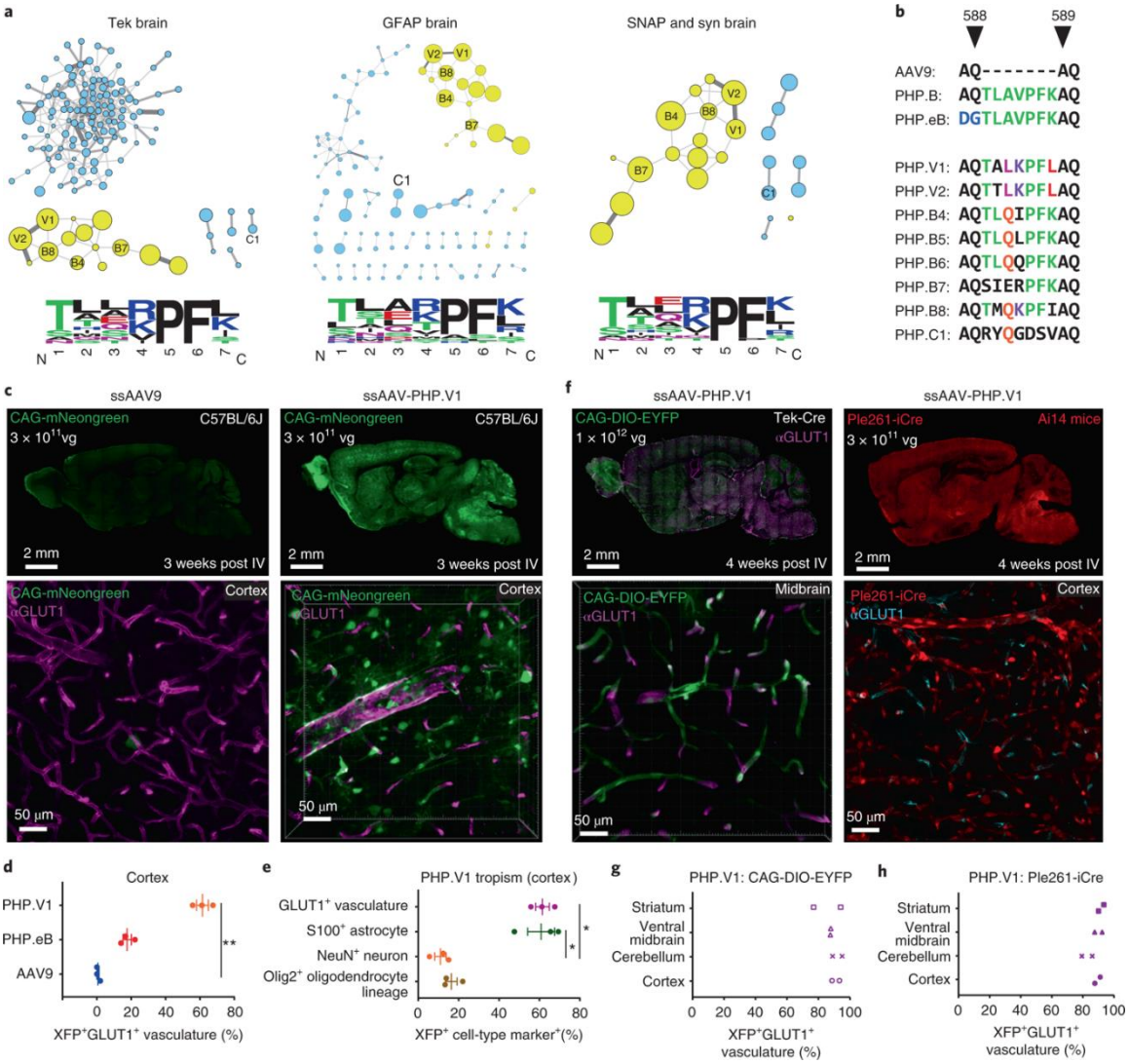


Figure 3.3: Selected AAV capsids form sequence families and include variants for brain-wide transduction of vasculature.

(a), Clustering analysis of variants from synthetic pool brain libraries after enrichment in Tek-Cre (left), GFAP-Cre (middle) and combined SNAP-Cre and Syn-Cre (right) selections. The size of the nodes represents relative enrichment in the brain. Thickness of the edges (connecting lines) represents the degree of relatedness. Distinct families (yellow) with the corresponding AA frequency logos (AA size represents prevalence and color encodes AA properties) are shown. **(b)**, The hepta AA insertion peptide sequences of AAV-PHP variants between AA positions 588–589 of AAV9 capsid are shown. AAs are colored by shared identity to AAV-PHP.B and eB (green) or among new variants (unique color per position). **(c)**, AAV9 (left) and AAV-PHP.V1 (right) mediated expression using ssAAV:CAG-mNeongreen genome (green, $n = 3$, 3 weeks of expression in C57BL/6J adult mice with 3×10^{11} vg i.v. dose per mouse, imaged under the same settings) in sagittal

sections of brain (top) with higher-magnification image from cortex (bottom). Magenta, αGLUT1 antibody staining for vasculature. **(d)**, Percentage of vasculature stained with αGLUT1 that overlaps with mNeongreen (XFP) expression in cortex. One-way analysis of variance (ANOVA) non-parametric Kruskal–Wallis test ($P = 0.0036$), and follow-up multiple comparisons using uncorrected Dunn's test ($P = 0.0070$ for AAV9 versus PHP.V1) are reported. $^{**}P \leq 0.01$ is shown, $P > 0.05$ is not shown; data are mean ± s.e.m., $n = 3$ mice per AAV variant, cells quantified from 2–4 images per mouse per cell type. **(e)**, Percentage of cells stained with each cell-type specific marker (αGLUT1, αS100 for astrocytes, αNeuN for neurons, and αOlig2 for oligodendrocyte lineage cells) that overlaps with mNeongreen (XFP) expression in cortex. Kruskal–Wallis test ($P = 0.0078$), and uncorrected Dunn's test ($P = 0.0235$ for neuron versus vascular cells, and 0.0174 for neuron versus astrocyte) are reported. $^{*}P \leq 0.05$ is shown, and $P > 0.05$ is not shown; data are mean ± s.e.m., $n = 3$ mice, cells quantified from 2–4 images per mouse per cell type. **(f)**, Vascular transduction by ssAAV-PHP.V1:CAG-DIO-EYFP in Tek-Cre adult mice (left) ($n = 2$, 4 weeks of expression, 1×10^{12} vg i.v. dose per mouse), and by ssAAV-PHP.V1:Ple261-iCre in Ai14 reporter mice (right) ($n = 2$, 3 weeks of expression, 3×10^{11} vg i.v. dose per mouse). Tissues are stained with αGLUT1 (magenta (left) and cyan (right)). **(g)**, Efficiency of vascular transduction (as described in (**d)**) in Tek-Cre mice ($n = 2$, mean from 3 images per mouse per brain region). **(h)**, Efficiency of vascular transduction in Ai14 mice ($n = 2$, a mean from 4 images per mouse per brain region).

AAV9 variants with enhanced BBB entry and CNS transduction

Given the dominance of the PHP.B family in the R2 selection, we characterized its most enriched member, harboring a TALKPFL motif and henceforth referred to as AAV-PHP.V1 (***Figure 3.3a,b***). Despite its sequence similarity to AAV.PHP.B, the tropism of AAV-PHP.V1 is biased toward transducing brain vascular cells (***Figure 3.3c*** and ***Supplementary Figure 3.S4a***). When delivered intravenously, AAV-PHP.V1 carrying a fluorescent reporter under the control of the ubiquitous CAG promoter transduces ~60% of $GLUT1^{+}$ cortical brain vasculature, compared with ~20% with AAV-PHP.eB and almost no transduction with AAV9 (***Figure 3.3c,d***). In addition to the vasculature, AAV-PHP.V1 also transduced ~60% of cortical $S100^{+}$ astrocytes (***Figure 3.3e***). However, AAV-PHP.V1 is not as efficient for astrocyte transduction as the previously reported AAV-PHP.eB (when packaged with an astrocyte specific GfABC1D promoter[132], ***Supplementary Figure 3.S4b***).

For applications requiring endothelial-cell-restricted transduction via i.v. delivery, AAV-PHP.V1 vectors can be used in three different systems: (1) in endothelial-cell-type specific Tek-Cre[246] mice with a Cre-dependent expression vector (***Figure 3.3f*** (left),***g*** and ***Supplementary Video 3.V1***), (2) in fluorescent reporter mice where Cre is delivered with an endothelial-cell-type specific MiniPromoter (Ple261)[129] (***Figure 3.3f*** (right),***h*** and ***Supplementary Figure 3.S4c–e***) and (3) in wild-type mice by packaging a self-complementary genome (scAAV) containing a ubiquitous promoter (***Supplementary Figure 3.4f***). The mechanism of endothelial-cell-specific transduction by AAV-PHP.V1 using scAAV genomes is unclear, but shifts in vector tropism when packaging scAAV genomes have been reported for another capsid[160].

Given the difference in tropism between AAV-PHP.V1 and AAV-PHP.B or AAV-PHP.eB, we characterized several additional variants within the PHP.B-like family. One variant, AAV-PHP.V2, harboring the TTLKPFL 7-mer sequence and differing by only one amino acid from AAV-PHP.V1, has a similar tropism (***Supplementary Figure 3.S5***).

AAV-PHP.V2 capsid

AAV-PHP.V2 was found at high abundance in R1 selection across all brain libraries and was highly enriched in R2 (***Figure 3.1d***, ***3.2g*** (right panel), ***3.3a,b, Supplementary Figure 3.S1e***). Given its sequence similarity, we predicted similar tropism to that of AAV-PHP.V1. We validated this *in vivo* in C57BL/6J adult mice (ssAAV-PHP.V2:CAG-mNeongreen genome, $3x10^{11}$ vg dose per adult mice, n = 3, ***Supplementary Figure 3.5a***), in Tek-Cre mice (ssAAV-PHP.V2:CAG-DIO-EYFP genome, 1×10^{12} vg dose per adult mouse, n = 2, ***Supplementary Figure 3.S5b***), and in GFAP-Cre mice (ssAAV-PHP.V2:CAG-DIO-EYFP, 1×10^{12} vg dose per adult mouse, n = 2, ***Supplementary Figure 3.S5c***).

Three other variants with sequences of roughly equal deviation from both AAV.PHP.V1 and AAV.PHP.B, AAV-PHP.B4 with TLQIPFK, AAV-PHP.B7 with SIERPFK and AAV-PHP.B8 with TMQKPFI (***Figures 3.3a,b*** and ***3.4a,b***) have PHP.B-

like tropism with biased transduction toward neurons and astrocytes (***Figure 3.4b*** and ***Supplementary Figure 3.S6a–c***). Similar variants among the spike-in library, AAV-PHP.B5 with TLQLPFK and AAV-PHP.B6 with TLQQPFK, also shared this tropism (***Figures 3.3b*** and ***3.4a,b, Supplementary Figure 3.S6a***).

To evaluate the performance of the *spike-in* library, we chose two highly enriched variants similarly placed in sequence space: AAV-PHP.B6 – TLQLPFK and AAV-PHP.B7 – TLQQPFK (***Supplementary Figure 3.S2d*** (middle panel), ***3.S3d***) that were previously identified in the *3-mer-s* PHP.B library[133] but never validated *in vivo*. At a modest dose of $1x10^{11}$ vg in C57BL/6J adult mice, these variants also display PHP.B-like tropism (***Figure 3.4a,b, Supplementary Figure 3.S6a***).

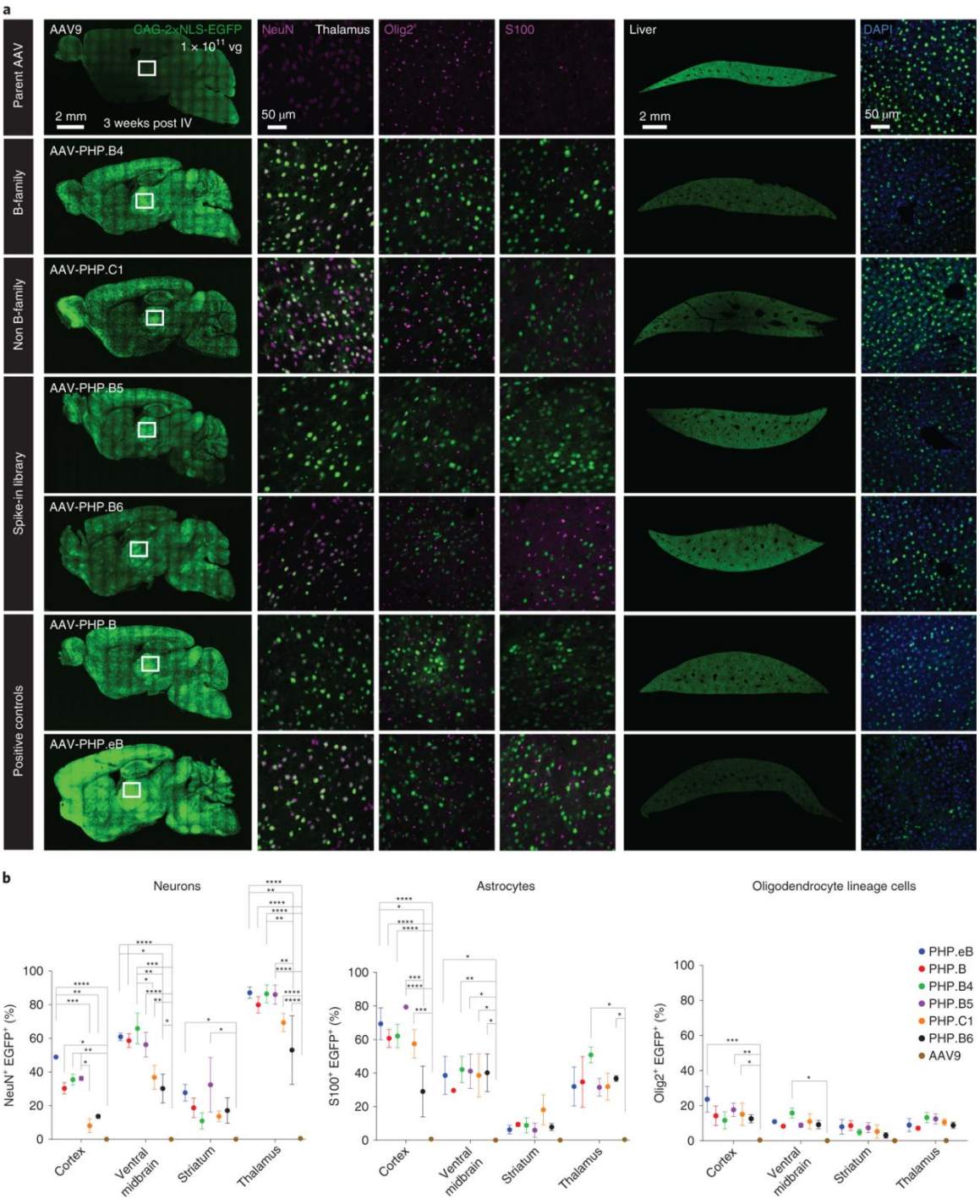


Figure 3.4: Characterization of round-2 brain libraries and identification of capsids with broad CNS tropism.

(a), Transduction by AAV-PHP.B4-B6 and C1 variants, as well as B, eB and AAV9 controls in sagittal brain and liver sections (each column was imaged under the same

settings). White box, thalamus (this is not the precise region of the figures to the right). Vectors are packaged with ssAAV:CAG-2xNLS-EGFP genome ($n = 3$ per group, 1×10^{11} vg i.v. dose per adult C57BL/6J mouse, 3 weeks of expression). Tissues are stained with cell-type specific markers (magenta): αNeuN for neurons, αS100 for astrocytes and αOlig2 for oligodendrocyte lineage cells. Liver tissues are stained with DAPI (blue). **(b)**, The percentage of $\alpha NeuN^+$, $\alpha S100^+$ and $\alpha Olig2^+$ cells with detectable nuclear-localized EGFP in the indicated brain regions are shown ($n = 3$ per group, 1×10^{11} vg dose). A two-way ANOVA with correction for multiple comparisons using Tukey's test is reported with adjusted *P* values (****$P \leq 0.0001$, ***$P \leq 0.001$, **$P \leq 0.01$, *$P \leq 0.05$, is shown, and $P > 0.05$ is not shown on the plot; 95% confidence interval (CI), data are mean ± s.e.m. The dataset comprises a mean of two images per region per cell-type marker per mouse).

We next investigated a series of variants selected to verify M-CREATE's predictive power outside this family. A highly enriched variant with an unrelated sequence, AAV-PHP.C1 harboring RYQGDSV (***Figures 3.3a,b*** and ***3.4a,b***), transduced astrocytes at a similar efficiency and neurons at lower efficiency compared to other tested variants from the B family (***Figure 3.4b***). Two variants found in high abundance in the R2 synthetic pool virus library and underrepresented in brain (with both codon replicates in agreement), AAV-PHP.X1 with ARQMDLS and AAV-PHP.X2 with TNKVGNI (***Supplementary Figure 3.S2b***, right), poorly transduced the CNS (***Supplementary Figure 3.S6b***). Two variants that we found in higher abundance in brain libraries from the PCR pool R2, AAV-PHP.X3 with QNVTKGV and AAV-PHP.X4 with LNAIKNI also failed to outperform AAV9 in the brain (***Supplementary Figure 3.S6d***).

Collectively, our characterization of these AAV variants suggests several key points. First, within a diverse sequence family, there is room for both functional redundancy and the emergence of alternative tropisms. Second, highly enriched sequences outside the dominant family are also likely to possess enhanced function. Third, buoyed by codon replicate agreement in the synthetic pool, a variant's enrichment across tissues may be predictive. Fourth, while the synthetic pool R2 library contains a subset of the sequences that are in the PCR pool R2 and may thereby lack some enhanced variants, those variants found exclusively within the PCR pool library are more likely to be false positives.

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

An AAV9 variant that specifically transduces neurons

Using NGS, we re-investigated a *3-mer-s* (s for substitution) PHP.B library generated by the prior CREATE methodology and that yielded AAV-PHP.eB[133] (***Figure 3.5a***). Briefly, the re-investigated *3-mer-s* PHP.B library diversified positions 587-597 of the AAV-PHP.B capsid (equivalent of 587-590 AA on AAV9) in portions of three consecutive AAs, (~40,000 total variants) (***Figure 3.4a***). Selections were performed in three Cre-transgenic lines: Vglut2-IRES-Cre for glutamatergic neurons, Vgat-IRES-Cre for GABAergic neurons, and GFAP-Cre for astrocytes.

We deep sequenced the libraries recovered from brain (using Cre-dependent PCR) and a R2 library from the livers of wild-type mice (processed via PCR for all capsid sequences regardless of Cre-mediated inversion) and identified 150–200 capsids enriched in brain tissue (***Figure 3.5b*** and ***Supplementary Figure 3.S7a,b***).

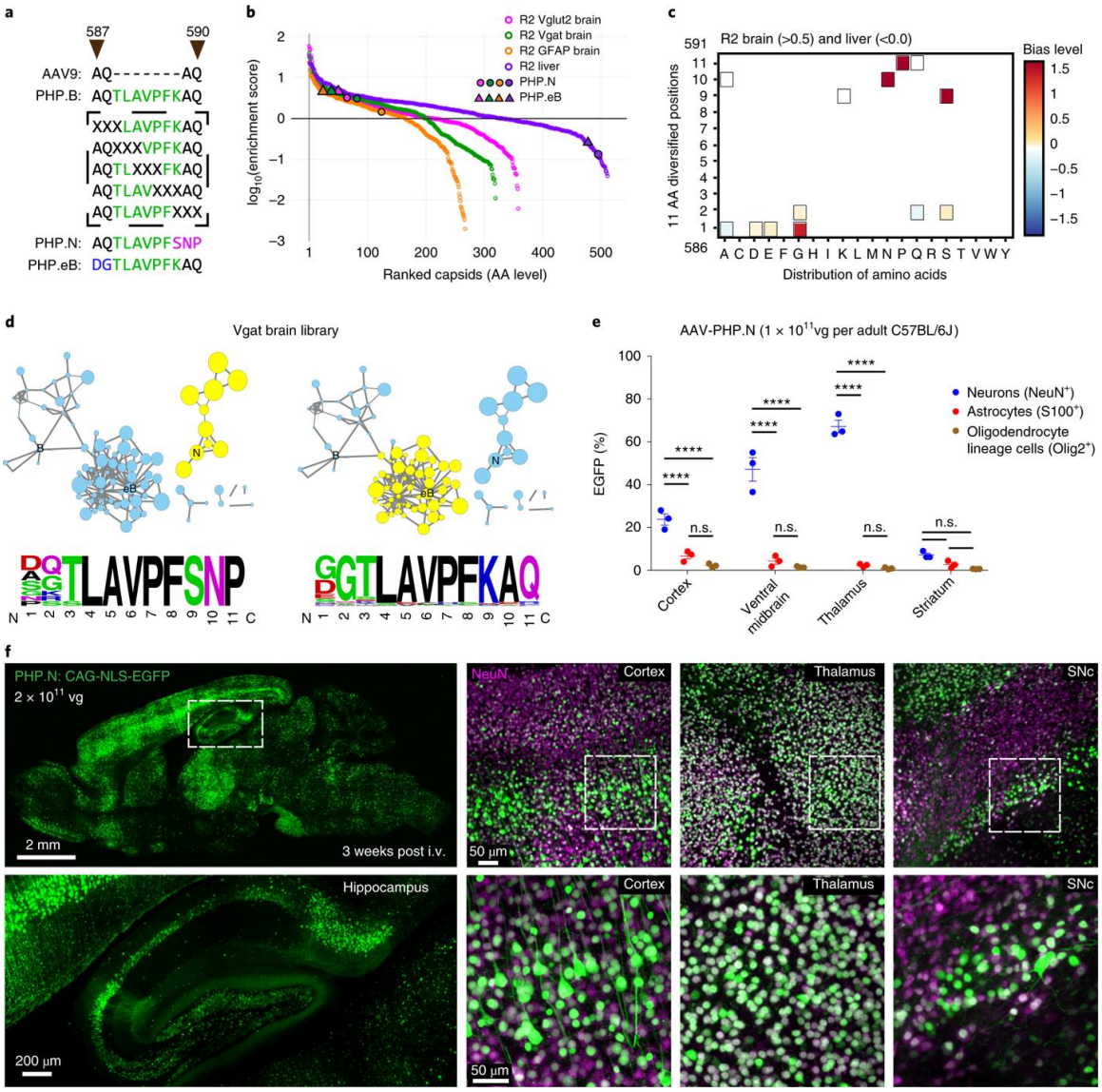


Figure 3.5: Recovery of AAV-PHP.B variants including one with high specificity for neurons.

(a), Design of the 3-mer-s PHP.B library with combinations of three AA diversification between AA 587–597 of AAV-PHP.B (corresponding to AA 587–590 of AAV9). Shared amion acid identity with the parent AAV-PHP.B (green) is shown along with unique motifs for AAV-PHP.N (pink) and AAV-PHP.eB (blue). **(b)**, Distributions of R2 brain and liver libraries (at the amino acid level) by enrichment score (normalized to R2 virus library, with variants sorted in decreasing order of enrichment score). The enrichment of AAV-PHP.eB and AAV-PHP.N across all libraries is mapped on the plot. **(c)**, Heat map represents the magnitude ($\log_2$(fold change)) of a given amino acid's relative enrichment or depletion at each position across the diversified region, only if statistical significance is reached on fold

change (boxed if $P \leq 0.0001$, two-sided, two-proportion *z*-test, *P* corrected for multiple comparisons using Bonferroni correction). Plot includes variants that were highly enriched in brain (>0.5 mean enrichment score, where mean is drawn across Vglut2, Vgat and GFAP, $n = 1$ library per mouse line (sample pooled from 2 mice per line)) and underrepresented in liver (<0.0) (32 amino acid sequences). **(d)**, Clustering analysis of enriched variants from Vgat brain library is shown. Node size represents the degree of depletion in liver. Thickness of edges (connecting lines) represents degree of relatedness between nodes. Two distinct families are highlighted in yellow and their corresponding amino acid frequency logos are shown below (amino acid size represents prevalence, and color encodes amino acid properties). **(e)**, The percentage of neurons, astrocytes and oligodendrocyte lineage cells with ssAAV-PHP.N:CAG-2xNLS-EGFP in the indicated brain regions is shown ($n = 3$, 1×10^{11} vg i.v. dose per adult C57BL/6J mouse, 3 weeks of expression, data is mean ± s.e.m., 6–8 images for cortex, thalamus and striatum, and 2 images for ventral midbrain, per mouse per cell-type marker using ×20 objective covering the entire regions). A two-way ANOVA with correction for multiple comparisons using Tukey's test gave adjusted *P* values reported as ****$P \leq 0.0001$, n.s. for $P > 0.05$, 95% CI. **(f)**, Transduction by ssAAV-PHP.N:CAG-NLS-EGFP ($n = 2$, 2×10^{11} vg i.v. dose per adult C57BL/6J mouse, 3 weeks of expression) is shown with NeuN staining (magenta) across three brain areas (cortex, SNc (substantia nigra pars compacta) and thalamus).

Variants that were enriched in brain and underrepresented in liver show a significant bias towards certain amino acids such as G, D and E at position 1; G and S at position 2 (which includes the AAV-PHP.eB motif, DG); and S, N and P at position 9, 10 and 11 (***Figure 3.5c*** and ***Supplementary Figure 3.S7c***; $P \leq 0.0001$, two-sided, two-proportion *z*-test, *P* values were corrected for multiple comparisons using Bonferroni correction). We clustered variants that were enriched in the brain according to their sequence similarities and ranked them by their underrepresentation in liver (represented by node size in clusters). A distinct family referred to as N emerged with the common motif SNP at positions 9–11 in the PHP.B backbone (***Figure 3.5d*** and ***Supplementary Figure 3.S7d***).

The core variant of the N-family cluster, with the AQTLAVPFSNP motif, was highly abundant in R1 and R2 selections, had higher enrichment score in Vglut2 and Vgat brain tissues compared to GFAP, and was underrepresented in liver tissue (***Figure 3.5b*** and ***Supplementary Figure 3.S7a–d***). Unlike AAV-PHP.eB, this variant (AAV-PHP.N) specifically transduced $NeuN^+$ neurons even when packaged with a ubiquitous CAG promoter, although the transduction efficiency varied across brain regions (from ~10–70%

in $NeuN^+$ neurons, including both $VGLUT1^+$ excitatory and $GAD1^+$ inhibitory neurons; ***Figure 3.5e,f*** and ***Supplementary Figure 3.S7e,f***).

Thus, by re-examining the *3-mer-s* library we identified several useful variants, including one with notable cell-type-specific tropism. While Vglut2-Cre and Vgat-Cre mice were used for *in* vivo selection, we didn't find variants that stood out for neuronal subtype-specific transduction of excitatory and inhibitory populations from our initial investigations on the NGS dataset. It is possible that a biological solution to this (stringent) selection was not present in this library.

Investigation of capsid families beyond the C57BL/6J mouse strain

The enhanced CNS tropism of AAV-PHP.eB and AAV-PHP.B relative to AAV9 is absent in a subset of mouse strains. Their CNS transduction is highly efficient in C57BL/6J, FVB/NCrl, DBA/2 and SJL/J, with intermediate enhancement in 129S1/SvimJ, and no apparent enhancement over AAV9 in BALB/cJ and several additional strains[134,164–168]. This pattern holds for the two variants from the PHP.B family that we characterized further, AAV-PHP.V1 and AAV-PHP.N (***Figure 3.6a*** and ***Table 3.T3***). These variants did not transduce the CNS in BALB/cJ, yet transduced the FVB/NJ strain (***Figure 3.6b***). AAV-PHP.V1 transduced human brain microvascular endothelial cell (HBMEC) culture, resulting in increased mean fluorescent intensity compared with that following AAV9 and AAV-PHP.eB transduction (***Supplementary Figure 3.S8a***) however, suggesting mechanistic complexity.

AAV Variants	Reference / Selection method	Tropism	Production	Rounds of evolution from parent capsid
PHP.B, B2, B3	Deverman et al, 2016 / CREATE	**B**road CNS transduction	Good	1 round from AAV9
PHP.A	Deverman et al, 2016 / CREATE	**A**strocyte transduction	Poor; prone to precipitate upon storage at 4ºC.	1 round from AAV9
PHP.eB	Chan et al, 2017 / CREATE	**E**nhanced **B**road CNS transduction	Good	2 rounds from AAV9 or 1 round from PHP.B
PHP.S	Chan et al, 2017 / CREATE	**S**ensory neuron transduction	Good	1 round from AAV9
PHP.V1, V2	Current study / M-CREATE	BBB **V**ascular cells and astrocytes transduction	Good	1 round from AAV9
PHP.B4, B7, B8,	Current study / M-CREATE	**B**road CNS transduction	Good	1 round from AAV9
PHP.B5, B6	Current study / M-CREATE and CREATE	**B**road CNS transduction	Good	2 rounds from AAV9 or 1 round from PHP.B
PHP.C1, C2, C3	Current study / M-CREATE	Broad **C**NS transduction across mouse strains	Good; PHP.C1 prone to precipitate upon storage at 4ºC.	1 round from AAV9
PHP.N	Current study / M-CREATE and CREATE	**N**euron transduction	Average	2 rounds from AAV9 or 1 round from PHP.B

PHP variants – named in memory of late Professor **Paul **H. P**atterson, Caltech.*

Table 3.T3: AAV-PHP vectors identified by CREATE and M-CREATE.

The table provides a summary of the variants that have been identified so far using CREATE and M-CREATE, along with their tropism and the evolutionary steps from the parent capsid that was involved in their discovery.

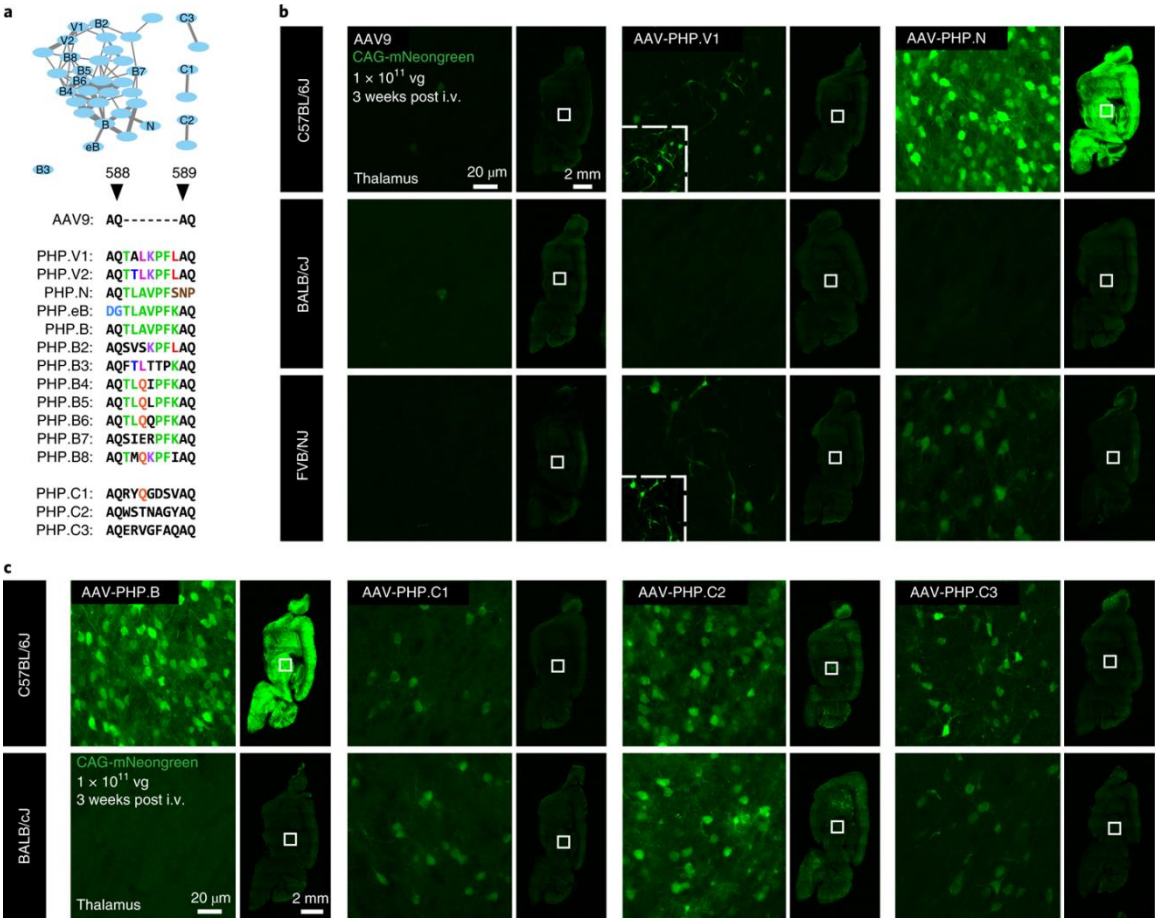


Figure 3.6: Tropism of variants from distinct families across mouse strains.

(a), Clustering analysis showing the brain-enriched sequence families of variants identified in prior studies (PHP.B-B3, PHP.eB) or in the current study (PHP.B4-B8, PHP.V1-2, PHP.C1-3). Thickness of edges (connecting lines) represents degree of relatedness between nodes. The amino acid sequences inserted between 588–589 (of AAV9 capsid) for all the variants discussed are shown below. **(b)**, Transduction of AAV9, AAV-PHP.V1 and AAV-PHP.N across the mouse strains C57BL/6J, BALB/cJ and FVB/NJ are shown in sagittal brain sections (right), along with a higher-magnification image of the thalamus brain region (left). **(c)**, Transduction by AAV-PHP.B, AAV-PHP.C1-C3 in C57BL/6J and BALB/cJ mice are shown in sagittal brain sections (right), along with a higher-magnification image of the thalamus brain region (left). **(b**,**c)**, White box, thalamus (this is not the precise area that is zoomed-in on the figure to the left). All sagittal sections and thalamus regions were acquired under same image settings. The insets in AAV-PHP.V1 are zoom-ins with enhanced brightness. The indicated capsids were used to package ssAAV:CAG-mNeongreen ($n = 2$–3 per group, 1×10^{11} vg i.v. dose per 6- to 8-week-old adult mouse, 3 weeks of expression. The data reported in **(b)** and **(c)** are from one experiment where all viruses were freshly prepared and titered in the same assay for dosage consistency. AAV-

PHP.C2 and AAV-PHP.C3 were further validated in an independent experiment for BALB/cJ, $n = 2$ per group).

Notably, M-CREATE revealed many non-PHP.B-like sequence families that enriched through selection for transduction of cells in the CNS. We tested the previously mentioned AAV-PHP.C1 (RYQGDSV), as well as AAV-PHP.C2 (WSTNAGY), and AAV-PHP.C3 (ERVGFAQ) (***Figure 3.6a***). These showed enhanced BBB crossing irrespective of mouse strain, with roughly equal CNS transduction in BALB/cJ and C57BL/6J (***Figure 3.6c*** and ***Supplementary Figure 3.S8b***). Collectively, these studies suggest that M-CREATE is capable of finding capsid variants with diverse mechanisms of BBB entry that do not exhibit strain specificity.

3.4 DISCUSSION

This work outlines the development and validation of the M-CREATE platform for multiplexed viral capsid selection. M-CREATE incorporates multiple internal controls to monitor sequence progression, minimize bias and accelerate the discovery of capsid variants with useful tropisms. Utilizing M-CREATE, we have identified both individual capsids and distinct families of capsids that are biased toward different cell-types of the adult brain when delivered intravenously. The outcome from 7-mer-i selection demonstrates the possibility of finding AAV capsids with improved efficiency and specificity towards one or more cell types. Patterns of CNS infectivity across mouse strains suggest that M-CREATE may also identify capsids with distinct mechanisms of BBB crossing. With additional rounds of evolution as shown in the *3-mer-s* selection, the specificity or efficiency of 7-mer-i library variants may be improved, as was observed with AAV-PHP.N or AAV-PHP.eB[133].

We believe that the variants tested *in vivo* and their families will find broad application in neuroscience, including studies involving the BBB[247], neural circuits[248], neuropathologies[249], and therapeutics[99]. AAV-PHP.V1 or AAV-PHP.N are well-suited for studies requiring gene delivery for optogenetic or chemogenetic manipulations[86], or in rare

monogenic disorders (targeting brain endothelial cells, for example GLUT1-deficiency syndrome, NLS1-microcephaly[249], or targeting neurons, for example mucopolysaccharidosis type IIIC[244]).

The outcomes from our experiments employing M-CREATE opens several promising lines of inquiry, such as the assessment of identified capsid families across species, the investigation of the mechanistic properties that underlie the ability to cross specific barriers (such as the BBB) or target specific cell populations and further evolution of the identified variants for improved efficiency and specificity. In addition, the datasets generated by M-CREATE could be used as training sets for in silico selection by machine-learning models. M-CREATE is presently limited by the low throughput of vector characterization *in vivo*; however, RNA-sequencing technologies[250] offer hope in this regard. In summary, M-CREATE will serve as a next-generation capsid-selection platform that can open directions in vector engineering and potentially broaden the AAV toolbox for various applications in science and in therapeutics.

3.5 METHODS

Plasmids

Library generation

The rAAV-ΔCap-in-cis-Lox2 plasmid (***Supplementary Figure 3.S1a***, plasmid available upon request at Caltech CLOVER Center) is a modification of the rAAV-ΔCap-in-cis-Lox plasmid[157]. For *7-mer-i* library fragment generation, we used the pCRII-9Cap-XE plasmid[157] as a template. The AAV2/9 REP-AAP-ΔCap plasmid (***Supplementary Figure 3.S1a***, plasmid available upon request at Caltech CLOVER Center) was modified from the AAV2/9 REP-AAP plasmid[157].

The rAAV-ΔCap-in-cis-Lox2 plasmid consists of three major elements that are flanked by AAV2 ITRs.

(i) UBC ubiquitous promoter driving the expression of fluorescent protein, mNeongreen, followed by a synthetic polyadenylation sequence. The mCherry expression cassette of the previous version of the plasmid was replaced by mNeonGreen cassette.

(ii) A portion of AAV2 rep gene that has the splicing sequences and AAV5 p41 promoter (1680-1974 residues of GenBank AF085716.1) followed by AAV9 cap gene. The prior version of this plasmid, rAAV-ΔCap-in-cis-Lox, has a short 12 bp sequence between restriction sites XbaI and AgeI at AA 450 and 592 of the AAV9 Cap gene. This was replaced by a 723 bp sequence of mRuby2 gene in-frame (acts as filler DNA) in the newer version of the plasmid.

(iii) SV40 polyadenylation sequence that is flanked by lox71 and lox66 sites. The minor changes were introduced to the prior version of the plasmid to facilitate ease of cloning and to visualize mammalian cell transfection. The Lox sites in these rAAV plasmids show modest levels of Cre-independent flipping. This was minimized during PCR-based capsid recovery by lowering the number of amplification cycles to a point where we cannot recover any rAAV capsids from the control DNA extracted from wild-type mice (i.e., lacking Cre expression) that were injected with the library.

The pCRII-9Cap-XE plasmid contains the AAV9 capsid gene sequence from AAs 450-592 and is flanked by XbaI and AgeI restriction sites.

The AAV2/9 REP-AAP-ΔCap plasmid has the five previously existing stop codons of AAV2/9 REP-AAP in addition to the deletion of AAs 450-592 of the AAV9 capsid sequence. These modifications did not affect vector production. The deletion of the overlapping fragment between the REP-AAP and rAAV-ΔCap-in-cis-Lox2 plasmids minimizes recombination between plasmids that could potentially generate AAV9 wild-type capsids during co-transfection in vector production.

Capsid characterization

AAV capsids

The AAV capsid variants with heptamer insertions or 11-mer substitutions were made between AA positions 587–597 of AAV-PHP.B capsid using the pUCmini-iCAP-PHP.B backbone[157] (Addgene ID: 103002).

ssAAV genomes

To characterize the AAV capsid variants, we used the single-stranded (ss) rAAV genomes. We used genomes such as pAAV:CAG-mNeonGreen[133] (Addgene ID: 99134104055), pAAV:CAG-NLS-EGFP[157] (equivalent version with one NLS is on Addgene ID: 104061), pAAV:CAG-DIO-EYFP[134] (Addgene ID: 104052), pAAV: GfABC1D-2xNLS-mTurquoise2[134] (Addgene ID: 104053) and pAAV:-Ple261-iCre[129] (Addgene ID: 49113).

pAAV:CAG-mNeonGreen[133] genome consists of a ubiquitous CMV-β-Actin-intron-β-Globin (CAG) hybrid promoter driving the expression of a fluorescent protein, mNeonGreen (equivalent plasmid, pAAV: CAG-eYFP[251]; Addgene ID: 104055). pAAV:CAG-NLS-EGFP[157] consists of NLS sequences at the N- and C-termini of EGFP and is driven by the CAG promoter. An equivalent version with one NLS is on Addgene (ID 104061). pAAV:CAG-DIO-EYFP[251] (Addgene ID: 104052) consists of a EYFP gene built in the reverse direction of the CAG promoter, and it is flanked by a pair of Cre-Lox sites (Lox P and Lox 2272) on either ends. In cells expressing Cre, the Cre-lox pair inverts EYFP enabling transcription and translation, followed by excision in the lox site to prevent re-inversion. pAAV: GfABC1D-2xNLS-mTurquoise2[251], referred to elsewhere as pAAV:GFAP-2xNLS-mTurquoise2 (Addgene ID: 104053), consists of NLS sequences at the N- and C-termini of mTurquoise2 and is driven by the astrocyte-specific promoter GfABC1D[132]. pAAV:Ple261-iCre[129] (Addgene ID 49113) contains an endothelial-cell-specific promoter driving the expression of iCre.

We packaged pAAV:CAG-XFP (mNeongreen) for characterizing AAV variants. However, when performing quantification of cell-types neurons, astrocytes, and oligodendrocytes, we use CAG-NLS-EGFP to restrict the expression to nucleus for easier

quantification using microscope images. GFAP-NLS-mTurq2 is used to quantify astrocytes. CAG-DIO-EYFP is used for Cre driver lines, due to the presence of lox sites in this plasmid.

scAAV genomes

To characterize the AAV capsid variant, AAV-PHP.V1, using self-complementary (sc) rAAV genomes, we used scAAV genomes from different sources. scAAV:CB6-EGFP was a gift from G. Gao (University of Massachusetts Medical School) and scAAV:CAG-EGFP[252] from Addgene (Addgene ID: 83279).

The self-complementary genome from Dr. Guangping Gao, scAAV:CB6-EGFP genome has a hybrid ubiquitous CB6 promoter (975 bp) comprising a CMV enhancer (cytomegalovirus immediate early enhancer), a chicken-β-actin promoter and hybrid intron, that drives the expression of EGFP. The genome has a rabbit globin poly A (127 bp) following the EGFP gene. The scAAV:CAG-EGFP[252] (Addgene ID:83279), vector uses a ubiquitous CMV-β-Actin-intron-β-Globin (CAG) hybrid promoter to drive the expression of EGFP.

AAV capsid library generation

Round-1 AAV capsid DNA library

Mutagenesis strategy

The randomized (21-base) heptamer insertion was designed using the NNK saturation mutagenesis strategy, involving degenerate primers containing mixed bases (Integrated DNA Technologies). N can be an A, C, G or T base, and K can be G or T. Using this strategy, we obtained combinations of all 20 amino acids at each position of the heptamer peptide using 33 codons, resulting in a theoretical library size of 1.28 billion at the level of amino acid combinations. The mutagenesis strategy for the 3-mer-s PHP.B library is described in our prior work[133].

Library cloning

The 480-bp AAV capsid fragment (450–592 amino acids) with the randomized heptamer insertion between amino acids 588 and 589 was generated by conventional PCR methods using the pCRII-9Cap-XE[157] template by Q5 Hot Start High-Fidelity 2X Master Mix (NEB; M0494S) with forward primer, XF, and reverse primer, 7xMNN-588i (***Table 3.T4***).

To avoid PCR-induced biases resulting from point mutations, recombination, and template switching, PCR amplification of the library was limited to 15 – 20 cycles and the reactions were scaled up to get the required yield. The resulting PCR products were run on a 1% agarose gel and extracted with a Zymoclean Gel DNA Recovery kit (Zymo Research; D4007). It is critical to avoid AAV contamination during this step by taking precautionary measures like using a clean gel-running box and freshly prepared 1× TAE buffer.

The rAAV-ΔCap-in-cis-Lox2 plasmid (6,960 bp) was linearized with the restriction enzymes AgeI and XbaI, and the amplified library fragment was assembled into the linearized vector at 1:2 molar ratio using the NEBuilder HiFi DNA Assembly Master Mix (NEB; E2621S) by following the NEB recommended protocol.

Primer name	**Sequence (5'-3')**	**Forward or reverse direction**
XF	ACTCATCGACCAATACTTGTACTATCTCTCTAGAAC	Forward
7xMNN-588i	GTATTCCTTGGTTTTGAACCCAACCGGTCTGCGCCTGTGCMNNMNNMNNMNNMNNMNNMNNTTGGGCACTCTGGTGGTTTGTG	Reverse
588-R2lib-F	CACTCATCGACCAATACTTGTACTATCTCTCT	Forward
588-R2lib-R	GTATTCCTTGGTTTTGAACCCAACCG	Reverse
11-mer-588i	GTATTCCTTGGTTTTGAACCCAACCGGTCTGCGCXXXXXXMNNMNNMNNMNNMNNMNNMNNXXXXXXACTCTGGTGGTTTGTG	Reverse
71F	CTTCCAGTTCAGCTACGAGTTTGAGAAC	Forward
CDF/R	CAAGTAAAACCTCTACAAATGTGGTAAAATCG	Forward/Reverse, see Methods
588i-lib-PCR1-6bpUID-F	CACGACGCTCTTCCGATCTAANNNNNNAGTCCTATGGACAAGTGGCCACA	Forward
588i-lib-PCR1-R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTTCCTTGGTTTTGAACCCAACCG	Reverse
1527	ACACTCTTTCCCTACACGACGCTCTTCCGATCTGACAAGTGGCCACAAACCACCAG	Forward
1532	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTCCTTGGTTTTGAACCCAACCG	Reverse
mNeonGreen-F	CGACACATGAGTTACACATCTTTGGCTC	Forward

mNeonGreen-R	GGAGGTCACCCTTGGTGGACTTC	Reverse
Mito-F	CCCAGCTACTACCATCATTCAAGT	Forward
Mito-R	GATGGTTTGGGAGATTGGTTGATGT	Reverse
CapF-56	ATTGGCACCAGATACCTG ACTCGTAA	Forward
Cre-R-57	GTCCAAACTCATCAATGTATCTTATCATGTCTG	Reverse
NGS-QC-F	AATGATACGGCGACCACCGAG	Forward
NGS-QC-R	CAAGCAGAAGACGGCATACGA	Reverse

Table 3.T4: Primers used in M-CREATE selection.

The table provides a list of primers used in M-CREATE across the different steps of the selection process as described in the Methods.

Library purification

The assembled library was then subjected to Plasmid Safe (PS) DNase I (Epicentre; E3105K) treatment, or alternatively, Exonuclease V (RecBCD) (NEB; M0345S) following the recommended protocols, to purify the assembled product by degrading the un-assembled DNA fragments from the mixture. The resulting mixture was purified with a PCR purification kit (DNA Clean and Concentrator kit, Zymo Research; D4013).

Library yield

With an assembly efficiency of 15–20% post-PS treatment, we obtained a yield of about 15–20 ng per 100 ng of input DNA per 20 μL of assembly reaction.

Quality control

The following four steps were carried out to ensure successful capsid library generation.

1. To validate successful assembly of the library, 1 ng of the final assembled library was transformed into *E. coli* SURE 2 Supercompetent Cells (Integrated Sciences; 200152). We checked for colonies on an LB/Agar plate containing carbenicillin antibiotic after overnight incubation at 37°C.

2. The DNA library was sequenced around the insertion site (Laragen; Sanger Sequencing). A non-biased library may match the diversity of the NNK/MNN motif (where N = 25% each of A, T, G, C; K = 50% each of G, T; M = 50% each of A, C) with some fluctuations across the diversified region.

3. To verify that the ITRs were intact, SmaI digestion was carried out as per the NEB recommended protocol (NEB; R0141S).

4. To validate successful transfection and assess the vector-production yield per 150 mm dish, 10 ng of *7-mer-i* library was used to transfect 293T producer cells (293T; ATCC CRL 3216). Uniform expression of mNeonGreen protein across HEK cells was observed, and an average yield of $0.1 – 1x10^{11}$ vg was obtained per 150 mm dish. Using the average yield per dish, we scaled up the vector production for *in vivo* selection (see ***Supplementary Figure 3.S2a***).

Round-2 AAV capsid DNA library

PCR pool design

To maintain proportionate pooling, we mathematically determined the fraction of each sample or library that needs to be pooled based on an individual library's diversity.

This process involved estimation of the diversity precluding noise and consideration of amplification of this diversity across samples by determining the area under the curve for the interval of high-confidence variants that falls in the higher RC range. We estimated the area under the curve (AUC) using the composite Simpson's rule by plotting all the recovered variants in a library (X-coordinate) to their read counts (RCs or copy number from deep sequencing data, Y-coordinate) (see ***Supplementary Figure 3.S1e***). To determine the definite intervals for AUC, we sorted the data based on the decreasing order of the RCs. Noticeably, the distribution has two phases, with a steadier slope of variants in the higher RC range, followed by a steep drop in the slope of the curve (~50-1000 fold lower RCs). By observation, this steeper side of the curve is predominant in sequencing

errors/ PCR mutations, hence we precluded this error dominant slope otherwise called noise from our AUC estimation. When comparing composite Simpson's rule with another function, such as composite trapezoidal rule, the difference was miniscule.

This area is then used to determine the fraction of an individual library that needs to be pooled into PCR pool library using the formula: [Area under the curve/ total number of libraries pooled].

The pooled sample was used as a template for further amplification with 12 cycles of 98 °C for 10 s, 60 °C for 20 s and 72 °C for 30 s by Q5 polymerase, using the primers 588-R2lib-F and 588-R2lib-R (***Table 3.T4***). Similar to R1 library generation, the PCR product was assembled into the rAAV-ΔCap-in-cis-Lox2 plasmid and the virus was produced.

The R1 libraries used to build R2 were the Cre-Lox flipped rAAV DNA from half of the mouse brains (~0.3 g) and portion of spinal cords (0.1-0.2g) from all Cre lines. The amount of tissue processed here was sufficient for complete capsid library recovery. The differentially pooled and amplified libraries (by *PCR pool* or *synthetic pool*) were assembled using Gibson assembly with a follow-up PS or Exonuclease V treatment (as described in R1 library generation). We validated successful library generation by transformation, Sanger sequencing, and an ITR SmaI digest.

For vector production, about 10 ng of the purified and assembled library was used to transfect each 150 mm dish of 293T cells, and we obtained a yield of about 6×10^{11} vg per 150 mm dish (i.e., the R2 yield was six times that of R1, unsurprisingly given these sequences have already produced well enough to survive R1 selection).

Synthetic pool design

As described in the PCR pool strategy, we chose high-confidence variants whose RCs were above the error-dominant noise slope from the plot of library distribution (***Supplementary Figure 3.S1e***). This came to about 9,000 sequences from all brain and

spinal-cord samples of all Cre lines. We used similar primer design as mentioned in the description of the R1 library generation. Primers XF and 11-mer-588i (***Table 3.T4***) were used. In 11-mer-588i primer, ‘XXXXXXMNNMNNNMNNNMNNNMNNNMNNXXXXXX' was replaced with unique nucleotide sequence of a heptamer tissue recovered variant (7xMNN) along with modification of two adjacent codons flanking on either end of the heptamer insertion site (6xX), which are residues 587–588 ‘AQ' and residues 589–590 ‘AQ' on AAV9 capsid. Since the spike-in library has 11-mer or 33-base oligonucleotide mutated variants, we used the same primer design where ‘XXXXXXMNNMNNNMNNNMNNNMNNNMNNXXXXXX' was replaced with a specific nucleotide sequence of a 33-base oligonucleotide variant. A duplicate of each sequence in this library was designed with different codons optimized for mammals. The primers were designed using a custom-built Python based script. The custom-designed oligopool was synthesized in an equimolar ratio by Twist Biosciences. The oligopool was used to minimally amplify the pCRII-XE Cap9 template over 13 cycles of 98 °C for 10 s, 60 °C for 20 s and 72 °C for 30 s. To obtain a higher yield for large-scale library preparation, the product of the first PCR was used as a template for the second PCR using the primers XF and 588-R2lib-R (described above) and minimally amplified for 13 cycles. Following PCR, we assembled the R2 *synthetic pool* DNA library and produced the virus as described in R2 *PCR pool* design.

AAV virus library production, purification and genome extraction

To prevent capsid mosaic formation of the 7-mer-i library in 293 T producer cells, we transfected only 10 ng of assembled library per 150-mm dish along with other required reagents for AAV vector production.

In addition to the 10 ng of library transfection per 150 mm dish of 293T producer cells, we transfected three plasmids: AAV2/9 REP-AAP-ΔCap, pUC18 and pHelper (genes encoding adenoviral proteins for AAV replication) at a ratio of 1:1:2. The plasmid pUC18 acts as a filler DNA to compensate for the low amount of library DNA in order to maintain

the N:P ratio required for optimal transfection using polyethylenimine (PEI, Polysciences; 24765-1) transfection). The cells and culture media were harvested at 60 h post-transfection to collect the viral particles. rAAV harvest and purification were performed as per the protocol[251]. The small amount of library DNA per plate and early cell harvest time are critical for reducing the possibility of mosaic capsid assemblies during vector production (similar considerations seen in prior reports[149,253,254]).

For *7-mer-i* library, the production was scaled up to 60 dishes (~$1.8x10^7$ cells/dish) and with ~10% transfected with the library, resulted in ~$1x10^8$ total transformants. For an NNK 7-mer library with ~$1x10^8$ total transformants, the number of unique variants is $9.99x10^7$ (See Bosley & Ostermeier (2005)[255], section 2.1.2 for mathematics).

For the rAAV DNA extraction from purified rAAV viral library, ~10% of the purified viral library was used to extract the viral genome by proteinase K treatment.

In order to degrade any contaminating DNA from the purified library, it was treated with DNase I enzyme (5 µl of 10 U/µl) (Sigma-Aldrich; 4716728001) in 100 µl of DNase I buffer and incubated for 1 h at 37°C. The enzyme was inactivated by adding 5 µl of 0.5 M EDTA at 70°C for 10 min. Following DNase I treatment, the capsid protein shell was digested by adding 120 µl of proteinase solution containing 5 µl of 20 µg/µl of proteinase K and incubated at 50°C overnight. To inactivate the proteinase K, the mixture was boiled at 95°C. The extracted rAAV library DNA was then concentrated and purified using phenol chloroform and ethanol. An equal volume of Phenol:Chloroform:Isoamyl Alcohol 25:24:1, pH 8.0 (~250 µl; ThermoFisher Scientific; 15593031) was added and vortexed for 30 s. The mixture is incubated for 5 min at room temperature (RT) before centrifugation at 15,000 rpm for 10 min at 4°C. The upper aqueous phase was separated and mixed with an equal volume of chloroform and vortexed for 30 s. Following 5 min incubation at RT, centrifuge at 15,000 rpm for 10 min at 4°C. The upper aqueous phase was separated and one-tenth volume of 3M sodium acetate (pH 5.2) along with 2 µl Co-Precipitant Pink (Bioline; BIO-37075) and 2.5 volumes of ice cold 100% ethanol was added before vortexing for 30 s. The mixture was incubated for at least 1 hr at -20°C before

centrifugation at 15,000 rpm for 15 min at 4°C. The pellet was air dried and resuspended in TE buffer. The DNA concentration was determined using the Qubit ssDNA assay.

Animals

All animal procedures performed in this study were approved by the California Institute of Technology Institutional Animal Care and Use Committee (IACUC), and we have complied with all relevant ethical regulations. The C57BL/6J (000664), Tek-Cre[246] (8863), SNAP25-Cre[256] (23525), GFAP-Cre[257] (012886), Syn1-Cre[258] (3966), and Ai14[259] (007908) mouse lines used in this study were purchased from the Jackson Laboratory (JAX). The i.v. injection of rAAVs was into the retro-orbital sinus of 6- to 8-week-old male or female mice. For testing the transduction phenotypes of novel rAAVs, 6- to 8-week-old, male C57BL/6J or Tek-Cre or Ai14 mice were randomly assigned. The experimenter was not blinded for any of the experiments performed in this study.

In vivo selection

The *7-mer-i* viral library selections were carried out in different lines of Cre transgenic adult mice: Tek-Cre, SNAP25-Cre and GFAP-Cre for the R1 selections, and those three plus Syn1-Cre for the R2 selections. Male and female mice, 6- to 8-weeks-old, were i.v. administered with a viral vector dose of 2×10^{11} vg per mouse for the R1 selection, and a dose of 1×10^{12} vg per mouse for the R2 selection. The dose was determined on the basis of the virus yield, which was different across selection rounds (***Supplementary Figure 3.S2a***). Both genders were used to recover capsid variants with minimal gender bias. Two weeks' post-injection, mice were euthanized and all organs including brain were collected, snap frozen on dry ice, and stored at −80 °C.

rAAV genome extraction from tissue

Optimization

For rAAV genome extraction from tissues, we used both the Trizol method (Life Technologies; 15596) and the QIAprep Spin Miniprep kit (Qiagen, Inc; 27104) according to the manufacturers' recommended protocols, and found the Trizol method to be more efficient (see ***Supplementary Figure 3.S1f,g,i***). The total rAAV genome recovery from 0.1 g of mouse liver was quantified by quantitative PCR using the primers mNeonGreen-F and mNeonGreen-R, which binds to the mNeonGreen gene of the ssAAV-ΔCap-in-cis-Lox2 genome (see ***Table 3.T4***). As an internal control, we also quantified the amount of mitochondrial DNA (a surrogate for the recovery of circular, episomal genomes) using primers Mito-F and Mito-R (see ***Table 3.T4***). Although the percentage of viral DNA per 1 ng total extracted DNA was about 1.5 fold higher with the QIAprep kit than with the Trizol method, the overall recovery was lower with the QIAprep kit.

The extracted viral genome was digested with a restriction enzyme, such as SmaI (found within the ITRs), as it appeared to help improve rAAV genome recovery by PCR with ~4 fold change (see Supplementary Fig. 1h, ΔCT ~ 2, 1 CT = 2-fold difference in DNA copies, fold change = ($2^{\Delta CT}$) = 4). This was analyzed by quantitative PCR with Cre− primers, CapF-56 and Cre-R-57 (see ***Table 3.T4***).

rAAV genome extraction with the Trizol method

Half of a frozen brain hemisphere (0.3 g approx.) was homogenized with a 2 mL glass homogenizer (Sigma Aldrich; D8938) or a motorized plastic pestle (Fisher Scientific; 12–141–361, 12–141–363) (for smaller tissues) or beads using BeadBug homogenizers (1.5–3.0 mm zirconium or steel beads per manufacturer recommendations) (Homogenizers, Benchmark Scientific, D1032-15, D1032-30, D1033-28) and processed using Trizol as described in our prior work[157]. The extracted DNA by Trizol method was then treated with 3–6 μL of 10 μg μL^{-1} RNase Cocktail Enzyme Mix (ThermoFisher Scientific; AM2286) to remove RNA. The mixture was also digested with SmaI restriction enzyme to improve rAAV genome recovery by PCR (***Supplementary Figure 3.S1h***). The treated mixture was then finally purified with a Zymo DNA Clean and Concentrator kit

(D4033). From deep-sequencing data analysis, we observed that the amount of tissue processed was sufficient for rAAV genome recovery.

rAAV genome recovery by Cre-dependent PCR

rAAV genomes with Lox sites flipped by Cre recombination were selectively recovered and amplified using PCR with primers that yield a PCR product only if the Lox sites are flipped (***Supplementary Figure 3.S1b***). We used the primers 71F and CDF/R and amplified the Cre-recombined genomes over 25 cycles of 98 °C for 10 s, 58 °C for 30 s and 72 °C for 1 min, using Q5 DNA polymerase (***Table 3.T4***).

Total rAAV genome recovery by PCR (Cre-independent)

To recover all rAAV genomes from a tissue, we used the primers XF and 588-R2lib-R to amplify the genomes over 25 cycles of 98 °C for 10 s, 60 °C for 30 s and 72 °C for 30 min, using Q5 DNA polymerase (***Table 3.T4***).

Sample preparation for NGS

We processed the DNA library, the virus library and the tissue libraries following *in vivo* selection to add flow cell adaptors around the diversified heptamer insertion region (***Supplementary Figure 3.S1b***).

Preparation of rAAV DNA and viral DNA library

The Gibson-assembled rAAV DNA library and the DNA extracted from the viral library were amplified by Q5 DNA polymerase using the primers 588i-lib-PCR1-6bpUID-F and 588i-lib-PCR1-R that are positioned around 50 bases from the randomized heptamer insertion on the capsid, and that contain the Read1 and Read2 flow cell sequences on the 5′ end (***Table 3.T4***). The primer, 588i-lib-PCR1-6bpUID-F: 5’ CACGACGCTCTTCCGATCTAANNNNNNAGTCCTATGGACAAGTGGCCACA-3’ used to minimally amplify DNA and virus libraries for NGS has 6 nucleotides long UID

(unique identifier) “NNNNNN” that sits after 19 nucleotides of Read-1 sequence used in NGS “5’-CACGACGCTCTTCCGATCT” and linker “AA”. The sequence after UID “AGTCCTATGGACAAGTGGCCACA” is the region that anneals to the AAV9 capsid. UID is an optional feature for NGS data analysis to identify potential PCR amplification errors. However, this feature wasn’t utilized in the NGS data analysis in this study to maintain consistency with the primers used in rAAV genome recovery from tissues which lacks this UID feature (primers 71F and CDF/R, See ***Table 3.T4***). The UID or any kind of overhangs seemed to affect the PCR based recovery from tissue. Presumably, the primer thermostability have a key role to play in very low amount of extracted rAAV genomes from tissues.

Using 5–10 ng of template DNA in a 50 µL reaction, the DNA was minimally amplified for 4 cycles of 98 °C for 10 s, 60 °C for 30 s and 72 °C for 10 s. The mixture was then purified with a PCR purification kit. The eluted DNA was then used as a template in a second PCR to add the unique indices (single or dual) via the recommended primers (NEB; E7335S, E7500S, E7600S) in a 12-cycle reaction using the same temperature cycle as described above. The samples were then sent for deep sequencing following additional processing and validation.

The PCR products post indices addition were run on a freshly prepared 2% low-melting-point agarose gel (ThermoFisher Scientific; 16520050) for better separation and recovery of the approx. 120 bp DNA band on the gel. Before sending the sample for NGS, the nucleotide diversity at the randomized 7-mer position was verified by Sanger sequencing. If needed, an optional PCR was carried out to send sufficient sample for Sanger sequencing using 15 – 20 cycles of 98°C for 10 s, 60°C for 30 s, and 72°C for 10 s with the primers NGS-QC-F and NGS-QC-R (see ***Table 3.T4***). Upon validation, the libraries were sent for deep sequencing using the Illumina HiSeq 2500 System (Millard and Muriel Jacobs Genetics and Genomics Laboratory, Caltech; Integrative Genomics Core, City of Hope).

Preparation of rAAV tissue DNA library

The PCR-amplified rAAV DNA library from tissue (see sections: *rAAV genome recovery by Cre-dependent PCR* and *total rAAV genome recovery by PCR (Cre-independent)*) was further amplified with a 1:100 dilution of this DNA as a template to the primers 1527 and 1532 that are positioned around 50 bases from the randomized heptamer insertion on the capsid, and that contain the Read1 and Read2 sequences on the 5' end (see ***Table 3.T4***). The DNA was amplified by Q5 DNA polymerase for 10 cycles of 98 °C for 10 s, 59 °C for 30 s, and 72 °C for 10 s. The mixture was purified with a PCR purification kit. The eluted DNA was then used as a template in a second PCR to add the unique indices (single or dual) using the recommended primers (NEB; E7335S, E7500S, E7600S) in a ten-cycle reaction with the same temperature cycle as described above (for DNA and virus library preparation), and followed additional processing and validation before sequencing.

The PCR products post indices addition were run on a freshly prepared 2% low-melting-point agarose gel (ThermoFisher Scientific; 16520050) for better separation and recovery of the approx. 120 bp DNA band on the gel. Before sending the sample for NGS, the nucleotide diversity at the randomized 7-mer position was verified by Sanger sequencing. If needed, an optional PCR was carried out to send sufficient sample for Sanger sequencing using 15 – 20 cycles of 98°C for 10 s, 60°C for 30 s, and 72°C for 10 s with the primers NGS-QC-F and NGS-QC-R (see ***Table 3.T4***). Upon validation, the libraries were sent for deep sequencing using the Illumina HiSeq 2500 System (Millard and Muriel Jacobs Genetics and Genomics Laboratory, Caltech; Integrative Genomics Core, City of Hope).

In vivo characterization of AAV vectors

Cloning AAV capsid variants

The AAV capsid variants were cloned into a pUCmini-iCAP-PHP.B backbone (Addgene ID: 103002) using overlapping forward and reverse primers with 11-base oligonucleotide substitution (in case of *7-mer-i* variants, the flanking amino acids from AAV9 capsid AA 587–588 'AQ' and AA 589–590 'AQ' were subjected to codon

modification) that spans from the MscI site (at position 581 AA) to the AgeI site (at position 600 AA) on the pUCmini plasmid. The primers were designed for all capsid variants using a custom Python script and cloned using standard molecular techniques. The designed primers cover the entire fragment that is inserted into the linearized pUCmini-iCAP-PHP.B backbone. Hence these primers are simply self-annealed using PCR to synthesize double-stranded DNA fragment without the use of a template DNA. They are amplified by Q5 Hot Start High-Fidelity 2X Master Mix for 20 cycles of 98 °C for 10 s, 60 °C for 30 s and 72 °C for 10 s. This fragment was then assembled into the MscI/AgeI digested pUCmini-iCAP-PHP.B backbone by the Gibson assembly method. There is a second MscI site on the backbone; however, this was blocked by methylation. The assembled plasmids were then transformed into NEB Stable competent *E. coli* (New England Biolabs; C3040H), and colonies were selected on carbenicillin/ampicillin-LB agar plates. A list of primers used to clone AAV-PHP variants is provided (***Table 3.T5***).

Variant Name	Amino acid motif of variants with "AQ" overhangs from AAV9: AA587-88 and AA589-90. (11-mer)	Nucleotide sequence of 11-mer	Forward primer (5'-3')	Reverse primer (5'-3')
AAV-PHP.V1 (Addgene ID: 127847)	AQTALKPFLAQ	GCCCAAACCGCCCT CAAACCCTTCCTCG CACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAAACCGCCCT CAAACCC	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCGA GGAAGGGTTTGAGGGCGGT TTGGGC
AAV-PHP.V2 (Addgene ID: 127848)	AQTTLKPFLAQ	GCCCAAACCACCCT CAAACCCTTCCTCG CACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAAACCACCCT CAAACCC	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCGA GGAAGGGTTTGAGGGTGGT TTGGGC
AAV-PHP.B4 (Addgene ID: 127849)	AQTLQIPFKAQ	GCCCAAACGTTGCA GATTCCTTTTAAGGC ACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAAACGTTGCA GATTCCT	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCCTT AAAAGGAATCTGCAACGTTT GGGC
AAV-PHP.B5	AQTLQLPFKAQ	GCCCAAACCCTCCA ACTCCCCTTCAAAG CCCAA	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAAACCCTCCA ACTCCCC	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCTTGGGCTTT GAAGGGGAGTTGGAGGGTT TGGGC
AAV-PHP.B6	AQTLQQPFKAQ	GCCCAAACTTTGCA GCAGCCGTTTAAGG CACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAAACTTTGCA GCAGCCG	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCCTT AAACGGCTGCTGCAAAGTTT GGGC
AAV-PHP.B7	AQSIERPFKAQ	GCCCAAAGCATCGA AAGACCCTTCAAAG CACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAAAGCATCGA AAGACCC	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCTTT GAAGGGTCTTTCGATGCTTT GGGC
AAV-PHP.B8	AQTMQKPFIAQ	GCCCAAACCATGCA AAAACCCTTCATCG CACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAAACCATGCA AAAACCC	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCGA TGAAGGGTTTTTGCATGGTT TGGGC
AAV-PHP.C1	AQRYQGDSVAQ	GCCCAAAGGTATCA GGGTGATTCTGTTG CACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAAAGGTATCA GGGTGAT	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCAA CAGAATCACCCTGATACCTT TGGGC
AAV-PHP.C2	AQWSTNAGYAQ	GCCCAATGGTCGAC AAACGCTGGTTACG CACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAATGGTCGAC AAACGCT	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCGT AACCAGCGTTTGTCGACCAT TGGGC
AAV-PHP.C3	AQERVGFAQAQ	GCCCAAGAGCGTGT AGGTTTCGCACAGG CACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAAGAGCGTGT AGGTTTC	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCCT GTGCGAAACCTACACGCTCT TGGGC
AAV-PHP.N (Addgene ID: 127851)	AQTLAVPFSNP	GCGCAGACCCTAGC TGTCCCTTTTTCGAA CCCT	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCGCAGACCCTAGC TGTCCCT	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCAGGGTTCG AAAAAGGGACAGCTAGGGTC TGCGC
AAV-PHP.X1	AQARQMDLSAQ	GCCCAAGCCAGACA AATGGACCTCAGCG CACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAAGCCAGACA AATGGAC	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCGC TGAGGTCCATTTGTCTGGCT TGGGC
AAV-PHP.X2	AQTNKVGNIAQ	GCCCAAACCAACAA AGTCGGCAACATCG CACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAAACCAACAA AGTCGGC	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCGA TGTTGCCGACTTTGTTGGTT TGGGC
AAV-PHP.X3	AQQNVTKGVAQ	GCCCAACAGAACGT AACGAAGGGTGTGG CACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAACAGAACGT AACGAAG	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCCA CACCCTTCGTTACGTTCTGT TGGGC
AAV-PHP.X4	AQLNAIKNIAQ	GCCCAACTCAACGC TATCAAGAACATCG CACAG	GGAGTCCTATGGACAAG TGGCCACAAACCACCAG AGTGCCCAACTCAACGC TATCAAG	TTCCTTGGTTTTGAACCCAA CCGGTCTGCGCCTGTGCGA TGTTCTTGATAGCGTTGAGT TGGGC

Table 3.T5: Sequence motifs of AAV-PHP variants

The table provides the sequence motif information for all the new AAV-PHP variants identified in this study, and the primer information to clone them into pUCmini-iCAP-PHP.B

backbone (Addgene ID: 103002) as described in Methods. The variants from *7-mer-i* and *3-mer-s* libraries were cloned as 11-mer substitution.

AAV vector production

Using an optimized protocol[134] (*Chapter 2 in thesis*), we produced AAV vectors from 5–10 150-mm plates, which yielded sufficient amounts for administration to adult mice.

AAV vector administration, dosage and expression time

AAV vectors were administered intravenously to adult male mice (6–8 weeks of age) via retro-orbital injection at doses of 1×10^{11}–10×10^{11} vg with 3–4 weeks of *in vivo* expression times unless mentioned otherwise in the figures or legends.

The AAV doses are determined by the experimental needs. CAG-NLS-GFP related experiments for quantification were done at medium dose of $1x10^{11}$ vg given this was the dose previously determined for AAV-PHP.eB characterization. Otherwise, the non-NLS genome related experiments were done at $3x10^{11}$ vg, with the exception of Cre-driver lines (GFAP-Cre or Tek-Cre), or a lower strength promoter containing genome (GFAP-NLS-mTurq) where the dose was $1x10^{12}$ vg. The high dose was chosen to understand the full potential of the new vectors in these systems.

All experiments with vectors carrying CAG, a strong ubiquitous promoter, were incubated for 3 weeks. The 4 week incubations are those that involved expression from Cre driver lines or cell-type specific promoter where it is generally recommended for a longer wait time. The 2 week incubations are those where the vectors carried self-complementary genomes with strong ubiquitous promoters.

Tissue processing

After 3 weeks of expression (unless noted otherwise), the mice were anesthetized with Euthasol (pentobarbital sodium and phenytoin sodium solution, Virbac AH) and transcardially perfused with 30–50 mL of 0.1 M phosphate buffered saline (PBS) (pH 7.4),

followed by 30–50 ml of 4% paraformaldehyde (PFA) in 0.1 M PBS. After this procedure, all organs were harvested and post-fixed in 4% PFA at 4 °C overnight. The tissues were then washed and stored at 4 °C in 0.1 M PBS and 0.05% sodium azide. All solutions used for this procedure were freshly prepared. For the brain and liver, 100-µm thick sections were cut on a Leica VT1200 vibratome.

For vascular labeling, the mice were anesthetized and transcardially perfused with 20 mL of ice-cold PBS, followed by 10 mL of ice-cold PBS containing Texas Red-labeled Lycopersicon Esculentum (Tomato) Lectin (1:100, Vector laboratories, TL-1176) or DyLight 594 labeled Tomato Lectin (1:100, Vector laboratories, DL-1177), and then placed in 30 mL of ice-cold 4% PFA for fixation.

Immunohistochemistry

Immunohistochemistry was performed on 100-µm-thick tissue sections to label different cell-type markers such as NeuN (1:400, Abcam, ab177487) for neurons, S100 (1:400, Abcam, ab868) for astrocytes, Olig2 (1:400; Abcam, ab109186) for oligodendrocyte lineage cells and GLUT-1 (1:400; Millipore Sigma, 07-1401) for brain endothelial cells using optimized protocols.

Tissue sections, typically 100-µm thick, were first incubated in blocking buffer (10% normal donkey serum, 0.1% Triton X-100, and 0.01% sodium azide in 0.1 M PBS, pH 7.4) with primary antibodies at appropriate dilutions for 24 h at RT on a rocker. The primary antibodies used in this study were rabbit S100 (1:400, Abcam, ab868), rabbit Olig2 (1:400; Abcam, ab109186), rabbit NeuN (1:400, Abcam, ab177487), and rabbit GLUT-1 (1:400; Millipore Sigma, 07-1401). After primary antibody incubation, the tissues were washed 1 – 3 times with wash buffer 1 (0.1% Triton X-100 in 0.1 M PBS buffer, pH 7.4) over a period of 5 – 6 h in total. The tissues were then incubated in blocking buffer with the secondary antibodies at appropriate dilutions for 12 – 24 h at RT and then washed in three times in 0.1 M PBS, pH 7.4 over a total duration of 5 – 6 h. The secondary antibody was Alexa Fluor 647 AffiniPure donkey anti-rabbit IgG (H+L) (Jackson ImmunoResearch Lab,

711-605-152). When performing DNA staining, 4′,6-Diamidine-2′-phenylindole dihydrochloride (DAPI, Sigma Aldrich, 10236276001) is used at a 1:1000 dilution in 0.1 M PBS, pH 7.4 and incubated with tissues for 15 minutes followed by a single wash for 10 minutes in 0.1 M PBS, pH 7.4. The DAPI and/or antibody-stained tissue sections were mounted with ProLong Diamond Antifade Mountant (ThermoFisher Scientific, P36970).

Hybridization chain reaction (HCR)-based RNA labeling in tissues

Fluorescence in situ hybridization chain reaction (FITC-HCR) was used to label excitatory neurons with VGLUT1 and inhibitory neurons with GAD1 to characterize the AAV capsid variant AAV-PHP.N in brain tissue using an adapted third-generation HCR[260] protocol.

To characterize the AAV capsid variant AAV-PHP.N in brain tissue, HCR method was sought to label excitatory and inhibitory neurons. Fluorescence *in situ* hybridization chain reaction (FITC-HCR) was used to label excitatory neurons with VGLUT1 and inhibitory neurons with GAD1. Adapting the third-generation HCR[260], we designed 13 probe sets for each target by using custom-made software (https://github.com/GradinaruLab/HCRprobe).

After 3 weeks of expression, the mice were transcardially perfused and fixed as described earlier (Section D. Tissue processing). To minimize RNase enzyme exposure in fixed tissues, following overnight fixation in 4% PFA, the tissues were washed and stored at 4°C in 0.1 M RNase-free PBS and 0.05% sodium azide. The harvested brains were henceforth handled with care to avoid exposure to RNase using reagents such as RNAlater stabilization solution/RNase-free PBS/ RNaseZap (ThermoFisher Scientific, AM7021, AM9624, AM9780). Once the harvested brains were sagittally sliced to 100-µm thick sections, we performed FITC-HCR to detect both genes. We permeabilized tissue slices with 0.1% Triton X-100 in 0.1 M RNase-free PBS for 1 h at RT and pre-hybridized them in hybridization solution (10% dextran sulfate and 10% ethylene carbonate in 2xSSC buffer (saline-sodium citrate)) for >30 min at 37ºC. The designed probes were diluted in

hybridization solution to get a final concentration of 2 nM. The tissue sections were then subjected to hybridization with the probes overnight at 37ºC. Following this, the sections were washed with pre-warmed wash buffer (10% ethylene carbonate in 2xSSC) at 37ºC for 30 min twice, followed by 2xSSC at RT for 30 min twice. Amplification with hairpin pairs (Molecular Technologies, CA) were performed in amplification buffer (10x dextran sulfate in 2xSSC); hairpins were snap-cooled at 95ºC for 90 s, followed by RT for 30 min, and diluted with amplification buffer (60 nM). Tissues were then incubated in this amplification buffer with hairpins overnight at RT with gentle agitation. Once the amplification was done, samples were briefly washed with 2xSSC and mounted in Prolong Diamond for imaging.

Imaging and image processing

All images in this study were acquired either with a Zeiss LSM 880 confocal microscope using the objectives Fluar ×5 0.25 M27, Plan-Apochromat ×10 0.45 M27 (working distance, 2.0 mm), and Plan-Apochromat ×25 0.8 Imm Corr DIC M27 multi-immersion; or with a Keyence BZ-X700 microscope.

To prevent any imaging artifacts resulting from multiple fluorescence spectral overlap, the fluorescence excitation and emission spectra were kept distinct following the recommended linear unmixed acquisition of individual colors. A far-red fluorescent dye was chosen for any additional marker staining to keep the imaging parameters distinct from *in vivo* fluorescent expression thereby preventing any spectral overlap across detector channels. The tissues were routinely monitored for auto fluorescence or imaging artifacts before acquisition, and imaging parameters were adjusted if needed. The imaging parameters were cross-checked with tissues lacking *in vivo* transduction to avoid any imaging artifacts. The regions used for the images were closely matched across experimental groups to minimize bias during comparisons.

The acquired images were processed in the respective microscope software Zen Black 2.3 SP1 (Zeiss), BZ-X Analyzer (Keyence), Keyence Hybrid Cell Count software

(BZ-H3C), ImageJ, Imaris (Bitplane) and with Photoshop CC 2018 (Adobe). The images were compiled in Illustrator CC 2018 (Adobe).

Tissue clearing

Brain hemispheres were cleared using iDISCO[261] method and tissues over 500 µm thickness were optically cleared using ScaleS4(0)[170].

To demonstrate the ability of PHP.V1 to transduce the vasculature across thick tissues, such as a mouse-brain hemisphere, we assessed tissue from Tek-Cre mice 4 weeks' post administration. The brain hemisphere was stained with the primary antibody, Anti-GFP (1:200, Aves Labs, GFP-1020), and the secondary antibody, goat anti-Chicken IgY, Alexa Fluor 633 (1:200, ThermoFisher Scientific, A-21103), and cleared via the iDISCO protocol[261]. For imaging, a commercial light-sheet microscope (Lavision BioTec) with a custom objective lens (4×) was used[262]. The resulting image files were reorganized by a custom MATLAB script to allow stitching with TeraStitcher[263]. For 3-D visualization, Imaris (Bitplane) was used.

For images of peripheral organs such as heart, muscle, spleen that were over 500 microns in thickness, optical clearing was performed by incubating the tissues in 5 ml of *Scale*S4(0)[170] solution overnight at RT with gentle agitation, then mounted the tissues in fresh *Scale*S4(0) solution and imaged under the confocal microscope.

Tissue processing and imaging for quantification of rAAV transduction in vivo

For quantification of rAAV transduction, 6- to 8-week-old male mice were i.v. injected with the virus, which was allowed to express for 3 weeks (unless specified otherwise). The mice were randomly assigned to groups and the experimenter was not blinded. The mice were perfused and the organs were fixed in PFA. The brains and livers were cut into 100-µm-thick sections and immunostained with different cell-type-specific antibodies, as described above. The images were acquired either with a ×25 objective on a Zeiss LSM 880 confocal microscope or with a Keyence BZ-X700 microscope; images that

were compared directly across groups were acquired and processed with the same microscope and settings.

For quantification of PHP.B-family variant transduction in tissues, the images were acquired using 25x objective with 1x digital zoom on a Zeiss LSM 880 confocal microscope. With n=3 mice per variant, images were acquired across 4 brain regions – cortex, striatum, ventral midbrain and thalamus and tissues were stained with 3 cell type markers (NeuN, Olig2, and S100). For each mouse, 2 images per brain region per cell type marker were acquired, and the mean were plotted.

For PHP.N transduction analysis, the images were acquired using 20x objective on Keyence BZ-X700 microscope. With n=3 mice, images across 4 brain regions - cortex, striatum, ventral midbrain and thalamus were acquired to cover the entire brain regions for 3 cell type markers (NeuN, Olig2, and S100). This involved 6-8 images to cover cortex, thalamus and striatum, and 2 images to cover ventral midbrain per mouse per cell-type marker. For each mouse, across each region, the mean from the images were plotted.

For PHP.V GLUT1^{+} transduction analysis, the images were acquired using 25x objective with 1x digital zoom on a Zeiss LSM 880 confocal microscope. Each distinct blood vessel in the image with GLUT1^{+} staining and XFP expression was determined as positive for transduction. Quantification of expression from the CAG-mNeonGreen vector was performed across the cortex (n = 3 per group). Each data point is drawn from the mean of 3-2 images per mouse. Different brain regions were quantified for Tek-Cre and Ai14 mouse experiments with n = 2 per group. For cortex, cerebellum, striatum, and ventral midbrain, the mean was plotted from 3-4 images per mouse per region.

In vitro characterization of AAV vectors

Human brain microvascular endothelial cells (HBMEC) (ScienCell Research Laboratories, cat. no. 1000) were cultured as per the instructions provided by the vendor. HBMEC were cultured from a frozen stock vial in fibronectin-coated T-75 flask (7,000–

9,000 cells per cm^2 seeding density) using the endothelial cell medium (cat. no. 1001). The cells were subcultured in fibronectin-coated 48-well plates (0.95 cm^2 growth area) at the recommended seeding density and incubated at 37 °C for ~24–48 h until the cells were completely adherent with ~70–80% confluence. The viral vectors packaging pAAV:CAG-mNeongreen were added to the cell culture at a dose of either 1×10^8 or 1×10^{10} vg per well (3 wells per dose per vector). The medium was changed 24 h later, and the culture was assessed for fluorescence expression at 3 d post infection. Per vendor recommendation, the culture medium was changed every other day to maintain the cell culture.

Data analysis

Quantification of rAAV vector transduction

Manual counting was performed with the Adobe Photoshop CC 2018 Count Tool for cell types in which expression and/or antibody staining covered the whole cell morphology. The Keyence Hybrid Cell Count software (BZ-H3C) was used where the software could reliably detect distinct cells in an entire dataset. To maintain consistency in counting across different markers and groups, one person was assigned to quantify across all groups in all brain areas. Manual counting was performed for GLUT-1-stained blood vessels and expression of the ssAAV:CAG-mNeonGreen and ssAAV:CAG-DIO-EYFP, where the efficiency was calculated as the percentage of XFP^+ vessels relative to the GLUT-1 staining. Manual counting was also performed to quantify nuclear or soma stained cells, including NeuN-, Olig2-, and S100-stained cells. The efficiency was calculated as the percentage of XFP^+ cells relative to cell-$marker^+$ cells.

Keyence Hybrid Cell Count software (BZ-H3C) was used to quantify expression of nuclear localized AAV genomes in liver hepatocytes that co-localized with the DNA stain, DAPI; and also for the study involving ssAAV:GFAP-2xNLS-mTurquoise2 genomes with S100 cell marker.

The mean fluorescence intensity across microscopic images were quantified using ImageJ software. The images were processed for background subtraction and using the Threshold operation, the mean fluorescence intensity was measured. The experimenter was not blinded during data analysis.

NGS data alignment and processing

The raw fastq files from NGS runs were processed with custom-built scripts that align the data to AAV9 template DNA fragment containing the diversified region 7xNNK (for R1) or 11xNNN (for R2 since it was synthesized as 11xNNN).

The pipeline to process these datasets involved filtering the dataset to remove the low-quality reads by using the deep sequencing quality score for each sequence. The variant sequences were then recovered from the sequencing reads by searching for the flanking template sequences, and extracting the nucleotides of the diversified region (perfect string match algorithm). The quality of the aligned data was further investigated to remove any erroneous sequences (such as ones with stop codons). The raw data was plotted (as shown in ***Supplementary Figure 3.S1e***) to study the quality of recovery across every library. Based on the RC distribution, we adapted a thresholding method to remove plausible erroneous mutants that may have resulted from PCR or NGS based errors. The assumption is that if there is a PCR mutation or NGS error on the recovered parent sequence, the parent must have existed at least one round earlier than the erroneous sequence, and thus a difference in RCs should exist.

For R1 tissue libraries, we observed a steep drop in the slope of the distribution curve following a long tail of low count sequences, and were found to be rich in sequences that are variations of the parents in the higher counts range. We manually setup a threshold for RCs to remove such erroneous mutants. The thresholded data were then processed differently based on the experimental needs as described elsewhere using custom Python based scripts.

For R2 tissue libraries from *PCR pool* and *synthetic pool*, given the smaller library size compared to R1, we thresholded the data in two steps. We only considered the tissue recovered sequences that were present in the respective input DNA and virus library (after removing lower count variants from input libraries following the same principle as R1 tissue libraries). This step partially removed the long tail of low count reads. As a second step, we applied the thresholding that was described for R1 tissue libraries.

While it is plausible that true variants may be lost during thresholding, this method minimized false positives as the low count mutants in tissue and virus libraries often seemed to have very high enrichment score (as RCs are normalized to input library). In other words, thresholding allowed selective investigation on enriched variants that had a higher-confidence in their NGS RCs.

As an alternative to our manual thresholding method, an optional error correction method called "Collapsing" was built to further validate the outcome from filtered datasets. This method starts at the lowest count variants (variants of count 1) and searches for potential parent variants that are off by one nucleotide but have at least 2-fold higher counts (fold change = ($2^{\Delta CT}$) where CT is PCR cycle threshold). This error correction method then transfers the counts of these potential erroneous sequences to their originating sequences and repeats recursively until all sequences have been considered. On applying this error correction to our thresholded data, an additional ~0.002-0.03% of sequences were captured (compared to >19% captured by thresholding), confirming that our thresholding strategy was largely successful.

NGS data analysis

The aligned data were then further processed via a custom data-processing pipeline, with scripts written in Python.

The enrichment scores of variants (total, *N*) across different libraries were calculated from the read counts (RCs) according to the following formula:

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

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

Since the DNA and virus libraries were not completely sampled unlike the tissue libraries, we assigned an estimated RC for variants that were not present in the input library but were present in the output library. For instance, R1 virus library is the input library to the R1 tissue libraries. The estimated RC is defined as a number that is lower than the lowest RC in the library with the assumption that these variants were found at a relatively lower abundance than the variants recovered from the deep sequencing. In virus libraries, since RC of 1.0 was the lowest, we assigned all missing variants an estimated RC of 0.9. We use this method to calculate the enrichment score of the R1 tissue libraries which is normalized to R1 virus library (***Figure 3.1d***). This was done to represent libraries across two selection rounds consistently. Although, the individual enrichment score among R1 variants didn't add a significant value to the variants selected for R2 selection as described in the criteria to separate signal vs noise in R1 using the RCs.

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

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

Read count_i is raw copy number of a variant i. Mean is the mean of read counts of all variants across a specific library. The s.d. is the s.d. of read counts of all variants across a specific library.

The plots generated in this article were using the following software: Plotly, GraphPad PRISM 7.05, Matplotlib, Seaborn and Microsoft Excel 2016. The AAV9 capsid structure (PDB 3UX1)[264] was modeled in PyMOL.

Heat map generation

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

Each heatmap uses both an expected (input) distribution of amino acid sequences and an output distribution. The output distribution must be a list of sequences and their count, and the input distribution can be either a list of sequences and their count, or an expected amino acid frequency from a template, such as NNK. For both input and output, the total count of amino acids in each position is tallied in accordance to each sequence's count and then divided by the total sum of counts, giving a frequency of each amino acid at each position. Then, the log2 fold change is calculated between the output and the input. For amino acids with a count of 0 in either the input or output, no calculation is performed. In order to distinguish between statistically significant amino acid biases, a statistical test was performed using the statsmodels Python library. For the case where there are two amino acid counts, a two-sided, two-proportion z-test was performed; for comparing the output amino acid count to an expected input frequency from a template, a one-proportion z-test was performed. All p-values were then corrected for multiple comparisons using Bonferroni correction. Only bias differences below a significance threshold of 1e-4 are then outlined on the heatmap; all other (insignificant) squares are left empty.

Clustering analysis

Using custom scripts written in MATLAB (version R2017b; MathWorks), we determined the reverse Hamming distances representing the number of shared amino acids between two peptides. Cytoscape (version 3.7.1[265]) software was then used to cluster the variants. The amino acid frequency plot representing the highlighted cluster was created using Weblogo (Version 2.8.2)[266,267].

The reverse Hamming distances (representing the number of shared AAs between two peptides) was determined for all unique capsid variants with greater than 10 count and

greater than 2.5-fold enrichment after R2 selection. This process iteratively compares each variant with all other variants within the group. Capsid variants were then clustered by their reverse Hamming distances using Cytoscape. The minimum reverse Hamming distance for visualization was chosen manually based on sequence similarity.

For the amino acid frequency plots, the number on the bottom represents the position of the diversified motif starting from 1. The size of the amino acid in the stack reflects the proportion of unique clones in which the AA appears at that specific position in the motif. The color code is based on the AA properties. The positively charged residues K, R, and H are in blue. The negatively charged residues D and E are in red. The amide containing polar residues Q, and N are in magenta. The polar residues T, and S, are in green. The hydrophobic residues A, L, V, I, P, F, M, and W are in black.

Statistics and reproducibility

Statistical tests were performed using GraphPad PRISM or Python scripts. All correlation analyses reported were carried out using a linear least-squares regression method by an inbuilt Python function from SciPy library ‘scipy.stats.linregress', and the coefficient of determination (R^2) is reported. Tests evaluating the significance of amino acid bias were done using statsmodels Python library. A one-proportion *z*-test for a library versus known template frequency (NNK), and two-proportion *z*-test for two-library comparisons were performed. *P* values are corrected for multiple comparisons using a Bonferroni correction. For datasets with two experimental group comparisons, a Mann–Whitney test was used and two-tailed exact *P* values are reported. For more than two experimental group comparisons with one variable, a one-way ANOVA non-parametric Kruskal–Wallis test with multiple comparisons using uncorrected Dunn’s test was performed. Exact *P* values are reported from both tests (unless indicated otherwise). For experimental group comparisons with two variables, a two-way ANOVA with Tukey’s test for multiple comparisons reporting corrected *P* values were performed with 95% CI.

All quantitative data reported in graphs are from biological replicates (mouse or tissue culture replicates), where each data point from a biological replicate is the mean from technical replicates (raw data such as images of a specific brain region). Statistical analyses were performed on datasets with at least three biological replicates. Error bars in the figures denote s.e.m. All experiments were validated in more than one independent trial unless otherwise noted.

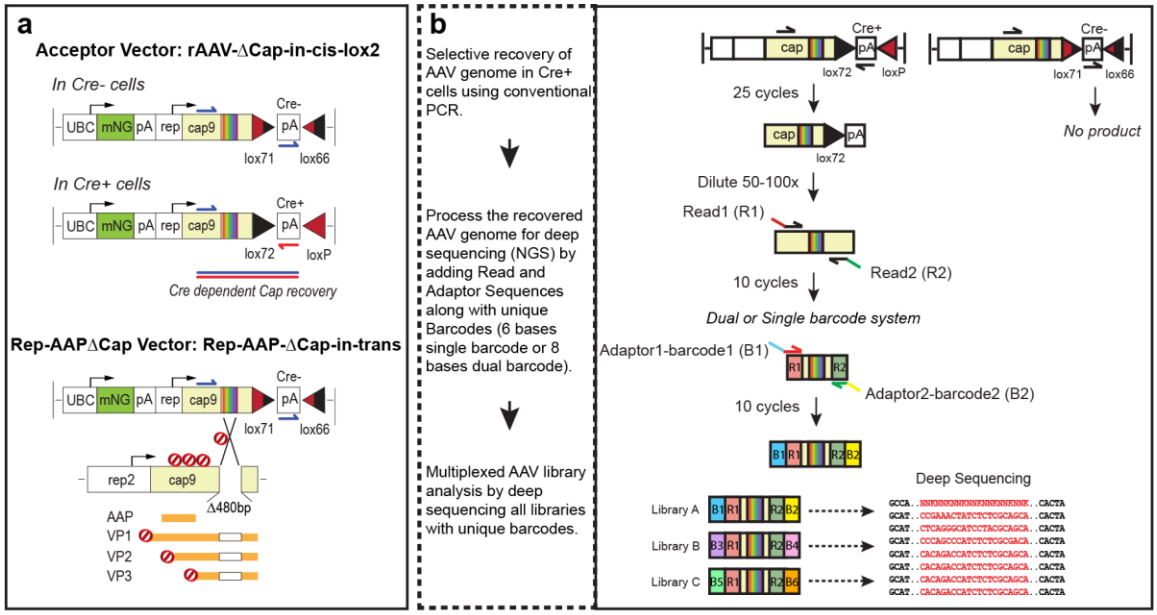
Data availability

The NGS datasets using the synthetic pool and PCR pool selection methods that are reported in this article are available under the SRA accession code: PRJNA610987. The following vector plasmids are deposited on Addgene for distribution (http://www.addgene.org) AAV-PHP.V1: 127847, AAV-PHP.V2: 127848, AAV-PHP.B4: 127849, and AAV-PHP.N: 127851, and viruses may be available for commonly packaged genomes. Other plasmids or viruses not available at Addgene may be requested from Caltech, CLOVER Center (http://clover.caltech.edu/). GenBank: AAV-PHP.V1: MT162422, AAV-PHP.V2: MT162423, AAV-PHP.N: MT162424, AAV-PHP.C1: MT162425, AAV-PHP.C2: MT162426, AAV-PHP.C3: MT162427, AAV-PHP.B4: MT162428, AAV-PHP.B5: MT162429, AAV-PHP.B6: MT162430, AAV-PHP.B7: MT162431 and AAV-PHP.B8: MT162432.

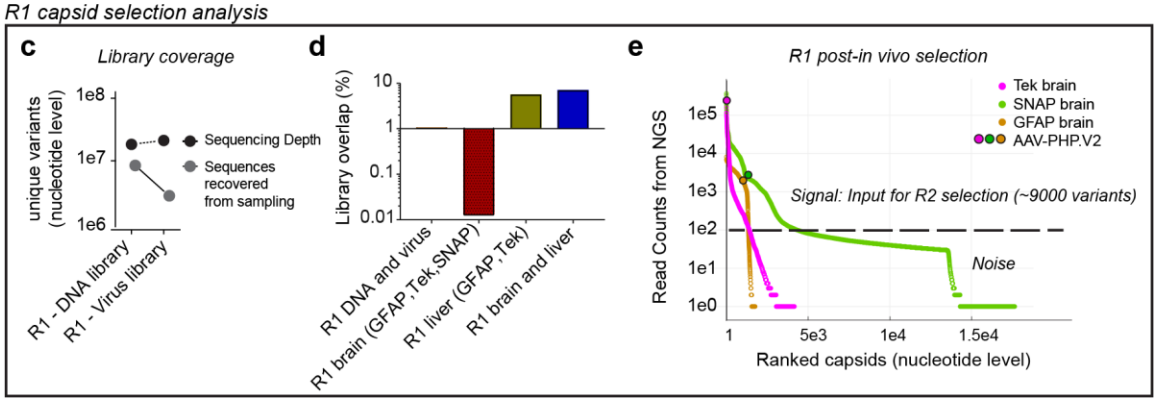
Code availability

The codes used for M-CREATE data analysis were written in python or MATLAB and are made available on GitHub: https://github.com/GradinaruLab/mCREATE. The custom MATLAB scripts to generate HCR probes is accessible through GitHub on a different repository: https://github.com/GradinaruLab/HCRprobe.

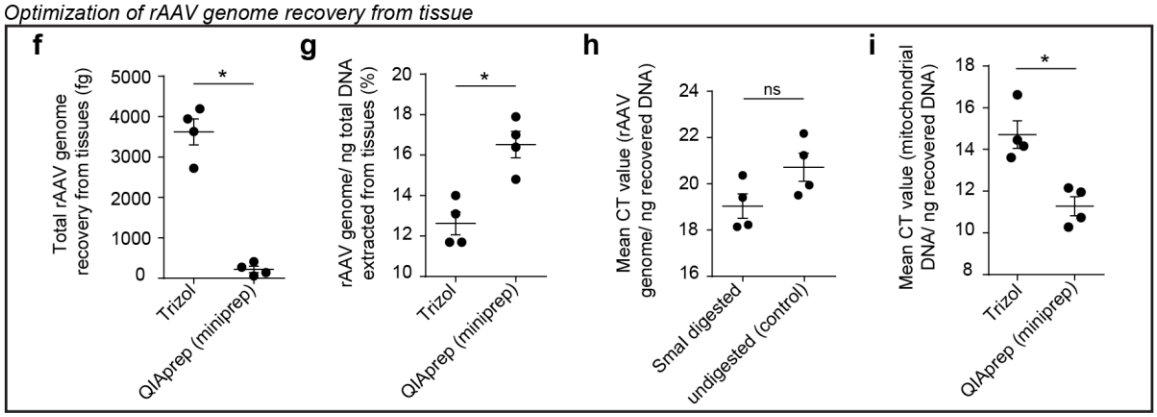
3.6 SUPPLEMENTARY INFORMATION



R1 capsid selection analysis

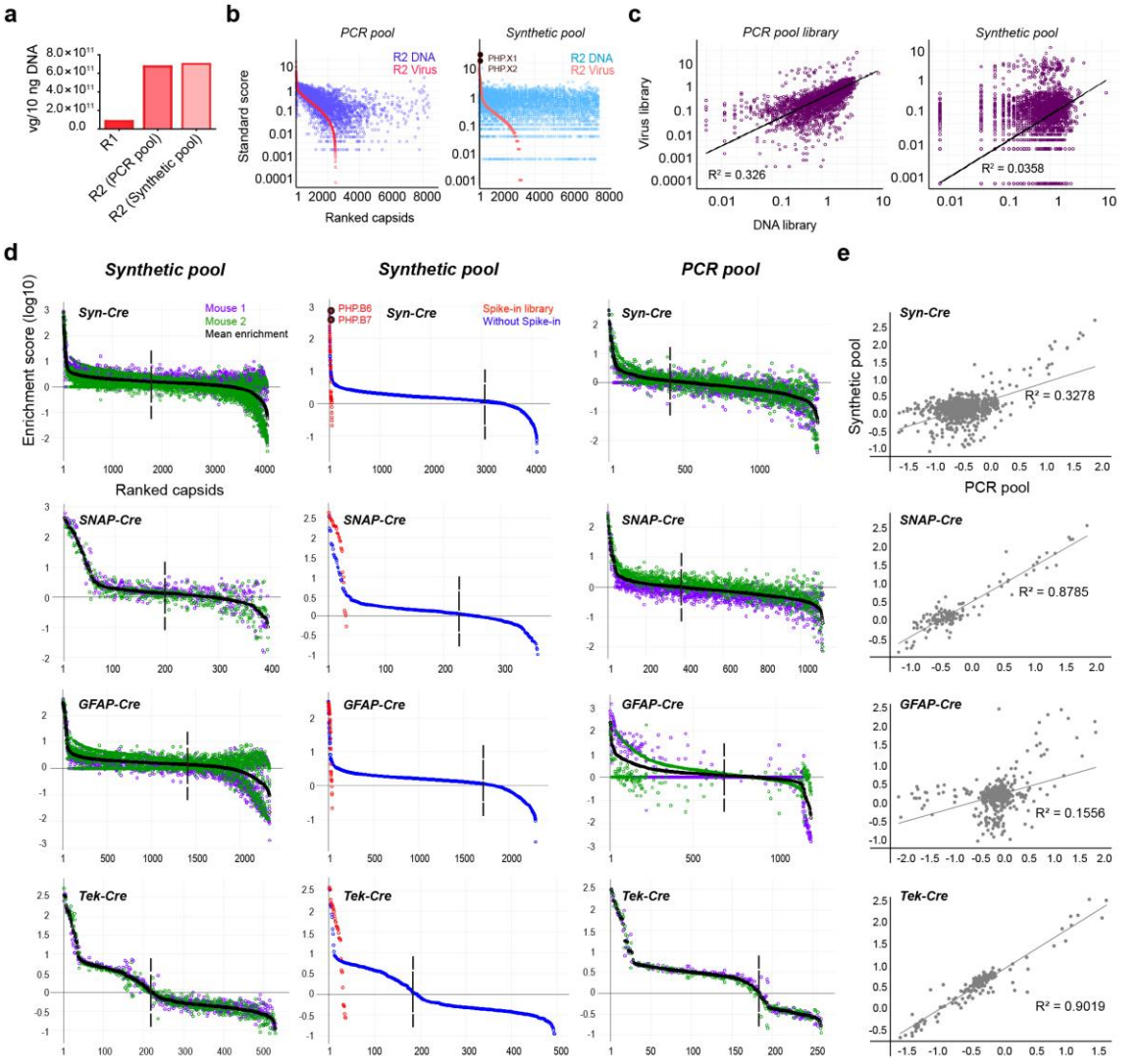


Optimization of rAAV genome recovery from tissue



Supplementary Figure 3.S1: Extended Schematic For Multiplexed-CREATE And Analysis Of Round-1 Selection.

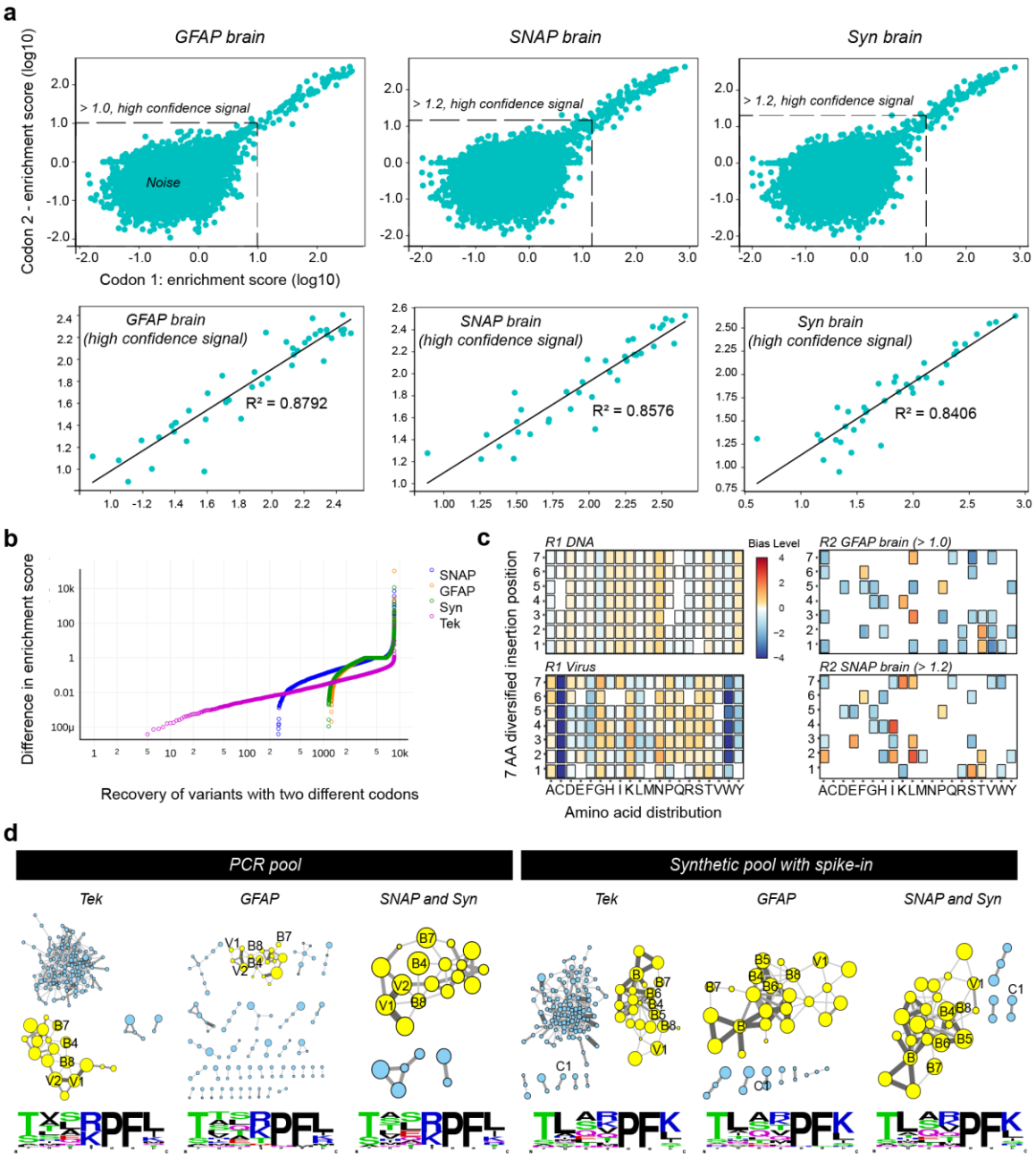
(a), Diagram of the genetic switch used in M-CREATE. The Acceptor Vector shows the position of the forward and reverse primers between the Lox sites that are used for selective recovery of capsids from the Cre^{+} cells. The Rep-AAPΔCap vector shows a deletion of 480 bp in *cap* gene in addition to the stop codons that are designed to prevent synthesis of VP1, VP2, and VP3 proteins. AAP protein translation is unaffected by these modifications. **(b)**, Schematic of the protocol to selectively recover rAAV genomes from the target population using the Cre-Lox flipping strategy and preparation of the sample for deep sequencing. **(c)**, The library coverage for R1 DNA and virus libraries obtained from specific sequencing depths. **(d)**, The percentage of variant overlap within the sampled DNA and virus, or across different Cre lines within tissues, or across tissues from R1 selection. **(e)**, The distributions of AAV capsid RCs for libraries recovered by NGS from brain tissue across different Cre transgenic mice post R1 selection. The dotted line is illustrative only and roughly separates the signal from noise (see Methods for estimation of signal versus noise) where signal in this context represents the input for the R2 selection. **(f)**, rAAV genome recovery from tissues using different treatments are shown with total rAAV genome recovery from 0.1 g of liver, **(g)**, Percentage of rAAV genomes recovered per ng of total extracted DNA, and **(h)**, The CT value (cycle threshold from qPCR) of rAAV genome extracted by trizol that were treated with SmaI restriction enzyme or untreated and **(i)**, CT value of mitochondrial DNA (internal control for smaller genome recovery, fold change = 10.79 ($2^{\Delta CT}$)) recovered from 1 ng of total DNA from liver tissue. In **(f-i)**, $n = 4$ mice; 2 from GFAP-Cre line and 2 from Tek-Cre line, each data point is drawn from the mean of three technical replicates, error bar is mean ± S.E.M., Mann-Whitney test, two-tailed (exact P-value of 0.0286 (*$P \leq 0.05$), in **(f, g, i)**, and 0.1143 (n.s., $P > 0.05$, CI 95%) in **(h)**). The data reported **(f, g, i)** are from one independent trial, and **(h)**, from three independent trials.



Supplementary Figure 3.S2: Analysis Of 7-mer-i rAAV Libraries From Round-2 Selections.

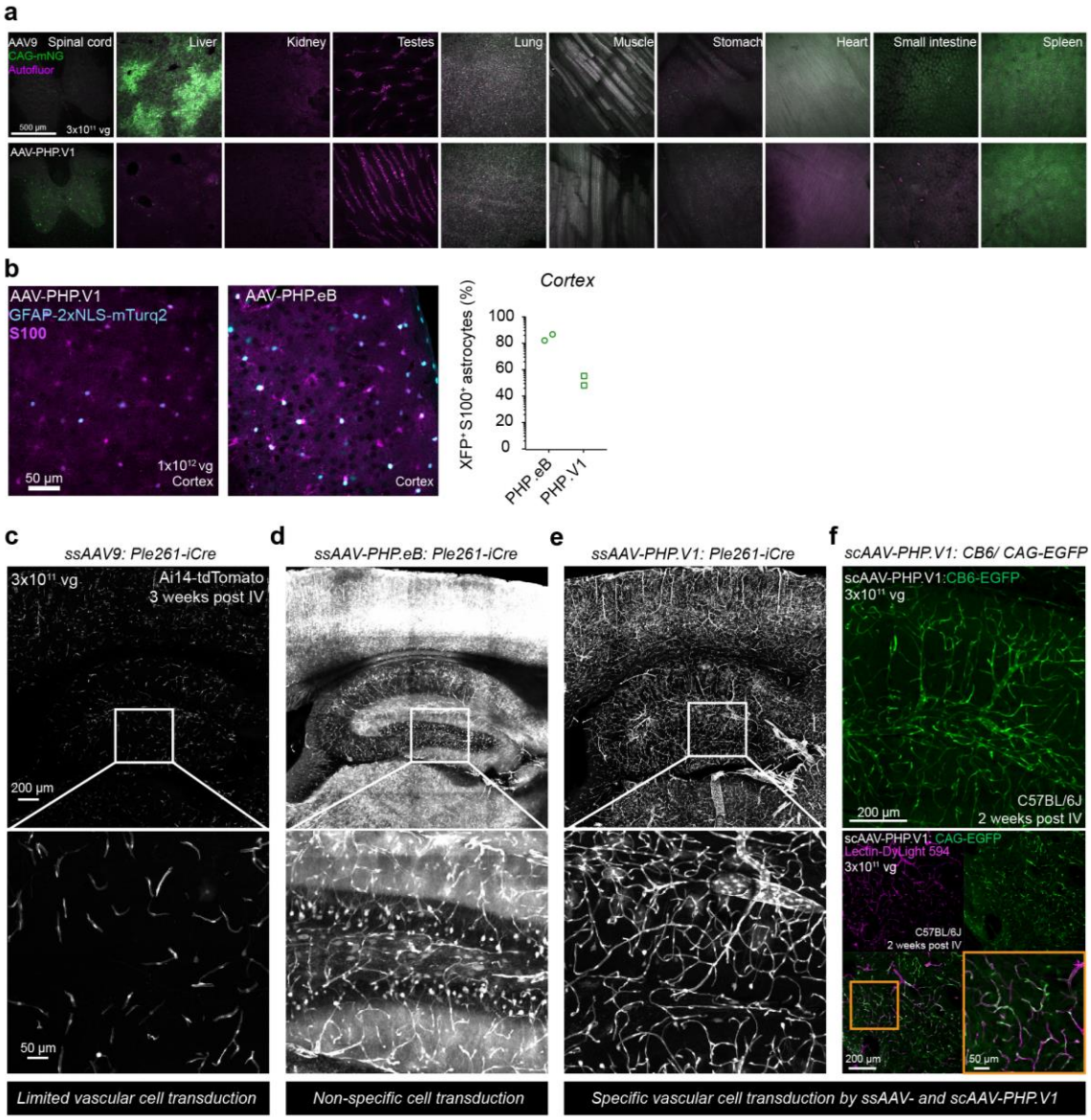
(a), The vector yields obtained per 10 ng of capsid DNA library across R1 and R2 vector productions. **(b)**, Distributions of the DNA and virus libraries produced by the *synthetic pool* and *PCR pool* methods by the standard score of NGS RCs. The variants in virus libraries are sorted by the decreasing order of standard score (SS) and their scores from respective DNA libraries are mapped onto them. **(c)**, Correlations between the SS of RCs for the DNA and virus libraries (n = 1 per library) produced by the *synthetic pool* and *PCR pool* methods is determined by linear least-squares regression, and the regression line (best fit) and R^2 representing the coefficient of determination is shown. **(d)**, Distributions of capsid libraries from brain tissue of two mice (purple and green) used in each Cre line selection, as produced by the *synthetic pool* (left) and *PCR pool* (right) designs. The distribution of *spike-in* library introduced in the *synthetic pool* library design

is shown in red (center). **(e)**, Correlations of enrichment scores of variants from the brain libraries (n = 2 per Cre line, mean is plotted) produced by *synthetic pool* and *PCR pool* methods is determined by the same method described in **(c)**.



Supplementary Figure 3.S3: Analysis Of Round-2 7-mer-i Tissue Libraries From Synthetic Pool And PCR Pool Methods.

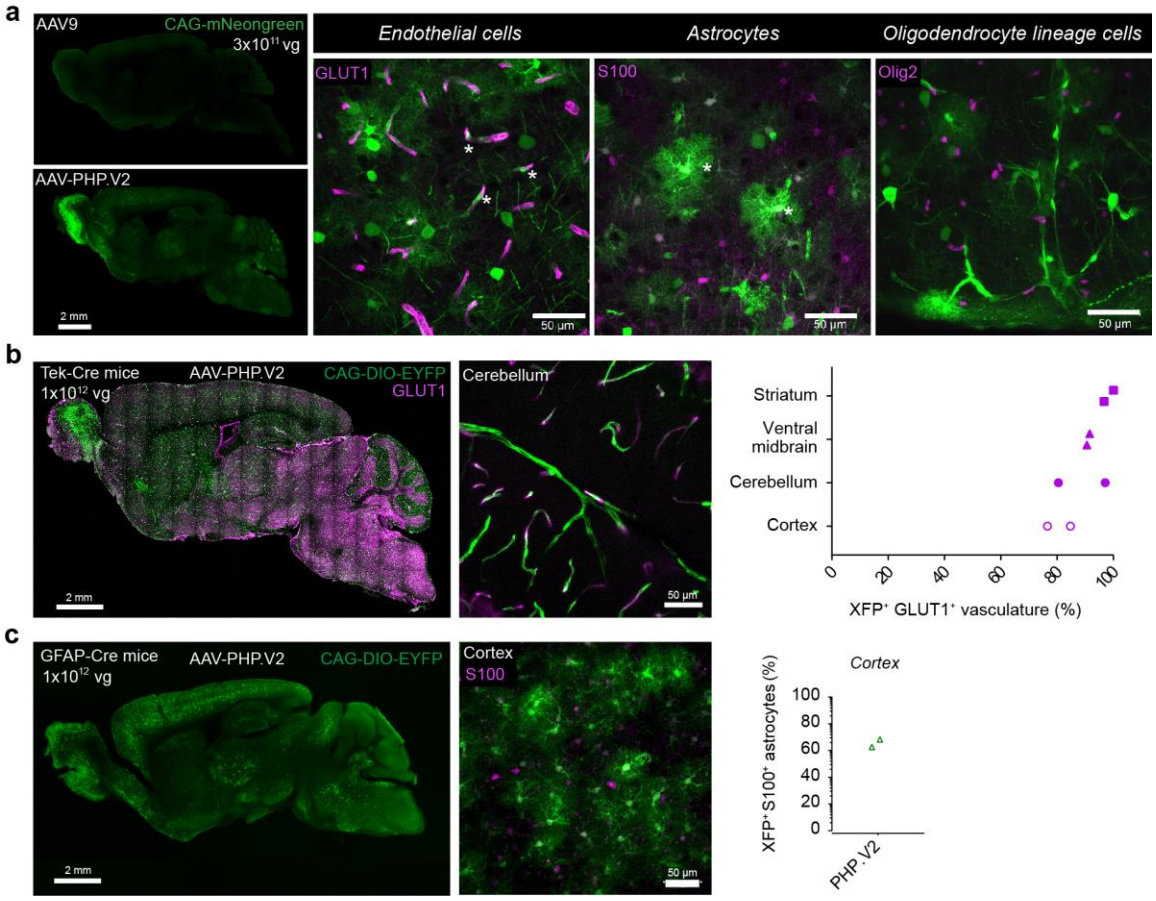
(a), Correlation analysis between the enrichment score (log10) of two alternate codon replicates of variants from the GFAP-Cre (left), SNAP-Cre (center), and Syn-Cre (right) brain libraries by linear least-squares regression (n = 2 per Cre line, mean is plotted). The dotted line separates the high-confidence signal from noise. High confidence signal (below) is assessed by a linear regression line (best fit) and R^2 represents the coefficient of determination. **(b)**, The difference in enrichment score between the two codon replicates of a variant, across different brain libraries, with over 8000 variants recovered in replicates. **(c)**, Heatmaps represent the magnitude (log2 fold change) of AA bias in "output" library 1 normalized to "input" library 2 that reach statistical significance (boxed if P-value ≤ 0.0001, two-sided, two-proportion z-test, except in R1 DNA normalized to known NNK template where one-proportion z-test was performed, and P-values corrected for multiple comparisons using Bonferroni correction) is shown. R1 DNA library normalized to NNK template (top left, ~9 million sequences), R1 virus normalized to R1 DNA libraries (bottom left, ~10 million sequences), R2 GFAP library with enrichment score above 1.0 in brain normalized to R2 virus (top right, 20 sequences,) and R2 SNAP library with enrichment score above 1.2 normalized to R2 virus (bottom right, 17 sequences) are shown (n = 1 for DNA, virus, and n = 2 for brain libraries). **(d)**, Clustering analysis of positively enriched variants from Tek, GFAP, and combined neuron brain libraries (SNAP and Syn) *by PCR pool* design, and by *synthetic pool* design with *spike-in* library are shown with size of nodes representing their relative enrichment in brain, and the thickness of edges (connecting lines) representing the extent of shared AA identity between nodes. A distinct family is highlighted in yellow with the corresponding AA frequency logo below (AA size reflects prevalence and color coded based on AA properties).



Supplementary Figure 3.S4: AAV-PHP.V1 Efficiently Targets The Brain Vasculature.

(a), Expression of AAV9 (above) and AAV-PHP.V1 (below) packaging ssAAV:CAG-mNeonGreen across all organs is shown (n=3, $3x10^{11}$ vg dose per adult C57BL/6J mouse, 3 weeks of expression). The background auto fluorescence is represented in magenta. **(b)**, Expression in cortical astrocytes ($S100^{+}$) after i.v. delivery of AAV-PHP.V1 (left) and AAV-PHP.eB (right) capsids carrying ssAAV:GfABC1D-2xNLS-mTurquoise2 ($1x10^{12}$ vg dose per adult mouse, 4 weeks of expression). Percentage of cortical $S100^{+}$ cells that overlapped with mTurquoise2 expression is quantified ($n = 2$, each data point is mean from 3 images per mouse). **(c)**, Expression of AAV9, AAV-PHP.eB **(d)**, AAV-PHP.V1 **(e)** packaging ssAAV:Ple261-iCre transgene in Ai14-

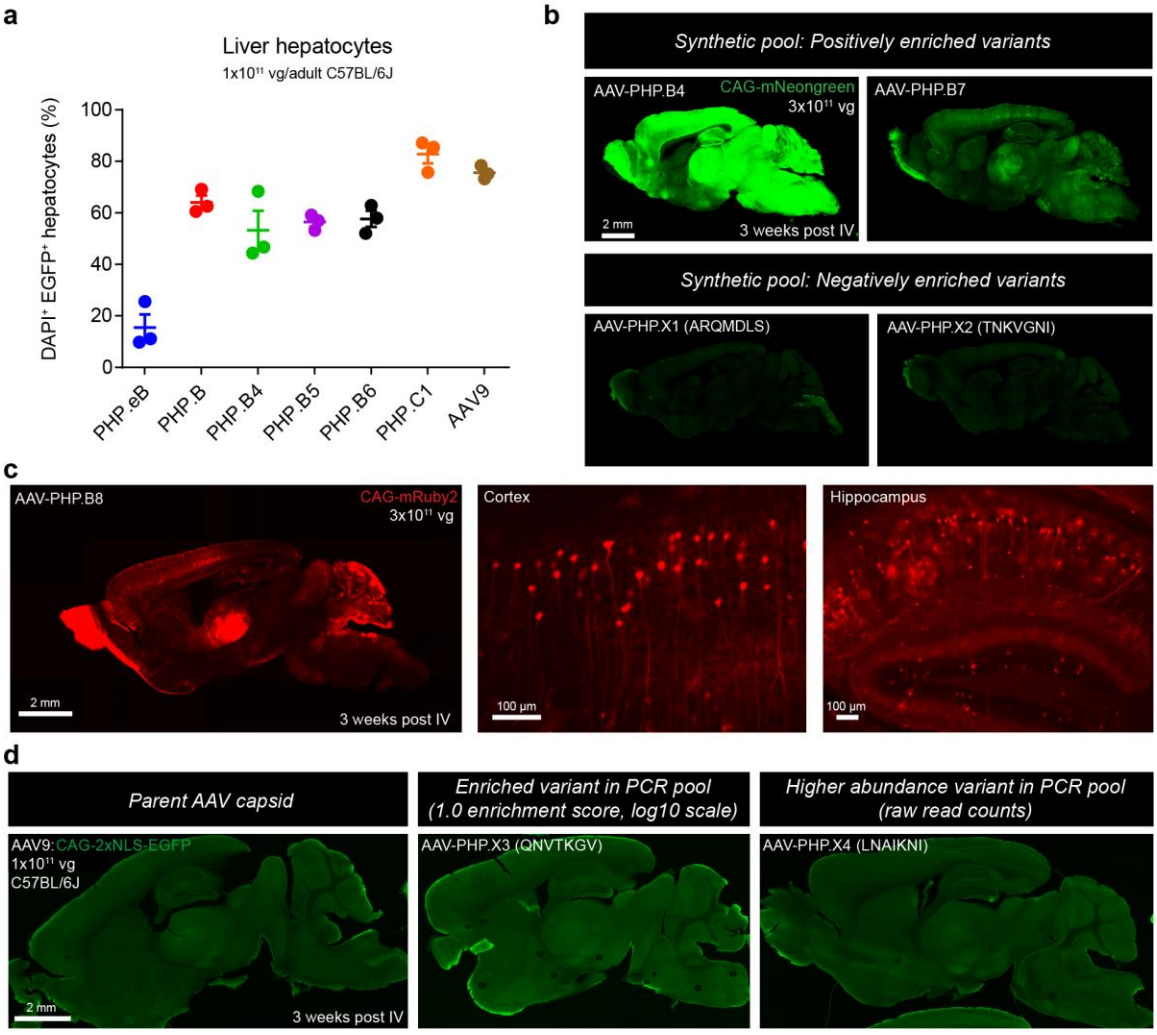
tdTomato reporter adult mouse (n=2–3 per group, $3x10^{11}$ vg dose per adult mouse, 3 weeks of expression). **(f)**, Expression of AAV-PHP.V1 carrying self-complementary (sc) scAAV:CB6-EGFP (above) and scAAV:CAG-EGFP (below). Magenta represents the lectin DyLight 594 staining (n=2-3, $3x10^{11}$ vg dose per adult C57BL/6J mouse, 2 weeks of expression). Experiments in **(c-e)** are reported from one independent trial from a fresh batch of viruses, and tittered in the same assay for dosage consistency, **(e)** and **(f)** validated in two independent trials (n = 2 per group).



Supplementary Figure 3.S5: AAV-PHP.V2 Variant Exhibits Biased Transduction Towards Brain Vascular Cells.

(a), Transduction of mouse brain by the AAV-PHP.V2 variant and control AAV9, carrying the ssAAV:CAG-mNeonGreen (n = 3, $3x10^{11}$ vg i.v. dose per C57BL/6J adult mouse, 3 weeks of expression) is shown. The sagittal brain images (left) are imaged under the same settings (also matched to the settings of sagittal brain images in Fig. 3c). Higher-magnification images of AAV-PHP.V2 transduced brain sections stained with αGLUT or αS100 or αOlig2 (magenta) are shown. **(b),** Transduction of brain vasculature by AAV-PHP.V2 carrying ssAAV:CAG-DIO-EYFP (green) in Tek-Cre adult mice (left, $1x10^{12}$ vg i.v. dose per mouse, 4 weeks of expression) is shown, and its efficiency (right) is

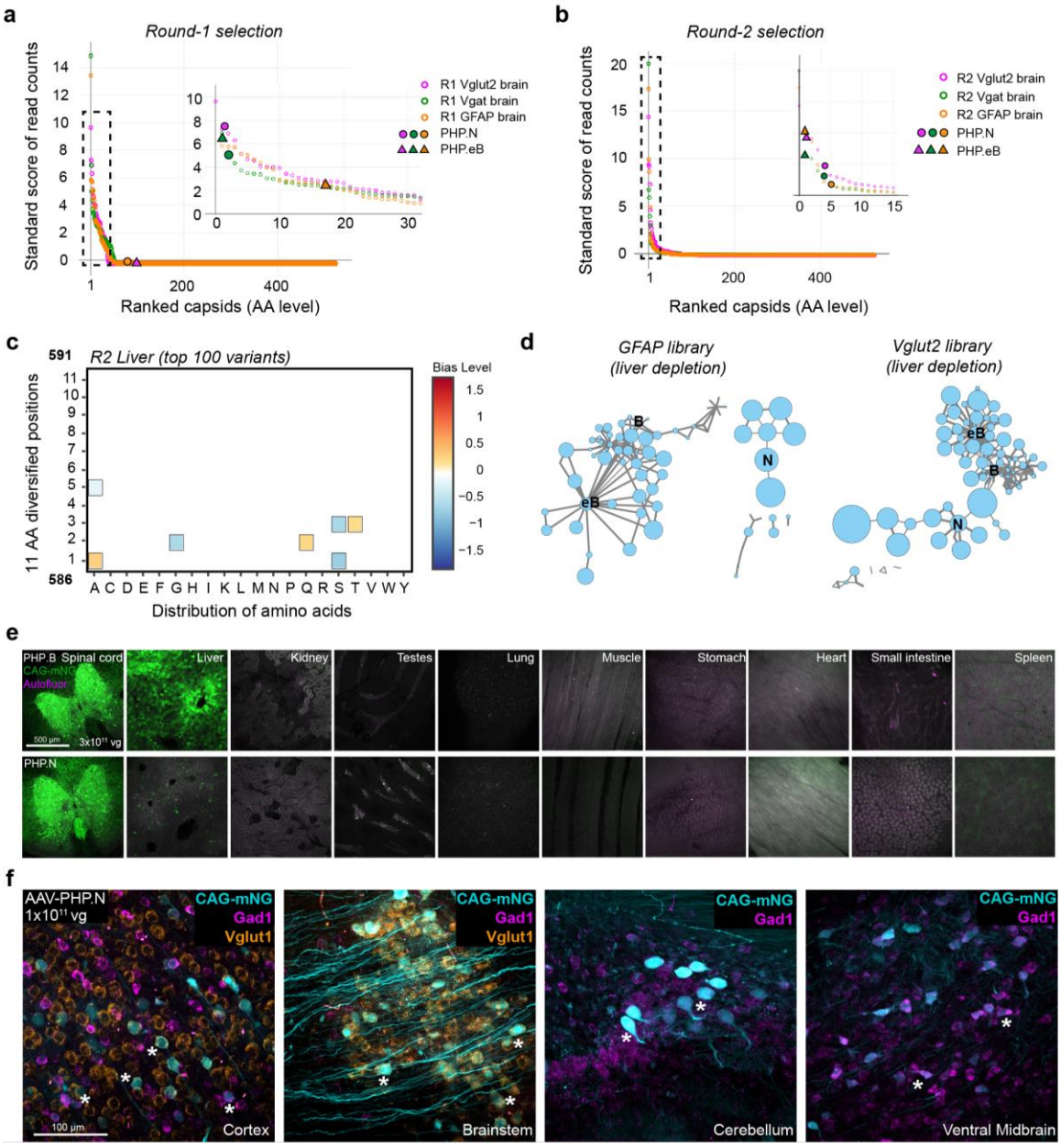
determined by the overlap of αGLUT1 staining (red) with EYFP expression across different brain areas (n = 2, mean of 3 images per brain region per mouse) **(c)**, Transduction of astrocytes by AAV-PHP.V2 in GFAP-Cre adult mouse ($1x10^{12}$ vg i.v. dose per mouse, 4 weeks of expression) is shown. Percentage of cortical S100+ cells that overlapped with EYFP expression is quantified (n = 2, mean of 3 images per mouse).



Supplementary Figure 3.S6: Further Validation Of Synthetic Pool And PCR Pool Variants Demonstrates Higher Confidence In Synthetic Pool NGS Data.

(a), Transduction levels of liver hepatocytes quantified as the percentage of $DAPI^+$ cells that are $EGFP^+$ (*n* = 3, vectors packaged with ssAAV:CAG-2xNLS-EGFP, $1x10^{11}$ vg i.v. dose per adult C57BL/6J mouse, 3 weeks of expression, mean±S.E.M, 4 images per mouse per group. One-way ANOVA non-parametric Kruskal-Wallis test (approximate P-value of 0.0088), and follow-up multiple comparisons using uncorrected Dunn's test (P-value of 0.0353 for PHP.eB versus PHP.B, 0.0005 for

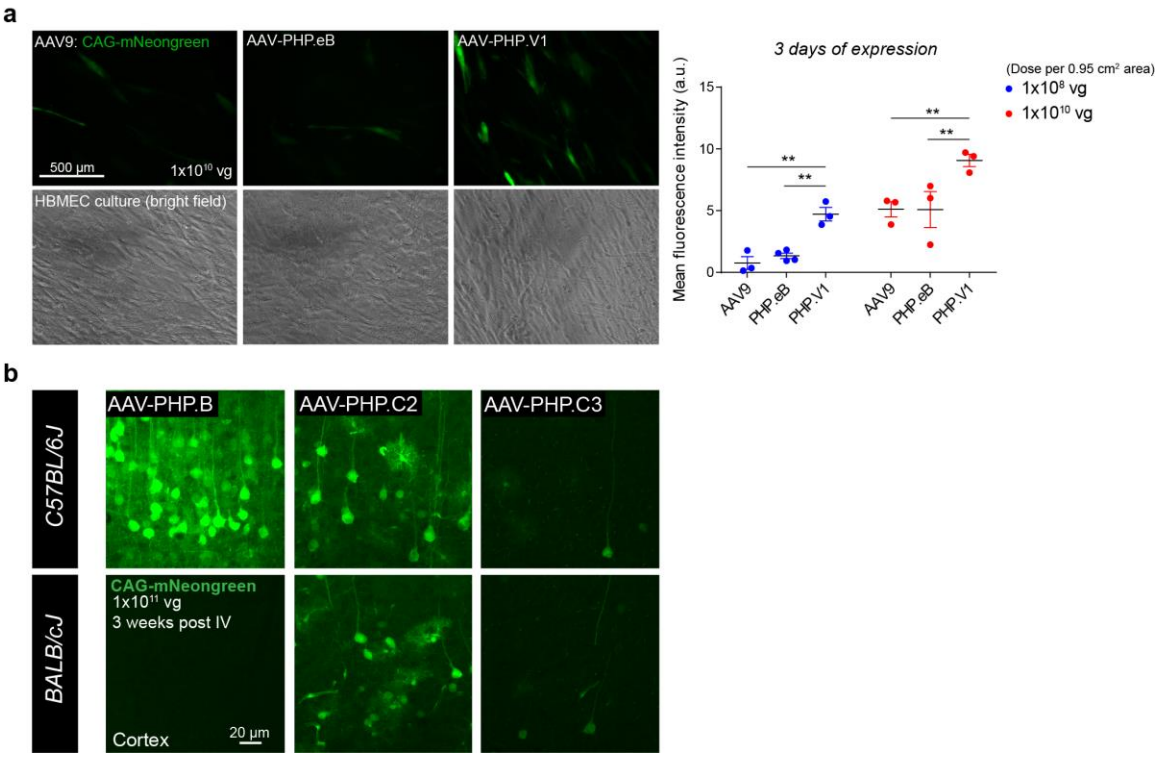
PHP.eB versus PHP.C1, 0.0025 for PHP.eB versus AAV9, 0.0179 for PHP.B4 versus PHP.C1, 0.0253 for PHP.B5 versus PHP.C1, 0.0414 for PHP.B6 versus PHP.C1) is performed. **(b)**, Transduction of brain tissue by AAV-PHP.B4, B7, AAV-PHP.X1 (ARQMDLS), and AAV-PHP.X2 (TNKVGNI) packaging ssAAV:CAG-mNeonGreen genome (n = 3, $1x10^{11}$ vg i.v. dose per adult C57BL/6J mouse, 3 weeks of expression), imaged under the same settings as that of AAV9 and AAV-PHP.V1 sagittal brain images in Fig. 3c. **(c)**, Transduction of the brain by AAV-PHP.B8 using the ssAAV:CAG-mRuby2 genome (n = 3, $3x10^{11}$ vg i.v. dose per adult C57BL/6J mouse, 3 weeks of expression). **(d)**, Transduction of AAV9 (left), AAV-PHP.X3 (QNVTKGV) (middle) and AAV-PHP.X4 (LNAIKNI) (right) vectors packaging ssAAV:CAG-2xNLS-EGFP (n = 2, $1x10^{11}$ vg i.v. dose per adult C57BL/6J mouse, 3 weeks of expression). **(a-d)** data is reported from one independent trial.



Supplementary Figure 3.S7: Evolution Of The AAV-PHP.B Capsid By Diversifying Amino Acid Positions 587-597.

(a), Distributions of R1 and **(b)**, R2 brain libraries (at AA level, SS of RCs sorted in decreasing order of scores) is shown. The SS for AAV-PHP.N and AAV-PHP.eB across libraries are mapped on the zoomed-in view of this plot (dotted line box). **(c)**, Heatmap of AA distributions across the diversified region of the enriched variants from R2 liver library (top 100 sequences) normalized to the R2 virus (input library). **(d)**, Clustering analysis of enriched variants from GFAP and Vglut2 brain libraries are shown with size of nodes

representing their relative depletion in liver, and the thickness of edges (connecting lines) representing their relative identity between nodes. **(e)**, Expression of AAV-PHP.B (above) and AAV-PHP.N (below) packaged with ssAAV:CAG-mNeonGreen across all organs is shown (n = 3, $3x10^{11}$ vg i.v. dose per adult C57BL/6J mouse, 3 weeks of expression). The background auto fluorescence is in magenta. **(f)**, Transduction of mouse brain by the AAV-PHP.N variant, carrying the CAG promoter that drives the expression of mNeonGreen (*n* = 3, $1x10^{11}$ vg i.v. dose per C57BL/6J adult mouse, 3 weeks of expression) is shown. Fluorescence *in situ* hybridization chain reaction (FITC-HCR) was used to label excitatory neurons with Vglut1 and inhibitory neurons with Gad1. Few cells where EGFP expression co-localized with specific cell markers are highlighted by asterisks symbol.



Supplementary Figure 3.S8: Investigation Of AAV-PHP Variants Across Different Mouse Strains And In Vitro Human Brain Microvascular Endothelial Cells.

(a), Transduction of AAV9, AAV-PHP.eB and AAV-PHP.V1 in human brain microvascular endothelial cell culture (HBMEC) is shown. The vectors were packaged with ssAAV:CAG-mNeongreen. The mean fluorescence intensity across the groups were quantified (n=3 tissue culture wells of 0.95 cm^2 surface area per group, 3 images per well per group per dose was imaged after three days of expression, doses $1x10^8$ vg and $1x10^{10}$ vg per 0.95 cm^2 surface area). A two-way ANOVA with correction for multiple comparisons using Tukey's test gave adjusted P-value of 0.0051 for AAV9 versus PHP.V1, 0.0096 for PHP.eB versus PHP.V1, 0.8222 for

AAV9 versus PHP.eB for $1x10^8$ vg, and 0.0052 for AAV9 versus PHP.V1, 0.0049 for PHP.eB versus PHP.V1, 0.9996 for AAV9 versus PHP.eB for $1x10^{10}$ vg (**P ≤ 0.01, is shown and P > 0.05 is not shown on the plot; mean ± S.E.M., 95% CI). **(b)**, The transduction of cortex brain region by AAV-PHP.B, AAV-PHP.C2 and AAV-PHP.C3 across two different mouse strains: C57BL/6J and BALB/cJ are shown. The vectors were packaged with ssAAV:CAG-mNeongreen (*n* = 2-3 per group, $1x10^{11}$ vg i.v. dose per adult mouse, 3 weeks of expression), and imaged under the same settings. The data reported in **(a,b)** are from one independent trial where all viruses were freshly prepared and tittered in the same assay for dosage consistency, with additional validation for AAV-PHP.C2 and AAV-PHP.C3 in an independent trial for BALB/cJ.

Supplementary Video 3.V1: Brain-Wide transduction of endothelial cells upon systemic delivery of the AAV-PHP.V1 capsid.

ssAAV-PHP.V1:CAG-DIO-EYFP vector was systemically delivered at a dose of 1×10^{12} vg per adult Tek-Cre mouse (*n* = 2). After 4 weeks of expression, mice were transcardially perfused and fixed with 4% PFA. Fixed brain hemispheres (one per mouse) were subjected to staining with αGFP primary and Alexa Fluor 633 secondary along with tissue clearing as described in the iDISCO protocol (the other fixed hemispheres from the same experiment were sliced sagitally (100-µm thickness), stained with αGLUT1, imaged and quantified to validate expression; data shown in ***Figure 3.4f,g***). One of the cleared brain hemisphere was imaged using a commercial light-sheet microscope (Lavision BioTec) with a custom objective lens (×4). The resulting image files were reorganized by a custom MATLAB script to allow stitching with TeraStitcher. For three-dimensional visualization, Imaris (Bitplane) was used. The data are reported from one independent trial.

Supplementary Dataset 3.D1: 7-mer-i spike-in library recovery in brain tissue across Cre transgenic lines.

Sheet 1 contains the list of peptides included in the *7-mer-i spike-in* library, along with their predicted enrichment in brain tissue as per prior study and their validation using the new method, M-CREATE. Sheet 2 includes the enrichment scores of the *spike-in* library in brain tissue across different Cre transgenic lines.

Chapter 4

FUTURE DIRECTIONS

4.1 EXPANSION OF AAV VECTOR TOOLKIT

The AAV-PHP vectors identified so far using CREATE and M-CREATE were predominantly focused on CNS cell types given the lack of precise, non-invasive gene delivery tools for neuroscience applications, and a lot remains to be done to develop a complete AAV-CNS toolkit. For instance, there are no good AAV vectors for targeting other widely studied CNS cell types such as microglia or oligodendrocyte by intravenous delivery[39,268,269]. In addition, there is room for further improvement of the AAV-PHP vectors' efficiency and specificity, especially for AAV-PHP.C-like variants that may have the potential for translation across species.

With the M-CREATE method in place, we can now push forward to apply similar parallel selections across other cell types of interest such as the peripheral nervous system. Although CREATE selected AAV-PHP.S vector from GFAP-Cre (astrocyte) selection, and is used for applications requiring gene delivery to sensory and enteric nervous system[134,270], there is room for improving the selection using M-CREATE with the hope to find vectors with higher efficiency and specificity. And one could expand this further to do similar selections to obtain cell-type or organ specific variants, provided there are Cre transgenic lines that are highly specific to the cell populations of interest. Often times, this may not be feasible, and may require alternate strategies to supplement for the loss of selection pressures across different biological systems, or even require additional rounds of evolution, or rational or semi rational designs to achieve the desired vector properties.

4.2 INVESTIGATION OF MECHANISM OF ENGINEERED VECTORS

While the directed evolution approach is very promising to yield the vectors with desired properties, understanding mechanism of how these vectors gain entry to cell types of interest can eventually allow us to understand how these vectors may perform across species. For instance, AAV-PHP.B vectors that performed really well in select strains in rodents uses a cell surface protein receptor, *ly6a*, the homolog of which is not seen in humans[164,165,167,271]. This is a very useful information to the community who would invest in the possibility of translation of such vectors in gene therapy applications.

The engineered AAV-PHP.C variants identified by M-CREATE have distinct amino acid signatures. This opens up the possibility that they may act on different cell surface protein receptors, or in other words, may have different mechanisms of entry into cells, and this may or may not necessarily coincide across species. Hence for translation purposes, studies similar to the ones conducted to identify *ly6a* would be very useful to make an informed decision on the choice of the AAV-PHP.C vectors for translation. In addition to the protein target identification, we could also make progress towards understanding the kinetics of vector biodistribution *in vivo* using new imaging modalities such as PET (positron emission tomography)[272] than the traditional methods involving quantification of viral genomes or expression of a protein over time.

4.3 DEVELOPMENT OF TECHNOLOGIES TO EMPOWER M-CREATE

While M-CREATE speeds up the selection process and the possible outcomes from a given experiment, this method is hit by the bottleneck of characterizing individual variants *in vivo*. Given the high yield of enriched variants or families of variants that may all be interesting in different aspects, it adds a huge burden on cost, labor, and time to screen them individually. Hence there is a need to increase the throughput of the screening process.

To this end, one could use the output from M-CREATE selection and feed the selected library of variants into one of a fast-screen technique such as the single-cell RNA sequencing[250] or by RNA probe based hybridization methods[260], either of which may have the power to rank variants based on their tropism, transduction efficiency and specificity *in vivo*.

The high-confidence outcome from M-CREATE can also serve as a good learning dataset for machine learning or *in silico* pipelines that can capture information that were not previously identified in our data analysis. There are different algorithms that have been used to perform *in silico* evolution or protein structure predictions which could potentially be useful to apply to our datasets[69,273–275]. Such approaches can improve our understanding on the design of our selections or the hidden outcomes and alleviate some of the labor intensive bench-work involved in the traditional large scale selection studies.

4.4 POTENTIAL APPLICATIONS OF AAV VECTOR TOOLKIT

The AAV-PHP vectors identified by M-CREATE could have broad applications in neuroscience, and can be put to immediate use in studies involving rodents. For instance, previously, AAV-PHP.B vectors were used in basic neuroscience studies in mice[163,189,191], in neurological disease models[158], and in preclinical gene therapy applications[161,276,277]. And one can hope to see similar applications from vectors like AAV-PHP.N or AAV-PHP.Cs. The AAV-PHP.V variants will open up new application in the field that studies BBB malformations, that are widely associated with various neurodegenerative diseases including Alzheimer's[247,249], or aging studies[278], or other neuropathologies[215,279,280]. For applications beyond rodents, further investigations are required for AAV-PHP.Cs across species to determine their use as gene delivery vector.

BIBLIOGRAPHY

1. Bisht, S., Bhakta, G., Mitra, S. & Maitra, A. pDNA loaded calcium phosphate nanoparticles: highly efficient non-viral vector for gene delivery. *Int. J. Pharm.* **288**, 157–168 (2005).
2. Ghosh, P. S., Kim, C.-K., Han, G., Forbes, N. S. & Rotello, V. M. Efficient Gene Delivery Vectors by Tuning the Surface Charge Density of Amino Acid-Functionalized Gold Nanoparticles. *ACS Nano* **2**, 2213–2218 (2008).
3. Borchard, G. Chitosans for gene delivery. *Adv. Drug Deliv. Rev.* **52**, 145–150 (2001).
4. Kean, T., Roth, S. & Thanou, M. Trimethylated chitosans as non-viral gene delivery vectors: Cytotoxicity and transfection efficiency. *J. Controlled Release* **103**, 643–653 (2005).
5. Köping-Höggård, M. *et al.* Chitosan as a nonviral gene delivery system. Structure–property relationships and characteristics compared with polyethylenimine in vitro and after lung administration in vivo. *Gene Ther.* **8**, 1108–1121 (2001).
6. Thanou, M., Florea, B. I., Geldof, M., Junginger, H. E. & Borchard, G. Quaternized chitosan oligomers as novel gene delivery vectors in epithelial cell lines. *Biomaterials* **23**, 153–159 (2002).
7. Martin, M. E. & Rice, K. G. Peptide-guided gene delivery. *AAPS J.* **9**, E18–E29 (2007).
8. Buck, J., Grossen, P., Cullis, P. R., Huwyler, J. & Witzigmann, D. Lipid-Based DNA Therapeutics: Hallmarks of Non-Viral Gene Delivery. *ACS Nano* **13**, 3754–3782 (2019).
9. Dang, J. M. & Leong, K. W. Natural polymers for gene delivery and tissue engineering. *Adv. Drug Deliv. Rev.* **58**, 487–499 (2006).
10. Pack, D. W., Hoffman, A. S., Pun, S. & Stayton, P. S. Design and development of polymers for gene delivery. *Nat. Rev. Drug Discov.* **4**, 581–593 (2005).

11. Van Bruggen, C., Hexum, J. K., Tan, Z., Dalal, R. J. & Reineke, T. M. Nonviral Gene Delivery with Cationic Glycopolymers. *Acc. Chem. Res.* **52**, 1347–1358 (2019).

12. Cevher, E., Sezer, A. D. & Çağlar, E. Ş. Gene Delivery Systems: Recent Progress in Viral and Non-Viral Therapy. *Recent Adv. Nov. Drug Carr. Syst.* (2012) doi:10.5772/53392.

13. Wu, P. *et al.* Non-viral gene delivery systems for tissue repair and regeneration. *J. Transl. Med.* **16**, 29 (2018).

14. Tomlinson, E. & Rolland, A. P. Controllable gene therapy pharmaceutics of non-viral gene delivery systems. *J. Controlled Release* **39**, 357–372 (1996).

15. Gascón, A. R., Pozo-Rodríguez, A. del & Solinís, M. Á. Non-Viral Delivery Systems in Gene Therapy. *Gene Ther. - Tools Potential Appl.* (2013) doi:10.5772/52704.

16. Vannucci, L., Lai, M., Chiuppesi, F., Ceccherini-Nelli, L. & Pistello, M. Viral vectors: a look back and ahead on gene transfer technology. *New Microbiol.* **36**, 1–22 (2013).

17. Thomas, C. E., Ehrhardt, A. & Kay, M. A. Progress and problems with the use of viral vectors for gene therapy. *Nat. Rev. Genet.* **4**, 346–358 (2003).

18. Lotze, M. T. & Kost, T. A. Viruses as gene delivery vectors: Application to gene function, target validation, and assay development. *Cancer Gene Ther.* **9**, 692–699 (2002).

19. Kost, T. A. & Condreay, J. P. Recombinant baculoviruses as mammalian cell gene-delivery vectors. *Trends Biotechnol.* **20**, 173–180 (2002).

20. Cone, R. D. & Mulligan, R. C. High-efficiency gene transfer into mammalian cells: generation of helper-free recombinant retrovirus with broad mammalian host range. *Proc. Natl. Acad. Sci.* **81**, 6349–6353 (1984).

21. Anderson, W. F. September 14, 1990: The Beginning. *Hum. Gene Ther.* **1**, 371–372 (1990).

22. Hacein-Bey-Abina, S. *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.* **118**, 3132–3142 (2008).

23. Wu, X., Li, Y., Crise, B. & Burgess, S. M. Transcription Start Regions in the Human Genome Are Favored Targets for MLV Integration. *Science* **300**, 1749–1751 (2003).

24. Hacein-Bey-Abina, S. *et al.* A Serious Adverse Event after Successful Gene Therapy for X-Linked Severe Combined Immunodeficiency. *N. Engl. J. Med.* **348**, 255–256 (2003).

25. Check, E. Gene therapy put on hold as third child develops cancer. *Nature* **433**, 561–561 (2005).

26. Barquinero, J., Eixarch, H. & Pérez-Melgosa, M. Retroviral vectors: new applications for an old tool. *Gene Ther.* **11**, S3–S9 (2004).

27. McTaggart, S. & Al-Rubeai, M. Retroviral vectors for human gene delivery. *Biotechnol. Adv.* **20**, 1–31 (2002).

28. Anson, D. S. The use of retroviral vectors for gene therapy-what are the risks? A review of retroviral pathogenesis and its relevance to retroviral vector-mediated gene delivery. *Genet. Vaccines Ther.* **2**, 9 (2004).

29. Vargas, J. E. *et al.* Retroviral vectors and transposons for stable gene therapy: advances, current challenges and perspectives. *J. Transl. Med.* **14**, 288 (2016).

30. Biasco, L., Baricordi, C. & Aiuti, A. Retroviral Integrations in Gene Therapy Trials. *Mol. Ther.* **20**, 709–716 (2012).

31. Cronin, J., Zhang, X.-Y. & Reiser, J. Altering the Tropism of Lentiviral Vectors through Pseudotyping. https://www.ingentaconnect.com/content/ben/cgt/2005/00000005/00000004/art00003 (2005) doi:info:doi/10.2174/1566523054546224.

32. Zufferey, R. *et al.* Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery. *J. Virol.* **72**, 9873–9880 (1998).

33. Dull, T. *et al.* A Third-Generation Lentivirus Vector with a Conditional Packaging System. *J. Virol.* **72**, 8463–8471 (1998).

34. Miyoshi, H., Blömer, U., Takahashi, M., Gage, F. H. & Verma, I. M. Development of a Self-Inactivating Lentivirus Vector. *J. Virol.* **72**, 8150–8157 (1998).

35. Lewis, P., Hensel, M. & Emerman, M. Human immunodeficiency virus infection of cells arrested in the cell cycle. *EMBO J.* **11**, 3053–3058 (1992).

36. Poeschla, E. M., Wong-Staal, F. & Looney, D. J. Efficient transduction of nondividing human cells by feline immunodeficiency virus lentiviral vectors. *Nat. Med.* **4**, 354–357 (1998).

37. Naldini, L. *et al.* In Vivo Gene Delivery and Stable Transduction of Nondividing Cells by a Lentiviral Vector. *Science* **272**, 263–267 (1996).

38. Abbas-Terki, T., Blanco-Bose, W., Déglon, N., Pralong, W. & Aebischer, P. Lentiviral-Mediated RNA Interference. *Hum. Gene Ther.* **13**, 2197–2201 (2002).

39. Maes, M., Colombo, G., Schulz, R. & Siegert, S. Targeting microglia with lentivirus and AAV: recent advances and remaining challenges. *Neurosci. Lett.* 134310 (2019) doi:10.1016/j.neulet.2019.134310.

40. Lois, C., Hong, E. J., Pease, S., Brown, E. J. & Baltimore, D. Germline Transmission and Tissue-Specific Expression of Transgenes Delivered by Lentiviral Vectors. *Science* **295**, 868–872 (2002).

41. Koike-Yusa, H., Li, Y., Tan, E.-P., Velasco-Herrera, M. D. C. & Yusa, K. Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. *Nat. Biotechnol.* **32**, 267–273 (2014).

42. Moffat, J. *et al.* A Lentiviral RNAi Library for Human and Mouse Genes Applied to an Arrayed Viral High-Content Screen. *Cell* **124**, 1283–1298 (2006).

43. Aiuti, A. *et al.* Lentiviral Hematopoietic Stem Cell Gene Therapy in Patients with Wiskott-Aldrich Syndrome. *Science* **341**, (2013).

44. Biffi, A. *et al.* Lentiviral Hematopoietic Stem Cell Gene Therapy Benefits Metachromatic Leukodystrophy. *Science* **341**, 1233158–1233158 (2013).

45. Yáñez-Muñoz, R. J. *et al.* Effective gene therapy with nonintegrating lentiviral vectors. *Nat. Med.* **12**, 348–353 (2006).

46. Milone, M. C. & O'Doherty, U. Clinical use of lentiviral vectors. *Leukemia* **32**, 1529–1541 (2018).

47. Wilson, J. M. Adenoviruses as Gene-Delivery Vehicles. *N. Engl. J. Med.* **334**, 1185–1187 (1996).

48. Benihoud, K., Yeh, P. & Perricaudet, M. Adenovirus vectors for gene delivery. *Curr. Opin. Biotechnol.* **10**, 440–447 (1999).

49. Parks, R. J. *et al.* A helper-dependent adenovirus vector system: Removal of helper virus by Cre-mediated excision of the viral packaging signal. *Proc. Natl. Acad. Sci.* **93**, 13565–13570 (1996).

50. Ledgerwood, J. E. *et al.* Chimpanzee Adenovirus Vector Ebola Vaccine. *N. Engl. J. Med.* **376**, 928–938 (2017).

51. Singh, S., Kumar, R. & Agrawal, B. Adenoviral Vector-Based Vaccines and Gene Therapies: Current Status and Future Prospects. *Adenoviruses* (2018) doi:10.5772/intechopen.79697.

52. Barouch, D. H. & Nabel, G. J. Adenovirus Vector-Based Vaccines for Human Immunodeficiency Virus Type 1. *Hum. Gene Ther.* **16**, 149–156 (2005).

53. Schiedner, G. *et al.* Genomic DNA transfer with a high-capacity adenovirus vector results in improved in vivo gene expression and decreased toxicity. *Nat. Genet.* **18**, 180–183 (1998).

54. Dmitriev, I. *et al.* An Adenovirus Vector with Genetically Modified Fibers Demonstrates Expanded Tropism via Utilization of a Coxsackievirus and Adenovirus Receptor-Independent Cell Entry Mechanism. *J. Virol.* **72**, 9706–9713 (1998).

55. Shayakhmetov, D. M., Papayannopoulou, T., Stamatoyannopoulos, G. & Lieber, A. Efficient Gene Transfer into Human CD34+ Cells by a Retargeted Adenovirus Vector. *J. Virol.* **74**, 2567–2583 (2000).

56. Hartman, Z. C., Appledorn, D. M. & Amalfitano, A. Adenovirus vector induced innate immune responses: Impact upon efficacy and toxicity in gene therapy and vaccine applications. *Virus Res.* **132**, 1–14 (2008).

57. Yang, Y., Su, Q. & Wilson, J. M. Role of viral antigens in destructive cellular immune responses to adenovirus vector-transduced cells in mouse lungs. *J. Virol.* **70**, 7209–7212 (1996).

58. Marshall, E. Gene Therapy Death Prompts Review of Adenovirus Vector. *Science* **286**, 2244–2245 (1999).

59. Raper, S. E. *et al.* Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer. *Mol. Genet. Metab.* **80**, 148–158 (2003).

60. Srivastava, A., Lusby, E. W. & Berns, K. I. Nucleotide sequence and organization of the adeno-associated virus 2 genome. *J. Virol.* **45**, 555–564 (1983).

61. Samulski, R. J., Berns, K. I., Tan, M. & Muzyczka, N. Cloning of adeno-associated virus into pBR322: rescue of intact virus from the recombinant plasmid in human cells. *Proc. Natl. Acad. Sci.* **79**, 2077–2081 (1982).

62. Xie, Q. *et al.* The atomic structure of adeno-associated virus (AAV-2), a vector for human gene therapy. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 10405–10410 (2002).

63. Douar, A.-M., Poulard, K., Stockholm, D. & Danos, O. Intracellular Trafficking of Adeno-Associated Virus Vectors: Routing to the Late Endosomal Compartment and Proteasome Degradation. *J. Virol.* **75**, 1824–1833 (2001).

64. Nonnenmacher, M. & Weber, T. Intracellular Transport of Recombinant Adeno-Associated Virus Vectors. *Gene Ther.* **19**, 649–658 (2012).

65. Berry, G. E. & Asokan, A. Cellular transduction mechanisms of adeno-associated viral vectors. *Curr. Opin. Virol.* **21**, 54–60 (2016).

66. Atchison, R. W., Casto, B. C. & Hammon, W. M. ADENOVIRUS-ASSOCIATED DEFECTIVE VIRUS PARTICLES. *Science* **149**, 754–756 (1965).

67. Pereira, D. J., McCarty, D. M. & Muzyczka, N. The adeno-associated virus (AAV) Rep protein acts as both a repressor and an activator to regulate AAV transcription during a productive infection. *J. Virol.* **71**, 1079–1088 (1997).

68. Maurer, A. C. *et al.* The Assembly-Activating Protein Promotes Stability and Interactions between AAV's Viral Proteins to Nucleate Capsid Assembly. *Cell Rep.* **23**, 1817–1830 (2018).

69. Ogden, P. J., Kelsic, E. D., Sinai, S. & Church, G. M. Comprehensive AAV capsid fitness landscape reveals a viral gene and enables machine-guided design. *Science* **366**, 1139–1143 (2019).

70. Janik, J. E., Huston, M. M. & Rose, J. A. Adeno-associated virus proteins: origin of the capsid components. *J. Virol.* **52**, 591–597 (1984).

71. Rescue and Replication Signals of the Adeno-associated Virus 2 Genome | Elsevier Enhanced Reader. https://reader.elsevier.com/reader/sd/pii/S0022283685703981?token=D1C4D765FE5A5FD55D98E0DB79DFB209CA63F1D5863619DBE752C76BEF61C421F7640342BDC14C41682ADA629AEEE4F4 doi:10.1006/jmbi.1995.0398.

72. Bohenzky, R. A. & Berns, K. I. Interactions between the termini of adeno-associated virus DNA. *J. Mol. Biol.* **206**, 91–100 (1989).

73. Yan, Z., Zak, R., Zhang, Y. & Engelhardt, J. F. Inverted Terminal Repeat Sequences Are Important for Intermolecular Recombination and Circularization of Adeno-Associated Virus Genomes. *J. Virol.* **79**, 364–379 (2005).

74. Linden, R. M., Ward, P., Giraud, C., Winocour, E. & Berns, K. I. Site-specific integration by adeno-associated virus. *Proc. Natl. Acad. Sci.* **93**, 11288–11294 (1996).

75. Hoggan, M. D., Blacklow, N. R. & Rowe, W. P. Studies of small DNA viruses found in various adenovirus preparations: physical, biological, and immunological characteristics. *Proc. Natl. Acad. Sci. U. S. A.* **55**, 1467–1474 (1966).

76. Hoggan, M. D. Adenovirus associated viruses. *Prog. Med. Virol. Fortschritte Med. Virusforsch. Progres En Virol. Medicale* **12**, 211–239 (1970).

77. McLaughlin, S. K., Collis, P., Hermonat, P. L. & Muzyczka, N. Adeno-associated virus general transduction vectors: analysis of proviral structures. *J. Virol.* **62**, 1963–1973 (1988).

78. Hermonat, P. L. & Muzyczka, N. Use of adeno-associated virus as a mammalian DNA cloning vector: transduction of neomycin resistance into mammalian tissue culture cells. *Proc. Natl. Acad. Sci. U. S. A.* **81**, 6466–6470 (1984).

79. Muzyczka, N. Use of Adeno-Associated Virus as a General Transduction Vector for Mammalian Cells. in *Viral Expression Vectors* (ed. Muzyczka, N.) 97–129 (Springer, 1992). doi:10.1007/978-3-642-75608-5_5.

80. Vance, M. A., Mitchell, A. & Samulski, R. J. AAV Biology, Infectivity and Therapeutic Use from Bench to Clinic. *Gene Ther. - Princ. Chall.* (2015) doi:10.5772/61988.

81. Wu, Z., Yang, H. & Colosi, P. Effect of Genome Size on AAV Vector Packaging. *Mol. Ther.* **18**, 80–86 (2010).

82. Van Vliet, K. M., Blouin, V., Brument, N., Agbandje-McKenna, M. & Snyder, R. O. The role of the adeno-associated virus capsid in gene transfer. *Methods Mol. Biol. Clifton NJ* **437**, 51–91 (2008).

83. McCarty, D. M., Young, S. M. & Samulski, R. J. Integration of Adeno-Associated Virus (AAV) and Recombinant AAV Vectors. *Annu. Rev. Genet.* **38**, 819–845 (2004).

84. Carter, P. J. & Samulski, R. J. Adeno-associated viral vectors as gene delivery vehicles. *Int. J. Mol. Med.* **6**, 17–44 (2000).

85. Schultz, B. R. & Chamberlain, J. S. Recombinant Adeno-associated Virus Transduction and Integration. *Mol. Ther. J. Am. Soc. Gene Ther.* **16**, 1189–1199 (2008).

86. Vlasov, K., Van Dort, C. J. & Solt, K. Chapter Eleven - Optogenetics and Chemogenetics. in *Methods in Enzymology* (eds. Eckenhoff, R. G. & Dmochowski, I. J.) vol. 603 181–196 (Academic Press, 2018).

87. Iyer, S. M. *et al.* Virally mediated optogenetic excitation and inhibition of pain in freely moving non-transgenic mice. *Nat. Biotechnol.* **32**, 274–278 (2014).

88. Zingg, B. *et al.* AAV-Mediated Anterograde Transsynaptic Tagging: Mapping Corticocollicular Input-Defined Neural Pathways for Defense Behaviors. *Neuron* **93**, 33–47 (2017).

89. Krook-Magnuson, E. & Soltesz, I. Beyond the hammer and the scalpel: selective circuit control for the epilepsies. *Nat. Neurosci.* **18**, 331–338 (2015).

90. Lawlor, P. A. *et al.* Novel rat Alzheimer's disease models based on AAV-mediated gene transfer to selectively increase hippocampal Abeta levels. *Mol. Neurodegener.* **2**, 11 (2007).

91. Potashkin, J. A., Blume, S. R. & Runkle, N. K. Limitations of Animal Models of Parkinson's Disease. *Parkinson’s Disease* vol. 2011 e658083 https://www.hindawi.com/journals/pd/2011/658083/ (2011).

92. Jucker, M. The benefits and limitations of animal models for translational research in neurodegenerative diseases. *Nat. Med.* **16**, 1210–1214 (2010).

93. 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).

94. Wang, D., Zhang, F. & Gao, G. CRISPR-Based Therapeutic Genome Editing: Strategies and In Vivo Delivery by AAV Vectors. *Cell* **181**, 136–150 (2020).

95. Matharu, N. *et al.* CRISPR-mediated activation of a promoter or enhancer rescues obesity caused by haploinsufficiency. *Science* **363**, (2019).

96. Lau, C.-H., Ho, J. W.-T., Lo, P. K. & Tin, C. Targeted Transgene Activation in the Brain Tissue by Systemic Delivery of Engineered AAV1 Expressing CRISPRa. *Mol. Ther. Nucleic Acids* **16**, 637–649 (2019).

97. Zhang, Y. *et al.* Enhanced CRISPR-Cas9 correction of Duchenne muscular dystrophy in mice by a self-complementary AAV delivery system. *Sci. Adv.* **6**, eaay6812 (2020).

98. Grames, M. S. *et al.* Cre-dependent AAV vectors for highly targeted expression of disease-related proteins and neurodegeneration in the substantia nigra. *FASEB J.* **32**, 4420–4427 (2018).

99. Lykken, E. A., Shyng, C., Edwards, R. J., Rozenberg, A. & Gray, S. J. Recent progress and considerations for AAV gene therapies targeting the central nervous system. *J. Neurodev. Disord.* **10**, 16 (2018).

100. Naso, M. F., Tomkowicz, B., Perry, W. L. & Strohl, W. R. Adeno-Associated Virus (AAV) as a Vector for Gene Therapy. *BioDrugs Clin. Immunother. Biopharm. Gene Ther.* **31**, 317–334 (2017).

101. Wang, D., Tai, P. W. L. & Gao, G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat. Rev. Drug Discov.* **18**, 358–378 (2019).

102. Kaplitt, M. G. *et al.* Long-term gene expression and phenotypic correction using adeno-associated virus vectors in the mammalian brain. *Nat. Genet.* **8**, 148–154 (1994).

103. Flotte, T. *et al.* A phase I study of an adeno-associated virus-CFTR gene vector in adult CF patients with mild lung disease. *Hum. Gene Ther.* **7**, 1145–1159 (1996).

104. Cideciyan, A. V. *et al.* Human gene therapy for RPE65 isomerase deficiency activates the retinoid cycle of vision but with slow rod kinetics. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 15112–15117 (2008).

105. Hauswirth, W. W. *et al.* Treatment of Leber Congenital Amaurosis Due to RPE65 Mutations by Ocular Subretinal Injection of Adeno-Associated Virus Gene Vector: Short-Term Results of a Phase I Trial. *Hum. Gene Ther.* **19**, 979–990 (2008).

106. Bainbridge, J. W. B. *et al.* Effect of Gene Therapy on Visual Function in Leber's Congenital Amaurosis. *N. Engl. J. Med.* **358**, 2231–2239 (2008).

107. Ylä-Herttuala, S. Endgame: Glybera Finally Recommended for Approval as the First Gene Therapy Drug in the European Union. *Mol. Ther.* **20**, 1831–1832 (2012).

108. 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).

109. Maguire, A. M. *et al.* Efficacy, Safety, and Durability of Voretigene Neparvovec-rzyl in RPE65 Mutation-Associated Inherited Retinal Dystrophy: Results of Phase 1 and 3 Trials. *Ophthalmology* **126**, 1273–1285 (2019).

110. Russell, S. *et al.* Efficacy and safety of voretigene neparvovec (AAV2-hRPE65v2) in patients with RPE65-mediated inherited retinal dystrophy: a randomised, controlled, open-label, phase 3 trial. *Lancet Lond. Engl.* **390**, 849–860 (2017).

111. Mahajan, R. Onasemnogene Abeparvovec for Spinal Muscular Atrophy: The Costlier Drug Ever. *Int. J. Appl. Basic Med. Res.* **9**, 127–128 (2019).

112. Mendell, J. R. *et al.* Single-Dose Gene-Replacement Therapy for Spinal Muscular Atrophy. *N. Engl. J. Med.* **377**, 1713–1722 (2017).

113. Lisowski, L., Tay, S. S. & Alexander, I. E. Adeno-associated virus serotypes for gene therapeutics. *Curr. Opin. Pharmacol.* **24**, 59–67 (2015).

114. Summerford, C. & Samulski, R. J. Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions. *J. Virol.* **72**, 1438–1445 (1998).

115. Terminal N-Linked Galactose Is the Primary Receptor for Adeno-associated Virus 9. https://www.jbc.org/content/286/15/13532.long.

116. Summerford, C., Bartlett, J. S. & Samulski, R. J. AlphaVbeta5 integrin: a co-receptor for adeno-associated virus type 2 infection. *Nat. Med.* **5**, 78–82 (1999).

117. Qing, K. *et al.* Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2. *Nat. Med.* **5**, 71–77 (1999).

118. Hamilton, B. A., Li, X., Pezzulo, A. A., Alaiwa, M. H. A. & Zabner, J. Polarized AAVR expression determines infectivity by AAV gene therapy vectors. *Gene Ther.* 1 (2019) doi:10.1038/s41434-019-0078-3.

119. Meyer, N. L. *et al.* Structure of the gene therapy vector, adeno-associated virus with its cell receptor, AAVR. *eLife* **8**, e44707 (2019).

120. Zhang, R. *et al.* Divergent engagements between adeno-associated viruses with their cellular receptor AAVR. *Nat. Commun.* **10**, 1–11 (2019).

121. Zhang, R. *et al.* Adeno-associated virus 2 bound to its cellular receptor AAVR. *Nat. Microbiol.* **4**, 675–682 (2019).

122. Pillay, S. *et al.* An essential receptor for adeno-associated virus infection. *Nature* **530**, 108–112 (2016).

123. Pillay, S. *et al.* Adeno-associated Virus (AAV) Serotypes Have Distinctive Interactions with Domains of the Cellular AAV Receptor. *J. Virol.* **91**, (2017).

124. Cetin, A., Komai, S., Eliava, M., Seeburg, P. H. & Osten, P. Stereotaxic gene delivery in the rodent brain. *Nat. Protoc.* **1**, 3166–3173 (2006).

125. Powell, S. K., Rivera-Soto, R. & Gray, S. J. Viral expression cassette elements to enhance transgene target specificity and expression in gene therapy. *Discov. Med.* **19**, 49–57 (2015).

126. Paterna, J.-C., Moccetti, T., Mura, A., Feldon, J. & Büeler, H. Influence of promoter and WHV post-transcriptional regulatory element on AAV-mediated transgene expression in the rat brain. *Gene Ther.* **7**, 1304–1311 (2000).

127. Choi, J.-H. *et al.* Optimization of AAV expression cassettes to improve packaging capacity and transgene expression in neurons. *Mol. Brain* **7**, 17 (2014).

128. 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).

129. de Leeuw, C. N. *et al.* rAAV-compatible MiniPromoters for restricted expression in the brain and eye. *Mol. Brain* **9**, (2016).

130. McCarty, D. M. Self-complementary AAV Vectors; Advances and Applications. *Mol. Ther.* **16**, 1648–1656 (2008).

131. Griffin, J. M. *et al.* Astrocyte-selective AAV gene therapy through the endogenous GFAP promoter results in robust transduction in the rat spinal cord following injury. *Gene Ther.* **26**, 198–210 (2019).

132. Lee, Y., Messing, A., Su, M. & Brenner, M. GFAP promoter elements required for region-specific and astrocyte-specific expression. *Glia* **56**, 481–93 (2008).

133. 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).

134. Challis, R. C. *et al.* Systemic AAV vectors for widespread and targeted gene delivery in rodents. *Nat. Protoc.* **14**, 379 (2019).

135. Jovičić, A. *et al.* Comprehensive Expression Analyses of Neural Cell-Type-Specific miRNAs Identify New Determinants of the Specification and Maintenance of Neuronal Phenotypes. *J. Neurosci.* **33**, 5127–5137 (2013).

136. Lagos-Quintana, M. *et al.* Identification of Tissue-Specific MicroRNAs from Mouse. *Curr. Biol.* **12**, 735–739 (2002).

137. Xie, J. *et al.* MicroRNA-regulated, Systemically Delivered rAAV9: A Step Closer to CNS-restricted Transgene Expression. *Mol. Ther.* **19**, 526–535 (2011).

138. Rabinowitz, J., Chan, Y. K. & Samulski, R. J. Adeno-Associated Virus (AAV) Versus Immune Response. *Viruses* **11**, 102 (2019).

139. Cocas, L. A. *et al.* Cell Type-Specific Circuit Mapping Reveals the Presynaptic Connectivity of Developing Cortical Circuits. *J. Neurosci.* **36**, 3378–3390 (2016).

140. Zacchigna, S., Zentilin, L. & Giacca, M. Adeno-associated virus vectors as therapeutic and investigational tools in the cardiovascular system. *Circ. Res.* **114**, 1827–1846 (2014).

141. Steines, B. *et al.* CFTR gene transfer with AAV improves early cystic fibrosis pig phenotypes. *JCI Insight* **1**, e88728 (2016).

142. Rocca, C. J., Ur, S. N., Harrison, F. & Cherqui, S. rAAV9 combined with renal vein injection is optimal for kidney-targeted gene delivery: conclusion of a comparative study. *Gene Ther.* **21**, 618–628 (2014).

143. Pulicherla, N. *et al.* Engineering liver-detargeted AAV9 vectors for cardiac and musculoskeletal gene transfer. *Mol. Ther. J. Am. Soc. Gene Ther.* **19**, 1070–1078 (2011).

144. Shen, S. *et al.* Engraftment of a Galactose Receptor Footprint onto Adeno-associated Viral Capsids Improves Transduction Efficiency. *J. Biol. Chem.* **288**, 28814–28823 (2013).

145.Sen, D. Improving clinical efficacy of adeno associated vectors by rational capsid bioengineering. *J. Biomed. Sci.* **21**, (2014).

146.Lee, E. J., Guenther, C. M. & Suh, J. Adeno-associated virus (AAV) vectors: Rational design strategies for capsid engineering. *Curr. Opin. Biomed. Eng.* **7**, 58–63 (2018).

147.Davis, A. S. *et al.* Rational design and engineering of a modified adeno-associated virus (AAV1)-based vector system for enhanced retrograde gene delivery. *Neurosurgery* **76**, 216–225; discussion 225 (2015).

148.Lutz, S. Beyond directed evolution - semi-rational protein engineering and design. *Curr. Opin. Biotechnol.* **21**, 734–743 (2010).

149.Maheshri, N., Koerber, J. T., Kaspar, B. K. & Schaffer, D. V. Directed evolution of adeno-associated virus yields enhanced gene delivery vectors. *Nat. Biotechnol.* **24**, 198–204 (2006).

150.Körbelin, J. *et al.* A brain microvasculature endothelial cell-specific viral vector with the potential to treat neurovascular and neurological diseases. *EMBO Mol. Med.* **8**, 609–625 (2016).

151.Dalkara, D. *et al.* In vivo-directed evolution of a new adeno-associated virus for therapeutic outer retinal gene delivery from the vitreous. *Sci. Transl. Med.* **5**, 189ra76 (2013).

152.Excoffon, K. J. D. A. *et al.* Directed evolution of adeno-associated virus to an infectious respiratory virus. *Proc. Natl. Acad. Sci.* **106**, 3865–3870 (2009).

153.Byrne, L. C. *et al.* In vivo directed evolution of AAV in the primate retina. *JCI Insight* (2020) doi:10.1172/jci.insight.135112.

154.Kwon, I. & Schaffer, D. V. Designer Gene Delivery Vectors: Molecular Engineering and Evolution of Adeno-Associated Viral Vectors for Enhanced Gene Transfer. *Pharm. Res.* **25**, 489–499 (2008).

155. Deverman, B. E., Ravina, B. M., Bankiewicz, K. S., Paul, S. M. & Sah, D. W. Y. Gene therapy for neurological disorders: progress and prospects. *Nat. Rev. Drug Discov.* **17**, 767 (2018).

156. Ojala, D. S., Amara, D. P. & Schaffer, D. V. Adeno-associated virus vectors and neurological gene therapy. *Neurosci. Rev. J. Bringing Neurobiol. Neurol. Psychiatry* **21**, 84–98 (2015).

157. 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).

158. 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).

159. Jackson, K. L., Dayton, R. D., Deverman, B. E. & Klein, R. L. Better Targeting, Better Efficiency for Wide-Scale Neuronal Transduction with the Synapsin Promoter and AAV-PHP.B. *Front. Mol. Neurosci.* **9**, (2016).

160. 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).

161. 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).

162. 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).

163. Bedbrook, C. N., Yang, K. K., Robinson, J. E., Gradinaru, V. & Arnold, F. H. Machine learning-guided channelrhodopsin engineering enables minimally-invasive optogenetics. *bioRxiv* 565606 (2019) doi:10.1101/565606.

164. Hordeaux, J. *et al.* The Neurotropic Properties of AAV-PHP.B Are Limited to C57BL/6J Mice. *Mol. Ther.* **26**, 664–668 (2018).

165. Huang, Q. *et al.* Delivering genes across the blood-brain barrier: LY6A, a novel cellular receptor for AAV-PHP.B capsids. *PLOS ONE* **14**, e0225206 (2019).

166. Matsuzaki, Y. *et al.* Neurotropic Properties of AAV-PHP.B Are Shared among Diverse Inbred Strains of Mice. *Mol. Ther.* **27**, 700–704 (2019).

167. Hordeaux, J. *et al.* The GPI-Linked Protein LY6A Drives AAV-PHP.B Transport across the Blood-Brain Barrier. *Mol. Ther.* **27**, 912–921 (2019).

168. Batista, A. R. *et al.* Ly6a Differential Expression in Blood–Brain Barrier Is Responsible for Strain Specific Central Nervous System Transduction Profile of AAV-PHP.B. *Hum. Gene Ther.* (2019) doi:10.1089/hum.2019.186.

169. Samulski, R. J. & Muzyczka, N. AAV-Mediated Gene Therapy for Research and Therapeutic Purposes. *Annu. Rev. Virol.* **1**, 427–451 (2014).

170. Hama, H. *et al.* Sca*l*eS: an optical clearing palette for biological imaging. *Nat. Neurosci.* **18**, 1518–1529 (2015).

171. Yang, B. *et al.* Single-cell phenotyping within transparent intact tissue through whole-body clearing. *Cell* **158**, 945–958 (2014).

172. Mano, T. *et al.* Whole-Brain Analysis of Cells and Circuits by Tissue Clearing and Light-Sheet Microscopy. *J. Neurosci.* **38**, 9330–9337 (2018).

173. Fosque, B. F. *et al.* Labeling of active neural circuits in vivo with designed calcium integrators. *Science* **347**, 755–760 (2015).

174. Dana, H. *et al.* Sensitive red protein calcium indicators for imaging neural activity. *eLife* **5**, e12727 (2016).

175. Hancock, J. F., Cadwallader, K., Paterson, H. & Marshall, C. J. A CAAX or a CAAL motif and a second signal are sufficient for plasma membrane targeting of ras proteins. *EMBO J.* **10**, 4033–4039 (1991).

176. Kim, J. *et al.* mGRASP enables mapping mammalian synaptic connectivity with light microscopy. *Nat. Methods* **9**, 96–102 (2012).

177. Kim, E. & Sheng, M. PDZ domain proteins of synapses. *Nat. Rev. Neurosci.* **5**, 771–781 (2004).

178. Reed, S. E., Staley, E. M., Mayginnes, J. P., Pintel, D. J. & Tullis, G. E. Transfection of mammalian cells using linear polyethylenimine is a simple and effective means of producing recombinant adeno-associated virus vectors. *J. Virol. Methods* **138**, 85–98 (2006).

179. Wright, J. F. Transient Transfection Methods for Clinical Adeno-Associated Viral Vector Production. *Hum. Gene Ther.* **20**, 698–706 (2009).

180. Xiao, X., Li, J. & Samulski, R. J. Production of High-Titer Recombinant Adeno-Associated Virus Vectors in the Absence of Helper Adenovirus. *J. Virol.* **72**, 2224–2232 (1998).

181. Ayuso, E. *et al.* High AAV vector purity results in serotype- and tissue-independent enhancement of transduction efficiency. *Gene Ther.* **17**, 503–510 (2010).

182. Grieger, J. C., Choi, V. W. & Samulski, R. J. Production and characterization of adeno-associated viral vectors. *Nat. Protoc.* **1**, 1412–1428 (2006).

183. Zolotukhin, S. *et al.* Recombinant adeno-associated virus purification using novel methods improves infectious titer and yield. *Gene Ther.* **6**, 973–985 (1999).

184. Gray, S. J. *et al.* Production of Recombinant Adeno-Associated Viral Vectors and Use in In Vitro and In Vivo Administration. *Curr. Protoc. Neurosci. Editor. Board Jacqueline N Crawley Al* **CHAPTER**, Unit4.17 (2011).

185. Yardeni, T., Eckhaus, M., Morris, H. D., Huizing, M. & Hoogstraten-Miller, S. Retro-orbital injections in mice. *Lab Anim.* **40**, 155–160 (2011).

186. Lerner, T. N. *et al.* Intact-Brain Analyses Reveal Distinct Information Carried by SNc Dopamine Subcircuits. *Cell* **162**, 635–647 (2015).

187. Tervo, D. G. R. *et al.* A Designer AAV Variant Permits Efficient Retrograde Access to Projection Neurons. *Neuron* **92**, 372–382 (2016).

188. Cai, D., Cohen, K. B., Luo, T., Lichtman, J. W. & Sanes, J. R. Improved tools for the Brainbow toolbox. *Nat. Methods* **10**, 540–547 (2013).

189. Zelikowsky, M. *et al.* The Neuropeptide Tac2 Controls a Distributed Brain State Induced by Chronic Social Isolation Stress. *Cell* **173**, 1265-1279.e19 (2018).

190. Allen, W. E. *et al.* Global Representations of Goal-Directed Behavior in Distinct Cell Types of Mouse Neocortex. *Neuron* **94**, 891-907.e6 (2017).

191. Hillier, D. *et al.* Causal evidence for retina-dependent and -independent visual motion computations in mouse cortex. *Nat. Neurosci.* **20**, 960–968 (2017).

192. Chang, R. B., Strochlic, D. E., Williams, E. K., Umans, B. D. & Liberles, S. D. Vagal Sensory Neuron Subtypes that Differentially Control Breathing. *Cell* **161**, 622–633 (2015).

193. Williams, E. K. *et al.* Sensory Neurons that Detect Stretch and Nutrients in the Digestive System. *Cell* **166**, 209–221 (2016).

194.Bruegmann, T. *et al.* Optogenetic control of heart muscle in vitro and in vivo. *Nat. Methods* **7**, 897–900 (2010).

195.Guettier, J.-M. *et al.* A chemical-genetic approach to study G protein regulation of β cell function in vivo. *Proc. Natl. Acad. Sci.* **106**, 19197–19202 (2009).

196.Jain, S. *et al.* Chronic activation of a designer G_q-coupled receptor improves β cell function. *J. Clin. Invest.* **123**, 1750–1762 (2013).

197.Hua Li, J. *et al.* A Novel Experimental Strategy to Assess the Metabolic Effects of Selective Activation of a Gq-Coupled Receptor in Hepatocytes In Vivo. *Endocrinology* **154**, 3539–3551 (2013).

198.Gradinaru, V. Overriding sleep. *Science* **358**, 457–457 (2017).

199.Robinson, J. E. & Gradinaru, V. Dopaminergic dysfunction in neurodevelopmental disorders: recent advances and synergistic technologies to aid basic research. *Curr. Opin. Neurobiol.* **48**, 17–29 (2018).

200.Ran, F. A. *et al.* In vivo genome editing using Staphylococcus aureus Cas9. *Nature* **520**, 186–191 (2015).

201.Yang, Y. *et al.* A dual AAV system enables the Cas9-mediated correction of a metabolic liver disease in newborn mice. *Nat. Biotechnol.* **34**, 334–338 (2016).

202.Senís, E. *et al.* CRISPR/Cas9-mediated genome engineering: An adeno-associated viral (AAV) vector toolbox. *Biotechnol. J.* **9**, 1402–1412 (2014).

203.Yang, Q. *et al.* AAV-based shRNA silencing of NF-κB ameliorates muscle pathologies in mdx mice. *Gene Ther.* **19**, 1196–1204 (2012).

204.Kotterman, M. A. & Schaffer, D. V. Engineering adeno-associated viruses for clinical gene therapy. *Nat. Rev. Genet.* **15**, 445–451 (2014).

205. Xu, R. *et al.* Quantitative comparison of expression with adeno-associated virus (AAV-2) brain-specific gene cassettes. *Gene Ther.* **8**, 1323–1332 (2001).

206. Chamberlain, K., Riyad, J. M. & Weber, T. Expressing Transgenes That Exceed the Packaging Capacity of Adeno-Associated Virus Capsids. *Hum. Gene Ther. Methods* **27**, 1–12 (2016).

207. Broderick, J. A. & Zamore, P. D. MicroRNA therapeutics. *Gene Ther.* **18**, 1104–1110 (2011).

208. Nayak, S. & Herzog, R. W. Progress and prospects: immune responses to viral vectors. *Gene Ther.* **17**, 295–304 (2010).

209. Gao, K. *et al.* Empty virions in AAV8 vector preparations reduce transduction efficiency and may cause total viral particle dose-limiting side effects. *Mol. Ther. - Methods Clin. Dev.* **1**, 9 (2014).

210. Mingozzi, F. & High, K. A. Immune responses to AAV vectors: overcoming barriers to successful gene therapy. *Blood* **122**, 23–36 (2013).

211. Strobel, B., Miller, F. D., Rist, W. & Lamla, T. Comparative Analysis of Cesium Chloride- and Iodixanol-Based Purification of Recombinant Adeno-Associated Viral Vectors for Preclinical Applications. *Hum. Gene Ther. Methods* **26**, 147–157 (2015).

212. Gray, S. J. *et al.* Preclinical Differences of Intravascular AAV9 Delivery to Neurons and Glia: A Comparative Study of Adult Mice and Nonhuman Primates. *Mol. Ther.* **19**, 1058–1069 (2011).

213. Chakrabarty, P. *et al.* Capsid Serotype and Timing of Injection Determines AAV Transduction in the Neonatal Mice Brain. *PLOS ONE* **8**, e67680 (2013).

214. Maguire, C. A. *et al.* Mouse Gender Influences Brain Transduction by Intravascularly Administered AAV9. *Mol. Ther.* **21**, 1470–1471 (2013).

215. Chen, Y. H., Chang, M. & Davidson, B. L. Molecular signatures of disease brain endothelia provide new sites for CNS-directed enzyme therapy. *Nat. Med.* **15**, 1215–1218 (2009).

216. Franks, K. M. *et al.* Recurrent Circuitry Dynamically Shapes the Activation of Piriform Cortex. *Neuron* **72**, 49–56 (2011).

217. Resendez, S. L. *et al.* Visualization of cortical, subcortical and deep brain neural circuit dynamics during naturalistic mammalian behavior with head-mounted microscopes and chronically implanted lenses. *Nat. Protoc.* **11**, 566–597 (2016).

218. Treweek, J. B. *et al.* Whole-body tissue stabilization and selective extractions via tissue-hydrogel hybrids for high-resolution intact circuit mapping and phenotyping. *Nat. Protoc.* **10**, 1860–1896 (2015).

219. Mahmood, T. & Yang, P.-C. Western Blot: Technique, Theory, and Trouble Shooting. *North Am. J. Med. Sci.* **4**, 429–434 (2012).

220. Greenbaum, A., Jang, M. J., Challis, C. & Gradinaru, V. Q&A: How can advances in tissue clearing and optogenetics contribute to our understanding of normal and diseased biology? *BMC Biol.* **15**, 87 (2017).

221. Richardson, D. S. & Lichtman, J. W. Clarifying Tissue Clearing. *Cell* **162**, 246–257 (2015).

222. Gradinaru, V., Treweek, J., Overton, K. & Deisseroth, K. Hydrogel-Tissue Chemistry: Principles and Applications. *Annu. Rev. Biophys.* **47**, 355–376 (2018).

223. Treweek, J. B. & Gradinaru, V. Extracting structural and functional features of widely distributed biological circuits with single cell resolution via tissue clearing and delivery vectors. *Curr. Opin. Biotechnol.* **40**, 193–207 (2016).

224. Day, R. N. & Davidson, M. W. The fluorescent protein palette: tools for cellular imaging. *Chem. Soc. Rev.* **38**, 2887–2921 (2009).

225. Gray, D. A. & Woulfe, J. Lipofuscin and Aging: A Matter of Toxic Waste. *Sci. Aging Knowl. Environ.* **2005**, re1 (2005).

226. Kupferschmidt, D. A., Cody, P. A., Lovinger, D. M. & Davis, M. I. Brain BLAQ: Post-hoc thick-section histochemistry for localizing optogenetic constructs in neurons and their distal terminals. *Front. Neuroanat.* **9**, (2015).

227. Petri, K. *et al.* Comparative next-generation sequencing of adeno-associated virus inverted terminal repeats. *BioTechniques* **56**, 269–273 (2014).

228. Masters, J. R. & Stacey, G. N. Changing medium and passaging cell lines. *Nat. Protoc.* **2**, 2276–2284 (2007).

229. Anesthesia Induction and Maintenance | Protocol. https://www.jove.com/science-education/10263/anesthesia-induction-and-maintenance.

230. Huang, X. *et al.* AAV2 production with optimized N/P ratio and PEI-mediated transfection results in low toxicity and high titer for in vitro and in vivo applications. *J. Virol. Methods* **193**, 270–277 (2013).

231. Lock, M. *et al.* Rapid, Simple, and Versatile Manufacturing of Recombinant Adeno-Associated Viral Vectors at Scale. *Hum. Gene Ther.* **21**, 1259–1271 (2010).

232. Gage, G. J., Kipke, D. R. & Shain, W. Whole Animal Perfusion Fixation for Rodents. *JoVE J. Vis. Exp.* e3564 (2012) doi:10.3791/3564.

233. Park, J.-J. & Cunningham, M. G. Thin Sectioning of Slice Preparations for Immunohistochemistry. *JoVE J. Vis. Exp.* e194 (2007) doi:10.3791/194.

234. Iulianella, A. Cutting Thick Sections Using a Vibratome. *Cold Spring Harb. Protoc.* **2017**, pdb.prot094011 (2017).

235. Wu, Z., Asokan, A. & Samulski, R. J. Adeno-associated virus serotypes: vector toolkit for human gene therapy. *Mol. Ther. J. Am. Soc. Gene Ther.* **14**, 316–327 (2006).

236. Daya, S. & Berns, K. I. Gene Therapy Using Adeno-Associated Virus Vectors. *Clin. Microbiol. Rev.* **21**, 583–593 (2008).

237. Gaj, T., Epstein, B. E. & Schaffer, D. V. Genome Engineering Using Adeno-associated Virus: Basic and Clinical Research Applications. *Mol. Ther.* **24**, 458–464 (2016).

238. Bartlett, J. S., Kleinschmidt, J., Boucher, R. C. & Samulski, R. J. Targeted adeno-associated virus vector transduction of nonpermissive cells mediated by a bispecific F(ab'gamma)2 antibody. *Nat. Biotechnol.* **17**, 181–186 (1999).

239. Davidsson, M. *et al.* A systematic capsid evolution approach performed in vivo for the design of AAV vectors with tailored properties and tropism. *Proc. Natl. Acad. Sci.* **116**, 27053–27062 (2019).

240. Grimm, D. *et al.* In vitro and in vivo gene therapy vector evolution via multispecies interbreeding and retargeting of adeno-associated viruses. *J. Virol.* **82**, 5887–5911 (2008).

241. Ying, Y. *et al.* Heart-targeted adeno-associated viral vectors selected by in vivo biopanning of a random viral display peptide library. *Gene Ther.* **17**, 980–990 (2010).

242. Müller, O. J. *et al.* Random peptide libraries displayed on adeno-associated virus to select for targeted gene therapy vectors. *Nat. Biotechnol.* **21**, 1040–1046 (2003).

243. Pekrun, K. *et al.* Using a barcoded AAV capsid library to select for clinically relevant gene therapy vectors. *JCI Insight* **4**, (2019).

244.Tordo, J. *et al.* A novel adeno-associated virus capsid with enhanced neurotropism corrects a lysosomal transmembrane enzyme deficiency. *Brain* **141**, 2014–2031 (2018).

245.Ojala, D. S. *et al.* In Vivo Selection of a Computationally Designed SCHEMA AAV Library Yields a Novel Variant for Infection of Adult Neural Stem Cells in the SVZ. *Mol. Ther. J. Am. Soc. Gene Ther.* **26**, 304–319 (2018).

246.Kisanuki, Y. Y. *et al.* Tie2-Cre transgenic mice: a new model for endothelial cell-lineage analysis in vivo. *Dev. Biol.* **230**, 230–242 (2001).

247.Sweeney, M. D., Zhao, Z., Montagne, A., Nelson, A. R. & Zlokovic, B. V. Blood-Brain Barrier: From Physiology to Disease and Back. *Physiol. Rev.* **99**, 21–78 (2019).

248.Betley, J. N. & Sternson, S. M. Adeno-Associated Viral Vectors for Mapping, Monitoring, and Manipulating Neural Circuits. *Hum. Gene Ther.* **22**, 669–677 (2011).

249.Sweeney, M. D., Sagare, A. P. & Zlokovic, B. V. Blood-brain barrier breakdown in Alzheimer disease and other neurodegenerative disorders. *Nat. Rev. Neurol.* **14**, 133–150 (2018).

250.Hwang, B., Lee, J. H. & Bang, D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp. Mol. Med.* **50**, 96 (2018).

251.Challis, R. C. *et al.* Systemic AAV vectors for widespread and targeted gene delivery in rodents. *Nat. Protoc.* 1 (2019) doi:10.1038/s41596-018-0097-3.

252.Paulk, N. K. *et al.* Bioengineered Viral Platform for Intramuscular Passive Vaccine Delivery to Human Skeletal Muscle. *Mol. Ther. - Methods Clin. Dev.* **10**, 144–155 (2018).

253.Schmit, P. F. *et al.* Cross-Packaging and Capsid Mosaic Formation in Multiplexed AAV Libraries. *Mol. Ther. - Methods Clin. Dev.* **17**, 107–121 (2020).

254. Koerber, J. T., Jang, J.-H. & Schaffer, D. V. DNA Shuffling of Adeno-associated Virus Yields Functionally Diverse Viral Progeny. *Mol. Ther.* **16**, 1703–1709 (2008).

255. Bosley, A. D. & Ostermeier, M. Mathematical expressions useful in the construction, description and evaluation of protein libraries. *Biomol. Eng.* **22**, 57–61 (2005).

256. Harris, J. A. *et al.* Anatomical characterization of Cre driver mice for neural circuit mapping and manipulation. *Front. Neural Circuits* **8**, 76 (2014).

257. 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).

258. Zhu, Y. *et al.* Ablation of NF1 function in neurons induces abnormal development of cerebral cortex and reactive gliosis in the brain. *Genes Dev.* **15**, 859–876 (2001).

259. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nat. Neurosci.* **13**, 133–140 (2010).

260. Choi, H. M. T. *et al.* Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).

261. Renier, N. *et al.* iDISCO: A Simple, Rapid Method to Immunolabel Large Tissue Samples for Volume Imaging. *Cell* **159**, 896–910 (2014).

262. Dodt, H.-U. *et al.* Ultramicroscopy: development and outlook. *Neurophotonics* **2**, 041407 (2015).

263. Bria, A. & Iannello, G. TeraStitcher - a tool for fast automatic 3D-stitching of teravoxel-sized microscopy images. *BMC Bioinformatics* **13**, 316 (2012).

264. DiMattia, M. A. *et al.* Structural Insight into the Unique Properties of Adeno-Associated Virus Serotype 9. *J. Virol.* **86**, 6947–6958 (2012).

265.Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).

266.Schneider, T. D. & Stephens, R. M. Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.* **18**, 6097–6100 (1990).

267.Crooks, G. E. WebLogo: A Sequence Logo Generator. *Genome Res.* **14**, 1188–1190 (2004).

268.Powell, S. K. *et al.* Characterization of a novel adeno-associated viral vector with preferential oligodendrocyte tropism. *Gene Ther.* **23**, 807–814 (2016).

269.Cucchiarini, M., Ren, X. L., Perides, G. & Terwilliger, E. F. Selective gene expression in brain microglia mediated via adeno-associated virus type 2 and type 5 vectors. *Gene Ther.* **10**, 657–667 (2003).

270.Rajendran, P. S. *et al.* Identification of peripheral neural circuits that regulate heart rate using optogenetic and viral vector strategies. *Nat. Commun.* **10**, 1–13 (2019).

271.Batista, A. R. *et al. Ly6a* Differential Expression in Blood–Brain Barrier Is Responsible for Strain Specific Central Nervous System Transduction Profile of AAV-PHP.B. *Hum. Gene Ther.* **31**, 90–102 (2020).

272.Seo, J. W. *et al.* Positron emission tomography imaging of novel AAV capsids maps rapid brain accumulation. *Nat. Commun.* **11**, 1–13 (2020).

273.Yang, K. K., Wu, Z. & Arnold, F. H. Machine learning in protein engineering. *ArXiv181110775 Q-Bio* (2018).

274.Wu, Z., Kan, S. B. J., Lewis, R. D., Wittmann, B. J. & Arnold, F. H. Machine learning-assisted directed protein evolution with combinatorial libraries. *Proc. Natl. Acad. Sci.* **116**, 8852–8858 (2019).

275. Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. *Nature* **577**, 706–710 (2020).

276. György, B. *et al.* Gene Transfer with AAV9-PHP.B Rescues Hearing in a Mouse Model of Usher Syndrome 3A and Transduces Hair Cells in a Non-human Primate. *Mol. Ther. - Methods Clin. Dev.* **13**, 1–13 (2019).

277. Lim, J.-A. *et al.* Intravenous Injection of an AAV-PHP.B Vector Encoding Human Acid α-Glucosidase Rescues Both Muscle and CNS Defects in Murine Pompe Disease. *Mol. Ther. Methods Clin. Dev.* **12**, 233–245 (2019).

278. Chen, M. B. *et al.* Brain Endothelial Cells Are Exquisite Sensors of Age-Related Circulatory Cues. *Cell Rep.* **30**, 4418-4432.e4 (2020).

279. Wolburg, H., Noell, S., Mack, A., Wolburg-Buchholz, K. & Fallier-Becker, P. Brain endothelial cells and the glio-vascular complex. *Cell Tissue Res.* **335**, 75–96 (2009).

280. Munji, R. N. *et al.* Profiling the mouse brain endothelial transcriptome in health and disease models reveals a core blood–brain barrier dysfunction module. *Nat. Neurosci.* **22**, 1892–1902 (2019).